Huang, J., Chen, B., Yan, Z., Ounis, I. and Wang, J. (2022) GEO: A computational design framework for automotive exterior facelift. *ACM Transactions on Knowledge Discovery from Data*

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Copyright © 2022 Association for Computing Machinery. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in ACM Transactions on Knowledge Discovery from Data https://doi.org/10.1145/3578521

http://eprints.gla.ac.uk/287845/

Deposited on: 14 December 2022

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

GEO: A Computational Design Framework for Automotive Exterior Facelift

JINGMIN HUANG, University of Glasgow, UK
BOWEI CHEN*, University of Glasgow, UK
ZHI YAN, University of Technology of Belfort-Montbeliard, France
IADH OUNIS, University of Glasgow, UK
JUN WANG, University College London, UK

Exterior facelift has become an effective method for automakers to boost the consumers' interest in an existing car model before it is redesigned. To support the automotive facelift design process, this study develops a novel computational framework – Generator, Evaluator, Optimiser (GEO), which comprises 3 components: a StyleGAN2-based design generator that creates different facelift designs; a convolutional neural network (CNN)-based evaluator that assesses designs from the aesthetics perspective; and a recurrent neural network (RNN)-based decision optimiser that selects designs to maximise the predicted profit of the targeted car model over time. We validate the GEO framework in experiments with real-world datasets and describe some resulting managerial implications for automotive facelift. Our study makes both methodological and application contributions. First, the generator's mapping network and projection methods are carefully tailored to facelift where only minor changes are performed without affecting the family signature of the automobile brands. Second, two evaluation metrics are proposed to assess the generated designs. Third, profit maximisation is taken into account in the design selection. From a high-level perspective, our study contributes to the recent use of machine learning and data mining in marketing and design studies. To the best of our knowledge, this is the first study that uses deep generative models for automotive regional design upgrading and that provides an end-to-end decision-support solution for automakers and designers.

CCS Concepts: • **Applied computing** → *Business intelligence*; • **Computing methodologies** → *Neural networks*.

Additional Key Words and Phrases: Automotive design, exterior facelift, design generation, aesthetics evaluation, decision optimisation

ACM Reference Format:
Jingmin Huang, Bowei Chen, Zhi Yan, Iadh Ounis, and Jun Wang. 2022. GEO: A Computational Design Framework for Automotive Exterior Facelift. 1, 1 (December 2022), 21 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding author

Authors' addresses: Jingmin Huang, University of Glasgow, Glasgow, UK, 2421107h@student.gla.ac.uk; Bowei Chen, University of Glasgow, Glasgow, UK, Bowei.Chen@glasgow.ac.uk; Zhi Yan, University of Technology of Belfort-Montbeliard, France, Zhi.Yan@utbm.fr; Iadh Ounis, University of Glasgow, Glasgow, UK, Iadh.Ounis@glasgow.ac.uk; Jun Wang, University College London, London, UK, Jun.Wang@cs.ucl.ac.uk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2022 Association for Computing Machinery.
Manuscript submitted to ACM

1 INTRODUCTION

> "... which brings us on to the new Golf GTI. Well, VW says it is new — in reality it is a facelift of the current MK 7 — in the hope that existing owners will feel compelled to sell their old model and sign on the dotted line of whatever nonsensical finance arrangement the beancounters have come up with this time. "
>
> - Jeremy Clarkson
> *The Clarkson Review: 2017 VW Golf GTI*

The automotive sector is a major industrial and economic force worldwide. Automotive markets are very well developed in many countries, and cars are becoming more homogeneous in prices and functions. Therefore, exterior styling has played an important role in any market success [1, 2]. According to J.D. Power [3], which surveyed consumers who had just bought a car, exterior styling has been considered a strong motivation for their purchase. In fact, it has become one of the most critical determinants for automotive sales, and a hot spot for market competition.

Most automakers launch a new generation of a given car model every six to eight years. 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*) [4]. 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. 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.

Building on the work from [5, 6], which apply deep learning algorithms to assist the automotive aesthetic 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 proposed framework contains three key components – a *design generator*, a *design evaluator* and a *decision optimiser* – hence we simply call it *GEO*. The design generator proposes various upgrading schemes for automotive exterior facelift; the evaluator assesses the designs from the aesthetic perspective; and the decision optimiser selects the best design that maximises the overall profit for the automaker. Specifically, the design generator is based on the state-of-the-art StyleGAN2 model [7, 8] 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. 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 [9, 5] 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 [10, 11, 12] to

Fig. 1. Examples of car front facelifts (designs before and after the facelifts) for Audi A3 and Lexus NX.

facilitate the learning of discriminating features. The decision optimiser is specified for selecting the best facelift plans, maximising profit over time. By deploying the recurrent neural network (RNN) [13] to estimate the market share shifts caused by the facelift, the optimiser compares the possible gains of different plans and then selects the ones with the highest returns.

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 [5, 6, 14], 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) [15] 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.

The rest of this article is organised as follows. Section 2 reviews the related literature. Section 3 introduces our proposed computational framework for automotive exterior facelift. In Section 4, we introduce the datasets we have used, the experimental settings and our analysis of results. Finally, Section 5 concludes the paper.

2 RELATED WORK

Until recently, product aesthetic design had for a long time been assumed to be a field of human endeavour, despite computer science researchers having made various attempts over decades to break into it. The very early attempts were focused on manipulating 2D design shapes, which were represented by coordinate points and the curves between them [16]. Later studies followed similar approaches but augmented such a manipulation with more dimensions to

obtain more detailed designs. For instance, restructuring the product design in the 3D space or creating additional rich colour choices [17, 18]. However, as pointed out in [19], the inclusion of new dimensions amplified the manipulation difficulties, thus lessening the design flexibility. Such a trade-off between generation realism and flexibility has for long been a challenge for subsequent researchers.

This trade-off has now been largely reconciled. Recent deep generative models only need hundreds of latent variables to obtain various highly realistic designs. The generative adversarial network and variational autoencoder (VAE) [20] families are the two most popular deep generative frameworks nowadays; GAN has the advantage of generating more realistic results [21]. A classical GAN frame involves two subnetworks, termed generator and discriminator. The former is used for generating fake images, while the latter is responsible for discriminating between real and fake images. As the generator and discriminator have adversarial roles, intuitively, in the game theory sense, the training procedure can be perceived as a zero-sum game between the generator and discriminator.

A few existing studies have attempted to apply deep generative models to automotive design. In the very first study, the authors of [19] demonstrated that by training the VAE model with adequate data, the trained model is capable of generating diverse car designs. The generator can automatically produce cars in various body types, brands, and viewpoints with no need to carefully select numerous control variables. In their next study [5], Pan et al. proposed a more comprehensive design framework that included both the deep generative models for design generation and deep neural networks for design evaluation. Such a framework is functional in both offering various candidate designs and evaluating them from an aesthetic perspective. However, these existing studies deployed generative models to produce novel designs, in which design freedom constraints have not been considered. As [22] pointed out, in real automotive markets, designers need to achieve a balance between design freedom and brand recognition. Thus, regional design upgrading, known as design facelift, is much needed in automotive aesthetic design. From the technical perspective, this is more challenging due to the feature entanglement problem [23, 24]. That is, the change of a single latent value can cause global changes in the resulting design. Furthermore, a theoretical analysis [23] has indicated that the disentangled representations cannot be resolved through the unsupervised learning approach if no inductive biases are provided.

The newly developed style-based generators, namely StyleGAN [7] and StyleGAN2 [8], have performed outstandingly on the controllability of generation, while easing the entanglement problem. Similar to our application needs, Liu et al. [25] proposed an objective function-based method for spatial modification, which changes regional content by manipulating the "style" values fed to the synthesis network. On the other hand, to resolve the problem of embedding a given image in the trained StyleGAN, both [26] and [8] have proposed methods for projections, which have the inverse purposes of image generation.

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 [27, 28, 29]. 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. Psychologically, the perception of aesthetics (known as aesthetic emotion) involves a series of neural activations for visual processing and emotional arousal. The complicated nature of this issue made the prediction of aesthetic sensation challenging before the rise of deep models in the last decade. These days, it is common to apply convolutional neural networks (CNNs) [30] for aesthetic evaluation tasks. CNNs are well-known for their end-to-end

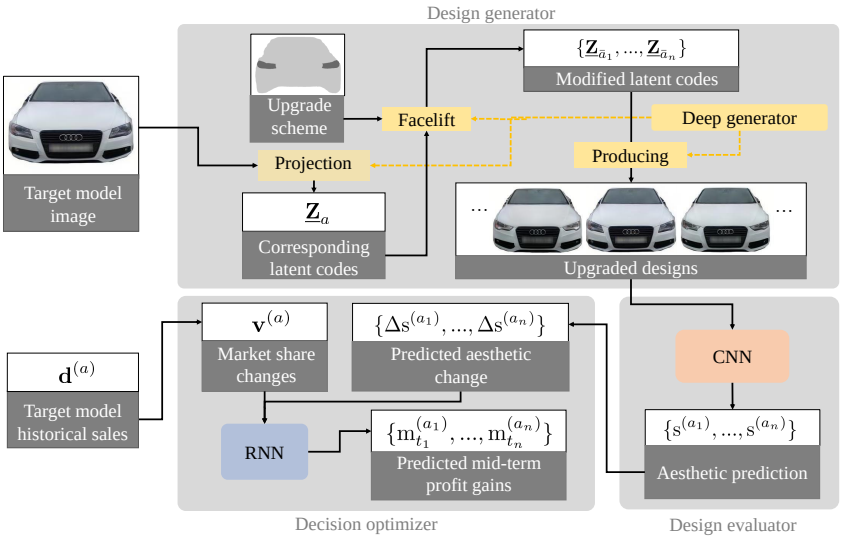


Fig. 2. Schematic view of the GEO framework.

prediction power, which has achieved impressive record-breaking results on several computer vision tasks [31, 32, 33, 34]. Several existing studies [35, 36, 37] attempted to apply the CNN model to the categorisation of fashion products, achieving a remarkable accuracy. On the other hand, several studies relied on CNN-based models to predict the aesthetic level of durable products such as cars [38, 5, 6]. 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, appealing and innovative cars. Inspired by these studies, the development of an automotive aesthetic evaluator becomes relatively straightforward.

Motivated by the aforementioned advances, our present study 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 [5] and [6], our proposed framework differs in two major aspects. First, our 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.

3 THE GEO FRAMEWORK

As illustrated in Fig. 2, the proposed GEO framework consists of a design generator, a design evaluator, and a decision optimiser. The generator proposes different facelift designs; the evaluator assesses the designs from the aesthetic perspective; and the decision optimiser selects the designs that maximise mid-term profit for the automaker. In the following, the technical details of these three components will be introduced in turn.

3.1 The Design Generator

Our generator is based on the StyleGAN2 model [8] as it is the state-of-the-art deep generative model that has performed extremely well in synthesising high-resolution images [39]. 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. 3 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-connected layers, responsible for mapping an input latent code $\mathbf{z}^{(0)}$ to an intermediate code $\mathbf{z}^{(l)}$. The superscript indicates the position of the latent code within the mapping network. For instance, the intermediate latent code $\mathbf{z}^{(l)}$ is the output from the lth layer. The synthesis network, denoted by $G_{\text{ST}}(\cdot)$, 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 [8], 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. 3, 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 $O_k = g_{\text{ST}}\left(O_{k-1}, \underline{Z}_{[k]}\right)$, where O_k represents the output feature maps from the kth generation block, $g_{\text{ST}}(\cdot)$ denotes the computation of the entire generation block, and $\underline{Z}_{[k]}$ represents the latent code fed to the kth generation block. As the right part of the synthesis network in Fig. 3 shows, unlike the original StyleGAN model, StyleGAN2 replaces the progressive growing strategy with "skip connections", then the final output image is the sum of all the generation block results:

$$G_{\text{ST}}(\underline{Z}) = \sum_{k=1}^{K} f_{\text{UP}}\Big(f_{\text{RGB}}(O_k),\ K-k\Big), \tag{1}$$

where $f_{\text{RGB}}(\cdot)$ represents the "To RGB" module that converts feature maps to images, each feature map has the size of $2^{k+1} \times 2^{k+1}$, $f_{\text{UP}}(\cdot)$ 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:

$$G_{\text{ST}}^{-1}(C_a) = \underset{\underline{Z}}{\text{argmin}}\ D_{\text{PER}}\left(G_{\text{ST}}(\underline{Z}), C_a\right), \tag{2}$$

where $D_{\text{PER}}(\cdot)$ is the distance measure between images. It should be noted that for the projection, 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) [40], 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:

$$\mathbb{E}_{C\sim\mathcal{X}_{\text{test}}} D_{\text{PER}}\left[G_{\text{ST}}\left(G_{\text{ST}}^{-1}(C)\right) - C\right], \tag{3}$$

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.

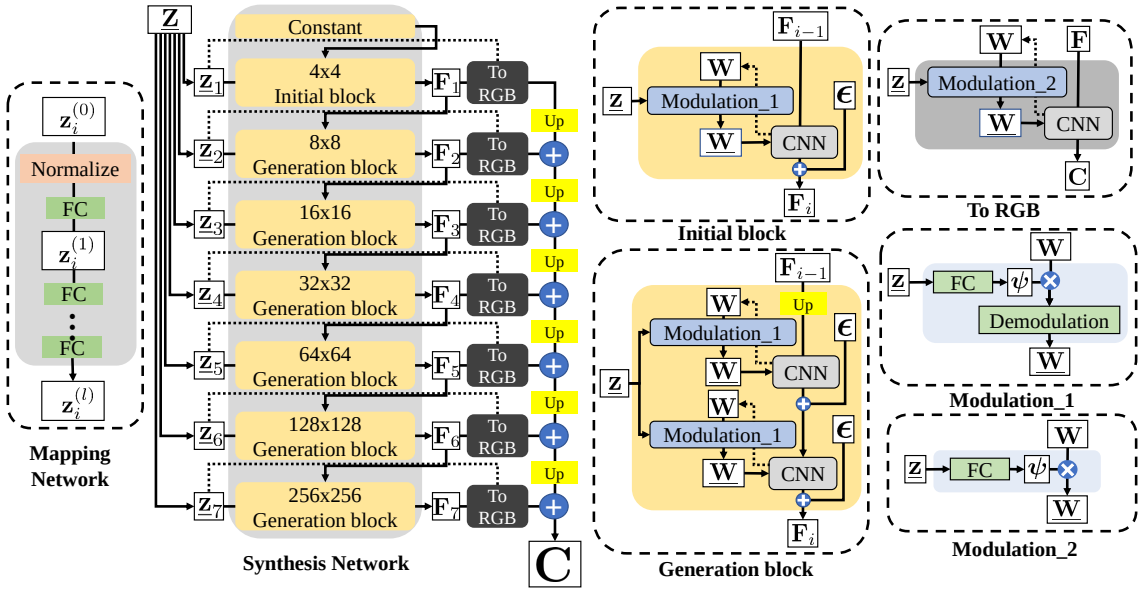


Fig. 3. Structure of the generator in StyleGAN2. The StyleGAN/StyleGAN2 models comprise two main parts, the mapping network and the synthesis network. The former transforms the raw latent code to the intermediate latent code, while the latter generates images/designs according to the given intermediate latent codes.

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) [41] 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:

$$\mathbb{E}_{\mathbf{z}_i,\mathbf{z}_j\sim P_z} f_{\text{FID}}\left(G_{\text{ST}}(\underline{\mathbf{Z}}_{\mathbf{z}_i,\mathbf{z}_j}), \mathcal{X}_{\text{train}}\right), \tag{4}$$

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

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:

$$\max_{\underline{Z}} \left[S_{\text{AE}}\left(G_{\text{ST}}(\underline{Z})\right) - D_{\text{PER}}\left(B_a \odot C_a, B_a \odot G_{\text{ST}}(\underline{Z})\right)\right], \tag{5}$$

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. (5) is not a proper objective function. First, automotive designers would prefer to have more candidate designs rather than a single "best" suggestion [6]. 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. (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 $\{\underline{Z}_1, ..., \underline{Z}_N\}$, we look for a design that satisfies the following objective:

$$\min_{\underline{Z}} \left|\underline{Z}_{[:k]} - \underline{Z}_{i[:k]}\right| + D_{\text{PER}}\Big(B_a \odot C_a, B_a \odot G_{\text{ST}}(\underline{Z})\Big), \tag{6}$$

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 ith inspiration latent matrix. We adopt the L^1 norm since in our trials it achieves more stable results than the Euclidean distance [42]. Previous studies have shown that early generation blocks control general-features and later ones regulate the fine-features [7, 8]. 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. (6), we then rank the designs obtained according to their aesthetic scores and select the best candidates accordingly:

$$\underset{i}{\text{argmax}}\ S_{\text{AE}}\left(G_{\text{ST}}(\underline{Z}_i)\right), \text{ for } \underline{Z}_i \in \mathcal{Z}. \tag{7}$$

3.2 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 [43, 9, 5] and facial recognition/prediction studies [44, 45, 46, 47], 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:

$$\min \mathbb{E}_{a\sim\mathcal{X}_{\text{train}}}\left[\ell_{\text{aes}}^{(a)} + \ell_{\text{rec}}^{(a)}\right], \tag{8}$$

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 $\left(s^{(a)} - \underline{s}^{(a)}\right)^2$, 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 [10, 11, 12] setting, further enhancing the learning of discriminating features in classification. Given the design C_a under car model y_a, and $\mathbf{x}_a$ as its corresponding feature vector extracted from the convolutional backbone network, the jth classification values before softmax can be expressed as $h_{\text{rec}}^{(a,j)} = \mathbf{w}_j^\intercal \mathbf{x}_a + b_j$, where $\mathbf{w}$ and b represent the output weight vector and bias values, respectively. These variables are constrained in the angular loss setting: $||\mathbf{w}_j|| = 1$, $b_j = 0$, and $||\mathbf{x}_a|| = \alpha$, where α is a given constant. This makes each $h_{\text{rec}}^{(a,j)}$ only depends on the angle size between $\mathbf{w}_j$ and $\mathbf{x}_a$, and converts the whole softmax computing into the following formula:

$$\ell_{\text{rec}}^{(a)} = -\log\left\{\frac{\exp\left(\alpha\cos\left(\theta_{y_a,a} + \beta\right)\right)}{\exp\left(\alpha\cos\left(\theta_{y_a,a} + \beta\right)\right) + \sum_{j\neq y_a}\exp\left(\alpha\cos\theta_{j,a}\right)}\right\}. \tag{9}$$

Here we adopt [12]'s setting, which incorporates a constant margin penalty β in the target class's angle, making $h_{\text{rec}}^{(a,y_a)} = \alpha\cos\left(\theta_{y_a,a} + \beta\right)$, 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 [10, 11].

3.3 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 [48]. 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 $\{d_0^{(a)}, ..., d_T^{(a)}\}$, we compute its market share change vector $\mathbf{v}_T^{(a)} := \{v_1^{(a)}, ..., v_T^{(a)}\}$ 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 $f_{\text{MS}}(\mathbf{v}_{t-1}^{(a)}, \Delta s_t^{(a)})$, 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 $\mathbb{E}_{a,t\sim\mathcal{X}_{\text{train}}}\left[v_t^{(a)} - \tilde{v}_t^{(a)}\right]^2$. 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:

$$\tilde{v}_t^{(\hat{a})} = \begin{cases} f_{\text{MS}}\left(\tilde{\mathbf{v}}_{t-1}^{(a)}, \Delta s^{(\hat{a})}\right), & \text{if } t = \eta \\ f_{\text{MS}}\left(\tilde{\mathbf{v}}_{t-1}^{(a)}, 0\right), & \text{otherwise} \end{cases}, \tag{10}$$

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 $\tilde{\mathbf{v}}_{t-1}^{(a)}$, not $\mathbf{v}_{t-1}^{(a)}$, to denote the input vector since here inputs are also simulated results (except for $t = 1$ where $\mathbf{v}_0^{(a)}$ is based on actual records).

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

$$m_\eta^{(\hat{a})} = d_0^{(a)} \cdot \sum_{r=1}^{T} \left[\prod_{t=1}^{r} \tilde{v}_t^{(\hat{a})} - \prod_{t=1}^{r} \tilde{v}_t^{(a)} \right], \tag{11}$$

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

4 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.

4.1 Datasets

Fig. 4 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 styling[1]. 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) focus on modifying the design

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

Audi A4
Saloon
218,847 (19.6%)
"Not a particularly daring design"
www.parkers.co.uk/audi/a4

VW Jetta
Saloon
23,170 (2.1%)
"Dull Image"
www.parkers.co.uk/volks-wagen/jetta

Peugeot 508
Saloon
17,651 (1.6%)
"Exterior design is bland"
www.parkers.co.uk/peuge-ot/508

VW Phaeton
Saloon
1,060 (0.1%)
"Comparatively dull styling"
www.carbuyer.co.uk/volks-wagen/phaeton

Peugeot 2008
SUV
72,298 (2.0%)
"Awkward styling"
www.parkers.co.uk/peuge-ot/2008

SKODA Yeti
SUV
76,749 (2.1%)
"Not to everyone's taste"
www.parkers.co.uk/skoda/yeti

Toyota RAV4
SUV
69,125 (1.9%)
"Not as stylish as rivals"
www.parkers.co.uk/toyota/rav4

Ford EcoSport
SUV
41,438 (1.1%)
"Looks awkward"
www.parkers.co.uk/ford/eco-sport

Vauxhall Astra
Hatchback
674,362 (4.9%)
"Ordinary looks"
www.parkers.co.uk/vauxh-all/astra

BMW 1 Series
Hatchback
312,354 (2.3%)
"Styling not to everyone's taste"
www.parkers.co.uk/bmw/1-series

Honda Jazz
Hatchback
233,144 (1.7%)
"Dull styling"
www.parkers.co.uk/honda/jazz

Hyundai i10
Hatchback
206,492 (1.5%)
"Lacks appeal"
www.parkers.co.uk/hyundai/i10

Fig. 4. 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 forth row shows the critiques of the car model's aesthetic design; the fifth row provides the reference (i.e., URL) of the consumer reviews.

Table 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

of the headlights since they are the most frequently upgraded features in automotive refreshments. As shown in Table 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 [49]. 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.

4.2 Experimental Settings

We use the StyleGAN2 model for the facelift design generation [8]. 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 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) [7] 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 overplay 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) [13] is used.

4.3 Analysis of Results

As shown in Fig. 5, 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 [7]; and (ii)

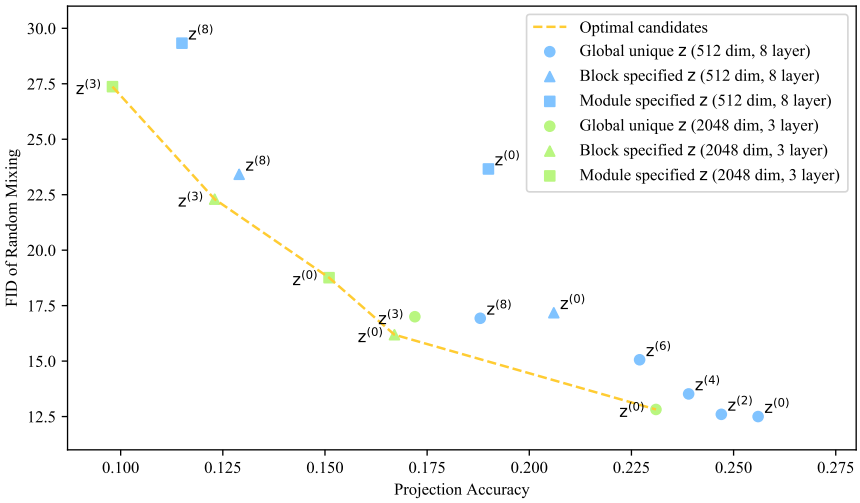


Fig. 5. Comparison of candidate latent spaces, where the superscript indicates the latent layer position among the mapping network.

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 [7, 8], 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., $\mathbf{z}^{(0)}$) would have better FID scores but lower projection accuracy than a later layer (e.g., $\mathbf{z}^{(7)}$). We interpret this trade-off with the concept of design space. The output space corresponding to $\mathbf{z}^{(8)}$ (the intermediate latent codes) can be seen as a design space with a large size. The stacking of additional mapping network layers (such as $\mathbf{z}^{(7)}$, $\mathbf{z}^{(6)}$) 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. 5 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. 5 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

Table 2. Comparison of our design upgrading algorithm with existing methods.

Method	Mapping setting	Projection accuracy	FID of random mixing
Style-based [25]	8×512	0.120	60.60
Our intermediate-based†	8×512	0.190	23.66
Style-based [25]	3×2048	0.119	91.57
Our intermediate-based	3×2048	0.151	18.76

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

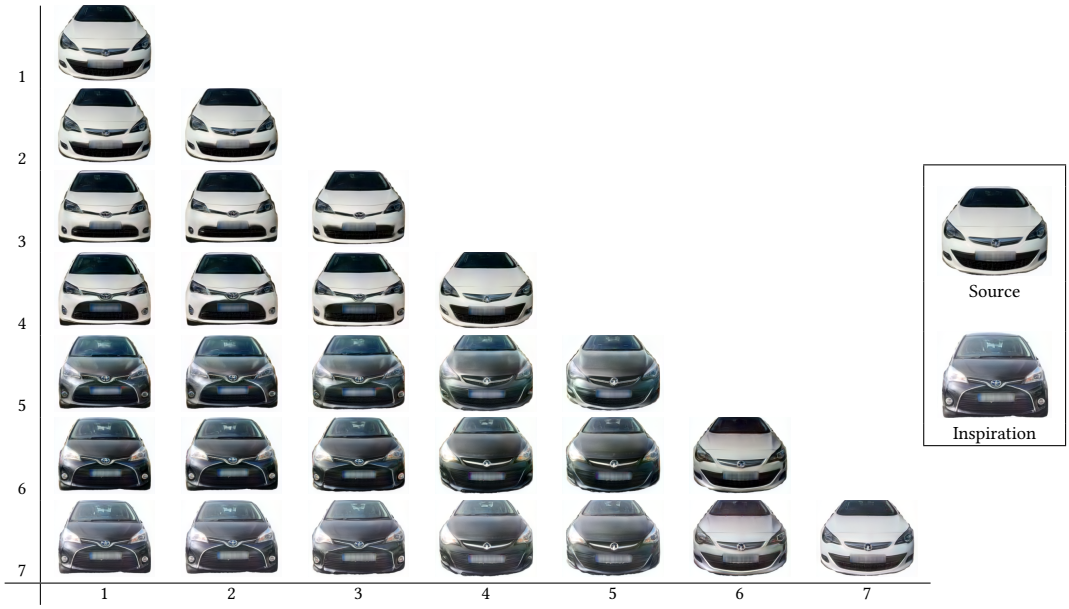


Fig. 6. Comparison of car front designs of replacing different generation block latent codes, where a column is the replacement starting block and a row is the replacement end block.

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.

In Table 2, we further compare our design upgrading algorithm with the method proposed by [25], 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 low 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. 6 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

Fig. 7. Empirical examples of car front aesthetic scores predicted by the trained evaluator.

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. 7 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:

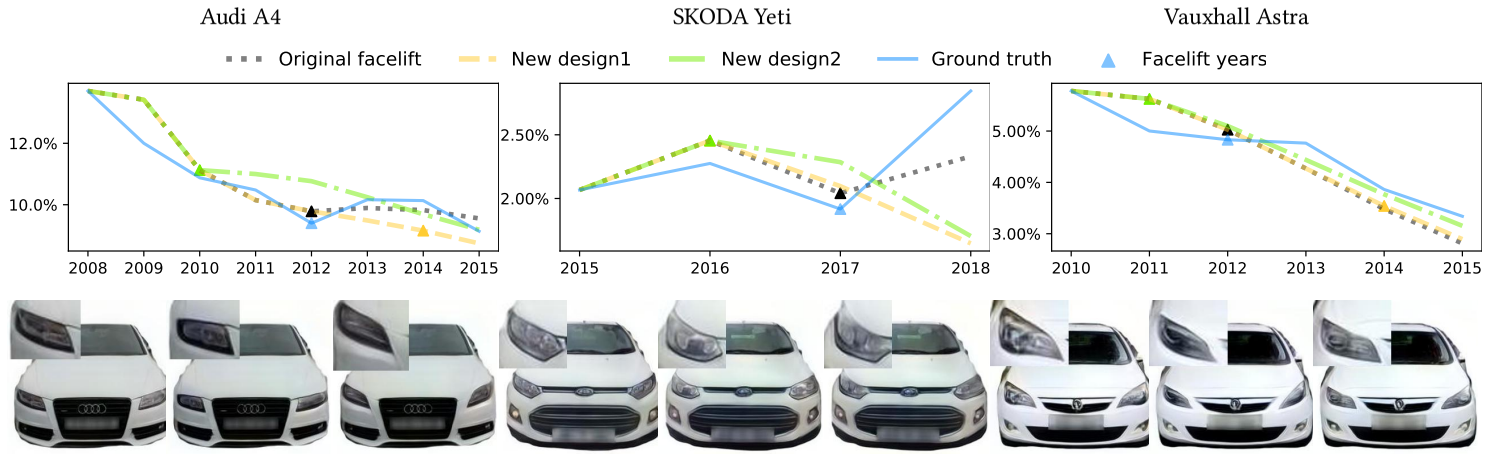
> "...can't disguise its dated design, and the Punto looks bland ...making Fiat's Punto look really rather old..." [2]
>
> "...the ix20 rather blends into the crowd compared to its more stylish rivals..." [3]
>
> "...the BMW M2 is a car that's huge fun from behind the wheel, stupendously quick and relatively 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. 8 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. 8). In particular, the second plot of Fig. 8 shows that the car models' market share 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 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

[2]URL:www.autoexpress.co.uk/fiat/punto/interior
[3]URL:heycar.co.uk/hyundai/ix20
[4]URL:www.autoexpress.co.uk/bmw/2-series/105480/used-bmw-m2-review
[5]URL:www.carbuyer.co.uk/jaguar/xf

Table 3. Examples of optimised designs.



(1.394)	(1.382)	(1.439)	(1.186)	(1.199)	(1.232)	(1.280)	(1.268)	(1.307)

Car type	Before facelift		Real facelift rise		Proposed facelift rise†		Market share rise‡	
	Mean	STD	Mean	STD	Mean	STD	Mean(%)	STD(%)
Sedan	1.393	0.022	-0.079	0.061	0.102	0.044	15.224	20.341
SUV	1.297	0.155	0.113	0.263	0.141	0.036	14.139	14.432
Compact	1.330	0.155	-0.026	0.210	0.126	0.072	36.757	42.389

† For each model, the top 10% facelift designs are selected as the proposed set

‡ The real facelift designs are used as baselines to infer the gained market shares here

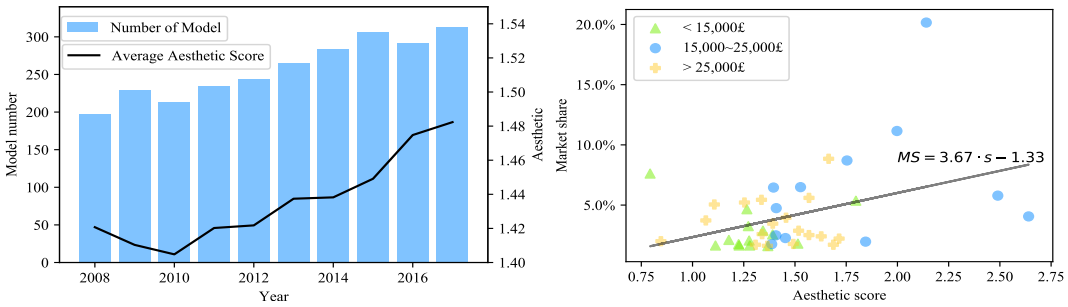


Fig. 8. Analysis of the aesthetic scores: (left) Time series plot of the average aesthetic score and the sales of the targeted car models from 2008 to 2017; (right) Aesthetic score VS market shares. By comparing the model's market shares and their aesthetic scores in recent years, we find that the market share is correlated to their aesthetic levels, where the Pearson test results are as follows: $\rho = 0.39$ with a p-value=0.009.

Table 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 $\frac{\text{Estimated 10 year sales}}{10\times\text{Starting year sales}}$.

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, 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. 9 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. 9 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. 9, 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 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.

5 CONCLUSIONS

We have developed a new machine learning-based framework (i.e., GEO), which can assist automakers when it comes to the cars' aesthetic design. Unlike the existing works, our study focused on the scenario of automotive facelifts,

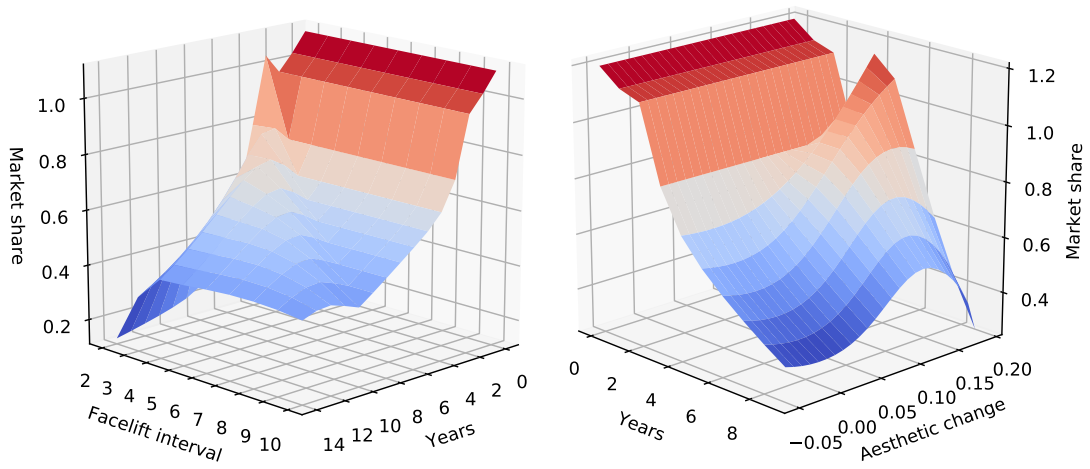


Fig. 9. Effects of facelift: (left) mid-term market share evolution when adopting different facelift years. (right) market share simulation according to aesthetic changes.

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 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 decision 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, our study does naturally have some limitations, which can be addressed in future research work. 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 aesthetic design upgrading can be perceived as a crucial application scenario for the disentanglement problem. Second, the nature of aesthetic perception is pretty complex. Future user studies could investigate how different aesthetic attributes affect the overall aesthetic perception. Third, when computing market changes, we did not take into account the effects of competition from other car models, while the modern automotive market is typically a zero-sum game. It would also be an interesting direction to study aesthetic design competition in different segmented markets.

ACKNOWLEDGMENTS

The first author acknowledges the Adam Smith Business School and the School of Computing Science of University of Glasgow's funding support for this research. The second author would like to thank the funding support of the Region Bourgogne Franche Comté Mobility Grant, Nvidia through its Accelerated Data Science Grant and Google Cloud through its Academic Research Grant.

REFERENCES

[1] Peter H Bloch. 1995. Seeking the ideal form: Product design and consumer response. *Journal of Marketing*, 59, 3, 16–29. DOI: https://doi.org/10.1177/002224299505900302.

[2] Marielle E H Creusen and Jan P L Schoormans. 2005. The different roles of product appearance in consumer choice. *Journal of Product Innovation Management*, 22, 1, 63–81. DOI: https://doi.org/10.1111/j.0737-6782.2005.00103.x.

[3] J D Power. 2016. Initial quality study: Reliability still tops list of purchase factors. (2016). Retrieved 01/25/2018 from Retrieved from http://www.jdpower.com/cars/articles/jd-power-studies/2016-initial-quality-study-reliability-still-tops-list-purchase.

[4] Bruce Blonigen, Christopher Knittel, and Anson Soderbery. 2013. Keeping it fresh: Strategic product redesigns and welfare. Technical report. National Bureau of Economic Research, Cambridge. DOI: https://doi.org/10.3386/w18997.

[5] Yanxin Pan, Alexander Burnap, Jeffrey Hartley, Richard Gonzalez, and Panos Papalambros. 2017. Deep design: Product aesthetics for heterogeneous markets. *International Conference on Knowledge Discovery and Data Mining*, 1961–1970. DOI: https://doi.org/10.1145/3097983.3098176.

[6] Alex Burnap, John R. Hauser, and Artem Timoshenko. 2019. Design and evaluation of product aesthetics: A human-machine hybrid approach. *ArXiv e-prints*, 1907.07786. DOI: https://doi.org/10.2139/ssrn.3421771.

[7] Tero Karras, Samuli Laine, and Timo Aila. 2018. A style-based generator architecture for generative adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 4396–4405. DOI: https://doi.org/10.1109/cvpr.2019.00453.

[8] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of StyleGAN. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8107–8116. DOI: https://doi.org/10.1109/cvpr42600.2020.00813.

[9] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. 2016. Photo aesthetics ranking network with attributes and content adaptation. *European Conference on Computer Vision*, 662–679. DOI: https://doi.org/10.1007/978-3-319-46448-0_40.

[10] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. 2017. SphereFace: Deep hypersphere embedding for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 6738–6746. DOI: https://doi.org/10.1109/cvpr.2017.713.

[11] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. 2018. CosFace: Large margin cosine loss for deep face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 5265–5274. DOI: https://doi.org/10.1109/cvpr.2018.00552.

[12] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. ArcFace: Additive angular margin loss for deep face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 4685–4694. DOI: https://doi.org/10.1109/cvpr.2019.00482.

[13] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Conference on Empirical Methods in Natural Language Processing*, 1724–1734. DOI: https://doi.org/10.3115/v1/d14-1179.

[14] Alex Burnap, Yanxin Pan, Ye Liu, Yi Ren, Honglak Lee, Richard Gonzalez, and Panos Y. Papalambros. 2016. Improving design preference prediction accuracy using feature learning. *Journal of Mechanical Design*, 138, 7, 071404. DOI: https://doi.org/10.1115/1.4033427.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial networks. *Advances in Neural Information Processing Systems*, 2672–2680. DOI: https://doi.org/10.1145/3422622.

[16] Tahira N. Reid, Richard D. Gonzalez, and Panos Y. Papalambros. 2010. Quantification of perceived environmental friendliness for vehicle silhouette design. *Journal of Mechanical Design*, 132, 10, 1–12. DOI: https://doi.org/10.1115/1.4002290.

[17] István Kókai, Jörg Finger, Randall C. Smith, Richard Pawlicki, and Thomas Vetter. 2007. Example-based conceptual styling framework for automotive shapes. *Eurographics Workshop on Sketch-based Interfaces and Modeling*, 37. DOI: https://doi.org/10.1145/1384429.1384440.

[18] Gunay Orbay, Luoting Fu, and Levent Burak Kara. 2015. Deciphering the influence of product shape on consumer judgments through geometric abstraction. *Journal of Mechanical Design*, 137, 8, 81103. DOI: https://doi.org/10.1115/1.4030206.

[19] Alexander Burnap, Ye Liu, Yanxin Pan, Honglak Lee, Richard Gonzalez, and Panos Y. Papalambros. 2016. Estimating and exploring the product form design space using deep generative models. *Design Automation Conference*, 1–13. DOI: https://doi.org/10.1115/detc2016-60091.

[20] Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. *International Conference on Learning Representations*. DOI: https://doi.org/10.48550/arxiv.1312.6114.

[21] Chunwei Tian, Xuanyu Zhang, Jerry Chun-Wen Lin, Wangmeng Zuo, and Yanning Zhang. 2022. Generative adversarial networks for image super-resolution: A survey. *ArXiv e-prints*, 2204.13620. arXiv: 2204.13620. Retrieved from http://arxiv.org/abs/2204.13620.

[22] Alexander Burnap, Jeffrey Hartley, Yanxin Pan, Richard Gonzalez, and Panos Y. Papalambros. 2016. Balancing design freedom and brand recognition in the evolution of automotive brand styling. *Design Science*, 2, 1–28. DOI: https://doi.org/10.1017/dsj.2016.9.

[23] Francesco Locatello, Stefan Bauer, Mario Lucie, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. 2019. Challenging common assumptions in the unsupervised learning of disentangled representations. *International Conference on Machine Learning*, 7247–7283. DOI: https://doi.org/10.48550/arxiv.1811.12359.

[24] Vincenzo Lomonaco, Lorenzo Pellegrini, Pau Rodriguez, Massimo Caccia, Qi She, Yu Chen, Quentin Jodelet, Ruiping Wang, Zheda Mai, David Vazquez, German I. Parisi, Nikhil Churamani, Marc Pickett, Issam Laradji, and Davide Maltoni. 2022. CVPR 2020 continual learning in computer vision competition: Approaches, results, current challenges and future directions. *Artificial Intelligence*, 303, 103635. DOI: https://doi.org/10.1016/j.artint.2021.103635.

[25] Yunfan Liu, Qi Li, Zhenan Sun, and Tieniu Tan. 2020. Style intervention: How to achieve spatial disentanglement with style-based generators? *ArXiv e-prints*, 2011.09699. DOI: https://doi.org/10.48550/arxiv.2011.09699.

[26] Rameen Abdal, Yipeng Qin, and Peter Wonka. 2019. Image2StyleGAN: How to embed images into the StyleGAN latent space? *IEEE International Conference on Computer Vision*, 4431–4440. DOI: https://doi.org/10.1109/iccv.2019.00453.

[27] Kohli Rajeev and Ramesh Krishnamurti. 1987. A heuristic approach to product design. *Management Science*, 33, 12, 1523–1533. DOI: https://doi.org/10.1287/mnsc.33.12.1523.

[28] Anil Kaul and Vithala R. Rao. 1995. Research for product positioning and design decisions: an integrative review. *International Journal of Research in Marketing*, 12, 4, 293–320. DOI: https://doi.org/10.1016/0167-8116(94)00018-2.

[29] Leyuan Shi, Sigurdur Ólafsson, and Qun Chen. 2001. An optimization framework for product design. *Management Science*, 47, 12, 1681–1692. DOI: https://doi.org/10.1287/mnsc.47.12.1681.10243.

[30] Y. Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 11, 2278–2324. DOI: https://doi.org/10.1109/5.726791.

[31] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 1–14. DOI: https://doi.org/10.48550/arxiv.1409.1556.

[32] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*, 1–9. DOI: https://doi.org/10.1109/cvpr.2015.7298594.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. DOI: https://doi.org/10.1109/cvpr.2016.90.

[34] Chunwei Tian, Menghua Zheng, Wangmeng Zuo, Bob Zhang, Yanning Zhang, and David Zhang. 2022. Multi-stage image denoising with the wavelet transform. *Pattern Recognition*, 109050. DOI: https://doi.org/10.1016/j.patcog.2022.109050.

[35] Jonas Teuwen and Nikita Moriakov. 2020. Convolutional neural networks. In *Handbook of Medical Image Computing and Computer Assisted Intervention*. Kevin Zhou, Daniel Rueckert, and Gabor Fichtinger, editors. Academic Press, Amsterdam, 481–501. DOI: https://doi.org/10.1016/b978-0-12-816176-0.00025-9.

[36] John Paul A. Madulid and Paula E. Mayol. 2019. Clothing classification using the convolutional neural network inception model. *International Conference on Information Science and Systems*, 3–7. DOI: https://doi.org/10.1145/3322645.3322646.

[37] Yian Seo and Kyung shik Shin. 2019. Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications*, 116, 328–339. DOI: https://doi.org/10.1016/j.eswa.2018.09.022.

[38] Yanxin Pan, Alex Burnap, Ye Liu, Honglak Lee, Richard Gonzalez, and Panos Papalambros. 2016. A quantitative model for identifying regions of design visual attraction and application to automobile styling. *Internation Design Conference*.

[39] Chunwei Tian, Yixuan Yuan, Shichao Zhang, Chia Wen Lin, Wangmeng Zuo, and David Zhang. 2022. Image super-resolution with an enhanced group convolutional neural network. *Neural Networks*, 153, 373–385. DOI: https://doi.org/10.1016/j.neunet.2022.06.009.

[40] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. *IEEE Conference on Computer Vision and Pattern Recognition*, 586–595. DOI: https://doi.org/10.1109/cvpr.2018.00068.

[41] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in Neural Information Processing Systems*, 6627–6638. DOI: https://doi.org/10.5555/3295222.3295408.

[42] Phillip Isola, Jun Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-to-image translation with conditional adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 5967–5976. DOI: https://doi.org/10.1109/cvpr.2017.632.

[43] S. Chopra, R. Hadsell, and Y. LeCun. 1997. Learning a similarity metric discriminatively, with application to face verification. *IEEE Conference on Computer Vision and Pattern Recognition*, 539–546. DOI: https://doi.org/10.1109/cvpr.2005.202.

[44] Yael Eisenthal, Gideon Dror, and Eytan Ruppin. 2006. Facial attractiveness: Beauty and the machine. *Neural Computation*, 18, 1, 119–142. DOI: https://doi.org/10.1162/089976606774841602.

[45] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski. 2008. Data-driven enhancement of facial attractiveness. *ACM Transactions on Graphics*, 27, 3. DOI: https://doi.org/10.1145/1360612.1360637.

[46] Douglas Gray, Kai Yu, Wei Xu, and Yihong Gong. 2010. Predicting facial beauty without landmarks. *European Conference on Computer Vision*, 434–447. DOI: https://doi.org/10.1007/978-3-642-15567-3_32.

[47] Lingyu Liang, Luojun Lin, Lianwen Jin, Duorui Xie, and Mengru Li. 2018. SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. *International Conference on Pattern Recognition*, 1598–1603. DOI: https://doi.org/10.1109/icpr.2018.8546038.

[48] María José Moral and Jordi Jaumandreu. 2007. Automobile demand, model cycle and age effects. *Spanish Economic Review*, 9, 3, 193–218. DOI: https://doi.org/10.1007/s10108-006-9014-y.

[49] Jingming Huang, Bowei Chen, Lan Luo, Shigang Yue, and Iadh Ounis. 2021. DVM-CAR: A large-scale automotive dataset for visual marketing research and applications. *ArXiv e-prints*, 2109.00881. DOI: https://doi.org/10.48550/arxiv.2109.00881.