Deep Design: Product Aesthetics for Heterogeneous Markets

Yanxin Pan
University of Michigan
yanxinp@umich.edu

Alexander Burnap
University of Michigan
aburnap@aburnap.edu

Jeffrey Hartley
General Motors Corporation
jeff.l.hartley@gm.com

Richard Gonzalez
University of Michigan
gonzo@umich.edu

Panos Y. Papalambros
University of Michigan
pyp@umich.edu

ABSTRACT

Aesthetic appeal is a primary driver of customer consideration for products such as automobiles. Product designers must accordingly convey design attributes (e.g., 'Sportiness'), a challenging proposition given the subjective nature of aesthetics and heterogeneous market segments with potentially different aesthetic preferences. We introduce a scalable deep learning approach that predicts how customers across different market segments perceive aesthetic designs and provides a visualization that can aid in product design. We tested this approach using a large-scale product design and crowdsourced customer data set with a Siamese neural network architecture containing a pair of conditional generative adversarial networks. The results show that the model predicts aesthetic design attributes of customers in heterogeneous market segments and provides a visualization of these aesthetic perceptions. This suggests that the proposed deep learning approach provides a scalable method for understanding customer aesthetic perceptions.

KEYWORDS

Product Aesthetics; Heterogeneous Markets; Deep Learning; Crowdsourcing; Automobile Design

ACM Reference format:
Yanxin Pan, Alexander Burnap, Jeffrey Hartley, Richard Gonzalez, and Panos Y. Papalambros. 2017. Deep Design: Product Aesthetics for Heterogeneous Markets. In *Proceedings of KDD '17, Halifax, NS, Canada, August 13-17, 2017,* 10 pages.
DOI: 10.1145/3097983.3098176

1 INTRODUCTION

Aesthetic appeal is of critical importance for product design, as it not only attracts customer attention, but assists in conveying design attributes (e.g., 'luxurious,' 'sporty,' 'well-proportioned') that are meaningful to the customer [1, 8]. Conveying these aesthetic attributes is particularly important for the automotive industry, as underpinned by the most respected industry assessments (e.g., J.D. Power Initial Quality Study [42], J.D. Power APEAL Study [9]) and internal confidential studies at General Motors.

Specifically, exterior styling is always in the top two or three reasons for purchase, year after year. It is also a prominent reason for not considering a vehicle for purchase or for rejecting it as a finalist. This pattern has been found not just in developed markets, such as the US, but in emerging markets such as India [19]. Understanding these aesthetic preferences remains an important and ongoing challenge for product designers.

The major challenge behind this understanding is the "heterogeneity" of diverse customers across various markets and the inherent subjectivity of their aesthetic perceptions. This challenge is especially vital for customer-centered product designs such as automobiles, as these designs require product differentiation across market segments. This challenge is further exacerbated given the globalized nature of modern automobile design, with customers often geographically and culturally distant from the designers. Accordingly, product designers use a variety of qualitative and quantitative methods to assess aesthetic preferences across market segments, with examples including design theme clinics, focus groups, customer surveys, design reviews, and Kansei engineering [20]. The primary goal of these methods is to understand the reasons "why" the customer perceives a design concept as being aesthetically appealing or unappealing. Ideally, a designer could identify specific regions of the physical product design that contribute to the customer's perception of design attributes. Identification of these regions, called "salient design regions" [10], can provide valuable insight during the design process.

While these methods may capture in-depth customer rationale for aesthetic perceptions, they have two main drawbacks. First, customers often cannot articulate accurately why they like or dislike a design [22, 40]. Second, they are not scalable due to being labor and resource intensive, particularly as multinational enterprises often deal with hundreds or thousands of heterogeneous markets.

In this work, we aim to understand perceptions of aesthetic design attributes across customers from heterogeneous markets, and to do this at the scale consistent with a global company. Specifically, we aim to answer three fundamental questions in the context of product design:

(1) Does the product design achieve the desired aesthetic design attributes for a given market segment?
(2) What are the product's salient design regions for a given design attribute?
(3) How do salient design regions differ across different market segments?

KDD '17, August 13-17, 2017, Halifax, NS, Canada

978-1-4503-4887-4/17/08. . . $15.00
DOI: 10.1145/3097983.3098176

We propose a deep learning approach for prediction of aesthetic design attributes for a given customer, that allows interpretation of the reasons "why" a design is perceived as appealing or unappealing. This approach uses a deep learning architecture that captures the heterogeneity of customer perceptions across aesthetic design attributes. Moreover, this approach has the capacity to analyze large-scale data, such that customer heterogeneity may be accurately modeled across market segments. Importantly, this approach enables visual interpretation of results by identifying regions of the product design that are relevant for a given design attribute.

We conduct an study to test this deep learning approach using 179,000 2D images of vehicles in the last decade, 3,302 customer profiles as well as 33,020 data points of customer perceptions of aesthetic design attributes crowdsourced using an online web application. Our results show that we are indeed able to predict diverse customer perceptions over design attributes, as well as visually interpret the reasons underlying customer perceptions.

The main contribution of this research is providing an approach to interpreting aesthetic design appeal for design concepts across heterogeneous markets. This approach is scalable to hundreds or thousands of markets, an important consideration for multinational enterprises engaged in product design. Methodological contributions include a novel deep Siamese neural network architecture using conditional generative adversarial networks, trained using multimodal data including 2D images, numerical labels, and large-scale crowdsourced aesthetic response data.

The rest of this paper is structured as follows: Section 2 discusses previous efforts that quantitatively analyze product aesthetics. Section 3 introduces the research approach as well as the deep learning model and its interpretation algorithm. Section 4 details the experimental setup, describes the data sets and presents results showing aesthetic perceptions across market segments. Section 5 discusses how this work contributes to the product design, as well as its limitations and future directions. Section 6 provides an overall conclusion.

2 RELATED WORK

We review related work from the engineering and product design communities, as well as recent advances in deep learning for aesthetic styling.

2.1 Product Design Aesthetics

The engineering and product design communities have studied factors affecting the styling attributes of designs using both experimental approaches and modeling approaches.

Experimental approaches to understand implicit customer perceptions of aesthetic design attributes often employ eye-tracking. The earliest pioneering work with such eye-tracking dates back to 1935, when an experiment recorded eye-gaze fixation across regions in artistic pictures [5]. Eye-tracking methods have been successfully applied in various domains such as optimizing the layout of product placement in advertisements [11], web page layouts [4, 45], and consumer choice under pressure [37]. For product design specifically, eye-tracking has been applied to design representations [17, 35], relations with vehicle face components [10, 46], and design diagram assessment [39].

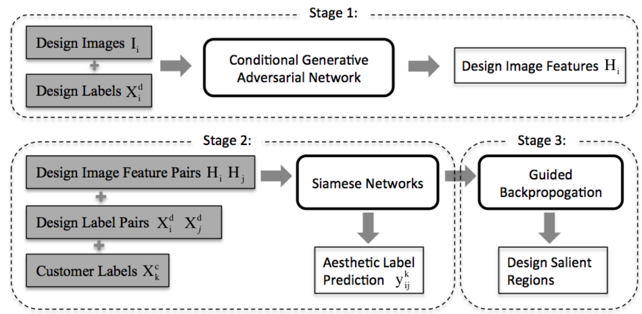

Figure 1: Overview of the proposed deep learning approach for aesthetic design appeal prediction for heterogeneous customers. Grey boxes represent the inputs, white boxes represent outputs, and rounded corner boxes represent the model or algorithm.

Aesthetic modeling approaches have conventionally relied on hand-crafted design representations such as a set of parametric control points to manipulate vehicle silhouettes [25, 27, 28, 34], representations created implicitly using finite shape grammars [18, 30], representations corresponing to a change in intensity [31], or representation manipulated by a set of geometrical handles [2, 47]. In spite of their high interpretability and successful applications, the fidelity of these models are bounded by the realism and flexibility of the design representation [3]. Recently, design research has hybridized modeling approaches with experimental based approaches such as assessments of vehicle face attributes with Kansei engineering and eye-tracking [6]. These aesthetic models has been used to optimize automobile design; examples include 2D vehicle side view silhouettes [25, 34] and 2D vehicle faces [27, 32]. In addition, recent work has extended this research into 3D [36] and virtual reality representations [44].

2.2 Deep Learning for Aesthetics

Deep learning has emerged as a state-of-art approach to model large datasets that include hierarchical data such as 2D images, including recent work in aesthetic styling. These studies frame styling problems as a supervised learning task by using the object's style as labels. Examples include image style recognition [15], comparing illustrative style using stylistic labels [13], and learning high-level judgments of urban perception [21, 24]. Unsupervised learning approaches have also covered cases where stylistic labels are not available, such as style comparison [12], and style transfer to generate paintings in styles of artists from Van Gogh to Picasso [14]. These works validate the possibility of capturing aesthetic-related information using features in deep learning models.

At the same time, however, deep learning has interpretation challenges as these methods create highly nonlinear interactions among the input variables. From a technical standpoint, commonly used optimization objectives such as predictive accuracy or likelihood may no longer provide sufficient criteria for interpretation of deep learning prediction models.

These interpretation challenges have been the focus of recent work in deep learning. A widely-used approach to interpreting a

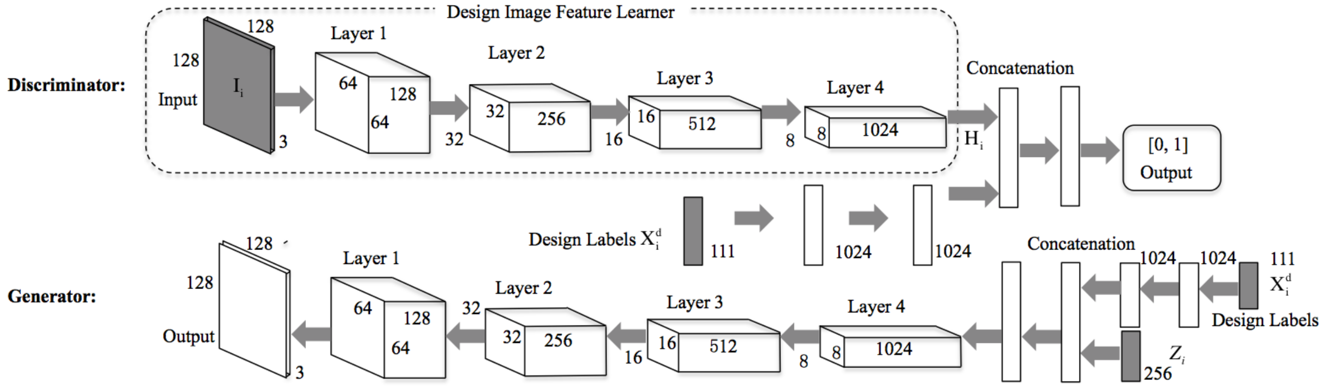

Figure 2: Discriminator and generator in conditional generative adversarial network. Grey boxes represent inputs and white boxes represents convolutional layers in discriminator and upsampling layers.

deep learning model is showing the patches of the 2D image with highest response on feature detectors. Researchers have successfully built a deep learning model to predict safety of urban scenes and interpreted the patterns of safety by visualizing those patches [29]. Unlike urban scenes, small differences in curves or shapes can dramatically change the aesthetic of a product design, thus image patches are too coarse to reveal the salient design regions. A deep convolutional neural network can also be interpreted by computing an approximate inverse mapping of the neural network such as deconvnet, which uses "switch" variables to record the position of pooling operations so that the irreversible pooling operator can be approximately reversed [48]. This method has been used to visualize aesthetic attributes for product design [26], however, this approach is limited because it relies on the selection of neurons to be visualized. Interpretation of deep learning classifiers can also be accomplished by learning an interpretable model locally around the prediction [38]. However, this approach may fail in interpreting product aesthetics because the underlying model is likely highly non-linear even in the locality of the prediction and this may lead to a biased interpretation. The interpretation technique used in our approach, Guided Backpropagation [41], facilitates interpretation by visualizing the salience map. This saliency map captures the salient region in any shape and does not rely on the selection of neurons. Moreover, backpropagation can handle a high degree of nonlinearity.

3 RESEARCH APPROACH

We introduce a research approach that develops a deep learning model to predict and interpret how a customer or market segment perceives a product design concept according to a given aesthetic design attribute (e.g., 'Appealing', 'Sporty'). As shown in Figure 1, the overall research approach consists of three stages:

(1) Stage 1 converts design images to a lower-dimensional feature representation by training a conditional generative adversarial network (cGAN) on the distribution of design images given the design labels (e.g., brand, bodytype).
(2) Stage 2 trains a Siamese network with a pair of cGANs to predict how a customer will perceive a design for a given design attribute, for example, whether a ('Rich,' 'Male') perceives a '2014 Range Rover' as 'Sporty.'
(3) Stage 3 uses guided backpropagation to obtain the saliency map of the Siamese network, then filters the saliency map to discover salient design regions for the given design attribute. This stage allows visual interpretation of predictions of the deep learning model; for example, providing an account for why a ('Rich', 'Male') perceives a '2014 Range Rover' as 'Sporty.'

We next formalize customers, design concepts, and aesthetic design attributes, followed by discussing these three stages of the research approach in detail. Denote the i-th design $\mathbf{D}_i$ as represented by its image $\mathbf{I}_i$ and its design labels $\mathbf{X}_i^d$, i.e. $\mathbf{D}_i = \{\mathbf{I}_i, \mathbf{X}_i^d\}$. Denote the k-th customer $\mathbf{X}_k^c$ as a one-hot encoded vector of customer variables. For each tuple $(\mathbf{D}_i, \mathbf{D}_j, \mathbf{X}_k^c)$, there is a corresponding label y_{ij}^k, with $y_{ij}^k = 1$ having the interpretation that customer k prefers design i over design j for the given design attribute. For example, if the design attribute is 'Sportiness,' then $y_{ij}^k = 1$ corresponds to customer k perceiving design i as more sporty than design j, and $y_{ij}^k = 0$ corresponds to customer k perceiving design i as less sporty than design j.

3.1 Conditional Generative Adversarial Network

The generative adversarial network (GAN) is a generative model consisting of two components, a *discriminator* $\mathcal{D} : \mathbf{I} \longrightarrow [0, 1]$ and a *generator* $\mathcal{G} : \mathbf{Z} \longrightarrow \mathbf{I}$, where $\mathbf{Z} \in \mathbb{R}^{n_z}$ is a noise vector used to seed the generator. The value of $\mathbf{Z}$ is sampled from a noise distribution $p_{\mathbf{z}}(\mathbf{z})$, which is a standard Gaussian distribution in this work. The *discriminator* and *generator* are posed in an adversarial game. The *discriminator* aims to distinguish between real samples from the training data and fake samples generated by the *generator*, while the *generator* aims to generate samples that can not be distinguished by the *discriminator*. This adversarial game is obtained by using a min-max value function as the objective:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \quad \left(\mathbb{E}_{\mathbf{I}\sim p_I(I)}\left[\log \mathcal{D}\left(\mathbf{I}\right)\right] + \mathbb{E}_{\mathbf{Z}\sim p_z(\mathbf{z})}\left[\log\left(1-\mathcal{D}\left(\mathcal{G}\left(\mathbf{Z}\right)\right)\right)\right]\right) \tag{1}$$

Our work extends conventional generative adversarial networks with a conditional architecture, termed a conditional generative adversarial network (cGAN). In this architecture, there is a set of variables $\mathbf{X}^d$ that are believed to be relevant to the image $\mathbf{I}$, and the cGAN aims to capture the relationship between the image and this external information. The *generator* $\mathcal{G}$ and *discriminator* $\mathcal{D}$ in the cGAN model can be redefined as following:

$$\begin{aligned} \mathcal{G}: &\quad (\mathbf{Z}\times\mathbf{X}^d) \rightarrow \mathbf{I} \\ \mathcal{D}: &\quad (\mathbf{I}\times\mathbf{X}^d) \rightarrow [0,1] \end{aligned} \tag{2}$$

The *generator* $\mathcal{G}$ defines a conditional distribution $p_g(\mathbf{I}|\mathbf{X}^d)$, enabling conditioning of the generative model with contextual information $\mathbf{X}^d$. In this work, $\mathbf{X}^d$ are the design labels of an automobile, such as brand, body type, color, and viewpoints. By varying the value of $\mathbf{X}^d$ in the generator, the design labels of the generated sample can be explicitly controlled. More importantly, conditioning on design labels will prompt the model to focus on learning the features describing the appearance of the design instead of the known semantic features of the the design labels. These design labels are detailed in Table 1. In this way, the features extracted from cGAN are more relevant to our later predictive task of capturing aesthetic appeal as will be discussed later.

Similar to the GAN, the generator $\mathcal{G}$ and discriminator $\mathcal{D}$ in cGAN are posed in an adversarial game by a minmax value function:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \quad \left(\mathbb{E}_{(\mathbf{I},\mathbf{X}^d)\sim p_I(I,\mathbf{X}^d)}\left[\log \mathcal{D}\left(\mathbf{I},\mathbf{X}^d\right)\right] + \mathbb{E}_{\mathbf{X}^d\sim p_{\mathbf{X}^d},z\sim p_z(\mathbf{z})}\left[\log\left(1-\mathcal{D}\left(\mathcal{G}\left(\mathbf{Z},\mathbf{X}^d\right),\mathbf{X}^d\right)\right)\right]\right) \tag{3}$$

Conventionally, the discriminator aims to assign a positive label to the training samples $(\mathbf{I}_i, \mathbf{X}_i^d)$, and a negtive label to generated samples $\mathcal{G}(\mathbf{Z}_i, \mathbf{X}_i^d)$, $i = 1, 2, \dots, n$. The objective function of a conventional discriminator is:

$$J_{\mathcal{D}} = -\frac{1}{n}\Big(\sum_{i=1}^{n} \log \mathcal{D}(\mathbf{I}_i, \mathbf{X}_i^d) + \sum_{i=1}^{n} \log(1-\mathcal{D}(\mathcal{G}(\mathbf{Z}_i,\mathbf{X}_i^d),\mathbf{X}_i^d))\Big) \tag{4}$$

We extend this formulation by forcing the discriminator to capture the link between the design images and their design labels by modifying the loss function [33]. Specifically, we penalize the discriminator when it assigns a positive label to an incorrect training sample $(\mathbf{I}_i, \mathbf{X}_r^d)$, where r is randomly drawn from $[1, 2, \dots, n]$ and $r \neq i$.

$$J_{\mathcal{D}} = -\frac{1}{n}\Big(\sum_{i=1}^{n} \log \mathcal{D}(\mathbf{I}_i, \mathbf{X}_i^d) + \frac{1}{2}\Big(\sum_{i=1}^{n} \log(1-\mathcal{D}(\mathbf{I}_i,\mathbf{X}_r^d)) + \sum_{i=1}^{n} \log(1-\mathcal{D}(\mathcal{G}(\mathbf{Z}_i,\mathbf{X}_i^d),\mathbf{X}_i^d))\Big)\Big) \tag{5}$$

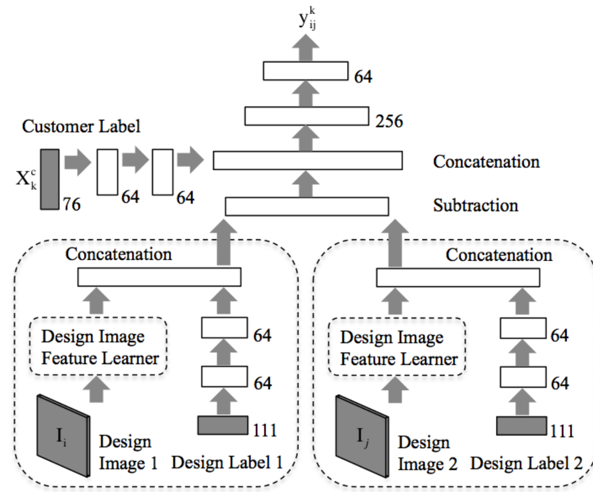

Figure 3: Siamese network of identical conditional generative adversarial networks, with conditioning on design and customer labels. This structure is used to model a customer's aesthetic perception y_{ij}^k for a given design attribute.

Moreover, we maximize the probability assigned by the discriminator to the sample generated by the generator, resulting in the following generator loss function:

$$J_{\mathrm{G}} = -\frac{1}{n}\sum_{i=1}^{n} \log \mathcal{D}\left(\mathcal{G}(\mathbf{Z}_i, \mathbf{X}_i^d), \mathbf{X}_i^d\right) \tag{6}$$

We use deep neural networks for the discriminator and generator. Their architectures are similar to each other as shown in Figures 2. In the discriminator, the input design images and design labels are processed separately by several layers before they are concatenated together. The grey boxes represent the inputs, white boxes represent the output of fully connected layers, and rectangular prisms represent either the output of convolutional layers with filter size 5×5 and ReLu layers in the *discriminator* or deconvolutional layers with filter size 5×5 in the *generator*. The output of the fourth layer in the discriminator (denoted by the dotted box in Figure 2) is then used as the feature representation of the design images.

Though there are simpler models to extract image features, cGAN is used here for several reasons. First, cGAN is a generative model that provides a visual sanity check of whether the cGAN is capturing the distribution of vehicles in the 2D image space, while non-generative models may fall to provide such a visualization. Second, a distinguishing difference between cGAN and other generative models is that cGAN learns the conditional distribution of the vehicle images given the design labels. In this way, the cGAN can focus more on the image features other than design labels whose relationship with product aesthetic preference can be predicted and interpreted using simpler models [20]. Third, using such a generative model allows the generation of new product designs with desired aesthetic attributes.

3.2 Siamese Network

Siamese neural networks are a class of neural network architectures that contain two or more identical subnetworks [7]. These identical subnetworks share the same architecture as well as the same parameters and weights. Siamese neural networks are common for modeling similarity or a relation between two comparable inputs, for example, verifying handwritten signatures. The Siamese structure offers several technical advantages, including requiring fewer parameters to estimate so is less likely to overfit the data.

The structure of the Siamese network used in this work is given in Figure 3. The "design image feature learner," or the feature representation given by the bottom four layers of the discriminator from the cGAN (see in Figure 2), are used as the Siamese network's subnetworks. This 2D image feature representation is then connected with a feature representation of the design labels. After subtracting between concatenated 2D image and design label features, the result is then concatenated with features of customer labels. This concatenated feature vector is then passed through two fully connected layers before the binary prediction task. The objective we used to train the entire model is the cross entropy $\mathbf{J}_s$ of this task.

$$\mathbf{J}_s = -\frac{1}{n}\Big(\sum_{i,j,k} y_{ij}^k \log(\sigma(f(\mathbf{I}_i, \mathbf{I}_j, \mathbf{X}_i^d, \mathbf{X}_j^d, \mathbf{X}_k^c)))+ (1-y_{ij}^k)\log(1-\sigma(f(\mathbf{I}_i, \mathbf{I}_j, \mathbf{X}_i^d, \mathbf{X}_j^d, \mathbf{X}_k^c)))\Big) \tag{7}$$

where $\sigma(\cdot)$ is the sigmoid function, and $f(\cdot)$ represents the Siamese network.

3.3 Guided Backpropagation

Guided backpropagation computes a saliency map for a trained neural network [41]. This saliency map is used to visualize which pixels/regions of an input image are most important for a a neural network's prediction. The key idea behind guided backpropagation is to compute the gradient of the neural network's prediction with respect to the input image with fixed weights. This determines which pixels in the design image are sensitive to the prediction label, or in other words, which pixels can significantly affect the prediction even with small perturbations.

Compared with other visualization methods, guided backpropagation has the ability to produce sharp visualizations of salient image regions. This sharpness is particularly important for our task as shapes and edges of product designs are a major contributor to a customer's aesthetic perception [23]. Accordingly, we use guided backpropagation to visualize the trained Siamese network from Section 3.2. This allows interpretation of which regions of a design most contribute to a customer's perceptual response over aesthetic design attributes.

Guided backpropagation is an extension of conventional backpropagation. The primary difference is how the gradient is backpropagated through "neurons," in which we always assume as linear rectifier units, $y(x) = max(x, 0) = x \cdot [x > 0]$, where $[\cdot]$ is the indicator function. In conventional backpropagation, the gradient of the rectifier's output with respect to its input is defined as follows: $\frac{dy}{dx} y(x) = [x > 0]$. Backpropagation of the error signal δ_i through the rectifier is $\delta_{i-1} = \delta_i \cdot [x > 0]$. Instead, in guided backpropagation, the error signal is $\delta_{i-1} = \delta_i \cdot [x > 0] \cdot [\delta_i > 0]$ when passing through the rectifier. This results in guided backpropagation only passing positive error to positive inputs, such that the error signal is guided not only by the input from the layer below the rectifier, but also by the error signal from layers above the rectifier.

Based on the obtained saliency map, we define salient regions by thresholding on saliency map values. This threshold is a hyperparameter chosen by designers, who have domain expertise in this area. A higher threshold results in salient regions with higher levels of sensitivity, while lower thresholds allow more holistic visualization of salient regions.

Table 1: Design labels

Label Names	Dim.	Label Value
Year	15	2000-2014
Make	48	Land Rover, Nissan, etc.
Model	23	Range Rover Sport, Rogue Select, etc.
Body type	20	SUV, Sedan, etc.
View Point	2	$[\sin\theta, \cos\theta]$, where θ is the angle.
Color	3	RGB

Table 2: Customer labels

Label Names	Dim.	Label Value
Age	1	0 - 99
Gender	3	Female, Male, Prefer not to say.
Income Level	20	$0 - $200,000+
House Region	5	Metropolitan, Suburban, etc.
Family Size	10	0 - 20
Current Car Brand	48	Audi, Cadiliac, BMW, etc.

4 STUDY

We conducted a study to test whether the deep learning research approach introduced in Section 3 can be used to understand aesthetic perceptions of customers in heterogeneous markets. Specifically, we aim to capture customer perceptions of four pairs of aesthetic design attributes: 'Sporty' vs. 'Conservative', 'Luxurious' vs. 'Basic', 'Innovative' vs. 'Traditional', and 'Appealing' vs. 'Unappealing', for the sport utility vehicles (SUV) designed on the U.S. market from 2010 to 2014, followed by visual interpretation of salient regions of these SUVs according to customer perceptions of sportiness.

4.1 Data

Four data sources are used from different modalities: (i) 2D design images, (ii) design labels (e.g., bodytype), (iii) customer labels crowdsourced using an online interactive web application, and (iv) aesthetic perception data for a given customer and set of designs.

4.1.1 Design Data. The full design data set consists of 2D images and design labels corresponding to semantic information about these images. This data set contains 179,702 2D images of vehicle designs on the U.S. market from 2000 to 2014. Each design image has corresponding design labels as listed in Table 1. The full design data set was used to train the conditional generative adversarial network described in subsection 3.1.

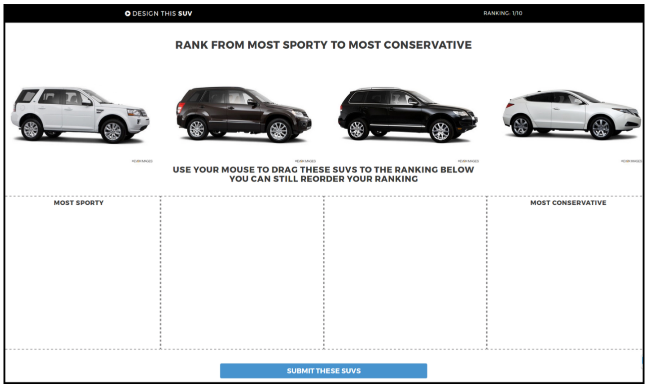

Figure 4: A snapshot of the ranking page in the crowdsourcing web application.

The SUV data set consists of 13,464 2D images and labels of SUV design on the U.S. market from 2010 to 2014, which covers 373 SUV models from 29 brands. This data set was used to collect customer aesthetic perceptions.

4.1.2 Customer Data. A crowdsourcing web application was developed to collect customer aesthetic perceptions for the four pairs of design attributes. Customers first landed on a home page that described the aesthetic perception task. They were then directed to a data collection page as shown in Figure 4, in which they were asked to rank four randomly selected SUVs from the same viewpoint along one randomly selected semantic differential such as 'Sporty' vs. 'Conservative.' The order of this semantic differential was randomly flipped for each customer to counterbalance for ordering biases; however, a single customer always saw the same semantic differential and the same ordering.

Customers were asked to complete 10 rankings, with different viewpoints and SUVs for each ranking. Upon completion of 10 rankings, they were redirected to a survey page, in which they were asked to answer questions about themselves, populating customer labels as listed in Table 2.

A total of 3,302 respondents were collected through General Motors' respondent panels. These respondents had bought an SUV in the past 5 years. Respondents were drawn from several sources and compensated in a variety of ways ranging from no compensation to a Sweepstakes entry to win a $500 gift card. Figure 5 shows the statistics of these customer data. Demographics include 61.6% of the respondents being female. We note that this data has a broad distribution of customers labels, suggesting it contains numerous heterogeneous market segments.

4.2 Procedure

The procedure involved three steps: (1) data preprocessing, (2) training of the cGAN to estimate the feature representation of design images, and (3) training of the Siamese network, containing two cGANs, as well as incorporating customer label and aesthetic perceptions data.

4.2.1 Data Preprocessing. For each attribute (e.g., "Sporty"), respondents who evaluated that attribute were split into training and validation datasets at an 80%/20% ratio. Splitting on the users themselves ensured that the task being assessed was more general than splitting on the evaluations of the respondents. The aesthetic perception data obtained during crowdsourcing was converted from a ranking of four designs to a binary comparison format. This conversion generated a single binary choice pair for each ranking, by taking the first and last design from the ranking. These binary comparisons of design were assigned the label '1' to the pair $[(\mathbf{I}_i, \mathbf{X}_i^d), (\mathbf{I}_j, \mathbf{X}_j^d)]$ if vehicle i was ranked higher than vehicle j, otherwise '0.' Only the "first" rank and "last" rank are used instead of all pairwise binary choices with a ranking of 4 SUVs.

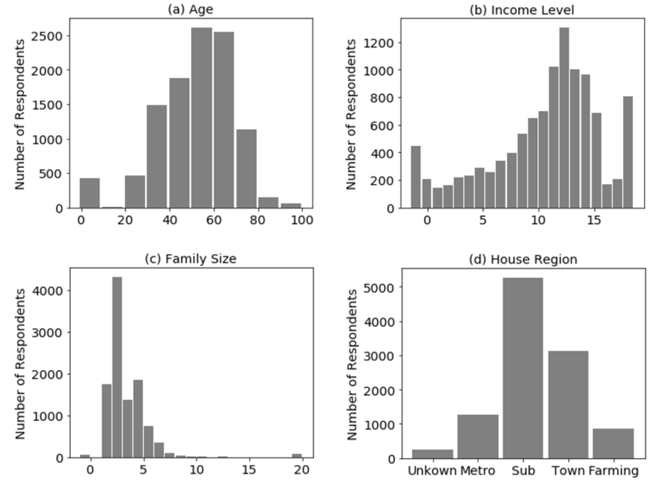

Figure 5: The customer data distribution of (a) Age, (b) Income Level, (c) Family Size, and (d) Housing/Living Region, where "Metro" means "Metropolitan", "Sub" means "Suburban", "Town" means "Small Town", and "Farming" means "Farming Area".

4.2.2 Conditional Adversarial Network Training. Though the study focuses on predicting design attributes such as 'Sportiness' of SUVs for customers in heterogeneous markets, we trained the cGAN using the design data containing 2D images from vehicles of 20 body types (e.g., sedans, trucks). This captures the notion that there are commonalities in aesthetic appearance among all types of vehicles. For example, all vehicles have headlights and wheels. As a result, features learned from all vehicles may better capture the appearance of SUVs as opposed to only training the cGAN with images of SUVs.

The cGAN was trained using the ADAM optimizer [16] to minimize the loss functions of the *discriminator* and *generator*, i.e $\mathbf{J}_{\mathcal{D}}$ and $\mathbf{J}_{\mathcal{G}}$ in Equations (4) and (5). Specifically, we propagate the gradients of the loss function of the *discriminator* $\mathbf{J}_{\mathcal{D}}$ once, then propagate the gradients of the loss function of the *generator* $\mathbf{J}_{\mathcal{G}}$ twice. This sequential training procedure aims at avoiding the discriminator improving too quickly relative to the generator. Moreover, we maintain disentanglement of 2D images and their conditioned labels by training on combinations of real/wrong images with real/wrong labels as described in [33].

Figure 6: Randomly generated vehicle designs from the cGAN generator. These images provide evidence the cGAN is capturing the data distribution of vehicles, particularly with more realism than similar approaches by the authors such as variational autoencoders.

Figure 6 shows randomly generated vehicle images using the cGAN generator. Though not the focus on this work, these images provide a sanity check that the cGAN is capturing the distribution of vehicles in the 2D image space. Note that these images, while plausibly real, do not exist in the training data set. The model required cGAN hyperparameter tuning to achieve aesthetic realism, noting that conventional metrics such as sample loss or pixel-wise distance have been shown to produce images of poor aesthetic quality [43].

4.2.3 Siamese Network Training. We trained the Siamese network by minimizing the negative log likelihood given in Equation (7), using the ADAM optimizer over minibatches of training data [16]. Training was improved by updating only portions of the Siamese network to maintain relative information flow between portions of the cGAN and the randomly initialized portions of the Siamese network. Moreover, we applied batch normalization for every convolutional layer in Figure 3.

4.3 Model Accuracy

The Siamese network achieves different testing accuracies depending on the design attribute as given in Table 3. As a sanity check, a Siamese network with the same architecture as shown in Figure 3, but *without* pretrained 2D image features from the cGAN, achieves lower prediction accuracy on all four design attributes. This suggests the Siamese network architecture is learning how a given customers perceives SUV design attributes.

4.4 Visualization of Aesthetic Saliency

We turn our attention to visualizing the model in order to interpret "why" a customer perceives a SUV across aeshtetic design attributes

	Siamese Net with Image Features, Design labels, and Customer labels	Siamese Net with Design labels and Customer labels
Attribute	Accuracy (Std.Dev)	Accuracy (Std. Dev)
Sporty	75.07 (0.33)	69.17 (0.15)
Appealing	67.29 (0.18)	64.82(0.24)
Innovative	75.44 (0.39)	74.89(0.09)
Luxurious	75.09 (0.13)	74.53 (0.18)

Table 3: Averaged prediction accuracy and its standard deviation on hold-out test data using the Siamese Net with image features, design labels, customer labels or only with the design and customer labels. Average and standard deviation were calculated from 5 random training and testing splits common to each method.

such as 'Sporty.' Moreover, we demonstrate that we are able to perform this visual interpretation for customers in differing market segments.

In particular, we analyze salient regions of a 2014 Land Rover Range Rover Sport for the design attribute: 'Sporty'. From internal research in General Motors, one market segment of the Range Rover Sport is suburban women who opt for a classy SUV. Another market segment is rich men over 40 who want to project proclivities for off-road adventures. By filtering our customer data according to these criteria, we obtain two separate datasets, one for each predefined market segment. Among customers who ranked the 'Sportiness' of the 2014 Range Rover Sport, there were 15 women living in suburban regions with a family size larger than 2. Similarly, there were 12 men with an age greater than 40 and annual income more than $50,000.

Figure 7 shows salient regions for each market segment, corresponding to the 'Sportiness' of the 2014 Range Rover Sport. To obtain these regions, we computed the saliency map of the market segments using guided backpropagation as detailed in Section 3.3, then filtered pixels in the saliency map using a threshold of $[-3\sigma, 3\sigma]$, where σ is the standard deviation of pixel values in the saliency map. In other words, only pixels with an absolute value larger than 3σ are considered salient pixels.

5 CONTRIBUTION TO PRODUCT DESIGN

The high-level goal of this research is to address the three design questions introduced in Section 1, in the context of the proposed deep learning model. These design questions are addressed below using quantitative metrics, as well as qualitative interpretation using input from designers and marketers at General Motors:

(1) Does the product design achieve desired aesthetic design attributes for a given market segment?

As detailed in Section 4.3, the Siamese network was able to predict the design attribute 'Sporty' to 75.07% accuracy, the design attribute 'Appealing' to 67.29%, the design attribute 'Innovative' to 75.44 %, and the design attribute 'Luxurious' to 75.09%, using a hold-out testing dataset. This is evidence that the proposed approach has

Figure 7: Visualization of salient design regions for the 2014 Range Rover Sport. The first row shows salient regions for 'Suburban' 'Women,' while the second row shows salient regions for 'Rich' 'Men' 'Over 40."

utility in helping product designers and executives understand whether given design achieves a desired aesthetic design attribute.

Many design decisions rely on the designer's ability to predict how those choices will affect the perceived design attributes. Along with the many decisions each designer makes in developing the design, these decisions also include executive design reviews and selection of designs. The prediction obtained using our approach not only has relatively high prediction accuracy but also captures the heterogeneity of the market, which can help company decision makers understand how each design will be perceived in the multiple markets which it is aimed. Moreover, our approach allows high capacity and flexibility of testing a large number of new designs within a brief period of time, while traditional market researches (e.g., focus groups, surveys) require much more time and resources and introduce confidentiality issues, all of which our approach avoids.

(2) Where do salient design regions exist on the product for a given design attribute?

As shown in Figure 7, we are able to visualize salient regions of a 2014 Range Rover for the aesthetic design attribute 'Sportiness.' These regions are shown to differ depending on the perceiver's demographics and presumably, viewpoint. Identification of these salient design regions can help designers interpret and better understand which elements of the design are most responsible for the customer's perception. Such information is incredibly important to designers as they relate physical design details to psychological customer reactions.

(3) How do salient design regions differ across different market segments?

As shown in Figure 7, there are some commonalities between the salient regions for suburban women and rich men over 40. For example, the design of the lower front face (shown in the front view) and the side mirrors (shown in the front and rear view) are common salient regions for both market segments. There are also common regions which are not salient regions for both market segments such as the lower part of the side doors (shown in the side view).

There are also interesting differences between the two market segments. In general, the salient regions of suburban women cover a larger proportion of the design than those of rich men over 40. This indicates either that suburban women are more sensitive to design appearance details than rich men over 40, or that they are processing the stimuli more as Gestalts than as individual elements.

These design details include the shape of the back of the car, as shown in the images in the second column (from the 30 degree isometric viewpoint). Also, in line with lifestyle differences, these suburban women with a family seem to be more attentive to rear seat headroom (see the third and fourth columns of Figure 7); these customers may be more likely to have rear seat passengers and may be assessing this functionality as part of their overall assessment. Salient design regions help the designer learn the general relationships between his or her design actions and the perceptual results, which adds to the long term skill development of the design community. When coupled with other marketing analysis and cognitive study (e.g., eye tracking), our approach will enable deeper insights into the market segments and what differentiates their aesthetic reactions.

5.1 Limitations and Future Work

The aesthetic design processes at global enterprises use a number of approaches to understand perceptions of design attributes for heterogeneous markets, with approaches related to the current work including design theme studies and focus groups. In these approaches, customers from various market segments around the world assess baseline and concept designs on design attributes using in-person design stimuli, with follow-up discussion in focus groups.

While these approaches are able to gather rich customer response data, they are not scalable to hundreds or thousands of distinct market segments across the world. This offers promising opportunity for the proposed research as a complement to existing aesthetic design approaches at multinational product design companies. Advancing this deep learning approach into practice, however, requires overcoming a number of limitations.

First, in contrast to many machine learning tasks focused on increasing prediction accuracy, such as optimal ad placement for advertising companies such as Google, understanding the underlying factors affecting heterogeneous use perceptions are most important for this work. For example, how the ordering of perceptual stimuli affect the construction of customer preferences, which may suggest the layout of such information presenting design options to customers. This provides an opportunity for machine learning algorithms such as the one used here to inform design process.

This work may thus provide a test bed for design-specific and psychological questions, such as differences between binary choice and partial ranking tasks under various mediums. This may be particularly relevant given the ongoing shift to internet-based information seeking of customers. Important in this direction are consistency metrics that measure the difference between the salient regions of known similar customers. Such metrics have proven challenging, due to the mismatch between common quantitative metrics for model evaluation and the realism of designs encoded using generative models [43].

Second, the prediction accuracy of the current work must be increased before designers may have full confidence in predicted answers to aesthetic design questions. Increasing this prediction accuracy may take a number of directions. Collecting more customer data may significantly improve accuracy, both by simply having more data, but perhaps more importantly, by having more customer variables. One can imagine customer variables, such as 'hobbies' and 'environmental consciousness,' may provide a much richer representation of customers with regards to their aesthetic preferences.

Architectural changes to this deep learning approach, beyond the Siamese network, may provide additional opportunities for improved accuracy. While details are not reported, a number of similar approaches were attempted before the current architecture was selected. Pretrained neural networks did not result in useful image features, and in fact, reduced aesthetic prediction accuracy below baseline levels due to the increase in parameters. Similarly, generative approaches such as the use of variational autoencoders did not improve prediction accuracies. The authors suggest this is likely due to the low realism of the current state-of-the-art of generative modeling. In this direction, recent results in stacking multiscale generative adversarial networks has shown impressive capture of the underlying data distribution [49]. Moreover, changing the prediction task itself to better capture the human perception process will likely improve accuracy; for example, changing to a ranking output task.

Third, validation of the proposed deep learning approach requires additional study. High prediction accuracy does not necessarily lead to valid answers to design questions [43]. For example, learned feature representations may lead to highly distributed encodings that are efficient for separation of data in the feature space rather than localized encodings that more representative of human perceptions over design. A possible direction to validate our approach is to cross-reference findings from design theme clinics and focus groups, or use experiment-based methods such as eye-tracking [17, 35]. The generalizability of the proposed approach can be validated by the studies on other products besides vehicles.

There are many interesting future directions. For example, the generative model used in our approach provides a possibility of using the deep learning model to generate new designs with the desired aesthetic attributes.

6 CONCLUSION

Aesthetic appeal is of critical importance to customer-centric product designs such as automobiles. This creates an ongoing challenge for designers that aim to understand the factors influencing a customer's aesthetic perception over design attributes. Exacerbating this challenge is the scale at which such an understanding is undertaken, with global enterprises designing for hundreds or thousands of heterogeneous market segments.

We have introduced a research approach to predict and interpret customer perceptions of design attributes for heterogeneous markets. Specifically, we build on recent advances in deep learning and develop a Siamese neural network containing a pair of conditional generative adversarial networks. This model takes as input 2D design images and associated labels, customer data corresponding to heterogeneous market segments, and the perceptions of these customers across aesthetic design attributes.

A study was conducted to assess the utility of this research approach. A dataset consisting of automotive vehicles from 2000-2014, as well as customer data collected using an online crowdsourcing web application, was used to train the Siamese network. Our results show that this research approach is indeed able to predict design attributes across customers belonging to heterogeneous market segments. Further, we show visual interpretation of customer perceptions of design attributes for various market segments.

While this approach shows that the proposed research approach is viable in the context of scalable understanding of customer perceptions to aesthetic product design, a number of limitations must be overcome before this approach may be advanced to practice. At the same time, many of these limitations may be mitigated by recent advances in other areas of deep learning, as well as complementary approaches already used at multinational design enterprises.

ACKNOWLEDGMENTS

This research has been partially supported by General Motors Corporation, and the National Science Foundation under Grant No. CMMI-1266184. This support is gratefully acknowledged. We also acknowledge General Motors Corporation for valuable insights and access to data used in this study.

REFERENCES

[1] Daniel E. Berlyne. 1971. *Aesthetics and Psychobiology.* Appleton-Century-Crofts, East Norwalk, CT, US. http://doi.apa.org/psycinfo/1973-00821-000

[2] 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 Journal* 2, 9 (2016).

[3] 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. (Aug. 2016), V02AT03A013. DOI: http://dx.doi.org/10.1115/DETC2016-60091

[4] Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. 2009. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI conference on human factors in computing systems.* ACM, 21–30.

[5] Guy Thomas Buswell. 1935. How people look at pictures: a study of the psychology and perception in art. (1935).

[6] Y.-M. Chang, C.-P. Chu, and M.-Y. Ma. 2013. Exploring the Visual Cognitive Features on the Design of Car Based on the Theory of Eye-Tracking Technology. *Przeglfjd Elektrotechniczny* 89, 1b (2013), 143–146.

[7] Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1. IEEE, 539–546.

[8] Del Coates. 2003. *Watches Tell More Than Time: Product Design, Information, and the Quest for Elegance.* McGraw-Hill London.

[9] Joseph Dobrian. 2016. 2016 U.S. APEAL Study: Safety Features Score Big, Boosting New-Vehicle Appeal. http://www.jdpower.com/cars/articles/jd-power-studies/2016-us-apeal-study-results. (2016).

[10] Ping Du and Erin F. MacDonald. 2014. Eye-Tracking Data Predict Importance of Product Features and Saliency of Size Change. *Journal of Mechanical Design* 136, 8 (2014), 081005.

[11] Andrew T. Duchowski. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4 (2002), 455–470.

[12] Takahiko Furuya, Shigeru Kuriyama, and Ryutarou Ohbuchi. 2015. An unsupervised approach for comparing styles of illustrations. In *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on.* IEEE, 1–6.

[13] Elena Garces, Aseem Agarwala, Diego Gutierrez, and Aaron Hertzmann. 2014. A similarity measure for illustration style. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 93.

[14] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).

[15] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. 2013. Recognizing image style. *arXiv preprint arXiv:1311.3715* (2013).

[16] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[17] Brandeis H. Marshall, Shweta Sareen, John A. Springer, and Tahira Reid. 2014. Eye tracking data understanding for product representation studies. In *Proceedings of the 3rd annual conference on Research in information technology.* ACM, 3–8.

[18] Jay P McCormack, Jonathan Cagan, and Craig M Vogel. 2004. Speaking the Buick language: capturing, understanding, and exploring brand identity with shape grammars. *Design Studies* 25, 1 (Jan. 2004), 1–29. DOI: http://dx.doi.org/10.1016/S0142-694X(03)00023-1

[19] General Motors. 2014. *General Motors Internal Research.*

[20] Mitsuo Nagamachi. 1995. Kansei engineering: a new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics* 15, 1 (1995), 3–11.

[21] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and César Hidalgo. 2014. Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* 779–785.

[22] Richard E Nisbett and Timothy D Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological review* 84, 3 (1977), 231.

[23] 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 (2015), 081103.

[24] Vicente Ordonez and Tamara L Berg. 2014. Learning high-level judgments of urban perception. In *European Conference on Computer Vision.* Springer, 494–510.

[25] Seth Orsborn, Jonathan Cagan, and Peter Boatwright. 2009. Quantifying Aesthetic Form Preference in a Utility Function. *Journal of Mechanical Design* 131, 6 (2009), 061001. DOI: http://dx.doi.org/10.1115/1.3116260

[26] Yanxin Pan, Alexander 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. In *Proceedings of the 2016 Internation Design Conference.* Dubrovnik, Croatia.

[27] Jean-Franois Petiot and Antoine Dagher. 2010. Preference-Oriented Form Design: Application to Carsfi Headlights. *International Journal on Interactive Design and Manufacturing* 5, 1 (Oct. 2010), 17–27. DOI: http://dx.doi.org/10.1007/s12008-010-0105-5

[28] Emilie Poirson, Jean-Franois Petiot, Ludivine Boivin, and David Blumenthal. 2013. Eliciting user perceptions using assessment tests based on an interactive genetic algorithm. *Journal of Mechanical Design* 135, 3 (2013), 031004.

[29] Lorenzo Porzi, Samuel Rota Bulò, Bruno Lepri, and Elisa Ricci. 2015. Predicting and understanding urban perception with convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia.* ACM, 139–148.

[30] Michael J. Pugliese and Jonathan Cagan. 2002. Capturing a rebel: modeling the Harley-Davidson brand through a motorcycle shape grammar. *Research in Engineering Design* 13, 3 (2002), 139–156.

[31] Daniele Quercia, Neil Keith O'Hare, and Henriette Cramer. 2014. Aesthetic capital: what makes london look beautiful, quiet, and happy?. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.* ACM, 945–955.

[32] Charlie Ranscombe, Ben Hicks, Glen Mullineux, and Baljinder Singh. 2012. Visually decomposing vehicle images: Exploring the influence of different aesthetic features on consumer perception of brand. *Design Studies* 33, 4 (2012), 319–341.

[33] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative adversarial text to image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*, Vol. 3.

[34] 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 (2010), 101010. DOI: http://dx.doi.org/10.1115/1.4002290

[35] Tahira N. Reid, Erin F. MacDonald, and Ping Du. 2013. Impact of product design representation on customer judgment. *Journal of Mechanical Design* 135, 9 (2013), 091008.

[36] Yi Ren, Alexander Burnap, and Panos Papalambros. 2013. Quantification of perceptual design attributes using a crowd. In *DS 75-6: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 6: Design Information and Knowledge, Seoul, Korea, 19-22.08. 2013.*

[37] Elena Reutskaja, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel. 2011. Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review* 101, 2 (2011), 900–926.

[38] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why Should I Trust You?: Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 1135–1144.

[39] Anne Ruckpaul, Thomas Frstenhfer, and Sven Matthiesen. 2015. Combination of Eye Tracking and Think-Aloud Methods in Engineering Design Research. In *Design Computing and Cognition '14*, John S. Gero and Sean Hanna (Eds.). Springer International Publishing, Cham, 81–97.

[40] David H Silvera, Robert A Josephs, and R Brian Giesler. 2002. Bigger is better: The influence of physical size on aesthetic preference judgments. *Journal of Behavioral Decision Making* 15, 3 (2002), 189–202.

[41] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2014. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* (2014).

[42] John Tews. 2016. J.D. Power Initial quality Study 2016. http://www.jdpower.com/press-releases/2016-us-initial-quality-study-iqs. (2016).

[43] Lucas Theis, Aäron van den Oord, and Matthias Bethge. 2015. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844* (2015).

[44] Noah Tovares, Peter Boatwright, and Jonathan Cagan. 2014. Experiential Conjoint Analysis: An Experience-Based Method for Eliciting, Capturing, and Modeling Consumer Preference. *Journal of Mechanical Design* 136, 10 (2014), 101404.

[45] Qiuzhen Wang, Sa Yang, Manlu Liu, Zike Cao, and Qingguo Ma. 2014. An eye-tracking study of website complexity from cognitive load perspective. *Decision support systems* 62 (2014), 1–10.

[46] Sonja Windhager, Florian Hutzler, Claus-Christian Carbon, Elisabeth Oberzaucher, Katrin Schaefer, Truls Thorstensen, Helmut Leder, and Karl Grammer. 2010. Laying eyes on headlights: Eye movements suggest facial features in cars. *Collegium antropologicum* 34, 3 (2010), 1075–1080.

[47] M. E. Yumer, Siddhartha Chaudhuri, J. K. Hodgins, and Levent Burak Kara. 2015. Semantic Shape Editing Using Deformation Handles. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2015)* 34, 4 (2015).

[48] Matthew D Zeiler and Rob Fergus. 2014. Visualizing and understanding convolutional networks. In *European conference on computer vision.* Springer, 818–833.

[49] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, and Dimitris Metaxas. 2016. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. *arXiv preprint arXiv:1612.03242* (2016).