Economic Value of Visual Product Characteristics

Ankit Sisodia, Vineet Kumar, Alex Burnap
Yale School of Management, ankit.sisodia@yale.edu, vineet.kumar@yale.edu, alex.burnap@yale.edu

June 2022

Demand models typically use structured data for estimating the value of product characteristics. However, for several product categories such as automobiles, consumers emphasize that visual characteristics of the product are significant demand drivers. Since visual characteristics are typically in high-dimensional unstructured data (e.g., product images), this poses a challenge to incorporate them in demand models. We introduce a method that enables estimation of demand using visual characteristics, by building on the BLP demand model with recent advances in disentangled representation learning Our method also overcomes the challenge of not having supervised signals, which are required for good disentanglement, by using the demand model as supervisory signal. We discover independent and human interpretable visual characteristics directly from product image data, while simultaneously estimating equilibrium demand in a competitive automobile market in the UK. We conduct a counterfactual analysis using a recent dramatic change in the visual design language of BMW cars, and show our predicted results align with actual changes in BMW market share. To our best knowledge, this work is the first to link visual product characteristics with demand–in other words, to quantify the economic value of design.

Key words: visual analytics, deep learning, demand models

> Exterior look/design is the top reason shoppers avoid a particular vehicle (30%), followed by cost (17%).
>
> *–JD Power Avoider Study 2015*

1. Introduction

Products have long been represented as bundles of characteristics (Lancaster 1966), with both functional and form characteristics serving as value drivers (Rosen 1974). Visual characteristics comprising product form are often a primary factor in a product's market success (Jindal et al. 2016, Veryzer Jr 1993). Visual characteristics are designed by firms for everything from communicating intended product differentiation and segmentation (Bloch 1995, Homburg et al. 2015), signaling brand equity (Aaker 1997), and of course, to make appealing products that consumers choose (Creusen and Schoormans 2005, Norman 2013).

Despite their importance, however, visual characteristics have received relatively little attention compared to structured product characteristics (e.g., size or brand) in widely-adopted quantitative models of consumer demand. This has arguably been due to two

related factors. First, quantitatively representing product form is inherently challenging, as even low-fidelity product form representations (e.g., even wireframes that capture outer structure) may require hundreds, if not thousands, of highly interdependent variables.[1] This high dimensionality of images poses significant modeling challenges for conventional demand models that typically operate with 10's of structured product characteristics. Second, even if high-dimensional product form representations were tractable, a lack of methods that can causally link visual product characteristics with economic primitives such as willingness-to-pay or consumer demand.

Our research aims to estimate the economic value of visual product characteristics. We aim towards overcoming the two aforementioned challenges by integrating two methodologies that have typically been distinct in the fields of quantitative (econometric) models and deep learning. This leads to the following set of research questions that are both substantive and methodological.

From a substantive perspective, we seek to understand whether any of the multiple visual characteristics impact choices and market outcomes, and if so, to what degree. It is also useful to understand how consumers trade off between structured and visual characteristics. Second, we aim to examine whether products that are "close" in the space of structured characteristics also close in the space of visual characteristics. Third, we aim to understand how substitution occurs with products close in structured characteristics compared to the space of visual characteristics.

From a methodological perspective, we seek to understand whether demand models can improve the deep learning component of our integrated model. Namely, we ask whether a demand model describing consumer choices can be used as a supervisory signal in the absence of "ground truth" to enable better discovery of such visual characteristics. Similarly, in the opposite direction, we seek to understand whether the deep learning component of this integrated model can improve the properties of the demand model; namely, whether preference parameters obtained by incorporating visual characteristics differ from a baseline that includes only structured characteristics.

Our overall model itself may be viewed as an integrated model comprised of two component models, an econometric demand model and disentanglement-based machine learning

[1] In many industries such as furniture and automobiles, industry-standard product form representations are images, scaling this challenge to millions of interdependent variables. For example, even a 1000 pixel $\times$ 1000 pixel black and image is 1 million variables, each having a value of 0 or 1.

model. At a high level, the disentanglement-based machine learning model identifies visual characteristics from images of products, while the demand model identifies how those visual characteristics (alongside other more conventional structured product characteristics) affect a demand system comprising a competitive market of firms and heterogeneous consumers.

For the demand model component of our overall model, we build upon the well-regarded and standard demand model of Berry et al. (1995) (BLP) to allow consumers to have preferences over visual product characteristics. BLP has been popular across a spectrum of applications in marketing and economics for a number of reasons; among them are that the approach can be used with market-level data on products and outcomes, it accommodates unobservable product characteristics, allows for observable and unobservable consumer heterogeneity, and importantly accounts for endogeneity of prices (which can depend on the unobservable product characteristics).

For the machine learning component of our overall model, we need a method that can obtain visual characteristics from product image data. This task is challenging due to the aforementioned high dimensionality of images. However, to integrate visual elements with the demand model, it is necessary to obtain a lower-dimensional representation. One may first consider obtaining these lower-dimensional factors using classical linear approaches like PCA, or similarly, classical nonlinear approaches like autoencoders (Bengio et al. 2013). However, the resulting factors typically have no semantic meaning to humans (i.e., are not interpretable). Alternatively, we can pre-specify select characteristics, like in Zhang et al. (2021). However, we might not be able to identify what aspects of the images directly correspond to any given characteristic, and such an approach also cannot generate counterfactual visual characteristics.

We instead leverage recent developments in an area of deep learning called "disentangled representation learning," which aims at identifying the "true" underlying visual characteristics from high-dimensional images in a manner that is semantically interpretable to humans. Our deep learning model is built on variational autoencoders (VAE) (Kingma and Welling 2014), and thus comprised of an "encoder" and "decoder." The "encoder" takes an image of a product as input, which undergoes a sequence of highly non-linear transformations to output a lower-dimensional latent code corresponding to the visual characteristics. The "decoder" then takes that latent code as input and outputs a reconstruction of the

original image. This process thus represents the content of a high-dimensional image using a low-dimensional latent code. To promote disentanglement of the learned visual characteristics, we extend the VAE using additional losses to enforce learned visual characteristics to be uniquely represented, statistically independent, and minimally lossy relative to the original product image data. While the machine learning literature has shown this approach does not guarantee our goal of identifying the "true" semantically interpretable visual characteristics in a purely unsupervised manner (i.e., using only product image) (Locatello et al. 2019), one of our contributions is to overcome this issue by integrating the deep learning component of our model by using the demand model as a supervisory signal.

Finally, our integrated demand-and-disentangled-representation-learning model has two distinct specifications, which we term "open loop" and "closed loop." In the open loop specification, the (vector of) visual characteristics obtained are included as additional product characteristics in the utility specification and the demand model is estimated once. In contrast, in the closed loop specification, we jointly use the demand model as a supervisory signal in order to obtain a disentangled representation of visual characteristics. The demand model provides useful supervision, since it allows consumers to respond to visual characteristics (in addition to structured characteristics), and incorporate them in their choice process, which in turn must match market outcomes. Thus, improving the loss of the demand model could potentially benefit the visual disentanglement as well.

We next test our integrated model using an empirical application. Specifically, we evaluate the impact of visual characteristics on consumer preferences and aggregate demand for the automobile market in the UK. We focus on the automobile market because these represent important and substantial purchases for consumers, and it is a product where visual characteristics are likely to have an impact.[2] Also, as we have detailed earlier, automobile buyers report that visual design is an important criterion in purchase.[3] The data used here spans the period 2008-2017 and includes a total of 45 brands and 407 models, ranging in price from £5,761 (Hyundai Amica) to £124,130 (Audi R8). We separately obtain structured data on prices, product characteristics and market outcomes, and unstructured image data corresponding to the "front view" of those automobile models. We estimate

[2] According to Leader Bank (https://www.leaderloancenter.com/the-2nd-most-expensive-thing/), an automobile is typically the most expensive consumer product other than a house

[3] According to JD Power 2015 US Avoider Study (https://www.jdpower.com/business/press-releases/2015-us-avoider-study), exterior design is the top reason shoppers avoid a particular vehicle).

both the open loop and closed loop variants of the model, and do so with and without incorporating visual characteristics into consumer preferences.

Our proposed approach obtain 3 visual characteristics that are human interpretable from the image in both the open and closed loop specifications of our model. These characteristics can be interpreted as body shape, boxiness (presence of angular versus curved features), and grille size. We find that the price elasticities are more negative in the presence of visual characteristics in the model, compared to the baseline BLP estimates. Also, when using the baseline OLS and IV Logit (Berry 1994) models with visual characteristics included, the number of products with inelastic demand is lower. We also conduct a prediction task to obtain predicted market share when entirely new products are introduced, to examine the impact of visual characteristics.

We lastly use our model to simulate a counterfactual corresponding how changing the visual form of a product(s) impacts economic valuation and ultimately demand for the firm's products. While most changes in automoboile visual characteristics are relatively mild "facelifts" and "refreshes", dramatic variations do occasionally come up in which a firm chooses to make dramatic changes. A recent example is BMW's change in "design language" in Model Year 2020, in which BMWs changed their iconic and distinctive "kidney bean" front grille, a mainstay amongst BMWs continual product line of coupes, sedans, and CUV/SUVs. We conduct a counterfactual where BMW introduces such a visual design change earlier, and then evaluate market outcomes after this introduction. This counterfactual enables us to quantify the value of design, and also to identify how the substitution patterns change when the new design is introduced, and compare that to the case where the prior design was continued.

The rest of this paper is structured as follows: Section 2 discusses related literature, Section 3 introduces the proposed integrated economics-and-machine learning model, Section 4 details the empirical application of the approach, Section 5 simulates a counterfactual in visual design, and Section 6 concludes this work with a discussion and opportunities for future work.

2. Literature Review

Our work is related to three broad streams of literature: First, it is related to the BLP class of demand models in economics. Second, it builds on "disentangled" representation

learning in the machine learning literature to find human-interpretable patterns within unstructured data corresponding to the visual characteristics we seek. Third, it is related to the use of unstructured data in empirical modeling within marketing.

2.1. Demand Models

The random coefficient logit model described in Berry et al. (1995) has been extensively used for estimating demand for differentiated products in economics as well as marketing. It addresses the potential endogeneity of price by using an instrumental variable approach similar to Berry (1994). It does so by assuming that a consumer's utility for a good depends not only on its observed product characteristic (available as data to econometrician) but also on unobserved product characteristics (observed by the market participants such as firms and consumers but not observed by the econometrician). The unobserved product characteristics are correlated with price. Besides addressing price endogeneity, this model showed that allowing flexible substitution patterns to reflect the heterogeneity in customer tastes for observed product characteristics is important when estimating demand elasticities. Finally, another key aspect of this model is the ability to work with aggregate data at market level in which a large number of differentiated products are available.

Several papers have extended the BLP class of demand models. Nevo (2000) allowed for interaction of observed heterogeneity with product characteristics in addition to only considering unobserved heterogeneity. It also showed how to include product fixed effects into the BLP class of models. Petrin (2002) and Berry et al. (2004) provide an approach to incorporate micro data along with market data to estimate BLP models. They show that not only does incorporating micro data directly allow us to obtain more richer substitution patterns but also reduce the number of instruments required to solve the problem. Finally, Grieco et al. (2021) allow the average unobserved quality to change across markets. This allows them to study the evolution of market power over time.

We add to this rich literature by allowing for consumer's utility to also depend on visual characteristics of products. These visual characteristics are learned from product images by using deep learning methods. Thus, we make available additional structured data to the econometrician that was earlier present only in the form of unstructured data and hence were unobservable to the econometrician.

2.2. Disentangled Representation Learning

Representation learning broadly refers to using machine learning to re-represent data from its original observed representation into a different (typically lower-dimensional) representation with properties now better suited to a given task (Bengio et al. 2013). For example, the classical model of principle component analysis (PCA) may be viewed under this lens if one wanted to re-represent data into lower-dimensional linear subspace spanned by vectors that best explain the original representations variance; or likewise, the classical model of an autoencoder, which aims at nonlinearly compressing data while minimizing information loss of the new representation. In this work, we aim to re-represent high-dimensional unstructured data (i.e., images of automobiles) into representations that are not just lower-dimensional or uncorrelated, but are also semantically interpretable to humans for use in demand models. Recent advances in an area of deep learning known as "disentangled" representation learning seeks exactly such interpretable representations.

Most recent advances in disentanglement representation learning methods are built using deep generative models that both re-represent the original observed data to a corresponding lower-dimensional representation, as well as enable the generation of new data. Within deep generative models, variational autoencoders (VAEs) (Kingma and Welling 2014) have been most widely adopted for disentangled representation learning. This is due to the VAE being autoencoder, and thus naturally suited for nonlinearly re-representating high-dimensional data, but also due to the VAE being a probabilistic extension of the classical autoencoder that adopts a variational Bayesian formulation of the autoencoder. In our case, this latter is useful as it enables us to model full posterior uncertainty over the visual characteristics for each corresponding product image in a principled manner, while simultaneously enabling us to enforce disentanglement via statistical independence over the learned visual characteristics in aggregate.

Several methods accordingly extend the VAE in this manner by reformulating and/or reweighing the two terms in the traditional VAE, namely, the reconstruction accuracy between and observed image and the reconstructed image from the encoding and decoding process, and the KL-divergence of the learned characteristics from a reference prior distribution. (Higgins et al. (2017) showed that trading off these two terms by upweighting the KL-divergence term promoted disentanglement. This was further extended by Burgess

et al. (2017), Kim and Mnih (2018), and Chen et al. (2018), who decomposed this KL-divergence term into several sub-terms to give even more fine-grained control over the disentanglement process which we elaborate formally in Section 3.2.

One key challenge of any disentanglement method is that, with purely unsupervised methods, there is no theoretical guarantee for learning unique disentangled representations Locatello et al. (2019). In other words, we need some form of relevant supervision to identify independent and semantically interpretable visual characteristics. To address this challenge, Locatello et al. (2020) showed that a small number of labelled examples with even potentially imprecise and incomplete labels is sufficient to perform model selection to learn disentangled representations. Further, Sisodia et al. (2022) showed that structured product characteristics can also serve as a supervisory signal. An advantage of using structured product characteristics is that this data is typically available in marketing datasets. This means that there is no need for additional labeling. In this paper, we address this theoretical challenge by using a demand model as a supervisory signal. Implicitly, we assume that demand is a function of visual characteristics as well as other known observed product characteristics.

2.3. Empirical Models with Unstructured Data

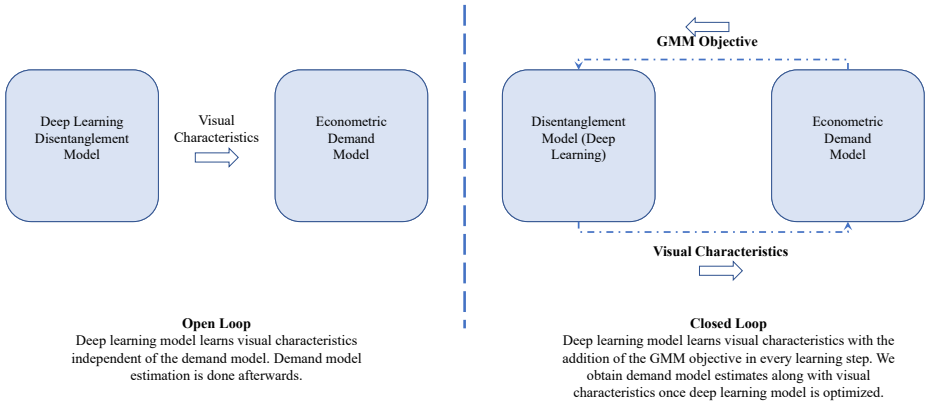
Unstructured data (text, images etc.) are rich in content and very high dimensional, which makes it challenging to tractably incorporate in marketing models. There are several approaches to incorporate visual characteristics to study their effect on demand. First, we can define a specific set of visual or textual attributes of interest. For example, in a study on AirBnB images, Zhang et al. (2021) examines specific interpretable properties of images like composition, color or aspect ratio that can be derived from the image. Dew et al. (2021) combine textual and image data to augment visual design of new brand logos. Timoshenko and Hauser (2019) identify consumer needs for new products from textual review data. Liu et al. (2020) study how brands visual identity can be listened in using online images. Similarly, Zhang and Luo (2018) defines the visual characteristics of images uploaded by customers to Yelp to study their effect on restaurant exit, and Malik et al. (2019) defines the visual characteristics of profile pictures from an online professional social network to study beauty premium in career progression.[4] Burnap et al. (2019) uses images to augment

[4] For example, a researcher studying the effect of hair color on a worker's chance of getting promoted might conclude that hair color has explanatory power. However, hair color is correlated with race and the explanation behind the worker's promotion might possibly be due to race.

how designers develop the aesthetics of products. Overall, in this approach, the selected characteristics typically have to be human labeled. We also rely on researcher judgment, which creates a challenge if they select an unimportant visual characteristics while omitting an important one. A second approach is to use classical statistical methods like PCA or autoencoders (Bengio et al. 2013), which can reduce the dimensionality automatically, however the disadvantage is that the characteristics obtained are not interpretable.

Our disentanglement approach aims to achieve the advantages of both the above, automatically learning (discovering) and quantifying visual characteristics without human labeling, but also obtaining independent and semantically-interpretable visual characteristics. Our closed loop model also allows the deep learning and demand models to learn from each other, with a view to improving the outcomes of each model.

Figure 1 An Integrated Model of Econometric Demand and Disentangled Representation Learning



3. Methodology

We now propose a method for estimating the economic value of visual product characteristics. Our method may be at a high level conceptualized as a integrated model comprised of two component models, an econometric demand model and disentanglement-based machine learning model.

Figure 1 diagrams an overview of our integrated approach. At a high level, the disentanglement-based machine learning model identifies visual characteristics from images of products, while the demand model identifies how those visual characteristics (alongside other more conventional structured product characteristics) affect a demand system comprising a competitive market of firms and heterogeneous consumers. We next formalize the

two components of our integrated model separately, followed by their integration using two distinct model specifications. Section 3.1 describes the demand model, including its supply-side and demand-side assumptions, as well as parameter estimation. Section 3.2 describes the disentanglement-based deep learning model that identifies visual characteristics from product images. Section 3.3 describes the proposed integrated model, and specifically, the "open loop" and "closed loop" specifications of this model. We provide a table of notation in Table 1.

Table 1 Table of Notation

Demand Model		Disentanglement Model	
Symbol	Meaning	Symbol	Meaning
j	Products	$\mathbf{m}$	image
t	Markets	$\mathbf{v}$	visual characteristic
i	Consumers	$\rho(\mathbf{v})$	prior distribution
f	Firms	$\rho_\theta(\mathbf{m}\|\mathbf{v})$	Decoder neural net
T	Number of Markets	$q_\phi(\mathbf{v}\|\mathbf{m})$	Encoder neural net
J_t	Number of products in market t	θ	Decoder's parameters
I_t	Number of consumers in market t	ϕ	Encoder's parameters
F_t	Number of firms in market t	c	Ground-truth factor of variation
ζ	Model Parameters	$\mathbf{v}_{\text{inf}}$	informative visual characteristic
ζ_1	Linear demand-side parameters	$L(\theta, \phi; \mathbf{m}, \mathbf{v})$	Total Loss
ζ_2	Non-linear common parameters	$\mathbf{E}_{q_\phi(\mathbf{v}\|\mathbf{m})}\left[\log \rho_\theta(\mathbf{m}\|\mathbf{v})\right]$	Reconstruction Loss
ζ_3	Linear supply-side parameters	$I_q(\mathbf{v}, \mathbf{m})$	Mutual Information Loss
p_{jt}	Price	$KL\left[q(\mathbf{v})\|\|\prod_{j=1}^{J} q(v_j)\right]$	Total Correlation Loss
c_{jt}	Marginal Cost	$\sum_{j=1}^{J} KL\left[q(v_j)\|\|\rho(v_j)\right]$	Dimension KL Divergence Loss
x_{jt}	Observed product characteristic	λ_1	Weight on Total Correlation Loss
v_{jt}	Visual product characteristic	λ_2	Weight on GMM Objective
U_{ujt}	Indirect utility		
δ_{jt}	Mean utility		
μ_{ijt}	Heterogeneous utility		
ϵ_{ijt}	Idiosyncratic taste shock		
d_{ijt}	Choice indicator		
s_{ijt}	Choice probability		
s_{jt}	Market share		
ξ_{jt}	Demand-side structural error		
ω_{jt}	Supply-side structural error		
Z^D	Demand Instruments		
Z^S	Supply Instruments		
W	Weighting matrix		
g	Sample Moments		

3.1. Demand Model

We describe the BLP demand model (Berry et al. 1995) with the specification laid out in Berry et al. (1999).

3.1.1. Consumers In each market $t = 1, \ldots, T$, there are J_t differentiated goods and I_t consumers. For each market, we observe average quantities, prices and product characteristics for all J_t products.

Consistent with the standard BLP model, the indirect utility of consumer i from purchasing product j in market t is a function of observed product characteristics $\mathbf{x_{jt}}$, unobserved product-market characteristics ξ_{jt}, price p_{jt}, consumer characteristics $\nu_{\mathbf{it}}$ and unknown parameters, ζ. The total number of observed characteristics for the product is K. We use the specification written in Equation (1). Here, price p_{jt} is endogeneous, since it could be based on the unobserved product-market characteristics ξ_{jt}, and hence correlated with it. The indirect utility is specified as:

$$U_{ijt} = \mathbf{x_{jt}}\overline{\beta_1} - \alpha p_{jt}/y_{it} + \xi_{jt} + \sum_k (\sigma^k_{\beta_1} x^k_{jt} \nu^k_{it}) + \epsilon_{ijt} \tag{1}$$

where $\mathbf{x_{jt}}$ or observed product characteristics only includes structured product characteristics, p_{jt} is the price of product j in market t, y_{it} is the income of the consumer i in market t, ξ_{jt} is the unobserved product-market characteristic, ν^k_{it} represents consumers i's taste for characteristic k in market t, and finally, ϵ_{ijt} denotes a mean-zero idiosyncratic taste shock. The unobserved product-market characteristics can reflect hard to quantify aspects of the product such as quality or style. The unobserved product characteristics can be decomposed into visual product characteristics $\mathbf{v_{jt}}$ and rest of the unobserved product-market characteristics $\tilde{\xi}_{jt}$. This decomposition is written in Equation 2. We assume that there is no heterogeneity in preferences over visual characteristics. In a demand model without visual characteristics, this would reduce to the typical BLP structural error ξ_{jt}. Note that, even after we account for visual characteristics, the remaining unobserved product-market characteristics $\tilde{\xi}_{jt}$ may still contain some unobserved visual characteristic, as well as any other aspects of unobservable quality.

$$\xi_{jt} = \mathbf{v_{jt}}\overline{\beta_2} + \tilde{\xi}_{jt} \tag{2}$$

Each consumer i in market t has unit demand. Consumers choose among $J_t = \{0, 1, \ldots, J_t\}$ discrete choices including the outside good, denoted by $j = 0$. The utility of the outside good represents the choice of not purchasing any product in the market and is given by $U_{i0t} = \epsilon_{i0t}$. Consumers select the alternative (including outside good) with the highest utility:

$$d_{ijt} = \begin{cases} 1 \text{ if } U_{ijt} > U_{ikt} \text{ for all } k \neq j \\ 0 \qquad\quad \text{otherwise} \end{cases} \tag{3}$$

Note that as in BLP, we can decompose the indirect utility in Equation (1) into a mean utility, $\delta_{\mathbf{jt}}$ in Equation (4) and a deviation from that mean, μ_{ijt}, in Equation (5).

$$\delta_{jt}(\mathbf{x_{jt}}, p_{jt}, \xi_{jt}; \zeta_{\mathbf{1}}) = \mathbf{x_{jt}}\overline{\beta_{\mathbf{1}}} + \mathbf{v_{jt}}\overline{\beta_{\mathbf{2}}} + \tilde{\xi}_{jt} \tag{4}$$

$$\mu_{ijt}(\mathbf{x_{jt}}, p_{jt}, \nu_{\mathbf{ijt}}, y_i; \zeta_{\mathbf{2}}) = -\alpha p_{jt}/y_{it} + \sum_k (\sigma^k_{\beta_1} x^k_{jt} \nu^k_{it}) + \epsilon_{ijt} \tag{5}$$

where $\mathbf{x_{jt}}$ or observed product characteristics only includes structured product characteristics, $\mathbf{v_{jt}}$ is the visual product characteristics, p_{jt} is the price of product j in market t, y_{it} is the income of the consumer i in market t, $\tilde{\xi}_{jt}$ is the unobserved product-market characteristic, ν^k_{it} represents consumers i's taste for characteristic k in market t, and finally, ϵ_{ijt} denotes a mean-zero idiosyncratic taste shock.

We denote $\zeta = (\zeta_{\mathbf{1}}, \zeta_{\mathbf{2}})$, a vector of all the parameters in the model. The vector $\zeta_{\mathbf{1}}$ contain the linear parameters or the mean preference on $\mathbf{x_{jt}}$, i.e. $\overline{\beta_{\mathbf{1}}}$ and on $\mathbf{v_{jt}}$, i.e. $\overline{\beta_{\mathbf{2}}}$. These preferences are common across all consumers. The vector $\zeta_{\mathbf{2}}$ contain the nonlinear parameters or the standard deviation from mean preference i.e. σ_{β_1} as well as the term on the price α. These nonlinear parameters introduce heterogeneity in preferences over structured product characteristics.

Using the standard assumption that ϵ_{ijt} are i.i.d. with the Type I extreme value distribution, the probability s_{ijt} that consumer i chooses product j in market t is given by Equation (6) below.

$$s_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\Sigma_{k \in J_t} \exp(\delta_{kt} + \mu_{ikt})} \tag{6}$$

Aggregate market shares s_{jt} are obtained by integrating over the choices made by heterogeneous consumers as represented in Equation (7).

$$s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{\Sigma_{k \in J_t} \exp(\delta_{kt} + \mu_{ikt})} dFi \tag{7}$$

3.1.2. Firms We assume that automobile firms, indexed by f and part of a set F_t, play a static, full information, simultaneous move pricing game each period. Firms choose the price levels of all their models with the objective of maximizing overall profit. We specify a constant marginal cost c_{jt} for a product j in market t. The pricing first order condition for vehicle j is given by Equation (8).

$$s_{jt} + \Sigma_{j \in J_t}(p_{jt} - c_{jt}) \frac{\partial s_{jt}}{\partial p_{jt}} = 0 \tag{8}$$

We parameterize the marginal costs as written below in Equation (9).

$$c_{jt} = x_{jt}\gamma_1 + w_{jt}\gamma_2 + \omega_{jt} \tag{9}$$

where x_{jt} are product characteristics, w_{jt} are observable cost-shifters and ω_{jt} are unobserved cost-shifters. We can estimate the marginal costs for each product when we solve the supply model jointly with the demand model. We do not explicitly include visual characteristics in the supply-side and so they are assumed to be part of the unobservables.

3.1.3. Instruments In this demand model, we assume that a consumer's utility depends up on the observed product characteristics as well as unobserved (to the researcher) product characteristics. Firms observe these unobserved product characteristics and set then set prices, which implies that price is endogenous and necessitates the use of instruments. There are multiple options for instruments. First, we could use *exogenous cost shifters*. These are valid if we assume that firms respond to cost shifts by changing prices, and not by changing product characteristics. Second, we could use *observed product characteristics* other than price. This would be valid if we make a timing assumption that firms first set observed product characteristics, then observing the "unobserved" product characteristics (structural error), and then set prices. This assumption would be supported by the observation that firms change prices frequently, whereas product characteristics are altered less frequently. Third, we could use Hausman instruments, a common example of which includes prices in other markets, if we have multiple markets that have the same product. Finally, we could use observed product characteristics of other products. One example of these instruments is referred to as *BLP instruments* in which we take sums of characteristics of other products made by the same firm and sums of characteristics of all other firms. We present them in Equation (10). Another example instead uses differences

in characteristics. They are referred to as *differentiation IVs* (Gandhi and Houde 2019). To construct *differentiation IVs*, we define the difference in characteristic space between product j and product k (in market t) as $d_{jkt} = d_{kt} - d_{jt}$ for each characteristic in x_{jt}. Using the differences in characteristic space, we construct local and quadratic forms of *differentiation IVs*. Both these set of *differentiation IVs* assume that rival as well as non-rival products close to the focal product in the characteristics space affect the firm's pricing decision more than products far away. We present these instruments in Equation (11) and Equation (12). Conlon and Gortmaker (2020) found that *differentiation IVs* perform better than *BLP instruments* in simulation exercises. This result is not surprising because of the somewhat localized nature of competition in the product characteristic space. Moreover, differentiation IVs also reduces the problem of weak instruments that might be present in *BLP instruments*. In our main analysis, we use the local form of *differentiation IVs*,

$$Z_{BLP} = \{1, x_{jt}, w_{jt}, \Sigma_{j \in J_t \ \{j\}} 1, \Sigma_{j \notin J_t} 1, \Sigma_{j \in J_t \ \{j\}} x_{jt}, \Sigma_{j \notin J_t} x_{jt}\} \tag{10}$$

$$Z_{local} = \{1, x_{jt}, w_{jt}, \Sigma_{j \in J_t \ \{j\}} 1(|d_{jkt}| < \sigma_l(d)), \Sigma_{j \notin J_t} 1(|d_{jkt}| < \sigma_l(d)\} \tag{11}$$

where $\sigma_l(d)$ denotes the standard deviation of the distance in the characteristic space.

$$Z_{quadratic} = \{1, x_{jt}, w_{jt}, \Sigma_{j \in J_t \ \{j\}} d^2_{jkt}, \Sigma_{j \notin J_t} d^2_{jkt}\} \tag{12}$$

With the addition of demand instruments Z^D_{jt}, we construct demand-side moment conditions of the form $E[\tilde{\xi}_{jt} Z^D_{jt}] = 0$. Similarly, we also construct supply-side moment conditions of the form $E[\omega_{jt} Z^S_{jt}] = 0$ using supply instruments Z^S_{jt}.

3.1.4. GMM Estimator We construct a GMM estimator using both supply-side and demand-side moment conditions.

$$g(\theta) = \begin{bmatrix} \frac{1}{N} \Sigma_{jt} E[\tilde{\xi}_{jt} Z^D_{jt}] \\ \frac{1}{N} \Sigma_{jt} E[\omega_{jt} Z^S_{jt}] \end{bmatrix} \tag{13}$$

We construct a nonlinear GMM estimator for ζ with some weighting matrix W in Equation (14). We solve this problem twice. First, we obtain a consistent estimate of W and then an efficient GMM estimator.

$$\min_{\theta} \ g(\zeta)' W g(\zeta) \tag{14}$$

3.2. Disentangled Representation Learning

We now present a method to discover independent and human interpretable visual characteristics from product images. We refer readers to Sisodia et al. (2022) for more details. We have a dataset $\mathbf{m}$ of images. We assume that they are generated from a distribution parameterized by visual characteristics $\mathbf{v}$. The generative model is a combination of the prior $\rho(\mathbf{v})$ set to an isotropic unit Gaussian $\mathcal{N}(0,1)$ and a decoder neural net $\rho_\theta(\mathbf{m}|\mathbf{v})$. The true posterior is intractable as in variational Bayesian inference (Blei et al. 2017) and so it is approximated as $\log q_\phi(\mathbf{v}|\mathbf{m}) = \log \mathcal{N}(\mathbf{v}; \boldsymbol{\mu_d}, \boldsymbol{\sigma_d}^2\mathbf{I})$ where $\boldsymbol{\mu_d}$ and $\boldsymbol{\sigma_d}$ are the mean and the s.d. of the approximate posterior. The loss for the original VAE is written in Equation (15). We refer readers to Kingma and Welling (2014) for its detailed derivation.

$$\underbrace{L(\theta,\phi;\mathbf{m},\mathbf{v})}_{\text{VAE Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{v}|\mathbf{m})}\left[\log \rho_\theta(\mathbf{m}|\mathbf{v})\right]}_{\text{Reconstruction Loss}} + \underbrace{KL\left[q_\phi(\mathbf{v}|\mathbf{m})||\rho(\mathbf{v})\right]}_{\text{Regularizer Term}} \tag{15}$$

In Equation (16), we decompose the regularizer term in Equation (15) into three terms (Chen et al. 2018, Hoffman and Johnson 2016, Kim and Mnih 2018). We follow the $\beta-$TCVAE method (Chen et al. 2018) by imposing a heavier penalty on the total correlation loss term. We provide an intuition behind each of the loss terms below.

$$\begin{aligned}\underbrace{L(\theta,\phi;\mathbf{m},\mathbf{v})}_{\text{Open Loop Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{v}|\mathbf{m})}\left[\log \rho_\theta(\mathbf{m}|\mathbf{v})\right]}_{\substack{\text{Reconstruction}\\ \text{Loss}}} + \underbrace{I_q(\mathbf{v},\mathbf{m})}_{\substack{\text{Mutual}\\ \text{Information}\\ \text{Loss}}} \\ + \lambda_1 \underbrace{KL\left[q(\mathbf{v})||\prod_{j=1}^{J} q(v_j)\right]}_{\substack{\text{Total Correlation}\\ \text{Loss}}} + \underbrace{\sum_{j=1}^{J} KL\left[q(v_j)||\rho(v_j)\right]}_{\substack{\text{Dimension-Wise}\\ \text{KL Divergence Loss}}}\end{aligned} \tag{16}$$

Reconstruction Loss: Penalizing this term encourages the input data $\mathbf{m}$ to be as similar to the reconstructed output $\hat{\mathbf{m}}(\mathbf{v})$ as possible. This means that we want the discovered visual characteristics to have the necessary information so that the reconstructed output is as close as possible to the input image.

Mutual Information Loss: Penalizing this term means encouraging the visual characteristics $\mathbf{v}$ store as little information about the product image $\mathbf{m}$ as possible from an information-theoretic point of view (Achille and Soatto 2018). Although it seems counter

intuitive that penalizing this term would encourage disentangled representation, but it allows the visual characteristics to not store any nuisance information

Total Correlation Loss: Penalizing this term encourages the discovered visual characteristics $\mathbf{v}$ to be statistically independent (Watanabe 1960). A zero loss on this term means that KL divergence is zero and the discovered visual characteristics are statistically independent. A $\lambda_1 > 1$ penalty on this term means that we are trading off reconstruction accuracy for more statistically independent disentangled representations.

Dimension-Wise KL Loss: Penalizing this term encourages the distribution of each visual characteristic of every datum to be close the prior distribution. The prior is typically assumed to be Gaussian. This term promotes a continuous latent space, which allows generation from a smooth and compact region of latent space.

3.3. Integrated Model of Visual Characteristics Discovery and Demand

In this section, we describe our proposed integrated model of econometric demand and disentangled representation learning, referred to integrated model hereafter. Our integrated model furthermore has two different model specifications, which we call **open loop** and **closed loop**.

In both open and closed loop model specifications, we assume that a consumer's utility depends upon observed product characteristics (including structured and visual product characteristics), unobserved product characteristics, price as well as an idiosyncratic taste. The primary difference between the open loop and closed loop model specification is whether the demand model and disentangle representation learning model is separable or not. We later test the two specifications of of integrated model against a baseline approach. The baseline approach (Model B) is the same as the one described in Berry et al. (1995), and does not incorporate visual characteristics.

In the first model specification, **open loop**, we first automatically discover the visual characteristics using disentangled representation learning without specifying the number or nature of those characteristics. The discovered characteristics are obtained for each product, and then included as visual product characteristics in the demand model. These visual characteristics are thus similar to structured product characteristics and can be treated as data during the estimation of the demand model. We can obtain estimates of consumer preferences for these visual characteristics similar to how preferences over structured characteristics are ordinarily obtained.

The open loop class of demand model relies on an unsupervised method for finding disentangled representations. Although the unsupervised method for finding disentangled representations has been found to work in practice, Locatello et al. (2019) proved out that there was no theoretical guarantee to learn disentangled representations without supervision. When the visual characteristics are known ex-ante, then we refer those characteristics as ground truth. When ground truth for visual characteristics is available, then the deep learning literature suggests using the learned visual characteristics to predict the known labels of ground truth Locatello et al. (2020).

In the second specification, called **closed loop**, we allow for the disentanglement representation learning to use the demand model as a supervisory signal. Although supervisory signals used in the ML literature always rely on knowing ground truth about the visual characteristics, such an approach would not be suitable here since the purpose of out framework is to specifically discover these visual characteristics. Our idea of using the demand model to supervise the disentanglement model addresses this theoretical challenge by supervising the VAE-based method to learn disentangled representations on the demand model discussed in the Section 3.1.

The rationale for supervising on a demand model is that it captures how consumers respond to changes in both structured and visual product characteristic space. This endogenous response of consumers making choices and ensuring that those choices are consistent with observed market outcomes provides us with additional variation that the disentanglement model can learn from. Supervising on a demand model means that the demand model's loss (or objective) function is included in the disentanglement model's objective function. The modified loss equation is specified in Equation (17). Whereas in both open and closed loop specifications, the visual characteristics are incorporated similarly into the demand model, in the open loop the visual characteristics discovered at the completion of the disentanglement model do not change during the estimation process. However, in the closed loop, as the demand model is estimated, the parameters (and objective function) of the learned demand model impact the estimation and results of the disentanglement model, and therefore the discovered visual characteristics are continually updated to be consistent with the demand model. In short, the discovered visual characteristics and the estimates from the demand model will be different depending on whether we use open loop or closed loop.

We detail the implementation details of how the demand model is incorporated within the deep learning model in Table 1.

$$\underbrace{L(\theta,\phi;\mathbf{m},\mathbf{v})}_{\text{Closed Loop Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{v}|\mathbf{m})}\left[\log p_\theta(\mathbf{m}|\mathbf{v})\right]}_{\substack{\text{Reconstruction}\\ \text{Loss}}} + \underbrace{I_q(\mathbf{v},\mathbf{m})}_{\substack{\text{Mutual}\\ \text{Information}\\ \text{Loss}}} + \lambda_1 \underbrace{KL\left[q(\mathbf{v})||\prod_{j=1}^{J} q(v_j)\right]}_{\substack{\text{Total Correlation}\\ \text{Loss}}} + \underbrace{\sum_{j=1}^{J} KL\left[q(v_j)||p(v_j)\right]}_{\substack{\text{Dimension-Wise}\\ \text{KL Divergence Loss}}} + \lambda_2 \underbrace{GMM(\mathbf{v})}_{\substack{\text{Demand Model}\\ \text{Objective}}} \tag{17}$$

Algorithm 1 Closed Loop Estimation Algorithm

Initialize visual characteristics
Estimate demand model and store ($\overline{\beta}$)
do ▷ Training Epochs Start
  do ▷ Steps within Epoch
    Learn $v_j t$ for images in a batch
    Compute $\tilde{\xi}_{jt}$ using the $v_j t$ and $\overline{\beta}$
    Compute $E[\tilde{\xi}_{jt} Z_{jt}]$ and add it to the overall loss term
  while end of epoch ▷ Epoch Ends
  Obtain visual characteristics of entire data
  Update mean coefficients ($\overline{\beta}$)
  Input $\overline{\beta}$ in subsequent epoch
while *epochs* ≤ 200 ▷ Training Ends

3.3.1. Hyperparameter Selection using UDR Metric Both open and closed loop models require us to make modeling choices in the form of hyperparameters, which impact the estimation process but are not parameters estimated with the model (e.g., number of training epochs). In both model specifications, we have a number of hyperparameters, which are detailed in Table 2.

In the open loop model, we have a hyperparameter λ_1, which is the weight on the total correlation loss within the disentanglement loss (Chen et al. 2018). In the closed loop model, in addition to λ_1, we have a hyperparameter λ_2, which represents the weight of the

demand loss when incorporated into the loss of the disentanglement model. For instance, a higher value of λ_2 will weigh or prioritize the demand loss more relative to disentanglement loss terms like mutual information or reconstruction loss, which could reduce the quality of disentanglement. See Section 3.2 for more details.

Hyperparameter selection additionally requires we define a metric for model selection. We use the Unsupervised Disentanglement Ranking (UDR) (Duan et al. 2020) metric for hyperparameter selection. This metric measures the robustness of disentangled representations to variance at different starting points. It relies on the assumption that for a particular dataset, a disentangling VAE will converge on the same disentangled representation (up to *permutation*, *sign inverse*, and *subsetting*). [5]We select the hyperparameters λ_1 and λ_2 corresponding to the highest UDR. Moreover, this metric does not require access to the ground truth data generative process unlike other metrics such as β-VAE metric (Higgins et al. 2017), the FactorVAE metric (Kim and Mnih 2018), Mutual Information Gap (MIG) (Chen et al. 2018) and DCI Disentanglement scores (Eastwood and Williams 2018). We describe the steps to calculate UDR below in Table 3.

3.3.2. Neural Net Architecture Figure 2 shows the detailed neural net architecture. Our architecture is a modified version of the one used in Burgess et al. (2017). We modify the architecture to use images of 128×128 pixels as well as to incorporate a demand model. We use Convolutional Neural Net (CNNs) to construct the encoder neural net because we are working with images. We stack a sequence of CNN layers in the encoder neural net so that we learn high-level concepts for images. We then use 2 fully-connected (FC) layers to first flatten the output of the sequence of CNN layers and then reduce the number of dimensions in order to learn a maximum of 20 visual characteristics. The decoder neural net is simply the transpose of the encoder neural net, and is designed to reconstruct the image from the 20-dimensional latent visual characteristics. Finally, we feed the discovered visual characteristics in conjunction with structured product characteristics

[5]
1. Permutation: The same ground truth factor may be encoded by a model with two different seed values at a different index position.
2. Sign inverse: A model with two different seed values may learn to encode the values of the generative factor in the opposite order to each other.
3. Subsetting: A model from one seed value may learn a subset of the factors that the a model with a different seed value has learned. This is because different seed values may encourage a different number of latents to be switched off in the two models.

Table 2 Hyperparameter Selection (Open and Closed Loop)

Step	Description
1	Fix all the hyperparameters except λ_1 (and λ_2 in case of closed loop)
2	Fix[‡1] batch size=64, number of visual characteristics=20, learning rate=0.0005, and the number of epochs=200
3	Sweep over $\lambda_1 = [1, 5, 10, 20, 30, 40, 50]$ (and $\lambda_2 = [0, 5, 10, 20, 30, 40, 50]$ in case of closed loop)
4	Obtain disentangled representations for every combination of λ_1 (and λ_2 in case of closed loop)
5	Calculate Unsupervised Disentanglement Ranking (UDR)[‡2] Duan et al. (2020)
6	Select λ_1 (and λ_2 in case of closed loop) for disentanglement model with the highest UDR.
7	Obtain the learned visual characteristics $\mathbf{v}$ for both open loop and closed loop

[‡1] Intuition for Hyperparameters:

1. Batch Size: On the one hand, if a very low value for batch size is used, then the model takes longer to converge. On the other hand, if a very high value for batch size is used, then the model loses its generalizability beyond the training set.
2. Number of Visual Characteristics: On the one hand, if a very low number of visual characteristics is specified, then the model would force multiple factors of variation to be coded into a single visual characteristic. On the other hand, if a very high number of visual characteristics is specified, then the model would encourage a single factor of variation would be split into multiple visual characteristics.
3. Learning Rate: On the one hand, if a very low learning rate is used, then the model can get stuck on a local minima. On the other hand, if a very high learning rate is used, then the model may overshoot the minima.
4. Number of Epochs: On the one hand, training for a very low number of epochs may lead the model not to converge. On the other hand, training for a very high number of epochs may lead to overfitting the training data.

[‡2] See Section 3.3 for details related to UDR

to estimate demand. We add the GMM objective from the demand model to the loss of the neural network. This ensures that the learned visual characteristics not only help in disentanglement but also help in finding a lower GMM objective for the demand model. Note that, the demand model is not estimated using neural nets and so not included part of the nerual net architecture.

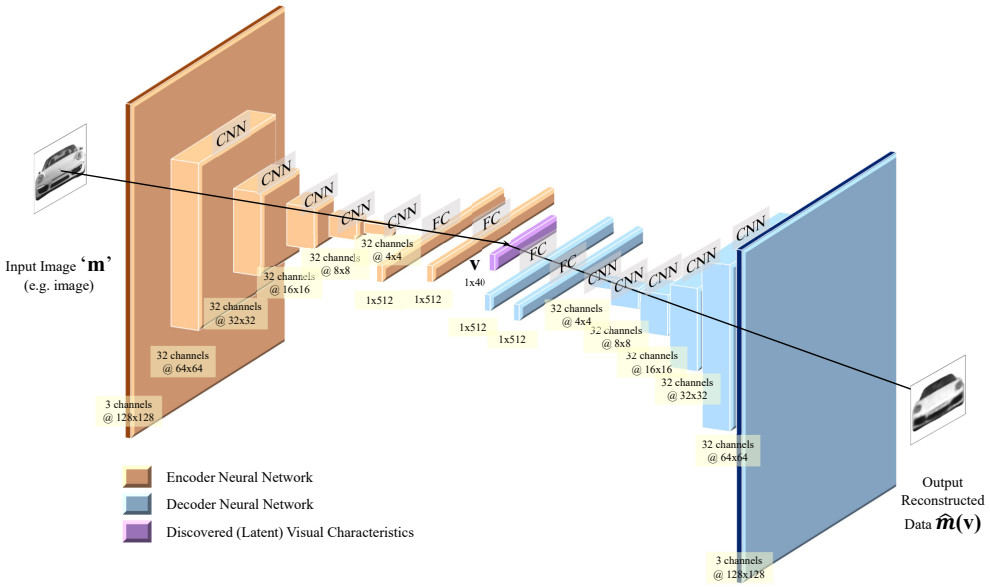
Table 3 UDR Algorithm

Step	Description
1	For each trained model $\tau(\lambda)$, perform $\kappa = 10$ pairwise comparisons
2	Pairwise comparisons: models trained with the same λ but with different seed values
2	Calculate the $UDR_{\tau_{s_1}\tau_{s_2}}$, where τ_{s_1} and τ_{s_2} index the model τ learned with two different seed values
3	Calculate $UDR_{\tau_{s_1}\tau_{s_2}}$ score as similarity matrix $R_{\tau_{s_1}\tau_{s_2}}$ where each entry is the Spearman correlation between the responses of individual latent units of the two models.
4	Calculate absolute value of the similarity matrix as $\lvert R_{\tau_{s_1}\tau_{s_2}} \rvert$
5	Compute the score $UDR_{\tau_{s_1}\tau_{s_2}}$ for each pair of models‡
6	Compute the final score UDR_{τ} for model τ by taking the median of $UDR_{\tau_{s_1}\tau_{s_2}}$

‡ $UDR_{\tau_{s_1}\tau_{s_2}} = \frac{1}{v_{\inf_a}+v_{\inf_b}} \left[\Sigma_b \frac{r_a^2 I_{KL}(b)}{\Sigma_a R(a,b)} + \Sigma_a \frac{r_b^2 I_{KL}(a)}{\Sigma_b R(a,b)} \right]$

where a and b index the latent units of models τ_{seed_1} and τ_{seed_2}, respectively, $r_a = max_a R(a,b)$ and $r_b = max_b R(a,b)$. I_{KL} indicates an *informative* visual characteristics within a model and $v_{\inf}$ is the number of such characteristics: $v_{\inf_a} = \Sigma_a I_{KL}(a)$ and $v_{\inf_b} = \Sigma_b I_{KL}(b)$

Figure 2 Schematic Illustration of Proposed Approach



4. Empirical Application: UK Automobile Market

4.1. Data

We compiled a data set covering 2008 through 2017 consisting of automobile characteristics, market shares and their images from the UK. Our primary source of data is manufacturer suggested retail prices (MSRP), and characteristics of all automobiles sold in the UK from

2008-2017 that we obtain from Parker's. We obtain information on sales (in 1000's) and images of the automobiles from DVM-CAR (Huang et al. 2021). Market research studies have shown that 70% of consumers identify and judge automobiles by the appearance of headlights and grille located on the face of the automobile.[6] So we only select the images of the front face of the automobiles and ignore other views. Since our sales data comes at the make-model level, we choose the entry-level trim's product characteristic as the product characteristic. This is similar to similar to other studies of the automobile market (Berry et al. (1995), Petrin (2002), Berry et al. (2004)). We have product characteristics for number of seats, number of doors, weight, horsepower, length, width, miles per gallon, luggage capacity and dummy variables for whether the automobile has automatic transmission, front wheel or rear wheel or four wheel drivetrain. The price variable is the list price (in £1000's) for the entry-level model. Prices in all years are deflated to 2015 UK using the consumer price index.

We supplemented the Parker's information with additional information, including vehicle country of production and company ownership information. We also supplemented additional information from the Office of National Statics, UK. We gathered the price of ultra low sulphur petrol per gallon and ultra low sulphur diesel per gallon as well as the number of households in the UK. Similar to Berry et al. (1995), we calculated miles per UK pound (MP£) as miles per gallon divided by the price per gallon. We measure the market size as the number of households in the UK. We only use 'HP/Weight' and 'MP£' to construct the local form of *differentiation IVs*.

In Table 4, we display summary statistics for the products at the make-model-year level. There are 2439 observations in our sample and a total of 409 distinct models. The variables include quantity (in units of 1000), price (in £000 units), the ratio of horsepower to weight (in HP per 10 lbs.), the number of ten mile increments one could drive for one £ of gasoline (MP£), tens of miles per gallon (MPG), and size (measured as length times width). We provide sales-weighted means for each variable. We see that automobiles have improved in terms of both power and fuel efficiency over these ten years. Prices for automobiles has also gone up substantially from 2008 to 2017.

[6] URL: https://www.wsj.com/articles/SB114195150869994250

Table 4 Descriptive Statistics

Market	No. of Models	Quantity	Price	HP/Wt	Size	MPG	MP£
2008	230	6.433	16.978	0.366	1.242	4.359	0.754
2009	244	6.382	16.765	0.358	1.233	4.585	0.874
2010	231	7.169	17.125	0.358	1.247	4.685	0.787
2011	225	7.251	17.211	0.358	1.263	4.787	0.734
2012	236	7.498	17.352	0.359	1.268	4.942	0.766
2013	244	8.235	17.299	0.362	1.273	5.077	0.815
2014	252	8.831	17.501	0.371	1.276	5.192	0.892
2015	256	9.321	18.474	0.384	1.292	5.266	1.038
2016	263	9.268	19.742	0.398	1.310	5.242	1.064
2017	258	8.693	20.504	0.405	1.324	5.136	0.992
All	2439	7.952	18.060	0.374	1.277	4.977	0.889

In Table 5, we provide a range of continuous demand characteristics along by providing the minimum, median and maximum values. The least expensive automobile in our sample is the Hyundai Amica while the Audi R8 is the most expensive. Similarly, Citroen Dispatch has the lowest ratio of horsepower to weight and Nissan GT-R has the highest ratio of horsepower to weight. Smart fortwo is the smallest automobile and Audi SQ7 is the largest automobile in terms of space. Finally, the Vauxhall VXR8 is the least fuel-efficient automobile while the Lexus CT is the most fuel-efficient automobile.

Table 5 Range of Continuous Demand Characteristics

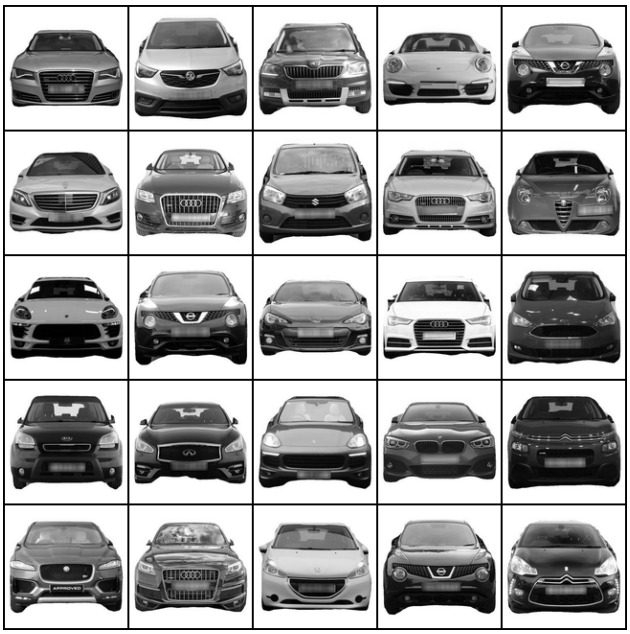
Variable	0th Percentile	50th Percentile	100th Percentile
Price	5.762	20.475	103.420
	Hyundai Amica	Volkswagen Touran	Audi R8
HP/Wt	0.2147	0.4052	1.3589
	Citroen Dispatch	Toyota Prius+	Nissan GT-R
Size	0.7495	1.3257	1.7380
	Smart fortwo	Mitsubishi Grandis	Audi SQ7
MPG	18.00	44.00	75.43
	Vauxhall VXR8	Suzuki SX4	Lexus CT
MP£	3.181	7.856	14.684
	Maserati Granturismo	Ford EcoSport	Kia Niro

In Figure 3, we display images of 25 automobiles present in our dataset. Note that, we converted color images to grayscale for our study (sales are also not available separately by color). For each image, we have its associated make, model, year, structured product characteristics and price.

4.2. Results

4.2.1. Disentanglement For model training, we initially set model hyperparameters including the number of epochs=200, batch size=64, number of latent space dimen-

Figure 3 Sample of Automobile Images



sions=20, learning rate=0.0005, and the threshold value on KL loss term=0.01. Table 2 in the methodology section explains the intuition behind these terms. We select the hyperparameters corresponding to the weight on the total correlation term λ_1, and the weight on the demand loss term λ_2 by using the UDR metric described in the methodology section. Higher values of the UDR metric correspond to better disentanglement and discovery of independent visual characteristics. Table 6 details the UDR metrics for the top 5 (high UDR) and bottom 3 (low UDR) configurations. We find that the best model corresponds to a positive weight on the demand loss term. The best configuration was with the weight on the total correlation term or $\lambda_1 = 30$ and the weight on the demand loss term or $\lambda_2 = 40$. This finding implies that supervising the disentanglement model on the demand model (closed loop) aids in finding visual characteristics, as measured by the UDR metric. The best disentanglement model with no supervision (open loop) corresponds to the weight on the total correlation term or $\lambda_1 = 30$ and the weight on the demand loss term or $\lambda_2 = 0$.

Table 6 Disentanglement Model Evaluation

Weight on total correlation loss	Weight on demand loss	UDR (Best to Worst)	Model Type
30	**40**	**0.382**	**Closed Loop**
30	30	0.381	Closed Loop
30	10	0.378	Closed Loop
30	**0**	**0.377**	**Open Loop**
40	30	0.371	Closed Loop
...	...	...	...
1	0	0.042	Open Loop
1	50	0.041	Closed Loop
1	10	0.040	Closed Loop

We show the discovered visual characteristics in Figure 4. Each row in the image corresponds to a visual characteristics. In each row, we change the value of one visual characteristic while fixing the value of all the other characteristics. We find five visual characteristics of a automobile's front view to be informative. Rest of the visual characteristics were uninformative i.e. changing the visual characteristic produces no change in the image. We interpret these visual characteristics as written below, subject to validation by human coders.

1. Body Shape: This characteristic can be interpreted as ratio of width to height. Automobiles with a low value on this visual characteristic have a low width to height ratio and vice-versa.
2. Color: Automobiles with a low value on this visual characteristic are darker and vice-versa. We ignore this characteristic in the demand model as automobiles come in multiple colors and sales data is only available at make-model level.
3. We ignore this characteristic as this is entangled with color
4. Boxiness: Automobiles scoring low on this characteristic have a box-like shape and vice-versa.
5. Grille: Automobiles scoring low on this characteristic have a smaller grille and vice-versa.

Figure 5a shows the density plot of the discovered visual characteristics. Figure 5b shows the visualization of the correlation between the three discovered visual characteristics.

4.2.2. Demand Model In this section, we compare a demand model only including structured product characteristics with a demand model including both structured product characteristics as well as visual product characteristics. Following Berry et al. (1995), we first present results using OLS Logit and IV Logit. We include the ratio of horsepower

Figure 4 Discovered Visual Characteristics

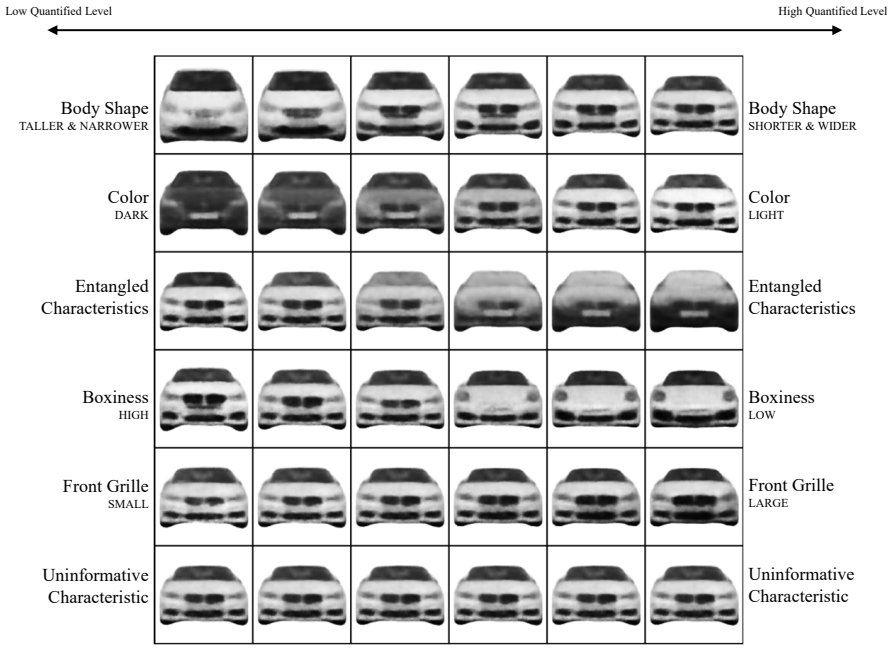
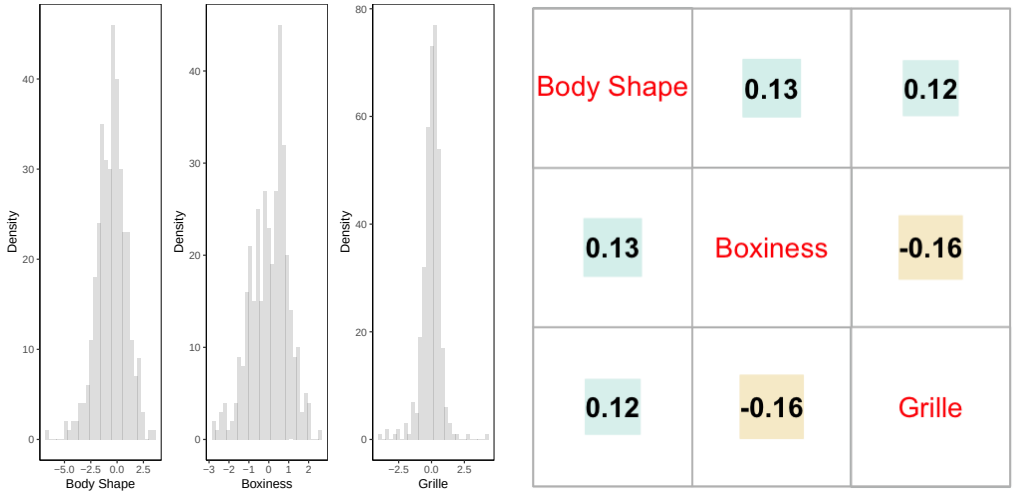


Figure 5 Summary of Discovered Visual Characteristics
(a) Density Plot of Visual Characteristics **(b) Visualization of Correlation Matrix**



to weight (HP/Weight), miles per pound (MP£), and size as structured product characteristics. We also include three visual characteristics: body shape, boxiness and grille. As expected, Table 7 shows that adding differentiated IVs as instruments for price reduces the number of products facing inelastic demand. Moreover, we note that adding visual characteristics to the demand model reduces the number of products facing inelastic demand.

Table 7 OLS Logit & IV Logit Model Results

	OLS Logit Demand		IV Logit Demand	
	Baseline	Closed Loop	Baseline	Closed Loop
Constant	−10.08	−8.89	−12.23	−10.84
	(0.29)	(0.31)	(0.55)	(0.54)
HP/Weight	−0.84	−1.63	1.69	0.85
	(0.23)	(0.25)	(0.68)	(0.67)
MP£	2.02	1.83	1.73	1.62
	(0.14)	(0.14)	(0.17)	(0.17)
Price	−0.02	−0.02	−0.07	−0.07
	(0.003)	(0.003)	(0.01)	(0.01)
$v_{bodyshape}$		0.23		0.30
		(0.03)		(0.03)
$v_{boxiness}$		0.08		−0.06
		(0.03)		(0.04)
v_{grille}		0.14		0.11
		(0.05)		(0.05)
Space	0.18	−0.19	2.11	1.58
	(0.18)	(0.18)	(0.44)	(0.43)
Observations	2,439	2,439	2,439	2,439
No. of Inelastic Demands	2,263	2,195	628	583

Next, we compare different specifications using the random coefficients demand model in Table 8. The first specification does not include visual characteristics. The second specification includes visual characteristics when the demand model was supervised on the disentanglement model (closed loop). The third specification includes visual characteristics when the demand model was not supervised on the disentanglement model (open loop). Figure 6 plots the distribution of price elasticity for a specification only including structured product characteristics (baseline) and a specification including both structured product characteristics as well as visual product characteristics (closed loop).

Table 8 Demand Model Estimates

		Baseline (1)	Closed Loop (2)	Open Loop (3)
Means ($\overline{\beta}'s$)	Constant	−11.76	−10.76	−10.72
		(20.00)	(38.89)	(33.82)
	HP/Weight	2.10	2.02	1.92
		(44.90)	(100.49)	(83.46)
	MP£	1.90	1.94	1.94
		(30.54)	(13.31)	(26.04)
	Space	1.46	0.37	0.57
		(19.48)	(19.83)	(24.15)
	$v_{bodyshape}$		0.35	0.33
			(0.17)	(0.17)
	$v_{boxiness}$		−0.15	−0.16
			(1.05)	(0.93)
	v_{grille}		0.12	−0.06
			(0.13)	(0.20)
Standard Deviation ($\sigma'_\beta s$)	Constant	0.00	0.00	0.00
		(22.03)	(33.97)	(27.51)
	HP/Weight	0.00	0.00	0.00
		(76.96)	(169.42)	(135.94)
	MP£	0.77	1.09	1.00
		(61.89)	(38.56)	(60.91)
	Space	3.34	4.79	4.48
		(14.47)	(12.17)	(11.40)
Term on Price	(-p/y)	−16.71	−20.79	−20.49
		(73.27)	(129.94)	(136.36)
Supply-Side Terms	Constant	2.96	3.07	3.06
		(0.67)	(0.77)	(0.84)
	ln(HP/Weight)	1.03	0.96	0.97
		(0.71)	(0.86)	(0.89)
	ln(MPG)	−0.22	−0.19	−0.19
		(1.15)	(1.07)	(1.17)
	ln(Space)	2.55	2.35	2.37
		(4.41)	(3.24)	(4.06)
	Trend	−0.01	−0.01	−0.01
		(0.09)	(0.04)	(0.06)
Mean Own-Price Elasticity		−2.35	−2.80	−2.75
S.D. Own-Price Elasticity		0.55	0.65	0.64
No. of Inelastic Demands		0	0	0
Mean Markup		0.46	0.39	0.40
GMM Objective		10.92	5.47	5.73

Figure 6 Density Plot - Price Elasticities

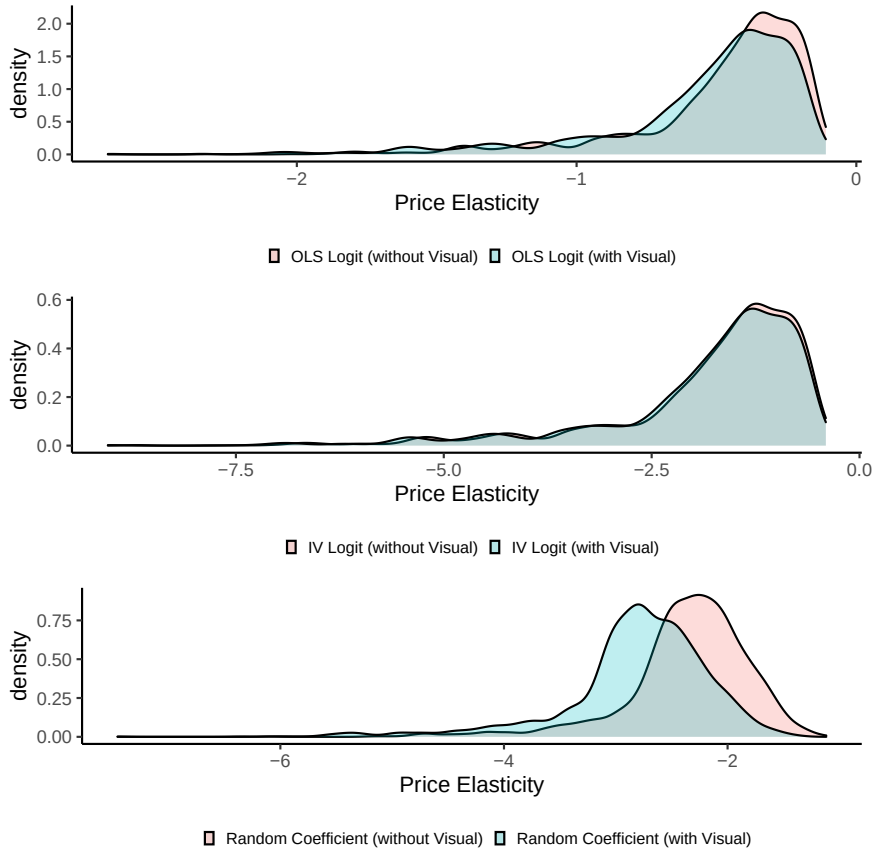


Figure 7 Evolution of BMW Kidney Bean Grille

BMW Front Grille
(2013-2018)

BMW Front Grille
(2018 - Present)

5. Counterfactual Analysis: BMW Kidney Bean Grille

> Admitting he had "noticed" and "been hurt" by the criticism of the enormous grille – which has sparked countless memes on the internet and social media – BMW design chief Adrian van Hooydonk says he understands why both buyers and the design community had been shocked.
>
> *– Car Sales Blog, June 26, 2019*

While visual characteristics are important across products (automobile models), most products would not have dramatic variation across time or markets. However, there are cases where firms choose to make dramatic changes in visual design. A recent example is BMW's change in "design language" in 2018, which was very polarizing. Historically, BMWs have always had a grille known as the "kidney beans," distinctively a visual cue for all BMW models (Bangle 2001). For model year 2018, this longstanding mainstay visual design was disrupted by dramatically elongating the grille, with many commentators having a negative opinion on the new lung-shaped grille imposed on much of the BMW product line.

We conduct a counterfactual analysis where BMW introduces such a visual design change earlier at the end of our data period (2017), and then evaluate market outcomes following this change. This counterfactual enables us to compare BMW's market shares for the original and the new design. We also seek to examine how the results differ for models that include visual characteristics and those that do not.

We detail the steps to conduct the counterfactual analysis. First, we evaluate the case where no change happens in BMW's "design language" which represents the true outcome we observe. We compute equilibrium prices and market shares for the entire market in 2017 using the demand estimates from the baseline model (without visual characteristics)

and closed loop model (with visual characteristics). We report BMW's market shares in 2017 in column (2) and column (3) of Table 9.

Next, we examine the case where BMW changes its "design language" in 2017. For the closed loop model (with visual characteristics), we change the levels of the visual characteristic v_{grille} and assume no change in unobserved product characteristic $\tilde{\xi}_{jt}$ corresponding to BMW models. (Recall that $\tilde{\xi}_{jt}$ captures unobservable product characteristics *excluding* the visual characteristics). The counterfactual visual design can be specified by changing the discovered visual characteristics, specifically the characteristic v_{grille} for a change in grille. We then recompute the equilibrium prices and market shares for the entire market in 2017 using the demand estimates from model (with visual characteristics). We report BMW's counterfactual market shares in 2017 in column (4) of Table 9. Note that, with a baseline model, when the BMW changes its "design language", there is a change in the unobserved product characteristic ξ_{jt} corresponding to BMW models, which implicitly also captures visual characteristics (and any other unobservables at the product market level). However, it is unknown what part of ξ_{jt} represents each of the discovered visual features, and therefore it is unclear how to adjust the unobserved product characteristic ξ_{jt} to account for the change in BMW's visual characteristics. Finally, we report the ratio of BMW's market shares with the counterfactual design change to BMW's market shares with no design change in column (5) of Table 9. We note that our model predicts an increase in the market share of BMW models with this new design change. This finding is directionally and qualitatively consistent with the actual increase in sales for Model Year 2018 (by 26%) for the BMW X-Series, where the change is grille was most pronounced.

Table 9 Counterfactual Design Change for BMW automobiles

		Predicted Share in 2017			
	Original Data	Baseline	Closed Loop (No Design Change)	Closed Loop (Counterfactual Design Change)	Ratio of (4) to (3)
Make-Model	(1)	(2)	(3)	(4)	(5)
1-Series	9.93e-04	9.87e-04	9.21e-04	1.03e-03	1.12
2-Series	5.67e-04	5.51e-04	5.10e-04	5.68e-04	1.11
3-Series	1.13e-03	1.04e-03	9.42e-04	1.05e-03	1.11
4-Series	6.26e-04	5.50e-04	4.87e-04	5.42e-04	1.12
5-Series	5.61e-04	4.62e-04	4.03e-04	4.47e-04	1.11
6-Series	4.76e-05	3.25e-05	2.89e-05	3.20e-05	1.11
7-Series	5.51e-05	3.51e-05	3.33e-05	3.68e-05	1.11
M2	1.03e-04	8.56e-05	7.39e-05	8.21e-05	1.11
M3	3.50e-05	2.34e-05	2.20e-05	2.44e-05	1.11
M4	7.07e-05	4.76e-05	4.54e-05	5.03e-05	1.11
X1	4.90e-04	4.52e-04	4.07e-04	4.53e-04	1.11
X3	2.91e-04	2.38e-04	2.06e-04	2.29e-04	1.11
X4	7.85e-05	6.05e-05	5.18e-05	5.75e-05	1.11
X5	2.33e-04	1.56e-04	1.38e-04	1.53e-04	1.11
X6	4.91e-05	3.20e-05	3.05e-05	3.38e-05	1.11

6. Discussion & Conclusion

Visual product characteristics are important to quantify in a semantically-interpretable manner and incorporate into traditional demand models used in marketing and economics. However, this is a challenging task especially when the model is required to discover visual characteristics automatically. We propose a model that combines a disentanglement VAE-based deep learning model that discovers visual characteristics with a BLP demand model to evaluate the economic impact of visual characteristics. Methodologically we compare an open loop approach where the demand model just incorporates the visual characteristics similar to structured characteristics, to a closed loop approach where the results of the demand model can help train the disentanglement model to obtain better visual representations.

We obtain 3 distinct visual characteristics (body shape, grille and boxiness) obtained from the closed loop integrated model. We also find that after incorporating visual characteristics, our estimates of elasticity are more negative relative to the baseline without visual characteristics. We also conduct a counterfactual to examine the market impact of a visual design change by an auto manufacturer (BMW), and find that market shares in the counterfactual (in the new equilibrium) are higher for BMW, implying that the design change had a positive impact.

Our research has certain limitations. First, although we use a standard demand model like BLP, we do not have access to any individual consumer-level microdata. Thus, we are not able to evaluate an approach like Berry et al. (2004). Second, the product images we use in the model are currently limited to the front view of the automobile. It is likely that the side and rear view might matter as well to buyers, and could be potentially useful to study. Third, similar to most demand models, we assume exogeneity of product characteristics, which include visual characteristics as well.

The framework presented here can be readily extended along several aspects. First, our model can be applied across a broader class of demand models than considered in this paper. In marketing applications, consumer-level data is commonplace and it would be useful to extend our approach with such data. Second, it would be interesting to explore where consumers form 'visual consideration sets' in addition to consideration sets that are usually based on structured characteristics Mehta et al. (2003). Finally, it would be helpful to study the impact of visual design across other product categories.

References

Aaker JL (1997) Dimensions of brand personality. *Journal of Marketing Rresearch* 34(3):347–356.

Achille A, Soatto S (2018) Emergence of invariance and disentanglement in deep representations. *Journal of Machine Learning Research* 19(1):1947–1980.

Bangle C (2001) The ultimate creativity machine. how bmw turns art into profit. *Harvard Business Review* 79(1):47–55.

Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35(8):1798–1828.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

Berry S, Levinsohn J, Pakes A (1999) Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review* 89(3):400–430.

Berry S, Levinsohn J, Pakes A (2004) Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy* 112(1):68–105.

Berry ST (1994) Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* 242–262.

Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational inference: A review for statisticians. *Journal of the American Statistical Association* 112(518):859–877.

Bloch PH (1995) Seeking the ideal form: Product design and consumer response. *Journal of marketing* 59(3):16–29.

Burgess C, Higgins I, Pal A, Matthey L, Watters N, Desjardins G, Lerchner A (2017) Understanding disentangling in β-vae. *Workshop on Learning Disentangled Representations at the 31st Conference on Neural Information Processing Systems*.

Burnap A, Hauser JR, Timoshenko A (2019) Design and Evaluation of Product Aesthetics: A Human-Machine Hybrid Approach. *Available at SSRN 3421771* .

Chen RTQ, Li X, Grosse RB, Duvenaud DK (2018) Isolating sources of disentanglement in variational autoencoders. *Advances in Neural Information Processing Systems*, 2615–2625.

Conlon C, Gortmaker J (2020) Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics* 51(4):1108–1161.

Creusen ME, Schoormans JP (2005) The different roles of product appearance in consumer choice. *Journal of product innovation management* 22(1):63–81.

Dew R, Ansari A, Toubia O (2021) Letting logos speak: Leveraging multiview representation learning for data-driven branding and logo design. *Marketing Science* .

Duan S, Matthey L, Saraiva A, Watters N, Burgess C, Lerchner A, Higgins I (2020) Unsupervised model selection for variational disentangled representation learning. *International Conference on Learning Representations*.

Eastwood C, Williams CK (2018) A framework for the quantitative evaluation of disentangled representations. *International Conference on Learning Representations*.

Gandhi A, Houde JF (2019) Measuring substitution patterns in differentiated-products industries. Technical report, National Bureau of Economic Research.

Grieco PL, Murry C, Yurukoglu A (2021) The evolution of market power in the us auto industry. Technical report, National Bureau of Economic Research.

Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, Lerchner A (2017) beta-vae: Learning basic visual concepts with a constrained variational framework. *International Conference on Learning Representations*.

Hoffman MD, Johnson MJ (2016) Elbo surgery: yet another way to carve up the variational evidence lower bound. *Workshop in Advances in Approximate Bayesian Inference, Neural Information Processing Systems*.

Homburg C, Schwemmle M, Kuehnl C (2015) New product design: Concept, measurement, and consequences. *Journal of marketing* 79(3):41–56.

Huang J, Chen B, Luo L, Yue S, Ounis I (2021) Dvm-car: A large-scale automotive dataset for visual marketing research and applications. *arXiv preprint arXiv:2109.00881* .

Jindal RP, Sarangee KR, Echambadi R, Lee S (2016) Designed to succeed: Dimensions of product design and their impact on market share. *Journal of Marketing* 80(4):72–89.

Kim H, Mnih A (2018) Disentangling by factorising. *International Conference on Machine Learning*, 2649–2658.

Kingma DP, Welling M (2014) Auto-encoding variational bayes. *International Conference on Learning Representations*.

Lancaster KJ (1966) A new approach to consumer theory. *Journal of Political Economy* 74(2):132–157.

Liu L, Dzyabura D, Mizik N (2020) Visual listening in: Extracting brand image portrayed on social media. *Marketing Science* 39(4):669–686.

Locatello F, Bauer S, Lučić M, Rätsch G, Gelly S, Schölkopf B, Bachem OF (2019) Challenging common assumptions in the unsupervised learning of disentangled representations. *International Conference on Machine Learning*, 4114–4124.

Locatello F, Tschannen M, Bauer S, Rätsch G, Schölkopf B, Bachem O (2020) Disentangling factors of variations using few labels. *International Conference on Learning Representations*.

Malik N, Singh PV, Srinivasan K (2019) A dynamic analysis of beauty premium. *Available at SSRN 3208162* .

Mehta N, Rajiv S, Srinivasan K (2003) Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science* 22(1):58–84.

Nevo A (2000) Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics* 395–421.

Norman D (2013) *The design of everyday things: Revised and expanded edition* (Basic books).

Petrin A (2002) Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy* 110(4):705–729.

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy* 82(1):34–55.

Sisodia A, Burnap A, Kumar V (2022) Automatically discovering product characteristics impacting consumer preferences. *Available at SSRN 4151019* .

Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. *Marketing Science* 38(1):1–20.

Veryzer Jr RW (1993) Aesthetic response and the influence of design principles on product preferences. *Advances in Consumer research* 20(1).

Watanabe S (1960) Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development* 4(1):66–82.

Zhang M, Luo L (2018) Can consumer-posted photos serve as a leading indicator of restaurant survival? evidence from yelp. *Available at SSRN 3108288* .

Zhang S, Lee D, Singh PV, Srinivasan K (2021) What makes a good image? airbnb demand analytics leveraging interpretable image features. *Management Science* .