



Huang, Jingming (2023) *Deep visual generation for automotive design upgrading and market optimising*. PhD thesis.

<https://theses.gla.ac.uk/83877/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Deep Visual Generation for Automotive Design Upgrading and Market Optimising

JINGMIN HUANG

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow



University
of Glasgow

SEPTEMBER, 2023

Abstract

The rising levels of homogeneity of modern cars in terms of price and functions has made exterior styling increasingly vital for market success. Recently, researchers have attempted to apply deep learning, especially deep generative models, to automotive exterior design, which has enabled machines to deliver diverse novel designs from large-scale data. In this thesis, we argue that recent advancements in deep learning techniques, particularly in deep generation, can be utilised to facilitate different aspects of automotive exterior design, including design generation, evaluation, and market profit predicting. We conducted three independent studies, each providing tailored solutions to specific automotive design scenarios. These include: a study focused on adapting the latest deep generative model to achieve regional modifications in existing designs, and evaluating these adjustments in terms of design aesthetic and prospective profit changes; another study dedicated to developing a predictive model to assess the modernity of existing designs regarding the future fashion trends; and a final study aiming to incorporate the distinctive shape characteristics of a cheetah into the side view designs of cars. This thesis has four main contributions. First, the developed DVM-CAR dataset is the first large-scale automotive dataset containing designs and marketing data over 10 years. It can be used for different types of research needs from multiple disciplines. Second, given the inherent constraints in automotive design, such as the need to maintain “family face”, and the fact that unconstrained design generation can be seen as a special form of regional modification, our research distinctively focuses on the regional modifications to existing designs, a departure from existing studies. Third, our studies are the first works that integrate the design modules with market profit optimisation. This reforms the traditional product design optimisation frameworks by replacing the abridged design profiles with graphical designs. Finally, the proposed data-driven measures offer effective approaches for automotive aesthetic evaluation and market forecasting, including approaches that can make assessments from a dynamic perspective by examining the evolving fashion trends.

Acknowledgements

When the day finally came, it felt more surreal than I had anticipated. Looking back, my Ph.D. journey was challenging, with moments of doubt where I had often imagined the joy of graduation. Completing my Ph.D. holds a great meaning for me, and I'm deeply grateful to those who supported and guided me throughout the journey.

Thank you to my supervisory team and co-authors for your invaluable guidance and support throughout my doctoral studies. I feel truly fortunate to have received what I perceive to be the best doctoral training for me. This experience has shaped, and will continue to influence, both my research and my life.

Thank you, Adam Smith Business School and School of Computing Science. I am deeply grateful for the scholarship that I've received, which enabled me to complete my Ph.D. I also appreciate the assistance and support from the faculty and administrative staff over the years.

Thank you, my dear friends. During the Ph.D. journey, there are many moments of loneliness and frustration. It's hard to imagine how I would have gotten through those challenging and disheartening times without your support and encouragement.

Thank you, my family. You have consistently provided me with a stable foundation, allowing me to focus on my Ph.D. studies. Thank you, my son, Mumu. Your arrival has been the greatest gift of my life thus far.

Lastly, Bowei, while our culture implies words can't fully express deep gratitude, I still want to sincerely thank you for all the guidance and support that you provided.

Contents

1	Introduction	1
1.1	Motivations	2
1.2	Research Hypothesis	4
1.3	Contributions	4
1.4	Framework Overview	6
1.5	Source Publications	8
1.6	Thesis Outline	8
2	Background	11
2.1	Automotive Exterior Styling	11
2.2	Preliminaries of Deep Learning	18
2.3	Deep Learning for Automotive Styling	30
2.4	Summary	32
3	DVM-CAR: An Automotive Database for Deep Visual Marketing	33
3.1	Introduction	33
3.2	Data Collection and Preparation	35
3.3	Dataset Description	38
3.4	Quality Control: Towards a Good Research Dataset	38
3.5	Dataset Application Examples	42
3.5.1	Understanding Automotive Exterior Aesthetics Design	42
3.5.2	AI-Powered Automotive Exterior Design	43
3.5.3	Visual-Based Used Car Pricing	45
3.6	Summary	47

4	GEO: Integrate Deep Generation with Profit Optimisation	48
4.1	Introduction	48
4.2	The GEO Framework	51
4.2.1	Task Definition and Framework Overview	51
4.2.2	The Design Generator	52
4.2.3	The Design Evaluator	56
4.2.4	The Decision Optimiser	57
4.3	Experiments	58
4.3.1	Datasets	60
4.3.2	Experimental Settings	60
4.3.3	Analysis of Results	61
4.4	Summary	67
5	Trendiness: Design for Upcoming Fashion	70
5.1	Introduction	70
5.2	Problem Statement	73
5.3	Proposed Methods for Design Evaluation and Design Assistant	75
5.3.1	Stage I: Trendiness Evaluator	75
5.3.2	Stage II: Antiquated Design Analytics	78
5.3.3	Stage III: Image Completion-Based Design Upgrading	80
5.3.4	Stage IV: Ranking of Candidate Designs	82
5.4	Data and Experiment Settings	83
5.5	Results and Discussion	84
5.5.1	Validation of Trendiness Measure	84
5.5.2	Interpretation of Upgrading Results	87
5.5.3	Discussion	88
5.6	Summary	91
6	Animism: Design with Bio-Inspired Novelty	92
6.1	Introduction	92
6.2	Methodology	95
6.2.1	Running Cheetah Image Processing	96
6.2.2	Car Side-View Image Processing	99
6.2.3	Bio-Inspired Car Side-View Design Generation	102
6.3	Experimental Settings	105

6.3.1	Key Steps of the Targeted Saloon Models	105
6.3.2	Using of Sobel Edge Detection	105
6.3.3	Optimal SMM Selection	106
6.4	Empirical Investigation	108
6.4.1	Pretest of Schema Activation	109
6.4.2	Bio-Inspired Aesthetics Design	111
6.5	Summary	117
7	Conclusions and Future Work	118
7.1	Conclusion and Contributions	118
7.2	Directions for Future Work	119
7.3	Potential Impact and Transferability	120

List of Figures

1.4.1 Overview of the automotive design upgrading framework.	6
2.2.1 Structure of a two layer MLP.	19
2.2.2 Schematic view of using the Mask R-CNN to segment car body from its original image.	28
3.2.1 Illustration of car images in our dataset: (a) angles of car image labelling; (b) example of background removal for Audi A2 from angle 270 degree; (c) example of background removal for Audi A2 from angle 45 degree.	36
3.2.2 Schematic view of the DVM-CAR dataset, which consists of two parts: image data and table data. Image and table data can be joined or connected via model, adv and image IDs.	36
3.2.3 Summary of image data in the DVM-CAR dataset: (a) the number of car models across years and the average images for each model; (b) the annual percentage composition of images by body type.	37
3.2.4 Four categories of variables in our dataset.	37
3.4.1 Categorisation of data issues reported by participating business and computer science researchers.	41
3.5.1 Illustration of application examples: (a) modernity scores of Land Rover models registered in 2015; (b) distribution of the predicted modernity scores of car models from 2000 to 2015; (c) new BMW front designs generated by CycleGAN according to given facial shapes; (d) automobile design sketches generated by CycleGAN.	44
3.5.2 Simulated depletion surface of residual values.	46
4.1.1 Examples of car front facelifts (designs before and after the facelifts) for Audi A3 and Lexus NX.	49

4.2.1 Schematic view of the GEO framework.	53
4.2.2 Structure of the generator in StyleGAN2. The StyleGAN/StyleGAN2 models comprise two main parts, the mapping network and the synthesis network. The former transforms the raw latent code to the intermediate latent code, while the latter generates images/designs according to the given intermediate latent codes.	54
4.2.3 Targeted car models for facelift exterior design: the yellow masks highlight the areas to be modified by our proposed framework; the first row shows the car model name; the second row shows the car model type; the number in the third row shows the total sales of the car model from 2007 to 2017 and the number in the brackets shows the car model's market share (i.e., the division of the model's sales to the total sales of the model's segment market); the fourth row shows the critiques of the car model's exterior design; the fifth row provides the reference (i.e., URL) of the consumer reviews.	59
4.3.1 Comparison of candidate latent spaces, where the superscript indicates the latent layer position among the mapping network.	62
4.3.2 Comparison of car front designs of replacing different generation block latent codes, where a column is the replacement starting block and a row is the replacement end block.	64
4.3.3 Empirical examples of car front aesthetic scores predicted by the trained evaluator.	65
4.3.4 Analysis of the aesthetic scores: (left) Time series plot of the average aesthetic score and the sales of the targeted car models from 2008 to 2017; (right) Aesthetic score VS market shares . By comparing the model's market shares and their aesthetic scores in recent years, we find that the market share is correlated to their aesthetic levels, where the Pearson test results are as follows: $\rho = 0.39$ with a p -value=0.009.	65
4.3.5 Effects of facelift : (left) mid-term market share evolution when adopting different facelift years. (right) market share simulation according to aesthetic changes.	68
5.2.1 Illustrations of aesthetic space, fashion evolution and the proposed trendiness measure. Dots represent the sold car designs in the market. The three dimensions are named PCs 1-3 to analogise the principal components of the principal component analysis.	74
5.3.1 Schematic view of the proposed design assistant approach - (i) Trendiness Evaluator: a CNN-based ordinal regression for trendiness prediction; and (ii) Design Upgrader: image completion-based region-directed design upgrading.	76

5.3.2 An example of antiquated design recognition: (a) the original image; (b) the recognised semantic area; (c) the saliency map from the Guided Backpropagation; (d) the saliency map after thresholding and blur; (e) the balanced saliency map; (f) the image with covering mask.	78
5.5.1 Illustration of identified antiquated regions. The green masks indicated the semantic areas that are highlighted by the saliency methods. Note that the demonstrated semantic areas are shaped by the Mask R-CNN.	87
5.5.2 Empirical examples of three updated car models: Lexus IS 2016, Land Rover Range Rover 2017, SEAT Toledo 2017. As the changes on Lexus IS demonstrate, the changes on the main grill would alter the brand's family faces, which is problematic for real application. Nevertheless, these samples demonstrate the modifications achieved via the image completion algorithm.	89
6.1.1 Side-view design comparison for Audi A4 2018 (left) vs 2019 (right) facelift.	93
6.2.1 Schematic view of the proposed bio-inspired design framework.	95
6.2.2 Schematic view of the running cheetah image processing module.	96
6.2.3 Illustration of HED architecture for the running cheetah edge detection.	97
6.2.4 The most representative motions of the running cheetah.	99
6.2.5 Schematic view of the car image processing module.	100
6.2.6 Schematic view of the new car side design generation module.	102
6.2.7 Illustration of CycleGAN structure.	104
6.3.1 Summary of the key steps of the proposed design framework.	106
6.3.2 Robustness check of the ranking for frames in "1410 to 1440".	108
6.4.1 Example of the control condition in the survey.	109
6.4.2 Example of the cheetah condition in the survey.	110
6.4.3 Users' responses to aesthetics design, where ** represents the p value less than 0.01.	110
6.4.4 Illustration of new designs of BMW 3 Series generated by the proposed framework.	112
6.4.5 Example of design selection for BMW 3 Series in the user survey study.	113
6.4.6 Comparison of original and new designs for BMW 3 Series, where the blue box highlights the shoulder curve and the red box highlights the waist curve; the width (W), the height (H) and the central position (C) of each box in the car side-view image are presented.	114
6.4.7 Original and new designs of the targeted premium saloons.	115
6.4.8 Original and new designs of the targeted mainstream saloons.	116

List of Tables

- 2.2.1 Notation list for MLP introduction. 20
- 2.2.2 Notation list for CNN introduction. 25
- 2.2.3 Notation list for GAN introduction. 29
- 2.3.1 The list of key literature for this thesis. 30

- 3.3.1 Description of table data in the DVM-CAR dataset. 39
- 3.3.2 Summary of car specification and sales variables included in the related market-
ing studies and our proposed DVM-CAR dataset. 40
- 3.5.1 Withdrawn chances within the next three years of car groups with different
modernity scores. 45

- 4.3.1 Summary of the used datasets. 60
- 4.3.2 Comparison of our design upgrading algorithm with existing methods. 63
- 4.3.3 Examples of optimised designs. 66
- 4.3.4 Number of facelifts released in ten years. 67

- 5.5.1 Trendiness VS market share changes. 86
- 5.5.2 Comparison of upgrading on semantic/non-semantic areas. 88

- 6.3.1 Selection of the SMM. 107

Chapter 1

Introduction

With the development of the automotive industry, cars are becoming more homogeneous in terms of price and functions. Consequently, the exterior styling and aesthetic design plays a more determining role for market success. The exterior styling in particular has been reported as the most important motivation in purchasing mainstream and luxury automotive brands (NADA, 2016).

In the past decades, marketing researchers have employed a number of quantitative methods to answer some key questions regarding product design: What are the visual attributes of the product that determine sales success (Bloch, 1995; Creusen and Schoormans, 2005; Krishna et al., 2016)? How can these design attributes influence consumer choice (Jia et al., 2015; Hung and Chen, 2012)? Are there insights that can be used to design the appearance of a product or an advertisement containing specific visual messages such as aesthetic or symbolic values (DiSalvo and Gemperle, 2003; Jindal et al., 2016; Rubera, 2015)? How should one search for the optimal market positions for new products that maximise profits (Rajeev and Krishnamurti, 1987; Kaul and Rao, 1995; Shi et al., 2001)? Although the mainstream publications have focused on statistical models, such models alone are incapable of dealing with large-scale unstructured data such as images and video. Some of the major concerns around the existing works include the definition of the visual aspects of products, the evaluation of presented designs, and the lack of actual design plans in product design optimisation (Aggarwal and McGill, 2007; Heisley and Levy, 1991; Landwehr et al., 2011b; Maeng and Aggarwal, 2017; Wedel and Pieters, 2008).

Motivated by the recent advancements of machine learning in visual content processing and generation, a group of design researchers applied deep learning technologies to the exterior

design of products, with a primary focus on automotive styling during the conceptual design phase (Burnap et al., 2016b; Pan et al., 2017; Burnap et al., 2021). This phase is the initial stage of new car model development, typically without considering engineering constraints or challenges (Burnap et al., 2016a).

The conventional methods rely heavily on human-defined features, which require time and vast domain knowledge. Deep learning approaches work by quantifying digital images and automatically learning valuable features from the data, and they are thus capable of generating realistic-looking images or reaching semantic understanding (LeCun et al., 2015). Studies have affirmed that these learned feature detectors perform well on human subjective experience-related tasks, such as aesthetic evaluation (Lu et al., 2015; Segalin et al., 2017), and can produce highly realistic designs (Zhu et al., 2017; Karras et al., 2020).

Being greatly motivated by the advancements of deep learning on visual content understanding and generation, in the present PhD thesis we focus on applying deep learning algorithms to automotive exterior design and analytics. In particular, it provides upgraded designs where design constraints are considered and integrates deep generation into product design optimisation frameworks.

1.1 Motivations

The automotive industry is a critical sector of the modern global economy and accounts for 3.65% of world GDP (Saberi, 2018). As a product that has emerged for over a hundred years, cars are becoming more homogeneous in price and function, which makes their design, especially the exterior design, a more important determinant of market share (Jindal et al., 2016). The exterior styling in particular has been identified as one of the determinative factors of consumer choice in the automotive market (NADA, 2016).

Consumer purchase behaviours are widely recognised to be frequently driven by perceptual stimuli (Scott, 2000), and among various product marketing strategies, exterior design is one of the most direct and effective. The importance of product appearance has long been acknowledged by marketers and researchers (Sherman and Hoffer, 1971; Bloch, 1995). Numerous examples illustrate how manufacturers have achieved market success through outstanding exterior designs. For instance, Apple, as one of the most successful corporations globally, continues to dominate the market by creating products with unique yet concise exterior styles (Shelley, 2015). Despite the critical role of exterior styling in marketing, the related research is surprisingly scant in comparison to other influential factors. This research gap is largely due to the

lack of automated techniques for processing unstructured data types such as images or videos, which significantly hampers progress in the field. Traditionally, handling visual content has demanded the acquisition of task-relevant pattern detectors through feature engineering (LeCun et al., 2015), while the creation of innovative exterior designs has often been seen as the exclusive domain of human creativity (Sbai et al., 2019).

In recent years, however, a group of pioneering studies have employed the new deep learning methods in automotive design (Burnap et al., 2016c; Pan et al., 2017; Burnap et al., 2021). Unlike the traditional approach, predictive models based on convolutional neural networks (CNNs) (LeCun et al., 2015) can, after being trained with appropriate data, make end-to-end evaluations of exterior designs regarding different aesthetic attributes (Pan et al., 2016). Moreover, the deep generative models can be employed to obtain various realistic-looking design images (Burnap et al., 2016b).

However, these pioneering works have not taken into account the constraints of designs, which require the developed designs to retain specific appearances (Burnap et al., 2016a). For instance, modern automakers adopt similar visual design cues, called family face, to their models for branding purposes. Thus, the designs should preserve these family faces. Moreover, as the complete redesign of an existing car model costs billions of dollars, automakers often adopt mid-generational refreshes (i.e., facelifts) to reboot sales, where only regional changes are allowed on existing designs (Blonigen et al., 2013; Greim, 2017). From the technical perspective, such a regional modification is more challenging than a novel creation, since the latter can be seen as a special case of the former in which all areas can be modified (Bertalmio et al., 2000, 2003). In other words, generative algorithms, which are capable of generating novel designs and modifying designs regionally, are needed for automotive exterior design.

Furthermore, as existing studies have primarily been design-focused, none has connected automotive aesthetic design with the concept of product design optimisation. Studies of product design optimisation focus on developing algorithmic frameworks that can infer optimal design profiles for maximising profit, but no real designs are offered (Rajeev and Krishnamurti, 1987; Kaul and Rao, 1995; Nair et al., 1995; Shi et al., 2001). Consequently, the complexity of real scenarios is underestimated. For example, evaluating the aesthetic attributes of a product is considered straightforward and precise for design profiles. In real cases, however, the evaluation of candidate designs, termed “theme clinics” (Pan et al., 2017), is a costly and lengthy process. Moreover, these works assume that all ranges of aesthetic attributes are achievable when using design profiles, and so no restrictions are assigned to the probable design space (Nair et al.,

1995; Shi et al., 2001). All these restrictions restrain the industrial impact of design optimisation studies. It would therefore be innovative for a proposed product design optimisation framework to provide vivid designs rather than design profiles.

Given this situation, we apply machine learning methods, particularly deep learning approaches, to regional design upgrading and apply predictive models, such as recurrent neural networks (RNNs) (LeCun et al., 2015), to profit estimation. We aim to develop comprehensive frameworks that deliver photographic upgrade designs while offering optimal marketing solutions. This thesis is driven by four key questions: (i) Can effective generative models be developed to upgrade existing car appearance designs regionally (e.g., headlamps, grills and body curves), incorporating new visual features to improve market performance? (ii) Can machine learning be used effectively to automatically learn from image data (together with other data formats) to determine which designs are superior in aesthetics and what visual attributes are more influential to the design attractiveness? (iii) Can the design generation be incorporated into a profit optimisation framework that can assess novel designs in terms of the potential market profit? (iv) Can methods be developed that incorporate nature-inspired elements (e.g., animal-inspired features) into new car styling to make their appearance more sporty?

1.2 Research Hypothesis

This thesis argues that the recent advancements in deep learning can be used to help solve a series of decision-making problems in automotive styling upgrading and extend the traditional product design optimisation frameworks with design modules. In particular, deep generative models can be deployed to produce realistic-looking automotive designs, and CNNs can be employed for automotive aesthetic evaluation. In addition, dynamic models such as RNNs, supported by historical sales and price data, can predict the potential market gains or losses from adopting a specific design.

1.3 Contributions

As this thesis lies in the overlap between computer science, marketing, and design studies, We worked on some independent studies that develop different solutions for specific automotive design scenarios. These studies are not intended as fundamental methodological contributions to new machine learning algorithms, but rather as application contributions. Each independent study has tailored methodological developments, including new computational frameworks, evaluation metrics and scenario-specified algorithm modification. They contribute to both com-

puter science (primarily) and marketing.

The developed DVM-CAR dataset is the first large-scale automotive dataset that contains design data (images) and marketing data (e.g., car specifications and sales) over 10 years. It has been shared via a specifically developed webpage in the hope of facilitating the studies in relevant fields. As an interdisciplinary dataset, it can be used for different types of analytics, ranging from descriptive to diagnostic, predictive and perspective, by researchers from computer science, marketing and design communities.

These studies provide detailed explorations of practicable methods for regional design modifications, which, within the context of this thesis, concern design upgrading. In contrast to previous works, which focus on generating novel designs without considering constraints, the design upgrading methods allow users to make partial changes to designs while retaining the rest unchanged. These methods tackle challenges in maintaining “family face” and executing “facelift” changes for automotive design generation. Such approaches presents additional technical challenges, as unconstrained generation can be viewed as a subset of constrained generation when no constraints are applied. In total, three generative adversarial networks (GAN) ([Goodfellow et al., 2014](#)) based approaches have been proposed and demonstrated for the regional modifications needs. These approaches also offer valuable directions for future studies in product design generation.

Unlike previous works ([Kong et al., 2016](#); [Ren et al., 2017](#); [Pan et al., 2017](#)), which have primarily adopted subjective ratings as evaluation labels, our studies provide a group of data-driven aesthetic-related measures. In particular, deep models that evaluate the aesthetic from the dynamic perspective are developed by examining the concepts of fashion trends and modernness. Developing such a time-dependent aesthetic measure is a novel approach.

From a marketing perspective, to the best of our knowledge, our studies are the first to integrate design generation, evaluation, and profit optimisation. Furthermore, ours are the initial works to develop frameworks suggesting photographic design schemes instead of merely profiling design attributes. The inclusion of new design modules has made the study framework more complex and yet more practicable to design optimisation. Our works are clear demonstrations of how to employ advanced deep learning algorithms as particular modules in comprehensive frameworks for design optimisation. It is hoped that our studies can serve as groundwork for prospective investigations that adopt deep generative methods in product design optimisation.

1.4 Framework Overview

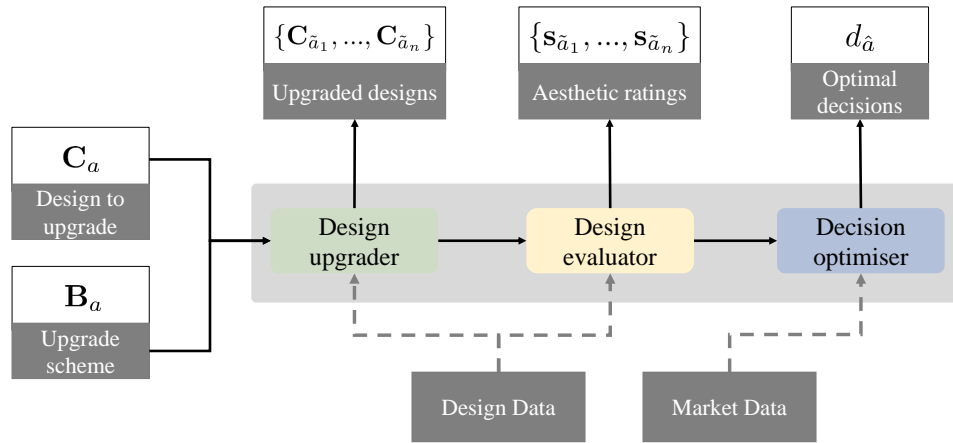


Figure 1.4.1: Overview of the automotive design upgrading framework.

To develop machine learning-based methods that offer diverse upgraded automotive designs and decision support (illustrated in Fig. 1.4.1), this thesis proposes a general product design optimisation framework with five main components: market data, design data, design upgrader, design evaluator and decision optimiser.

Design data refers to automotive exterior images, which are used for training generative or evaluative models. There are two reasons for using static images as design data. First, as most of the advancements in recent deep learning concern images, using static images as design data means many probable computational models. Second, as we aim to provide alternative choices to automakers for conceptual design selection, where sketches are still employed as the mainstream medium, these images can fulfil the evaluation needs with realistic-looking cars.

Market data refers to a collection of variables, including new car price, historical sales and model specifications. These variables are primarily used in decision optimisation settings or sales estimation. As the two most critical variables for marketing or economic studies, over two decades of new car prices and sales are collected, which is sufficient for typical studies. Other factors that influence consumers purchase decisions, such as body type, engine power and gas emissions, are also included.

As its name suggests, the **design upgrader** (or generator) specialises in offering various upgraded schemes. We use the word “upgrading” because, as a refinement of the previous studies, the design constraints are considered in our studies. Only the intended areas/regions of

an existing car design will be modified, while the rest will be left untouched. It is worth noting again that, upgrading is technically more challenging than merely generating, as an upgrading-oriented algorithm can easily apply to novel design generation (i.e., take a large intended area), but not vice versa. To achieve realistic-looking design images, we focus on applying the latest deep generative models, especially the GAN. Functionally, the design upgrader module takes two types of inputs: the candidate design image to upgrade and the corresponding binary mask that indicates the intended areas. It then offers diverse upgraded designs.

The **design evaluator** module evaluates the car designs in terms of aesthetics. Automakers commonly adopt the “theme clinics” periodically to collect subjective ratings for new designs (Burnap et al., 2021). Whenever the design team want external opinions on their new designs, they spend a few weeks hiring participants to collect aesthetic feedback. This is costly in both time and finance. Thus, an automatic module that can provide subjective ratings would be useful to the design team. Over the past decade, CNNs have achieved record-breaking performance on visual discriminative and predictive tasks (LeCun et al., 2015). In particular, they have achieved breaking results on aesthetic prediction tasks (Ren et al., 2017), which were believed to be human endeavours. Hence, in this thesis, CNNs are employed as the end-to-end design evaluator models, providing immediate aesthetic ratings regarding design images. Furthermore, the evaluator is also used to spotlight the vulnerable design features for design upgrading. The gradient-based saliency methods are employed to highlight visual cues that are influential to the objective aesthetic attribute.

The **decision optimiser** is responsible for two objectives: first, estimating the mid- to long-term profits (i.e., the market share shifts) yielded by the given design; second, choosing the market decisions that bring the highest profits. In this thesis, the former is achieved by using the historical sales and evaluation results to train predictive models for estimating the mid- to long-term market share changes. The latter involves exploring the possible action space to specify optimal launch times for the new designs. These two functions are often simplified as one in traditional product design optimisation studies, as designs are represented by design profiles (Green et al., 1981; Rajeev and Krishnamurti, 1987; Shi et al., 2001). As this proposed framework involves more design modules than classical studies, it is overwhelming to include complex settings for a single module; thus, we mainly adopt methods with straightforward settings for decision optimisers.

1.5 Source Publications

Most of the materials presented in chapters 3 to 6 are primarily based on the following published or working studies.

- Chapter 3 DVM-CAR: An Automotive Database for Deep Visual Marketing
Jingming Huang, Bowei Chen, Lan Luo, Shigang Yue, and Iadh Ounis. (2022). "DVM-CAR: A large-scale automotive dataset for visual marketing research and applications." In *Proceedings of IEEE International Conference on Big Data*.
- Chapter 4 GEO: Integrate Deep Generation with Profit Optimisation
Jingming Huang, Bowei Chen, Zhi Yan, Iadh Ounis, and Jun Wang. (2023). "GEO: A computational design framework for automotive exterior facelift." *ACM Transactions on Knowledge Discovery from Data*.
- Chapter 5 Trendiness: Design for Upcoming Fashion
Jingming Huang, Hao Xu, and Bowei Chen. (2023). "Revamping automobile fronts for trendiness." *Working paper*.
- Chapter 6 Animism: Design with Bio-Inspired Novelty
Bowei Chen, Jingming Huang, and Lan Luo. (2023). "Does that car want to give me a ride? Bio-inspired automotive aesthetics design." *Working paper*.

1.6 Thesis Outline

The rest of this thesis is organised as follows:

- The first part of Chapter 2 (Background) reviews the marketing, psychology, design and computer science literature regarding product appearance and car styling. In particular, it (i) reviews how the contemporary market shapes automotive styling characteristics in different eras; (ii) lists the functionalities of product appearance in the market sales; (iii) discusses the psychological mechanism, namely anthropomorphism, involved in automotive design perception; (iv) shows that automotive market success is largely determined by the exterior design; (v) indicates that the increasingly intense competition has compelled automakers to increase their expenditure on exterior design; and (vi) points out a missing point of existing product design optimisation studies in which no graphic designs

are delivered. The second part of the chapter introduces deep learning technologies by explaining the fundamental concepts and classical models for different tasks. In particular, it details the common concepts of the neural network, the classical multilayer perceptron, the convolutional neural network and the evolution of deep generative models. Furthermore, the last part reviews the most relevant studies of the use of deep learning in automotive exterior design.

- Chapter 3 (DVM-CAR Dataset) introduces the first large-scale automotive dataset, which contains 1.4 million images from 899 car models and corresponding specification and sales information from over 10 years in the UK market. The development of this dataset fulfils the growing need for a comprehensive automotive dataset for visual-related marketing research and applications. Moreover, it can be used for different types of analytics, ranging from descriptive to diagnostic, predictive and perspective.

Serving as the groundwork for later studies, this chapter describes how we solve the data difficulty by constructing our own dataset. The developed dataset offers a sound data base for my PhD studies, enabling us to carry out various automotive styling-related studies and reconsider the exterior design problem from the profit optimisation perspective.

- Chapter 4 (GEO) focuses on the facelift of front designs. It extends the automotive exterior design framework to a product design optimisation problem. The proposed novel computational framework provides an end-to-end decision-support solution for automotive designers and manufacturers. With careful selection of the latent space and training examples for the generator, innovative facelift designs for a car front can be generated from its competitors, while maintaining the models branding characteristics. Furthermore, the decision optimiser ensures the selected designs have the maximum predicted revenue before the redesign of the given car model.

This chapter plays a critical role in my PhD, as it divides automotive exterior design into three tasks and embeds profit optimisation into the framework. It seeks to determine (i) how to use deep generative models to obtain partially upgraded automotive designs, (ii) how to assess the effectiveness of a generative model in exterior design upgrading, (iii) how to use a data-driven model to evaluate car aesthetics, and (iv) how to estimate the long-term profit changes caused by exterior design changes.

- Chapter 5 (Trendiness) investigates the deep learning-based automotive design from perspectives not considered in Chapter 4. In particular, it discusses the impacts of modernness

in product styling and offers solutions to upgrade the cars front if it appears dated. The main innovation is to have the system learn, define and score modernness from existing car front images and, in turn, produce unpresented but highly scored modern designs. Additionally, the chapter demonstrates that image-completion algorithms, which predict missing content based on the global semantic content, can be applied as the design upgrader with no feature entanglement problems.

This chapter addresses a number of questions not answered in Chapter 4. It outlines a relatively straightforward framework mainly comprising the design upgrader and design evaluator. Specifically, it has answered the following research questions: (i) How should designed be upgraded regionally without incurring feature entanglement? (ii) Is it possible to compute the fashion trends in the automotive market and score car designs from the fashion evolution perspective? (iii) How should design parts that are perceived as antiquated be identified?

- Chapter 6 (Animism) presents the first study that develops a data-driven framework for bio-inspired automotive exterior design. Building upon previous research on anthropomorphism and schema congruity as a basis for product evaluation, it examines whether presenting a sports sedan as a cheetah (which can run fast) affects consumers perception of the automobile. By integrating CNNs and GANs, we propose a novel computational framework to incorporate body curves derived from a running cheetah into the exterior design of sports sedans.

Unlike the previous chapters, this chapter focuses on upgrading the cars side designs, demonstrating that automotive styling upgrading is not limited to particular viewpoints. Moreover, as design originality is constrained in previous data-driven methods, this work tries to create novel bio-inspired design features and incorporate them into automotive styling. In summary, this chapter seeks to determine (i) how to obtain novel design features that are not sourced from existing designs and whether it is possible to take inspiration from the natural world, and (ii) how to provide upgrade solutions for the side design of cars.

- Chapter 7 concludes the thesis and discusses possible future works.

Chapter 2

Background

This chapter first reviews existing studies regarding automotive exterior styling in different disciplines and then introduces the essential concepts and models in deep learning. Moreover, the most relevant studies, which recently applied deep learning to automotive styling, are also reviewed. The rest of this chapter is organised as the following: Section 2.1 briefly reviews existing literature in marketing and psychology regarding automotive styling. Section 2.2 introduces the essential concepts and knowledge of the most applied deep models. Section 2.3 discuss the most relevant studies of utilising deep learning in automotive exterior design, and Section 2.4 gives a short summary.

2.1 Automotive Exterior Styling

Since the appearance of the first car in human history in 1885, cars have become an essential component of modern life (Happian-Smith, 2001). The automotive industry has nowadays become a crucial sector of the world economy. According to the International Organisation of Motor Vehicle Manufacturers' report, more than 71 million vehicles were produced in the year 2021.¹ The general car industry occupies around 3.65% of the world's GDP, approximately 2.75 trillion Euros (Saber, 2018). The massive economic scale of the automotive industry underscores the significance of researching automotive exterior designs. Through a review of relevant literature, subsequent parts will explore the significance of automotive design, including historical and functional perspectives.

¹<https://www.oica.net>

The Evolution of Automotive Appearance

To start, a brief historical examination of the evolution of automotive styling proposed by [Jaafarnia and Adele \(2011\)](#) is introduced. This history helps illustrate the increasing importance of a car's design to its market success.

According to [Jaafarnia and Adele \(2011\)](#), the history of automotive development can be assorted into the following seven eras based on the car's exterior traits. (1) *Invention era (1885 - 1896)*: this is the very first period that cars have just appeared. In this phase, cars appear in irregular shapes not far from the unplanned combination of engine and wheels. Designers obviously had not taken the exterior styling into consideration at this stage as cars are far from the mature goods for sale. (2) *Innovation era (1897 - 1907)*: compared with the previous era, cars in this era were way more complex in structure. As novel creations under rapid transformation, cars in this era had various looks. As they were still considered goods for the niche market ([Wolf, 1996](#)), aesthetics was still unconsidered for exterior design. (3) *Manufacturing era (1908-1919)*: thanks to Henry Ford's invention of the assembly line for large-scale production, the prices of cars in this era were much decreased. Therefore, making cars more affordable for fairly wealthy families. The structure of cars became more unified in this phase, while it is hard to tell whether this was due to the styling consideration or for the convenience of production. (4) *Capsule era (1920-1930)*: driven by market preference, the most significant characteristic of cars in this era was the enclosing of the body, which separated the inside spaces from the externals. This made the car design even more unified than the previous cars, although the general shapes were still extremely boxy ([O'connell, 1998](#)). The aesthetic had obviously been considered in exterior design. (5) *Classic era (1931-1948)*: contrasted to the last era, the exterior design of cars in this phase was intended to be more aerodynamic, thus containing more large curves and big-sized arches. Furthermore, the entire car bodies were more unified as headlights were mostly integrated into the body. (6) *Integration era (1949-1967)*: the exterior styling had become one of the automakers' most critical considerations, and some automakers even incorporated the exaggerated symbolic design features into the exterior design to achieve marketing success (e.g., the missile design of Cadillac Cyclone). The majority of car bodies were made of the one-piece component ([Down, 2010](#)). Motivated by the fashion trends, the entire shape of cars had become very low and lengthy, which is fairly similar to modern car designs. (7) *Modern era (1968-present)*: the exterior styling is now becoming the most crucial factor for automakers establishing new models. Automakers no longer build cars with exaggerated semantic features but incorporate diverse design philosophies rooted in various sources ([Dant and Martin, 2001](#)),

such as the natural world or wildlife. To sum up, the evolution of car styling shows due to the increasing maturity of the market, modern cars are relatively homogeneous in terms of price and functions, while factors such as exterior styling are becoming determinants for market success.

Why Product Appearance Matter

Although the evolution of automotive exterior design clearly demonstrates continuous industry innovation based on market preferences, academia still lacks a comprehensive understanding of the precise role product appearance plays for consumers. Despite the limited research, some studies, notably Bloch (1995) and Creusen and Schoormans (2005), have attempted to categorise and summarise the functionalities by which product appearance influences consumption. These key functionalities will be outlined in this subsection.

Drawing from Bloch (1995) and Creusen and Schoormans (2005)'s work, five key aspects could be outlined as to how product appearance impacts consumer behaviour. (1) *Aesthetic*: aesthetic value is the most natural function of the product appearance, which refers to the intuitive feeling when consumers see the appearance of products. In many purchase circumstances, the difference in aesthetic perception could determine consumers' final selections between goods (Ahmed et al., 1450). In the long run, as consumers see the durable goods repeatedly, beautiful designs can continuously act as positive stimuli or rewards for consumers, while ill-looking designs serve inverted (Hsu et al., 2018). These short and long-term effects illustrate why aesthetic design is essential for modern marketing. (2) *Attention grabbing*: things with flashy colours or unique forms naturally attract humans. Consumers tend to buy the products that could grab their eyeballs when their involvement is low (Schoormans and Robben, 1997), for instance, buying day-to-day goods such as bread. In such scenarios, attention-grabbing becomes the most critical functionality of product appearance design, especially when the products are mainly sold in supermarkets where thousands of other products are also displayed. (3) *Symbolic value*: being influenced by the public culture, product appearances are often associated with symbolic meanings (Simonson and Schmitt, 1997). For example, in modern design language, flashy colours are associated with youthfulness, sharp-cornered shapes are tagged as masculinity, and roundness designs are considered feminine. As consumers purchase goods with symbolic values that match their self-images, manufacturers tend to develop products in the shape that match target market segments. (4) *Indicating functional traits*: by learning from experience, people are pretty gifted at inferring functional or ergonomic traits of products just by examining the appearance of products (Dawar and Parker, 1994). A vacuum cleaner with a big size in the market would be believed with large power, and toys with more unified designs are

perceived as more solid. As online shopping is becoming extremely popular, the indication of functional traits on appearance is increasingly vital for modern marketing. (5) *Categorisation*: on the grounds that the human cognitive system relies on schema for object recognition, the categorisation of goods is much based on the products' visual cues. The term typicality refers to the degree of similarity toward the most representative design (Veryzer Jr and Hutchinson, 1998). Previous studies show that consumers favour typical-looking goods that have some design novelties (Hekkert et al., 2003). That is, the appearance of products must be designed to facilitate their visual categorisation. Based on these functionalities, exterior designing naturally becomes the key competition for fashion products such as clothes or shoes. As mentioned before, exterior styling is, in particular, vital for the most expensive durable goods in daily use – cars.

Anthropomorphism in Automotive Aesthetic Perceiving

In academic circles, considerable research exists on "Anthropomorphism", a phenomenon where consumers perceive the front end of a car as the face of a living creature. This perspective provides an interesting perspective of how car exterior design influences consumer purchasing decisions. A review of the related literature is provided in the following.

The early findings were from Aggarwal and McGill (2007), where they found when lab participants were asked to read a description depicting the displayed car in the first person, participants showed a clear preference for cars with smiling faces (i.e., with saucer shape grills) than the frown faces. The authors interpreted this as a result of schema congruity, so when people are induced to perceive the car alive, they become more sensitive to angry expressions. In a similar study but without considering the schema part, Landwehr et al. (2011b) investigated the emotional expressions that consumers perceived from different car designs. Their study revealed that emotional expressions could be conveyed by a car's grill and headlights as these parts are semantically associated with the mouth and eyes. Moreover, the perception of friendliness is limited to the grille, while aggressiveness can be communicated with both grille and headlights. Interestingly, they found that consumers did not dislike but actually preferred cars with clear aggressive expressions, which was later confirmed in other literature. As group living creatures, humans are evolved to be extremely skilful for reading facial expressions or inferring from static facial cues. Inspired by evolutionary psychology studies that show humans naturally rate the faces with higher width-to-height ratios as more dominant and aggressive, Maeng and Aggarwal (2017) surveyed people's evaluations of car fronts that had been morphed wider. They also reported that similar to face perceiving, car fronts with larger width-to-height ratios are rated as more dominant and preferred by participants.

Unlike the just mentioned studies that rely on survey or lab studies, [Purucker et al. \(2014\)](#) employed the eye-tracking device to test whether people show eye gaze patterns similar to facial perception when seeing car fronts. They found that consumers would spontaneously notice the aggressive-looking parts and showed avoidance of these areas in the later gazing, which is identical to the ocular moments when humans perceive angry faces. A piece of more robust evidence was indicated by the blood-oxygen measure study on cortex activities. The well-known fusiform face area (FFA) is named for its specification of human face perception. With the fMRI device, [Ku \(2014\)](#) showed the FFA area of participants would also be activated when car front images were displayed to the participants, but not the temporoparietal junction and medial prefrontal areas. Their findings revealed that the anthropomorphism of car faces is a bottom-up procedure that occurs unconsciously.

All the aforementioned studies show that the fronts of cars are perceived as their faces by consumers. It could be an exciting question to know whether this is a market-driven result that the face-like designs could have marketing benefits, but little literature has been found on this.

Automotive Styling for Marketing

Perhaps the most compelling evidence demonstrating the value of car exterior design comes from market surveys, direct viewpoints from the automotive industry, and studies on the impact of appearance on sales. In this subsection, a review of such content will be presented, serving as the most direct evidence of the value of car exterior design.

According to a public talk delivered by the marketing CEO of Audi, the most successful German luxury automobile brand, 60 per cent of its consumer's purchasing decisions are ascribed to the appearance/design of cars ([Kreuzbauer and Malter, 2005](#)). On the other hand, as one of the most influential car media, J.D. Power keeps surveying the key factors affecting consumer decisions. In recent years, the survey responses of tens of thousands of the US new vehicle buyers suggest that exterior styling has counted 57 and 71 per cent as the most critical motivation in purchasing mainstream and luxury auto brands, respectively ([NADA, 2016](#)).

Regarding the sales impact perspective, marketing researchers have a long history of investigating the relationships between automotive aesthetics and market performance. By investigating the effects of style changes on the market shares of several brands in the US, [Sherman and Hoffer \(1971\)](#) found that it is a key demand driver for the producers, and it can be more effective in promoting the market share of smaller producers. Their study manually assigned the design refreshment into four levels, from "trim changes" to "completely new", and operated the regres-

sion analysis on the historical automotive sales from the year 1955 to 1966. They reported that the change of exterior design is a key determinant for market performance shifting, especially in the premium market. Later, in [Hoffer and Reilly \(1984\)](#)'s study, by similarly categorising the degree of styling change into eight grades, the authors employed the paired-difference model on the prior data (from the year 1954 to 1982) to test whether car sales shifted accordingly with appearance changes. It was further confirmed that notable styling changes result in remarkable sales shifts, and notably, the market has continuously become more responsive to design changes in past decades.

As one of the most influential prior studies, [Jindal et al. \(2016\)](#) compared the market shares of cars with a different design emphasis, namely, the form, functional and ergonomic dimensions. They took car reviews from popular websites like J.D. Power to compute car models' ratings on these dimensions and analysed the models' historical market performance. Their results showed car models that have been launched for years but with superior styling could perform much better in market sales than contemporary cars that emphasise functional or ergonomic advantages. It is interpreted that, unlike the functional or ergonomic advantages that need regular updating, the superiors on exterior styling could last for a long time, making the styling emphasised cars succeed in the long competition.

Another stream of studies has attempted to identify and label the specific aesthetic genres of design that determine their performance in the market. [Talke et al. \(2009\)](#) presented that cars with innovative outfits are likely to occupy higher market shares, and the promotion brought by the new designs often takes place very soon after launching. Meanwhile, [Landwehr et al. \(2009\)](#) marked that two particular design factors, namely typicality and complexity, explain 42 per cent of the variation within six months of car sales records from the German Federal Transport Authority. In their 2011 study, [Landwehr et al. \(2011a\)](#) incorporating the typicality and complexity features extracted by computational algorithms into the car sales regression model improved the model's accuracy by 19%.

The Evolution of Fashion Change

As suggested by the history of automotive exterior development, the understanding of car design's aesthetic appeal should be tied to its respective era. With consumer preferences for car appearances undergoing gradual but consistent evolution, no car model can thrive with an unchanged look. This subsection will review the corresponding literature.

In the study by [Blijlevens et al. \(2009\)](#), it is observed that consumers often prefer products

that convey a sense of modernity in their exterior design, as reflected by their frequent use of modernity-related terms to describe product appearances. According to [Moral and Jaumandreu \(2007\)](#), half of the newly launched cars would be dropped off within eight years. After reaching peak market shares, sales volumes tend to decline as products become outdated and lose their appeal. Moreover, [Greim \(2017\)](#) has argued that such a duration has been significantly shortened over the past decade, averaging from eight years to six years.

The short life-cycles intensify the competition between automakers, as new strategies are always needed to keep their market shares. Marketing research showed the superiority of exterior design could slow the deterioration of demands. [Talke et al. \(2009\)](#) found that models with innovative looking can have better sales through their whole life-cycles. [Jindal et al. \(2016\)](#) observed in the long-term, cars with superior exterior designs do much better in terms of sales than corresponding generation models either with better functions or ergonomics.

Hence, to maintain its market share, the development of a new product line or redesign of an existing model is perceived as the main solution with major economic costs ([Ernst et al., 2017](#)). Though these costs are closely guarded, [Blonigen et al. \(2013\)](#) concludes, on average, automakers spend about \$1.25 billion on a critical redesign, while developing a new model costs up to \$6 billion. Due to the cost constraints, automakers can hardly afford to redesign their models oftentimes, which results in the repeated adoption of mid-generational refreshes (that is, partially design upgrading). These mid-generational refreshes allow automakers to upgrade exterior designs while keeping the majority of existing production equipment. Motivated by the above facts, our studies focus on developing design upgrading/regionally modification algorithms.

The Missing of “Designs” in Design Optimisation

In existing research on automotive exterior design, there exists an interesting branch, namely product design optimisation. Within this subset, researchers do not focus on specific designs but rather endeavour to discuss optimisation algorithms for product market positioning, including exterior design. The aim is to enable manufacturers to use these algorithms to identify product designs that yield maximum market profits. In this subsection, we will review a group of articles from this field, as our subsequent study pertains to this domain.

Existing studies in the field of product design optimisation frequently treat design attributes as controllable factors, looking at them from the standpoint of profit optimisation. This line of research can be traced back to the 1980s when [Rajeev and Krishnamurti \(1987\)](#) suggested approaching new product design positioning as a share-of-choice problem, proposing a heuristic

dynamic programming-based algorithm in response. In their study, a developing product could be perceived as a profile with a collection of design attributes, and each attribute has many discrete levels. Consumers with heterogeneous needs would perceive the product with ideal attribute levels as more valuable (higher utility) and thus more likely to buy.

[Kaul and Rao \(1995\)](#) took the lens model as the theoretical base, which considers the observable design cues as critical indications of utilities and consumers making choices based on these perceived attributes. To this end, they proposed a mathematical framework that tries to optimise the potential market profit, taking consumers' heterogeneous needs as a scattered distribution in a high-dimensional space. Thus producers can infer the ideal positions for the product's design attributes that maximise the overall purchase likelihood. However, all these studies does not involve the actual generating of graphic designs.

Despite the extensive investigation of product design optimisation, existing studies have largely simplified the complexity by considering the attributes of design profiles as already-known values. In other words, the prior product design optimisation studies can only tell the best marketing positioning without offering actual graphic designs. Thus, challenging tasks such as accessing the aesthetic attributes of styling or generating vivid designs for selection were not concerns in these investigations.

Inspired by the recent works of applying deep learning to automotive styling ([Pan et al., 2017](#); [Burnap et al., 2021](#)), it has become a great opportunity to integrate deep generation methods with the existing product design optimisation frameworks, in other words, extend the conventional framework with the ability to produce graphic designs.

2.2 Preliminaries of Deep Learning

Basics of Artificial Neural Network

From the historical perspective, deep learning can be seen as an inherited approach following the idea of connectionism that tries to achieve artificial intelligence by organising artificial neural units together as networks, termed artificial neural networks (ANN), to perform various computational tasks ([LeCun et al., 2015](#)).

To start, a number of ANN terms should be introduced first. Neuron – the fundamental units of an ANN, where each one represents a mathematical function. They produce outputs based on the inputs from lower layers and pass values to the next layer. Layer – each layer comprises many neurons, and the stacking of layers results in a complete network. A typical ANN model

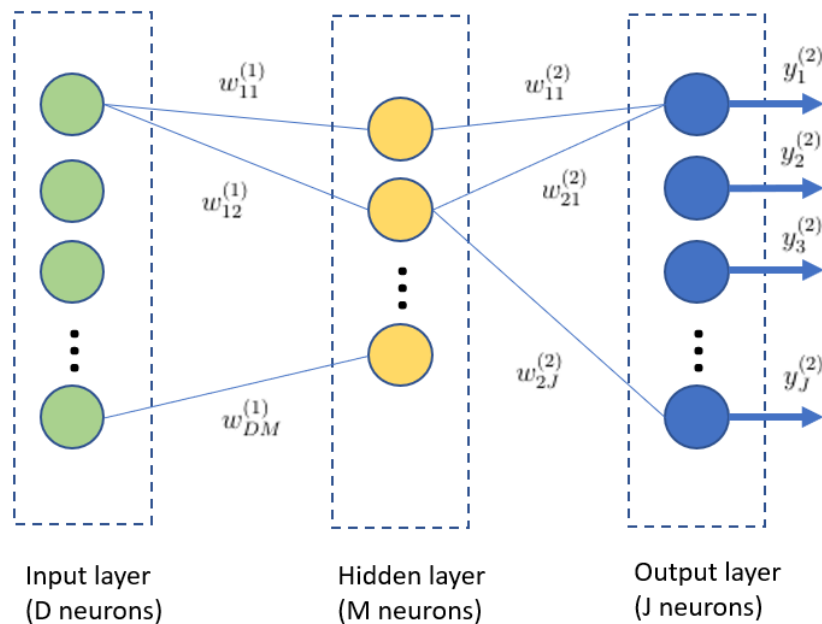


Figure 2.2.1: Structure of a two layer MLP.

consists of three types of layers, namely, the input, hidden and output layers. Weights – the term weight presents the strength value of the connection between neurons. Input layer – the input layer actually represents the input data. It means the dimension number of the input is also the number of neurons in the input layer. Hidden layer – any layers between input and output layers are termed as hidden layers. The hidden layer neurons would sum up the values from the lower layers, process the value via the built-in activation functions, and pass it to the next layer. Output layer – this type of layer is similar to hidden layers, but in many ANNs, its neurons would have different (or without) activation functions (Goodfellow et al., 2016).

The emergence of early ANNs is much based on two findings. The first is Walter Pitts and Warren McCulloch’s “logical neurons” (McCulloch and Pitts, 1943), which they design to imitate the firing of biological neurons, as a biological neuron would not fire until they receive the input stimuli exceed a certain threshold. The second is the “Hebb Rule” (Hebb, 1949), which summarise the principles of how biological neurons connects, namely, “fire together, wire together”. Based on these findings, the first well-known ANN model "Perceptron" is proposed by Rosenblatt (1958). This algorithm can be applied for simple tasks such as binary classification. However, it has been overly propagandised in its early stage, thus, the research community soon lost interests in this algorithm in the 1970s when people realised this algorithm cannot solve the problems such as XOR classification (Goodfellow et al., 2016). Since then, the de-

Table 2.2.1: Notation list for MLP introduction.

D	The number of neurons in the input layer
M	The number of neurons in the hidden layer
J	The number of neurons in the output layer
x	The input layer is made up by $\{x_1, x_2, \dots, x_D\}$, which represents the D dimension inputs. \mathbf{x} is its vector form with a additional 1 (we will explain why later). So \mathbf{x} is $\{1, x_1, x_2, \dots, x_D\}$
t_j	The target (label) value of the j th output
w	$w_{bc}^{(a)}$ is the connection weight between neuron a at hidden layer $a - 1$ and the neuron c at hidden layer a . $\mathbf{w}_c^{(a)}$ is the weight vector of neuron c at hidden layer a , in which the bias $b_c^{(a)}$ is included.
b	$b_c^{(a)}$ is the bias value of the neuron a at layer c . It can be recognised as weight value which represent the $w_{0c}^{(a)}$
σ	$\sigma^{(a)}(\cdot)$ is the activation function used in layer a
$\phi(\cdot)$	The softmax function
z	$z_b^{(a)}$ is the total input to the neuron b at hidden layer a
y	$y_b^{(a)}$ is the output value of the neuron b at layer a

velopment of ANN was significantly hampered until the popularity of multilayer perceptrons (MLP) and the backpropagation (Rumelhart et al., 1986) algorithm.

Before the recent rising of deep learning, MLPs were the most widely applied ANN models, which have multiple layers. In theory, MLP models are considered as “universal approximators” (Hornik et al., 1989) – an MLP model can be trained to approximate any measurable functions. However, the training of such hierarchical models remained unattainable until the invention of the backpropagation algorithm. In early 60’s there exist prior works of backpropagation (Kelley, 1960), these early methods were inefficient and could not obtain enough attention in communities. In the late 80’s, Rumelhart et al. (1986) demonstrated that by updating the neural weights according to their “error” gradients, hierarchical models such as MLP could be properly trained when sufficient data are given. As such algorithm is computed inversely to the feedforward processing, it termed as backpropagation. It has been shown with the backpropagation, MLPs could achieve reasonable performance on recognition and word prediction. The following parts will introduce the computation details in an ANN with essential formulas.

Feedforward Computing

First, detailed computing during the network feedforward phase needs to be introduced. In Fig. 2.2.1, the inputs and outputs of the m th hidden neuron can be expressed as:

$$\begin{aligned} z_m^{(1)} &= b_m + \sum_{d=1}^D w_{dm}^{(1)} x_d = \sum_{d=0}^D w_{dm}^{(1)} x_d = \mathbf{w}_m^{(1)\top} \mathbf{x}, \\ y_m^{(1)} &= \sigma^{(1)}(z_m^{(1)}). \end{aligned} \quad (2.1)$$

When the network is applied for **regression**, the j th output neuron can be expressed as:

$$y_j^{(2)} = z_j^{(2)} = \sum_{m=0}^M w_{mj}^{(2)} y_m^{(1)} = \mathbf{w}_j^{(2)\top} \mathbf{y}^{(1)}. \quad (2.2)$$

So there is no activation function for the output layer. This is because applying the activation function in the output layer would constrain the value of the output. When this model is applied for **classification**, the output of neuron j at the output layer can be expressed as:

$$y_j^{(2)} = \phi(z_j^{(2)}) = \phi\left(\sum_{m=0}^M w_{mj}^{(2)} y_m^{(1)}\right) = \phi(\mathbf{w}_j^{(2)\top} \mathbf{y}^{(1)}), \quad (2.3)$$

where ϕ as the **softmax** function with:

$$\phi(z_j^{(2)}) = \frac{e^{z_j^{(2)}}}{\sum_{i=1}^J e^{z_i^{(2)}}}. \quad (2.4)$$

It turns the original outputs into “probability values”. Before the popularity of deep learning algorithms, the **sigmoid** and **tanh** (hyperbolic tangent) functions were the most widely applied activation functions, which have the following definitions:

$$\begin{aligned} \text{sigmoid}(y) &= \frac{1}{1 + e^{-y}}, \\ \tanh(h) &= \frac{e^y - e^{-y}}{e^y + e^{-y}} = 2\sigma(2y) - 1. \end{aligned} \quad (2.5)$$

Loss Function and Backpropagation

The backpropagation (BP) (Rumelhart et al., 1986) algorithm is applied to update neural weights during training. The main idea is to adjust the network weights according to the deriva-

tion from prediction errors. These error values are generated according to the loss function, which measures how good the current prediction is.

Regression with Quadratic Loss

The quadratic function is the most widely applied loss function for regression tasks, which has the following form:

$$E = \sum_{n=1}^N e_n = \frac{1}{2} \sum_{n=1}^N \sum_{j=1}^J (t_j^{(n)} - y_j^{(n,2)})^2 = \frac{1}{2} \sum_{n=1}^N \sum_{j=1}^J \left\{ t_j^{(n)} - \sum_{m=0}^M w_{mj}^{(2)} \sigma^{(1)} \left(\sum_{d=0}^D w_{dm}^{(1)} x_d^{(n)} \right) \right\}^2, \quad (2.6)$$

where e_n is the cost value for a single data point $\mathbf{x}^{(n)}$. From such definition, we can have the gradient:

$$\frac{\delta e}{\delta z_j^{(2)}} = z_j^{(2)} - t_j = y_j^{(2)} - t_j, \quad (2.7)$$

which further indicates:

$$\frac{\delta z_j^{(2)}}{\delta w_{mj}^{(2)}} = y_m^{(1)}, \quad \frac{\delta z_j^{(2)}}{\delta y_m^{(2)}} = w_{mj}^{(2)}, \quad \frac{\delta y_m^{(1)}}{\delta z_m^{(1)}} = \frac{e^{-z_m^{(1)}}}{(1 + e^{-z_m^{(1)}})^2}. \quad (2.8)$$

Hence, the gradients of weights in the 2nd layer are:

$$\frac{\delta e}{\delta w_{mj}^{(2)}} = \frac{\delta e}{\delta z_j^{(2)}} \cdot \frac{\delta z_j^{(2)}}{\delta w_{mj}^{(2)}} = y_m^{(1)} (z_j^{(2)} - t_j) = y_m^{(1)} (y_j^{(2)} - t_j). \quad (2.9)$$

To calculate the gradients of the weight at 1st layer, as j is not given (which means all the $y^{(2)}$ neurons are computed), so they are:

$$\frac{\delta e}{\delta w_{dm}^{(1)}} = \sum_{j=1}^J \frac{\delta e}{\delta z_j^{(2)}} \frac{\delta z_j^{(2)}}{\delta y_m^{(1)}} \frac{\delta y_m^{(1)}}{\delta z_m^{(1)}} \frac{\delta z_m^{(1)}}{\delta w_{dm}^{(1)}} = \sum_{j=1}^J (y_j^{(2)} - t_j) \cdot w_{mj}^{(2)} \cdot \frac{e^{-z_m^{(1)}}}{(1 + e^{-z_m^{(1)}})^2} \cdot x_d. \quad (2.10)$$

For Classification (Cross-Entropy)

For classification, it is common to use cross-entropy (negative log-likelihood) to represent the loss value:

$$E = \sum_{n=1}^N e_n = - \sum_{n=1}^N \sum_{j=1}^J t_j^{(n)} \log(y_j^{(n,2)}), \quad (2.11)$$

where

$$e_n = - \sum_{j=1}^J t_j^{(n)} \log(y_j^{(n,2)}). \quad (2.12)$$

The application of cross-entropy requires the output $y^{(2)}$ to be probabilities, therefore the soft-max function is applied. And the gradient function of $z^{(2)}$ can be expressed as:

$$\frac{\delta e}{\delta z_j^{(2)}} = \sum_{i=1}^J \frac{\delta e}{\delta y_i^{(2)}} \frac{\delta y_i^{(2)}}{\delta z_j^{(2)}} = - \sum_{i=1}^J \frac{t_i}{y_i} \cdot \frac{\delta y_i^{(2)}}{\delta z_j^{(2)}}. \quad (2.13)$$

As $\{t_1, t_2, \dots, t_J\}$ is a one-hot vector, only $t_t = 1$ (as t is the targeting class), the above equation can be rewrite as:

$$\begin{aligned} \frac{\delta e}{\delta z_j^{(2)}} &= - \frac{t_t}{y_t^{(2)}} \cdot \frac{\delta y_t^{(2)}}{\delta z_j^{(2)}} = - \frac{t_t}{y_t^{(2)}} \cdot \left[\delta \frac{e^{z_t^{(2)}}}{\sum_i^J e^{z_i^{(2)}}} / \delta z_j^{(2)} \right] \\ &= - \frac{t_t}{y_t^{(2)}} \cdot \left[\frac{\delta e^{z_t^{(2)}}}{\delta z_j^{(2)}} \cdot \frac{1}{\sum_i^J e^{z_i^{(2)}}} - \frac{e^{z_t^{(2)}} e^{z_j^{(2)}}}{(\sum_i^J e^{z_i^{(2)}})^2} \right] \\ &= - \frac{t_t}{y_t^{(2)}} \cdot \left[y_t^{(2)} \cdot \mathbb{1}_{(j=t)} - y_t^{(2)} y_j^{(2)} \right] \\ &= t_t \cdot y_j^{(2)} - t_t \cdot \mathbb{1}_{(j=t)} \\ &= y_j^{(2)} - t_j. \end{aligned} \quad (2.14)$$

As the $\frac{\delta e}{\delta z_j^{(2)}}$ in classification is the same as the $\frac{\delta e}{\delta z_j^{(2)}}$ in regression, we have only one gradient formula.

MLP with Deep Layers

When we consider the MLP model with more than one hidden layer, the feedforward computing of Eq.2.2 can be generalised to:

$$y_j^{(c+1)} = \sigma(z_j^{(c+1)}) = \sigma(\mathbf{w}_j^{(c+1)\top} \mathbf{y}^{(c)}), \quad (2.15)$$

where

$$z_j^{(c+1)} = \sum_{m=0}^M w_{mj}^{(c+1)} y_m^{(c)}, \quad y_m^{(c)} = \sigma\left(\sum_{d=0}^D w_{dm}^{(c)} y_d^{(c-1)}\right). \quad (2.16)$$

Specifically:

$$\begin{aligned} \frac{\delta z_j^{(c+1)}}{\delta y_m^{(c)}} &= w_{mj}^{(c+1)}, \quad \frac{\delta y_m^{(c)}}{\delta w_{dm}^{(c)}} = y_d^{(c-1)}, \quad \frac{\delta y_m^{(c)}}{\delta z_m^{(c)}} = \frac{e^{-z_m^{(c)}}}{(1 + e^{-z_m^{(c)}})^2}, \\ \frac{\delta z_j^{(c+1)}}{\delta z_m^{(c)}} &= w_{mj}^{(c+1)} \cdot \frac{e^{-z_m^{(c)}}}{(1 + e^{-z_m^{(c)}})^2}, \quad \frac{\delta z_j^{(c+1)}}{\delta w_{dm}^{(c)}} = w_{mj}^{(c+1)} \cdot y_d^{(c-1)}. \end{aligned} \quad (2.17)$$

Based on Eq.23, we have:

$$\frac{\delta e}{\delta w_{mj}^{(c+1)}} = \frac{\delta e}{\delta z_j^{(c+1)}} \cdot \frac{\delta z_j^{(c+1)}}{\delta w_{mj}^{(c+1)}} = \frac{\delta e}{\delta z_j^{(c+1)}} \cdot y_m^{(c)}, \quad (2.18)$$

$$\frac{\delta e}{\delta z_m^{(c)}} = \sum_{j=1}^J \frac{\delta e}{\delta z_j^{(c+1)}} \cdot \frac{\delta z_j^{(c+1)}}{\delta z_m^{(c)}}, \quad (2.19)$$

$$\frac{\delta e}{\delta w_{dm}^{(c)}} = \frac{\delta e}{\delta z_m^{(c)}} \cdot y_d^{(c-1)}. \quad (2.20)$$

So we are able to calculate the $\frac{\delta e}{\delta z^{(c)}}$ values once the $\frac{\delta e}{\delta z^{(c+1)}}$ values in the previous layer are known.

Adopting ReLU

In modern convolutional neural networks, the ReLU is widely adopted as the activation function, which has the following expression:

$$\text{ReLU}(y) = \max(0, y) \quad (2.21)$$

The new backpropagation equations are:

$$\begin{aligned} \frac{\delta y}{\delta z} &= \mathbb{1}_{(z>0)} \\ \frac{\delta e}{\delta w_{dm}^{(1)}} &= \sum_{j=1}^J (y_j^{(2)} - t_j) \cdot \mathbb{1}_{(z_j^{(2)}>0)} \cdot w_{mj}^{(2)} \cdot \frac{e^{-z_m^{(1)}}}{(1 + e^{-z_m^{(1)}})^2} \cdot \mathbb{1}_{(z_m^{(1)}>0)} \cdot x_d \\ \frac{\delta e}{\delta w_{ab}^{(c)}} &= \sum_{d=1}^D \frac{\delta e}{\delta z_d^{(c+1)}} \cdot w_{bd}^{(c+1)} \cdot \mathbb{1}_{(z_a^{(c)}>0)} \cdot y_a^{(c-1)} \end{aligned} \quad (2.22)$$

Convolutional Neural Networks (Being applied from chapters 3 to 6)

Due to the overly exaggerated future of Artificial Intelligence, ANN-related studies have been treated as pseudoscience during the AI winter from 1985 to the 90s. Despite that, a group of researchers (such as Geoffrey Hinton, Yann LeCun) continued to work on the ANN models and achieved considerable advancements (Goodfellow et al., 2016) in this period.

Convolutional neural network (CNN), the most prevailing deep-learning algorithm currently, can be recognised as a derivation of MLP. Its history can date back to the 80s when Fukushima (1980) published the study regarding Neocognitron, an early version of convolutional neural network architecture. It allows the model to perform visual pattern recognition tasks. Due to the lack of proper training methods, the algorithms have not captured enough attention. The first modern CNN was the LeNet-5 model proposed by Lecun et al. (1998), which has three types of layers: the convolutional, pooling and dense layer. By trained with backpropagation, this model achieves outstanding performance on handwritten number recognition, thus widely used for handwritten check reading systems.

The recent boom of deep learning was started in 2012, by the achievements made by AlexNet from Krizhevsky et al. (2012), this CNN architecture which wins the 2012 ImageNet (Deng et al., 2009) competition with a 15.4% error rate on the Top-5 test (i.e., whether the answer is within the top 5 predictions). The result was so incredible that the whole of the computer-vision community had been attracted to the deep-learning approach. Since then, large amounts of CNN models with diverse structures have developed. Among these newly developed models, the models such VGG16 (Simonyan and Zisserman, 2015), GoogLeNet (Szegedy et al., 2015) and ResNet (He et al., 2016) are the most representative deep models used for predictive tasks.

Table 2.2.2: Notation list for CNN introduction.

$M^{(r)}$	The width of the r th layer output.
$N^{(r)}$	The height of the r th layer output.
$C^{(r,l)}$	The output feature map (matrix) from the r th layer l th kernel. The M and N represents the width and height of the output.
$K^{(r,l)}$	The l th kernel at layer r which has $P \times Q$ size. In modern CNN models, the applied kernels always have the square shape. That is, $P = Q$, and the P and Q are always odd numbers.

Convolutional Layer

As its name suggests, convolutional neural networks are characterised by convolutional layers. A convolutional layer has multiple kernels, and for each kernel, the computation is to scan the whole input matrix (or tensor) by shifting the kernel. Thus a feature map will be generated correspondingly for each kernel. The convolution is a mathematical operation on two functions to produce a third function, which can be written as $(f * g)(t)$. It can be expressed in integral form as:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\delta)g(t - \delta)d\delta, \quad (2.23)$$

or in discrete form as

$$(f * g)(t) = \sum_{\delta=-\infty}^{\infty} f(\delta)g(t - \delta). \quad (2.24)$$

Say if there is only one feature map from the $C^{(r-1)}$ layer, then the convolution process in the convolutional layer r can be expressed as:

$$C_{(m,n)}^{(r,l)} = (f * k)(m,n) = \sum_{a=-d}^d \sum_{b=-d}^d K_{a,b}C_{m-a,n-b}^{(r-1,1)}, \quad (2.25)$$

where $d = \lfloor \frac{P-1}{2} \rfloor$. As the convolutional process would produce a size-shrunk (compared with the input size) feature map, a specific method, namely padding, is used to maintain the map size. This method adds the 0s to the surround of the C. So m, n are equal to the input sizes, in case $m - a < 0$ or $m - a > M$, they will set to zeros.

Derivation Models from CNN

Besides the basic predictive tasks such as classification or regression, derivation models of the convolutional neural network have also achieved massive record-breaking results in various domains over the past decade (LeCun et al., 2015). Our studies have used and integrated different deep learning algorithms from several fields. In the following, we will particularly introduce the Mask R-CNN network, as it has been repeatedly used in our studies for image background removal.

Mask RCNN (Being applied in Chapter 3, 4, 5, 6)

The Mask R-CNN is a CNN-based image segmentation method. It is the state-of-the-art CNN proposed by (He et al., 2017), which aims to detect and segment object instances in digital images. It has been widely used in autonomous driving (Huang et al., 2020; Xia and

Sattar, 2019), multi-person pose estimation (Dong et al., 2019), and neural stem cell differentiation (Zhu et al., 2021). In this thesis, we often use it to detect car body in the given raw car image and remove the background accordingly. The Mask R-CNN has two stages: (i) the region proposal network (RPN) that creates the candidate object bounding boxes; and (ii) the detection stage in which segmentation masks are predicted. As exhibited in Figure 2.2.2, we use Mercedes C Class (2019 reg) to illustrate how does the Mask R-CNN work in our context. In the first stage, the RPN proposes region candidates that contain car objects. Technically speaking, each candidate is called the *region of interest*. The input image will first go through a series of convolutional layers that are well-trained in existing CNNs such as ResNet (He et al., 2016) and feature pyramid network (Lin et al., 2017). This will result in matrices called *feature maps* (Goodfellow et al., 2016), and each feature map is a matrix that topographically indicates the presence of a common feature in the input image. Feature maps are then converted into patches called *anchors*, and then be used in training the model. Within our context, each patch of the image refers to specific car shape regions. The RPN is often trained independently before fitting into the entire Mask R-CNN framework and its loss function \mathcal{L}_{RPN} is defined as

$$\mathcal{L}_{\text{RPN}} = \lambda_{\text{CLS}} \mathcal{L}_{\text{CLS}} + \lambda_{\text{BOX}} \mathcal{L}_{\text{BOX}}, \quad (2.26)$$

where \mathcal{L}_{CLS} is the classification loss, \mathcal{L}_{BOX} is the bounding-box loss, and λ_{CLS} and λ_{BOX} are the corresponding coefficients.

The classification loss \mathcal{L}_{CLS} is specified by the cross-entropy

$$\mathcal{L}_{\text{CLS}} = - \sum_{\tilde{i}} \left[p_{\tilde{i}} \log \{ \hat{p}_{\tilde{i}} \} + (1 - p_{\tilde{i}}) \log \{ 1 - \hat{p}_{\tilde{i}} \} \right], \quad (2.27)$$

where \tilde{i} is the index of anchors, $\hat{p}_{\tilde{i}}$ is the predicted probability that anchor \tilde{i} is belong to an object and $p_{\tilde{i}}$ is the corresponding ground truth label in $\{0, 1\}$.

The bounding-box loss \mathcal{L}_{BOX} is defined by Girshick (2015) as follows

$$\mathcal{L}_{\text{BOX}} = \sum_{\tilde{i}} \sum_{\tilde{j}} \left[\frac{1}{2} d_{(\tilde{i}, \tilde{j})}^2 \right]^{\mathbb{I}_{d_{(\tilde{i}, \tilde{j})} < 1}} \left[|d_{(\tilde{i}, \tilde{j})} - \frac{1}{2}| \right]^{1 - \mathbb{I}_{d_{(\tilde{i}, \tilde{j})} < 1}}, \quad (2.28)$$

where $d_{(\tilde{i}, \tilde{j})}$ is the difference between the predicted and ground-truth values of anchor \tilde{i} 's coordinate \tilde{j} .

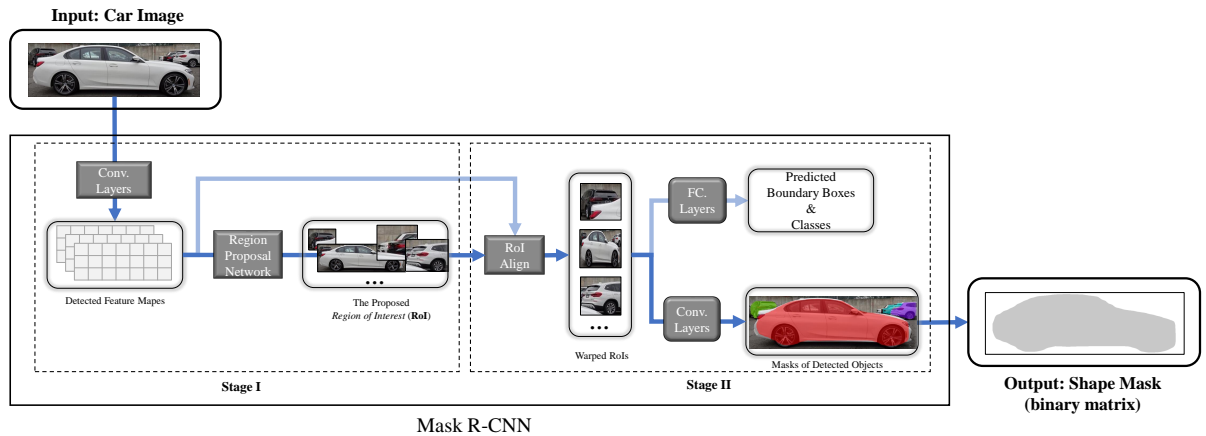


Figure 2.2.2: Schematic view of using the Mask R-CNN to segment car body from its original image.

In the second stage of the Mask R-CNN, our primary goal is to cover the shape of detected car objects. The detection stage then completes segmentation masks. The mask loss $\mathcal{L}_{\text{MASK}}$ is taken into account and can be specified by cross-entropy as follows

$$\mathcal{L}_{\text{MASK}} = - \sum_l \sum_k \left[q_l^{(k)} \log \{ \hat{q}_l^{(k)} \} + (1 - q_l^{(k)}) \log \{ 1 - \hat{q}_l^{(k)} \} \right], \quad (2.29)$$

where l is the index of pixel in the RoI, $\hat{q}_l^{(k)}$ and $q_l^{(k)}$ are the predicted probability and the ground truth of pixel l for class k .

Deep Generative Models (Being applied in Chapter 3, 4, 5, 6)

In contrast to the CNN algorithm that has a long history, the deep-generative models were not emerged until a few years ago, where Goodfellow et al. (2014) created the first generative adversarial network (GAN). GANs are used to generate new samples to approximate the probability distribution of the true data, and they solve unsupervised tasks by using a supervised loss as part of model training. As these algorithms are excellent at generating real-looking pictures, the studies of deep-generative models have become remarkably popular since then.

As GANs can achieve tremendous success in generating highly realistic images, they have been recently used for product appearance design. For example, Quan et al. (2018) developed a framework that can incorporate the given colour and styling features into clothing design. Sbai et al. (2019) investigated whether the addition of an explicit loss for creativity can drive GANs to produce original and compelling fashion designs.

The vanilla GAN network is mainly comprised of a generator (G) and a discriminator (D). The generator is used to generate images based on the given input z (latent vector); while the discriminator is applied to discriminate whether the input image is real or not. By learning the true data distribution from the training sets, the generative models can generate new images which resemble the real ones. The objective function of GAN is:

$$\min_G \max_D \mathbb{E}_{x \sim P_{data}} [\log(D(x))] + \mathbb{E}_{z \sim P_z} [\log(1 - D(G(z)))]. \quad (2.30)$$

The intuition of the function is straightforward. The $\log(D(x))$ and $\log(1 - D(x))$ represent the log probability of correct discrimination. So the $D(\cdot)$ is trained to max the objective function as it tries to identify whether the image is faked or not. Contrariwise, the $G(\cdot)$ is trained to max the function by maximising the $\mathbb{E}_{x \sim P_G} [\log(1 - D(x))]$ part. Moreover, the vanilla GAN can easily be extended to conditional GAN with the following modification:

$$\min_G \max_D \mathbb{E}_{x \sim P_{data}} [\log(D(x|y))] + \mathbb{E}_{z \sim P_z} [\log(1 - D(G(z|y)))]. \quad (2.31)$$

In the first phase of model training, $G(\cdot)$ is fixed, where the optimal $D_G^*(x)$ is:

$$D_G^*(x) = \frac{P_{data}(x)}{P_{data}(x) + P_G(x)}. \quad (2.32)$$

If we denote $P_{data}(x)$ as a , $P_G(x)$ as b , and $D(x)$ as y , then:

$$\frac{d[a * \log(y) + b * \log(1 - y)]}{dy} = 0, \quad \text{when } y = \frac{a}{a + b}. \quad (2.33)$$

In the second phase, $D(\cdot)$ is fixed, and we have:

$$\min_G \int_{P_{data}} P_{data}(x) \log \left(\frac{P_{data}(x)}{P_{data}(x) + P_G(x)} \right) dx + \int_{P_G} P_G(x) \log \left(\frac{P_G(x)}{P_{data}(x) + P_G(x)} \right) dx. \quad (2.34)$$

Table 2.2.3: Notation list for GAN introduction.

P_{data}	The distribution of real image
P_z	The distribution of the random variable z
x	Real image sample
$G(\cdot)$	Generator
$D(\cdot)$	Discriminator

Key Literature	Year	Description	Design Generating	Design Evaluating	Positioning Optimising
Rajeev and Krishnamurti (1987)	1987	Non-styling focus; marketing			✓
Kaul and Rao (1995)	1995	positioning focus; no actual			✓
Shi et al. (2001)	2001	designs			✓
Burnap et al. (2016c)	2016	Design studies; automotive		✓	
Burnap et al. (2016b)	2016	styling; deep learning; novel	✓		
Pan et al. (2017)	2017	design focusing	✓	✓	
Burnap et al. (2021)	2019		✓	✓	

Table 2.3.1: The list of key literature for this thesis.

This objective function will achieve the minimum at $P_{data}(x) = P_G(x)$.

2.3 Deep Learning for Automotive Styling

The research community has long time attempted to develop computational technologies for generating automotive designs. Some studies rely on the manipulation of points and lines in the 2D graph to generate uncomplicated car shapes and silhouettes and applied surveys to collect aesthetic feedback ([Reid et al., 2010](#)). More up-to-date studies advance these methods by creating designs in 3D spaces with more rich colour choices ([Kókai et al., 2007](#); [Orbay et al., 2015](#)). However, among the development of such traditional design methods, researchers are often challenged by the difficulties of setting the increasing number of control variables – in order to generate realistic designs, designers need to handle more additional parameter setting problems. As [Burnap et al. \(2016b\)](#) concluded, there exists a clear trade-off between the difficulties of manipulations and the design flexibility. Such challenges made the computational methods unsuitable for the conceptual design stage.

Despite the wide use of computational technologies such as CAD in the late stages of automotive design, computers are seldom involved in the early conceptual design stage ([Tovey et al., 2000](#)). Such conceptual design procedure is led by the aesthetic design team, indicating the quality of the resulting designs is much determined by the expertise of team members. The rough design ideas are presented in the form of sketches, which are more convenient for fast modification and selection. Design evaluations are done through consumer surveys to decide the designs for further investing, termed the theme clinics. Only the selected designs are then digitised using computer-aided design programs ([Coates, 2003](#)).

With the popularity of deep learning in the CS community, a group of researchers, especially the design research team from the University of Michigan, tried to apply these state-of-art methods to automotive aesthetic design. The first investigation that tried to deploy the deep

generative model on the automotive aesthetic design was from [Burnap et al. \(2016a\)](#). The authors used the variational autoencoders (VAEs) ([Kingma and Welling, 2014](#)) to map various car design images into a design space controlled by variables drawn from uniform distributions. By manipulating these input variables, the trained VAE is able to produce car designs from different viewpoints with diverse looks. The authors suggest that the most substantial advantage of using the deep generative model is the resolution of the design flexibility VS the realism trade-off, where the unimportant details are automatically filled by the machine. In comparison, the conventional methods generate realistic designs at the price of increasing the control difficulties with more input variables. Although the early VAE model results are perceived as inferior to modern methods, this prior attempt to apply the deep generative model is a clear milestone for the computational assistant product appearance design.

In their later work, [Pan et al. \(2017\)](#) firstly proposed the generator + evaluator framework for car exterior design automation. Different from their prior works, the conditional generative adversarial network (cGAN) ([Mirza and Osindero, 2014](#)) was employed as the design generator in this study to gain higher controllability of produced results. So the authors can specify the desired characteristics of brand, colour or bodytype for generation, as these variables had been conditioned during the GAN model training. On the other hand, the Siamese neural network was used to predict consumers' ratings from different market segments, as it is specified on learning contrastive features between inputs with similar looks. Moreover, the authors applied the gradient-based interpretation method to the evaluator network to highlight the influential design parts for design attribute perceiving.

In their recent work, [Burnap et al. \(2021\)](#), more oriented toward marketing research, the authors tried integrating three deep neural network models, VAE, GAN and CNN conceptually, by jointly interplaying with the proposed embedding space. In particular, the VAE model was responsible for mapping design images to embedding vectors with a significantly smaller number of dimensions ([Kingma and Welling, 2014](#)). The GAN model constructed design images on the embedded vectors, and the predictive model forecasted the design's appearance ratings. Apart from that, this study examined the actual values of adopting deep learning technologies in automotive design by interviewing real automotive designers from the general motors corporation. Unlike the periodic consumer research clinics, which are expensive, and time costly, the machine-based design evaluator can be continually available with no extra costs. On the other hand, the deep generative models can efficiently provide alternative choices for the design team during the conceptual design stage, unaffected by the human designers' aesthetic preferences.

2.4 Summary

In this chapter, we first reviewed the historical literature from marketing, design and psychology to further explain the background and motivation of this thesis. We particularly reviewed product design optimisation studies that proposed diverse mathematical solutions, where no actual product designs were offered in these studies. The chapter then introduced the basics of deep learning. In particular, it went through the classical artificial neural network models and the fundamental concepts. It then introduced more advanced models, such as CNNs and deep generative models. The last part presented the conventional computer science methods used in automotive styling design and discussed the constraints. Then it came to the recent trend of using deep learning technologies in automotive exterior design and went through the pioneer works done by the University of Michigan design team, which clearly showed the advantages of the recently advanced machine learning methods. The next chapter presents the proposed general framework and discusses the key challenges of the proposed studies.

Chapter 3

DVM-CAR: An Automotive Database for Deep Visual Marketing

In this chapter, we present our multidisciplinary initiative and create a large-scale automotive research dataset named DVM-CAR, which serves as a valuable database for my PhD studies and has been shared via the following website:

<https://deepvisualmarketing.github.io>.

Section 3.1 presents the motivation for DVM-CAR. Section 3.2 describes how the raw data is collected and processed. Section 3.3 gives an overview of the dataset. Section 3.4 explains how the dataset is designed to be practical for research. Section 3.5 illustrates three research samples that could be achieved with DVM-CAR, and Section 3.6 summarises the chapter.

3.1 Introduction

Over the past years, there has been a rising interest in applying data mining, machine learning and artificial intelligence to automotive research and applications, including autonomous driving, intelligent manufacturing, car appearance design, consumer analytics and sales modelling (Burnap et al., 2016c; Pan et al., 2016, 2017; Burnap et al., 2021). In the former two areas, automotive manufacturers are working actively with robotics research experts, whom they have been closely with over the decades in automobile engineering. The latter three areas contain significant business ingredients such as product design, marketing, operational research and economics. They have generated many interesting multidisciplinary research questions, and

solving them can have positive social consequences for automotive manufacturers, consumers, and other participants in the automotive ecosystem and even policy makers (Dant and Martin, 2001).

Despite the growing interest in car aesthetics analytics and design, the lack of applicable datasets has greatly hampered the usage of machine learning technologies in automotive design studies as data mining, machine learning and artificial intelligence are built upon data, in many cases, massive data (Yang et al., 2015; LeCun et al., 2015; Huang et al., 2020). Before our study, no publicly available datasets cover a wide range of car or other commercial product information, such as product appearance, model specifications and sales data, are extensively lacking. Meanwhile, data collection can often be a lengthy, difficult or expensive process for many researchers, particularly, researchers in the business community whose primary focuses are not programming or database management.

To this end, we develop a comprehensive car dataset, which researchers from different disciplines can use, by integrating different online product information. This chapter makes the following four major contributions:

- **Domain Application:** Our dataset meets the growing need of a comprehensive automotive dataset for visual-related marketing research and applications. For example, our dataset can be used for automotive exterior design and consumer analytics, which will benefit automotive manufacturers, help them better understand their targeted consumer segment preferences behind the purchase decisions, and then use the obtained consumer insights to generate new exterior designs powered by technologies. Our dataset can also be used for car sales modelling which benefits all the participants in the automotive ecosystem including car dealers, consumers, and marketers.
- **Big Data:** To the best of our knowledge, our dataset is the very first large-scale automotive dataset. It contains 1.4 million images from 899 different car models and corresponding specification and sales information from over ten years in the UK market. The four characteristics of big data (i.e., volume, variety, velocity, and veracity) are satisfied, which makes the dataset can be used for different types of analytics, ranging from descriptive to diagnostic, predictive and perspective. The dataset can be used by multidisciplinary researchers to solve different tasks. Many interesting insights, data-driven models and applications can be derived.
- **Database Design:** We conduct a survey study with the researchers working in either

business or computer science field. The data challenges these researchers meet can be categorised into three major issues: coverage, accessibility and quality. We design the proposed dataset by addressing these issues with the hope that our data can be researcher-friendly and thus can have a large impact in research and practice. Therefore, we also demonstrate a good practice of developing a dataset which can alleviate the common data issues faced by researchers.

- **Data Fusion:** Car images, model specification and sales information are collected from different online sources while they are merged and stored in a flexible and hierarchical structure that allows it to be easily expanded with new data and used by researchers. We also compare our dataset with several related existing publicly available datasets, indicating the differences and our improvements. Our work contributes to the multi-source data fusion as it includes different data formats from different online sources.

3.2 Data Collection and Preparation

Nowadays researchers often collect online data and process them as new datasets for their research. Datasets like ImageNet (Deng et al., 2009) and OpenImages (Kuznetsova et al., 2020) are characterised by their enormous size and have achieved massive success in computer science research and applications. The two existing car image datasets, Stanford-Car ¹(Krause et al., 2013) and CompCars ²(Yang et al., 2015), are also based on the web scraped contents. Motivated by these existing works, we develop the dataset by collecting and integrating data from different online sources. First, car images are collected from the popular automotive classified advertising platforms. These platforms are popular online marketplaces in the UK for buying and selling used cars, which host millions of car images and their selling prices for almost all car models from different automotive manufacturers. Second, car sales data is collected from the Driver and Vehicle Licensing Agency (DVLA), which is part of the Department for Transport, holding over 49 million driver records and over 40 million vehicle records in the UK. The DVLA publishes the statistics of newly registered vehicles in its seasonal reports. We extract and add the sales appropriately for various car models. Third, new car prices are collected from car review websites, which cover the selling prices of various car trims sold in past years.

In the preparation of data sharing, a number of data cleaning and fusion steps are performed. Images with unexpected contents are dropped, and remaining images are cropped according to

¹https://ai.stanford.edu/~jkrause/cars/car_dataset.html

²http://mmlab.ie.cuhk.edu.hk/datasets/comp_cars/index.html

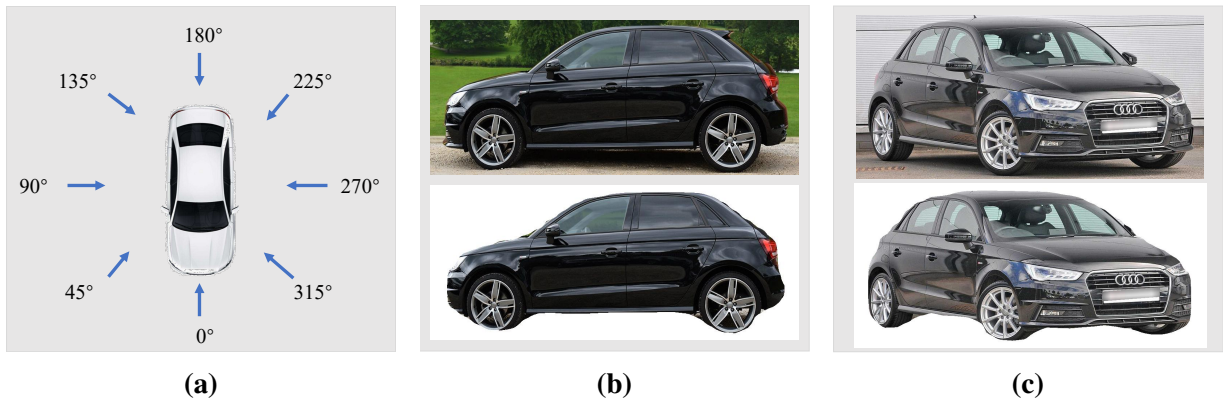


Figure 3.2.1: Illustration of car images in our dataset: (a) angles of car image labelling; (b) example of background removal for Audi A2 from angle 270 degree; (c) example of background removal for Audi A2 from angle 45 degree.

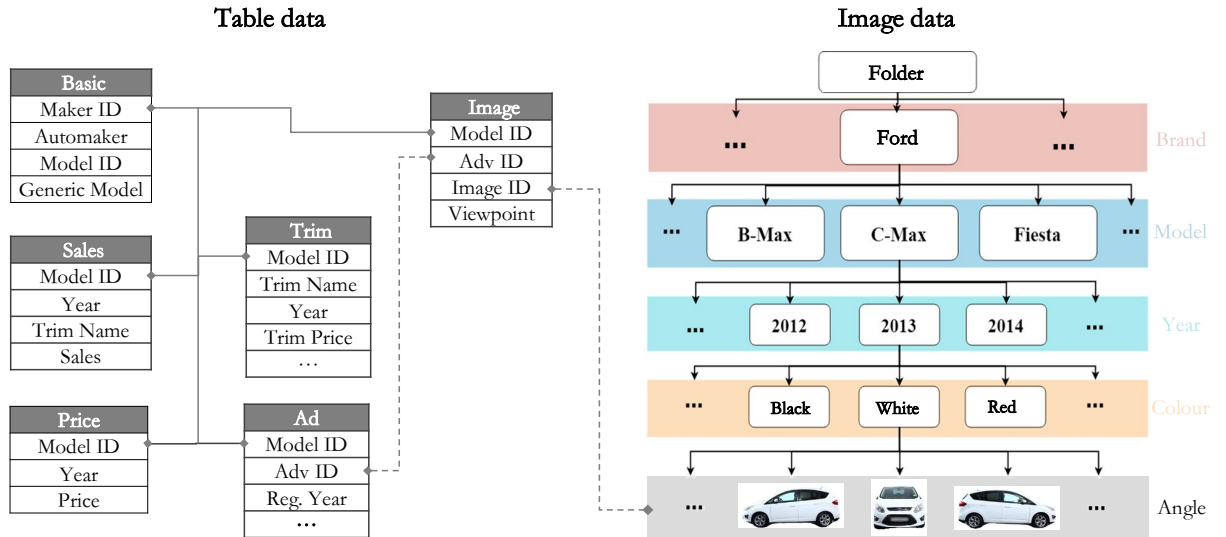


Figure 3.2.2: Schematic view of the DVM-CAR dataset, which consists of two parts: image data and table data. Image and table data can be joined or connected via model, adv and image IDs.

detected car positions. First, indicated by as Fig. 3.2.1 (a), a machine learning model is used to filter out non-exterior-viewing images. An ImageNet pre-trained convolutional neural network (CNN) (Lecun et al., 1998) is fine-tuned with manually prepared samples where pictures are labelled according to their qualities and observation viewpoints. The trained CNN classifies images according to observation viewpoints, only images taken from the eight targeting viewpoints are kept for later usage. This reduces the size of the raw image set from over six million to less than 1.5 million. Second, as Fig. 3.2.1 (b) illustrates, to comply with the General Data Pro-

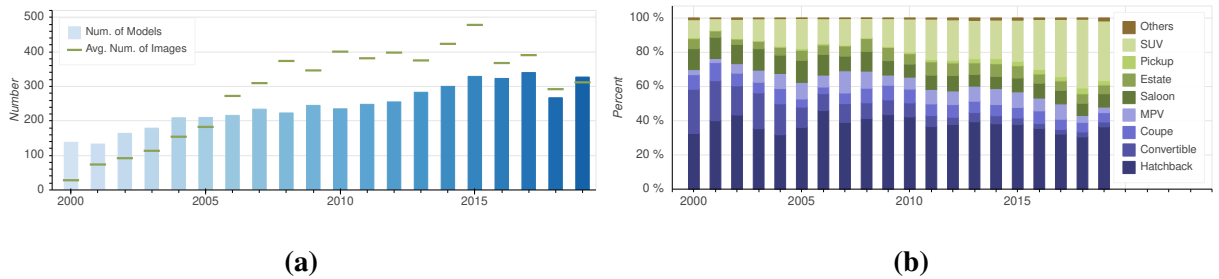


Figure 3.2.3: Summary of image data in the DVM-CAR dataset: (a) the number of car models across years and the average images for each model; (b) the annual percentage composition of images by body type.

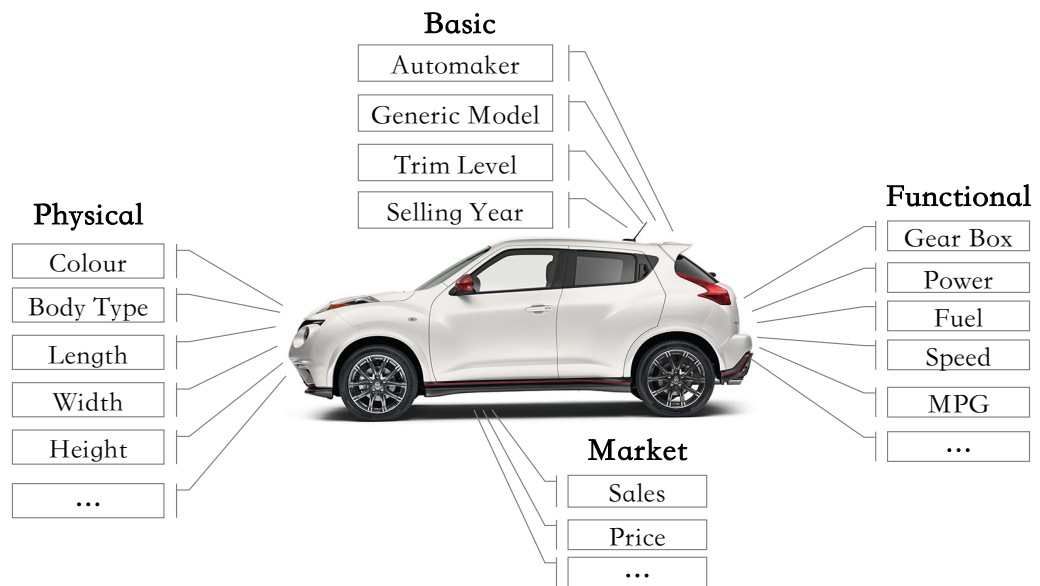


Figure 3.2.4: Four categories of variables in our dataset.

tection Regulation (GDPR) (Voigt and Von dem Bussche, 2017), we utilise Mask R-CNN (He et al., 2017) (trained on the COCO dataset (Lin et al., 2014)) to detect objects present in car images and to identify the positions of car bodies within these images. This approach enables the removal of regions not pertaining to the car bodies, ensuring the absence of extraneous objects, such as human figures, in the processed images. Figure 3.2.1 (c) exhibits the resulting images after detecting number plates using the algorithm from (Silva and Jung, 2018), followed by blurring the number plate regions. Beyond image processing, all non-visual content is transformed into attributes and organised into data tables. Ambiguous observation values are corrected or harmonised. Each car model is assigned a unique identifier (ID) to facilitate information integration. Information such as annual sales is compiled at a more abstract level. In the table,

images are associated with their observation viewpoints and source advertisement IDs.

3.3 Dataset Description

As our primary motivation to facilitate visual marketing research and forecasting applications, we call the proposed dataset **Deep Visual Marketing Car** (in short **DVM-CAR**), which is publicly available under the **CC BY-NC** license at:

<https://deepvisualmarketing.github.io>

As shown in Fig. 3.2.2, the DVM-CAR dataset consists of two parts: image data and table data. The part of **image data** contains 1,451,784 car images (in JPEG format with resolution 300×300) which are compressed in a ZIP file (13.6 GB file size). These images have been selected from eight observation viewpoints and stored under the categorisation “brand-model-year-colour”. This structure allows researchers to locate target images easily. The backgrounds of all car images are removed. A table is prepared for image indexing, so researchers can select the images via this table. In total, our image data covers 899 car models that sold in the UK market over the last 20 years. As Fig. 3.2.3 (a) shows, the automotive classified advertising platforms have more data for newly-launched models than older car models. In spite of this, it contains 138 car models sold in the year 2000, each with an average of 28 images. The wide longitude of the dataset makes the observation of long-term trends easy, thus, is particularly useful for economic and marketing analytics and forecasting. For instance, through Fig. 3.2.3 (b), it is shown that SUVs are becoming more popular and taking higher market shares over the last two decades. The part of **table data** (i.e., non-visual part) comprises six tables in the CSV format (156 MB file size), which are named the basic, sales, price, trim, ad and image tables. Together, these tables cover various variables and form a relational database (Codd, 1983), as each two of them can be joined via the primary or secondary keys. More details of these tables can be found in Table 3.3.1.

3.4 Quality Control: Towards a Good Research Dataset

During the data preparation, we deploy a survey using Qualtrics (www.qualtrics.com) to explore the most common data issues that people face in their research and practice. We collect responses from 54 researchers, including 26 participants with a computer science background and 28 with a business studies background, including economics, marketing and management. As presented in Figure 3.4.1, the reported issues from our survey study can be broadly divided

Table 3.3.1: Description of table data in the DVM-CAR dataset.

Name	Table description	Attribute name	Attribute description
Basic	It is mainly for indexing other tables.	Genmodel	Generic model name
		Genmodel ID	Generic model ID
		Automaker	Automaker name
		Automaker ID	Automaker ID
Ad	It contains more than 0.27 million used car advertisements information posted on the automotive classified advertising platforms, including variables like advertisement's creation time, used car registration year, cumulative mileage, selling price, etc.	Genmodel	Generic model name
		Genmodel ID	Generic model ID
		Maker	Automaker name
		Adv ID	Advertisement ID
		Adv year	Advertisement's creation year
		Adv month	Advertisement's creation month
		Colour	This car's colour
		Reg year	This car's first registration/selling year
		Bodytype	This car's body type
		Runned Miles	This car's runned mileage
		Engin size	This car's engin size
		Gearbox	This car's gearbox
		Fuel type	This car's fuel type
		Price	This car's selling price
		Seat num	This car's seats number
Door num	This car's doors number		
Image	It contains image data related information like predicted viewpoint and quality check result.	Genmodel ID	Generic model ID
		Image ID	Image ID
		Image name	Image name
		Predicted viewpoint	This image's predicted viewpoint
Price	It contains the entry-level new car prices. It is designed for people who only need the basic price of car models.	Quality check	Manually check result
		Genmodel	Generic model name
		Genmodel ID	Generic model ID
		Maker	Automaker name
		Year	Generic model's selling year
Sales	It contains car sales data of the UK market (based on the released statics from the DVLA).	Entry price	Generic model's entry-level price
		Genmodel	Generic model name
		Genmodel ID	Generic model ID
		Maker	Automaker name
Trim	It includes 0.33 million trim level information such as selling price, fuel type and engine size. It is designed for people who are interested in the price of car model at a specific trim level.	Year 2001 to 2020	Generic model's annual sales
		Genmodel	Generic model name
		Genmodel ID	Generic model ID
		Trim	Trim name
		Maker	Automaker name
		Year	Trim's selling year
		Price	Trim's price at selling year
		Gas emission	Trim's CO ₂ emission
Fuel type	Trim's fuel type		
Engine size	Trim's engine size		

into three categories: coverage, accessibility, and quality.

Coverage refers to the issue of a dataset not containing the needed information. This happens in two situations. First, the dataset is not comprehensive enough. It does not contain the attributes or variables that a researcher is interested in for a specific study. Second, the dataset does not cover enough records or samples. Our dataset is designed to have excellent coverage

Variables/features	Our dataset	Related work			
		Kukova et al.	Korenok et al.	Jindal et al.	Landwehr et al.
Sales/market share	✓		✓	✓	✓
Price	✓	✓	✓	✓	✓
Brand	✓	✓	✓		✓
Exterior features	✓	✓	✓	✓	✓
Fuel economy	✓	✓		✓	
Horsepower	✓	✓		✓	
Engine	✓	✓		✓	
Transmission	✓	✓		✓	
Equipment features	✓	✓			✓
Life-cycle/model year	✓		✓	✓	✓
Advertising			✓	✓	
Reliability		✓	✓		
Safety		✓	✓		
Driving/handling		✓		✓	
Ergonomics/rooming		✓		✓	
Interior		✓			

Table 3.3.2: Summary of car specification and sales variables included in the related marketing studies and our proposed DVM-CAR dataset.

of both variables and observed samples. For the former, as presented in Table 3.3.2, we design the dataset to contain many important car specifications and sales variables used in the related marketing studies. In addition to these variables, as also demonstrated in Fig. 3.2.1, we provide corresponding images for various car models in various angles over a long-term period. The existing public car image datasets, such as the Stanford-car and CompCars datasets, do not include the mentioned car specification and sales variables as they are mainly designed for computer vision tasks. In terms of volume, the dataset covers millions of registered cars in the UK market in the past decades.

Accessibility refers to the difficulties regarding data usage caused by closed or proprietary datasets or the underlying complex data structure. Our participants in the survey highlighted that many datasets require license purchase, and in many cases, these datasets are sold at remarkably high prices. Researchers find that it is difficult to secure funding to purchase their needed datasets. On the other hand, the free shared datasets are often presented in a researcher-unfriendly way. The developers of those datasets may either put too much trivial information or leave data unprocessed. Even worse, many datasets are shared online without providing basic descriptions. A lot of efforts have been made to improve the accessibility of our dataset. First, we create an easily accessible data webpage on GitHub where researchers can download our dataset and find the needed description and usage instructions. Second, there are also no

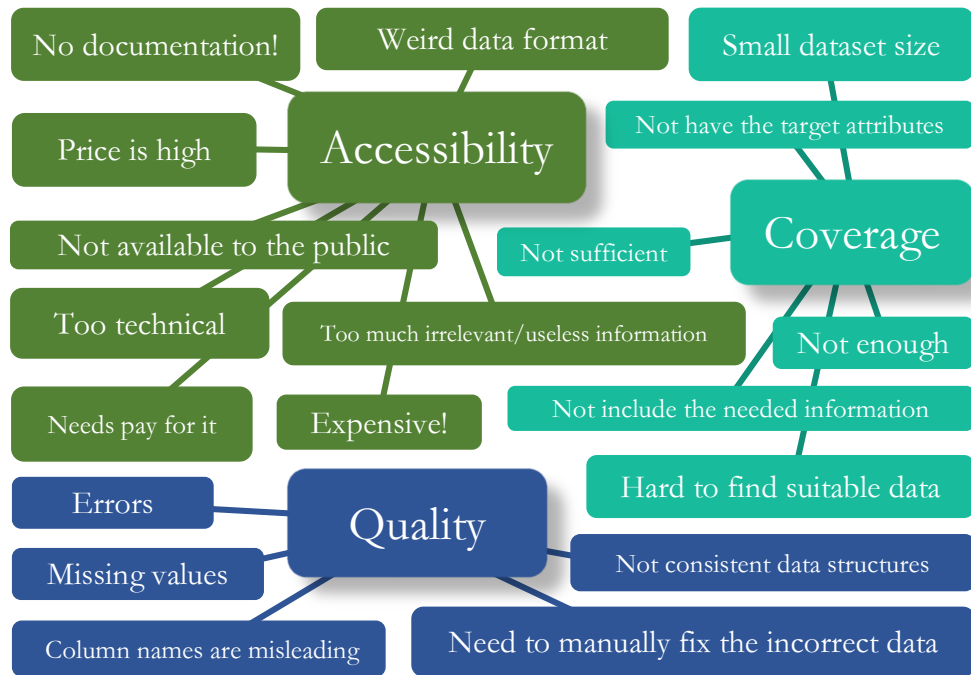


Figure 3.4.1: Categorisation of data issues reported by participating business and computer science researchers.

specific restrictions on our data usage. Moreover, we hope our data can be easily used by researchers from different backgrounds, so it is presented in a researcher-friendly manner. For example, all non-visual contents are processed into tabular attributes and organised into separate tables according to their categories. All the car models have been assigned a unique identifier (ID), which can be used for information integration and future data fusion. In addition, we remove the background of collected images, which can simplify the potential applications for researchers interested in automotive exterior design.

Quality refers to missing data and error values. If problems exist in the original data, they can hardly be resolved by researchers using the data. Therefore, a series of data cleaning procedures are adopted to ensure the resulting data quality. First, the ambiguous or inconsistent values are unified or corrected. For example, a car’s brand or name containing “Benz” is corrected into “Mercedes”. Second, data values at different granularities are aggregated according to certain groups. For instance, annual sales regarding various car trims, a trim level representing the equipment levels in a specific car model, are aggregated to the model levels. Third, uncertain contents are largely abandoned. The raw collected data has more than six million images, but most of them are abandoned in later processes for the purpose of quality control.

Besides the aspects above, ethics issues are thoughtfully reviewed throughout the data preparation. We carefully pre-process and reproduce image contents to strictly comply with the GDPR (e.g., removing image background and covering plate numbers). It is worth noting that the UK government has special laws to encourage research studies related to creating and using web content based datasets³. Therefore, we do not have copyright concerns as long as people use it for non-commercial purposes. As our dataset only contains car-related information (e.g., images, car model specification, sales), it seems unlikely it will generate negative societal impact.

3.5 Dataset Application Examples

This section briefly illustrates three application examples for demonstrating how the DVM-CAR dataset could be applied to business research and applications.

3.5.1 Understanding Automotive Exterior Aesthetics Design

Product aesthetics design is a determinant of consumer acceptance and product success (Hoffer and Reilly, 1984; Bloch, 1995; Schoormans and Robben, 1997; Jindal et al., 2016). Marketing scholars have discussed aesthetics from various aspects, including the influence of aesthetics on product differentiation and new product development, and specific determinants of consumer responses to aesthetics. For example, morphing techniques were used to quantify and incorporate aesthetics design into empirical car sales models (Landwehr et al., 2011a; Tseng et al., 2013). This is an important step in modelling the effect of aesthetics design on sales. However, the used method is still limited in its ability to process image data. Car visual attributes are only extracted by pre-defined feature extractors, which are coarse-grained, and their respective implications for sales analysis and car appearance design are limited.

The recent advancement of machine learning technologies has provided marketing researchers with new tools to investigate product aesthetics. Several studies applied deep neural networks such CNNs to interpret the perceived design features (Pan et al., 2016; Burnap et al., 2016c). Compared with the traditional quantitative methods used in marketing research, deep learning algorithms can automatically learn high-level representations of visual features from car image data. Thus, the deep models can be used end-to-end, which solely requires raw images and tagging data as labels. Such advantages of deep models make them widely applied for visual based predictive studies (LeCun et al., 2015).

The DVM-CAR dataset provides an excellent base for researchers to apply deep learning

³<https://www.gov.uk/guidance/exceptions-to-copyright>

to extract visual attributes from car images. For instance, Fig. 3.5.1 (b) shows the sample using car images to infer the modernity score (i.e., labels computed from car registration years). By fine-tuning the ImageNet pre-trained VGG'16 (Simonyan and Zisserman, 2015), the results show deep learning models are capable of predicting the design fashion for family cars. Besides, we can further recognise the design language or patterns from the inside gradients. Fig. 3.5.1 (a) displays the predicted scores for Land Rovers. Although all these models are from the same automaker and were sold in the same year, the trained deep model rates them with entirely different modernity scores. The second row of Fig. 3.5.1 (a) presents, with the help of visualisation methods such as guided back-propagation (Springenberg et al., 2014), researchers can identify specific designs (i.e., highlighted in blue colour) that make the appearance outdated.

Based on the car models' modernity predictions, we can further investigate the associations between cars' modernity of appearance and their market performance by using the sales records in the DVM-CAR dataset. As Table 3.5.1 illustrates, we find the models' future withdrawn chances are correlated to their modernity scores (Note: the numbers in brackets of Table 3.5.1 are the sample sizes of the groups). The car group with low modernity scores have a higher chance of being withdrawn from the market, while the high modernity group tend to survive longer. As essential marketing variables such as new car prices and sales are all covered by the dataset, users can investigate the association between cars' exterior styling and market performance from diverse perspectives.

3.5.2 AI-Powered Automotive Exterior Design

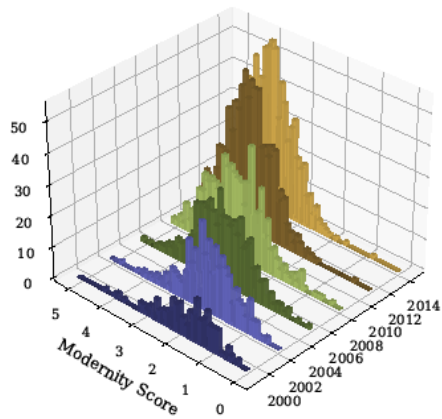
Anthropomorphism (Miesler et al., 2011; Waytz et al., 2014; Ku, 2014) refers to the attribution of human or animal characteristics to non-living objectives. As a common phenomenon in consumption environments, anthropomorphism has drawn wide attention from both social psychology and marketing researchers. Psychological studies (Purucker et al., 2014) reveal that our brain is highly specified for face perception, which is so evolved that we often perceive faces from non-living objects. A group of marketing studies extensively investigate how products with human-like or animal characteristics lead to face perception. Typically, in car related studies, existing investigations (Aggarwal and McGill, 2007; Landwehr et al., 2011b; Miesler et al., 2011) show consumers have a strong tendency to anthropomorphise the car front. However, in these studies, human or animal characteristics are predefined and only used in user surveys. They cannot be further deployed for facial feature recognition for unseen images. Thus, providing limited insights on car front appearance design for car manufacturers.

As a group of generative models with neural network structures, deep generative models can

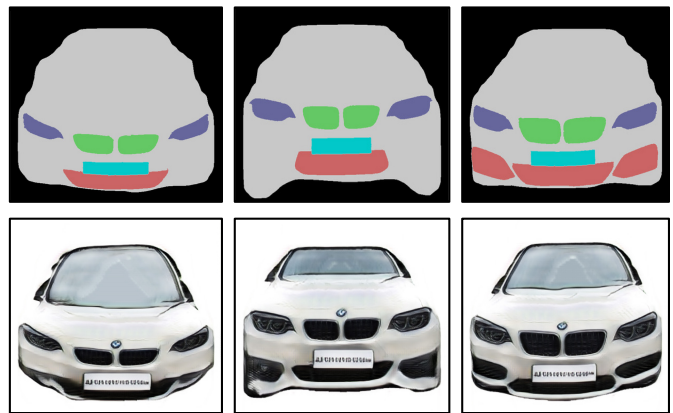
3.5. Dataset Application Examples



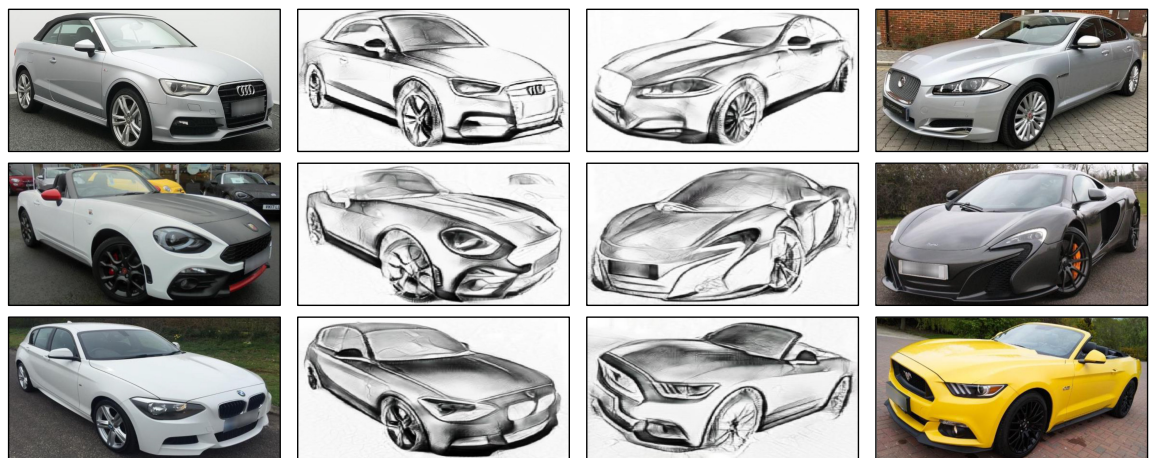
(a)



(b)



(c)



(d)

Figure 3.5.1: Illustration of application examples: (a) modernity scores of Land Rover models registered in 2015; (b) distribution of the predicted modernity scores of car models from 2000 to 2015; (c) new BMW front designs generated by CycleGAN according to given facial shapes; (d) automobile design sketches generated by CycleGAN.

Table 3.5.1: Withdrawn chances within the next three years of car groups with different modernity scores.

Year	Modernity score				
	[0,1)	[1,2)	[2,3)	[3,4)	[4,5]
2000	12.0%(25)	2.8% (36)	0.0% (10)	-(2)	-(0)
2003	23.1%(13)	10.7%(56)	0.0%(28)	0.0%(10)	-(0)
2006	-(4)	13.9%(36)	8.3%(60)	7.4%(27)	-(1)
2009	-(1)	33.3%(9)	16.1%(62)	3.3%(61)	15.8%(19)
2012	-(0)	-(1)	17.9%(28)	9.1%(88)	5.4%(56)
2015	-(0)	-(0)	18.2%(11)	9.0%(67)	7.3%(109)

learn to generate highly realistic representations from data, thus making them a hotspot in machine learning and statistics in past years (Goodfellow et al., 2014). These models have become extremely successful in the applications such as image translation and fake data generation. The 1.4 million car images of the DVM-CAR dataset make it extremely useful for generative model training. Two examples are provided in Fig. 3.5.1 (c) and (d). The former shows new BMW 3 Series front designs generated by cycle-consistent adversarial networks (in short CycleGAN) (Zhu et al., 2017) based on the fed semantic views. The latter presents sketches generated from real car images via the CycleGAN. These are from our working topics where we try to obtain bio-inspired designs by morphing the cars' layouts after cheetah faces.

3.5.3 Visual-Based Used Car Pricing

According to the Society of Motor Manufacturers and Traders (SMMT), around 8 million used cars are sold in the UK market each year.⁴ As these used cars have different specification such as mileage and maintenance conditions, their residual value prediction (or pricing) has become a challenging task (Englmaier et al., 2018; Huang et al., 2019). For example, the discontinuity nature of the used car prices was investigated (Englmaier et al., 2018) and the demand uncertainties have also been included in the item-specified pricing model (Huang et al., 2019). Fig. 3.5.2 displays the summed depletion surface for the 0.27 million collected used car records in DVM-CAR. As it reveals, our proposed DVM-CAR dataset allows business researchers to develop predictive models which take the used car's specification information to estimate its residual value.

Exterior aesthetics can be an important factor in used car pricing but aesthetics attributes (such as visual complexity or sporty) are predefined in the existing studies (Huang et al., 2019;

⁴<https://www.smm.co.uk>

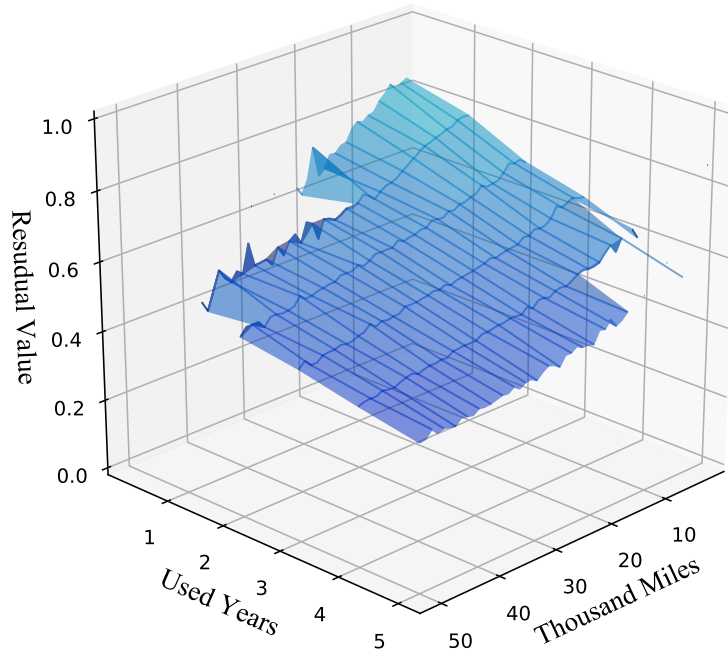


Figure 3.5.2: Simulated depletion surface of residual values.

[Englmaier et al., 2018](#)) since researchers lack comprehensive datasets which contain both image data and used car specification and sales data.

With the aid of the DVM-CAR dataset, the investigation of used car pricing becomes straightforward if appropriate visual features can be extracted and used as independent variables. We find metric learning algorithms, such as Siamese neural network ([Chopra et al., 1997](#)) and SphereFace ([Liu et al., 2017](#)) can be deployed for such tasks. Practical experience shows that the plain CNNs can be ineffective when trained on images sourced from numerous classes with similar layouts (such as faces or cars). In contrast, metric learning algorithms can learn the discriminative information for similar-looking object recognition by minimising the intra-class distance while maximising the inter-class separability. Previous studies have successfully extended these models on the photo and automotive aesthetics ranking and obtained promising results ([Kong et al., 2016](#); [Pan et al., 2017](#)) with large-scale data. Therefore, the scale advantage of the DVM-CAR can become a solid base to apply machine learning models to build predictive models for visual-based used car pricing, incorporating the exterior facts into the prediction.

3.6 Summary

This chapter demonstrates the design and development of a large-scale dataset for business research with online sources. On the one hand, this dataset, which is large and comprehensive in terms of product information, is a sound database for my later PhD studies. On the other hand, the development and sharing of this dataset have meaningful contributions to the research community, especially those interested in using artificial intelligence technologies. The targeted users of our dataset are business researchers and computer scientists who work in visual-related research and applications, particularly on (but not limited to) the topics of automotive exterior design, consumer analytics and sales prediction. It is worth noting that we would like to maintain and keep updating the dataset in the long term.

Chapter 4

GEO: Integrate Deep Generation with Profit Optimisation

This chapter proposes a unified computational framework – Generator, Evaluator, Optimiser (GEO) for the automotive exterior facelift to provide intelligent decision support to manufacturers and designers with the following three questions: (i) How to evaluate automotive designs in terms of aesthetics? (ii) How to perform regionally directed modifications on car image while leaving the unintended areas unchanged? (iii) How to estimate the long-term profit changes caused by design modifications? The rest of this chapter is organised as follows. Section 4.1 introduces the background and motivation for the GEO framework. Section 4.2 presents the technical details of each module in the framework. The first two parts of Section 4.3 describe the data and settings used in the study. Section 4.3's rest parts display the obtained experimental results and give discussions accordingly. Section 4.4 summarises this chapter.

4.1 Introduction

As mentioned in the literature review, a few existing studies have attempted to apply deep generative models to automotive design (Burnap et al., 2016c; Pan et al., 2017; Burnap et al., 2021). However, these existing studies deployed generative models to produce novel designs, in which design freedom constraints have not been considered. As Burnap et al. (2016a) pointed out, in real automotive markets, designers need to achieve a balance between design freedom and brand recognition.

Most automakers launch a new generation of a given car model every six to eight years.



Figure 4.1.1: Examples of car front facelifts (designs before and after the facelifts) for Audi A3 and Lexus NX.

This is a very long lifespan for a consumer good. Before the arrival of the next generation, automakers typically introduce some minor changes to an existing car model, which are known as *facelift* (or *mid-generational refresh*) (Blonigen et al., 2013). When consumers search for a new car, they may prefer to consider “the facelifted VW Golf” or “the facelifted BMW 3 Series”. Facelifts include upgrades to exterior styling, interior equipment, accessories, engine and safety options. For example, if a car model has a facelift, it may have a newly-designed front or rear bumper, LED lights or wheels, and the infotainment system may be upgraded with a bigger screen. Most of the time, a facelifted car will have a noticeably different look from the previous year’s model. Fig. 4.1.1 illustrates two real facelift samples from the market in the past. Facelifts have become an effective method of boosting a consumer’s interest in an existing car model before it is redesigned.

Thus, regional design upgrading, known as design facelift, is much needed in automotive exterior design. From the technical perspective, this is more challenging due to the feature entanglement problem (Locatello et al., 2019; Lomonaco et al., 2022). That is, the change of a single latent value can cause global changes in the resulting design. Furthermore, a theoretical analysis (Locatello et al., 2019) has indicated that the disentangled representations cannot be resolved through the unsupervised learning approach if no inductive biases are provided.

Building on the work from Pan et al. (2017) and Burnap et al. (2021), which apply deep

learning algorithms to assist the automotive exterior design, this paper proposes a computational framework for automotive exterior facelift, which provides intelligent decision support to automakers and designers. Our study aims to address the following research questions: Can a generative model be trained to present a design space for various automotive designs? How can such a model upgrade existing designs regionally with innovative design patterns? How can an upgraded design be evaluated? How can the profit shift caused by a given design change be estimated?

The newly developed style-based generators, namely StyleGAN (Karras et al., 2018) and StyleGAN2 (Karras et al., 2020), have performed outstandingly on the controllability of generation, while easing the entanglement problem. From the automaker's perspective, the ultimate goal of the facelift is to maximise the overall profit of the car models that have been launched, which could be seen as an optimisation problem. Despite the extensive study of optimisation problems in economics, game theory, and marketing studies, only a few existing studies have attempted to investigate product design from the optimisation perspective (Rajeev and Krishnamurti, 1987; Kaul and Rao, 1995; Shi et al., 2001). In these studies, simulation is based on high-level aesthetic attributes. Due to the lack of proper design generation methods, no realistic designs were provided as vivid samples. In our present study, we try instead to solve design optimisation directly on the various designs proposed by the generator.

To estimate the profit/utility caused by the design change, the evaluation of designs from the aesthetic perspective is required. Aesthetics, traditionally viewed as subjective, relate to the appreciation of beauty and artistry. This concept is intricate, influenced by individual preferences, cultural nuances, and historical contexts. The complexity of this makes predicting aesthetic sensations a challenge. Recently, with the advancements in deep learning technologies, researchers have tackled these challenges using data-driven approaches. By gathering crowd-sourced subjective ratings from individuals, machine learning models have been trained to predict the anticipated human response to specific visual stimuli (Kong et al., 2016; Ren et al., 2017; Pan et al., 2017). These days, it is common to apply convolutional neural networks (CNNs) (Lecun et al., 1998) for aesthetic evaluation tasks. CNNs are well-known for their end-to-end prediction power, which has achieved impressive record-breaking results on several computer vision tasks (Simonyan and Zisserman, 2015; Szegedy et al., 2015; He et al., 2016). Several existing studies (Teuwen and Moriakov, 2020; Madulid and Mayol, 2019; Seo and shik Shin, 2019) attempted to apply the CNN model to the categorisation of fashion products, achieving a remarkable accuracy. Researchers found that with sufficient images of cars and subjective labels, the CNN model could learn to make aesthetic ratings on diverse aspects such as sporty, appeal-

ing and innovative cars. Inspired by these studies, the development of an automotive aesthetic evaluator becomes relatively straightforward.

Motivated by the aforementioned advances, this chapter aims to develop a three-stage framework that can generate, evaluate and estimate profit changes of facelifted designs for various existing designs. Compared with the most relevant works, namely [Pan et al. \(2017\)](#) and [Burnap et al. \(2021\)](#), this proposed framework differs in two major aspects. First, this study focuses on offering regional design upgrades rather than new designs, where the primary shape of existing car models is untouched. Second, profit maximisation is incorporated into the pipeline, which attempts to locate the upgrades that lead to the highest profit. This makes our current study a product design optimisation investigation.

From a broader perspective, our study adds to the recent applications of machine learning and data mining in marketing and design studies. Different from previous studies, which have employed deep generative models for automotive exterior design ([Pan et al., 2017](#); [Burnap et al., 2021, 2016c](#)), our study is the first to investigate automotive facelift that focuses on regional design upgrades and considers the design selection from a revenue maximisation perspective. In terms of technology, the proposed computational framework provides an end-to-end decision-support solution for automakers and automotive designers. First, the design generator adopts a style-based generative adversarial network (GAN) ([Goodfellow et al., 2014](#)) architecture. By carefully selecting the latent space and training examples, innovative facelift designs for a car front can be generated while maintaining the car model’s family characteristics. Second, new aesthetic evaluation metrics are proposed to assess the car’s design analogous to the subjective human preferences. Last but not least, the decision optimiser tries to recognise the designs that can maximise mid-term revenues before the redesign of the target model.

4.2 The GEO Framework

4.2.1 Task Definition and Framework Overview

As shown in Fig. 4.2.1, our proposed framework is organised into three primary tasks, each representing a core challenge encountered by designers in the car’s conceptual phase.

Design generator—upgrading an existing car design: The goal is to enhance a current design while preserving most of its original elements. The design generator proposes various upgrading schemes for automotive exterior facelift. The applied model is based on the state-of-the-art StyleGAN2 ([Karras et al., 2018, 2020](#)) where we reform the mapping network and the projection

method to generate new facelift designs. Therefore, the design generator can generate many innovative designs on the targeted car front areas but keeps the “family” face of a given car model. For instance, a facelifted BMW 3 Series can have a more aggressive front bumper and LED headlights while still possessing the well-known BMW “kidney” grille.

Design evaluator—estimating the aesthetics of a car exterior design: While aesthetics is traditionally a subjective term associated with sensations of beauty and artistry, in our study we define “aesthetic” as a score indicating the visual appeal of a car’s appearance, derived from a data-driven approach that employs user reviews focused on the car’s exterior. The design evaluator is responsible for rating the designs from the aesthetic perspective. As modern car designs have similar shapes and layouts, it is difficult for the regular deep models, with limited samples, to learn discriminating features for aesthetic ratings. Inspired by previous studies (Kong et al., 2016; Pan et al., 2017) that solve the problem through the metric learning approach, a double-task training frame is proposed in this paper, incorporating an angular loss-based classification (Liu et al., 2017; Wang et al., 2018; Deng et al., 2019) to facilitate the learning of discriminating features.

Design evaluator—estimating profit changes from design modifications: Predicting the potential financial implications of various design changes is essential. However, few studies have ventured into this territory. In our study, the decision optimiser is tailored to choose the most effective facelift plans, aiming to maximise profits over time. We utilise a recurrent neural network (RNN) (Cho et al., 2014) to assess market share shifts associated with facelift adjustments. By analysing historical data linking car facelift aesthetic alterations to corresponding sales shifts, the model is trained to anticipate market share changes triggered by different aesthetic variations. The optimiser subsequently assesses the prospective advantages of different facelift plans, opting for those yielding the greatest potential returns.

4.2.2 The Design Generator

Our generator is based on the StyleGAN2 model (Karras et al., 2020) as it is the state-of-the-art deep generative model that has performed extremely well in synthesising high-resolution images. Meanwhile, compared with other GAN models, it has a flexible and enriching architecture that allows automotive designers to control design types, which results in more innovative designs.

As Fig. 4.2.2 shows, a StyleGAN2-based generator has two parts: a mapping network and a synthesis network. The mapping network, denoted by $f_{\text{MAP}}(\cdot)$, is implemented through l fully-

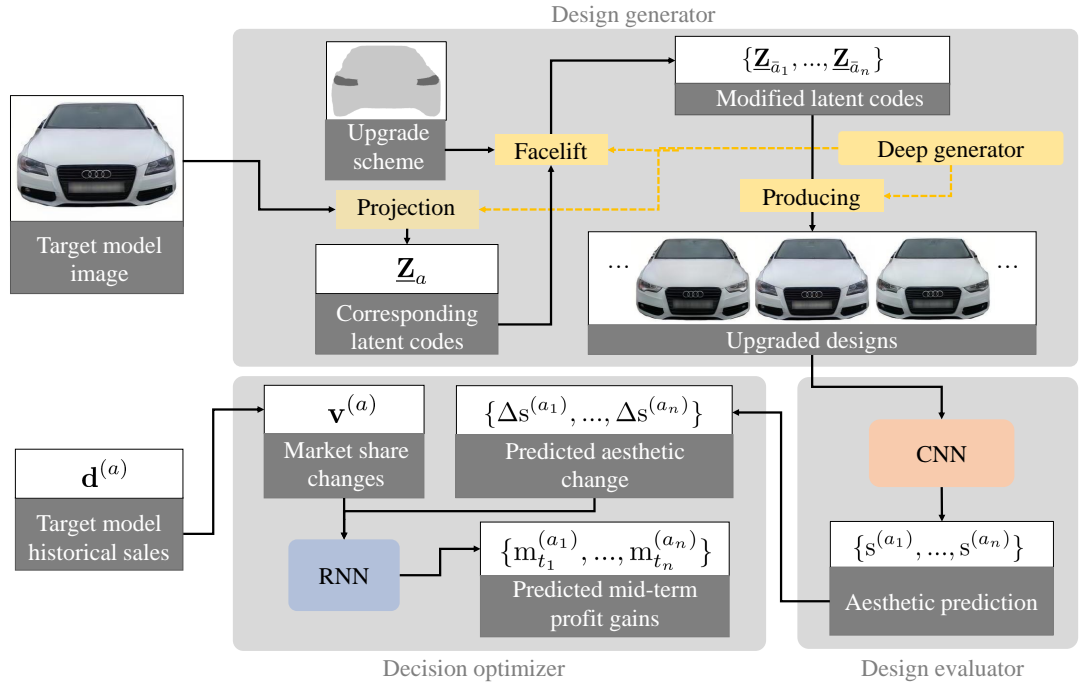


Figure 4.2.1: Schematic view of the GEO framework.

connected layers, responsible for mapping an input latent code $z^{(0)}$ to an intermediate code $z^{(l)}$. The superscript indicates the position of the latent code within the mapping network. For instance, the intermediate latent code $z^{(l)}$ is the output from the l th layer. The synthesis network, denoted by $G_{ST}(\cdot)$, comprises K generation blocks. As blocks or convolutional layers can take different intermediate latent codes to produce images, we denote all the fed latent codes by a matrix \underline{Z} . This is different from the original StyleGAN2 study (Karras et al., 2020), since the fed intermediate latent vectors in our design framework are not required to be the same across the generation blocks or convolutional layers.

Also illustrated by Fig. 4.2.2, each generation block (excepting the first one) in StyleGAN2 consists of two convolutional layers and one upsampling layer. To simplify, the computation of a single block is formulated as $O_k = g_{ST}(O_{k-1}, \underline{Z}_{[k]})$, where O_k represents the output feature maps from the k th generation block, $g_{ST}(\cdot)$ denotes the computation of the entire generation block, and $\underline{Z}_{[k]}$ represents the latent code fed to the k th generation block. As the right part of the synthesis network in Fig. 4.2.2 shows, unlike the original StyleGAN model, StyleGAN2 replaces the progressive growing strategy with “skip connections”, then the final output image

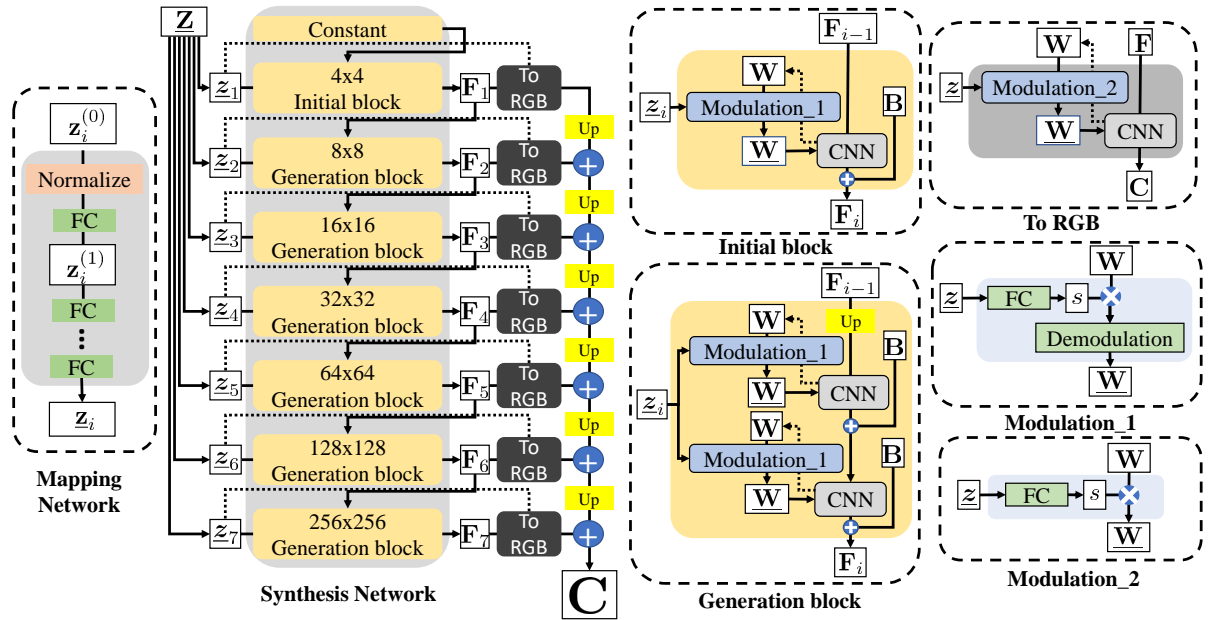


Figure 4.2.2: Structure of the generator in StyleGAN2. The StyleGAN/StyleGAN2 models comprise two main parts, the mapping network and the synthesis network. The former transforms the raw latent code to the intermediate latent code, while the latter generates images/designs according to the given intermediate latent codes.

is the sum of all the generation block results:

$$G_{ST}(\underline{Z}) = \sum_{k=1}^K f_{UP} \left(f_{RGB}(O_k), K - k \right), \quad (4.1)$$

where $f_{RGB}(\cdot)$ represents the “To RGB” module that converts feature maps to images, each feature map has the size of $2^{k+1} \times 2^{k+1}$, $f_{UP}(\cdot)$ is the upsampling function, and $K - k$ indicates the times needed to double the size, ensuring output channels from different blocks are all sized in $2^{K+1} \times 2^{K+1}$.

After a proper training, the resulting StyleGAN2 can be perceived as a design space for various automotive images. By searching for the latent codes that produce the most similar results, a given car design C_a can be represented by a distinct \underline{Z}_a through the inverse of the generation process $G^{-1}(\cdot)$, namely the projection method:

$$G_{ST}^{-1}(C_a) = \underset{\underline{Z}}{\operatorname{argmin}} D_{PER}(G_{ST}(\underline{Z}), C_a), \quad (4.2)$$

where $D_{PER}(\cdot)$ is the distance measure between images. It should be noted that for the projec-

tion, the obtained \underline{Z} can consist of latent codes from arbitrary layers of the mapping network. With regard to measuring the distance between images, we adopt learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018), which is a CNN-based score that measures the perceptual difference between two images.

To indicate the generator model’s ability to produce novel and realistic designs, we propose two new metrics related to the concept of design space, namely *domain size* and *domain quality*. The former evaluates the generator’s ability to generate innovative designs, and a larger-sized design space would contain more unseen designs. In the study, we apply the calculation of *projection accuracy* to measure the domain size, which is formulated as follows:

$$\mathbb{E}_{C \sim \mathcal{X}_{\text{test}}} D_{\text{PER}} \left[G_{\text{ST}} \left(G_{\text{ST}}^{-1}(C) \right) - C \right], \quad (4.3)$$

where C is an unseen design drawn from the test set $\mathcal{X}_{\text{test}}$, and $G_{\text{ST}}(G_{\text{ST}}^{-1}(C))$ is the reconstruction of C in the given design space. The projection accuracy measures how accurately an unseen design is represented in the learned generative system. Intuitively, if a generative system has a large design space for cars, it should be able to precisely represent unseen designs.

The domain quality metric indicates the quality of design generations. The quality problems have been extensively studied in deep generative models, where the Fréchet inception distance (FID) (Heusel et al., 2017) metric is widely used. We adopt a modified FID, namely *FID of random mixing*, to sample the overall generation quality since our proposed facelift tries to shift an existing latent code partially to another. It is formulated as follows:

$$\mathbb{E}_{\underline{z}_i, \underline{z}_j \sim P_z} f_{\text{FID}} \left(G_{\text{ST}}(\underline{z}_{z_i, z_j}), \mathcal{X}_{\text{train}} \right), \quad (4.4)$$

where $f_{\text{FID}}(\cdot)$ denotes the FID measure, $\mathcal{X}_{\text{train}}$ is the training set, and \underline{z}_{z_i, z_j} consists of vectors that result from the random mixing of \underline{z}_i and \underline{z}_j .

Inspired by the architecture of StyleGAN2, it is expected that regional design upgrades could be achieved by revising the corresponding \underline{Z} . Given a candidate design C_a and an intended upgrading scheme B_a (which is a binary matrix that indicates the image area to modify, it can have an arbitrary size and shape in the given image), our design generation objective can be formulated as follows:

$$\max_{\underline{Z}} [S_{\text{AE}}(G_{\text{ST}}(\underline{Z})) - D_{\text{PER}}(B_a \odot C_a, B_a \odot G_{\text{ST}}(\underline{Z}))], \quad (4.5)$$

where $S_{\text{AE}}(\cdot)$ is the aesthetic evaluation carried out by the evaluator, and \odot represents the Hadamard product. Then, we look for a latent matrix \underline{Z} that can maximise design aesthetics while mitigating the modifications of the unintended areas. The latter can be the areas related to the family signature of a given car model such as the BMW kidney grille. For convenience, in the rest of the paper we will refer to these unintended areas as the *fixed region*.

It is worth pointing out that, from the automotive facelift perspective, Eq. (4.5) is not a proper objective function. First, automotive designers would prefer to have more candidate designs rather than a single “best” suggestion (Burnap et al., 2021). Second, high-level modifications (e.g., shape) are preferred over fine-feature changes (e.g., colour, texture). If we do not specify further constraints, the upgrades will end up with changes to the fine-features rather than to the overall structure.

To obtain more candidate designs, Eq. (4.5) can be reformulated as a two-stage optimisation problem. In the first stage, we look for designs inspired by other latent codes while minimising the changes in the fixed region. Given a candidate design C_a and an inspiration latent matrix \underline{Z}_i from the set $\{\underline{Z}_1, \dots, \underline{Z}_N\}$, we look for a design that satisfies the following objective:

$$\min_{\underline{Z}} \left| \underline{Z}_{[:k]} - \underline{Z}_{i[:k]} \right| + D_{\text{PER}} \left(B_a \odot C_a, B_a \odot G_{\text{ST}}(\underline{Z}) \right), \quad (4.6)$$

where k indicates the target latent code places, $\underline{Z}_{[:k]}$ are the intermediate codes given to the first k blocks, and $|\underline{Z}_{[:k]} - \underline{Z}_{i[:k]}|$ is the L^1 distance between the design latent matrix and the i th inspiration latent matrix. We adopt the L^1 norm since in our trials it achieves more stable results than the Euclidean distance (Isola et al., 2017). Previous studies have shown that early generation blocks control general-features and later ones regulate the fine-features (Karras et al., 2018, 2020). Therefore, constraining the latent codes allows us to decide what types of design features to modify. Let \mathcal{Z} denote the set of designs from solving Eq. (4.6), we then rank the designs obtained according to their aesthetic scores and select the best candidates accordingly:

$$\underset{i}{\operatorname{argmax}} S_{\text{AE}}(G_{\text{ST}}(\underline{Z}_i)), \text{ for } \underline{Z}_i \in \mathcal{Z}. \quad (4.7)$$

4.2.3 The Design Evaluator

As modern car designs share similar layouts and forms, the design evaluator needs to learn discriminating features to distinguish between the proposed designs and facelifts. Inspired by the existing metric learning studies (Chopra et al., 1997; Kong et al., 2016; Pan et al., 2017) and

facial recognition/prediction studies (Eisenthal et al., 2006; Leyvand et al., 2008; Gray et al., 2010; Liang et al., 2018), where similar data challenges are faced, the double-task training strategy is adopted. The evaluator is trained simultaneously for aesthetic estimation and car model recognition to facilitate the learning of contrasting features between designs:

$$\min \mathbb{E}_{a \sim \mathcal{X}_{\text{train}}} \left[\ell_{\text{aes}}^{(a)} + \ell_{\text{rec}}^{(a)} \right], \quad (4.8)$$

where $\ell_{\text{aes}}^{(a)}$ and $\ell_{\text{rec}}^{(a)}$ are the aesthetics estimation and class recognition loss, and the $\ell_{\text{aes}}^{(a)}$ is expressed as $(s^{(a)} - \underline{s}^{(a)})^2$, representing the Mean Squared Error (MSE) between the predicted and ground truth aesthetic score. Unlike the traditional softmax, the $\ell_{\text{rec}}^{(a)}$ adopts the angular loss (Liu et al., 2017; Wang et al., 2018; Deng et al., 2019) setting, further enhancing the learning of discriminating features in classification. Given the design C_a under car model y_a , and x_a as its corresponding feature vector extracted from the convolutional backbone network, the j th classification values before softmax can be expressed as $h_{\text{rec}}^{(a,j)} = w_j^\top x_a + b_j$, where w and b represent the output weight vector and bias values, respectively. These variables are constrained in the angular loss setting: $\|w_j\| = 1$, $b_j = 0$, and $\|x_a\| = \alpha$, where α is a given constant. This makes each $h_{\text{rec}}^{(a,j)}$ only depends on the angle size between w_j and x_a , and converts the whole softmax computing into the following formula:

$$\ell_{\text{rec}}^{(a)} = -\log \left\{ \frac{\exp(\alpha \cos(\theta_{y_a, a} + \beta))}{\exp(\alpha \cos(\theta_{y_a, a} + \beta)) + \sum_{j \neq y_a} \exp(\alpha \cos \theta_{j, a})} \right\}. \quad (4.9)$$

Here we adopt Deng et al. (2019)'s setting, which incorporates a constant margin penalty β in the target class's angle, making $h_{\text{rec}}^{(a, y_a)} = \alpha \cos(\theta_{y_a, a} + \beta)$, thereby facilitating further discriminative feature learning by making the negative log-likelihood more sensitive to the angular distances. These settings force the model to use the vectors' directional differences rather than scale the differences to distinguish between classes, thus representing the class centres in the angular space. The use of angular loss has been empirically validated to maximise intra-class distance and minimise inter-class distance on tasks with comparable inputs (Liu et al., 2017; Wang et al., 2018).

4.2.4 The Decision Optimiser

The decision optimiser is proposed to select designs that maximise the expected mid-term revenues for the automaker before the redesign of the given car model. We use market share changes to measure and approximate revenue changes. Like other durable products, car models

have a typical life-cycle: a new car’s sales increase in the early years when a new generation is launched and then deteriorate over the rest of its lifespan until the facelift or redesign occurs (Moral and Jaumandreu, 2007). Based on this knowledge, we treat sales deterioration as a time-series process and thus rely on the RNN model to estimate how the future market share would evolve according to different facelift plans.

Given an existing design C_a and its historical market share records $\{d_0^{(a)}, \dots, d_T^{(a)}\}$, we compute its market share change vector $v_T^{(a)} := \{v_1^{(a)}, \dots, v_T^{(a)}\}$ according to the formula: $v_t^{(a)} = d_t^{(a)}/d_{t-1}^{(a)}$. The predicted share change $\tilde{v}_t^{(a)}$ at year t can be expressed as $f_{\text{MS}}(v_{t-1}^{(a)}, \Delta s_t^{(a)})$, where $f_{\text{MS}}(\cdot)$ represents the predictive RNN model and $\Delta s_t^{(a)}$ indicates the aesthetic change due to the design modification at year t . The model is simply trained to minimise the perdition error $\mathbb{E}_{a,t \sim \mathcal{X}_{\text{train}}} [v_t^{(a)} - \tilde{v}_t^{(a)}]^2$. After training, to estimate the market share changes driven by a suggested facelift $C_{\hat{a}}$ that launches at year η , the estimation setting becomes as follows:

$$\tilde{v}_t^{(\hat{a})} = \begin{cases} f_{\text{MS}}(\tilde{v}_{t-1}^{(a)}, \Delta s^{(\hat{a})}), & \text{if } t = \eta \\ f_{\text{MS}}(\tilde{v}_{t-1}^{(a)}, 0), & \text{otherwise} \end{cases}, \quad (4.10)$$

where $\Delta s^{(\hat{a})}$ represents the aesthetic scores change of the facelift design, and 0 means no design/aesthetic changes. It is worth noting that here we use $\tilde{v}_{t-1}^{(a)}$, not $v_{t-1}^{(a)}$, to denote the input vector since here inputs are also simulated results (except for $t = 1$ where $v_0^{(a)}$ is based on actual records).

Considering that the model will be redesigned at $T + 1$, we can adopt the baseline $\tilde{v}_T^{(a)}$, which represents the actual facelift to infer the overall profit change:

$$m_{\eta}^{(\hat{a})} = d_0^{(a)} \cdot \sum_{r=1}^T \left[\prod_{t=1}^r \tilde{v}_t^{(\hat{a})} - \prod_{t=1}^r v_t^{(a)} \right], \quad (4.11)$$

where $m_{\eta}^{(\hat{a})}$ represents the overall share difference when adopting facelift $C_{\hat{a}}$ at year η , and $d_0^{(a)} \cdot \prod_{t=1}^r v_t^{(a)}$ represents the predicted market share at year r .

4.3 Experiments

In this section, we introduce the datasets used, present the experimental settings for model training and testing, and discuss the analysis of the results.













	Audi A4 Saloon 218,847 (19.6%) "Not a particularly daring design" www.parkers.co.uk/audi/a4		Peugeot 2008 SUV 72,298 (2.0%) "Awkward styling" www.parkers.co.uk/peugeot/2008		Vauxhall Astra Hatchback 674,362 (4.9%) "Ordinary looks" www.parkers.co.uk/vauxhall/astra
	VW Jetta Saloon 23,170 (2.1%) "Dull Image" www.parkers.co.uk/volkswagen/jetta		SKODA Yeti SUV 76,749 (2.1%) "Not to everyone's taste" www.parkers.co.uk/skoda/yeti		BMW 1 Series Hatchback 312,354 (2.3%) "Styling not to everyone's taste" www.parkers.co.uk/bmw/1-series
	Peugeot 508 Saloon 17,651 (1.6%) "Exterior design is bland" www.parkers.co.uk/peugeot/508		Toyota RAV4 SUV 69,125 (1.9%) "Not as stylish as rivals" www.parkers.co.uk/toyota/rav4		Honda Jazz Hatchback 233,144 (1.7%) "Dull styling" www.parkers.co.uk/honda/jazz
	VW Phaeton Saloon 1,060 (0.1%) "Comparatively dull styling" www.carbuyer.co.uk/volkswagen/phaeton		Ford EcoSport SUV 41,438 (1.1%) "Looks awkward" www.parkers.co.uk/ford/ecosport		Hyundai i10 Hatchback 206,492 (1.5%) "Lacks appeal" www.parkers.co.uk/hyundai/i10

Figure 4.2.3: Targeted car models for facelift exterior design: the yellow masks highlight the areas to be modified by our proposed framework; the first row shows the car model name; the second row shows the car model type; the number in the third row shows the total sales of the car model from 2007 to 2017 and the number in the brackets shows the car model's market share (i.e., the division of the model's sales to the total sales of the model's segment market); the fourth row shows the critiques of the car model's exterior design; the fifth row provides the reference (i.e., URL) of the consumer reviews.

Table 4.3.1: Summary of the used datasets.

Dataset	DVM-CAR [†]	Edmunds [‡]
Period	2001-2020	2000-2019
Main content	1,451,784 images, 773 model sales	299,045 reviews
Number of automakers	60	46
Number of car models	899	905

[†] <https://deepvisualmarketing.github.io>

[‡] <https://www.kaggle.com/shreemunpranav/edmunds-car-review>

4.3.1 Datasets

Fig. 4.2.3 presents our targeted car models for facelift exterior design. We deliberately select car models of three popular types (i.e., hatchback, SUV and saloon), which have received criticism for their exterior styling¹. The aim of the experiments is to improve the market performance of the targeted models by suggesting good exterior designs in the facelift. For simplicity, the given scheme samples (indicated by the yellow masks in Fig. 4.2.3) focus on modifying the design of the headlights since they are the most frequently upgraded features in automotive refreshments. As shown in Table 4.3.1, two publicly available datasets are used in the experiments. The **DVM-CAR** dataset contains 1.4 million images from 8 different viewing angles of 899 car models as well as the corresponding model specification and sales information over more than ten years in the UK (Huang et al., 2022). For the generator training, 42,130 car front images are sampled from the DVM-CAR dataset to develop the design generator, evaluator and optimiser in the proposed GEO framework. The **Edmunds** dataset contains 299,045 car reviews of various automotive brands from edmunds.com between 2000 and 2019. Unlike many previous studies, which rely on surveys or lab experiments to collect subjective ratings, we estimate the car design aesthetic ratings by text mining the available consumer reviews. The aesthetic ratings obtained in this manner, together with the DVM-CAR dataset, are used to develop the design evaluator. For the aesthetic ratings extraction, the car reviews on models sold between 2007 and 2017 are extracted. After pairing the extracted aesthetic ratings with the car images from the DVM-CAR dataset, 15,213 images from 118 car models are used to train the evaluator.

4.3.2 Experimental Settings

We use the StyleGAN2 model for the facelift design generation (Karras et al., 2020). The generator is set with 7 generation blocks for 256×256 resolutions as well as 90% mixing regularisation. 2,000 out of 42,130 images are used for testing and the rest are used for model

¹E.g. see www.parkers.co.uk/audi/a4 regarding Audi A4

training. To enhance the generator’s upgrading ability on the target designs, each image of the targeted models is also augmented by rotating and flipping it with 20 replications. During the training procedure, we track the model’s performance using the FID and the Perceptual Path Length (PPL) (Karras et al., 2018) metrics. When there are no significant gains on the FID score, the training is stopped. The images in the test set and the twelve targeted designs are embedded into the latent space through the projection method, where both the MSE and the LPIPS metrics are included in the reconstruction loss part. We remove the optimisation of random noises during the projection since they would overlap their roles in latent spaces with small sizes. The facelift is implemented as a variation of the projection method, where the starting latent code is the one that represents the original design.

To prevent the overfitting problem, we adopt a 5-fold cross-validation in the evaluator training. Specifically, images are grouped according to their car models. This protects the trained evaluator from the bias of their model recognition. For the decision optimiser, we take each car model’s annual share in the segmented market (by body type) to indicate the market performance over time. The car models’ historical market shares and the aesthetic shifts over the years are used as inputs for RNN training, where the gated recurrent unit (GRU) (Cho et al., 2014) is used.

4.3.3 Analysis of Results

As shown in Fig. 4.3.1, the projection accuracy and the FID scores of random mixing are calculated across different latent space settings. We compare two different mapping network architectures: (i) 512 dimensions and 8 layers (Karras et al., 2018); and (ii) 2048 dimensions and 3 layers. In the “global unique z ” setting, all the generation blocks use one latent code. In the “block specified z ” setting, each generation block has a specific latent code, hence there are 7 different latent codes in total. In the “module specified z ” setting, each modulation has a latent code. Since the first block has only one modulation and the “To RGB” module shares the same latent code with the second CNN module in the block, there are 13 different latent codes in total.

Compared with the default latent space (Karras et al., 2018, 2020), different intermediate codes offer various choices for design subspaces characterised with different domain sizes and qualities. There is a clear trade-off between the projection accuracy and the FID score of random mixing when adopting different intermediate codes exists. Latent codes from an earlier layer (e.g., $z^{(0)}$) would have better FID scores but lower projection accuracy than a later layer (e.g., $z^{(7)}$). We interpret this trade-off with the concept of design space. The output space corresponding to $z^{(8)}$ (the intermediate latent codes) can be seen as a design space with a large

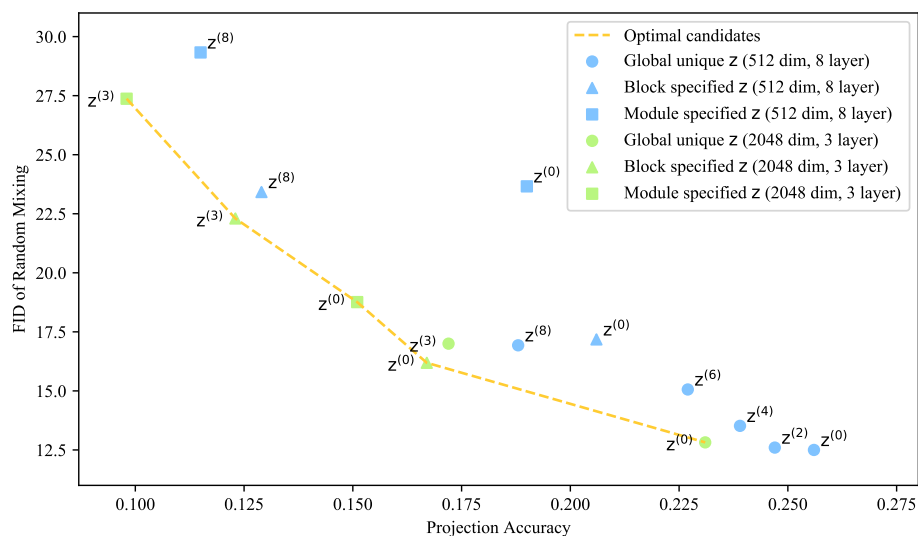


Figure 4.3.1: Comparison of candidate latent spaces, where the superscript indicates the latent layer position among the mapping network.

size. The stacking of additional mapping network layers (such as $z^{(7)}$, $z^{(6)}$) can reduce the space size, thereby resulting in denser subspaces for the car designs that increase the quality of the generated designs (i.e., with lower FIDs). On the other hand, Fig. 4.3.1 also shows that the use of incongruous z (i.e., the “block specified z ” and “module specified z ”) results in an increased projection accuracy and decreased FID scores. In particular, we suggest that the “disassociation of latent codes” (i.e., using different latent codes in different generation blocks) allows more novel designs that were not seen in the training set to appear in the test set, even though such novelty comes at the price of reduced quality. For example, if there are no white SUVs in the training set, then white SUVs would barely appear when using the global unique z setting, but the incongruous z would generate SUVs with various colours.

As Fig. 4.3.1 shows, a new setting of the mapping network (2048 dim, 3 layers) can improve the domain size as measured by the projection accuracy while retaining a similar domain quality measured by the FID score of random mixing. This new setting is inspired by the universal approximation theorem where a shallower network with more neurons in the hidden layer can also have a high approximate power. Hence, the new mapping network has fewer layers but reserves more space for novel designs. Overall, the results of our experiment confirm that a broader but shallower mapping network allows the synthesis network to generate further novel designs with higher qualities.

Table 4.3.2: Comparison of our design upgrading algorithm with existing methods.

Method	Mapping setting	Projection accuracy	FID of random mixing
Style-based (Liu et al., 2020)	8×512	0.120	60.60
Our intermediate-based [†]	8×512	0.190	23.66
Style-based (Liu et al., 2020)	3×2048	0.119	91.57
Our intermediate-based	3×2048	0.151	18.76

[†] Here the module specified $z^{(0)}$ is used for comparison

In Table 4.3.2, we further compare our design upgrading algorithm with the method proposed by Liu et al. (2020), as their study also develops a regional modification algorithm using StyleGAN. Unlike our intermediate latent-based method, their method directly manipulates style variables in StyleGAN. The reported results show that random modifications in style variables lead to distortions (indicated by the high FID scores) in the outputs, while our intermediate latent-based method can retain the design quality. This suggests that our method has better capabilities for generating unseen designs when exploring the design space, and hence is more suitable for the needs of the automotive facelift.

Fig. 4.3.2 provides an empirical example of an image matrix that compares the latent mixing results at different generation layers. When the inspiration latent codes are fed to the 1st-4th blocks (illustrated by the top four rows), the car’s overall design structure would be changed. When the 5th-7th blocks are fed with the inspiration codes (illustrated by the bottom three rows), the changes mainly happen to the texture and colours. Since a structural change is more preferred for automotive facelifts, our implemented facelift method would focus on modifying the first four generation blocks while leaving the rest unchanged.

As mentioned earlier, car front design aesthetics are evaluated by the design evaluator in terms of aesthetic scores. Fig. 4.3.3 provides several empirical examples of car front images with the predicted aesthetic scores, in which three car models receive low aesthetic scores (i.e., Toyota Aygo, Hyundai ix20, Fiat Punto) while the other three receive high scores (i.e., Volvo S90, BMW M2, and Jaguar XF). These results are consistent with the users’ aesthetics reviews collected in the Edmunds dataset. To illustrate this, we list a few examples of reviews as follows:

“...can’t disguise its dated design, and the Punto looks bland ...making Fiat’s Punto look really rather old...”²

²URL:www.autoexpress.co.uk/fiat/punto/interior



Figure 4.3.2: Comparison of car front designs of replacing different generation block latent codes, where a column is the replacement starting block and a row is the replacement end block.

“...the ix20 rather blends into the crowd compared to its more stylish rivals...”³

“...the BMW M2 is a car that’s huge fun from behind the wheel, stupendously quick and relatively low-key in its subtle appearance...”⁴

“The Jaguar XF...that car pulled Jaguar into the 21st century, rejecting the classic design language that had characterised the brand’s models since the fifties by replacing round headlights with sleek fastback looks and an aggressive new grille...”⁵

The first plot of Fig. 4.3.4 shows the average aesthetic score of the cars sold over the years. We find that the cars on the market are steadily becoming more aesthetic. It should be noted that the average rise of the aesthetic score is 0.008 per year, which is used in our decision optimiser’s hyper-parameter setting. We carry a simple linear regression analysis to investigate how the car models’ aesthetic levels are associated to their market performance (see the right-hand plot of Fig. 4.3.4). In particular, the second plot of Fig. 4.3.4 shows that the car models’ market share

³URL:heyicar.co.uk/hyundai/ix20

⁴URL:www.autoexpress.co.uk/bmw/2-series/105480/used-bmw-m2-review

⁵URL:www.carbuyer.co.uk/jaguar/xf



Figure 4.3.3: Empirical examples of car front aesthetic scores predicted by the trained evaluator.

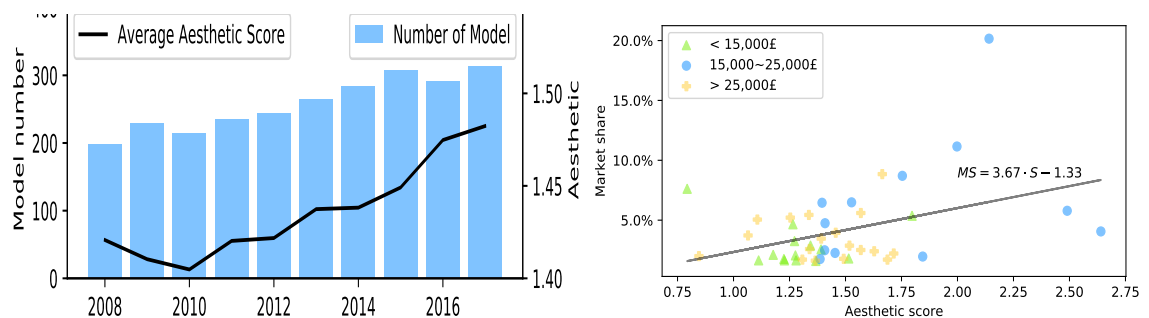


Figure 4.3.4: Analysis of the aesthetic scores: (left) Time series plot of the average aesthetic score and the sales of the targeted car models from 2008 to 2017; (right) Aesthetic score VS market shares. By comparing the model’s market shares and their aesthetic scores in recent years, we find that the market share is correlated to their aesthetic levels, where the Pearson test results are as follows: $\rho = 0.39$ with a p -value=0.009.

is significantly correlated to their aesthetic levels. Interestingly, car models with higher prices do not appear to be more aesthetic, which is not in line with our expectations.

Table 4.3.3 presents several design examples from three target models (i.e., Audi A4, Ford EcoSport, Vauxhall Astra) as well as the summarised statistics for all the targeted models. In order to perceive the variation caused by the design difference, for each model, two novel facelift designs (i.e., an inferior design and a superior design) are presented together with their original looks, and compared with their predicted market shares before the redesign (first row in the table). The original A4, EcoSport and Astra facelift designs are rated as 1.394, 1.330 and 1.208, respectively. Overall, the compact segment benefits the most from the proposed facelift designs,

Table 4.3.3: Examples of optimised designs.

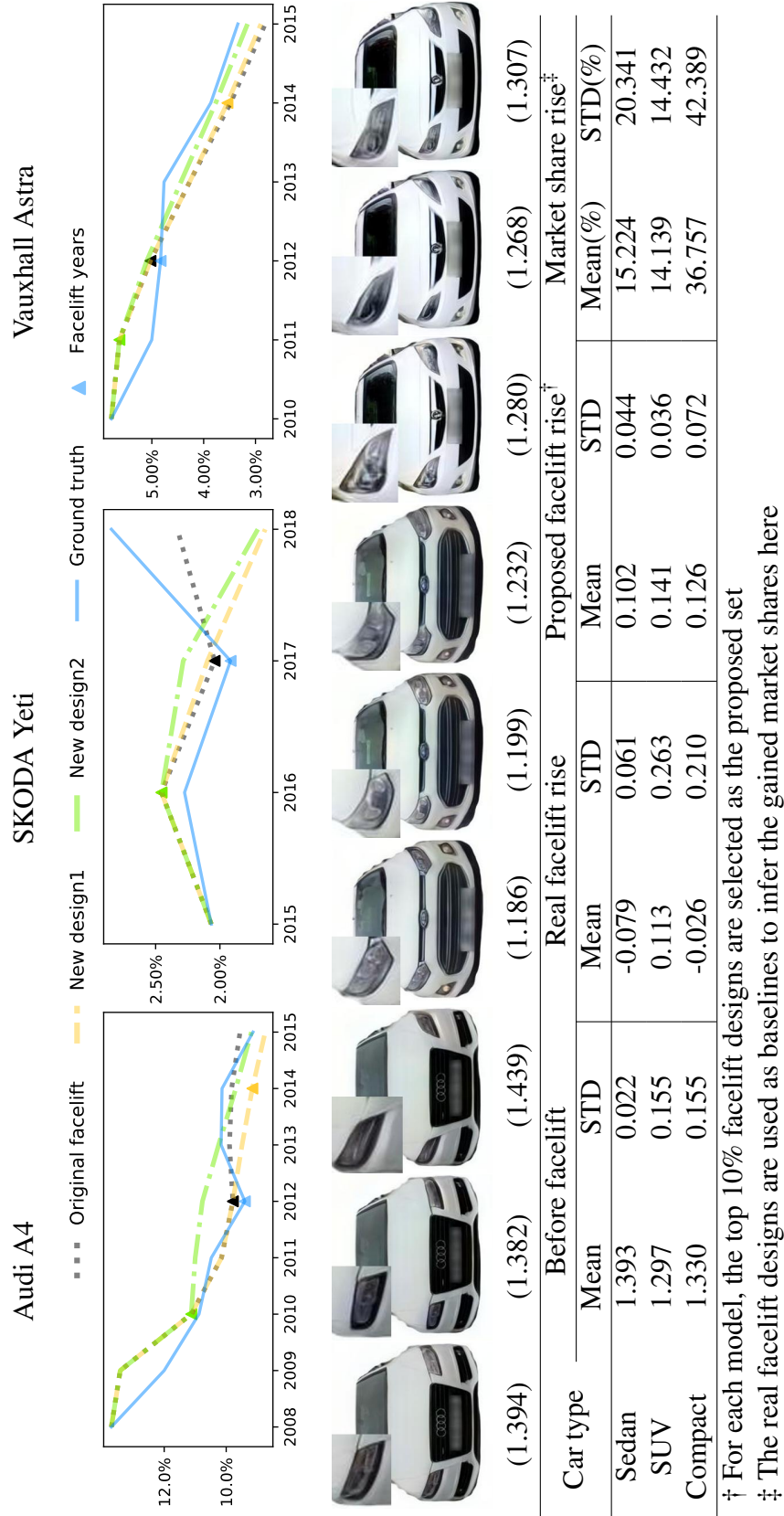


Table 4.3.4: Number of facelifts released in ten years.

Number of facelifts	10	5	3	2
Total profit	94.61%	94.49%	94.35%	94.22%

The profit is calculated by $\frac{\text{Estimated 10 year sales}}{10 \times \text{Starting year sales}}$.

with a 0.126 aesthetic rise, resulting in an average 36.757% increase in market share. However, the variance between the individual models is huge – most of the increment is contributed by Astra, which has an expected market gain of 209.555%. Based on the proposed facelift designs, SUVs have a significant aesthetic rise (0.141) but a moderate market increase (14.139%). Compared to the ground truth, the increase is not as significant as the aesthetic rise since the real facelift designs are remarkable – for instance, the Ford EcoSport has upgraded from 1.186 to 1.330. When comparing years for a suggested facelift for superior and inferior designs, the optimiser tends to delay the facelift for the weak designs. For instance, the new design 1 of A4 (i.e., scored 1.382) is suggested for a facelift in 2014, but its new design 2 (i.e., scored 1.439) is suggested to be facelifted in 2010.

Fig. 4.3.5 investigates the effects of the facelift interval and aesthetic change on the car model market share in the mid-term. The first plot of Fig. 4.3.5 shows that for facelifts with identical rising aesthetic scores, higher market shares can be expected if the facelifts are performed earlier, but such strategies can backfire, with a steeper market share deterioration in the long term. On the other hand, as demonstrated in the second plot of Fig. 4.3.5, according to our simulation of various facelifts, the stronger facelift will always result in higher mid-term gains, while lower aesthetic scores will lead to a mid-term loss.

Benchmarked with the expected annual aesthetic rise of 0.008, Table 4.3.4 investigates the optimal facelift frequency without cost constraints. Consistent with the observation that most automakers in the market release facelifted models annually or biennially, the optimal strategy is to have more frequent facelifts before the redesign of a given car model.

4.4 Summary

In this chapter, we have developed a new machine learning-based framework (i.e., GEO), which can assist automakers when it comes to the cars' exterior design. Unlike the existing works, this study focused on the scenario of automotive facelifts, which delivers regional upgrades for launched car models and views the design selection from a profit optimisation perspective. The proposed generator and facelift algorithm can incorporate novel styling features

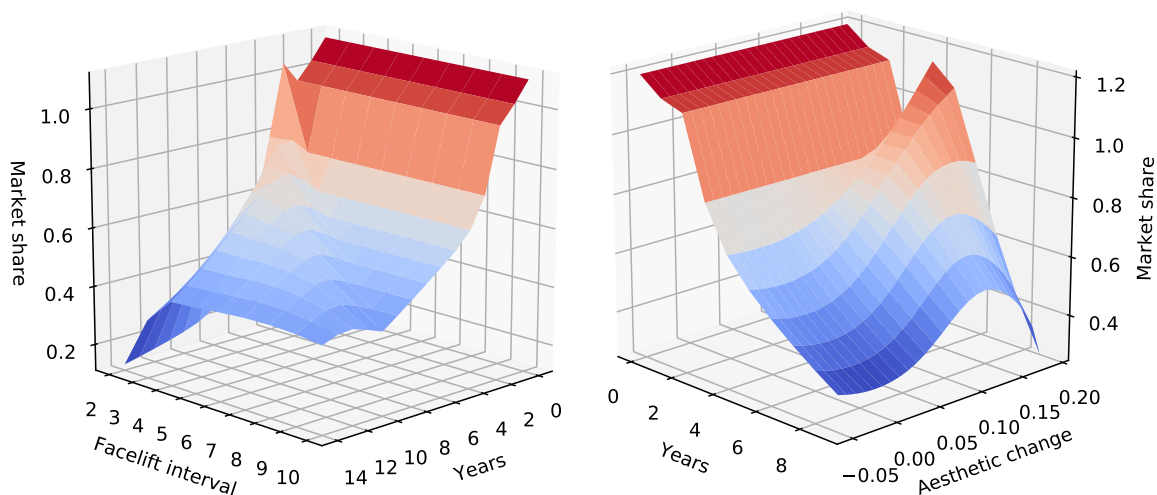


Figure 4.3.5: Effects of facelift: (left) mid-term market share evolution when adopting different facelift years. (right) market share simulation according to aesthetic changes.

into existing designs while maintaining the primary look of the car model. The metric learning-based evaluator can ease the challenge of evaluating the aesthetics of objects when they look similar, as the angular-loss guides the algorithm to focus on more discriminating features. For profit optimisation, based on the proposed aesthetic changes, the RNN-based optimiser simulates mid-term profit changes as a result of the released facelifts, hence providing automakers with intuitionistic support for decision making. In addition to the proposed GEO framework, we showed that the most desirable properties for design spaces are the domain size and quality. The former relates to the generator’s ability to generate novel designs, while the latter determines the quality of the generation. We proposed two corresponding metrics to these perspectives, which can be used as benchmarks for future product aesthetic studies. Finally, we explored how to improve the design space size and quality based on the StyleGAN2 model. We found that the rearrangement of mapping networks or the selection of different latent spaces can improve the StyleGAN2 generator’s performance in terms of both domain size and domain quality. Overall, our proposed framework provides automakers with a technology allowing them to better manage their facelift design process.

However, this study does naturally have some limitations, which can be addressed in later chapters. First, our study does not discuss how to achieve spatial disentanglement. However, deep generative-based design upgrading is strongly related to the feature disentanglement problem. In fact, product exterior design upgrading can be perceived as a crucial application scenario for the disentanglement problem. Second, we consider the market’s preference for car aesthet-

ics as static in this study, while in the real market, such preference keeps evolving with time. The study in the coming chapter will try to address these limitations by offering a solution for predicting future fashion trends in the automotive market, making our investigations to be more comprehensive in research.

Chapter 5

Trendiness: Design for Upcoming Fashion

This chapter introduces a four stage machine-learning system to revamp the outdated front views of cars. A main innovation is to have the system learn, define and score trendiness from thousands of historical car front images, which in turn produces massive unrepresented but highly scored modern designs. The rest of this chapter is organised as follows. Section 5.1 introduces the motivation for this chapter's study. Section 5.2 defines the concept of trendiness score within this study. Section 5.3 explains the new four-staged framework for automotive design upgrading. Section 5.4 describes the used data and experimental settings. Section 5.5 discuss the results and findings, and Section 5.6 summarise the chapter.

5.1 Introduction

As costly durable goods, cars are apparently under the influences of fashion trends, which indicates a life cycle of aesthetics and thus sales volume for the designs in the markets. Compared with quarterly and even monthly design updates of non-durable goods such as clothes, the adjustment speed of fashion or trendiness of motorcars is perceived to be much slower. According to [Reynolds \(1968\)](#), the establishment of lower and lengthy augments of cars took nearly three decades from the 1930s to the 1960s. And to the recent two decades, the aggressive appearance such as slant headlights are slowly adapted by the mainstream outfit designer of cars. The modernity/trendiness related terms are frequently used by consumers to describe product appearance ([Blijlevens et al., 2009](#)). The consequence of a rigid design would be losing market share and even die out in a short span of time. In fact, half of the newly launched cars were expected to be dropped off at most eight years according to [Moral and Jaumandreu \(2007\)](#), while

the most recent update such as Greim (2017) suggests that this product lifespan has been significantly shortened to an average of six years in the past decade. A classical life cycle of the sale volume of cars always consists of a peaked market shares and the afterward trough when getting outmoded and unattractive.

To catch up with fashion changes, upgrading an existing model or adopting a new successor product line are the most prevalent choices for automakers when facing outmoded car designs (Ernst et al., 2017). Through mid-generational refreshes, the exterior styling of extant designs could be upgraded while maintaining the general shapes and reserving existing production equipment. Even though the refreshes provide automakers additional chances for remedy, the aesthetic design for cars remains challenging. A failed judgement or estimation of future style trends not only brings actual economic cost and associated financial risks but will also harm the goodwill of the brand in the market for a non-trivial duration. Therefore, it is a real-world economic and business decision to predict aesthetic trends in the vehicle market and update designs with the evolution of fashion trends. However, conventional approaches that rely on surveys have the major limitation that the intuitive judgments or subjective predictions from participants are much biased by their personal experiences and perspectives. The historical unsuccessful models criticised for their weird designs have evidenced the limitations of individual predictions (Hekkert et al., 2003; Cadavid et al., 2016), though many of these came from the most experienced and talented professionals.

Motivated by the aforementioned facts, the study presented in this chapter has two main objectives. The first goal is to develop a metric that indicates how closely a design aligns with future car fashion trends. Recent studies indicate that deep models, when trained with aesthetic feedback from human participants, can effectively predict human aesthetic responses to visual stimuli (Lu et al., 2015; Kong et al., 2016; Ren et al., 2017; Pan et al., 2017). Although individual participants may have varying preferences and habits on a micro-level, the overall market assessment of modernity or trendiness is more macro and aggregate in nature. Given this, we aim to develop an efficient data-driven method that uses longitudinal observations of car designs to score a design's alignment with upcoming fashion trends."

Another critical goal of this chapter is to propose an alternative deep generative method for design upgrading with no feature entanglement issues. Different from the previous chapter, this study uses a novel design upgrading approach that employs the image completion algorithm. Image completion or image inpainting refers to the task of predicting the omitted content in an image based on the remaining information Bertalmio et al. (2000, 2003). It is widely used

for scenarios such as photo editing or de-captioning. Before the rise of deep learning models, image completion mainly relied on diffusion or patch-based methods, which usually assumed the missing content shares significant similarities with the surrounding areas (Bertalmio et al., 2003; Ballester et al., 2003; Barnes et al., 2009). Thus fill the missing area with similar colour and texture. The rise of deep learning models enables the models to generate more conceivable content in terms of semantics (Ren et al., 2015; Zheng et al., 2019; Zhao et al., 2020). By learning the content distribution from similar images, the data-driven completion models can fill the missing area with appropriate content that is very different in colour or texture from the surrounding. It matches our needs for design upgrading, which can replace outdated designs with more stylish looking.

In contrast to previous chapters, this chapter has several key contributions to the overall thesis. First, it introduces a dynamic approach for assessing car aesthetic ratings. The aesthetic measurement in Chapter 4 treated market preference for car exteriors as a static, time-independent factor. However, this doesn't reflect the real world, where aesthetic preferences constantly evolve. To address this, we developed an ordinal regression CNN model that can index designs according to their alignment with future aesthetic trends. This model can learn and predict the trendiness of car front visuals, helping identify outdated car models.

Second, this chapter presents a design upgrading method free from feature entanglement issues. The GAN-based design upgrading method proposed in Chapter 4 is limited by feature entanglement problems, where changes in certain regions can inadvertently cause changes in unintended areas. To overcome this, the design upgrader in this chapter utilises an image completion technique, generating a variety of images of the renovated designs. This assists automakers in their decision-making process.

Third, in contrast to the GEO framework, which estimates potential profit gains resulting from design upgrades, this chapter introduces a new module that aims to rank candidate upgraded designs focusing solely on the exterior perspective, but through a multi-objective aspect. Specifically, we transform the selection process into a Pareto problem, wherein we try to identify designs that are situated on the Pareto fronts for both aesthetic and trendiness dimensions. This approach could be regarded as a complementary approach to the profit estimation module, as it enables the sorting of designs based on pure exterior styling considerations.

5.2 Problem Statement

For clarity and precision, it is essential to have well-founded definitions of key concepts utilised in this study. This section introduces the definitions of a few crucial concepts relevant to “trendiness”.

Traditional approaches to describing the appearance of objects often begin by breaking down the overall impression of a product into various attributes, resulting in a mix of both physical and abstract aspects of its appearance. The term **aesthetic attribute** refers to a description or measurement of a specific visual trait that results from the combination of physical properties, as commonly adopted by previous research (Veryzer Jr, 1995; Hsiao et al., 2008; Blijlevens et al., 2009). Examples of aesthetic attributes include symmetry, colorfulness, and modernity, which have been frequently referenced in existing literature as indicators of a product’s visual characteristics. However, there is no guarantee that aesthetic attributes need to be independent of each other. In fact, many of them might be correlated. For instance, the attribute “sporty” could be similar to “dynamics”. From another aspect, the possible number of aesthetic attributes is likely to be vast, as it could include any arbitrary combination of basic physical traits.

Given the limitations of “aesthetic attribute”, dimension reduction methods, like Principal Component Analysis (PCA)¹, naturally emerge as potential solutions. The **principal components** in PCA encapsulate most of the variance in the original high-dimensional space while being orthogonal to each other. Essentially, these components represent a fusion and transformation of the original dimensions. Thus, if a significant volume of car aesthetic ratings across diverse aesthetic attributes is accumulated, then the most valuable aesthetic dimensions (principal components) could be obtained. Various car designs can then be represented as unique coordinates within an **aesthetic space** crafted from such aesthetic dimensions.

Fig. 5.2.1 visualises an imagined aesthetic space for car exterior designs, composed of three principal components. In this visualised aesthetic space, colours represent the production years of car designs. Those from the same year cluster closely due to the highly matured modern car market, where automakers, influenced by similar fashion ideologies during the same time period, adopt different but aesthetically similar styles for their products. Moreover, the central point of the car design distribution does not undergo dramatic or arbitrary changes; rather, it

¹PCA is a widely used dimensionality-reduction method to capture the major variance of high dimension data. As dimensional features in such data are highly covaried, where the PCA algorithm can use much fewer transformed features to capture the majority of variances (i.e., the information). The transformed features are termed principal components that rank in the order of the variance that they presented.

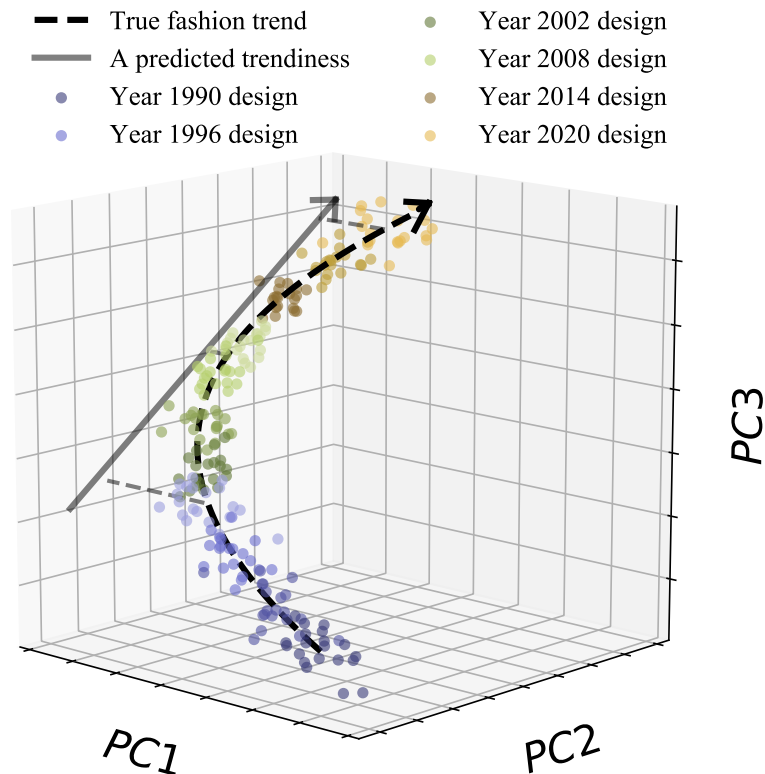


Figure 5.2.1: Illustrations of aesthetic space, fashion evolution and the proposed trendiness measure. Dots represent the sold car designs in the market. The three dimensions are named PCs 1-3 to analyse the principal components of the principal component analysis.

demonstrates a propensity to shift in a direction over a more prolonged duration. The concept of **trendiness direction**, as used in this study, refers to the direction in which the design distribution shifts at a particular point in time. Additionally, the **trendiness** represents the projection of a car design onto the trendiness direction. A higher trendiness degree indicates that the design aligns more closely with the prevailing trend.

The assumption that car design distribution would progress in a specific direction over an extended period is supported by existing studies, which affirm that motorcar fashion undergoes a gradual yet steadfast evolution. For instance, the gradual shift documented by Reynolds (1968) reveals that it took over thirty years for the market to self-perpetuate and transition towards lengthier and lower vehicles, a transformation that occurred between the 1930s and the 1960s. Extant studies (Zajonc, 1968; Reynolds, 1968; Hekkert et al., 2003) tend to describe the development of such a trend as a mutually reinforcing cycle where the designers attempt to meet the perceptions of their targeted consumers by adjusting the aesthetic attributes, while the launched

products will further reinforce consumers' preference towards the shifted/updated attributes.

With the defined concept of trendiness in car exterior design, this study is focused on four interrelated tasks. Initially, the development of a predictive model is targeted, which can indicate the degree of alignment of a particular car with the trendiness direction. Subsequently, the objective is to create a model that can pinpoint the areas of the exterior design contributing to a low trendiness rating. The third task involves the development of a deep generative model capable of regionally upgrading existing designs, thus, generating novel designs with a more contemporary appearance. Lastly, the goal is to formulate a selection method that identifies the candidate designs with the most appealing exterior styling.

5.3 Proposed Methods for Design Evaluation and Design Assistant

Fig. 5.3.1 presents a schematic view of the whole proposed approach. It is comprised of four main modules: (i) the first module access the trendiness of the designs by learning from car front images over the decade so that the outmoded car models can be identified; (ii) the second module decomposes image via neural network visual features, recognises and highlights the antiquated design regions that need to be renovated; (iii) the third module produces a range of renovated designs where the identified antiquated regions are upgraded; (iv) the last module is responsible of ranking the produced upgrades from aesthetics, quality and trendiness perspectives.

We first develop a prediction model that can discern the trendiness of designs across periods. The issue can be addressed by using the cars' production years as an index for trendiness. It is easily motivated by the linguistic definition of trendiness as the longer interval that the design was made from the present, the higher probability the design is likely to be outdated. The designs produced more recently are likely to carry more modern exterior features.

5.3.1 Stage I: Trendiness Evaluator

Over an extended period, a car's production year inherently reflects its design's trendiness; it is unlikely for the design of a car manufactured a decade ago to be more fashionable than a recently produced one. Hence, when a large collection of car images from various eras is available, the production years naturally serve as appropriate labels to denote trendiness. However, the "trendiness" intervals between any two consecutive years are not always constant and equivalent. Hence product years should not be directly used as the dependent variable for regression

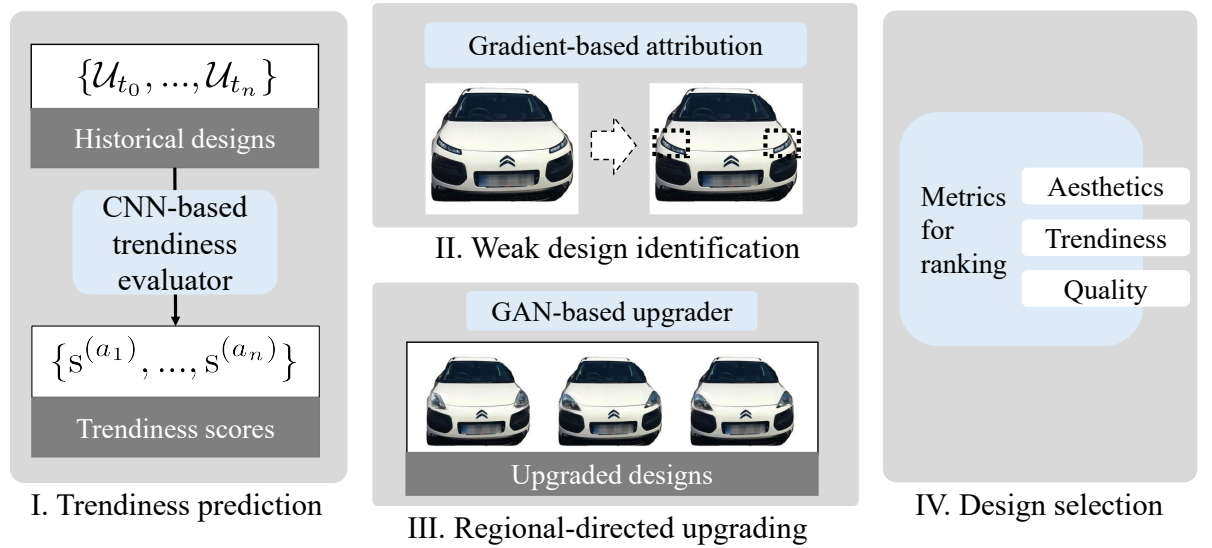


Figure 5.3.1: Schematic view of the proposed design assistant approach - (i) Trendiness Evaluator: a CNN-based ordinal regression for trendiness prediction; and (ii) Design Upgrader: image completion-based region-directed design upgrading.

models. Moreover, employing the years of production as categorical variables in classification will be based on an unrealistic strong assumption that largely dismiss the ordinal information. To overcome these issues, we then predict the trendiness score based on an ordinal regression model, which the yearly intervals are treated as proxies of contemporaneity differences that are not necessary numerically equivalent.

The core set-up of the ordered probit model, which assumes the existence of a “latent dependent variable”, is consistent with our defined trendiness. That is, there exists an unobserved function $f_{\text{TREND}}(\cdot)$ denoting the trendiness, and “determines” which production year a design should be assigned to. A higher score of $f_{\text{TREND}}(\cdot)$ indicates that the processed design has higher degrees of matching to the predicted trendiness for the year. In standard applications, such latent variable is just a tool for constructing the probabilities with the desired correlations without requiring actual meanings. While in this study, the trendiness score is an averaged estimation across images, expressed as:

$$f_{\text{TREND}}(d_a) = \mathbb{E}_{x_n \in d_a} [\boldsymbol{\beta}^\top v_n + \varepsilon_n], \quad (5.1)$$

where d_a is the a -th target design ($\forall a \in [1, A]$), here x_n represents an image of d_a , ε_n is the corresponding residual variance, v_n denotes the vector of aesthetic attributes extracted from x_n . In our study, v_n is extracted from a pre-trained CNN that with parameters $\boldsymbol{\theta}_{\text{CNN}}$. In a nutshell,

Eq.5.1 states that the trendiness score of the a -th target design is a fitted value of linear ordinal regression based on a vector of aesthetic attributes extracted from the set of images of the design (as the mean of the error term (ε_n) is expected as zero).

Two assumptions are made. First, we assume the error term ε follows a standard normal distribution for the designs introduced at the same year. Second, we assume the trendiness of designs is monotonically increasing across the periods around t . That means, given \mathcal{U}_{t_1} and \mathcal{U}_{t_2} represent two sets of car model designs produced at times t_1 and t_2 (Note: $t_2 > t_1$), for $d_{t_1} \in \mathcal{U}_{t_1}$ and $d_{t_2} \in \mathcal{U}_{t_2}$, it is expected that $\mathbb{E}[f_{\text{TREND}}(d_{t_2})] > \mathbb{E}[f_{\text{TREND}}(d_{t_1})]$.

Assuming that the car front designs from a set of J ordered but noncontinuous years can be characterised by J levels of trendiness, we have $J - 1$ threshold values, determined during training, given by $\boldsymbol{\lambda} := \lambda_1, \dots, \lambda_{J-1}$ to distinguish the J levels of trendiness.² Given a trendiness score $f_{\text{TREND}}(d_a)$, it will be assigned to the j -th trendiness level if $\lambda_{j-1} < f_{\text{TREND}}(d_a) \leq \lambda_j$. For an uncategorised design d_a , the probability that it belongs to j -th level can be expressed as:

$$\Pr(d_a \in \mathcal{U}_j) = F(\lambda_j - f_{\text{TREND}}(d_a)) - F(\lambda_{j-1} - f_{\text{TREND}}(d_a)), \quad (5.2)$$

where $F(\cdot)$ is the standard cumulative density function, and the training objective of the CNN is:

$$\max_{\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}_{\text{CNN}}} \mathbb{E}_{d_a \in \mathcal{U}_j} [\Pr(\lambda_j \geq f_{\text{TREND}}(d_a) \geq \lambda_{j-1})]. \quad (5.3)$$

Training and using of the trendiness evaluator. During model fitting, the optimal f_{TREND} and threshold values $\boldsymbol{\lambda}$ are determined based on their ability to best replicate the observed outcomes in the training data. Given specific f_{TREND} parameters, the $\boldsymbol{\lambda}$ values influence the likelihood of observing a particular ordinal category. By fine-tuning these parameters, the model adjusts the associated probabilities. The optimisation process then identifies parameter values that maximise the likelihood of reproducing the observed data. The specific settings used for training and testing are detailed in Section 5.4.

After obtaining the trendiness evaluator, we can label the designs with trendiness scores. Specifically, we assort the same year designs into three groups according to their trendiness scores, denoted as $\mathcal{U}_{t,l}$, $\mathcal{U}_{t,m}$, and $\mathcal{U}_{t,h}$, which represent designs sold in year t with low, medium, and high trendiness scores.

² $\lambda_0 = -\infty$ and $\lambda_J = +\infty$.

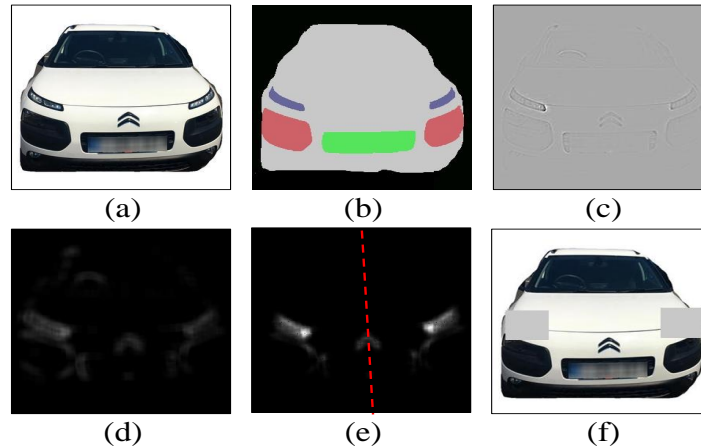


Figure 5.3.2: An example of antiquated design recognition: (a) the original image; (b) the recognised semantic area; (c) the saliency map from the Guided Backpropagation; (d) the saliency map after thresholding and blur; (e) the balanced saliency map; (f) the image with covering mask.

Properties of f_{TREND} as a trendiness score. As the latent variable from the ordinal probit model, the f_{TREND} represents the market’s “true” unobserved response regarding trendiness/modernity. This response is indirectly inferred from the shifts in market design styles over the years. Being a continuous ordinal variable, the intervals between its values are not uniform, which means there are no inherent meanings for 0 (i.e., no true zero point) and negative values. The range of the obtained f_{TREND} is unrestricted and can span any real number, depending on the model’s training duration. Theoretically, it can range from $-\infty$ to $+\infty$, but in practice, due to the limited number of epochs used to train the model, most of the design scores is centred around zero, with values typically ranging from single-digit negative to single-digit positive. Only a handful of designs exhibit very large or small values; however, this simply indicates that the presented designs have very high probabilities of belonging to the most modern or outdated groups.

5.3.2 Stage II: Antiquated Design Analytics

The primary task when upgrading extant designs is to identify the vulnerable regions that lower the trendiness score. Hence, a method that integrates the instance segmentation and gradient-based saliency map is proposed to locate these antiquated areas. First, the instance segmentation is applied to mask the critical design features. Due to the anthropomorphism phenomenon, historical studies show that the way consumers perceive certain design parts of car front designs is analogous to the perception of actual facial features of human faces, such as the headlights to the eyes (Welsh, 2006; Aggarwal and McGill, 2007; Windhager et al., 2008).

Modifying such semantically significant parts could substantially affect the perception of designs (Landwehr et al., 2011b). For this purpose, the instance method the Mask R-CNN (He et al., 2017) is applied to mark K non-overlapped semantic areas from each front image, denoted as:

$$\{B_{n,1}, \dots, B_{n,K}\} = f_{\text{MRCNN}}(x_n), \quad (5.4)$$

where each $B_{n,k}$ is a binary matrix that indicates the k -th corresponding semantic areas in x_n . Fig. 5.3.2 (b) illustrates an obtained semantic view.

Second, a saliency map is generated for each image of a vehicle to inform the importance of each semantic area. The resulted map is filled with gradient values between the final prediction and individual pixels. The SmoothGrad (Smilkov et al., 2017) and the Guided Backpropagation (Springenberg et al., 2014) are combined to generate saliency maps. The Guided Backpropagation omits the negative values during the backpropagation, thus highlighting areas most influential to the trendiness score. While the SmoothGrad method works by having multiple samplings that add random noises to the input image, therefore minimising the error caused by local gradient variance. The entire backpropagation process is formulated as:

$$\tilde{J}^{(n)} = \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, \sigma^2)} [-\partial f_{\text{TREND}}(x_n + \varepsilon) / \partial (x_n + \varepsilon)], \quad (5.5)$$

where $\tilde{J}^{(n)}$ denotes the raw saliency map obtained from the n -th car image, and ε here is a random error following the standard normal distribution. It is worth noting that the negative sign of Eq. 5.5 indicates that the weight and bias parameters in the trendiness predictor's last layer are inverted since we are looking for parts that contribute negatively to the trendiness. Fig. 5.3.2 (c) represents an obtained raw saliency map.

Following that, several image process methods are applied to further rinse the obtained saliency maps. Fig. 5.3.2 (d) and (e) perform the further rinsed saliency map after thresholding, blurring and balancing. These operations reduce the “noises” (i.e., random low value gradients) on the raw saliency map. As the obtained saliency maps could be imbalanced in many cases, a SIFT-based symmetry line detection method (Loy and Eklundh, 2006) is applied, which can produce two reflected saliency maps. The Hadamard product of these two saliency maps, $J^{(n)} := J_{\text{left}}^{(n)} \odot J_{\text{right}}^{(n)}$, is used for later computation.

The final step in this stage is to combine the semantic and gradient results to determine the most vulnerable parts. For design d_a that has N_a images available, the following selection

equation is used:

$$\operatorname{argmax}_i \frac{\sum_n^{N_a} \exp(\phi(J^{(n)} \odot B_{n,i}))}{\sum_n^{N_a} \sum_k^K \exp(\phi(J^{(n)} \odot B_{n,k}))}, \quad (5.6)$$

where $\phi(\cdot)$ presents the grand sum of the matrix. The whole equation is a softmax operation, which converts the problem as probability dependent decision. Thus, we simply choose the design part with the largest probability. Fig. 5.3.2 (f) demonstrates that the headlights are computed as the most vulnerable design parts for the target model. Each target design image x_n will be split into the masked part $x_n^- = x_n \odot B_{n,i^*}$ and the unmasked regions $x_n^+ = x_n \odot (\mathbf{1} - B_{n,i^*})$, here $\mathbf{1}$ indicate the matrix of ones.

5.3.3 Stage III: Image Completion-Based Design Upgrading

The study adopts the image completion approach to upgrade the previously identified out-moded parts. This approach will predict the omitted parts based on the perceived contents. With the given images $x_h \in \mathcal{U}_{t,h}$ and $x_l \in \mathcal{U}_{t,l}$ representing the ones drawn from the t -th year high and low trendiness groups, the scraped parts in antiquated designs x_l^- will be replaced with design features drawn from x_h . We formulate this objective function as a maximisation problem of the probability of $p(x_h^+ | x_l^-)$, which aims to optimise the probability of filling modern designs with the given x_l^- .

Inspired by the recent image completion studies (Wan et al., 2021; Yu et al., 2021; Zheng et al., 2021), the study divides the maximisation issue into dual tasks:

$$\begin{aligned} & p(x_h^+ | x_l^-) \\ & \approx p(x_h^+ | \tilde{x}_h^+) \cdot p(\tilde{x}_h^+ | \tilde{x}_l^-) \cdot p(\tilde{x}_l^- | x_l^-) \\ & = p(x_h^+ | \tilde{x}_h^+) \cdot p(\tilde{x}_h^+ | \tilde{x}_l^-) \\ & \approx p(x_h^+ | \tilde{x}_h^+) \cdot p(\tilde{x}_h^+ | \tilde{x}_h^-). \end{aligned} \quad (5.7)$$

where \tilde{x}_h^- is the thumbnail of x_h^- . The $p(\tilde{x}_h^- | x_h^-)$ is omitted for two reasons. First, it is well standardised to adopt the resizing of an image to its thumbnail during image processing. Meanwhile, the conditional entropy between the image and its thumbnail $H(\tilde{x}_h^- | x_h^-) = 0$, and $p(\tilde{x}_h^- | x_h^-) = 1$. For implementation, the simulation of $p(\tilde{x}_h^+ | \tilde{x}_h^-)$ uses the transformer network, while $p(x_h^+ | \tilde{x}_h^+)$ adopts the deconvolutional network.

Transformer for inference. The transformer network (in our case, the GTP-2 algorithm) is

applied to infer the missing parts, due to its superior performance in dealing with inter-dependent information. The transformer network was initially designed for the NLP tasks, the model has shown outstanding abilities for image processing, even though it mandates images in 1-d vectors (so 2-D images have to be flattened to 1D vectors for processing).

Via the attention mechanism, the model can compute attention scores to identify the critical dependencies among inputs. A transfer model is comprised of multiple blocks, given $X^{(l)} \in \mathbb{R}^{C_i \times n}$ (C_i and n represent the input's dimension and object/length number) as the input to l -th block, the computation of the block is:

$$X^{(l)} = f_{\text{BLOCK}} \left(f_{\text{SA}}(X^{(l-1)}) + X^{(l-1)} \right), \quad (5.8)$$

where $f_{\text{BLOCK}}(\cdot)$ presents a series of operations, including the LayerNorm and fully connected layers, to process the attention score. $f_{\text{SA}}(\cdot)$ presents the key computation of the self-attention score, denoted as:

$$f_{\text{SA}}(X^{(l-1)}) = W_o [f_{\text{ATTE}}^{(1)}(X^{(l-1)}) \frown \dots \frown f_{\text{ATTE}}^{(M)}(X^{(l-1)})], \quad (5.9)$$

where “ \frown ” denotes the concatenate operation, $W_o \in \mathbb{R}^{C_i \times M C_k}$ represent the learnable matrix that merge the scores from M “heads”. The computation of each “head” can be expressed as:

$$f_{\text{ATTE}}^{(m)}(X) = W_v^{(m)} X \left(f_{\text{SOFTMAX}} \left(\frac{(W_q^{(m)} X)^\top (W_k^{(m)} X)}{\sqrt{d_k}} \right) \right), \quad (5.10)$$

where $W_q, W_k, W_v \in \mathbb{R}^{C_k \times C_i}$ are algorithm learnable matrices, $W_q X$, $W_k X$ and $W_v X$ represent the query, key and value matrices. Such a unique self-attention mechanism permits the wide applications of the transformer models these days.

Deconvolutional network for rescale. The deconvolutional network is used to rescale the generated thumbnails to original sizes to fulfil the function of $p(x_h^+ | \tilde{x}_h^+)$. The deconvolutional network is titled due to the deconvolution layers, more formally known as transposed convolutional layers, which possess the same computation power of a vanilla convolutional layer but enlarges the inputs differently. They add padding between matrix entries rather than the surroundings. A transposed convolutional layer can generate output feature maps with larger sizes than input feature maps, thus commonly used to upsample the inputs. By following [Wan et al. \(2021\)](#)'s work, we train the deconvolutional network under the the generative adversarial net-

work (GAN) (Goodfellow et al., 2014) frame, which has the following objective:

$$\min_G \max_D \mathbb{E}_{x_i \sim P_{data}} [\log(D(x_i))] + \mathbb{E}_{x_n, \tilde{x}_n, x_n^+ \sim P_{data}} [\log(1 - D(G(\tilde{x}_n, x_n^+))) + \|x_n - G(\tilde{x}_n, x_n^+)\|_1], \quad (5.11)$$

where $D(\cdot)$ is the discriminator network, $G(\cdot)$ is the generator network and $\|\cdot\|_1$ indicates the L1 norm. Specifically, the L1 distance is added to enhance the reconstruction of the original image. Meanwhile, the \tilde{x}_n and x_n^+ are fed as the input to the generator network rather than the random latent variable.

5.3.4 Stage IV: Ranking of Candidate Designs

Applying deep generative models for an existing car model could deliver almost unlimited upgrade designs, which are diverse in image quality and design trendiness. This leads to a new challenge: how should we pick out the most suitable designs from such a large number of possibilities? A simple yet effective solution is to use a few criteria to rank the candidate designs. The trendiness score from the trained evaluator can be applied directly as one of the criteria. Meanwhile, two criteria are chosen to rank the generated upgrades: design aesthetics and image quality.

Although this study primarily focuses on promoting design trendiness, the general aesthetic level of the updates also needs to be guaranteed. Here we follow the approach used in Chapter 4, that estimates a car's aesthetic level based on car owners' reviews. Specifically, the sentiment method is employed to explore the positive and negative comments in reviews regarding cars' exterior styling, and the overall opinions are augmented as aesthetic scores from 0 to 1. Then a dual-task objective function, denoted \mathcal{L}_{aes} and \mathcal{L}_{rec} , is used for CNN training to enhance the learning of discriminative features for aesthetic inferring. The \mathcal{L}_{aes} for aesthetic regression uses the standard Mean Squared Error (MSE) to compute the error between truth and predicted aesthetic scores. \mathcal{L}_{rec} represents the angular loss for classification, by which the distances between categories are presented by cosine similarities $\cos \theta_j$ (here j indicates the class ID). Specifically, the variables in output neurons are all normalised, where input vectors' lengths are unified to α , weight vectors' lengths are unified to 1, and biases are set to 0. This loss function can be expressed as follows:

$$\mathcal{L}_{rec} = - \sum_{i=1}^N \log \left\{ \frac{\exp(\alpha \cos(\theta_{y_i} + \beta))}{\exp(\alpha \cos(\theta_{y_i} + \beta)) + \sum_{j \neq y_i} \exp(\alpha \cos \theta_j)} \right\}, \quad (5.12)$$

where N is the number of classes, y_i is the ground truth class for i th sample, and β is a constant margin penalty in the target class’s angle. Such a setting makes the distance only depends on the angle between the input image and neuron weight vectors. Existing studies (Liu et al., 2017; Wang et al., 2018; Deng et al., 2019) show the use of angular loss can minimise the intra-class distance and maximise the inter-class distances in angular space for classification.

The image quality, formally known as fidelity, indicates how realistic a produced car design is. Studies of adversarial perturbation show convolutional neural networks are not robust to noises. It is possible for a noisy-looking (i.e., low-quality) image to be labelled highly modern by the trendiness evaluator. Therefore, a measure that can assess the quality of an image is necessary to filter out poor generations. The widely used GAN generation criteria, such as Fréchet inception distance (FID) (Heusel et al., 2017) or Inception Score (IS) (Salimans et al., 2016), are not applicable in our scenario as they are designed for multiple images, while our study needs to assess individual design images. Therefore, a no-reference image quality assessment criterion is employed, which can make quality assessments without demanding the reference images. Specifically, the classical blind/referenceless image spatial quality evaluator (BRISQUE) (Sun et al., 2015) is used. It is a simple yet effective measure that infers image quality based on the distribution of locally normalised luminance coefficients.

5.4 Data and Experiment Settings

We use the DVM-CAR³ dataset (Huang et al., 2022) to validate the proposed approach. It contains 1,451,784 car images of 899 car models manufactured in the period from 2000 to 2019. Besides, the dataset includes all car models’ corresponding sale figures in the UK during the two decades. In experiments, we adopt the 61,827 car front images been 2000 to 2018 (i.e., 19 years).

For the trendiness prediction, four trendiness evaluators are trained accordingly for the four largest sub-markets: MPV, Saloon, Hatchback, and SUV. In total, 10,925 images from four discrete years (i.e., 2003, 2006, 2009, and 2012) are used, including 9,000 for training and 1,925 for testing. During ordinal model training, the images are labelled with four ordinal integers from 0 to 3 according to the cars’ production years from oldest to the most recent. The convolutional parts (no fully-connected layers) of VGG16 (Simonyan and Zisserman, 2015), which are pretrained by the ImageNet dataset (Deng et al., 2009), are adopted as the backbone for the evaluator, so the model can extract high-level visual features from input images. The

³<https://deepvisualmarketing.github.io>

ordinal regression layer takes these extracted visual features to make predictions of trendiness.

To obtain semantic views for the training of Mask R-CNN, 126 semantic view images from various brands are manually prepared. In each semantic image, the headlights, main and low grills are marked. The Mask R-CNN is fed with the original and semantic images to predict the positions of headlights, main and low grills. The trained Mask R-CNN processes all design images to obtain corresponding semantic views. Meanwhile, the gradient procedure is applied to car images for upgrading. For design scarping, masks larger than the semantic area are used to vacate the target parts. Thus, the whole antiquated parts are highlighted for upgrading. This allows the image completion algorithm to incorporate new design features with larger sizes.

To produce high trendiness designs in scraped areas, we used 26,899 car images from 2015 to 2017 with trendiness scores at the higher 50% rank compared to other designs in evaluator training. Meanwhile, 138 contemporary car models with the lowest 25% trendiness scores are chosen as the candidates for upgrading. We also include the setting of colour clustering to reduce the dimension size of inputs, which each pixel of an RGB image is converted from $256 \times 256 \times 3$ values to 1024 new values. During the training of the upgrader, the checkpoints' scoring on FID are tracked, which is widely used for generation quality indexing. The covered areas are chosen from the three semantic areas (i.e., the headlights, main and low grills). In our study, the rectangle masks are used as they only indicate the position of target semantic areas, while the similarly shaped masks could disclose the shape of the original missing parts. Furthermore, for data augmentation purposes, the training masks are randomly generated.

5.5 Results and Discussion

This section will present the designed experiments, results and corresponding discussions.

5.5.1 Validation of Trendiness Measure

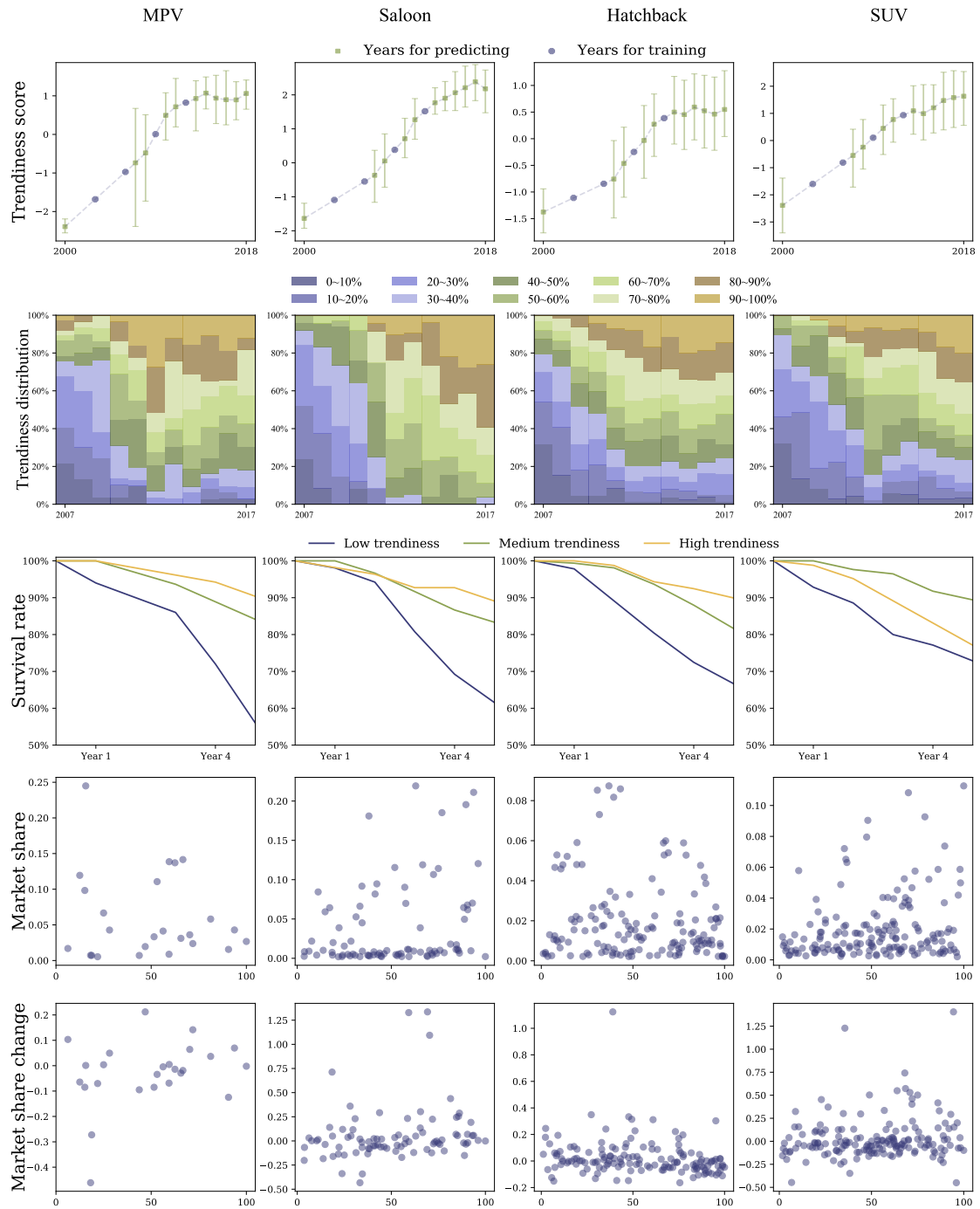
First, a series of data-based analyses were performed to validate if the obtained trendiness score is representative of the fashion trends. Analyses were based on four sub-markets, the MPV, Saloon, Hatchback and SUV markets, as they are the largest sub-markets for family cars. In some analyses, instead of the trendiness scores, we use the trendiness rank, the design's trendiness percentile among the same-year designs of the sub-market. E.g., if Audi A4 2016 has a trendiness score ranked as the top 10% of all 2016 Saloon cars, then its trendiness rank value is 10. Such rank can directly indicate the relative positioning of the designs on the "fashion distribution" of each period.

Trendiness score VS Time. As the trendiness score represents the design's projection position on the hypothetical trend of fashion, the average trendiness of the market is supposed to increase monotonically over the years. To verify this inference, as the first row of figures of Table 5.5.1 display, we compute the average trendiness score of all models over the observing periods, where error bars indicate the spans between the scores' first and third quartiles. Despite the evaluators being trained with the designs released before 2012, they can score the later designs (up to 2018) with higher trendiness ratings, which suggests that the trained evaluators can remain valid over six years. In general, the evaluator models give higher ratings to the actual designs that only appear after the training designs. The capability of predicting future fashion is a critical intention for the proposing of trendiness scoring. Due to the data limitation, the exact span of the prediction validation period has not been thoroughly examined in this study. Nevertheless, the implied predictive power of the evaluators is established. In rest analyses, the trendiness predictions obtained from the evaluators are used.

Trendiness rank VS Model length. As the developed trendiness measure indicates the evaluation of fashion, it is expected the designs labelled with relatively low trendiness scores of the year will disappear from the market over time, and the ones with high trendiness scores will occupy the market. The second row of Table 5.5.1 displays the distribution of car models according to their trendiness levels from 2007 to 2017. Here the year before 2007 or after 2017 is excluded as insufficient car models are included. The colour bar indicates the car models' trendiness levels among all designs. As the four subplots show, the cars with low trendiness scores only appear in the early years regardless of the sub-markets. It is noticed that the 2012 designs received unexpected high trendiness scores. We interpret this is a consequence of overfitting, where the ordinal regression model gives extremely high ratings to the 2012 designs, which is the last year in training data. Moreover, we conclude to see whether the trendiness is highly correlated to the car models' survival rate (i.e., the model is withdrawn or replaced) in the five coming years. In the third row of Table 5.5.1, the three trendiness levels are calculated based on the cars' trendiness ranking among the designs in the same year. The purple, green, and brown curves represent the survival rate of designs with the trendiness scores that are among the lowest 33.33%, medium 33.33% to 66.67%, and the highest 33.33% groups. All market drop-outs of the car models were marked with the lowest trendiness scores.

Trendiness rank VS Market performance. Besides, we also analyse whether the trendiness scores associate with the car models' market performance. Cars with higher scores shall perform better in the market sales. As it is also driven by other significant factors such as price, brand and functionalities, we expect moderate correlations between trendiness and mar-

Table 5.5.1: Trendiness VS market share changes.



	MPV	Saloon	Hatchback	SUV
MS line slope (β_{MS} in e-4)	-3.2	4.15	0.84	1.70
MS line intercept (C in e-2)	7.51	1.10	2.38	1.13
MS correlation (ρ) [†]	-0.15	0.26*	-0.13	0.22*
MS p-value [§]	0.48	0.02	0.13	0.01
MSC [†] line slope (β_{MSC} in e-3)	1.40	2.05	-0.83	1.21
MSC line intercept (C in e-2)	9.89	5.10	6.73	3.41
MSC correlation (ρ)	0.30	0.19	-0.18*	0.14
MSC p-value	0.15	0.08	0.03	0.08

[†] The market share change is computed as $\log \frac{\text{previous year market share}}{\text{this year market share}}$

[§] p-value from Pearson correlation



Figure 5.5.1: Illustration of identified antiquated regions. The green masks indicated the semantic areas that are highlighted by the saliency methods. Note that the demonstrated semantic areas are shaped by the Mask R-CNN.

ket shares. As the fourth row of Table 5.5.1 shows, a few significant results are found when comparing market shares with trendiness scores and ranks. As an alternative, it is rational to examine that cars with higher trendiness scores are more likely to gain higher market shares. We examine such a hypothesis by regressing the trendiness rank of designs with their corresponding changes in the market share. The four plots of the fifth row of Table 5.5.1 present the historical market share variances of models in the sub-markets according to their trendiness scores. The bottom parts of the table provide the supplementary statistics. As demonstrated, there exist only one negative significant correlation between trendiness and the cars' sales growth among the Hatchback market.

5.5.2 Interpretation of Upgrading Results

Antiquated parts recognition

To examine the importance of each semantic area in trendiness design, the recognised outmoded parts from the selected 138 car models are summarised. The three major design components, the headlights, main grill, and low grill, account for 52.89%, 32.56% and 14.56%, respectively. It suggests that the main grill determines the trendiness perception the most when component sizes are now considered (i.e., the main grills are much larger than other components in size). Fig. 5.5.1 demonstrates practical design samples with identified outdated parts. These highlighted areas are automatically segmented by the Mask R-CNN and identified as the most

Table 5.5.2: Comparison of upgrading on semantic/non-semantic areas.

	Non-semantic	Headlights	Main grill	Low grill
FID of upgradings	4.921	10.57	33.07	14.41
% of trendiness rise > 0.5	21.3%	44.8%	43.8%	53.1%
Avg. Trendiness rise	0.224	0.317	0.355	0.360

vulnerable part by Eq. 5.6 in gradient.

To further investigate the identified antiquated regions’ influence, trendiness changes caused by these semantic and non-semantic areas are compared. Specifically, for each semantic area, it is counted how likely an upgrading on the component can lead to a rise larger than 0.25 (defined as significant in this study) in trendiness. The upgrades on non-semantic areas, which have no overlap with semantic areas, are included for comparison. As the second and third rows of Table 5.5.2 show, compared with upgrades on non-semantic areas, the upgrades on headlights, main and low grills are more likely to bump up trendiness significantly. Moreover, as numbers of average trendiness rise show, the upgrades in semantic areas cause a considerably higher rise than in non-semantic areas, where the upgrades in low grills achieve the largest rise.

Upgrading Results

Fig. 5.5.2 gives multiple upgraded directions of the first three candidates in Fig. 5.5.1, where the antiquated parts are morphed. The general image quality deterioration caused by the design morphing can be indicated by the FID indices in Table 5.5.2, which are computed between upgrades on different semantic/non-semantic areas and the training images. The FID measures the average activation dissimilarity of CNN-extracted features between two image collections. Thus a high FID score shows the generation is remarkably different from the benchmark set, which indicates generations with poor quality. Apparently, the upgrades on the non-semantic areas have rarely affected image quality, which has the lowest FID score. However, the FID scores on semantic areas have large variances. The FID of headlights is 10.57, which is way lower than 33.07 of the main grills’ FID. The large FID of the main grill upgrades indicates the morphing on the main grills has quality issues – indicating the refilled content is unnatural.

5.5.3 Discussion

In general, our study indexes the aesthetic trendiness of cars in the automotive market, and upgrades the detected low trendiness designs using the machine learning approach with large-scale data. On the one hand, this study demonstrates that automotive trendiness could be characterised as a continuous scale that discriminates designs in different years. The simulated scale

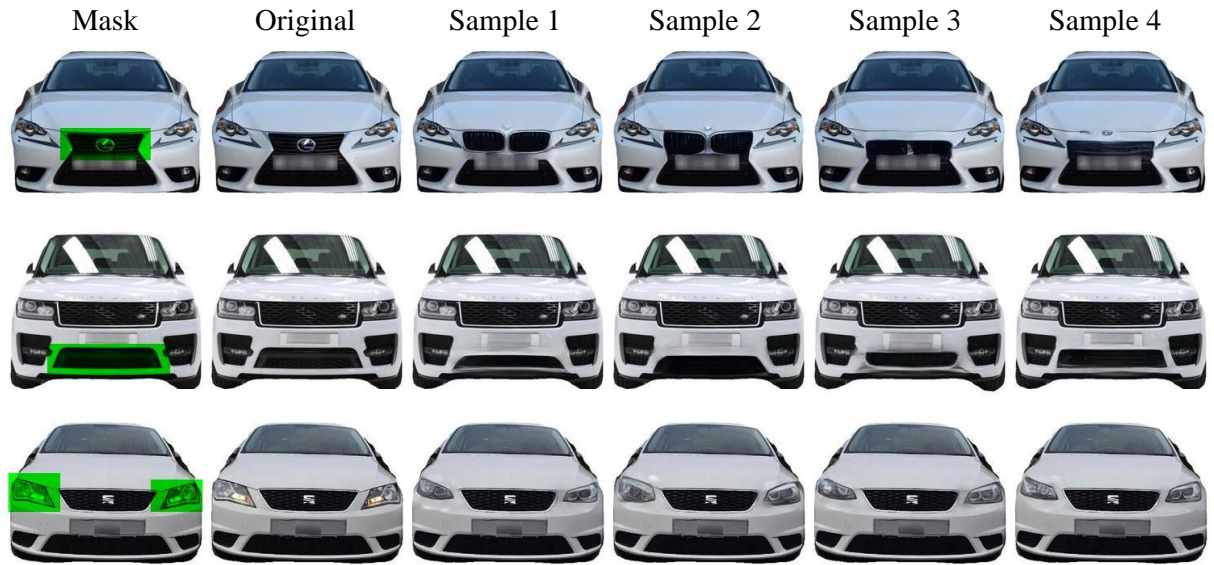


Figure 5.5.2: Empirical examples of three updated car models: Lexus IS 2016, Land Rover Range Rover 2017, SEAT Toledo 2017. As the changes on Lexus IS demonstrate, the changes on the main grill would alter the brand’s family faces, which is problematic for real application. Nevertheless, these samples demonstrate the modifications achieved via the image completion algorithm.

is applicable for a considerable length of duration and has noticeable associations with critical market performances, such as changes in its future market share and survival rates. On the other hand, we show that advanced image completion algorithms can be deployed to produce design solutions which upgrade antiquated designs regionally. This approach can adjust the perceived trendiness while maintaining the most current designs without being bothered by feature entanglement issues.

Regarding the analyses on the trendiness evaluator, first, in line with our expectations, the more recently launched models receive higher trendiness ratings from the trained evaluator. This indicates the existence of contemporaneity-related appearance features of cars, which allow the CNN models to infer their years of design. Second, the experimental results suggest the effectiveness of the proposed trendiness measure is not a side effect of overfitting, as later designs receive higher scores even though they did not appear in training samples. Considering the expected lifespan of car models is less than eight years, the mainstream designs in 2018 are very different from training samples selected from 2012. Such findings support the self-perpetuating characteristic of the automotive fashion evolution, concluded by (Reynolds, 1968). That is, despite the frequent turning over of car models, the direction of fashion evolution in the automobile

industry remains relatively stable, making it possible to predict future fashion trends.

Further analyses showed that the trendiness rise over the years due to the model replacement, and cars with low trendiness scores suffer from lower survival rates over time. In all sub-markets, designs with low trendiness scores would be gradually withdrawn and replaced by new models. As no statistical correlations are found between market sales and trendiness, the low survival rates of lowly scored cars are unlikely caused by the low market sales. Instead, the low survival rates attract the attention of the automakers who are keen to redesign or replace an existing design if the models look outdated and start to lose market shares, regardless of their temporary market shares. Noteworthy, no significant correlation is detected for the Hatchback market. A possible explanation for this finding would be due to the price differences of four main types of vehicles. Compared with others, hatchbacks are mostly cheaper so that the consumers focus more on its performance price ratio instead of appearance.

From the design upgrader perspective, the proposed semantic and gradient method can mark out front design parts that semantically associate with facial features and then determine the areas with the largest gradients. As predicted, morphing these semantic areas can result in larger improvements in trendiness scores rather than other areas, implying that the highlighted areas have more influence on trendiness perception. We interpret this as the result of anthropomorphism ([Aggarwal and McGill, 2007](#); [Landwehr et al., 2011b](#); [Welsh, 2006](#)) – as consumers tend to perceive car fronts as “car faces” and are sensitive to these semantic areas, automakers learned to signal the morphing of these parts when bringing new designs to the market. Future investigations can be done to justify this suppose.

Another key contribution of the design upgrader is to present the practicability of applying image completion algorithms to design upgrading. By removing the intended areas for upgrading, the state-of-the-art image completion algorithms can refill these scraped areas with natural-looking content while leaving the unintended areas untouched, thus achieving regional upgrading purposes. Results also displayed that via regional modifications, we can obtain morphed designs with higher trendiness, from which automakers can select candidate upgrades for facelift needs.

Admittedly, the study has its limitations. First of all, as a prior data-driven investigation for trendiness, no theoretical discussions regarding the underlying mechanism are provided in this study, which leaves many questions unclear. For example, does the observed trendiness result from the automakers’ collective market decisions? Does the evolution of trendiness source from the typicality, mere exposure effects or joint effects of both? Future studies can focus on

the theoretical parts of trendiness findings while being supported by results from practical data analysis. For antiquated design recognition, the branding designs (e.g., brand logo) are often included in the main grills for upgrading. This is problematic as such designs are cues for branding, which automakers avoid changing. Although the mistaken recognition can easily be addressed by manual effort (i.e., remove these family face parts manually), a better solution is needed to identify these family face cues automatically. In addition, the FID measure on different semantic areas shows the applied generative model performs poorly on specific parts (i.e., the main grills) – this also occurs in other extant studies. Similar results are obtained in other studies that use different deep generative models. This seems a general challenge for modern deep generative models to generate content with rigid structures or shapes such as a long straight line. This is interpreted as a limitation inherent in the mechanism of transposed convolutional layers, which can hardly arrange content in rigid forms globally.

5.6 Summary

This chapter proposed a data-driven approach that can learn, define and score the trendiness of car designs, meanwhile, identify the most antiquated design regions for upgrading. Besides, the generative modules offer numerous upgraded designs without changing the unintended design parts and choose the upgrades with more modern looks. The 2000 to 2018 UK automotive data were employed to develop and verify the developed measures and upgrades. Experimental results showed that the proposed trendiness ratings signify the model’s future chances of being withdrawn. Also, image completion algorithms, which have no feature entanglement issues, can be applied as the regional design upgrader.

This study has addressed a number of questions that remained from Chapter 4. These two chapters jointly demonstrate the effectiveness of applying deep generation for automotive front design upgrading. However, the question whether deep generative methods can effectively adjust car side-view designs remains unconvinced. Moreover, marketers might challenge whether original creativity exists in these upgrading methods, as data-driven models can only have inspiration from other existing car designs. To this end, a side-view focus study is proposed in the next chapter with cheetah-inspired design features.

Chapter 6

Animism: Design with Bio-Inspired Novelty

In this chapter, we propose a CycleGAN-based framework of bio-inspired automotive design for sports saloons. We tailor the deep learning methods such as HED (Xie and Tu, 2017) model to detect representative lines and curves of side-view images from the targeted saloons and the running cheetah, and integrate them to obtain upgraded novel designs. Specifically, Section 6.1 refers to the motivation for this bio-inspired study. Section 6.2 introduces the three-staged design process incorporating the cheetah feature into car design. Section 6.3 list the technical details for experiments. Section 6.4 presents the results we find through two user studies and gives discussions accordingly, and Section 6.5 summarises this chapter.

6.1 Introduction

Automotive anthropomorphism, in particular, the human tendency to see faces in car fronts, has been discussed considerably in prior research. Landwehr et al. (2011b) revealed that emotional expressions can be conveyed by a cars grill and headlights because they look like mouth and eyes. Purucker et al. (2014) evidenced that the way human perceive car fronts is similar to the way human perceive facial expression. Ku (2014) discovered that the anthropomorphism of a car's face is a bottom-up procedure as the fusiform face area in human brain is associated with seeing human features in car fronts. Based on the premise that car fronts are perceived in much the same way as human faces, Maeng and Aggarwal (2017) showed that with high face width-to-height ratio are perceived as more dominant, which in turn receive more positive



Figure 6.1.1: Side-view design comparison for Audi A4 2018 (left) vs 2019 (right) facelift.

evaluations.

Built upon this stream of research, we hypothesise that animals such as cheetah might activate the speedy schema if we were to draw design inspirations from it. As such, we adopt schema congruity (Aggarwal and McGill, 2007) as the theoretical foundation for examining the effectiveness and consequences of bio-inspired design. Our pretest in Section 6.4.1 shows that displaying a fast animal image alongside car side-view images without any alterations on the actual design indeed increase consumer’s evaluations on how sporty the automobile looks. The results are consistent with research on schema congruity for product evaluation (Fiske, 1982; Meyers-levy and Tybout, 1989; Aggarwal and McGill, 2007).

This chapter carries out the first empirical attempt to improve the mid-generational refresh on the side-view facelift design for the following two reasons. First, despite its importance in practice, enhancing the side-view of an automobile has rarely been discussed in the existing literature. To date, researchers have emphasised either the car front design (Aggarwal and McGill, 2007; Landwehr et al., 2011a,b) or redesigned the car holistically (Pan et al., 2016; Burnap et al., 2021). Second, if we were to capture the running cheetah’s swift and consecutive movements and integrate these formed inspirations into the aesthetic design, a side-view design would be the most relevant initial attempt.

When it comes to side-view facelift design, one particularly important feature is the exterior styling of shoulder and waist curves. Figure 6.1.1 provides an example of comparing the side-view designs of Audi A4 2019 facelift vs 2018.¹ We observe the following line changes in the 2019 facelift. First, the shoulder curve (or *bone line*²) – the line that starts from low on the front bumper and continues along the side through the fuel filler cap door – becomes shorter, which makes the facelift look more sporty and dynamic. Second, the waist curve – the character line on the door panel across the centre of the car – becomes curvier and with a larger slope, which enhances the sporty style. Our study aims to use deep generative models to modify shoulder

¹URL: https://www.youtube.com/watch?v=_uR11dkPkC0.

²URL: <https://www.youtube.com/watch?v=haiMna8NBQs>.

and waist curves of existing sports saloons from the inspirations learned from running cheetah. Both curves are perfectly suitable for our goal of facelift design because they do not change the overall shape and would allow our design framework retain the basic exterior styling and platform chassis of the car with aesthetic alterations.

The proposed design framework consists of three modules. The first module analyses a cheetah's movement and identifies the curves which can represent its body dynamics. The second module processes saloon side-view images, detects and extracts shoulder and waist curves for each car model in our study. The third module morphs the pre-processed lines from saloon and cheetah, and then generates the new exterior designs. Three deep learning algorithms – the mask region-based convolutional neural network (Mask R-CNN) (He et al., 2017), the holistically-nested edge detection (HED) algorithm (Xie and Tu, 2017), and the cycle consistency generative adversarial network (CycleGAN) (Zhu et al., 2017) – are used, tailored, and integrated into these three modules.

Specifically, the Mask R-CNN is used to locate the running cheetah in the video frames and to remove the background of sports saloons in their original side-view images. It is a state-of-the-art CNN proposed by He et al. (2017), which aims to detect and segment object instances in digital images. It has been widely used in autonomous driving (Huang et al., 2020; Xia and Sattar, 2019), multi-person pose estimation (Dong et al., 2019), and neural stem cell differentiation (Zhu et al., 2021). The HED is used to extract body curves from images of the running cheetah. It is a CNN-based algorithm that has been used in applications like semantic image segmentation (Chen et al., 2016; Yu et al., 2017) and salient object detection (Li and Yu, 2016; Li et al., 2017). Compared to traditional edge detection algorithms, the HED is more robust in edge detection that can effectively eliminate cheetah print. As the CycleGAN excels in image-to-image translation tasks like image style transfer and photo enhancement, its application in our design framework is two-fold: (i) generating sketches from real side-view images of saloon models; and (ii) generating new car designs in terms of near-realistic look images from sketches.

This research makes the following contributions to the extant literature. First, To the best of our knowledge, this is the very first study which develops a data-driven framework for bio-inspired automotive aesthetic design. As the proposed framework can also be extended to other design applications, the study contributes methodologically to developing a general bio-design framework of integrating various data-driven algorithms and frontier technologies. Second, unlike Pan et al. (2017) and Burnap et al. (2021) which also use deep learning for automotive aesthetics design, our design application have three specific focuses: bio-inspired, side-view,

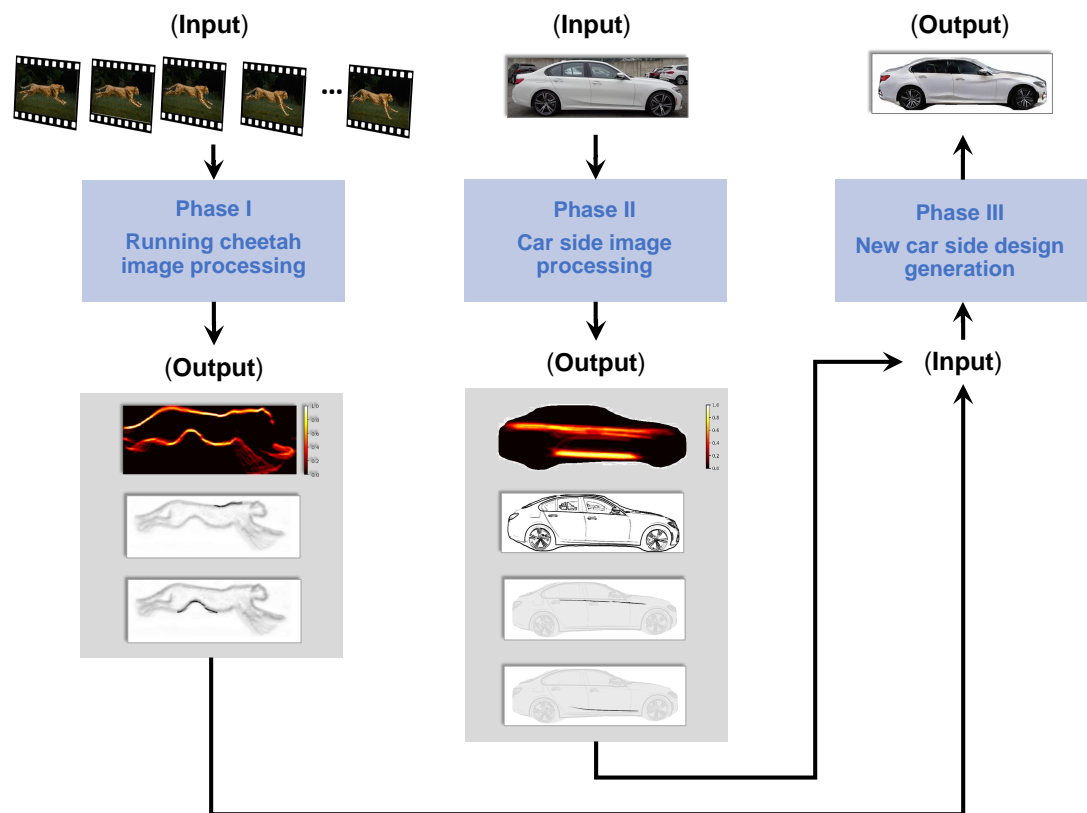


Figure 6.2.1: Schematic view of the proposed bio-inspired design framework.

and facelift. Last, compared with marketing studies in automotive aesthetics, our study employs the frontier deep learning techniques, and also fills the gap and contributes to the developing literature on automotive anthropomorphism and bio-inspired design.

6.2 Methodology

As presented in Figure 6.2.1, the proposed bio-inspired design framework contains three modules. In the first two modules, real-world side-view images of the running cheetah and car models are converted into sketches, respectively. The representative body curves of the cheetah and car models are extracted in the corresponding modules separately and in parallel. Sketches are widely used in the automotive design process to aid in the progression of a design solution and play an essential part in knowledge acquisition and representation (Bar-Eli, 2013). They can provide near-realistic look of car exterior with details, highlight the salient curves or lines, and avoid the influence of colour themes and light reflection. Then, in the third module, the bio-inspired new design sketches are generated and are then converted into the real-world

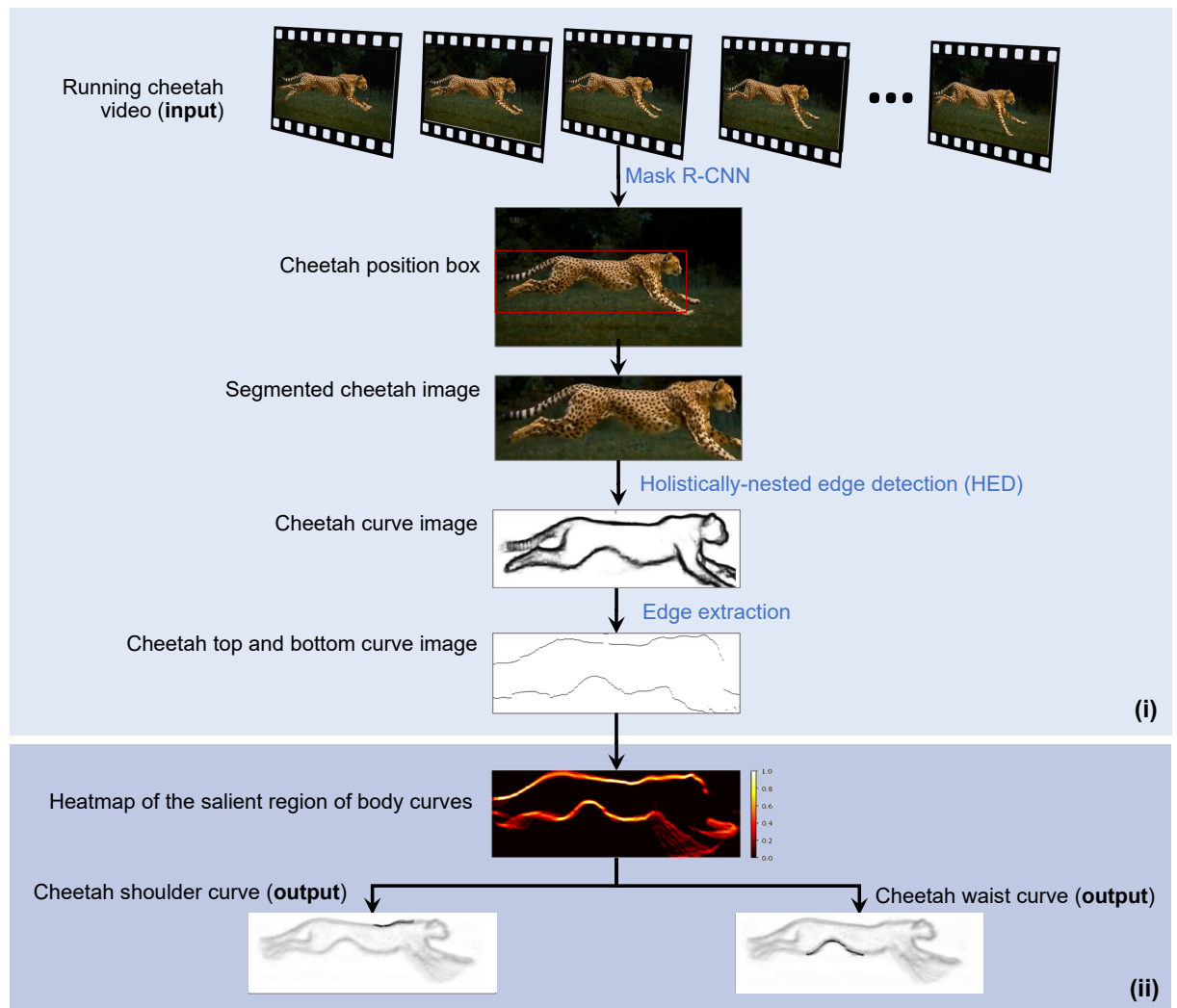


Figure 6.2.2: Schematic view of the running cheetah image processing module.

images which can accurately reflect reality and can help persuade consumers on production-ready viewing.³

6.2.1 Running Cheetah Image Processing

The first module aims to use deep learning to automatically learn the representative body curves from the world's fastest animal cheetah, and the key steps in the module are presented in Figure 6.2.2.

To form inspirations from the running cheetah, we need to capture its swift and consecutive

³URL: <https://www.creativebloq.com/inspiration/illustration-vs-photography-how-do-you-decide>

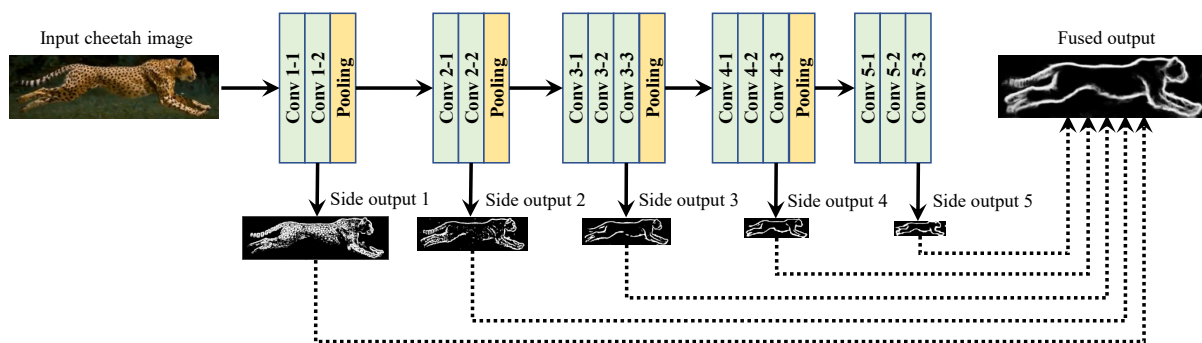


Figure 6.2.3: Illustration of HED architecture for the running cheetah edge detection.

movements. National Geographic magazine filmed the running cheetah with Phantom cameras in June 2012 (Smith, 2012). The mission of the project was to track alongside a cheetah run at full speed. As cheetah runs too fast for the eye to follow, they built a 410-foot dolly track for high-speed cameras so that there would be no wobbles. The process was to line everything up so that they would ensure that when the cheetah came out from the gate, everything was in place and they can capitalise on their opportunities. The video “Cheetahs on the Edge - Director’s Cut” was released by National Geographic magazine on YouTube to show cheetah’s movement. We download this video from YouTube and extract 8,222 frames containing the running cheetah using a moving window over a period of 4 minutes and 28 seconds.⁴ We then select the first frame in five consecutive frames so 1,645 images are sampled.

For a given image frame, the Mask-RCNN (He et al., 2017) is applied to detect the running cheetah. The algorithm takes the raw image as the input and its output includes the detected objects with bounding boxes, the classification results for the detected objects and the corresponding masks that cover the shape of the objects. As highlighted in Figure 6.2.2, the red rectangle is the bounding-box (in red colour) that shows the position of the detected running cheetah, then a segmented cheetah image is obtained.

Mathematically, the Mask R-CNN has the following joint loss function

$$\mathcal{L}_{\text{MASK R-CNN}} = \lambda_{\text{CLS}} \mathcal{L}_{\text{CLS}} + \lambda_{\text{BOX}} \mathcal{L}_{\text{BOX}} + \lambda_{\text{MASK}} \mathcal{L}_{\text{MASK}}, \quad (6.1)$$

where \mathcal{L}_{CLS} is the classification loss that predicts the classes of the detected objects, \mathcal{L}_{BOX} is the bounding-box loss that produces the bounding boxes for the objects, $\mathcal{L}_{\text{MASK}}$ is the mask loss that gives the mask labelling for the objects, and λ_{CLS} , λ_{BOX} , λ_{MASK} are the corresponding loss

⁴URL: https://youtu.be/THA_5cqAfcQ

coefficients, respectively. Technical details of these loss functions is introduced in Section 2.2.

Next, we extract the holistic body curves of the running cheetah into a sketch using the HED (Xie and Tu, 2017). This algorithm can effectively capture the general shape of the running cheetah while in the meantime can remove cheetah print in sketch generation. As shown in Figure 6.2.3, the HED is essentially a CNN but it unitises the features learned in different network layers by creating segmentation maps. These maps are called the *side outputs* (or *side-output layers*). They are network layer samples and do not change the original network architecture. Using side outputs in neural networks has proven to be useful in achieving desirable results in image processing (Farabet et al., 2013; Simonyan and Zisserman, 2015). The HED has a joint loss function which is a combination of side-outputs from distinct layers, which can be expressed as follows

$$\mathcal{L}_{\text{HED}} = \mathcal{L}_{\text{FUSE}} + \sum_m \mathcal{L}_{\text{SIDE}}^{(m)}, \quad (6.2)$$

where $m = 1, \dots, M$, is the m th side output, $\mathcal{L}_{\text{SIDE}}^{(m)}$ is the image-level loss function for side output m , and $\mathcal{L}_{\text{FUSE}}$ is the loss function at the fusion layer that learns the fusion weights of side outputs. The side-output loss $\mathcal{L}_{\text{SIDE}}^{(m)}$ is defined as the class-balanced cross-entropy

$$\mathcal{L}_{\text{SIDE}}^{(m)} = - \left[\left[1 - \frac{1}{N} \sum_n z_n \right] \left[\sum_n z_n \cdot \log \{ \hat{z}_n \} \right] + \frac{1}{N} \sum_n z_n \sum_n (1 - z_n) \log \{ 1 - \hat{z}_n \} \right], \quad (6.3)$$

where $n = 1, \dots, N$, is the n th pixel of the given image, $z_n \in \{0, 1\}$ tells whether the n th pixel is an edge pixel, \hat{z}_n is the model's prediction probability. Therefore, $\sum_n z_n$ is the number of edge pixels in the image, and the fractions $\frac{1}{N} \sum_n z_n$ and $(1 - \frac{1}{N} \sum_n z_n)$ are used to deal with the imbalance between edge and non-edge pixel numbers. The fusion loss $\mathcal{L}_{\text{FUSE}}$ is specified by the cross-entropy

$$\mathcal{L}_{\text{FUSE}} = - \left[\zeta \log \left\{ \alpha \left(\sum_m \delta_m \gamma_{\text{SIDE}}^{(m)} \right) \right\} + (1 - \zeta) \log \left\{ 1 - \alpha \left(\sum_m \delta_m \gamma_{\text{SIDE}}^{(m)} \right) \right\} \right], \quad (6.4)$$

where ζ is the corresponding ground truth binary edge map for the given image, $\alpha(\cdot)$ is the sigmoid function, $\gamma_{\text{SIDE}}^{(m)}$ is the activation out of side output m and δ_m is the corresponding fusion weight.

The top and bottom body curves are the most distinctive curves of the running cheetah, which also match with the sports saloon's bone line and character line from the side view. Different

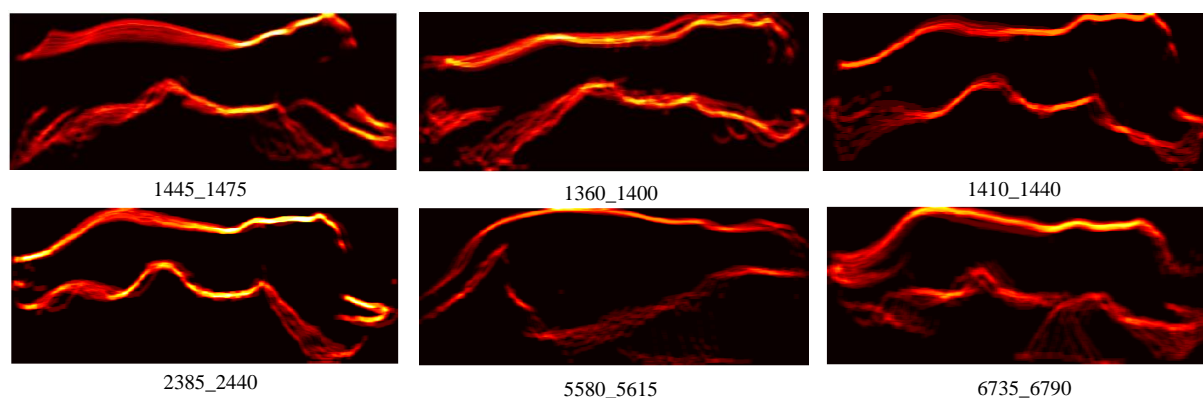


Figure 6.2.4: The most representative motions of the running cheetah.

cheetah curves of consecutive movements can be aggregated by a set of frame images, generating a heatmap of salient region, in which more common values are shown in brighter colours. We detect the most remarkable curves from all image frames to represent the cheetah’s sporty motions. The minimum Euclidean distance is used as the true distance between two frames, and the large frame distances (i.e., the highest 20%) are used as the split threshold to cluster frames. We call each frame cluster the *stable motion moment* (SMM), which contains roughly 10-20 frame images. Frames of the running cheetah in our collected data can be clustered into 14 SMMs. For the illustration purpose, 6 SMM examples are selected to present in Figure 6.2.4 and the first plot of each subfigure is the SMM.

The selected SMM undergoes further processing to extract the smooth body curves of the running cheetah. More specifically, it is divided into 32 columns based on its coordinates. The centres at the top and bottom are calculated for each column’s upper and lower curves. As depicted in the second plot of each subfigure in Figure 6.2.4, mass points for both the upper and lower curves can be obtained from each column. These points can then be used to fit the smooth body curves of the cheetah.

6.2.2 Car Side-View Image Processing

As shown in Figure 6.2.5, the second module processes the side-view images of different car models to identify the representative body lines. We deliberately use online publicly available car images rather than “cleaned” blueprints or advertising images from automakers so more variations can be learnt by the CycleGAN. As the Mercedes C Class shown in Figure 6.2.5, the original image contains the background that is irrelevant to aesthetics design. Therefore, in the first step, we segment the saloon from its background by using the *Mask R-CNN* (He

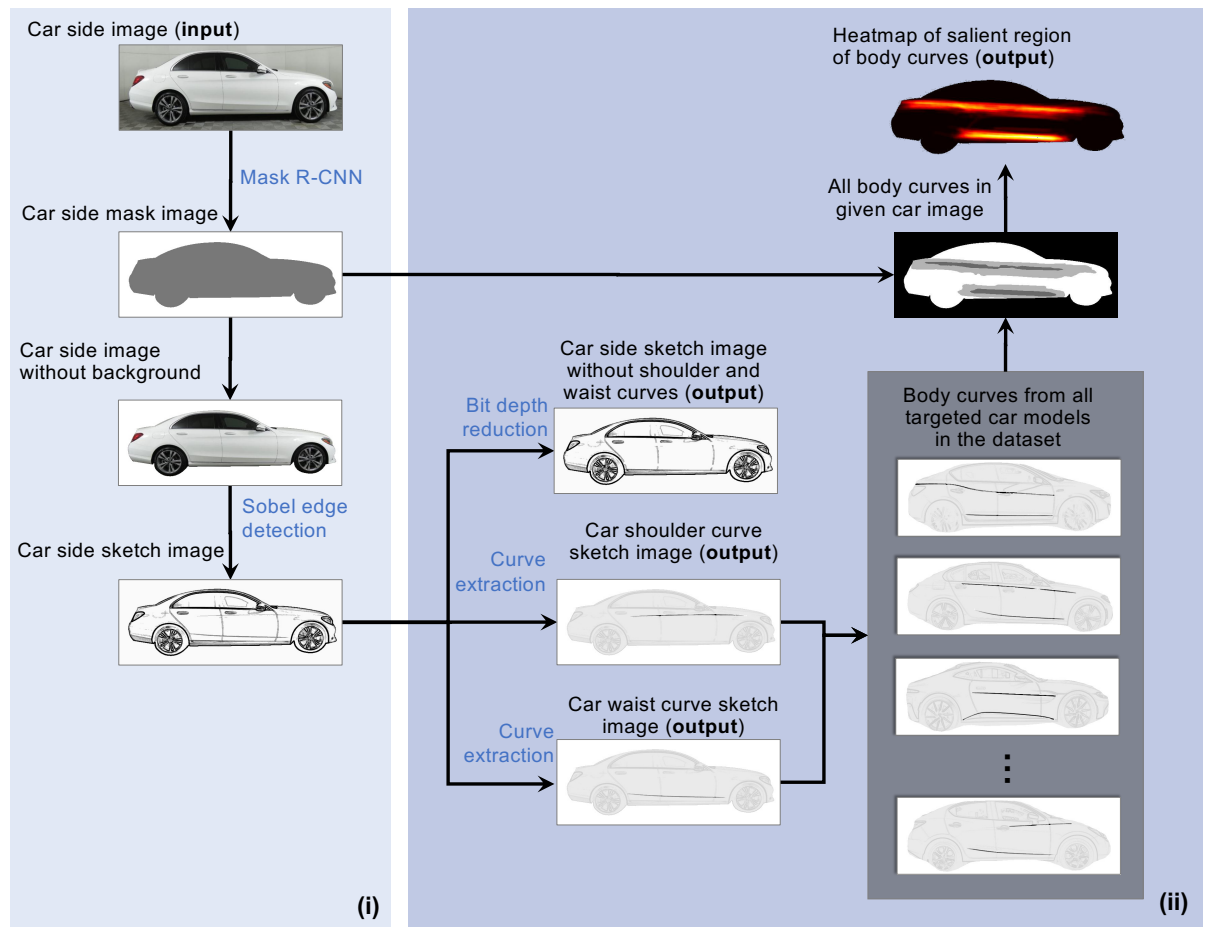


Figure 6.2.5: Schematic view of the car image processing module.

et al., 2017). Its mask output is a binary image (i.e., matrix) that indicates whether the car body presents in each pixel so that the position of the car can be identified. The element-wise product (or the Hadamard product) between the original image and the mask binary image can be computed, and then the car side-view image without background can be obtained. Intuitively, it assigns the pixels that lie outside the car to the white colour.

After removing the background, images from different car models are normalised into the same size. saloon's shape and colour texture are much simpler than cheetah so we can use the SED algorithm (Pratt, 1978) to generate the corresponding sketch image. Unlike the HED which has a lengthy training process, the SED can be trained quickly and can achieve a comparable excellent performance for saloon sketches generation. The input RGB image will be firstly

converted into grayscale by using the *ITU-R 601-2 luma transform*.⁵ Let R , G , and B denote the red, green and blue channel matrices of the input car side-view image without background, then the output grayscale image Θ can be obtained,

$$\Theta = \frac{1}{\lambda_R + \lambda_G + \lambda_B} \left[\lambda_R R + \lambda_G G + \lambda_B B \right], \quad (6.5)$$

where λ_R , λ_G , λ_B are constant coefficients. The Sobel operators are the kernels that are convolved with the grayscale image. Let Φ_1 denote the Sobel operator that responsible for horizontal edge detection and let Φ_2 denote Sobel operator that responsible for vertical edge detection. Then pixel (i, j) of the sketch image $\tilde{\Theta}$ can be obtained as below

$$\tilde{\Theta}(i, j) = \sqrt{\left[\sum_{i'=-\chi}^{\chi} \sum_{j'=-\chi}^{\chi} \Phi_1(i, j) \Theta(i-i', j-j') \right]^2 + \left[\sum_{i'=-\chi}^{\chi} \sum_{j'=-\chi}^{\chi} \Phi_2(i, j) \Theta(i-i', j-j') \right]^2}, \quad (6.6)$$

where $\chi = \lfloor \frac{v-1}{2} \rfloor$, $\lfloor \cdot \rfloor$ is the floor function, and v is the size of Sobel operator.

As we focus on morphing the shoulder and waist curves, each sketch can be decomposed into three images (or layers): a car side sketch image without body curves, a shoulder curve sketch image, and a waist curve sketch image. To reduce the image data “noise” and obtain salient body curves with less shades, the bit depth of sketch images is then reduced from 8 bits to 3 bits in the decomposition.⁶ To illustrate the possible places for curve incorporation, for each car model, the candidate areas are highlighted in colour, which results in the heatmap looking on the sketch image. The probability is computed according to the actual appearance frequencies of body curves in that area. The body curve sketches of different car models are combined and incorporated into the side-view mask image of the selected car model. An illustration example of Mercedes C Class (2019 reg) is presented in Figure 6.2.5. To obtain a more averaged distribution, the Gaussian smoothing process is applied to smooth the resulting image in order to obtain a heatmap.

⁵URL: <https://pillow.readthedocs.io/en/stable/reference/Image.html>

⁶The *bit depth* is an image processing measure for the level of precision, which quantifies how many unique colours are available in an image’s colour palette in terms of the number of bits. Images with a higher bit depth can encode more colours or shades – the 8-bit sketch image contains $2^8 = 256$ shades while the 3-bit sketch image only contains $2^3 = 8$ shades.

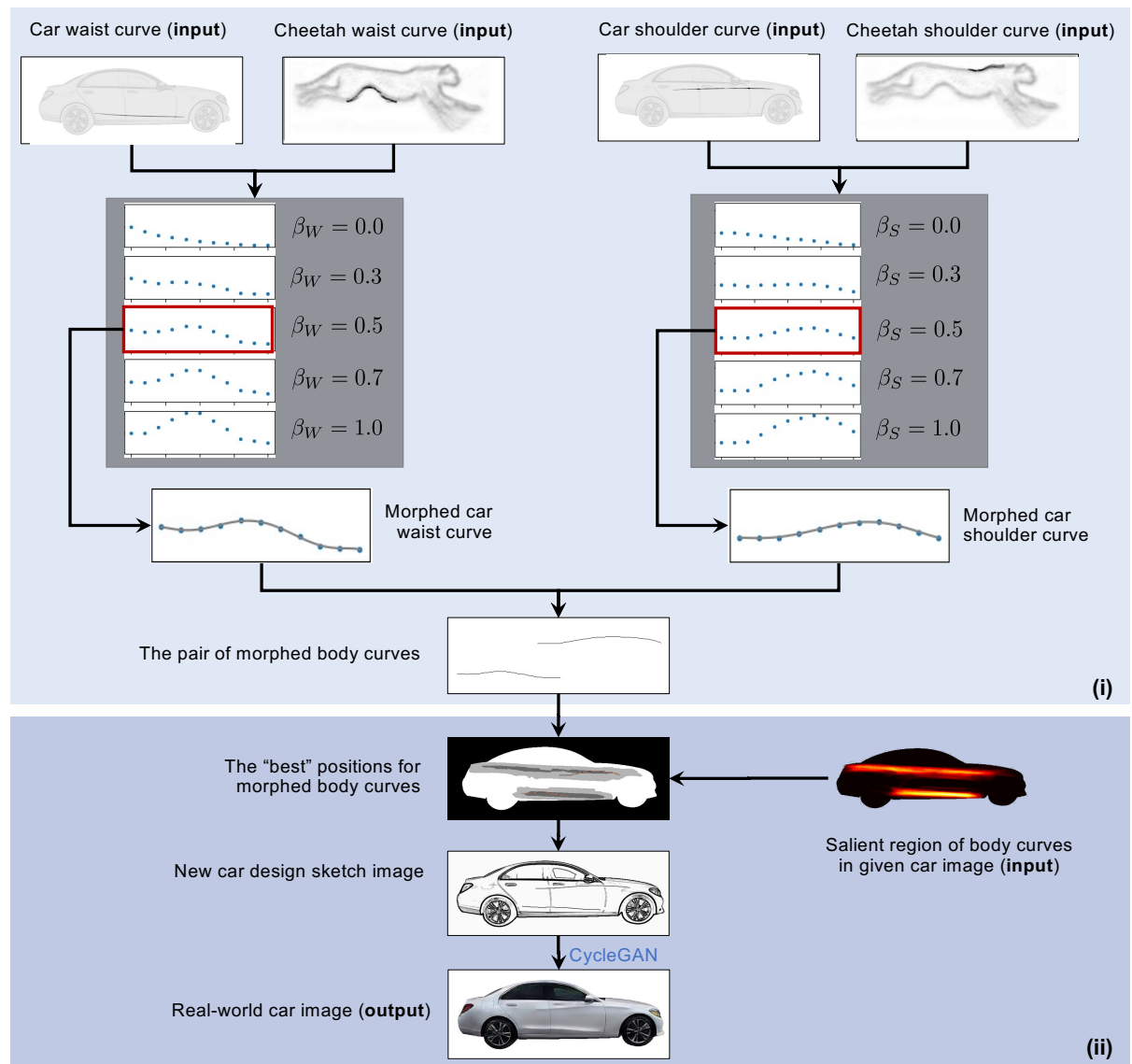


Figure 6.2.6: Schematic view of the new car side design generation module.

6.2.3 Bio-Inspired Car Side-View Design Generation

Figure 6.2.6 presents a schematic view of the final module. It generates a new car side-view aesthetics design by incorporating the body curves of running cheetah.

To morph selected saloon body curves with cheetah curves, they are firstly converted into vectors with the same length. Specifically, U points are sampled with a fixed step from the corresponding sketch images of body curves.⁷ For each vector, a point's value is its row index

⁷In experiments, we set $U = 10$.

of the sampled body curve in the sketch image. The converted body curve points are scaled using the z-score standardisation (Bishop, 2006) and are then morphed using linear interpolation. Let c_S and c_W denote the vectors of the standardised shoulder and waist curve points for the car, and let \tilde{c}_S and \tilde{c}_W denote the vectors of the standardised shoulder and waist curve points for the running cheetah. The vectors of the morphed body points for the shoulder and waist curves, denoted by r_S and r_W , respectively, can be obtained as follows:

$$r_S = (1 - \beta_S)c_S + \beta_S\tilde{c}_S, \quad (6.7)$$

$$r_W = (1 - \beta_W)c_W + \beta_W\tilde{c}_W, \quad (6.8)$$

where β_S and β_W are the coefficients that control the morphing degrees of the shoulder and waist curves, respectively. They subject to the constraints: $0 \leq \beta_S \leq 1$, $0 \leq \beta_W \leq 1$.

Smoothing body curves can be obtained by applying polynomial fitting (Bishop, 2006) to the morphed body curve points. A polynomial function can be expressed as $\sum_h \delta_h u^h$, where $u = 1, \dots, U$, is the index of input data points, $h = 1, \dots, H$, is the h th degree of polynomial fitting, and δ_h is the coefficient of the h th degree.⁸ Minimising the following regression loss can obtain the coefficient estimates of the polynomial function that approximates the targeted body curve:

$$\mathcal{L}_{\text{POLYFIT}} = \sum_u \left[\sum_h \delta_h u^h - r(u) \right]^2, \quad (6.9)$$

where $r(u)$ can be $r_S(u)$ or $r_W(u)$.

Then morphed smoothing shoulder and waist curves can be incorporated into the car sketch image based on the heatmap of the car model's salient region. In order to find the best positions for the new curve, we adopt a grid search method. Each pair of morphed curves are re-scaled into ten different sizes and then rotated further to have ten different degrees. Therefore, we obtain 100 different re-scaled and rotated pairs of body curves. The overlaps between these body curves and the heatmap are computed, which conceptually represent the total joint probability of curve pixel allocation. The "best" positions of the body curves in the given car sketch image is the pairs with the maximum total sum of the probabilities from the overlap on the heatmap. As illustrated in Figure 6.2.6, the pairs are highlighted in the red colour. Once the "best" positions are determined, the car side sketch image of bio-inspired design can be generated.

⁸In experiments, we set $H = 6$.

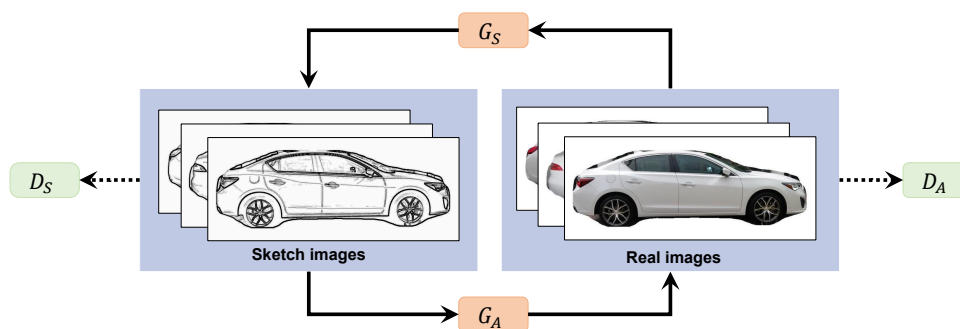


Figure 6.2.7: Illustration of CycleGAN structure.

Based on the output sketch image, the real-world car image can be generated using the CycleGAN (Zhu et al., 2017), which is specifically good for image-to-image translation tasks. In our framework, we aim to translate new car design from the domain of sketch images into the domain of real-world images. We use \mathcal{A} to denote the domain of the real-world images of car models and \mathcal{B} to denote the domain of corresponding sketches. Simply, CycleGAN contains two GANs (Goodfellow et al., 2014): one for the real-world images and one for the sketches. Each GAN has a generator and a discriminator. The former generates fake images while the latter detect if the generated images are fake or not. The two generators ensure the translation cycle between the real-world images and the sketches. The structure of CycleGAN is illustrated in Figure 6.2.7.

Specifically, we denote G_A for the generator that maps from \mathcal{B} to \mathcal{A} , G_B for the generator that maps from \mathcal{A} to \mathcal{B} , D_A for the discriminator that detects if a generated image from G_A is faked or not, and D_B for the discriminator that detects if a generated image from G_B . CycleGAN adopts the basic idea of GANs by constructing a minmax game between discriminator(s) and generator(s). Therefore, the training of CycleGAN is equivalent to solving the following optimisation problem

$$\min_{G_A, G_B, D_A, D_B} \max \mathcal{L}_{\text{GAN-A}} + \mathcal{L}_{\text{GAN-B}} + \mathcal{L}_{\text{CYCLE}}, \quad (6.10)$$

where $\mathcal{L}_{\text{GAN-A}}$ is the loss function for GAN that generates and detects real-world images, $\mathcal{L}_{\text{GAN-B}}$ is the loss function for GAN that generates and detects sketches, and $\mathcal{L}_{\text{CYCLE}}$ is the loss function that ensures the cycle consistency of image translation between two domains. We

adopt the specifications used by [Zhu et al. \(2017\)](#) for these loss functions as follows:

$$\mathcal{L}_{\text{GAN-A}} = \mathbb{E}_{A \in \mathcal{A}} \left[\log\{D_A(A)\} \right] + \mathbb{E}_{B \in \mathcal{B}} \left[\log\{1 - D_A(G_A(B))\} \right], \quad (6.11)$$

$$\mathcal{L}_{\text{GAN-B}} = \mathbb{E}_{B \in \mathcal{B}} \left[\log\{D_B(B)\} \right] + \mathbb{E}_{A \in \mathcal{A}} \left[\log\{1 - D_B(G_S(A))\} \right], \quad (6.12)$$

$$\mathcal{L}_{\text{CYCLE}} = \mathbb{E}_{A \in \mathcal{A}} \left[\|G_A(G_B(A)) - A\|_1 \right] + \mathbb{E}_{B \in \mathcal{B}} \left[\|G_B(G_A(B)) - B\|_1 \right], \quad (6.13)$$

where $\|\cdot\|_1$ is the L^1 norm.

6.3 Experimental Settings

6.3.1 Key Steps of the Targeted Saloon Models

As discussed in Section 6.2.2, the car image processing module identifies the representative body lines from side-view images of various car models. Fig. 6.3.1 displays the original side-view images of the ten targeted saloon models and their intermediate processing results from each step.

6.3.2 Using of Sobel Edge Detection

As described in Section 6.2, the SED takes raw car images as the input and generates sketch images. In our experiments, we follow [Pratt \(1978\)](#) and [Gonzalez et al. \(2004\)](#), and adopt the following settings:

$$\lambda_R = 299, \quad \lambda_G = 587, \quad \lambda_B = 114,$$

$$\Phi_1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

The SED with the above settings works very well in our context. It captures lines and shapes of the car in the given raw image, and converts them into sketches accordingly. Other more advanced edge detection algorithms (e.g., the HED) can also be used in our framework to replace the SED. However, according to Occam’s razor ([Schaffer, 2015](#)), the simpler algorithm or method (i.e., the one with fewer parameters) is to be preferred. In our study, compared to the HED, the SED can achieve comparable performance for car sketch generation while it is simpler and computationally faster.

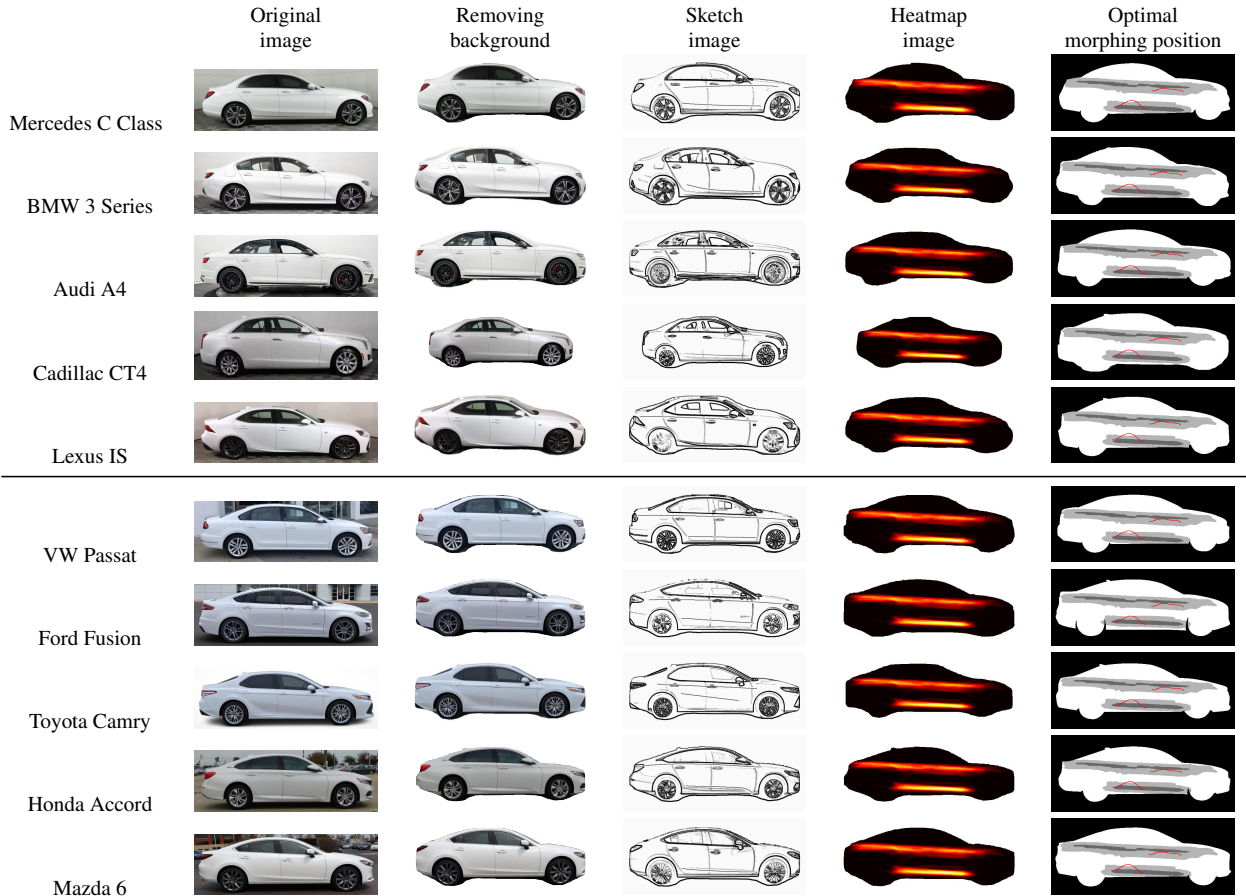


Figure 6.3.1: Summary of the key steps of the proposed design framework.

6.3.3 Optimal SMM Selection

As described earlier in Section 6.2.1, 1,645 frames of the running cheetah can be obtained from our collected video data, and they can be clustered into 14 SMMs. Each SMM can be represented or visualised as a heatmap image by stacking consecutive curve images.

Three metrics are used to evaluate different aspects of SMMs. First, the *noise* metric can capture the appearance of noise. Within our context, the noise refers to the discontinuity of cheetah body curves. In some images, the curves might be regionally discontinued due to the blueness of the cheetah body shape in the original images, when there is no clear colour contrast between the background and the cheetah body. We use the noise metric to measure the cleanness of the stacked curves because it is pixel sensitive so the metric signs the SMM with discontinuous

Table 6.3.1: Selection of the SMM.

SMM	Noise	Rank	Density	Rank	Straightness	Rank	Total rank
Frames 1360 to 1400	0.303	11	0.24	12	1.669	10	12
Frames 1410 to 1440	0.164	2	0.16	2	1.56	5	1
Frames 1445 to 1475	0.271	8	0.171	3	1.529	3	4
Frames 1860 to 1905	0.23	3	0.2	6	1.349	1	2
Frames 1910 to 1955	0.33	14	0.24	11	1.456	2	10
Frames 2385 to 2440	0.248	6	0.216	9	1.672	11	8
Frames 3590 to 3630	0.263	7	0.202	7	1.676	12	9
Frames 3760 to 3810	0.245	5	0.216	8	1.532	4	6
Frames 3815 to 3870	0.319	12	0.256	13	1.604	7	11
Frames 5580 to 5615	0.152	1	0.118	1	1.657	9	3
Frames 5885 to 5920	0.287	10	0.189	5	1.586	6	7
Frames 5925 to 5985	0.231	4	0.178	4	1.643	8	5
Frames 6735 to 6790	0.28	9	0.239	10	1.725	14	13
Frames 6795 to 6850	0.329	13	0.263	14	1.72	13	14

curves. Mathematically, it can be computed as follows:

$$\text{noise} = \frac{\sqrt{\pi/2}}{6(W-2)(H-2)} \sum_{w,h} (\Upsilon * X), \quad (6.14)$$

where X is the input image (single channel), W and H are the width and height of the input image, π is the mathematical constant, $*$ represents the convolution operation, and Υ is the kernel matrix, given by

$$\Upsilon = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & 2 \\ 1 & -2 & 1 \end{bmatrix}.$$

Second, the *density* metric can measure the degree of concentration, that is, how compact a cheetah body curve point shown in specific positions. A higher density score here indicates the cheetah has held the corresponding body posture for a longer period during the motion, therefore resulting in more curves stacked on the same positions. The density metric can be computed as

$$\text{density} = \frac{\sum_{w,h} \mathbb{I}_{X_{(w,h)} > \varepsilon}}{W \cdot H}, \quad (6.15)$$

where ε is a threshold constant, $\mathbb{I}_{X_{(w,h)} > \varepsilon}$ indicates whether the pixel $X_{(w,h)}$ is larger than ε .

Last, the *straightness* metric can measure the average Euclidean distance between curve

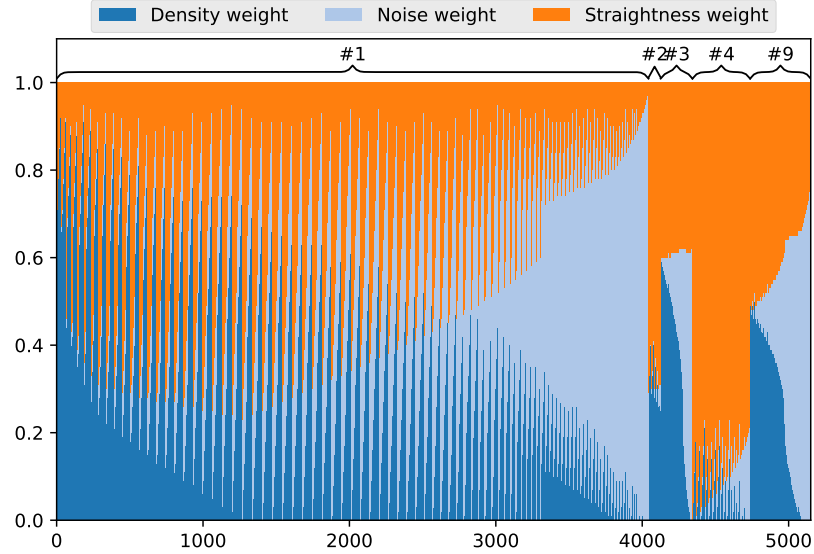


Figure 6.3.2: Robustness check of the ranking for frames in “1410 to 1440”.

points. It indicates how relatively straight an extracted curve is and whether it contains discontinued parts. We design this measure because our analysis shows many existing car body curves are relatively straight, compared to cheetah. The straightness metric can be computed as

$$\text{straightness} = \frac{1}{\tilde{N}} \sum_{\tilde{n}=2}^{\tilde{N}} \left[(x_{\tilde{n}} - x_{\tilde{n}-1})^2 + (y_{\tilde{n}} - y_{\tilde{n}-1})^2 \right]^{1/2}, \quad (6.16)$$

where x and y are the point coordinate for \tilde{N} curve points sorted according to their x position.

Table 6.3.1 presents the computed three metrics for clustered 14 SMMs of our cheetah data, their corresponding rankings, and the final ranking. Here the final ranking is calculated based on the equal weights of three metrics so frames in “1410 to 1440” are selected as the best SMM. We also check the robustness of this selection by investigating different weight combinations of metrics. Figure 6.3.2 shows that frames in “1410 to 1440” rank in the first position more than 80% of cases, which verifies our SMM selection.

6.4 Empirical Investigation

This section presents the results of our empirical investigation. We start with a pretest to examine whether activating the cheetah schema indeed prompts schema congruity of sports saloons as we hypothesised in Section 6.2.1. We then run another study to establish observational

evidence for the proposed design framework by automatically generating different aesthetics designs of sports saloons and to evaluate consumer's preference.

6.4.1 Pretest of Schema Activation

We test the hypothesis if consumers would rate the same automobile more sporty or more dynamic looking when we activate the cheetah schema. A total of 220 survey participants were recruited via the Amazon Mechanical Turk (in short MTurk). To ensure quality of our responses, all the participants met the following criteria: (i) resides in the United States; and (ii) has more than 500 tasks completed and approved on the MTurk, with approval rate no less than 99%.

Pretest Design

Our pretest has a two by one between-subject design. Participants were asked to evaluate the design of a new car, and they were randomly assigned to the cheetah and control conditions. In both conditions, we display an identical car image to the participants and tell them that the engineers have worked to improve the overall appearance of the car (see Figures 6.4.1-6.4.2). The only difference between the two conditions is that in the cheetah condition, as shown in Figure 6.4.2, participants were displayed with both car and cheetah images, and they were told that the overall appearance of the car has been inspired and modelled after cheetah, the fastest land animal on earth. Next, participants in both conditions evaluated the appearance of the car on four 9-point Likert scale (i.e., 1 = "Strongly Disagree" and 9 = "Strongly Agree") questions regarding whether: "(i) this car can accelerate very fast; (ii) this car is exotic; (iii) this car is fast; and (iv) this car is powerful". On average, participants took 118 seconds to complete the survey. Responses from eight participants were dropped because the duration of their response time was



In what follows, you will be asked to evaluate the displayed car image.

The engineers have worked to improve the overall appearance of this car. We would like to get your opinions. Please take a look and let us know what you think.

Figure 6.4.1: Example of the control condition in the survey.



In what follows, you will be asked to evaluate the displayed car image.

The engineers have worked together with bionics experts such that the overall appearance of this car have been inspired and modeled after *cheetah*, the fastest land animal on earth. We would like to get your opinions. Please take a look and let us know what you think.

Figure 6.4.2: Example of the cheetah condition in the survey.

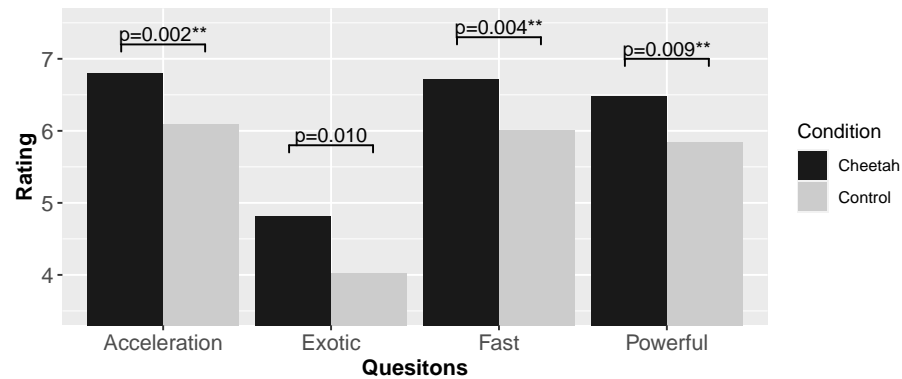


Figure 6.4.3: Users' responses to aesthetics design, where ** represents the p value less than 0.01.

either less than a half minute or more than an hour.

Results

Figure 6.4.3 presents the results of our pretest. We discover that, although the exactly same car image was shown in both conditions, participants in the cheetah condition perceived the car to accelerate faster ($\text{Mean}_{\text{acceleration}}^{(\text{cheetah})} = 6.80$ vs $\text{Mean}_{\text{acceleration}}^{(\text{control})} = 6.09$ and $p = 0.002$), more exotic ($\text{Mean}_{\text{exotic}}^{(\text{cheetah})} = 4.81$ vs $\text{Mean}_{\text{exotic}}^{(\text{control})} = 4.02$ and $p = 0.010$), faster ($\text{Mean}_{\text{fast}}^{(\text{cheetah})} = 6.72$ vs $\text{Mean}_{\text{fast}}^{(\text{control})} = 6.00$ and $p = 0.004$) and more powerful ($\text{Mean}_{\text{powerful}}^{(\text{cheetah})} = 6.48$ vs $\text{Mean}_{\text{powerful}}^{(\text{control})} = 6.00$ and $p = 0.009$).

5.84 and $p = 0.009$). Overall, the pretest provides support for our conjecture that activating the cheetah schema enhances consumers' perception about the sportiness of car image, even when the appearance of the car was kept constant. Such encouraging results from our pretest lead us to continue our investigation into whether consumers indeed prefer the aesthetics of sports saloon more when we employ a bio-inspired automobile design as detailed in Section 6.2.

6.4.2 Bio-Inspired Aesthetics Design

We then conduct a proof of concept study in which we ask participants to select their most preferred aesthetics design for ten popular sports saloons. For each car model, we provide a spectrum of aesthetic designs ranging from its original look to low/moderate/considerable changes in shoulder and waist curves generated from the deep generative model as described in Section 6.2.

We selected ten popular sports saloons (five luxury and five economy models) based on the 2019 automobile rankings from U.S. News and JD Power.⁹ Specifically, our empirical investigation includes Mercedes C Class, BMW 3 Series, Audi A4, Cadillac CT4, and Lexus IS for luxury sports saloons; and VW Passat, Ford Fusion, Toyota Camry, Honda Accord and Mazda 6 for economy sports saloons.

Our proposed design framework mainly uses the CycleGAN for design generation. Different to conventional supervised learning problems where there is a metric to evaluate the model's performance for the in-sample and out-of-sample data, training the CycleGAN that can generalise well in design generation is a challenging task. Thus, we perform two operations in model training to avoid the situation that the CycleGAN only learns the design patterns from the targeted models (i.e., over-fitting). First, we collect car side-view images of 50 different saloon models from 36 automakers from autotrader.com. For each model, 20 side-view images are sampled from the cars registered between 2010 and 2019. Second, to further incorporate more design patterns and variations, the collected images are augmented with random cropping which results in 280,000 car images in total (i.e., 20x14 possible positions per each image). For the reader's reference, the key processing steps of the proposed design framework for the targeted saloon models are provided in Figure 6.3.1.

Next, we generated 20 morphed images for each of the ten car models in our focal study, with the morphing parameter β ranging from 0 and 0.7. Figure 6.4.4 provides three cheetah-inspired

⁹URL: <https://cars.usnews.com/cars-trucks/rankings/used/2019-luxury-small-cars> (U.S.News); and <https://www.jdpower.com/cars/saloons/10-most-popular-midsize-cars> (J.D.Power).

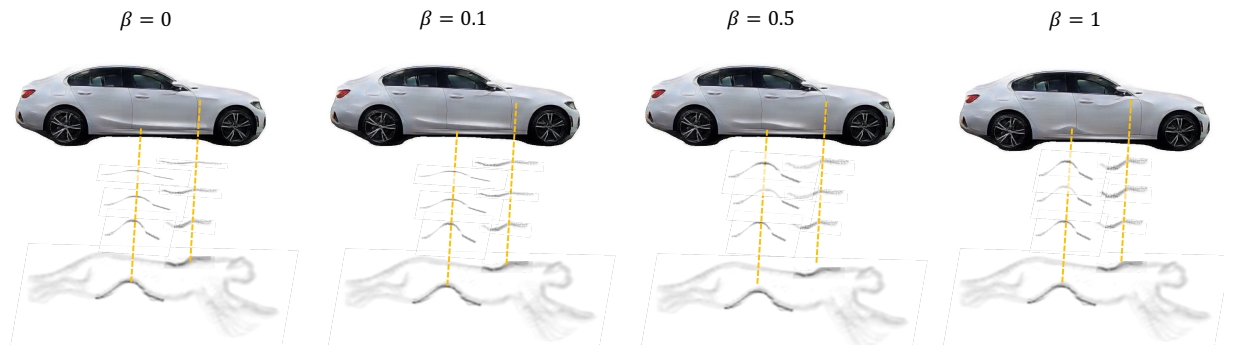


Figure 6.4.4: Illustration of new designs of BMW 3 Series generated by the proposed framework.

design examples for BMW 3 series, including the design when $\beta = 1$. One can easily describe the differences of modifications between the designs when the value of morphing parameter β increases. When more cheetah features are incorporated (i.e., increasing β), the shoulder and waist curves of the car become shorter and curvier like the running cheetah. Therefore, Figure 6.4.4 shows that by adjusting the morphing parameter, our framework can quantify the design generation and can support the decision-making in the automotive design process for designers by offering a number of new designs inspired by cheetah. Also, although 100% morphing is achievable (i.e., $\beta = 1$), it is very unlikely that automotive manufacturers would use the exact same shape of curves from the running cheetah in automotive design. We further created a survey that contains ten pages, with each page corresponding to one of the sports car models in our study. In each page, we present a drag bar in which the survey participant can move the bar on the slider back and forth to different locations to view different aesthetic designs of the car model. Each participant is shown 21 different designs of the focal car model, with 0 being the car's current look and 20 being the most significant change in this look as generated based on the approach outlined in Section 6.2 above.

Study Design

We carry out a survey study to explore whether consumers prefer the cheetah-inspired automobile designs more than their original look. A total of 212 participants were recruited via the MTurk. To ensure high quality responses to our survey, only participants satisfying the following criteria were allowed to take part in our study: (i) the participants are located in the United States; and (ii) they have more than 500 approved tasks on MTurk, with approval rates not less than 99%. Our survey included a video-based instruction and two attention check questions to ensure that the participants understood and paid attention to our survey questions. The par-

Merced C Class

Please move the bar on the slider back and forth to different locations to view different aesthetic designs of the car model, where 0 represents the car model's current look and 20 represents the most significant change in its look. Please move the bar on the slider to the spot that corresponds to the most appealing design to you.

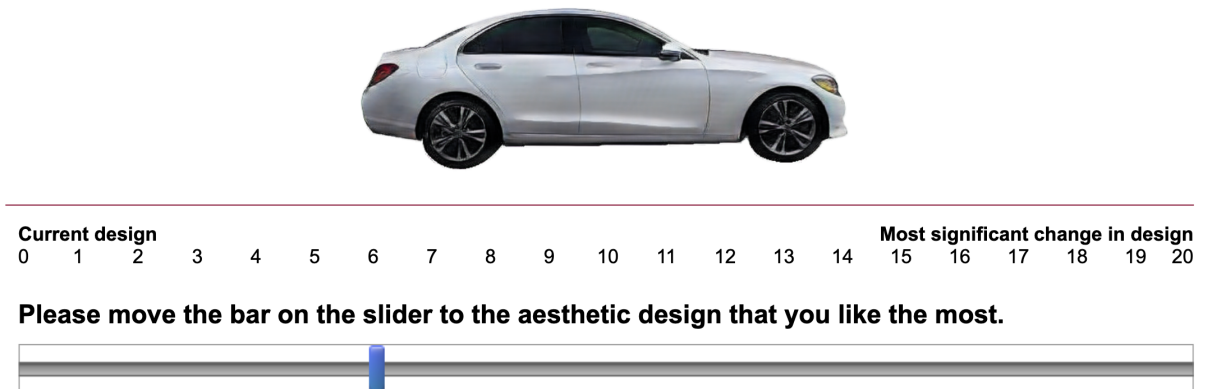


Figure 6.4.5: Example of design selection for BMW 3 Series in the user survey study.

Participants would be disqualified to take part in the study if they provided wrong answers to the attention check questions. During the survey, participants were displayed with our targeted ten saloon models and were told to identify the most appealing designs to them for each of the ten sports saloons in our study. For each model, they were instructed to drag the scroll bar back and forth until they found the most appealing look to them.¹⁰ The participants took 313.66 seconds on average to complete the survey. The survey duration of all participants appeared to be reasonable, with no one providing the same responses to all ten sports saloons in our study.

Figure 6.4.5 presents an example of design selection for BMW 3 Series. Participants can choose their preferred design using the scroll bar, where 0 represents the current design (i.e., $\beta = 0$) and 20 represents the most significant change in design (i.e., $\beta = 0.7$). Thus, we set the upper bound 0.7 for β in the empirical study to allow a certain level of cheetah features can be incorporated as well as to have a finer scale of survey questions in exploring consumer response.

Results

We compare the original design and the new design preferred by the participants, which is measured by taking the average of their morphing responses. In order to better quantify and compare the original and the new designs, we detect the shoulder and waist curves using

¹⁰URL: <https://youtu.be/-DJls9uy-oA>

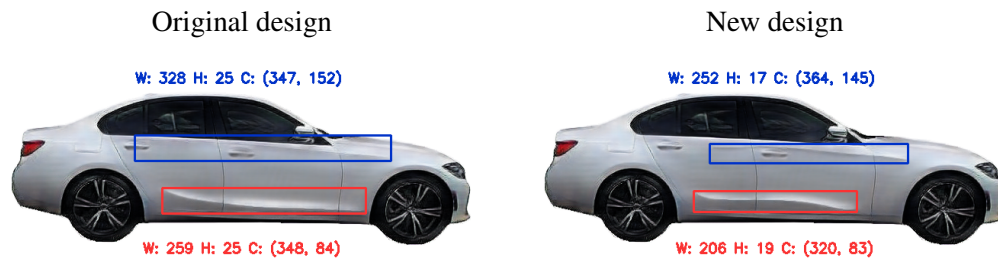


Figure 6.4.6: Comparison of original and new designs for BMW 3 Series, where the blue box highlights the shoulder curve and the red box highlights the waist curve; the width (W), the height (H) and the central position (C) of each box in the car side-view image are presented.

rectangular boxes, and report their position related measures such as width (W), height (H) and central position (C). It should be noted that the central position contains both horizontal and vertical values of the box's centroid. Figure 6.4.6 presents an illustration example of BMW 3 Series. Notable changes can be seen in the new design, where waist and shoulder curves have a smaller width and a shorter height. The central position of the waist curve moves towards the front of the car while the waist curve moves a little backward. It is our belief that such changes have made the car appear to be more aesthetically.

In Figures 6.4.7 and 6.4.8, we compare the preferred new designs with the original designs for our target premium and mainstream sedans, respectively. To more clearly highlight the differences, the rectangle boxes are omitted in these figures. However, the measurements for the waist and shoulder curves, including width, height, and central position, are still displayed.

Saloon models can be broadly categorised into three types based on their original shoulder curve design: **Short Shoulder Curves:** This category encompasses the BMW 3 Series and Mazda 6, which have shoulder curves that situate near the front wheel. The redesigned curves for both are even shorter and have more pronounced curvature. Specifically, the curve shifts slightly forward for the BMW 3 Series and moves backward for the Mazda 6. **Medium-Length Shoulder Curves:** The Mercedes C Class and Toyota Camry belong to this group. Their shoulder curves are moderately long, positioned around the centre of the car body. The redesigned versions of these models see a significant reduction in both the length and height of these curves, imbuing the cars with a sportier appearance. **Long Shoulder Curves:** Saloons in this category, particularly the Audi A4, VW Passat, and Ford Fusion, have extensive shoulder curves, accounting for more than 70% of the car body's length. These curves span from the car's front to its

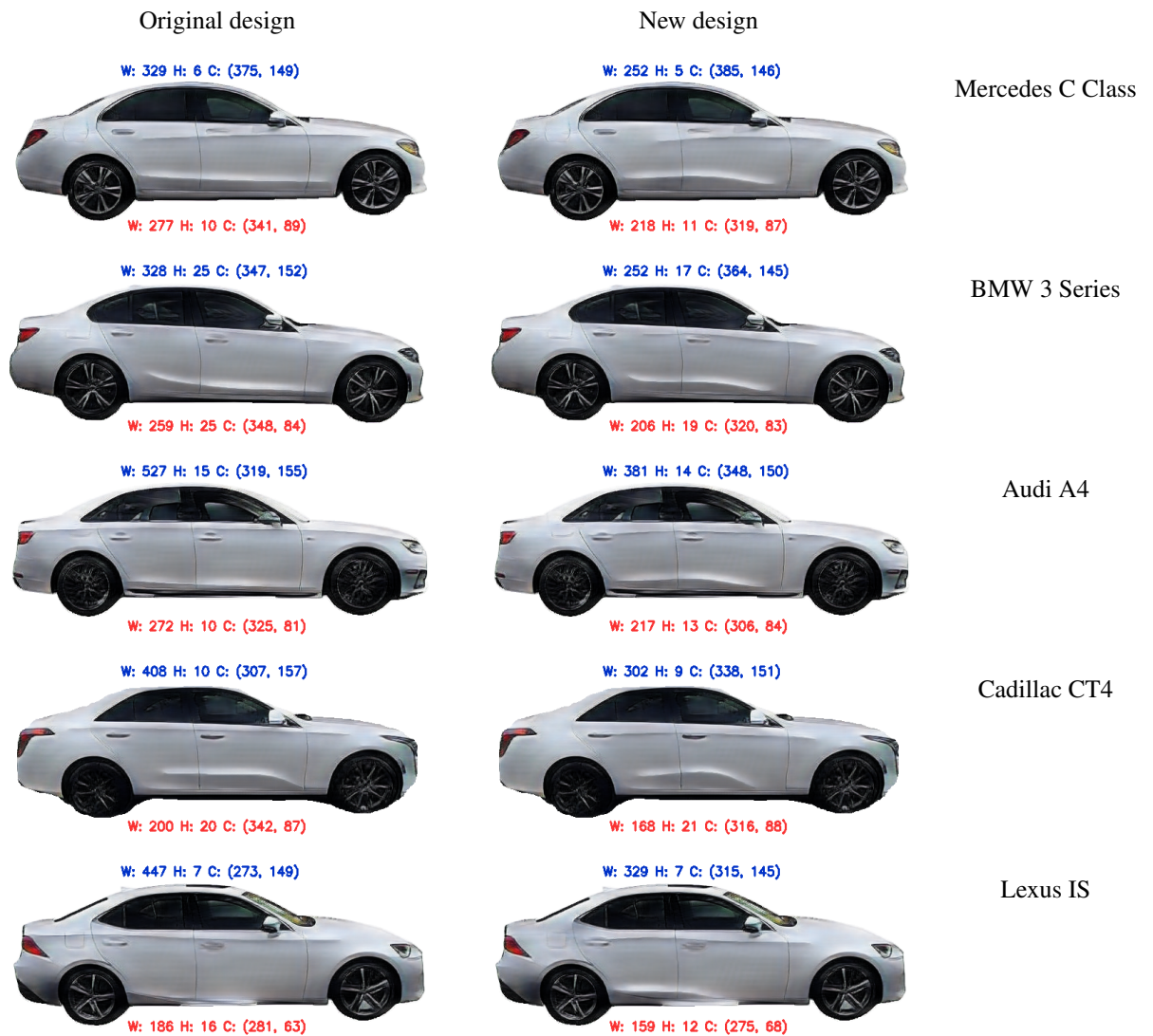


Figure 6.4.7: Original and new designs of the targeted premium saloons.

rear. Redesigns for these models experience a considerable length reduction of around 25%. Such categorisation provides a framework to understand design modifications in relation to the original shoulder curve structure.

Regarding the waist curve, the target sedan models can also be grouped into three categories: **Short and Low Waist Curves:** Lexus IS fall into this category. They feature waist curves that are shorter in length and situated lower. The redesigned models see reductions in both the widths and heights of these waist curves, with their positions shifting upwards. **Moderate-Length but High Waist Curves:** Toyota Camry represents this group. Its waist curve is moderately long



Figure 6.4.8: Original and new designs of the targeted mainstream saloons.

but occupies the highest position among the models. The redesign leads to reductions in both its waist curve's width and height, while the position descends. **Various-Length Waist Curves:** The remaining sedan models make up this category. In their new designs, the width of the waist curves generally reduces. However, the height of these waist curves either increases or remains unchanged with the exception of the BMW 3 Series, which sees a height reduction. Overall, participants' preferred designs consistently lean towards shorter and curvier shoulder and waist curves, combined with a more distinct slope.

In comparison to the original models, some sedans in our study exhibit more pronounced

modifications based on consumer preference, while others display subtler, yet discernible changes. Intriguingly, participants preferred a slightly greater morphing intensity for premium sedans than mainstream ones ($Mean_{premium} = 0.311$ vs. $Mean_{mainstream} = 0.297$, $p = 0.087$). This indicates a stronger inclination towards the cheetah curve design for luxury vehicles. Moreover, the sportier sedan models, characterized by more pronounced shoulder and waist curves, registered even higher morphing preferences ($Mean_{sporty} = 0.327$ vs. $Mean_{regular} = 0.304$, $p = 0.014$), signifying the versatility of our design approach across sedan categories.

6.5 Summary

As the last piece of my PhD works, this chapter proposes a bio-inspired automotive aesthetics design upgrading using CycleGAN. We provide a proof-of-concept design process using the example of sports saloon facelifts as inspired by the fastest land animal on earth, cheetah. Results show that consumers indeed find the cheetah-inspired new automobile design to be more aesthetically appealing. This study has two main contributions. First, unlike the previous chapters, this chapter focuses on upgrading the side-view designs, showing deep learning-based design upgrading is not confined to particular viewpoints. Second, as the originality is restricted in data-driven models, our developed approach offers a method of acquiring novel design features that are not sourced from existing designs.

Chapter 7

Conclusions and Future Work

This chapter will summarise the key findings from my PhD and the primary contributions of this thesis. Meanwhile, it discusses a few future research directions.

7.1 Conclusion and Contributions

As mentioned in Chapter 1, the present PhD thesis aims to apply deep learning algorithms to automotive aesthetic design and analytics by developing frameworks that can both deliver graphic upgraded designs and offer optimal marketing decision suggestions. In particular, four research questions are proposed to address: (i) How to use generative models to upgrade extant car appearance designs regionally? (ii) How to develop data-driven models that can automatically evaluate car designs in aesthetics? (iii) How to incorporate the design generating into a profit optimisation framework, which can assess unseen designs in terms of the potential market profit? (iv) How to incorporate visual cues such as animal-inspired features into the automotive design to have novel designs? Through Chapters 3 to 6, four of studies have been done to address these research questions. In particular:

- In Chapter 3 (DVM-CAR Dataset), we developed the very first large-scale automotive dataset that contains 1.4 million images from other car specifications. On the one hand, the developed dataset offers a sound data base for my PhD, which supports various automotive styling-related studies. On the other hand, the development of DVM-CAR dataset meets the growing need for a comprehensive automotive dataset for applying machine learning to product design, which can be used for different types of analytics, ranging from descriptive to diagnostic, predictive and perspective.

- In Chapter 4 (GEO), we developed a three-module framework that contains a design generator, design evaluator and decision optimiser, which provides end-to-end decision-support solutions for automotive designers and manufacturers. Different from previous studies that employed deep generative models for automotive design generating, our study focuses on offering regional design upgrades rather than new designs, where the primary shape of existing car models is untouched. Meanwhile, this study is the first work incorporating design generating into a profit optimisation framework.
- Chapter 5 (Trendiness) added to the discussion of the impacts of trendiness/modernness in product styling and offered solutions to upgrade the car's front when it is outmoded. Different from Chapter 4, the image completion model is adopted for design upgrading, which is not affected by the feature entanglement problem. The main innovation is to have the system learn, define and score trendiness from existing car front images, which in turn produces massive unrepresented but highly scored modern designs.
- In Chapter 6 (Animism), we proposed an approach that can identify noteworthy curves from the cheetah body and incorporate them into the automotive side's design. It is the first study that developed a data-driven computational method for bio-inspired automotive aesthetic design. As a complementary investigation to the previous two studies, it focuses on upgrading the car's side designs, showing that automotive design upgrading is not limited to particular viewpoints. Moreover, this study demonstrates a practical way of having originality in design that is not constrained by the used training samples.

7.2 Directions for Future Work

- Tailored Optimisation Algorithms for Variant Settings

Although we have developed an optimisation algorithm for design selection, the proposed algorithm is rather straightforward regarding the problem's complex nature. It is foreseeable that with a few prior works that integrate deep generation into product design optimisation, more researchers will be attracted and interested in joining the relevant studies. Therefore, developing more sophisticated algorithms for the decision-making of automotive facelifts is a promising direction with many possible topics. For example, we can alter the optimisation problem to a constraint one by estimating the achievable design space. That is, researchers can use the trained GAN model to compute the facelift's market positioning range in the aesthetic space for optimal solution searching. Or have the multiplayer

games setting by clustering cars in market subgroups, where each group member is the contender for others within the same submarket. In short, the integration of generative methods could make product optimisation problems more tangible and exciting.

- In Chapter 5, we try to predict the future aesthetic preference shift in the automotive market without discussing the underline reasons or mechanisms for fashion changes. Psychological studies regarding typicality, such as [Hekkert et al. \(2003\)](#), show consumers have a strong tendency to prefer products with highly representative looks. [Landwehr et al. \(2009\)](#) marked that typicality, together with complexity, explains 42 per cent of the variation within six months of car sales records from the German Federal Transport Authority. It is likely that the evolution of fashion changes in the automotive market is driven by the shifting of typicality – as time pass, the newly launched cars would gradually change consumers’ images of cars in various submarkets. To this end, researchers can apply deep models to construct prototypical designs in different years and then investigate whether such changes could explain the shifts in automotive sales.

- Aesthetics in Used Car Valuation

My PhD studies show that deep neural network models can successfully identify the critical aesthetic factors for new car sales when large-scale data is available. However, little is known to what extent aesthetic factors determine the depreciation of cars. Existing studies ([Englmaier et al., 2018](#); [Huang et al., 2019](#)) found the depreciation of car values is affected by multiple aspects, such as the discontinuity in prices and the fluctuating demands. Particularly, investigations can be done to verify the widely held arguments that whether design with more complex styles makes a car more vulnerable to value depreciation and do the showy colours make cars unattractive in the used market.

7.3 Potential Impact and Transferability

From a research standpoint, our study can be viewed as extending traditional product design optimisation by creating a graphical design space using deep generation techniques for optimal design searches. In contrast to our approach, traditional research methods use attribute profiles to represent designs, using scores to highlight design characteristics like sportiness and modernity. As this approach simplifies the task by bypassing the need for graphical designs, it is less appealing to practitioners. Automakers typically prefer tangible design solutions that can inform their decision-making processes. Given these factors, we anticipate that our approach, which in-

tegrates deep generative models with the product design optimisation framework, will garner significant interest from the broader research community.

Regarding transferability, our application of deep learning to car exterior design generation and evaluation provides a model that can be replicated across different commodities. While interest in machine learning has grown within business academia, substantial research and tangible applications are still sparse. Our approach exemplifies both rigorous research and the practical use of advanced machine learning technologies in the product design domain. The performance of deep learning models often depends on the amount of data available. Our methods used for studying car exterior design can easily be adapted to other products, with the main challenge being data collection and preparation. From this viewpoint, the limited use of machine learning in business often stems from reluctance to share valuable research datasets. Our work in collecting, processing, and structuring automotive data could set a precedent for analogous research.

Bibliography

- Aggarwal, P. and McGill, A. L. (2007). Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *Journal of Consumer Research*, 34(4):468–479.
- Ahmed, R. R., Parmar, V., and Amin, M. A. (1450). Impact of product packaging on consumer's buying behavior. *European Journal of Scientific Research*, 120(2):1450–216.
- Ballester, C., Caselles, V., and Verdera, J. (2003). Disocclusion by joint interpolation of vector fields and gray levels. *Multiscale Modeling and Simulation*, 2(1):80–123.
- Bar-Eli, S. (2013). Sketching profiles: Awareness to individual differences in sketching as a means of enhancing design solution development. *Design Studies*, 34:472–493.
- Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. (2009). PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24.
- Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C. (2000). Image inpainting. *Conference on Computer Graphics and Interactive Techniques*, 71(7):417–424.
- Bertalmio, M., Vese, L., Sapiro, G., and Osher, S. (2003). Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889.
- Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.
- Blijlevens, J., Creusen, M. E. H., and Schoormans, J. P. L. (2009). How consumers perceive product appearance: The identification of three product appearance attributes. *International Journal of Design*, 3(3):27–35.
- Bloch, P. H. (1995). Seeking the ideal form: Product design and consumer response. *Journal of Marketing*, 59(3):16–29.

- Blonigen, B., Knittel, C., and Soderbery, A. (2013). Keeping it fresh: Strategic product redesigns and welfare. Technical report, National Bureau of Economic Research, Cambridge.
- Burnap, A., Hartley, J., Pan, Y., Gonzalez, R., and Papalambros, P. Y. (2016a). Balancing design freedom and brand recognition in the evolution of automotive brand styling. *Design Science*, 2:1–28.
- Burnap, A., Hauser, J. R., and Timoshenko, A. (2021). Design and Evaluation of Product Aesthetics: A Human-Machine Hybrid Approach. *SSRN*, page 1907.07786.
- Burnap, A., Liu, Y., Pan, Y., Lee, H., Gonzalez, R., and Papalambros, P. Y. (2016b). Estimating and exploring the product form design space using deep generative models. *Design Automation Conference*, pages 1–13.
- Burnap, A., Pan, Y., Liu, Y., Ren, Y., Lee, H., Gonzalez, R., and Papalambros, P. Y. (2016c). Improving design preference prediction accuracy using feature learning. *Journal of Mechanical Design*, 138(7):071404.
- Cadavid, A. M., Ruiz, S., and Maya, J. (2016). What is an ugly car? A grounded theory approach on a webpage’s comments. *International Conference on Design and Emotion*, pages 449–458.
- Chen, X., Duan, Y., Houthoofd, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). InfoGAN: Interpretable representation learning. *Advances in Neural Information Processing Systems*, pages 2172–2180.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734.
- Chopra, S., Hadsell, R., and LeCun, Y. (1997). Learning a similarity metric discriminatively, with application to face verification. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 539–546.
- Coates, D. (2003). *Watches tell more than time: Product design, information*. McGraw-Hill Companies.
- Codd, E. F. (1983). A relational model of data for large shared data banks. *Communications of the ACM*, 26(1):64–69.

- Creusen, M. E. H. and Schoormans, J. P. L. (2005). The different roles of product appearance in consumer choice. *Journal of Product Innovation Management*, 22(1):63–81.
- Dant, T. and Martin, P. J. (2001). By car: Carrying modern society. *Ordinary Consumption*, 44(0):143–157.
- Dawar, N. and Parker, P. (1994). Marketing universals: Consumers' use of brand name, price, physical appearance, and retailer reputation as signals of product quality. *Journal of Marketing*, 58(2):81.
- Deng, J., Dong, W., Socher, R., Li, L., Kai, L., and FeiFei, L. (2009). ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.
- Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4685–4694.
- DiSalvo, C. and Gemperle, F. (2003). From seduction to fulfillment: The use of anthropomorphic form in design. *International Conference on Designing Pleasurable Products and Interfaces*, pages 67–72.
- Dong, J., Jiang, W., Huang, Q., Bao, H., and Zhou, X. (2019). Fast and robust multi-person 3D pose estimation from multiple views. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.
- Down, B. (2010). *Art deco and british car design: The airline cars of the 1930s*. Veloce Publishing Ltd.
- Eisenthal, Y., Dror, G., and Ruppin, E. (2006). Facial attractiveness: Beauty and the machine. *Neural Computation*, 18(1):119–142.
- Englmaier, F., Schmöller, A., and Stowasser, T. (2018). Price discontinuities in an online market for used cars. *Management Science*, 64(6):2754–2766.
- Ernst, F. U., Sebastian, K., and Thomas, V. (2017). Automotive styling: Supporting engineeringstyling convergence through surface-centric knowledge based engineering. *International Conference on Engineering Design*, 4:139–148.

- Farabet, C., Couprie, C., Najman, L., and LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1915–1929.
- Fiske, S. (1982). Schema-triggered affect: Applications to social perception. *Affect and cognition: Annual Carnegie Mellon Symposium on Cognition*, pages 55–78.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202.
- Girshick, R. (2015). Fast R-CNN. *IEEE International Conference on Computer Vision*.
- Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004). *Digital image processing using MATLAB*. Pearson Education India.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*, pages 2672–2680.
- Gray, D., Yu, K., Xu, W., and Gong, Y. (2010). Predicting facial beauty without landmarks. *European Conference on Computer Vision*, pages 434–447.
- Green, P. E., Carroll, J. D., and Goldberg, S. M. (1981). A general approach to product design optimization via conjoint analysis. *Journal of Marketing*, 45(3):17.
- Greim, L. (2017). *The impact of the model life cycle on the residual car value in the leasing industry*. PhD thesis, University of Twente.
- Happian-Smith, J., editor (2001). *An introduction to modern vehicle design*. Elsevier.
- He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask R-CNN. *IEEE International Conference on Computer Vision*, pages 2980–2988.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.
- Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. Science Editions, New York.

- Heisley, D. D. and Levy, S. J. (1991). Autodriving: A photoelicitation technique. *Journal of Consumer Research*, 18(3):257.
- Hekkert, P., Snelders, D., and Van Wieringen, P. C. (2003). 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology*, 94(1):111–124.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in Neural Information Processing Systems*, pages 6627–6638.
- Hoffer, G. E. and Reilly, R. J. (1984). Automobile styling as a shift variable: An investigation by firm and by industry. *Applied Economics*, 16(2):291–298.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- Hsiao, S. W., Chiu, F. Y., and Chen, C. S. (2008). Applying aesthetics measurement to product design. *International Journal of Industrial Ergonomics*, 38(11-12):910–920.
- Hsu, C. L., Chen, Y. C., Yang, T. N., Lin, W. K., and Liu, Y. H. (2018). Does product design matter? Exploring its influences in consumers' psychological responses and brand loyalty. *Information Technology and People*, 31(3):886–907.
- Huang, G., Luo, H., and Xia, J. (2019). Invest in information or wing it? A model of dynamic pricing with seller learning. *Management Science*, 65(12):5556–5583.
- Huang, J., Chen, B., Luo, L., Yue, S., and Ounis, I. (2022). DVM-CAR: A large-scale automotive dataset for visual marketing research and applications. *IEEE International Conference on Big Data*, pages 4130–4137.
- Huang, X., Wang, P., Cheng, X., Zhou, D., Geng, Q., and Yang, R. (2020). The apolloScape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2702–2719.
- Hung, W. K. and Chen, L. L. (2012). Effects of novelty and its dimensions on aesthetic preference in product design. *International Journal of Design*, 6(2):81–90.

- Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5967–5976.
- Jaafarnia, M. and Adele, B. (2011). Tracing the evolution of automobile design: Factors influencing the development of aesthetics in automobiles from 1885 to the present. *International conference on Innovative Methods in Product Design*, pages 8–12.
- Jia, H. M., Park, C. W., and Pol, G. (2015). Cuteness, nurturance, and implications for visual product design. In *The Psychology of Design: Creating Consumer Appeal*, pages 168–179. Routledge, Abingdon.
- Jindal, R., Sarangee, K., Echambadi, R., and Lee, S. (2016). Designed to succeed: Dimensions of product design and their impact on market share. *Journal of Marketing*, 80(4):72–89.
- Karras, T., Laine, S., and Aila, T. (2018). A style-based generator architecture for generative adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4396–4405.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of StyleGAN. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8107–8116.
- Kaul, A. and Rao, V. R. (1995). Research for product positioning and design decisions: An integrative review. *International Journal of Research in Marketing*, 12(4):293–320.
- Kelley, H. J. (1960). Gradient theory of optimal flight paths. *ARS Journal*, 30(10):947–954.
- Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. *International Conference on Learning Representations*.
- Kókai, I., Finger, J., Smith, R. C., Pawlicki, R., and Vetter, T. (2007). Example-based conceptual styling framework for automotive shapes. *Eurographics Workshop on Sketch-based Interfaces and Modeling*, page 37.
- Kong, S., Shen, X., Lin, Z., Mech, R., and Fowlkes, C. (2016). Photo aesthetics ranking network with attributes and content adaptation. *European Conference on Computer Vision*, pages 662–679.

- Korenok, O., Hoffer, G., and Millner, E. (2010). Non-price determinants of automotive demand: Restyling matters most. *Journal of Business Research*, 63(12):1282–1289.
- Krause, J., Stark, M., Deng, J., and Feifei, L. (2013). 3D object representations for fine-grained categorization. *IEEE International Conference on Computer Vision*, pages 554–561.
- Kreuzbauer, R. and Malter, A. J. (2005). Embodied cognition and new product design: Changing product form to influence brand categorization. *Journal of Product Innovation Management*, 22(2):165–176.
- Krishna, A., Cian, L., and Sokolova, T. (2016). The power of sensory marketing in advertising. *Current Opinion in Psychology*, 10:142–147.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pages 1097–1105.
- Ku, S. (2014). Is this car looking at you? How anthropomorphism predicts fusiform face area activation when seeing cars. *PloS one*, 9(12):1–14.
- Kukova, M., Diels, C., Jordan, P., Franco-Jorge, M., Anderson, J., and Kharouf, H. (2016). Do we really know which vehicle attributes are important for customers? *International Conference on Design Emotion*.
- Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Mallocci, M., Kolesnikov, A., Duerig, T., and Ferrari, V. (2020). The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision*, 128(7):1956–1981.
- Landwehr, J. R., Labroo, A., and Herrmann, A. (2011a). Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science*, 30(3):416–429.
- Landwehr, J. R., Labroo, A. A., and Herrmann, A. (2009). The pervasive effect of aesthetics on choice: Evidence from a field study. *ACR North American Advances*.
- Landwehr, J. R., McGill, A., and Herrmann, A. (2011b). It’s got the look: The effect of friendly and aggressive “facial” expressions on product liking and sales. *Journal of Marketing*, 75(3):132–146.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436.

- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Leyvand, T., Cohen-Or, D., Dror, G., and Lischinski, D. (2008). Data-driven enhancement of facial attractiveness. *ACM Transactions on Graphics*, 27(3).
- Li, G., Xie, Y., Lin, L., and Yu, Y. (2017). Instance-level salient object segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 247–256.
- Li, G. and Yu, Y. (2016). Deep contrast learning for salient object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016-Decem:478–487.
- Liang, L., Lin, L., Jin, L., Xie, D., and Li, M. (2018). SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. *International Conference on Pattern Recognition*, pages 1598–1603.
- Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, L. (2014). Microsoft COCO: Common objects in context. *European Conference on Computer Vision*, pages 740–755.
- Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 936–944.
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6738–6746.
- Liu, Y., Li, Q., Sun, Z., and Tan, T. (2020). Style intervention: How to achieve spatial disentanglement with style-based generators? *ArXiv e-prints*, page 2011.09699.
- Locatello, F., Bauer, S., Lucie, M., Rätsch, G., Gelly, S., Schölkopf, B., and Bachem, O. (2019). Challenging common assumptions in the unsupervised learning of disentangled representations. *International Conference on Machine Learning*, pages 7247–7283.
- Lomonaco, V., Pellegrini, L., Rodriguez, P., Caccia, M., She, Q., Chen, Y., Jodelet, Q., Wang, R., Mai, Z., Vazquez, D., Parisi, G. I., Churamani, N., Pickett, M., Laradji, I., and Maltoni, D. (2022). CVPR 2020 continual learning in computer vision competition: Approaches, results, current challenges and future directions. *Artificial Intelligence*, 303:103635.

- Loy, G. and Eklundh, J. O. (2006). Detecting symmetry and symmetric constellations of features. *European Conference on Computer Vision*, pages 508–521.
- Lu, X., Lin, Z., Jin, H., Yang, J., and Wang, J. Z. (2015). Rating image aesthetics using deep learning. *IEEE Transactions on Multimedia*, 17(11):2021–2034.
- Madulid, J. P. A. and Mayol, P. E. (2019). Clothing classification using the convolutional neural network inception model. *International Conference on Information Science and Systems*, pages 3–7.
- Maeng, A. and Aggarwal, P. (2017). Facing dominance: Anthropomorphism and the effect of product face ratio on consumer preference. *Journal of Consumer Research*, 44(5):1104–1122.
- McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.
- Meyers-levy, J. and Tybout, A. M. (1989). Schema congruity as a basis for product evaluation. *Journal of Consumer Research*, 16:39–54.
- Miesler, L., Leder, H., and Herrmann, A. (2011). Isn't it cute: An evolutionary perspective of baby-schema effects in visual product designs. *International Journal of Design*, 5(3):17–30.
- Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. pages 1–7.
- Moral, M. J. and Jaumandreu, J. (2007). Automobile demand, model cycle and age effects. *Spanish Economic Review*, 9(3):193–218.
- NADA (2016). Value discovery: How automotive brand affects used vehicle prices. Technical report, National Automobile Dealers Association.
- Nair, S. K., Thakur, L. S., Wen, K.-w., Science, S. M., May, N., Nair, S. K., and Thakur, L. S. (1995). Near optimal solutions for product line design and selection: Beam search heuristics. *Management Science*, 41(5):767–785.
- O'connell, S. (1998). *The Car and British Society: Class, Gender and Motoring*. Manchester University Press.
- Orbay, G., Fu, L., and Kara, L. B. (2015). Deciphering the influence of product shape on consumer judgments through geometric abstraction. *Journal of Mechanical Design*, 137(8):81103.

- Pan, Y., Burnap, A., Hartley, J., Gonzalez, R., and Papalambros, P. (2017). Deep design: Product aesthetics for heterogeneous markets. *International Conference on Knowledge Discovery and Data Mining*, pages 1961–1970.
- Pan, Y., Burnap, A., Liu, Y., Lee, H., Gonzalez, R., and Papalambros, P. (2016). A quantitative model for identifying regions of design visual attraction and application to automobile styling. *International Design Conference*.
- Pratt, W. K. (1978). *Digital image processing*. Wiley-Interscience, New York.
- Purucker, C., Sprott, D. E., and Herrmann, A. (2014). Consumer response to car fronts: eliciting biological preparedness with product design. *Review of Managerial Science*, 8(4):523–540.
- Quan, H., Li, S., and Hu, J. (2018). Product innovation design based on deep learning and kansei engineering. *Applied Sciences*, 8(12):2397.
- Rajeev, K. and Krishnamurti, R. (1987). A heuristic approach to product design. *Management Science*, 33(12):1523–1533.
- Reid, T. N., Gonzalez, R. D., and Papalambros, P. Y. (2010). Quantification of perceived environmental friendliness for vehicle silhouette design. *Journal of Mechanical Design*, 132(10):1–12.
- Ren, J., Shen, X., Lin, Z., Mech, R., and Foran, D. J. (2017). Personalized image aesthetics. *IEEE international conference on computer vision*, pages 638–647.
- Ren, J. S., Yan, Q., Xu, L., and Sun, W. (2015). Shepard convolutional neural networks. *Advances in Neural Information Processing Systems*, pages 901–909.
- Reynolds, W. H. (1968). Cars and clothing: Understanding fashion trends. *Journal of Marketing*, 32(3):44.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):1–23.
- Rubera, G. (2015). Design innovativeness and product sales’ evolution. *Marketing Science*, 34(1):98–115.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1.

- Saberi, B. (2018). The role of the automobile industry in the economy of developed countries. *International Robotics and Automation Journal*, 4(3):179–180.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training GANs. *Advances in Neural Information Processing Systems*, pages 2234–2242.
- Sbai, O., Elhoseiny, M., Bordes, A., LeCun, Y., and Couprie, C. (2019). DesIGN: Design inspiration from generative networks. *European Conference on Computer Vision*, pages 37–44.
- Schaffer, J. (2015). What not to multiply without necessity. *Australasian Journal of Philosophy*, 93(4):644–664.
- Schoormans, J. P. and Robben, H. S. (1997). The effect of new package design on product attention, categorization and evaluation. *Journal of Economic Psychology*, 18(2-3):271–287.
- Scott, J. (2000). Rational Choice Theory. In *Understanding Contemporary Society: Theories of the Present*, pages 126–138. SAGE Publications.
- Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., Cristani, M., and Lepri, B. (2017). What your facebook profile picture reveals about your personality. *ACM Multimedia Conference*, pages 460–468.
- Seo, Y. and shik Shin, K. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications*, 116:328–339.
- Shelley, C. (2015). The nature of simplicity in Apple design. *The Design Journal*, 18(3):439–456.
- Sherman, R. and Hoffer, G. (1971). Does automobile style change payoff? *Applied Economics*, 3(3):153–165.
- Shi, L., Ólafsson, S., and Chen, Q. (2001). An optimization framework for product design. *Management Science*, 47(12):1681–1692.
- Silva, S. M. and Jung, C. R. (2018). License plate detection and recognition in unconstrained scenarios. *European Conference on Computer Vision*, pages 580–596.

- Simonson, A. and Schmitt, B. H. (1997). *Marketing aesthetics: The strategic management of brands, identity, and image*. Simon and Schuster.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, pages 1–14.
- Smilkov, D., Thorat, N., Kim, B., Viégas, F., and Wattenberg, M. (2017). SmoothGrad: Removing noise by adding noise. *ArXiv e-prints*, page 1706.03825.
- Smith, R. (2012). Cheetahs on the edge.
- Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *ArXiv e-prints*, page 1412.6806.
- Sun, T., Zhu, X., Pan, J. S., Wen, J., and Meng, F. (2015). No-reference image quality assessment in spatial domain. *Advances in Intelligent Systems and Computing*, 329(12):381–388.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9.
- Talke, K., Salomo, S., Wieringa, J. E., and Lutz, A. (2009). What about design newness? Investigating the relevance of a neglected dimension of product innovativeness. *Journal of Product Innovation Management*, 26(6):601–615.
- Teuwen, J. and Moriakov, N. (2020). Convolutional neural networks. In Zhou, K., Rueckert, D., and Fichtinger, G., editors, *Handbook of Medical Image Computing and Computer Assisted Intervention*, pages 481–501. Academic Press, Amsterdam.
- Tovey, M., Owen, J., and Street, P. (2000). Sketching and direct CAD modelling in automotive design. *Design Studies*, 21(6):569–588.
- Tseng, I., Cagan, J., Kotovsky, K., and Wood, M. (2013). Form function fidelity. *Journal of Mechanical Design*, 135(1):11006.
- Veryzer Jr, R. W. (1995). The place of product design and aesthetics in consumer research. *Advances in consumer research*, 22(1):631–645.
- Veryzer Jr, R. W. and Hutchinson, J. W. (1998). The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of consumer research*, 24(4):374–394.

- Voigt, P. and Von dem Bussche, A. (2017). *The EU general data protection regulation (GDPR): A practical guide*. Springer International Publishing, 1st edition.
- Wan, Z., Zhang, J., Chen, D., and Liao, J. (2021). High-fidelity pluralistic image completion with transformers. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4692–4701.
- Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., and Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274.
- Waytz, A., Heafner, J., and Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52:113–117.
- Wedel, M. and Pieters, R., editors (2008). *Visual marketing from attention to action*. Taylor & Francis Group, New York.
- Welsh, J. (2006). Why cars got angry. *Wall Street Journal*.
- Windhager, S., Slice, D. E., Schaefer, K., Oberzaucher, E., Thorstensen, T., and Grammer, K. (2008). Face to face: The perception of automotive designs. *Human Nature*, 19(4):331–346.
- Wolf, W. (1996). *Car mania: A critical history of transport*. Pluto Press.
- Xia, Y. and Sattar, J. (2019). Visual diver recognition for underwater human-robot collaboration. *IEEE International Conference on Robotics and Automation*, pages 6839–6845.
- Xie, S. and Tu, Z. (2017). Holistically-nested edge detection. *International Journal of Computer Vision*, 125(1-3):3–18.
- Yang, L., Luo, P., Loy, C. C., and Tang, X. (2015). A large-scale car dataset for fine-grained categorization and verification. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3973–3981.
- Yu, Y., Zhan, F., Wu, R., Pan, J., Cui, K., Lu, S., Ma, F., Xie, X., and Miao, C. (2021). Diverse image inpainting with bidirectional and autoregressive transformers. *ACM International Conference on Multimedia*, pages 69–78.

- Yu, Z., Feng, C., Liu, M. Y., and Ramalingam, S. (2017). CASENet: Deep category-aware semantic edge detection. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 1761–1770.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2p2):1.
- Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595.
- Zhao, L., Mo, Q., Lin, S., Wang, Z., Zuo, Z., Chen, H., Xing, W., and Lu, D. (2020). UCT-GAN: Diverse image inpainting based on unsupervised cross-space translation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5741–5750.
- Zheng, C., Cham, T.-J., and Cai, J. (2019). Pluralistic image completion. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1438–1447.
- Zheng, C., Cham, T.-j., Cai, J., and Phung, D. (2021). Bridging global context interactions for high-fidelity image completion. *ArXiv e-prints*.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE International Conference on Computer Vision*, pages 2223–2232.
- Zhu, Y., Huang, R., Wu, Z., Song, S., Cheng, L., and Zhu, R. (2021). Deep learning-based predictive identification of neural stem cell differentiation. *Nature Communications*, 12(1):1–13.