Market Structure Mapping with Interpretable Visual Characteristics

Ankit Sisodia
Purdue University, asisodia@purdue.edu

Vineet Kumar
Yale School of Management, vineet.kumar@yale.edu

June 2025

We develop a novel approach to market structure mapping that integrates visual product characteristics from image data alongside functional characteristics from structured data. Using disentangled representation learning, we extract five interpretable visual characteristics from vehicle front-view designs in the UK automobile market (2008-2017): body shape, color, grille height, boxiness, and grille width. Our analysis reveals that some visual characteristics (like body shape) are tightly constrained by functional requirements, while others (particularly grille dimensions) remain largely unconstrained by function, challenging the simplistic *form follows function* maxim. Our market structure maps demonstrate that product segments show substantially greater differentiation when incorporating visual characteristics, brands pursue distinct visual strategies, and crossover vehicles use visual design to bridge segment boundaries between vehicle classes despite functional similarities to conventional cars. Analysis of consumer comparison searches shows that pairs of vehicles with similar visual characteristics are more frequently compared, independent of their functional similarities. By capturing how visual design shapes competitive positioning, our approach provides a more comprehensive view of market structure than functional attributes alone.

Key words: market structure, visual analytics, deep learning, disentanglement

1. Notes

1. Introduction
2. Literature Review
3. Flow Chart of the Paper **work on this**
4. Disentanglement (finding the characteristics, validating them using survey respondents , showing summary stats, showing correlations) -¿ disentanglement is able to pick up things that human reviewers care about (from professional reviewers results in LLM)
5. Methodological Contribution: Disentanglement allows us to disentangle and isolate each factor of variation. One of them happens to be color. It is a factor that needs to be removed in order to get accurate market structure mapping. We need to write this argument in a more convincing way.
6. Motivate Google Trends (we need some measure of cross substitution. what are some potential measures - modelbased and databased. example of model based is BLP. example of data based is 2nd choice data. other examples are co-search / co-mention / consideration set / firms talking about rivals in ads etc. we need to cite all people doing this. we say that we want to make minimal assumptions and not have the model inform us. we need to see a review paper on the value of search data, maybe by Bart Bronnenberg. **work on this**
7. Google Trends Results - Predicting Top 2 rivals and Predicting the Entire Substitution Matrix (All models + top selling models)
8. Google Trends Insights - See how crossovers evolve over a long period of time on Google Trends (Cars or Car based or Truck based SUVs -¿ Co-search pattern over a longer period of time) : when did Google Trends start ; when did crossovers begin to take off in the UK
9. Google Trends Insights - See how people mostly search for cars within segment and not across segments: -¿ Descriptive - Product Segment (along with boxplots)
10. Case Study - Crossovers (along with boxplots)

You need some measure of cross substitution. What are some potential measures? Classify based on data and model. Model-based ones are Pradeep's paper, BLP etc. Data based ones are 2nd choice data / co-search / co-mention. Cite all the people. Also, cite people who has used it before. Bart Bronnenberg - value of search data (some survey paper). want to make minimal assumption and not have model inform us. consideration set, where do they talk about rivals.

1. Extracting visual characteristics
   - Method: Disentanglement
   - Finding: 5 visual characteristics
2. Validating visual characteristics are human-interpretable
   - Method: Open-ended surveys
   - Finding: High level of agreement among respondents on interpretability and quantification (COMMENT: maybe repeat interpretability like the watch paper survey)
3. Form & Physical Dimensions Relationship
   - Method: Correlation between visual (form) characteristics and physical dimensions
   - Finding: Bodyshape is correlated with wheelbase, length, width in a way that makes sense for one end of bodyshape to be hatchback-like and the other end to be sedan-like; Boxiness is correlated with height; heigh-width ratio showing that it captures tallness/cabin uprightness and is not related to length/wheelbase and so is not related to overall vehicle size; Grille height/width are almost uncorrelated with physical dimensions
4. Form & Function Relationship
   - Method: Correlation between visual (form) characteristics and functional characteristics
   - Finding: Form follows function in case of bodyshape and boxiness; Form does not follow function for grille height/width
5. Product Segmentation Analysis
   - Method: Classification accuracy and convex hull overlap analysis across segments
   - Finding: Combining visual and structured characteristics improves segment classification accuracy from 65-67% to 72%. Segments show only 5.6% overlap when both characteristic types are used, versus 18.2% (structured only) or 26.8% (visual only)
6. Reason for Separation of Product Segments
   - Method: Box plots of different functional and form char for different segments
   - Finding: Boxiness separates sedans (C,D,E) from suv (J)
7. Brand Strategy Analysis
   - Method: Area (Convex hull) share calculations in structured vs. visual space

- Finding: Brands employ distinct visual strategies - Audi occupies large structured space but small visual space, emphasizing consistent design language; Jeep shows opposite pattern, using visual variety for differentiation

8. Reason for Audi having small visual share v Jeep having large visual share
   - Method: Box plots of different functional and form char for different makes
   - Finding: Audi's grilles are very similar and thats why it occupes a very small visual space
9. Crossover Vehicle Analysis
   - Method: MDS mapping and Google Trends co-search analysis
   - Finding: Crossovers share structured characteristics with compact cars but visual characteristics with SUVs. Consumer search patterns show crossovers are compared more with SUVs (score: 141) than with compact cars (score: 45) [When did crossovers beging in UK, and see if we can show that using Google Trends?]
10. Reason for Compact Cars being different from Compact Crossovers in the visual space
    - Method: Box plots of different functional and form char for compact cars, crossover SUVs, and SUVs
    - Finding: Compact cars are different than crossover SUVs and SUVs in boxiness
11. Consumer Comparison Behavior
    - Method: Predicting top-2 comparison partners using Google Trends co-search data
    - Finding: Visual characteristics significantly improve prediction of which cars consumers compare.

2. Introduction

Market structure mapping represents a cornerstone methodology in competitive marketing strategy, offering critical insights that guide managerial decision-making across multiple domains. Firms leverage these maps to understand relative competitive positioning, which subsequently informs brand positioning, new product development, and strategic decisions regarding advertising and pricing (Rao et al. 1986, Yang et al. 2022). Effective competitive analysis through market structure mapping enables firms to identify market boundaries, recognize closest competitors, discover unoccupied market positions for innovation, and evaluate potential partnership or acquisition targets (DeSarbo et al. 2006).

Constructing informative market structure maps requires careful consideration of two foundational elements. First, analyses must achieve comprehensive market coverage, capturing the full spectrum of products competing within a defined market space. Second, they must adhere to the principle of substitutability, focusing exclusively on products that consumers genuinely consider as alternatives to one another. Traditional approaches have typically employed multidimensional scaling (MDS) or related techniques to represent product similarities, drawing on either decompositional methods that begin with aggregate relationships such as cross-elasticities, or compositional methods that start with disaggregate product characteristics (Shugan 2014).

Despite significant methodological advances in market structure analysis, a critical dimension of product differentiation remains underexplored: visual product characteristics. While functional attributes like performance metrics and technical specifications have been thoroughly incorporated into competitive analyses, the visual elements that significantly influence consumer perception and choice behavior have received comparatively little attention in formal market structure mapping. This oversight is particularly notable given substantial evidence that visual form not just enables product differentiation (Vlasic 2011), but that visual form is often a key driver of consumer decision making—accounting for up to 60% of consumer purchase decisions in some contexts (Kreuzbauer and Malter 2005). As the JD Power Avoider Study (2015) found:

> “Exterior look/design is the top reason shoppers avoid a particular vehicle (30%), followed by cost (17%).”

This industry data underscores how product appearance powerfully affects consumer attention, perceived quality, brand identity, and ultimately purchase decisions, yet these influences remain largely unaddressed in competitive structure analyses (Bloch 1995, Creusen and Schoormans 2005). This phenomenon is readily apparent in product comparisons. As Magnolfi et al. (2025) note:

> “The 2020 Toyota Camry and the 2020 MINI Clubman are very similar cars based on horsepower, fuel efficiency, passenger volume, and curb weight; but we suspect consumers would not identify the two cars as being near each other in product space.”

Recent computer science approaches have attempted to address this gap using image embedding techniques for product visuals. However, these methods exhibit critical limitations for market structure applications. First, they typically generate non-interpretable visual dimensions, producing latent spaces where axes cannot be directly interpreted as explicit visual attributes that managers can understand and act upon (Han et al. 2021, McAuley et al. 2015). Second, many approaches violate the substitutability principle by including non-competing products from multiple categories in the same analysis, resulting in distances that lack meaningful market interpretation (Li et al. 2024). Third, these methods offer limited control over the inclusion or exclusion of specific visual dimensions that may distort true market relationships, such as when identical product models in different colors appear distant despite being perfect functional substitutes.[1]

Effective integration of visual characteristics into market structure analysis requires addressing two additional critical requirements beyond current approaches. Visual dimensions must be human-interpretable, with clear semantic meaning that translates to actionable insights for marketing decision-makers. The methodology must also provide selective control over which visual characteristics are included in the analysis, allowing researchers to exclude characteristics that might distort true competitive relationships while retaining those that meaningfully differentiate products.

In this paper, we develop a novel approach to visualize market structure maps using both functional (structured) and aesthetic (visual) characteristics of products. We employ disentangled representation learning to automatically extract and quantify human-interpretable visual characteristics without manual coding or intervention. By combining these visual characteristics with traditional functional attributes, we create integrated market structure maps that provide a more comprehensive view of competitive positioning than either characteristic set alone could provide.

We apply our methodology to the United Kingdom automobile market from 2008 to 2017, analyzing over 2,400 observations spanning 379 distinct models. The automotive industry provides an ideal empirical setting for our approach, as vehicles compete simultaneously on both technical specifications and visual design elements, with substantial evidence that both dimensions significantly influence consumer choice. Through our disentanglement approach, we identify five interpretable visual characteristics from automobile front-view

[1] A simple sanity check shows that a standard image embedding places a red Ferrari almost as close to a red apple (correlation 0.655) as to a red Toyota (0.727), indicating that color can dominate semantic content (see Appendix C). This motivates our disentangled representation with interpretable visual characteristics and selective control over color, so that market structure reflects true design similarity rather than superficial visual cues.

designs: body shape, color, grille height, boxiness, and grille width. We confirm the validity and interpretability of these characteristics through human subject validation.

Using these characteristics, we construct market structure maps for the 2013 UK automobile market using multidimensional scaling. We analyze the correlation between functional and visual characteristics, finding that they capture largely orthogonal aspects of product differentiation, with limited correlation between most visual and functional attributes. This finding underscores the importance of incorporating both dimensions into comprehensive market structure analyses. Through conditional density analyses, we investigate the relationship between form and function, identifying which visual characteristics are tightly constrained by functional requirements and which allow greater design freedom. This analysis reveals a selective version of the “form follows function” maxim—certain visual characteristics (e.g., body shape) are strongly linked to functional attributes, while others (e.g., grille geometry) remain largely discretionary and therefore valuable for stylistic differentiation.

Our analysis reveals four key insights about competitive market structure. First, product segments show substantially greater differentiation when visual characteristics are included alongside functional attributes, challenging the conventional view that product categories are defined primarily by functional similarities. Second, brands employ distinct visual strategies, with some prioritizing consistent visual identity while allowing functional variation across models, and others pursuing the opposite approach. Third, we demonstrate how crossover vehicles use visual design cues to bridge traditional segment boundaries despite functional similarities to conventional cars. Finally, our analysis of consumer search behavior confirms that both visual and functional differences influence comparison shopping patterns.

The remainder of this paper is organized as follows. Section 3 reviews the literature on market structure mapping and visual product characteristics, positioning our contribution relative to existing approaches. Section 4 details our methodology for discovering visual characteristics using disentangled representation learning and constructing integrated market structure maps. Section 5 describes our empirical setting and presents the results of our analysis, including detailed discussion of the four key insights outlined above.

3. Literature Review

Market structure mapping is one of the primary and commonly used methods in competitive marketing strategy (Rao et al. 1986). It provides a visual representation of the competitive landscape, illustrating the relationships between different companies and their products. According to DeSarbo et al. (1993), competitive market structure represents how products and brands are perceived as substitutes by consumers, helping to identify competition levels within well-defined product markets.

Firms use it to understand the relative positions of their products with respect to their rivals in order to inform brand positioning; new product development; and product, advertising, and pricing strategies (Urban et al. 1984, DeSarbo et al. 1993, Bergen and Peteraf 2002, Lattin et al. 2003, DeSarbo et al. 2006). These maps serve several critical functions in marketing strategy. First, they help both researchers and practitioners understand market boundaries, where products within a market are likely to be clustered. Yang et al. (2022) conceptualize this as product-markets with boundaries that define competing brands, which may overlap across different markets. Second, maps can help identify the closest competitors to a firm, enabling effective competitive positioning DeSarbo et al. (2006). Third, they can point to market gaps where products are not present, suggesting where firms might focus innovation efforts. Fourth, they can help suggest potential partnerships or acquisition targets for firms.

It is worth differentiating between brand (or product) positioning maps and brand perceptual maps based on the data used. Positioning maps based on product characteristics show how a product (e.g., Toyota Corolla) is positioned relative to competitors in terms of objective attributes like reliability and fuel efficiency (Hauser and Shugan 1983, Blattberg and Wisniewski 1989). These maps typically use product characteristics as input. In contrast, perceptual maps use consumer perceptions (e.g., subjective judgments of luxury or value) as input, reflecting psychological associations rather than objective attributes (Green and Rao 1969, Hauser and Koppelman 1979). These maps are constructed from survey data, similarity ratings, or preference rankings to understand subjective evaluations, comparison dimensions, and consideration set formation.

The classic mapping approach is typically done using multidimensional scaling (MDS), which represents the distances between products (or brands) in a two-dimensional space (DeSarbo et al. 1993). The input to commonly used methods to map the market structure is a similarity matrix between all pairs of products. This approach emerged in the late 1960s, when researchers began considering market structures consisting of products existing as points in a multidimensional space that exhibited the properties of a geometric space. In these representations, proximity in the space represented competitive substitutability. Market structure methods in marketing can be broadly classified into decompositional and compositional approaches. Decompositional methods start with aggregate relationships such as cross-elasticities or similarities between pairs of products. Compositional methods, conversely, start with disaggregate data such as product characteristics and summarize that data to form a market structure. Despite their different starting points, both approaches are fundamentally data reduction tools that aim to represent complex market relationships in simplified forms (Shugan 2014).

Market structure methods in marketing have used a variety of data sources, such as panel-level scanner data (Erdem 1996), consumer search data (Ringel and Skiera 2016, Kim et al. 2011), online product reviews (Lee and Bradlow 2011, Tirunillai and Tellis 2014), social media engagement data (Liu et al. 2020a, Yang et al. 2022), co-occurrence of products in shopping baskets (Gabel et al. 2019) as well as co-occurrence of products in product reviews (Netzer et al. 2012). Reflecting the compositional method, our focus is on positioning maps derived from product characteristics, specifically a method allowing us to map both functional and visual characteristics. We later use consumer comparison data to validate and understand how competitive distance impacts consumer search.

Functional characteristics are relevant to the product's ability to impact its function or performance. Garvin (1987) defines them as the primary operating characteristics of a product, including its performance, features, and reliability, that enable it to deliver its intended benefits to users. For instance, the number of bedrooms in a house, the image resolution of a camera, or the fuel efficiency of a car would represent functional characteristics. Product design or visual form has been recognized as an opportunity for differential advantage in the marketplace. Aesthetic characteristics are defined by Bloch (1995) as those characteristics of a product that describe its appearance, style, and design, including its visual, auditory, and tactile attributes that contribute to its overall aesthetic appeal. The appearance of a product influences consumer product choice in several ways, including drawing attention, standing out from competitive offerings, perceived quality, experience, and brand identity (Aaker 1997, Desmet and Hekkert 2007). Creusen and Schoormans

(2005) provide a framework for understanding the impact of aesthetic characteristics on consumer behavior.

Whereas functional characteristics have been commonly used in market structure mapping, visual characteristics have not been used to develop market structure maps in a principled way. While some computer science literature uses embedding techniques for product images, these approaches fail to meet several critical requirements for effective market structure mapping. First, these embeddings often lack interpretability of visual dimensions. For example, work by Han et al. (2021) and Han and Lee (2025) uses learned embeddings where visual dimensions cannot be directly interpreted as explicit visual attributes. While these approaches attempt post-hoc interpretation by examining corners of their 2D map, this method of interpretation is fundamentally limited and subjective. Such post-hoc analysis fails to provide precise control over specific visual characteristics and doesn't guarantee that the identified dimensions represent independent visual characteristics rather than complex combinations of multiple characteristics. In contrast, our approach explicitly identifies interpretable visual characteristics, ensuring that each characteristic has clear semantic meaning that can be independently manipulated. Second, many existing approaches (McAuley et al. 2015, Zhang et al. 2020, Avas et al. 2024, Li et al. 2024) violate the substitutability principle by including products from multiple categories (e.g., tops, bottoms, shoes) that don't compete with each other in the same market. When these methods compute Euclidean distances between embeddings across non-competing categories and project them into a 2D space, the resulting axes lack meaningful market interpretation. The distances between, for instance, a shoe and a jacket have no substantive meaning in terms of consumer substitution patterns or competitive positioning, rendering such maps ineffective for strategic marketing decisions. Third, current methods provide limited control over inclusion or exclusion of specific visual dimensions. For instance, color might remain implicitly embedded in the representation, causing two identical car models in different colors to appear distant on the map despite being perfect substitutes from a market perspective. Our approach addresses these limitations by: (1) ensuring complete coverage of products within a defined market, (2) maintaining focus on substitutable products, (3) providing human-interpretable visual characteristics, and (4) allowing selective exclusion of certain characteristics (such as color) that might distort true market relationships. The primary challenge in quantifying visual characteristics in a human-interpretable manner and subsequently obtaining market structure maps from these characteristics marks the contribution of the present paper.

To address this gap, we develop an approach to visualize market structure maps using both functional and aesthetic (visual) characteristics of products. We employ representation learning, a machine learning sub-field that theorizes high-dimensional data is generated from low-dimensional factors. According to Bengio et al. (2013), the goal is to learn data representations that simplify the extraction of useful information for building predictive models. This paper focuses on disentangled representation learning, which aims to isolate meaningful factors of variation in data (Bengio et al. 2013).

We use this method to automatically learn and quantify human-interpretable visual characteristics of products without human labeling or intervention. For example, just as disentangled representation learning can separate shapes, sizes, colors, and positions in the dSprites dataset Higgins et al. (2017), we aim to isolate meaningful visual characteristics in product images. 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). To address this challenge, Locatello et al. (2020) showed that a small number of labeled examples with even potentially imprecise and incomplete labels is sufficient to perform model selection to learn disentangled representations. However, since our aim is to identify visual characteristics automatically without specifying them in advance, we use the approach outlined by Sisodia et al. (2024) to use the functional characteristics (also referred to as structured product characteristics) as supervisory signal for overcoming this well-known challenge in the deep learning literature.

The presence of underlying visual characteristics from our approach would inform managers on the reason why two products are located close together or further apart on the market structure map. By identifying the key visual characteristics, our approach has the potential to provide insights for managers looking to make informed decisions about product design in order to differentiate their products or reposition them in the market.

4. Methodology

Our method for discovering and mapping the visual product characteristics consists of two main components. First, we employ a disentanglement-based approach using Variational Autoencoders (VAEs) to obtain visual product characteristics that are independent and human-interpretable. However, disentanglement methods typically require ground truth visual characteristics as supervisory signals, which are not available in our case since we aim to discover these characteristics. To address this challenge, we use structured product characteristics as supervisory signals. Next, we combine the structured characteristics and the discovered visual characteristics to create competitive market structure maps using Multidimensional Scaling (MDS). In the following subsections, we discuss each of these components in detail, starting with the disentanglement approach using VAEs (Section 4.1), followed by the market structure mapping methodology (Section 4.2).

4.1. Disentanglement with Variational Autoencoder

Our approach to discovering visual product characteristics follows the approach used by (Sisodia et al. 2024). It employs the use of Variational Autoencoders (VAEs) (Kingma and Welling 2014), a class of deep generative models that learn to encode input data into a latent space, and simultaneously enable the generation of new data samples from this latent space. In line with previous research (Higgins et al. 2017, Burgess et al. 2017, Chen et al. 2018, Kim and Mnih 2018), we utilize these VAEs specifically for disentangled representation learning (Bengio et al. 2013), allowing us to identify statistically independent and semantically meaningful visual factors that vary across products in our dataset.

We consider a dataset of product images, each of which we assume is generated by an underlying distribution parameterized by visual characteristics. Our goal is to learn a low-dimensional representation of these visual characteristics that captures the most salient factors of variation in the product images. To achieve this, we utilize a VAE, consisting of an encoder network $q_\phi(\mathbf{z}|\mathbf{x})$, which compresses each high-dimensional product image $\mathbf{x}$ into a lower-dimensional latent space of visual characteristics $\mathbf{z}$, and a decoder network $p_\theta(\mathbf{x}|\mathbf{z})$, which reconstructs images from these latent representations. Both the encoder and decoder are deep neural networks, parameterized by ϕ and θ, respectively.

The VAE framework assumes a generative model where the latent visual characteristics $\mathbf{z}$ are first sampled from a prior distribution $p(\mathbf{z})$, set to an isotropic unit Gaussian $\mathcal{N}(0, \mathbf{I})$. Subsequently, product images are generated through the decoder distribution $p_\theta(\mathbf{x}|\mathbf{z})$. During training, the encoder network approximates the true posterior distribution $p(\mathbf{z}|\mathbf{x})$ with a variational distribution $q_\phi(\mathbf{z}|\mathbf{x})$, typically modeled as a multivariate Gaussian with diagonal covariance, i.e., $\log q_\phi(\mathbf{z}|\mathbf{x}) = \log \mathcal{N}(\mathbf{z}; \boldsymbol{\mu_d}, \boldsymbol{\sigma_d}^2\mathbf{I})$, where $\boldsymbol{\mu_d}$ and $\boldsymbol{\sigma_d}$ are the mean and standard deviation outputs of the encoder network.

While various disentanglement methods built on the VAE framework—such as β-VAE (Higgins et al. 2017, Burgess et al. 2017), FactorVAE (Kim and Mnih 2018), and β-TCVAE (Chen et al. 2018)—have shown promising results, achieving true disentanglement in these representations faces fundamental theoretical challenges. Specifically, Locatello's theorem (Locatello et al. 2019) showed that unsupervised disentanglement is fundamentally limited without additional structure or assumptions—commonly referred to as inductive biases. These biases can be implicit or explicit and are essential for learning meaningful and interpretable representations from limited data. Some examples of inductive biases include architectural choices (e.g., convolutional neural networks for image data), prior distributions for latent variables, regularization techniques, and data augmentation. Consequently, recent efforts in the deep learning literature have focused on improving disentanglement methods by utilizing benchmark datasets with known ground truth labels corresponding to each visual characteristic (Locatello et al. 2020). However, the very visual characteristics we aim to discover are precisely these ground truth labels.

Instead, we follow Sisodia et al. (2024) in using structured product characteristic (such as brand or price) as supervisory signals to address Locatello's theorem (Locatello et al. 2019). Structured product characteristics, such as brand, material, performance attributes, and price, can be informative for guiding the learning of disentangled representations due to their potential correlation with visual characteristics. For example, a product's brand can significantly influence its visual appearance. A luxury brand like Louis Vuitton may be associated with specific visual design elements that set them apart from other brands. Similarly, the price point of a product can often be reflected in its visual design, with higher-priced items frequently exhibiting more refined or elaborate visual features compared to lower-priced alternatives. The supervised loss term directly quantifies the discrepancy between predicted structured characteristics derived from the latent space and their actual observed values. The machine learning literature has shown that even weak supervision with a set of signals that has some correlation to the ground truth will help obtain a disentangled representation in practice (Locatello et al. 2020).

To promote disentanglement of the learned representations, we incorporate additional regularization terms following the β-TCVAE approach (Chen et al. 2018). We minimize the objective function in Equation 1 to learn a disentangled representation of the visual characteristics present in the product image data.

$$\underbrace{\mathcal{L}(\theta,\phi;\mathbf{x},\mathbf{z})}_{\text{Disentanglement Loss}} = \underbrace{-\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]}_{\substack{\text{Reconstruction}\\ \text{Loss}}} + \underbrace{I_q(\mathbf{z},\mathbf{x})}_{\substack{\text{Mutual}\\ \text{Information}\\ \text{Loss}}} + \beta \underbrace{KL\left[q(\mathbf{z})||\prod_{j=1}^{J} q(z_j)\right]}_{\substack{\text{Total Correlation}\\ \text{Loss}}}$$

$$+ \underbrace{\sum_{j=1}^{J} KL\left[q(z_j)||p(z_j)\right]}_{\substack{\text{Dimension-Wise}\\ \text{KL Divergence Loss}}} + \delta \underbrace{P(\widehat{\mathbf{y}(\mathbf{z})},\mathbf{y})}_{\substack{\text{Supervised}\\ \text{Loss}}} \tag{1}$$

The objective function in Equation 1 includes five terms.

The first term, the reconstruction loss (-$\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]$), evaluates how well the model can regenerate the original image $\mathbf{x}$ from its latent representation or visual characteristics $\mathbf{z}$. By minimizing this term, we ensure that the latent space retains the key visual elements needed to accurately reproduce each product image. This acts as a fidelity constraint, grounding the learning process in the observable input data.

The second term is the mutual information loss ($I_q(\mathbf{z},\mathbf{x})$). It is a measure of how much information the latent visual characteristic $\mathbf{z}$ contain about the input data $\mathbf{x}$. In the context of disentanglement $I_q(\mathbf{z},\mathbf{x}) = D_{KL}(q(\mathbf{z},\mathbf{x})||q(\mathbf{z})p(\mathbf{x}))$. Higher mutual information means the latent variables capture more information about the inputs. This helps ensure the representation is meaningful and useful. If this term is completely minimized, then the latent visual characteristics would be independent of the input image $\mathbf{x}$. At the same time, if this term is maximized without constraint, then the model might simply memorize the training data without learning useful structure.

Next, the total correlation loss ($KL\left[q(\mathbf{z})||\prod_{j=1}^{J} q(z_j)\right]$) measures the dependence between the individual dimensions of the latent representation $\mathbf{z}$. By penalizing the KL divergence between the joint distribution $q(\mathbf{z})$ and the product of its marginals $\prod_{j=1}^{J} q(z_j)$, this loss term promotes statistical independence among the learned visual characteristics. This is critical for disentanglement: we want each latent variable to reflect a distinct and independent factor of variation. By encouraging the joint posterior distribution to approximate a factorized distribution, the model learns a more disentangled latent structure. The hyperparameter β controls the strength of this penalty.

A fourth term, the dimension-wise KL divergence ($\sum_{j=1}^{J} KL\left[q(z_j)||p(z_j)\right]$), ensures that each individual latent dimension remains close to its prior distribution (typically a standard

Gaussian). This regularization serves multiple purposes: it helps smooth the latent space, discourages redundant encodings, and implicitly controls model complexity by allowing the model to "turn off" unnecessary latent dimensions.

Finally, we include a supervised loss term $(P(\widehat{\mathbf{y}(\mathbf{z})}, \mathbf{y}))$ that measures the discrepancy between the predicted supervisory signals $\widehat{\mathbf{y}(\mathbf{z})}$ and the actual observed signals $\mathbf{y}$. This allows us to steer the latent space using weak but informative signals found in structured product characteristics that could be potentially correlated with underlying visual characteristics. The hyperparameter δ balances the importance of this supervised objective with the other unsupervised loss terms.

Our model architecture, detailed in Appendix A, is a modified version of the one used by Burgess et al. (2017). We adapt the architecture to work with 128×128 pixel images and incorporate the supervisory signals from our model of market equilibrium and structured product characteristics. The architecture consists of an encoder neural network, a decoder neural network, and a supervised neural network, which work together to learn disentangled visual representations and predict the supervisory signals.

In addition to the model parameters that are learned during training, we also make modeling choices in the form of hyperparameters. These hyperparameters impact the estimation process but are not directly estimated with the model. The separation between parameters and hyperparameters is common in machine learning, as hyperparameters are typically set before training and control the learning process, while parameters are learned from data during training (Goodfellow et al. 2016, Murphy 2012, Bishop 2006). Examples of hyperparameters include the learning rate, batch size, and regularization strengths like β and δ in our model. The underlying logic is that hyperparameters define the model's capacity, regularization, and optimization settings, which need to be tuned separately from the model parameters to achieve the best performance and generalization. While there are techniques for automatically searching for optimal hyperparameter values, such as grid search, random search, and Bayesian optimization (Bergstra and Bengio 2012, Snoek et al. 2012), hyperparameters are typically not learned directly during the main model training process.

In our model, we have two key hyperparameters: β and δ. The hyperparameter β controls the weight of the total correlation loss within the disentanglement loss (Chen et al. 2018). This term encourages the model to learn statistically independent latent factors. The hyperparameter δ represents the weight of the supervised loss when incorporated into the overall loss function. A higher value of δ prioritizes the model's ability to predict the vector of supervisory signals, relative to other loss terms like mutual information or reconstruction loss. However, placing too much emphasis on the supervised loss could potentially reduce the quality of disentanglement. On the other hand, setting δ to zero implies that we are not addressing the impossibility theorem (Locatello et al. 2019) and thus have no theoretical guarantees for discovering disentangled visual characteristics.

To select the optimal values for β and δ, we follow the approach proposed by Locatello et al. (2020). We perform a grid search over a range of values for these hyperparameters and select the combination that yields the lowest 10-fold cross-validated supervised loss for each vector of supervisory signals. This approach allows us to find the best balance between the disentanglement and supervised objectives, tailored to each specific set of supervisory signals. To compare the quality of disentangled representations produced by different vectors of supervisory signals, we use the Unsupervised Disentanglement Ranking (UDR) metric proposed by Duan et al. (2020). UDR is an automated method that assesses the robustness of disentangled representations to variance at different starting points without requiring access to the ground truth data generative process. It relies on the assumption that for a particular dataset, a disentangling VAE will converge on the same disentangled representation up to certain isomorphic transformations. We select the hyperparameters β and δ based on the lowest supervised loss for each vector of supervisory signals and then choose the supervisory signals that yield the highest UDR score. The details of the UDR algorithm are provided in Appendix D.

4.2. Market Structure Mapping

Market structure maps are graphical representations of the positioning of products within a given market. These maps serve as strategic tools for businesses to better understand the competitive landscape and inform marketing, product development, and overall business strategy. By mapping the market, companies can identify gaps, potential opportunities for new products, and the relative positioning of their competitors.

For our market structure mapping, we employ Multidimensional Scaling (MDS) over alternative dimensionality reduction methods for several reasons:

1. Unlike methods such as PCA which preserve variance, MDS directly preserves distances, which aligns with the concept of competitive substitutability in market structure analysis (Cox and Cox 2000).

2. While we implement classical MDS (which is equivalent to PCA when using Euclidean distances), the framework allows extension to non-metric MDS if needed, providing flexibility for non-linear relationships (Borg and Groenen 2007, Kruskal 1964).
3. MDS directly optimizes the stress function to ensure the low-dimensional representation best preserves the original distances. This is in contrast to methods like t-SNE (Van der Maaten and Hinton 2008) or UMAP (McInnes et al. 2018), which have gained popularity in machine learning applications. While these methods are effective for visualizing complex, non-linear relationships and identifying clusters in high-dimensional data, they intentionally distort global relationships to emphasize local structure (Wattenberg et al. 2016). This distortion makes them less suitable for market structure analysis where preserving global competitive distances is critical. Additionally, the non-deterministic nature of t-SNE and its sensitivity to hyperparameters can lead to inconsistent visualizations (Van Der Maaten 2014), which is problematic for strategic decision-making based on market structure maps.
4. The consistent interpretation of distances in the MDS solution allows for direct comparison of competitive intensity between different product pairs, which is essential for market structure analysis (DeSarbo et al. 2006).

MDS finds a lower-dimensional representation of products such that the distances between them in the lower-dimensional space closely match their original dissimilarities. This is achieved by minimizing a stress function (in Equation 2), which measures the discrepancy between the original dissimilarities and the distances in the lower-dimensional space.

$$\text{Stress} = \sqrt{\frac{\sum_{i,j}(d_{ij} - \delta_{ij})^2}{\sum_{i,j} d_{ij}^2}} \tag{2}$$

where d_{ij} represents the Euclidean distance between products i and j in the low-dimensional space, and δ_{ij} represents their original dissimilarity in the characteristic space.

To apply MDS, we construct a dissimilarity matrix of pairwise distances between products based on their characteristics. We consider three different sets of product characteristics: functional characteristics (structured product characteristics) alone, visual product characteristics discovered through our disentanglement approach, and a combination of both. For each characteristic set, we calculate the dissimilarity matrix using the Euclidean distance, where $\delta_{ij} = \sqrt{\sum_{k=1}^{K}(x_{ik} - x_{jk})^2}$ and x_{ik} represents the k-th normalized characteristic of product i. We normalize each characteristic by subtracting its mean value across all products and dividing by its standard deviation. This normalization ensures all characteristics contribute equally to the dissimilarity measure regardless of their original scales.

The resulting market structure map positions products in a two-dimensional space where proximity indicates similarity in terms of the considered characteristics. By comparing maps generated from different characteristic sets, we can isolate the impact of visual characteristics on market structure beyond what is captured by functional characteristics alone.

5. Empirical Setting and Results

In this section, we present the data and the main findings of our study on leveraging visual product characteristics in market structure analysis. We focus on the automobile industry in the United Kingdom (UK), which is well-suited for this analysis given the importance of visual design in consumer purchase decisions and the competitive nature of the market. We begin by describing our dataset, which combines information on automobile characteristics and images. Next, we discuss the visual characteristics discovered through our disentangled representation learning approach. To validate the interpretability and quantification of these visual characteristics, we present the results of human subject surveys. We then construct market structure maps using multidimensional scaling (MDS) based on structured characteristics alone, visual characteristics alone, and a combination of both. These market structure maps provide insights into the differentiation between product-segments, how visual characteristics shape brand identity and within-brand differentiation, and the role of visual characteristics in crossovers making the leap from cars to SUVs. Overall, our results highlight the importance of considering interpretable visual characteristics alongside traditional structured attributes when examining competitive dynamics in the automobile industry.

5.1. Data

We compiled a data set covering 2008 through 2017 consisting of automobile characteristics, quantity sold and their images from the United Kingdom (UK). 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 up to 70% of consumers identify and judge automobiles by the appearance of headlights and grille located on the face of the automobile.[2]

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

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 average of product characteristic across trims. We collected manufacturer suggested retail prices (MSRP), and characteristics of all automobiles sold in the UK from 2008-2017 from Parker's. The price variable is the list price (in £1000's) for the entry-level trim. Prices in all years are deflated to 2015 UK using the consumer price index. We have product characteristics for weight, horsepower, length, width, height, wheelbase and miles per gallon. We supplemented the Parker's information with additional information on the make's country of origin. There are 2439 observations in our sample and a total of 379 distinct models spanning the years 2008 to 2017. In Table 1, we display summary statistics for the products at the make-model-year level for the 2013 market. The variables include quantity (in units of 1000), price (in £000 units), tens of miles per gallon (MPG), horsepower (in HP), weight (in 10 lbs.), the ratio of horsepower to weight (in HP per 10 lbs.), length (in 1000 inches), width (in 1000 inches), height (in 1000 inches), wheelbase (in 1000 inches), space (measured as length times width), as well as dummies for country-of-origin of the make, and dummies for segment membership of the make-model.

For our subsequent analysis, Segment A (Minicars) and B (Subcompact) are combined as they represent small-format vehicles designed for urban mobility, sharing core attributes such as compactness, affordability, and ease of use—making them substitutable in most consumer decisions. Segments C (Compact), D (Mid-Size), and E (Mid-Size Luxury) are grouped together as they form a coherent progression of traditional passenger cars that differ primarily in size and price point, but share common body styles and usage patterns. Segment J (SUV) is kept distinct due to its unique form factor, higher seating position, and the status, safety, and versatility signals it conveys—factors that set it apart in both buyer psychology and brand positioning. Segment M (MPV) is excluded, given its narrow, declining relevance and limited overlap with broader consumer choice sets. This structuring aligns with widely accepted segmentations.[3]

[3] See URL: https://www.acea.auto/figure/new-passenger-cars-by-segment-in-eu/

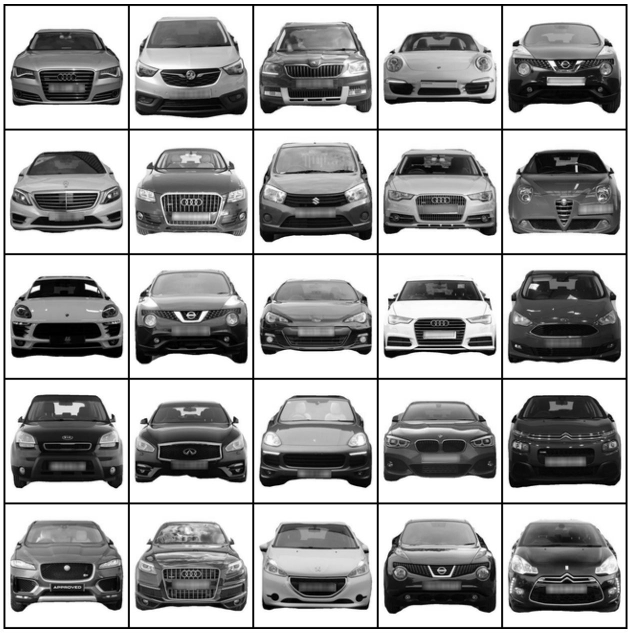
Table 1 Summary Statistics of the 2013 UK Auto Market

	Mean	St. Dev.	Minimum	Maximum
Total Quantity Sold	8,074.834	13,714.100	1	113,390
Price (in £000 units)	26.333	14.668	7.868	108.624
MPG (tens of miles per gallon)	4.999	1.058	2.250	7.200
Horsepower	154.960	84.949	5.000	555.500
Weight (in 10 lbs)	327.704	2.444	324.506	332.106
HP/Wt (in HP per 10 lbs.)	0.461	0.169	0.060	1.347
Length (in 1000 inches)	1.724	0.178	1.062	2.054
Width (in 1000 inches)	0.756	0.063	0.580	0.899
Height (in 1000 inches)	0.616	0.054	0.537	0.780
Wheelbase (in 1000 inches)	1.046	0.081	0.735	1.266
Space (length × width)	1.310	0.215	0.697	1.759
Vehicle Segment (Proportion):				
Segment A (Minicars)	0.112	0.316	0	1
Segment B (Subcompact)	0.120	0.326	0	1
Segment C (Compact)	0.162	0.369	0	1
Segment D (Mid-Size)	0.129	0.335	0	1
Segment E (Mid-Size Luxury)	0.071	0.257	0	1
Segment J (SUV)	0.241	0.428	0	1
Segment M (MPV)	0.166	0.373	0	1
Country of Origin (Proportion):				
France	0.108	0.311	0	1
Germany	0.241	0.428	0	1
Italy	0.058	0.234	0	1
Japan	0.220	0.415	0	1
South Korea	0.091	0.289	0	1
Sweden	0.041	0.200	0	1
UK	0.108	0.311	0	1
USA	0.058	0.234	0	1
Others	0.075	0.263	0	1

We choose to display the summary statistics for the 2013 market because it is the midpoint of 2008 and 2017 markets.

In Figure 1, we display images of 25 automobiles present in our dataset. Note that, we converted color images of size 128×128 to grayscale for our study (sales are also not available separately by color). Moreover, our goal is to extract visual characteristics that are related to the shape of the automobile and not related to the color. For each image, we have its associated make, model, year, structured product characteristics and price.

Figure 1 **Sample of Automobile Images**



5.2. Discovered Visual Characteristics

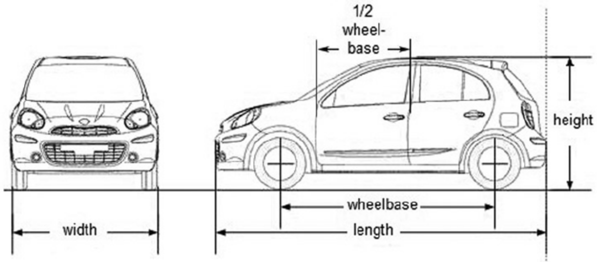
We learn the visual characteristics of each make-model sold in the UK between 2008 and 2017 using disentanglement representation learning. We compare the unsupervised approach to learn visual characteristics with supervised approaches. In the supervised approach, we train the learned visual characteristics to predict the supervisory signal associated with each make-model.[4] Figure 2 shows the dimensions of automobiles that serve as a subset of supervisory signals for disentanglement learning.

We follow the hyperparameter selection approach and the UDR metric described in the Methodology section (Section 4.1). From Table 2, we find that the visual characteristics learned from supervising on the combination of wheelbase, width, and height achieve the best disentanglement in terms of UDR.

[4] We use the following supervisory signals:

1. ’Make’ of the make-model
2. ’Country of Origin’ of the make
3. ’Segment’ of the make-model
4. ’Price’ of the make-model
5. ’Length’ of the make-model
6. ’Width’ of the make-model
7. ’Height’ of the make-model
8. ’Wheelbase’ of the make-model
9. Combination of structured product characteristics of the make-model (‘HP/Weight’, ‘MPG’, ‘Space’)
10. Combination of structured product characteristics of the make-model (‘Length’, ‘Width’, ‘Height’)
11. Combination of structured product characteristics of the make-model (‘Wheelbase’, ‘Width’, ‘Height’)

Figure 2 Dimensions of Automobiles



Note: This figure is sourced from Yap et al. (2013).

Table 2 Comparison of Different Supervisory Approaches

Number of Signals	Supervisory Signals	β	δ	UDR
3	Wheelbase, Width, Height	50	10	0.739
3	HP/Weight, MPG, Space	50	30	0.710
1	Price	50	30	0.708
1	Weight	50	40	0.708
1	Wheelbase	50	30	0.690
1	Width	50	5	0.689
3	Length, Width, Height	50	40	0.678
1	Length	50	40	0.666
0	Unsupervised β-TCVAE	50	0	0.658
1	Height	30	20	0.378
1	Country of Origin	10	10	0.139
1	Segment	10	10	0.134
1	Unsupervised VAE	1	0	0.073
1	Unsupervised AE	0	0	0.074
1	Make	1	1	0.072

$\beta \in [1,5,10,20,30,40,50]$ and $\delta \in [0,1,5,10,20,30,40,50]$.

We show the discovered visual characteristics in Figure 3 corresponding to the model with $\beta = 50$, $\delta = 10$ and the combination of 'Wheelbase, Width, and Height' as supervisory signals. Each row in the image corresponds to a visual characteristic. In each row, we change the value of one visual characteristic while fixing the value of all the other characteristics. Note that since we use a generative deep learning-based method, we can change the underlying learned visual characteristics and generate counterfactual images. The ability to generate counterfactual images allows us to interpret each visual characteristic as it isolates the effect of change in one visual characteristic while keeping the other characteristics fixed. We find five informative visual characteristics of an automobile's front view, while the rest were uninformative (i.e., changing the visual characteristic produces no change in the image). These informative characteristics are:

1. Body Shape: Automobiles scoring high on this characteristic have a narrower, more angular, and less rounded shape, resembling a sedan. Those scoring low have a wider, less angular, and more rounded shape, resembling a hatchback.
2. Color: Automobiles scoring low on this characteristic are lighter, while those scoring high are darker.
3. Grille Height: As the score of this visual characteristic increases, the grille becomes more prominent, larger, and more defined, with the top and bottom parts beginning to merge.
4. Boxiness: Automobiles scoring low on this characteristic have a high degree of boxiness, characterized by a taller, more upright, and narrower shape. Those scoring high have a lower degree of boxiness, with a lower, flatter, and wider appearance.
5. Grille Width: Automobiles scoring low on this characteristic have a narrower, less pronounced grille, while those scoring high have a wider, more prominent grille.

Relationship to Physical Dimensions To better understand our discovered visual dimensions, we examine their correlation with basic physical measurements such as wheelbase, weight, length, height, and width in Table 3. For instance, *Body Shape* (hatchback-like vs. sedan-like) has notable positive correlations with wheelbase (ρ=0.30), length (ρ=0.39), and weight (ρ=-0.33), consistent with hatchback-like profiles tending to have smaller wheelbase, lower length, and lighter. Likewise, its negative correlation with height-to-width ratio (ρ=-0.42) implies that cars with lower *Body Shape* scores appear taller and narrower—characteristics commonly associated with hatchbacks or smaller crossover-style cars.

Boxiness is strongly negatively correlated with vehicle height (ρ=-0.59) and height-to-width ratio (ρ=-0.49), indicating that cars with higher degree of boxiness (lower score on the visual characteristic of boxiness) tend to look taller and more upright from the front. In contrast, vehicles with lower degree of boxiness appear flatter and sleeker. Notably, *Boxiness* does not align closely with length or wheelbase, suggesting it captures more of a cabin “uprightness” or silhouette shape rather than overall vehicle size.

Meanwhile, *Grille Height* exhibits near-zero correlations with physical measures, implying that it reflects primarily stylistic design choices (e.g., tall vs. short front grilles) rather than purely functional or size-related factors. Finally, *Grille Width* is weakly but consistently positively correlated with physical measurements like wheelbase (ρ=0.12), length

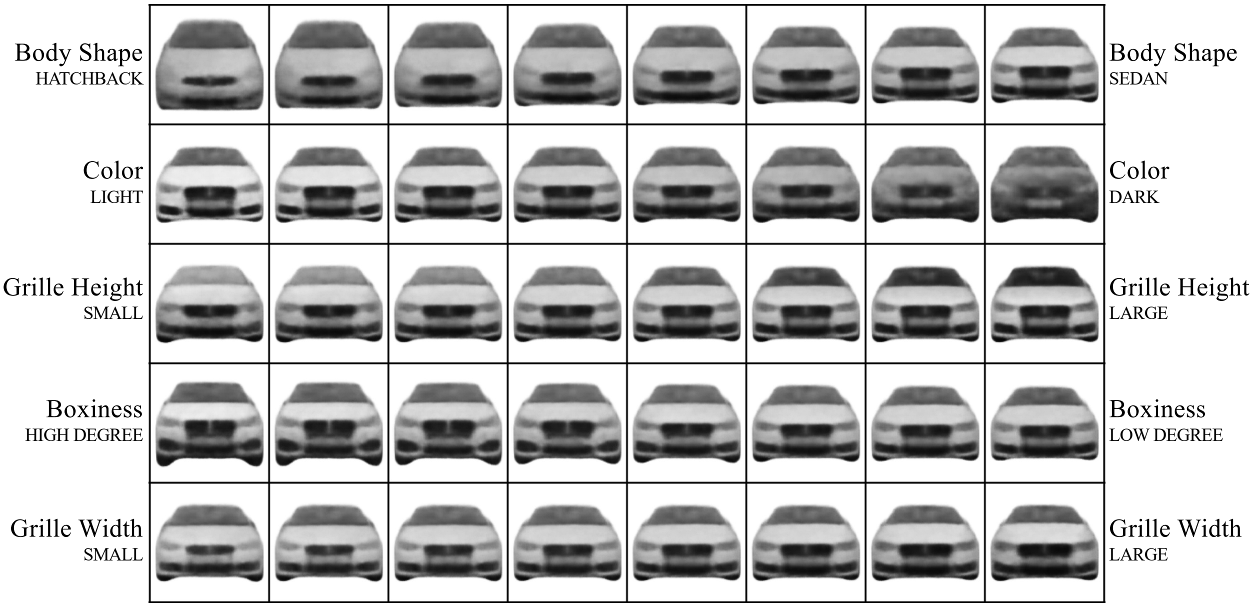
(ρ=0.12), and width (ρ=0.15). While this might hint at a slight tendency for grille width to increase with vehicle size, these correlations are too weak to establish a meaningful relationship between grille width and vehicle dimensions.

Table 3 Correlation Between Discovered Visual Dimensions and Physical Vehicle Measures

	Wheelbase	Weight	Length	Height	Width	Height/Width Ratio
Body Shape	0.30	0.33	0.39	-0.28	0.25	-0.42
Boxiness	0.05	-0.07	0.14	-0.59	0.02	-0.49
Grille Height	0.04	0.02	0.05	-0.04	0.03	-0.05
Grille Width	0.12	0.08	0.12	0.03	0.15	-0.09

Interpretability To validate the interpretability and quantification of the discovered visual characteristics, we conducted surveys with human respondents. The results show that the majority of respondents agreed with each other on the interpretation and with the algorithm on the quantification of the visual characteristics. Detailed information about these surveys can be found in Appendix E.

Figure 3 Discovered Visual Characteristics



Left to Right: Vary one visual characteristic, keeping all others fixed

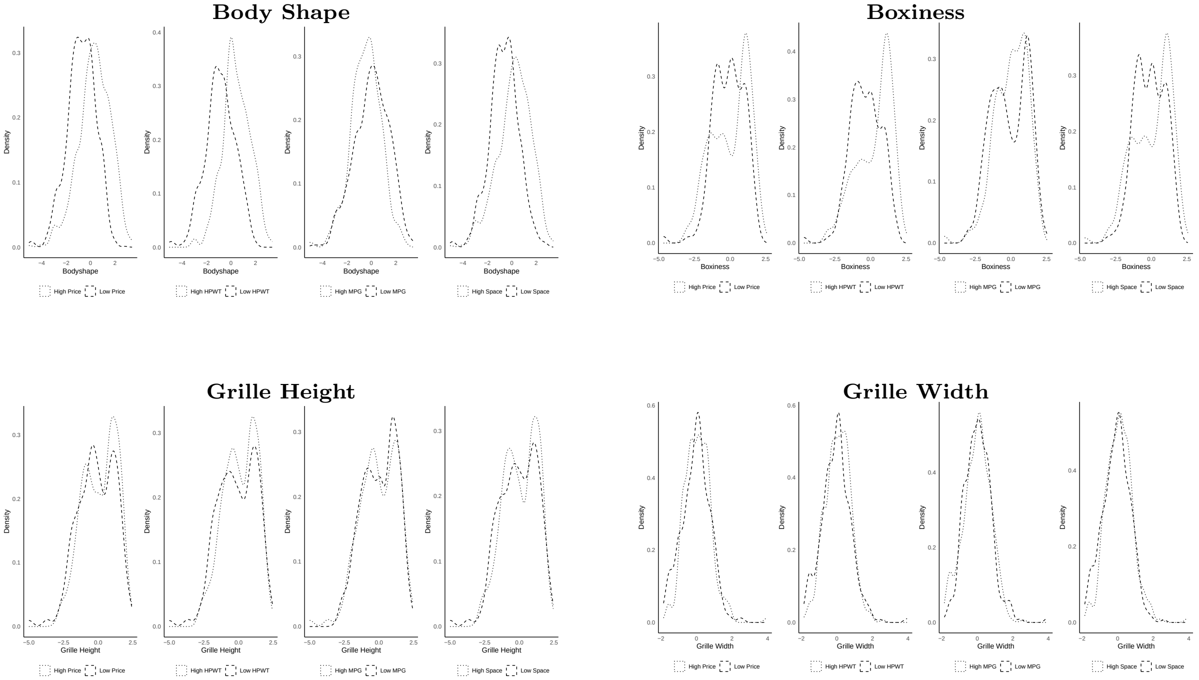
Correlation between Functional & Form Characteristics We analyzed the correlations between structured and visual product characteristics and found that they are weakly correlated with each other, as shown in Table 4. The visual product characteristics are largely uncorrelated with each other, with the exception of a weak correlation between boxiness and body shape. In contrast, the structured product characteristics exhibit correlations with each other. Notably, the structured product characteristics and visual product characteristics are only weakly correlated. This suggests that visual characteristics provide additional information not captured by structured characteristics, highlighting the potential value of incorporating visual information in market structure analysis. This empirical pattern aligns with experimental findings from Kang et al. (2019), who show that consumers make explicit trade-offs between visual form and functional attributes—supporting the idea that visual design holds independent value in consumer decision-making.

Table 4 Correlation Matrix

	Structured Characteristics				Visual Characteristics			
	Price	MPG	HP/Weight	Space	Boxiness	Body Shape	Grille Height	Grille Width
Price	1.00							
MPG	-0.60	1.00						
HP/Weight	0.74	-0.48	1.00					
Space	0.67	-0.47	0.36	1.00				
Boxiness	0.06	0.04	0.29	0.09	1.00			
Body Shape	0.50	-0.25	0.54	0.36	0.13	1.00		
Grille Height	0.11	0.03	0.12	0.05	0.04	-0.02	1.00	
Grille Width	0.07	-0.05	0.04	0.15	0.01	-0.12	-0.05	1.00

To probe these relationships in detail, we examine conditional density distributions of each discovered visual characteristic (*boxiness*, *body shape*, *grille height*, *grille width*), split by "low versus high" categories of structured attributes (i.e., *Price*, *HP/Weight*, *MPG*, and *Space*) in Figure 4. These plots reveal how design form varies when conditioning on different functional contexts. Using these plots, we ask whether *form follows function* entirely, or *does any degree of design freedom exist*? We conduct this initial analysis as proof of existence to set the groundwork for the several analysis we next carry out. We reason that if visual characteristics are indeed entirely constrained by structured characteristics, for example, if a vehicle's visual form only depended on the vehicle's underlying structure (e.g., frame, chassis, powertrain), then it would not even be possible for firms to compete using visual form. If visual characteristics instead are even partially unconstrained by structured characteristics, then it is possible for firms to differentiate themselves using visual characteristics and not structured characteristics alone. In fact, much prior work in automotive design has indicated that firms do have a significant degree of design freedom given: (1) the large heterogeneity in visual form amongst automotive vehicles; and (2) the fact that automotive firms spend 10's to 100's of millions on both eliciting consumer preferences over conceptual designs (Manoogian II 2013). Figure 4 yields the following insights:

Figure 4 **Discovered Visual Characteristics - Conditional Density Plots**



1. Vehicles with higher price, higher power (measured by horsepower per weight), and higher interior space all cluster toward the sedan-like *body-shape* visual characteristic as opposed to the hatchback-like *body-shape* visual characteristic. In fact, the more favorable aerodynamic characteristics of the traditional sedan shape—specifically its lower drag coefficient—make it a natural choice for manufacturers when designing higher-performance vehicles with greater horsepower-to-weight ratios (HP/WT) (Hucho 2013, Barnard 2001). Additionally, it is well known that sedans are generally designed to offer more interior space—particularly in the rear seat area—due to the presence of a separate trunk compartment.[5]. Hence, for body shape, *form largely follows function*. Sedans also tend to come at a higher price point than hatchbacks.[6]

[5] See URL: https://www.conceptcarcredit.co.uk/blog/why-a-cars-body-style-is-important/

[6] See URL: https://turo.com/blog/australia/gearheads/hatchback-vs-sedan/

2. Vehicles with higher price, higher power (measured by horsepower per weight), and higher interior space exhibit distinctly *bimodal* distributions in the boxiness factor, whereas vehicles with lower price, lower power, and lower interior space display tightly centered *unimodal* peaks at lower boxiness scores (i.e., more boxy). This pattern reflects underlying differences in segment composition. For instance, vehicles with lower space are dominated by hatchbacks from Segment A, B and C, as well as SUVs from Segment J, and MPVs from Segment M. Meanwhile, vehciles with higher space span a broader mix, including boxy SUVs (Segment J) and sleek sedans (Segment C, D, and E), leading to a bimodal distribution. A similar dynamic holds for power: lower HP/weight vehicles belong to hatchbacks, SUVs and MPVs, while higher HP/weight vehicles span both sedans and large SUVs. Together, these patterns suggest that while form appears tightly constrained by function in lower priced vehicles – most vehicles converge on boxy, upright designs, higher priced vehicles exhibit greater design variation.
3. Grille height and width show minimal functional coupling. Across all functional splits, the densities for *grille height* and *grille width* are nearly coincident. The front-end geometry is often wielded as a brand "face"(e.g., wide single-frame grilles for Audi, tall double-kidney grilles for BMW, spindle grilles for Lexus). Consumer research experiments confirm that such anthropomorphic "faces" drive perceptions of dominance or friendliness independent of technical merit (Landwehr et al. 2011). Hence, for grille dimensions, *form follows branding* more than function.

Taken together, the evidence supports a *selective* version of the "form follows function"' maxim: some visual factors (body shape, boxiness) are tightly tethered to physical constraints, while others (grille geometry) remain largely discretionary and therefore valuable for stylistic differentiation.

Disentanglement picks up things professional human reviewers write about? We use a language–model–based content analysis of 1,142 professional reviews sourced from AutoCar[7] to quantify how much professional reviews are devoted to visual design versus functionality.[8] Median word shares indicate that reviewers allocate roughly one quarter of each review to visual aspects (24.1%), with the remainder split between functional content (55.2%) and

[7] See URL https://www.autocar.co.uk/

[8] Prompt and coding rubric in Appendix G (LLM setup and guardrails).

Table 5 Prevalence of front-facing visual aspects in professional reviews

Aspect	Share of reviews
Bodyshape	88%
Boxiness	98%
Grille width	39%
Grille height	22%

other remarks (19.6%). Table **??** shows that body shape is mentioned in 88% of reviews and boxiness in 98%. By contrast, grille geometry appears less frequently (width 39%, height 22%). Together, these findings indicate that professional reviewers devote substantial attention to visual design, and that the disentangled visual characteristics we recover (bodyshape, boxiness, grille geometry) are indeed present in a large share of reviews.

5.3. Visual Market Structure Map

We apply Multidimensional Scaling (MDS) to create market structure maps based on structured product characteristics alone, visual characteristics alone, and a combination of both, as described in the Methodology section (Section 4.2). Following Berry et al. (1995), the structured product characteristics used are 'Price', 'HP/Weight (proxy for power)', MPG (proxy for fuel efficiency)', and 'Space (measures itself and is a proxy for safety)', while the visual product characteristics are 'Body Shape', 'Boxiness', 'Grille Height', and 'Grille Width'. Before implementing MDS, we normalize each characteristic by subtracting its mean value across all make-models and dividing by its standard deviation.

5.4. Insight 1: Does differentiation across product-segments increase when visual information is included?

Research on product categorization shows that when consumers first encounter an unfamiliar product they quickly assign it to a familiar mental *slot* on the basis of visual resemblance to known exemplars (Loken and Ward 1990, Sujan 1985, Sujan and Dekleva 1987). Bloch (1995) argues that such assignments rely on a mix of holistic impressions and atomistic cues. According to Veryzer Jr and Hutchinson (1998), holistic impressions reflect higher-order relational properties, such as the overall visual unity of the product (e.g., bodyshape, boxiness), while atomistic cues correspond to first-order component-level features (e.g., grille width and grille height). Prior research has demonstrated that visual features can effectively distinguish between vehicle types. In the computer vision domain, Dong et al. (2015) achieved high accuracy classifying vehicles into categories (sedans, SUVs, trucks, etc.) using front-facing images, while Yang et al. (2015) reached 54.1% accuracy predicting car types from front-facing images. However, these studies primarily focus on classification accuracy rather than examining how visual features specifically contribute to market differentiation beyond structured product characteristics. If consumers initially categorize vehicles based on visual cues as this research suggests, then the vehicle's *appearance* should be important to product-segment differentiation. Building on this insight, we test whether adding interpretable visual characteristics (body shape, boxiness, grille height, grille width) increases the differentiation across product-segments relative to a baseline that includes only structured characteristics.

We randomly split the 231 make-models belonging to the 2013 UK automobile market into an 80% training set and a 20% test set. Our primary goal was to evaluate whether incorporating visual characteristics of a car's front face (such as bodyshape, boxiness and grille dimensions) can enhance product-segment classification beyond what is achieved by using structured characteristics alone (price, MPG, etc.).[9] Specifically, we trained two machine learning algorithms—Random Forest and XGBoost—on three feature subsets: structured characteristics only, visual characteristics only, and a combined set of both. Table 6 reports the accuracy and Kappa (a reliability measure that accounts for chance agreement) of these six models on the held-out test set. We see that including both structured and visual characteristics, generally outperform the single-characteristics models for both Random Forest and XGBoost, reaching around 72% accuracy (with a Kappa near 0.66). Structured characteristics alone models yield moderate performance (65–67 accuracy), whereas visual characteristics alone only models produce lower accuracy (42–44%). Overall, these findings suggest that front-face visual characteristics do contribute additional predictive power, and combining them with structured data yields the strongest classification performance.

Next, we visualize this improvement in prediction accuracy using both structured and visual characteristics using market structure mapping. Figure 5 shows 3 market structure maps over the same market. From the market structure map using only structured product characteristics, we note that Segment (A and B) is clearly separated from both Segment (C, D, and E) and Segment J. However, Segment (C, D, and E) and J overlap. From the market structure map using only visual product characteristics, we note that Segment

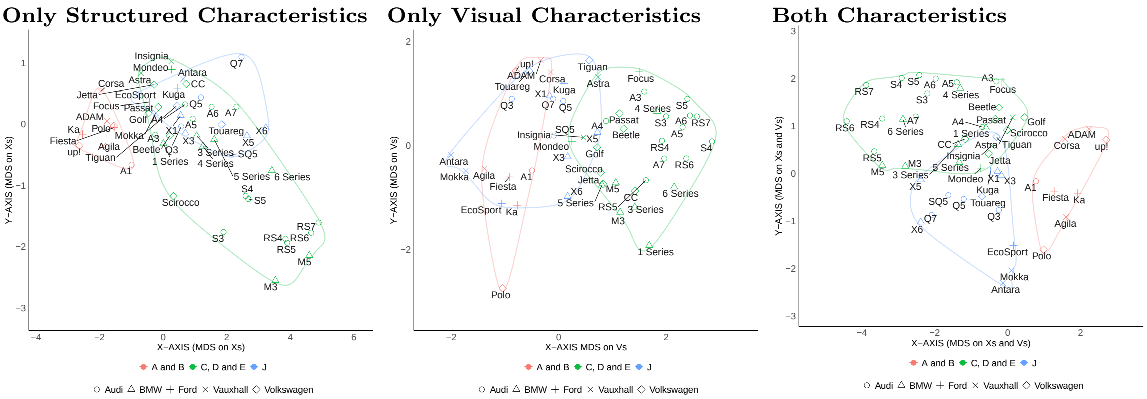
[9] We classify vehicles into the original industry-defined segments: A (Minicars), B (Subcompact), C (Compact), D (Mid-size), E (Mid-size Luxury), J (SUV), and M (MPV).

Table 6 **Classification Performance for Product Segment (Held-Out Test Dataset)**

Method	Features	Accuracy	95% CI	Kappa
Random Forest	Structured	0.6512	(0.4907, 0.7899)	0.5857
Random Forest	Visual	0.4186	(0.2701, 0.5787)	0.3109
Random Forest	Mixed	0.7209	(0.5633, 0.8467)	0.6649
XGBoost	Structured	0.6744	(0.5146, 0.8092)	0.6114
XGBoost	Visual	0.4419	(0.2908, 0.6012)	0.3294
XGBoost	Mixed	0.7209	(0.5633, 0.8467)	0.6658

(A and B) and Segment J overlap a lot. However, Segment (C, D, and E) has very little overlap with Segment (A and B) and Segment J. Interestingly, when we account for both structured and visual product characteristics, Segment (A and B), Segment (C, D, and E) and Segment J separate out significantly more. This means that if one considers only type of characteristic, then the market appears more competitive. However, if one includes both the characteristics, then the market seems less competitive because the segments seem substantially more differentiated. In other words, visual characteristics do indeed enable product differentiation outside of just structured characteristics.

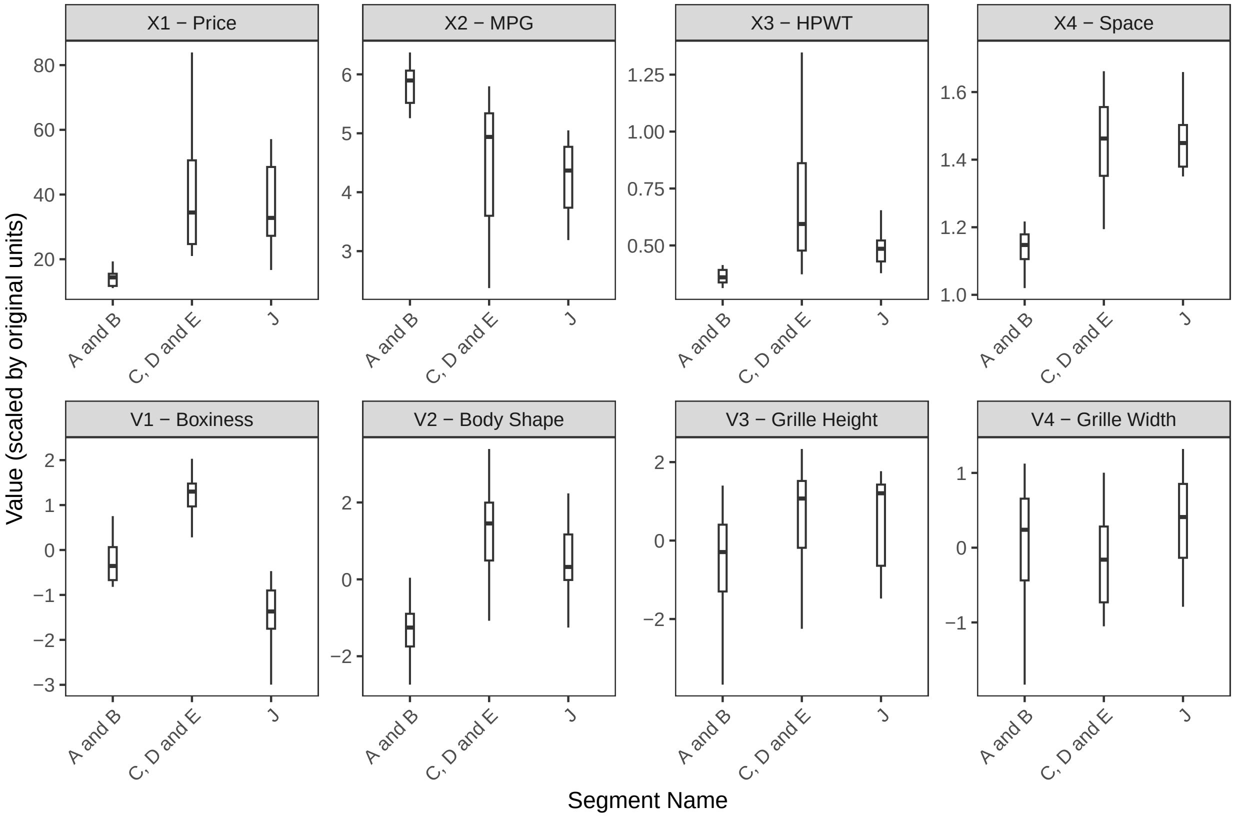
Figure 5 **(Color Online) Market Structure Map (MDS)**


Only Structured Characteristics
Only Visual Characteristics
Both Characteristics

To quantify how distinctly the three product-segments occupy the map, we adapt the convex-hull overlap metric used by Pironon et al. (2019). In their ecological-niche work, Pironon et al. (2019) construct minimum-convex polygons around species occurrences in a two-dimensional PCA climate space and measure the proportion of one polygon that intersects another. We apply the same geometry to the MDS coordinates of each product-segment: for every map (structured-only, visual-only, combined) we draw convex hulls around the Segment (A and B), Segment (C, D, and E) and Segment J points, compute the pairwise intersection areas, and express *overlap* as the fraction of total hull area that is shared. We find that in a map using only structured characteristics, 18.2% of the area is overlapping. Note that 26.8% of the area is overlapping in the market structure map created using only visual characteristics. Interestingly, when we create a market structure map using both structured and visual characteristics, then only 5.6% of the area is overlapping.

Building on these area-overlap results, we can investigate why the product-segments spread apart once we include both structured and visual characteristics by looking at the distributions of structured characteristics and visual characteristics in Figure 6. This analysis to study the underlying reason for differentiation and the market structure is possible only because disentanglement approach yields interpretable visual characteristics.

Figure 6 (Color Online) Boxplots of 8 Characteristics by Segment: Top 5 Makes in Market = 2013



Note: The box represents the interquartile range (IQR), with the horizontal line inside denoting the median, and whiskers extending to approximately ±1.5×IQR. Outliers beyond the whiskers are omitted for clarity.

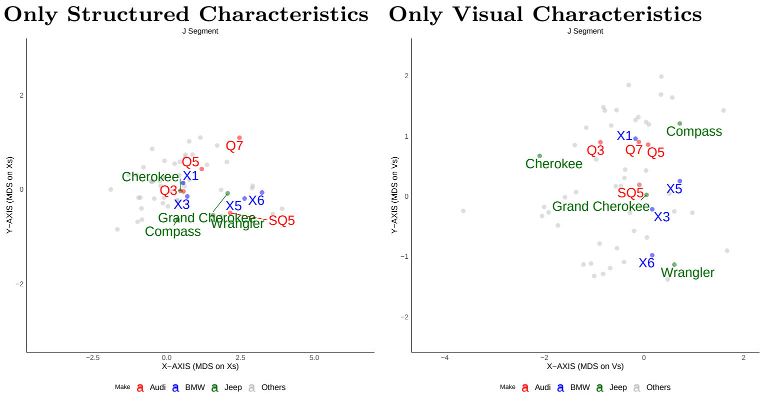
These plots show nuanced findings: Segment (C, D, and E) and Segment J, for example, have similar values across structured characteristics (price, MPG, HPWT, space), as well as similar values across the 'grille height' and 'grille width,' yet diverge substantially on the 'boxiness' visual characteristic. *In other words, the disentangled representation surfaces the specific traits (e.g., 'boxiness') that differentiating the Segment (C, D, and E) and Segment J segments.* This observation gives an explanation for why the combined (structured and visual) map leads to lower overlap than either a purely structured or purely visual map alone. This observation highlights how specific visual characteristics like 'boxiness' styling play a crucial role in product-segment-level differentiation.

5.5. Insight 2: Brands have different visual strategies

Decades of marketing research highlight visual design as a strategic balancing act between maintaining brand consistency that reinforces consumer recognition and introducing sufficient novelty to sustain consumer engagement Homburg et al. (2015), Burnap et al. (2016a). Early conceptual work shows that product form simultaneously communicates brand meaning and adds usage value, making coherent visual cues a source of competitive advantage Bloch (1995). At the same time, (Veryzer Jr and Hutchinson 1998) has shown that designs rated as both prototypical and unified are liked only up to the point where further repetition dulls aesthetic pleasure. Recent research confirms that brands vary systematically in how they resolve this tension: some (e.g., BMW) keep models tightly clustered on visual dimensions, trading variety for a stronger brand signal, whereas others (e.g., Lexus) spread out stylistically to capture heterogeneous tastes even if that weakens immediate recognition (Liu et al. 2017, Jindal et al. 2016). Recent literature has shown how deep-learning approaches can extract brand identity from image-based embeddings (Liu et al. 2020b). We investigate how different brands position their models across structured versus visual attributes. Complementing this, Dzyabura and Peres (2021) show that unsupervised visual inputs—like consumer-generated image collages—can reveal brand-level perceptual differences, underscoring the role of visual cues in brand positioning.

We first operationalize this by investigating the spread of each brand in the structured characteristic space as well the visual characteristic space. Our analysis focuses on the J segment (i.e., SUVs), and within this segment, the top 4 highest selling models of each brand in the UK market. [10] Figure 7 shows the market structure map of Segment J in the 2013 market with the models of these makes highlighted in different color.

Figure 7 **(Color Online) Segment J in 2013**



To quantify how much design space each brand occupies, we extend the same convex-hull geometry used in the segment-level analysis (Pironon et al. 2019). For each brand, we construct a minimum convex polygon around the MDS coordinates of its models within the J segment. We then compute the total area of this polygon as a measure of the brand's span in that characteristic space. We divide this by the total area of the convex polygon around the MDS coordinates of the entire J segment. When calculated separately for structured and visual dimensions, this "area share" provides an interpretable proxy for brand-level dispersion: a large area share in structured space suggests functional variety (e.g., differences in price or horsepower), whereas a large area share in visual space indicates greater stylistic diversity across models. This metric allows us to evaluate how tightly or loosely a brand positions its lineup and to compare strategic choices in visual versus structured differentiation.

Table 7 reports how much of the 2013 Segment J (SUV) *space* each manufacturer occupies when we look only at structured characteristics versus when we look only at visual characteristics. Audi's models command a significant 17.90% of the structured area share yet only 2.96% in the visual space. This suggests a brand strategy that emphasizes a strong,

[10] We choose the J segment both because it is one of the largest overall segments by consumer choice, as well as being practical for this analysis given the J segment has numerous brands with each brand having numerous models (make-model).

recognizable exterior *look* across models—while still giving consumers varying levels of performance, size, and price options. In contrast, Jeep reverses that pattern with 9.38% structured and 28.42% visual area share. Jeep keeps many of its models in a similar functional range while having a larger dispersion across the visual characteristics. We note that this in particular is likely due to the presence of the Jeep Wrangler, which has undergone relatively minor stylistic changes after it was commercialized for the civilian population from the WWII Jeep, a design that has persisted in appealing to a significant portion of buyers (Rosenbaum 2007).

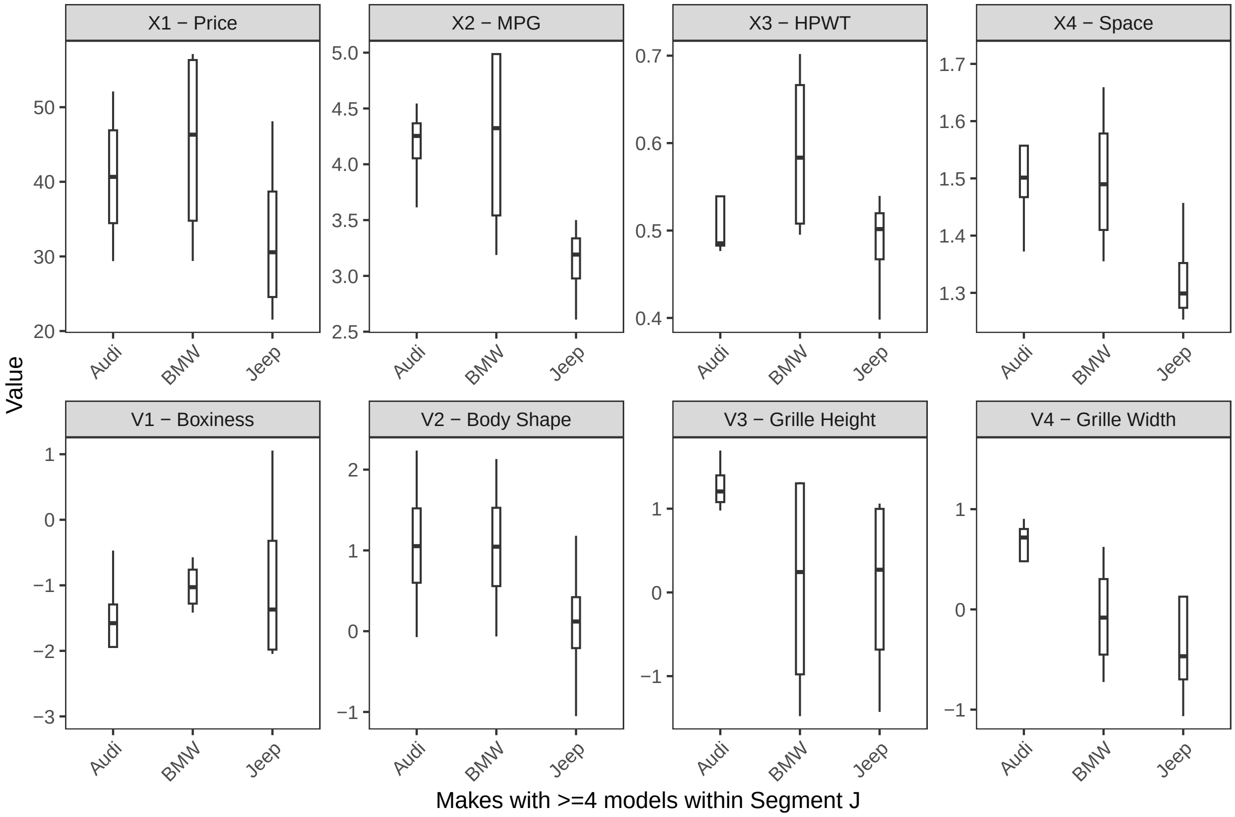
Table 7 Area Share of a Make in Structured Space & Visual Space in Segment J (SUV)

Make	Models	Area Share (Structured)	Area Share (Visual)	Ratio
Audi	4	17.90%	2.96%	6.05
BMW	4	6.35%	6.48%	0.98
Jeep	4	9.38%	28.42%	0.33

Figure 8 next investigates how each brand's SUVs distribute themselves across both structured and visual attributes. In particular, we focus on Audi and Jeep given their differences in area share. Audi's lineup covers a relatively broad range of structured characteristics but remains more tightly clustered on visual features (e.g., smaller Grille Height and Width ranges). It reflects Audi's deliberate emphasis on consistent brand styling, particularly via its now-iconic Singleframe grille design.[11] At the same time, the variation in Audi's 'Body Shape' scores demonstrates the use of visual design to differentiate models within its lineup. In other words, some visual characteristics are used to enforce brand consistency, whereas others are free to vary and used for differentiation within the product line. Jeep, by contrast, spans a narrower band of structured characteristics while exhibiting a broader dispersion in visual design (e.g., larger 'Boxiness' range), suggesting a deliberate strategy to differentiate models via visual characteristics rather than performance or size. The automatic extraction of interpretable visual characteristics helps us connect real-world design choices—like grille width, grille height, and overall boxiness as well as body shape—to each brand's distinct visual strategy, offering firms not just descriptive insights, but also prescriptive value in identifying which visual characteristics to standardize or vary across their lineup.

[11] Audi's grille design has become one of the most recognizable visual elements in the automotive world. The Singleframe grille, a term Audi has trademarked, evolved over decades and now anchors the front design of nearly all

Figure 8 (Color Online) Boxplots of 8 Characteristics by Makes with >=4 models in Segment J: Market = 2013



Note: The box represents the interquartile range (IQR), with the horizontal line inside denoting the median, and whiskers extending to approximately ±1.5×IQR. Outliers beyond the whiskers are omitted for clarity.

We repeat this exercise for Segment (C, D, and E) in Appendix F.

5.6. Insight 3: Crossovers – Making the Leap from Car to SUV

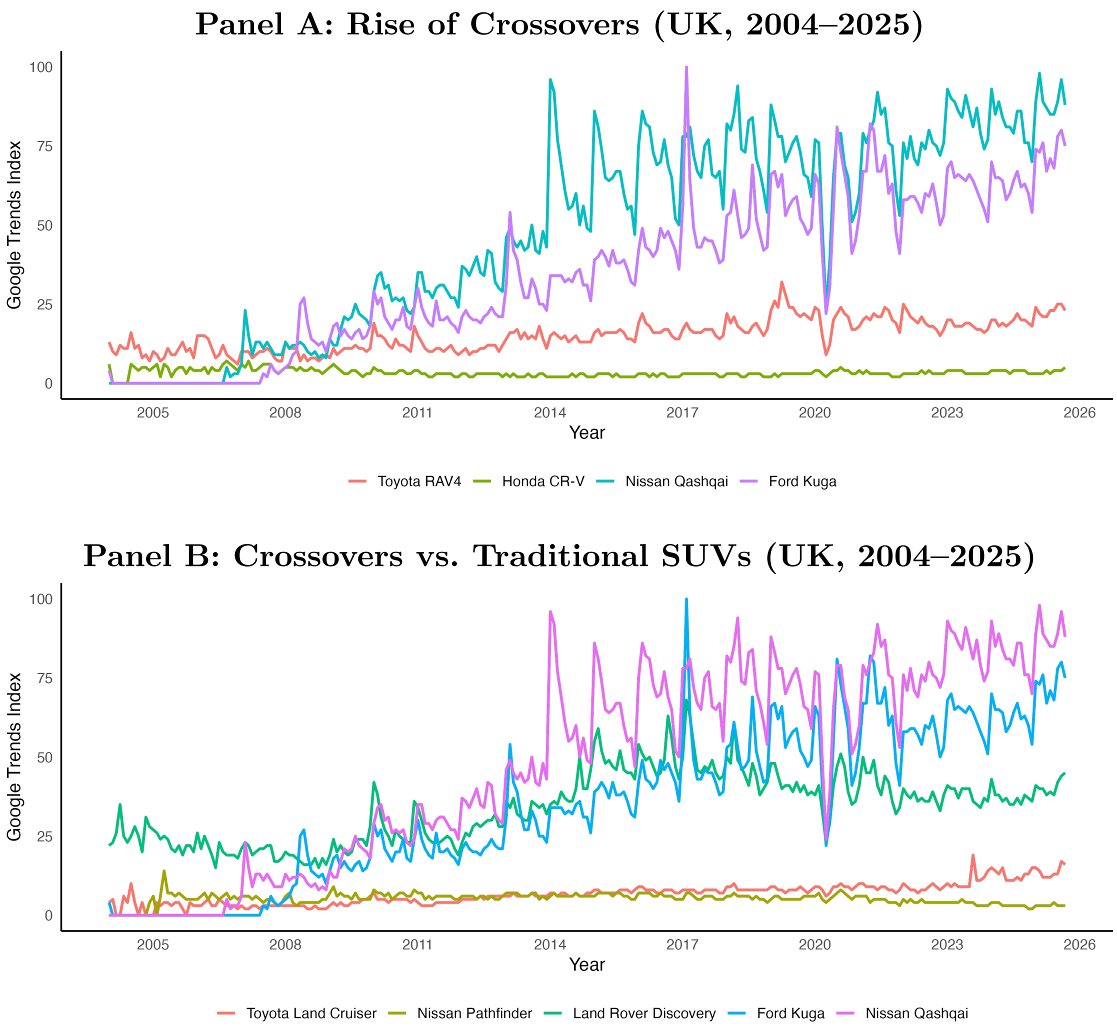
Crossovers date back to the 1990s, when automakers began experimenting with vehicles that combined the features of passenger cars and sport utility vehicles (SUVs). One of the first crossovers was the Toyota RAV4, introduced in 1994, which was constructed on a car-based (or sedan-based or unibody) platform but had the ground clearance and versatility of an SUV.[12] The crossover designation was applied to these new vehicle types within the broad SUV segment to distinguish them from the more rugged, truck-based (or body-on-frame) SUVs that had previously dominated the market.

its models. Designers such as Wolfgang Egger and Walter de Silva have described it as central to Audi's brand identity—balancing engineering rationality with emotional appeal. Even in the electric era, Audi has committed to retaining the grille as a symbolic and stylistic element. See quotes in Appendix H.

[12] See URL: https://www.autoevolution.com/news/the-origins-of-crossovers-and-the-main-reason-why-they-ve-become-so-popular-155237.html

To motivate why crossovers matter in our data, Figure 9 plots UK Google Trends interest since 2004. Panel A shows the diffusion of crossovers (e.g., Qashqai, Kuga, RAV4, CR-V), which rises sharply from the late 2000s onward. Panel B contrasts these with traditional, truck-based SUVs (e.g., Land Cruiser, Pathfinder, Discovery), whose search interest is comparatively flat over the same horizon. This time-series pattern foreshadows the shifts in competitive sets we study below.

Figure 9 **Google Trends Interest for Crossovers and Traditional SUVs in the UK (2004–2025)**



Crossovers differ from sedans (and hatchbacks) in several key ways. While sedans are typically designed for on-road driving and prioritize comfort and fuel efficiency, crossovers are designed to be more versatile and capable of handling light off-road driving. Crossovers typically have a higher ground clearance than sedans, which allows for better approach and departure angles, as well as improved visibility and sense of security for drivers. This is achieved by raising the suspension, using larger wheels and tires, and modifying the body structure to accommodate the increased height. The body is often strengthened with additional reinforcements, such as thicker steel, to improve durability.

While crossovers appear similar to truck-based (body-on-frame) SUVs in their visual characteristics, crossovers are built on car platforms with a unibody construction to ensure sedan-like ride and handling. Using a unibody platform also allows crossovers to have better fuel economy and lower emissions than truck-based SUVs.[13] For instance, the Ford Focus platform was used as a basis to develop the Ford Kuga crossover, and the Honda Civic platform served as the basis for the Honda CR-V. From a manufacturing perspective, crossovers are therefore likely to have similar components and structured characteristics (like engine, horsepower etc.) as cars.

Prior work has examined how vehicles blend visual cues from multiple categories to achieve hybrid design goals. (Orsborn et al. 2006) introduce shape grammars that combine class-specific rules to generate cross-over vehicles, such as designs that span traits of coupes and SUVs. (Burnap et al. 2016b) estimate a continuous distribution of automobile forms using deep generative models, enabling the morphing of designs across segments, though without quantifying how such morphs affect perceived segment identity. (Liu et al. 2017) introduce the concept of *cross-segment mimicry* (CSM), showing that economy cars adopting stylistic traits of luxury vehicles can enjoy greater consumer appeal. While their focus is on upward mimicry rather than SUV–car crossover, it highlights the strategic role of visual design in shaping perceived product category. Our contribution is to show that interpretable, disentangled visual characteristics can enable vehicles to blur traditional category boundaries—even when structured specifications remain nearly unchanged—providing evidence that visual form alone can enable a product to ‘cross over’ into a different competitive set.

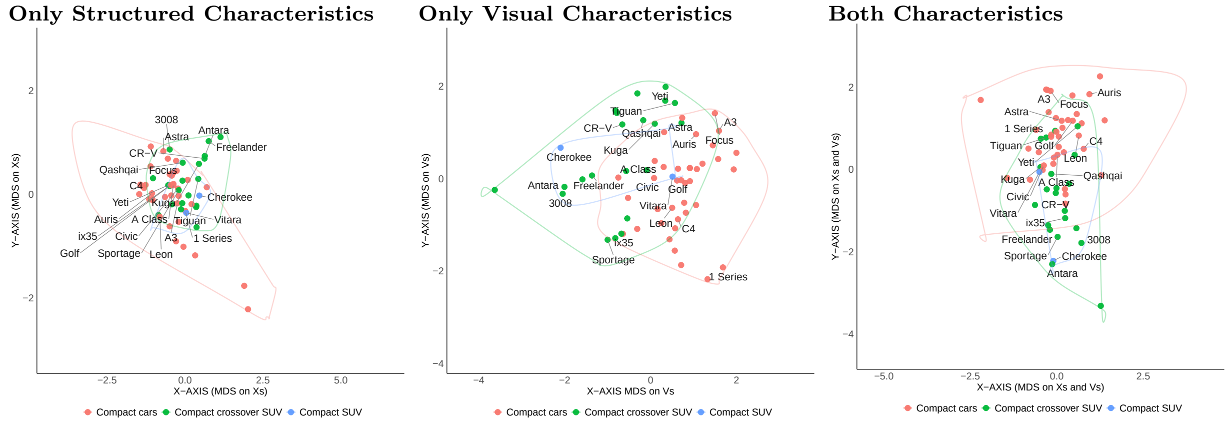
The rise in crossover sales in the UK has been significant in recent years, with crossovers becoming one of the fastest-growing segments of the automotive market. From 2010 to 2017, crossover sales in the UK grew from 95,942 to 612,180 units, an increase of over

[13] See URL: https://www.drivingline.com/articles/the-crossover-era-why-modern-suvs-are-better-enthusiast-vehicles-than-you-think/

538%, a significantly higher rate of growth than the entire market, which grew by 45% during this timeframe.

We evaluate how compact crossovers (based on Segment C platform but typically considered part of Segment J) compare to both compact cars (Segment C) and compact SUVs (Segment J) in our analysis. In Figure 10, we can see the impact of both structured and visual characteristics in positioning products across different segments. Compact car models share a substantial overlap in structured product characteristics with the compact crossover SUVs. For instance, Vauxhall Astra (a compact car) appears relatively close to the Vauxhall Antara (a compact crossover SUV) in the structured characteristics space (left panel), indicating similar attributes in terms of price, MPG, horsepower, and other functional features. However, when examining the visual characteristics map (middle panel), we observe that the Vauxhall Astra is positioned much further from the Vauxhall Antara than in the structured space, with many other vehicles positioned between them. This leads to two noteworthy insights. First, that with very similar structured characteristics, using a different visual design makes it possible for the product to ‘cross over’ into a different segment. Second, the competitive set of products that might be considered by consumers are likely to change due to such a change in visual appearance.

Figure 10 (Color Online) Body Type (2013): Compact Cars, Crossovers and SUVs

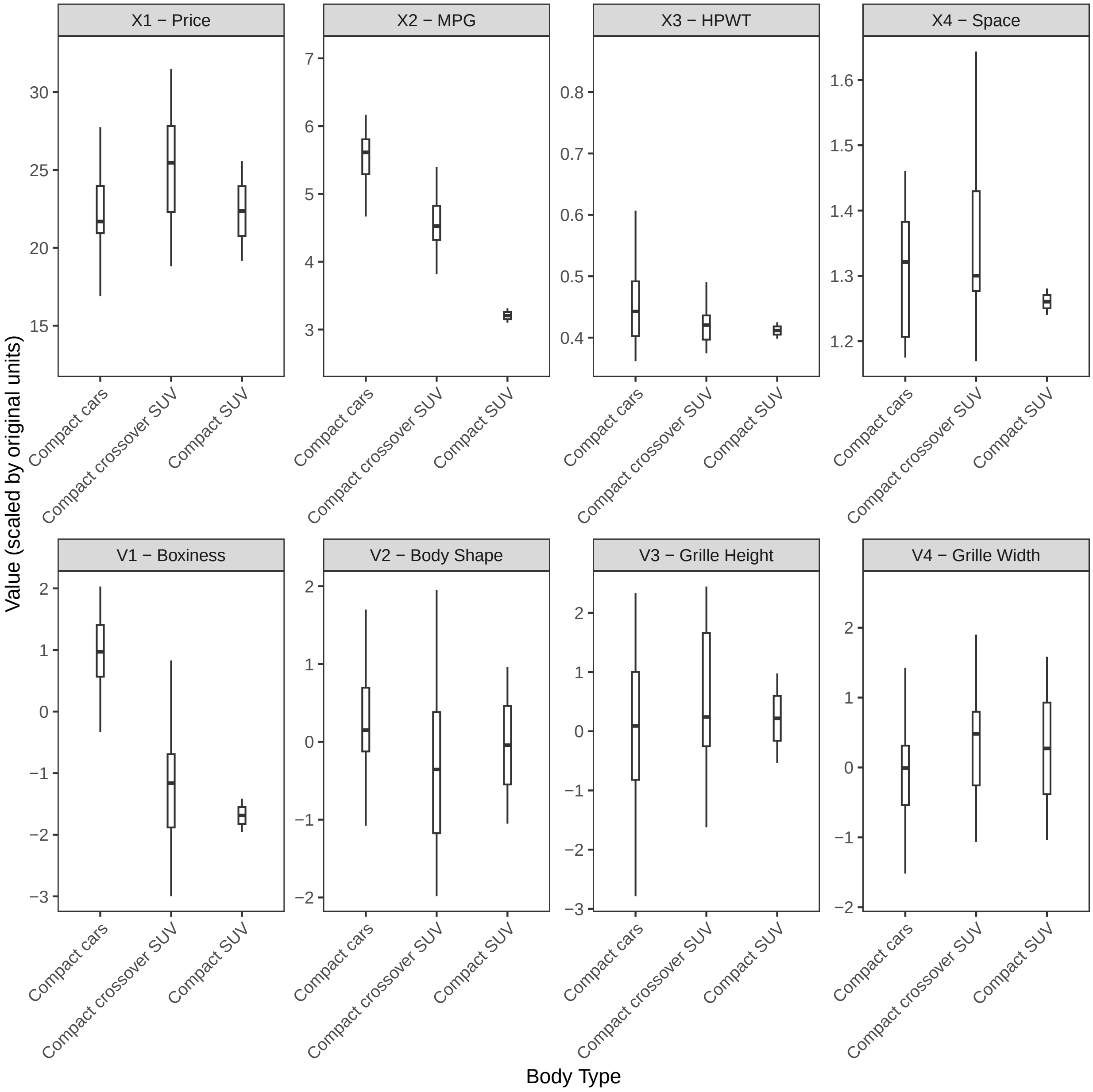


Note: Text labels display only the top 10 models with highest market share in each vehicle class.

To test the second insight, we evaluate how often consumers search for pairs of car models. We then evaluate whether they tend to compare crossovers more with other cars (that are similar in structured characteristics to them), or with other SUVs (that are similar in visual characteristics, but different in structured characteristics). For example, in 2016 the average Google Trend Score for a pair consisting of two compact crossover SUVs is 141; for a pair consisting of one compact car (Segment C) and one compact crossover SUV is 45; for a pair consisting of one compact crossover (unibody) SUV and one compact truck-based (body-on-frame) SUV is 10.[14] We find that consumers tend to compare crossovers mostly with SUVs, and do not seem to compare them with sedans. Overall, this finding points to the importance of visual characteristics (relative to structured characteristics) as being of high importance to consumers.

[14] The ordering is the same for other years.

Figure 11 (Color Online) Boxplots of 8 Characteristics by Class: Market = 2013



Note: The box represents the interquartile range (IQR), with the horizontal line inside denoting the median, and whiskers extending to approximately $\pm 1.5\times$**IQR. Outliers beyond the whiskers are omitted for clarity.**

We have not just quantified the visual appearance of a product, but our method allows us to identify and obtain human-interpretable or disentangled visual characteristics. This leads to further insights that would not be possible to obtain otherwise. In Figure 11, we provide boxplots showing the range of 4 structured and 4 visual characteristics across three plausibly related classes of cars: a) compact cars, b) compact crossover (car platform) SUVs and c) compact (truck platform) SUVs. We observe that for structured characteristics, the compact crossover SUVs have fuel efficiency (MPG), power (HPWT) and space closer to the compact cars than to the compact truck-based SUVs. However, when we examine the visual characteristics, the observations are more nuanced. We find that in terms of boxiness, the crossover (car-based) and non-crossover (truck-based) SUVs are very close to one another, but differ quite substantially from compact cars. Thus, having the disentangled visual characteristics allows us to identify how exactly these classes of automobile are similar or different in a human-interpretable manner.

5.7. Insight 4A: Do visual characteristics help in predicting the co–occurrence matrix?

Our ideal target is a *substitution matrix* derived from revealed preferences: the cross–elasticity matrix implied by a structural demand model (e.g., BLP for autos; Berry et al. 1995, Nevo 2001). Estimating such a matrix requires specifying utility over characteristics, solving for market shares, and then mapping the estimated preferences into substitution patterns across products. Rather than commit to a particular demand system, we take a complementary, data–driven route and predict a *co–occurrence matrix* that summarizes how often products are jointly considered.

Co–occurrences can be observed in several behavioral signals (e.g., joint searches, same–session page views, forum comparisons). We operationalize them using *Google Trends pairwise co–searches*: for product i and rival j in year t, we record the normalized query intensity for the joint string "i j" as s_{ijt} and stack these into a matrix $C_t = [s_{ijt}]$. Prior work shows that search signals are informative about demand and attention (e.g., Goel et al. 2010, Choi and Varian 2012), and pairwise co–searches provide direct evidence of juxtaposition in consumer information acquisition (Yang et al. 2022). While C_t is not a cross–elasticity matrix, it is a transparent, high–frequency proxy for joint consideration that is consistent with substitution in characteristics space.

Given a focal product–year (i,t), we form the candidate rival set J_{it} and observe pairwise co–search scores $\{s_{ijt} : j \in J_{it}\}$. Our prediction target is the (normalized) co–search intensity s_{ijt} and its incidence $\nVdash\{s_{ijt} > 0\}$. We estimate a two–part model: Stage A predicts incidence with gradient–boosted trees and probability calibration; Stage B predicts conditional intensity on the positives using gradient–boosted regression on $\log(1 + s_{ijt})$. All features are normalized (standardized) on the 2008–2015 training data and then applied to the 2016 validation and 2017 test sets. Hyperparameters are tuned on 2016, and performance is reported on 2017. Stage A (incidence) is evaluated with PR–AUC; Stage B (conditional intensity) with RMSE and MAE on the positive pairs.

We compare four visual representations: Disentangled (no color), VAE, Disentangled+color, and ResNet–50 embeddings reduced via PCA (top four PCs). Across specifications, the functional feature vector is fixed (price, horsepower–to–weight, MPG, space). We report three feature sets per representation: (i) functional only; (ii) visual only (the chosen visual representation, without functional variables); and (iii) combined (concatenating functional and visual). Each specification also includes pair indicators for same–make and same–segment.

Main results (2017 test). Table 8 shows that visual characteristics improve prediction of co–occurrences across all feature representations. In Stage A (incidence), adding visual features to functional features consistently raises PR–AUC (from 0.323 with functional only to the 0.34–0.38 range with visual or combined features). In Stage B (conditional intensity), visual–only models deliver the lowest error, reducing RMSE from 113.5 (functional) to 83–88 across representations (a 23–27% reduction), with MAE showing a similar decline (from 11.4 to about 8.1–8.4). Combined models sometimes help incidence but do not outperform visual–only on intensity. Overall, functional covariates are most useful for predicting whether any co–search occurs (Stage A), while visual similarity is especially informative about how much co–search mass concentrates among particular rivals (Stage B).

5.8. Insight 4B: Do visual or functional characteristics matter more when consumers compare models?

Consumers rarely evaluate all alternatives at once. They first reduce the universal set to form a *consideration set* and then choose from that set (Bettman 1979, Howard and Sheth 1969, Newell et al. 1972). This two–stage view implies that what matters is not only which products are ultimately chosen, but also which products are *jointly considered.* Since co–entry into a consideration set typically reflects perceived substitutability, products that end up side–by–side are expected to resemble each other in the product characteristics space (Berry et al. 1995). Moreover, products that are present together in individual consideration sets are therefore more likely to be compared with one another, as

Table 8 Co–search incidence and intensity (2017 test): effect of visual features

		Stage A: Incidence	Stage B: Intensity on positives	
Representation	Feature set	PR–AUC ↑	RMSE ↓	MAE ↓
Disentangled	Functional	0.323	113.5	11.4
	Visual	0.331	83.1	8.1
	Combined	0.351	104.7	10.2
VAE	Functional	0.323	113.5	11.4
	Visual	0.346	87.7	8.3
	Combined	0.372	111.8	11.1
Disentangled + color	Functional	0.323	113.5	11.4
	Visual	0.334	83.8	8.1
	Combined	0.364	97.3	9.4
ResNet–50 + PCA	Functional	0.323	113.5	11.4
	Visual	0.343	87.8	8.4
	Combined	0.378	109.3	10.9

Notes: Stage A predicts incidence ($s_{ijt} > 0$). Stage B predicts conditional intensity on positives. Training: 2008–2015; validation for tuning: 2016; test: 2017. Higher PR–AUC is better; lower RMSE and MAE are better.

the consumer narrows down and makes a final choice. This comparative evaluation within the consideration set reflects more serious purchase intent (Hauser and Wernerfelt 1990).

In this section, we test whether visual characteristics contribute to joint consideration over and above functional characteristics. However, we do not observe individual consideration sets. Instead, we look at how frequently products appear together in search, treating this joint appearance as evidence of side–by–side comparison. Following Yang et al. (2022), we proxy "being compared" with Google Trends *pairwise co–searches*. A query of the form "i j" in year t is interpreted as evidence that consumers juxtapose products i and j. For each focal product i and year t, we define a block consisting of all candidate rivals J_{it}, then rank all candidates j by their co–search score s_{ijt} and treat the top k as i's *observed comparison partners*. We use $k = 2$. This yields an interpretable, within–focal measure of how often products are considered side–by–side, consistent with the idea of co–entry into a consideration set.

Prediction task and split. We build a pairwise panel indexed by (i, j, t). Within each (i, t) block we learn scores $\hat{\pi}_{ijt}$ to rank rivals j. Models are trained on 2008–2015, tuned on 2016 (hyperparameters chosen to maximize validation NDCG@2), and evaluated on the held–out 2017 test set. We compare three feature sets—*functional*, *visual*, and *combined*—and four visual representations (Disentangled, VAE, Disentangled+color, and ResNet–50+PCA). We report top–of–ranking metrics that reward getting the *very top* of each block right: (i) a *Top–2 Hit Rate*, which simply counts how many of the top two observed rivals appear in the model's top two (0, 1, or 2, averaged across blocks); (ii) *MAP@2* (Mean Average Precision at cutoff 2) rewards *where* a correct partner appears in the top two. Unlike the Hit Rate (which treats ranks 1 and 2 equally), MAP@2 gives extra credit for placing a true partner first; (iii) *NDCG@2* (Normalized Discounted Cumulative Gain at 2), a rank–sensitive score that applies a logarithmic discount to lower ranks and normalizes by the block's ideal ordering. While both MAP@2 and NDCG@2 are rank–sensitive, only NDCG@2 is directly comparable across blocks with different candidate–set sizes because of this normalization (Manning 2008, Järvelin and Kekäläinen 2002). Next, we formalize these metrics.

Let J_{it} denote the candidate rivals for focal product i in year t. Let $y_{ijt} \in \{0,1\}$ indicate whether j is an observed comparison partner for (i,t), and let $\hat{\pi}_{ijt}$ denote the model score used to rank $j \in J_{it}$. Sorting $\{\hat{\pi}_{ijt}\}_{j\in J_{it}}$ in descending order yields recommendations $j_{(1)}, j_{(2)}, \ldots$. Our baseline evaluates with $k = 2$. Hyperparameters are chosen on the 2016 validation block to maximize NDCG@2.

$$\mathrm{HR@2}(i,t) = \tfrac{1}{2}\big(y_{ij_{(1)}t} + y_{ij_{(2)}t}\big),$$

$$\mathrm{MAP@2} = \mathrm{avg}_{(i,t)}\,\frac{1}{\min(2, m_{it})}\sum_{r=1}^{2} P@r(i,t)\, y_{ij_{(r)}t}, \qquad P@r(i,t) = \frac{1}{r}\sum_{q=1}^{r} y_{ij_{(q)}t},$$

$$\mathrm{NDCG@2} = \mathrm{avg}_{(i,t)}\,\frac{\sum_{r=1}^{2}\frac{2^{y_{ij_{(r)}t}}-1}{\log_2(r+1)}}{\mathrm{IDCG@2}(i,t)}.$$

Here, $j_{(r)}$ is the r-th highest–scoring rival according to the model's predictions $\hat{\pi}_{ijt}$, $y_{ij_{(r)}t} \in \{0,1\}$ indicates whether that rival is an observed comparison partner, $m_{it} = \sum_{j\in J_{it}} y_{ijt}$ is the number of true observed comparison partners in block (i,t), and IDCG@2(i,t) is the ideal DCG for block (i,t) with perfect ranking. All metrics are computed within each (i,t) block and then averaged across blocks.

2017 test results (point estimates). Tables 9–12 summarize performance for Random Forest (RF) and XGBoost–classification (XGB–Cls). Across all visual representations, adding visual information improves recovery of the observed comparison partners, especially for RF. For example, RF NDCG@2 rises from ≈ 0.49 with *functional* features to about 0.56 with *visual* features, with similar gains in HR@2 and MAP@2. XGB–Cls also benefits from visual features, though the gains are smaller.

Interestingly, the *combined* feature set does not always outperform visual features alone for RF models (e.g., Table 9: NDCG@2 of 0.557 vs. 0.552), suggesting that functional

Table 9 Top–2 rival prediction (2017 test) — Disentangled visual (no color)

Model	Feature set	HR@2	MAP@2	NDCG@2
RF	Functional	0.451	0.423	0.487
RF	Visual	0.538	0.512	0.557
RF	Combined	0.519	0.499	0.552
XGB–Cls	Functional	0.353	0.314	0.373
XGB–Cls	Visual	0.347	0.314	0.370
XGB–Cls	Combined	0.387	0.341	0.404

Table 10 Top–2 rival prediction (2017 test) — VAE visual

Model	Feature set	HR@2	MAP@2	NDCG@2
RF	Functional	0.451	0.423	0.487
RF	Visual	0.534	0.514	0.560
RF	Combined	0.523	0.507	0.559
XGB–Cls	Functional	0.353	0.314	0.373
XGB–Cls	Visual	0.343	0.320	0.374
XGB–Cls	Combined	0.362	0.323	0.385

Table 11 Top–2 rival prediction (2017 test) — Disentangled + color

Model	Feature set	HR@2	MAP@2	NDCG@2
RF	Functional	0.451	0.423	0.487
RF	Visual	0.545	0.520	0.566
RF	Combined	0.532	0.512	0.560
XGB–Cls	Functional	0.353	0.314	0.373
XGB–Cls	Visual	0.355	0.318	0.383
XGB–Cls	Combined	0.381	0.347	0.405

Table 12 Top–2 rival prediction (2017 test) — ResNet–50 + PCA (4 PCs)

Model	Feature set	HR@2	MAP@2	NDCG@2
RF	Functional	0.451	0.423	0.487
RF	Visual	0.536	0.512	0.557
RF	Combined	0.519	0.501	0.552
XGB–Cls	Functional	0.353	0.314	0.373
XGB–Cls	Visual	0.434	0.404	0.460
XGB–Cls	Combined	0.400	0.363	0.425

characteristics may add noise when predicting comparison behavior or that the model overfits with both feature types. For XGB–Cls, however, combined features consistently outperform either type alone.

Results using a stricter Top-1 evaluation criterion show that visual features consistently outperform functional features for RF models across all visual representations. However, the optimal feature combination differs between Top-2 and Top-1 evaluation for XGB, suggesting the models capture different aspects of comparison behavior at different ranking positions (see Appendix B).

5.9. Insight 5: Product Level Insights

Ankit: I am not comfortable with this section. The conceptual points are all valid. I worry about the disentanglement algorithm's precision in differentiating make-models. I don't think X5 and X6 look so different as the obtained characteristics seem to suggest.

Brands care about positioning their products in a consistent manner. The positioning maps we have shown can help us obtain insights into the closest competitive set for each product, which in turn are likely to be present together in consumers' consideration sets. These insights are important for both product line design and positioning for a single firm, as well as for competitive positioning with respect to rival firms.

We find concrete insights in the automobile market corresponding to the above points. In Figure 7, we see the market structure maps for SUVs (segment J). First, we note that the BMW X5 and X6 are very close in structured space, making it seem like the brand might face cannibalization risk with weakly differentiated products. However, when we examine the visual characteristics map, we note that there is significant differentiation between these models from the same manufactured, leading to reduced risk of cannibalization. Second, considering the BMW X3 as the focal product, in the map for structured product characteristics, we observe that the Jeep Cherokee seems like the closest competitor, but in the visual product space, the Jeep Grand Cherokee is the closest competitor. Thus, if firms focus only on structured characteristics, BMW might focus more on the Cherokee rather than the Grand Cherokee in making comparisons. However, consumers who value visual characteristics would be more likely to consider the BMW X3 in comparison with the Jeep Grand Cherokee, leading to a less effective marketing strategy.

In Table 13, we observe the structured product characteristics and the disentangled visual characteristics for these specific models discussed above. We can see the BMW X5 and X6 are close in structured characteristics, because other than fuel efficiency, the structured characteristics are within ±10% of each others. However, each of the visual characteristics are quite different for the models. Similarly, comparing the BMW X3 to the Jeep models, the structured characteristics reveal that it may be arguably equally close to the Jeep Cherokee or the Grand Cherokee, but the visual characteristics reveal that it is much closer to the Grand Cherokee on body shape, grille height and grille width.

Table 13 Product Characteristics for a Sample of Segment J (2013 Market)

	Structured Characteristics				Visual Characteristics			
Make Model	Price	MPG	HPWT	Space	Boxiness	Body Shape	Grille Height	Grille Width
BMW								
BMW X3	36.58	4.99	0.51	1.36	-1.42	0.77	-0.81	-0.72
BMW X5	56.05	3.66	0.65	1.55	-0.57	2.13	1.32	0.62
BMW X6	57.12	3.19	0.70	1.66	-1.23	1.33	-1.48	-0.36
Jeep								
Jeep Cherokee	25.57	3.10	0.40	1.28	-1.96	-1.05	0.98	1.58
Jeep Grand Cherokee	48.12	3.28	0.54	1.46	-2.05	1.18	-0.43	-0.58

6. Discussion & Conclusion

Our study introduces a novel approach to market structure analysis that incorporates visual product characteristics alongside traditional functional attributes. By employing disentangled representation learning, we automatically extract interpretable visual dimensions from product images without human labeling, enabling more comprehensive market structure mapping. Our analysis of the UK automobile market (2008-2017) yields several key insights with important implications for marketing theory and practice.

First, our analysis reveals the nuanced relationship between form and function—some visual characteristics (like body shape) are tightly constrained by functional requirements, while others (particularly grille dimensions) remain largely discretionary. This selective interpretation of "form follows function" provides evidence that manufacturers strategically deploy visual design elements across a spectrum from technically-constrained to stylistically-free dimensions.

Second, we find that incorporating visual characteristics significantly enhances differentiation across product segments. When both functional and visual characteristics are considered, product segments show substantially less overlap in market structure maps, revealing how visual design creates categorical boundaries that functional attributes alone cannot capture. This finding challenges the conventional view that product categories are defined primarily by functional similarities, highlighting instead the critical role of visual form in establishing competitive boundaries.

Third, our analysis reveals distinct visual strategies across automobile brands. Some manufacturers (e.g., Audi) maintain tight visual consistency while allowing functional variation across models, using design elements like the signature grille to reinforce brand identity. Others (e.g., Jeep) pursue the opposite approach, maintaining functional similarity while allowing greater visual differentiation across their lineup. These contrasting approaches reflect different strategic responses to the form-function relationship in product design.

Fourth, we demonstrate how crossover vehicles use visual design to bridge traditional segment boundaries despite functional similarities to conventional cars. This finding illuminates how manufacturers strategically employ visual cues to create new market positions without substantially altering mechanical platforms, facilitating category expansion with minimal technical investment.

Fifth, our analysis of consumer search behavior confirms that both visual and functional differences influence comparison shopping patterns. This provides direct evidence that consumers actively consider both dimensions when evaluating product alternatives, validating the importance of including both in comprehensive market structure analyses.

Thus, by providing a systematic approach for extracting interpretable visual characteristics, we enable researchers to incorporate product visuals into formal market structure analysis in a principled way. This addresses a significant gap in existing methodologies that either ignore visual aspects entirely or capture them in ways that lack interpretability or selective control.

The approach we describe here also offers potential advancements for demand estimation. Traditional methods like Berry et al. (1995) models effectively utilize functional characteristics but often overlook visual attributes that significantly influence consumer choices. Our technique enables these visual dimensions to be quantified and incorporated alongside functional characteristics, resulting in a more complete representation of the product space. As demonstrated by Magnolfi et al. (2025), augmenting conventional demand models with embeddings that capture additional product attributes significantly improves demand estimates by better representing how consumers actually perceive and compare products. Our disentangled visual characteristics, when used in conjunction with functional attributes rather than as replacements, can yield more accurate elasticity estimates and better predictions of substitution patterns, particularly in categories where aesthetic considerations drive purchase decisions.

While our study focuses on the automotive sector, the methodology is applicable across diverse product categories where visual design influences consumer choice. Industries like fashion, consumer electronics, home furnishings, and packaged goods could particularly benefit from this approach, as these sectors feature both significant visual differentiation and limited availability of comprehensive characteristic data.

Several limitations and opportunities for future research deserve mention. First, our analysis focuses on front-view images; incorporating multiple viewpoints could provide a more comprehensive representation of product design. Second, while we demonstrate correlations between visual design and market structure, future work could estimate the causal impact of specific visual characteristics on market outcomes through controlled experiments or quasi-experimental designs. Third, our approach could be extended to examine design evolution over time, potentially revealing how visual characteristics respond to competitive dynamics and technological constraints.

In conclusion, our research bridges the gap between unstructured visual data and formal market structure analysis, offering a novel approach that better captures the full spectrum of product differentiation. By incorporating interpretable visual characteristics alongside functional attributes, researchers and practitioners can develop more accurate representations of competitive market structures, potentially leading to improved forecasting, more effective product positioning, and deeper insights into the strategic role of product design in competitive markets.

References

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

Avas I, Allein L, Laenen K, Moens MF (2024) Align macridvae: multimodal alignment for disentangled recommendations. *European Conference on Information Retrieval*, 73–89 (Springer).

Barnard RH (2001) *Road vehicle aerodynamic design-an introduction.*

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.

Bergen M, Peteraf MA (2002) Competitor identification and competitor analysis: a broad-based managerial approach. *Managerial and decision economics* 23(4-5):157–169.

Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *Journal of machine learning research* 13(2).

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

Bettman JR (1979) An information processing theory of consumer choice. *(No Title)* .

Bishop CM (2006) Pattern recognition and machine learning. *Springer google schola* 2:1122–1128.

Blattberg RC, Wisniewski KJ (1989) Price-induced patterns of competition. *Marketing science* 8(4):291–309.

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

Borg I, Groenen PJ (2007) *Modern multidimensional scaling: Theory and applications* (Springer Science & Business Media).

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, Hartley J, Pan Y, Gonzalez R, Papalambros PY (2016a) Balancing design freedom and brand recognition in the evolution of automotive brand styling. *Design science* 2:e9.

Burnap A, Liu Y, Pan Y, Lee H, Gonzalez R, Papalambros PY (2016b) Estimating and exploring the product form design space using deep generative models. *International design engineering technical conferences and computers and information in engineering conference*, volume 50107, V02AT03A013 (American Society of Mechanical Engineers).

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

Choi H, Varian H (2012) Predicting the present with google trends. *Economic record* 88:2–9.

Cox TF, Cox MA (2000) *Multidimensional scaling* (CRC press).

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

DeSarbo WS, Grewal R, Wind J (2006) Who competes with whom? a demand-based perspective for identifying and representing asymmetric competition. *Strategic Management Journal* 27(2):101–129.

DeSarbo WS, Manrai AK, Manrai LA (1993) Non-spatial tree models for the assessment of competitive market structure: an integrated review of the marketing and psychometric literature. *Handbooks in operations research and management science* 5:193–257.

Desmet P, Hekkert P (2007) Framework of product experience. *International journal of design* 1(1):57–66.

Dong Z, Wu Y, Pei M, Jia Y (2015) Vehicle type classification using a semisupervised convolutional neural network. *IEEE transactions on intelligent transportation systems* 16(4):2247–2256.

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*.

Dumoulin V, Visin F (2016) A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285* .

Dzyabura D, Peres R (2021) Visual elicitation of brand perception. *Journal of Marketing* 85(4):44–66.

Erdem T (1996) A dynamic analysis of market structure based on panel data. *Marketing science* 15(4):359–378.

Gabel S, Guhl D, Klapper D (2019) P2v-map: Mapping market structures for large retail assortments. *Journal of Marketing Research* 56(4):557–580.

Garvin DA (1987) Competing on the eight dimensions of quality .

Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ (2010) Predicting consumer behavior with web search. *Proceedings of the National academy of sciences* 107(41):17486–17490.

Goodfellow I, Bengio Y, Courville A (2016) *Deep learning* (MIT press).

Green PE, Rao VR (1969) A note on proximity measures and cluster analysis.

Han S, Lee K (2025) Copyright and competition: Estimating supply and demand with unstructured data. *arXiv preprint arXiv:2501.16120* .

Han S, Schulman EH, Grauman K, Ramakrishnan S (2021) Shapes as product differentiation: Neural network embedding in the analysis of markets for fonts. *arXiv preprint arXiv:2107.02739* .

Hauser JR, Koppelman FS (1979) Alternative perceptual mapping techniques: Relative accuracy and usefulness. *Journal of marketing Research* 16(4):495–506.

Hauser JR, Shugan SM (1983) Defensive marketing strategies. *Marketing Science* 2(4):319–360.

Hauser JR, Wernerfelt B (1990) An evaluation cost model of consideration sets. *Journal of consumer research* 16(4):393–408.

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*.

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

Howard JA, Sheth JN (1969) *The theory of buyer behavior.* (John wiley & sons).

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* .

Hucho WH (2013) *Aerodynamics of road vehicles: from fluid mechanics to vehicle engineering* (Elsevier).

Järvelin K, Kekäläinen J (2002) Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)* 20(4):422–446.

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.

Kang N, Ren Y, Feinberg F, Papalambros P (2019) Form + function: Optimizing aesthetic product design via adaptive, geometrized preference elicitation. *arXiv preprint arXiv:1912.05047* .

Kim H, Mnih A (2018) Disentangling by factorising. *ICML*, 2649–2658.

Kim JB, Albuquerque P, Bronnenberg BJ (2011) Mapping online consumer search. *Journal of Marketing research* 48(1):13–27.

Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

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

Kreuzbauer R, Malter AJ (2005) Embodied Cognition and New Product Design: Changing Product Form to Influence Brand Categorization. *Journal of Product Innovation Management* 22(2):165–176, URL http://onlinelibrary.wiley.com/doi/10.1111/j.0737-6782.2005.00112.x/full, 00000.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25.

Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29(1):1–27.

Landwehr JR, McGill AL, Herrmann A (2011) It's got the look: The effect of friendly and aggressive "facial" expressions on product liking and sales. *Journal of marketing* 75(3):132–146.

Lattin JM, Carroll JD, Green PE (2003) *Analyzing multivariate data*, volume 1 (Thomson Brooks/Cole Pacific Grove, CA).

LeCun Y, Bengio Y, et al. (1995) Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361(10):1995.

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *Journal of Marketing Research* 48(5):881–894.

Li Z, Liu F, Wei Y, Cheng Z, Nie L, Kankanhalli M (2024) Attribute-driven disentangled representation learning for multimodal recommendation. *Proceedings of the 32nd ACM International Conference on Multimedia*, 9660–9669.

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

Liu L, Dzyabura D, Mizik N (2020b) Visual Listening In: Extracting Brand Image Portrayed on Social Media. *Marketing Science* 39(4):669–686, ISSN 0732-2399, URL http://dx.doi.org/10.1287/mksc.2020.1226, publisher: INFORMS.

Liu Y, Li KJ, Chen H, Balachander S (2017) The effects of products' aesthetic design on demand and marketing-mix effectiveness: The role of segment prototypicality and brand consistency. *Journal of Marketing* 81(1):83–102.

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, Poole B, Rätsch G, Schölkopf B, Bachem O, Tschannen M (2020) Weakly-supervised disentanglement without compromises. *International Conference on Machine Learning*, 6348–6359 (PMLR).

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*.

Loken B, Ward J (1990) Alternative approaches to understanding the determinants of typicality. *Journal of Consumer research* 17(2):111–126.

Maas AL, Hannun AY, Ng AY, et al. (2013) Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*, volume 30, 3 (Atlanta, GA).

Magnolfi L, McClure J, Sorensen A (2025) Triplet embeddings for demand estimation. *American Economic Journal: Microeconomics* 17(1):282–307.

Manning CD (2008) *Introduction to information retrieval* (Syngress Publishing,).

Manoogian II J (2013) Vehicle Design Process used at General Motors. 00000.

McAuley J, Targett C, Shi Q, Van Den Hengel A (2015) Image-based recommendations on styles and substitutes. *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 43–52.

McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* .

Murphy KP (2012) *Machine learning: a probabilistic perspective* (MIT press).

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Science* 31(3):521–543.

Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2):307–342.

Newell A, Simon HA, et al. (1972) *Human problem solving*, volume 104 (Prentice-hall Englewood Cliffs, NJ).

Orsborn S, Cagan J, Pawlicki R, Smith RC (2006) Creating cross-over vehicles: Defining and combining vehicle classes using shape grammars. *Ai Edam* 20(3):217–246.

Pironon S, Etherington TR, Borrell JS, Kühn N, Macias-Fauria M, Ondo I, Tovar C, Wilkin P, Willis KJ (2019) Potential adaptive strategies for 29 sub-saharan crops under future climate change. *Nature Climate Change* 9(10):758–763.

Rao VR, Sabavala DJ, et al. (1986) Measurement and use of market response functions for allocating marketing resources. *(No Title)* .

Ringel DM, Skiera B (2016) Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Science* 35(3):511–534.

Rosenbaum MS (2007) *The Jeep people: identity, consumption, and culture in a lifestyle community*. Ph.D. thesis, Indiana University.

Shugan SM (2014) Market structure research. *The History of Marketing Science*, 129–164 (World Scientific).

Sisodia A, Burnap A, Kumar V (2024) Generative interpretable visual design: Using disentanglement for visual conjoint analysis. *Journal of Marketing Research* URL https://doi.org/10.1177/00222437241276736.

Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25.

Sujan M (1985) Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of consumer research* 12(1):31–46.

Sujan M, Dekleva C (1987) Product categorization and inference making: Some implications for comparative advertising. *Journal of consumer research* 14(3):372–378.

Tirunillai S, Tellis GJ (2014) Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of marketing research* 51(4):463–479.

Urban GL, Johnson PL, Hauser JR (1984) Testing competitive market structures. *Marketing Science* 3(2):83–112.

Van Der Maaten L (2014) Accelerating t-sne using tree-based algorithms. *The journal of machine learning research* 15(1):3221–3245.

Van der Maaten L, Hinton G (2008) Visualizing data using t-sne. *Journal of machine learning research* 9(11).

Veryzer Jr RW, Hutchinson JW (1998) The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of consumer research* 24(4):374–394.

Vlasic B (2011) *Once Upon a Car: The Fall and Resurrection of America's Big Three Auto Makers–GM, Ford, and Chrysler* (William Morrow).

Wattenberg M, Viégas F, Johnson I (2016) How to use t-sne effectively. *Distill* 1(10):e2.

Yang L, Luo P, Change Loy C, Tang X (2015) A large-scale car dataset for fine-grained categorization and verification. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3973–3981.

Yang Y, Zhang K, Kannan P (2022) Identifying market structure: A deep network representation learning of social engagement. *Journal of Marketing* 86(4):37–56.

Yap AJ, Wazlawek AS, Lucas BJ, Cuddy AJ, Carney DR (2013) The ergonomics of dishonesty: The effect of incidental posture on stealing, cheating, and traffic violations. *Psychological science* 24(11):2281–2289.

Zhang Y, Zhu Z, He Y, Caverlee J (2020) Content-collaborative disentanglement representation learning for enhanced recommendation. *Proceedings of the 14th ACM conference on recommender systems*, 43–52.

Electronic Companion Supplement

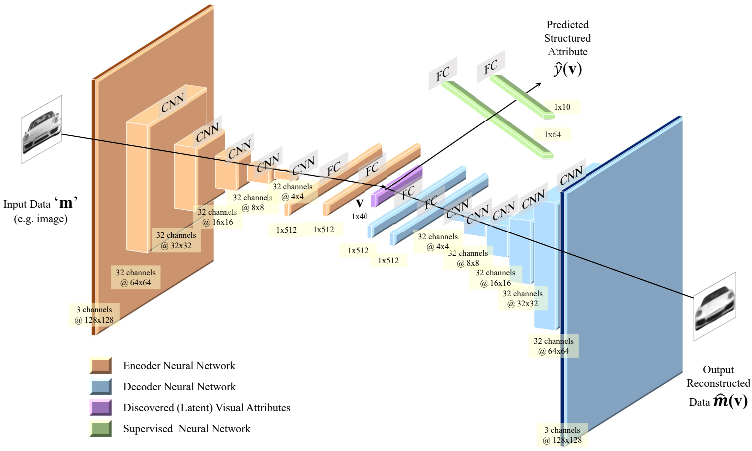
Appendix A: Neural Net Architecture

Figure EC.1 shows the detailed neural net architecture of our model. We modify the architecture proposed by Burgess et al. (2017) to accommodate 128 × 128 pixel images and incorporate the supervisory signals from our model of market equilibrium.

The encoder neural network uses a sequence of convolutional neural network (CNN) layers to learn high-level representations of the input images. CNNs are well-suited for working with image data, as they can effectively capture spatial hierarchies and learn translation-invariant features (LeCun et al. 1995, Krizhevsky et al. 2012). We stack multiple CNN layers in the encoder to progressively learn more complex and abstract visual concepts. The output of the final CNN layer is then flattened and passed through two fully-connected (FC) layers. The first FC layer reduces the dimensionality of the flattened representation, while the second FC layer further compresses the information into a compact set of latent visual characteristics, with a maximum of J dimensions.

The decoder neural network is designed to reconstruct the original image from the latent visual characteristics. Its architecture is essentially the transpose of the encoder network, consisting of FC layers followed by a sequence of transposed convolutional layers (Dumoulin and Visin 2016). The decoder takes the J-dimensional latent visual characteristics as input and gradually upsamples and expands the representation until it reaches the original image size of 128×128 pixels. Finally, the supervised neural network takes the discovered visual characteristics as input and predicts the vector of supervisory signals, which serve as labels for training the model. The supervised network allows the model to learn visual characteristics that are predictive of the supervisory signals, guiding the disentanglement process.

Figure EC.1 Model Architecture



Notes: The encoder neural net for the VAEs consisted of 5 convolutional layers, each with 32 channels, 4×4 kernels, and a stride of 2. This was followed by 2 fully connected layers, each of 512 units. The latent distribution consisted of one fully connected layer of 40 units parameterizing the mean and log standard deviation of 20 Gaussian random variables. The decoder neural net architecture was the transpose of the encoder neural net but with the output parameterizing Bernoulli distributions over the pixels. Leaky ReLU activations were used throughout, which help alleviate the vanishing gradient problem and improve the model's ability to learn complex representations (Maas et al. 2013). We used the Adam optimizer (Kingma and Ba 2014) with the learning rate 5e-4 and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We set batch size equal to 64. We train for 200 epochs to ensure convergence.

Appendix B: Robustness: Top-1 Rival Prediction

To assess robustness, we repeat the rival-prediction analysis using a stricter Top-1 criterion instead of Top-2. For RF models, visual features continue to outperform functional features across all visual representations, confirming the value of visual information in predicting comparison partners. However, the relative performance of combined versus visual-only features differs from Top-2: combined features now consistently achieve the highest performance for RF. For XGB models, the optimal feature combination varies across visual representations, suggesting that the models capture different aspects of comparison behavior depending on whether we focus on the top rival (Top-1) or the top two rivals (Top-2). Tables EC.1–EC.4 report these Top-1 results across all visual representations.

Table EC.1 Top–1 rival prediction (2017 test) — Disentangled visual (no color)

Model	Feature set	HR@1	MAP@1	NDCG@1
RF	Functional	0.477	0.477	0.477
RF	Visual	0.574	0.574	0.574
RF	Combined	0.604	0.604	0.604
XGB–Cls	Functional	0.345	0.345	0.345
XGB–Cls	Visual	0.272	0.272	0.272
XGB–Cls	Combined	0.311	0.311	0.311

Table EC.2 Top–1 rival prediction (2017 test) — VAE visual

Model	Feature set	HR@1	MAP@1	NDCG@1
RF	Functional	0.477	0.477	0.477
RF	Visual	0.591	0.591	0.591
RF	Combined	0.604	0.604	0.604
XGB–Cls	Functional	0.345	0.345	0.345
XGB–Cls	Visual	0.400	0.400	0.400
XGB–Cls	Combined	0.366	0.366	0.366

Table EC.3 Top–1 rival prediction (2017 test) — Disentangled + color

Model	Feature set	HR@1	MAP@1	NDCG@1
RF	Functional	0.477	0.477	0.477
RF	Visual	0.583	0.583	0.583
RF	Combined	0.609	0.609	0.609
XGB–Cls	Functional	0.345	0.345	0.345
XGB–Cls	Visual	0.400	0.400	0.400
XGB–Cls	Combined	0.391	0.391	0.391

Table EC.4 Top–1 rival prediction (2017 test) — ResNet–50 + PCA (4 PCs)

Model	Feature set	HR@1	MAP@1	NDCG@1
RF	Functional	0.477	0.477	0.477
RF	Visual	0.596	0.596	0.596
RF	Combined	0.600	0.600	0.600
XGB–Cls	Functional	0.345	0.345	0.345
XGB–Cls	Visual	0.285	0.285	0.285
XGB–Cls	Combined	0.353	0.353	0.353

Appendix C: Sanity Check: Color and Object Class

To probe what information a standard black-box visual representation is capturing, we conducted a simple sanity check using four car images and one non–car control image. Specifically, we considered front–view images of a red Toyota, red Ferrari, black Toyota, black Ferrari, and a red apple. We passed each image through a pre-trained image embedding model and computed the pairwise correlations between the resulting embeddings. Table **??** reports the resulting correlation matrix (values closer to 1 indicate higher similarity).

Two patterns are worth noting. First, the representation behaves sensibly within the automotive domain: cars of the same make (e.g., red Toyota vs. black Toyota, red Ferrari vs. black Ferrari) exhibit the highest off-diagonal similarities (0.906 and 0.833), and cross–make similarities are somewhat lower. This is consistent with the representation capturing stable, brand- and shape-related information.

Second, the non–car control image (the red apple) is not mapped far away from all cars. In particular, the red Ferrari–red Toyota similarity is 0.727, while the red Ferrari–red apple similarity is 0.655. Although the apple is less similar to the cars than they are to each other, the gap is modest and arguably “too close for comfort.” This suggests that the representation still places substantial weight on low-level visual properties such as color, shading, and curvature, which the red apple shares with the red Ferrari, even though they are semantically unrelated products.

From a market-structure perspective, this is problematic. If such entangled embeddings were used directly to construct a spatial map of competition, products could cluster by superficial attributes like color rather than by function or segment. For example, a “red cluster” might group economy sedans, high-performance sports cars, and even non-vehicle objects such as apples, obscuring the true competitive boundaries defined by body shape, size, and vehicle class.

This sanity check underscores the motivation for our disentanglement approach. By explicitly isolating color as a separate factor of variation (see Section 4), we ensure that the remaining dimensions (e.g., body shape, grille height, boxiness) reflect the underlying form of the vehicle. This, in turn, allows us to construct market-structure maps where proximity reflects genuine design similarity and competitive substitutability, rather than artifacts of shared paint color.

Appendix D: UDR Algorithm

The Unsupervised Disentanglement Ranking (UDR) metric, proposed by Duan et al. (2020), assesses the similarity of disentangled representations learned by different models. The key idea behind UDR is that if two models have learned similar disentangled representations, their informative latent dimensions should exhibit strong correlations.

To compute UDR for a pair of models i and j, we first calculate the correlation matrix R, where each entry $R(a,b)$ represents the correlation between latent dimensions a and b from models i and j, respectively (Equation EC.1). We then identify the most correlated latent dimension in model j for each dimension a in model i, denoted as r_a (Equation EC.2).

$$R(a,b) = cor(v_i(a), v_j(b)) \tag{EC.1}$$

$$r_a = \max_{b \in V(j)} corV(a,b) \tag{EC.2}$$

The UDR score for the pair of models, UDR_{ij}, is computed using Equation EC.3. This equation consists of two symmetric terms, each focusing on one model. The first term considers each informative latent dimension b in model j and calculates the ratio of the squared correlation of its most similar dimension in model i (r_b^2) to the sum of correlations between b and all dimensions in model i. This ratio is then multiplied by an indicator function $I_{KL}(b)$, which equals 1 if dimension b is informative (determined by its KL divergence from the prior distribution) and 0 otherwise. The second term follows the same process, but with the roles of models i and j reversed.

$$UDR_{ij} = \frac{1}{d_i + d_j}\left[\sum_{b \in Z(j)} \frac{r_b^2}{\sum_{a \in Z(i)} R(a,b)} I_{KL}(b) + \sum_{a \in Z(i)} \frac{r_a^2}{\sum_{b \in Z(j)} R(a,b)} I_{KL}(a)\right] \tag{EC.3}$$

The final UDR score is normalized by the total number of informative dimensions in both models ($d_i + d_j$) to ensure that having more informative dimensions does not automatically lead to a higher score. A perfect one-to-one correspondence between the informative dimensions of the two models would result in a UDR score of 1.

Appendix E: Validation Surveys for Interpretability and Quantification

To validate the interpretability of the discovered visual characteristics, we conducted a survey with 93 respondents after removing those who failed attention checks. Respondents were shown a sequence of five images for each visual characteristic, where the characteristic varied from left to right while keeping all other characteristics fixed. They were asked to describe how the car changed the most across the sequence of images. Figure EC.2 presents an example of the open-ended question posed to the respondents.

We then used language models (ChatGPT4 and Claude 3 Opus) to summarize the main themes that respondents agreed upon for each visual characteristic.[15]

Both language models agreed on the summaries, leading us to label the visual characteristics as follows:

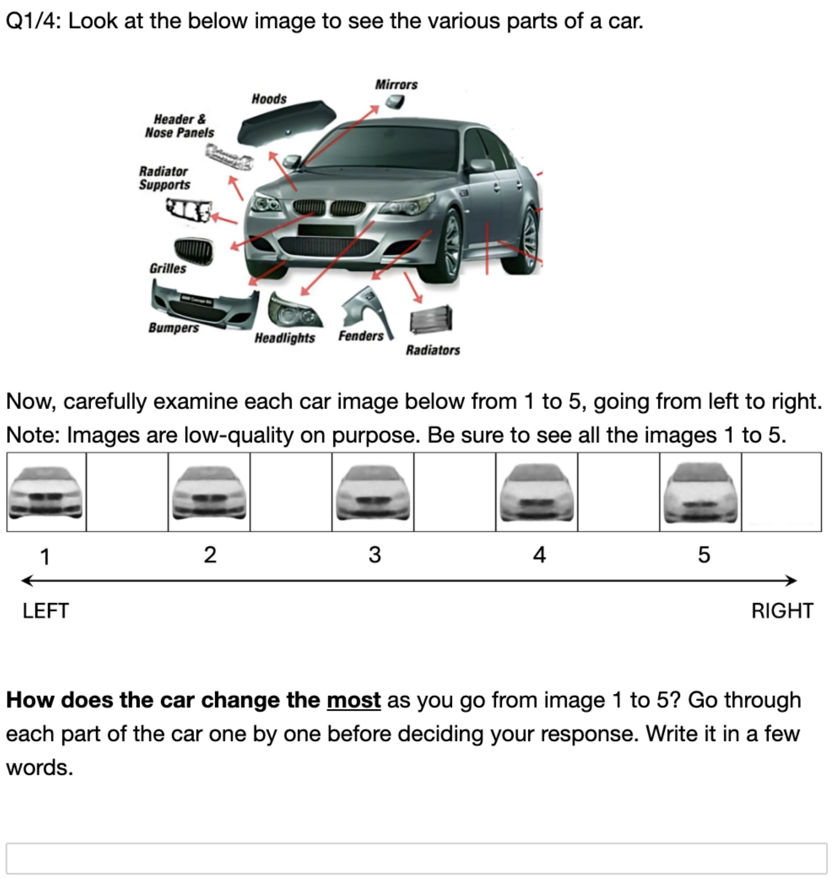
1. Body Shape: The LLM summary of survey respondents states that the "car appears to change in shape, particularly becoming narrower, less angular, and more rounded with each successive image."

2. Color: Automobiles scoring low on this characteristic are darker and vice-versa.

3. Grille Height: The LLM summary of survey respondents indicates that "grilles are become larger, darker, and more defined." Although this summary also mentions the windscreen becoming darker, which is entangled with the grille height, it captures the essence of the characteristic. We interpret this as automobiles scoring low on this characteristic have less prominent grilles, while those scoring high have more prominent, and larger grilles.

[15] LLM Prompt: Summarize the below responses and share the biggest theme that most respondents agree upon? `https://chat.openai.com/c/95ee3363-61a5-4604-8e15-4cd68ac1bae3` or `https://chatgpt.com/c/95ee3363-61a5-4604-8e15-4cd68ac1bae3` or `https://yalesurvey.ca1.qualtrics.com/jfe/form/SV_ai6c1wUD39kZP8O`

Figure EC.2 Interpretability Validation Survey Question



Q1/4: Look at the below image to see the various parts of a car.

Now, carefully examine each car image below from 1 to 5, going from left to right.
Note: Images are low-quality on purpose. Be sure to see all the images 1 to 5.

How does the car change the most as you go from image 1 to 5? Go through each part of the car one by one before deciding your response. Write it in a few words.

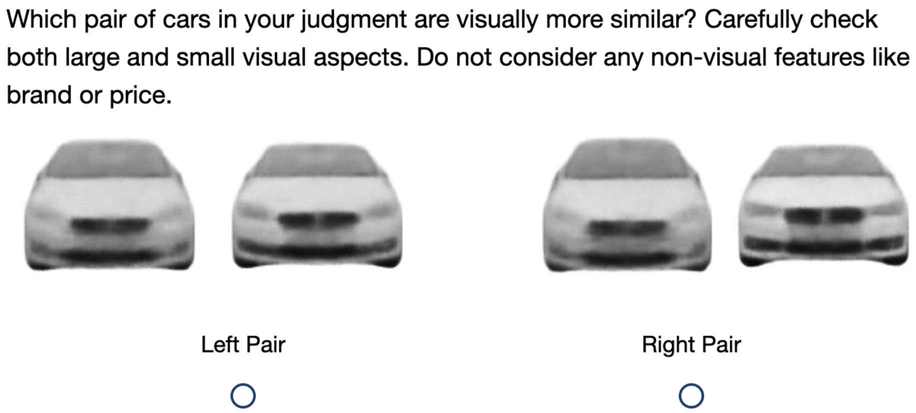
4. Boxiness: The LLM summary of survey respondents describes that the “car becomes lower, flatter, and wider as the sequence progresses.” We interpret this as automobiles scoring low on this characteristic have a high degree of boxiness, characterized by a taller, more upright, and narrower shape. In contrast, those scoring high on this characteristic have a lower degree of boxiness, with a lower, flatter, and wider appearance.

5. Grille Width: The LLM summary of survey respondents states that “grille width become smaller, narrower, and less pronounced as the sequence progresses.” Based on this, we interpret that automobiles scoring low on this characteristic have a wider, more prominent grille, while those scoring high have a narrower, less pronounced grille.

To further validate the quantification of the visual characteristics determined by our method, we conducted a second survey (Figure EC.3). In this survey, we presented respondents with several pairs of automobile images that differed only along one visual characteristic. Respondents were asked to select the pair of automobiles that they perceived as more similar. We then compared the responses to our algorithm’s quantification to assess consistency with human interpretation. For the characteristic we labeled as “Body Shape,” 97% of the 104 respondents agreed with the algorithm’s quantification scale. Similarly, for “Grille Height,” 98%

of the 107 respondents were in agreement. The characteristic we termed "Boxiness" had a 95% agreement rate among 103 respondents, while "Grille Width" saw 93% of the 104 respondents concurring with the algorithm's quantification. Overall, a strong majority of respondents (averaging 96% across four visual characteristics) agreed with the algorithm's quantification scale for the visual characteristics, demonstrating that our method's quantification aligns well with human perception.

Figure EC.3 Quantification Validation Survey Question



Which pair of cars in your judgment are visually more similar? Carefully check both large and small visual aspects. Do not consider any non-visual features like brand or price.

Left Pair ○

Right Pair ○

Appendix F: Additional Examples for Insight 2: Brands have different visual strategies

Figure EC.4 (Color Online) Segment (C, D, and E) in 2013

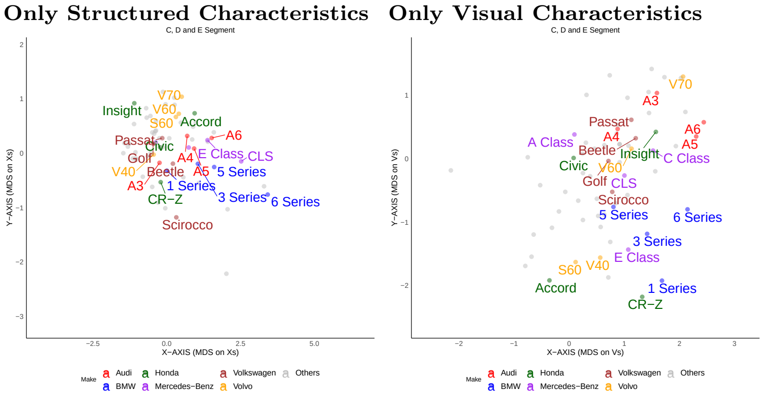
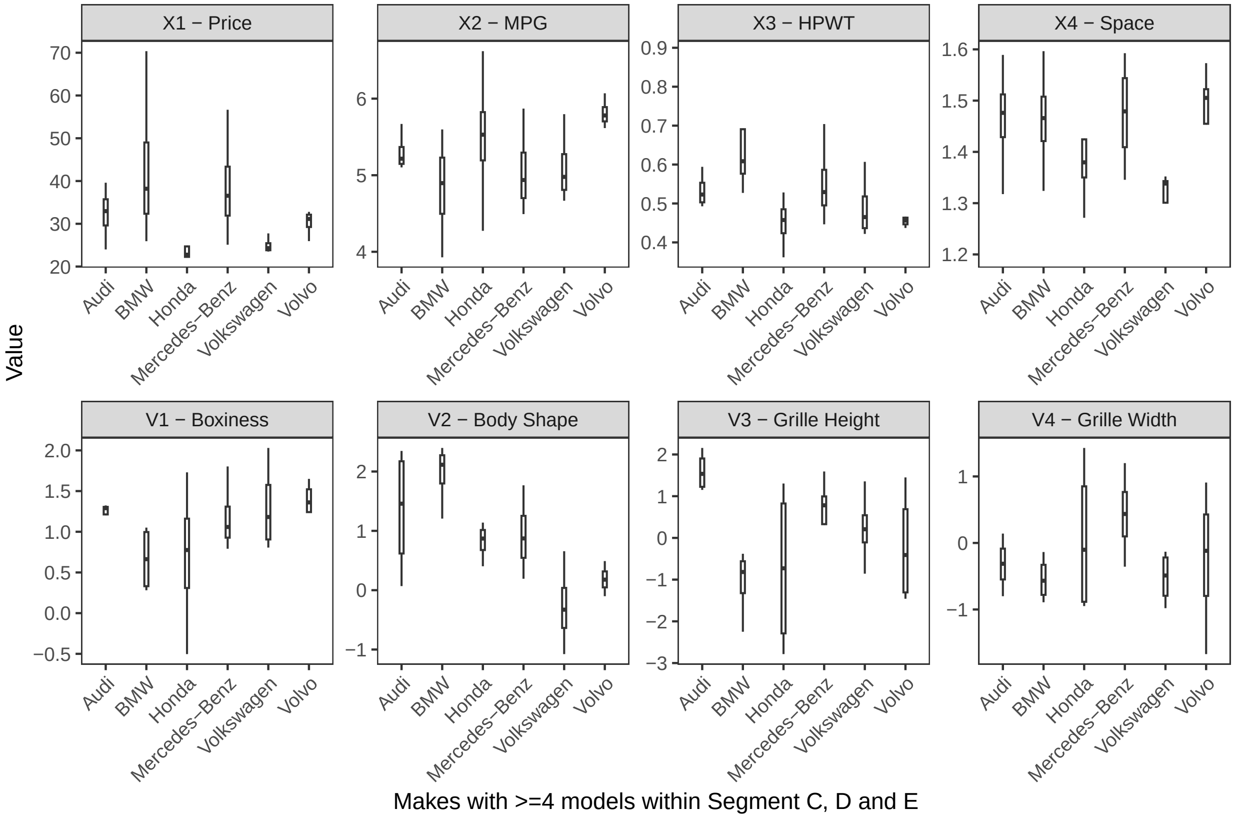


Table EC.5 Area Share of a Make in Structured Space & Visual Space in Segment (C, D, and E) (Sedans)

Make	Models	Area Share (Structured)	Area Share (Visual)	Ratio
Audi	4	3.00%	5.36%	0.56
BMW	4	6.19%	7.15%	0.87
Honda	4	16.26%	33.39%	0.49
Mercedes-Benz	4	4.24%	10.94%	0.39
Volkswagen	4	4.65%	2.05%	2.27
Volvo	4	1.13%	7.89%	0.14

Figure EC.5 (Color Online) Boxplots of 8 Characteristics by Makes with >=4 models in Segment (C, D, and E): Market = 2013



Note: The box represents the interquartile range (IQR), with the horizontal line inside denoting the median, and whiskers extending to approximately ±1.5×IQR. Outliers beyond the whiskers are omitted for clarity.

Appendix G: LLM Prompt

We used GPT–4o via API in a deterministic configuration (temperature = 0, top_p = 1, default penalties). Outputs were required to be valid JSON. No free–text reasoning was requested or stored. Below is the prompt (verbatim) used:

```
Think silently. Temperature=0.


From the review, output a structured JSON with:
```

```
- # of words: Overall, Functional, Visual, Other

- Valence scores (3 to +3): Overall, Functional, Visual, Other

- For front facing visual aspects: bodyshape_hatchback_to_sedan, grille_width, grille_height, boxy_vs_s


<Review Text>
```

Appendix H: Quotes

Ankit to Vineet: Which quotes should we keep? Is it okay to put them in footnotes like this?

> “The Audi grille is one of the most recognisable in the motoring world. Yet, have you noticed how this iconic grille has changed over the decades? Current Audi design chief, Wolfgang Egger, believes that the Audi front grille is a part of the brand’s symbolic identity. Today, current Audi car grilles take the shape of the Singleframe, a word Audi has trademarked to prevent any competitors using it. However, it took several evolutions before ending up in the form it is now.”
>
> https://www.zunsport.com/en/blog/audi-car-grille-history
>
> “Aesthetics play a crucial role in evoking emotions and perceptions in consumers. Some grilles exude strength and power, while others convey elegance and sophistication. Take, for instance, the iconic honeycomb grille of Audi, which has become synonymous with luxury and modernity. The intricate pattern of the grille portrays meticulous attention to detail, while the prominent Audi logo in the center reinforces the brand’s reputation for innovation and performance.”
>
> https://www.xy-tyj.com/a-news-the-influence-of-automotive-grilles-on-brand-identity
>
> “Audi has consistently demonstrated a commitment to design excellence, creating vehicles that are visually striking, modern, and timeless. The brand’s design philosophy emphasizes clean lines, precise proportions, and attention to detail. Audi’s iconic elements, such as the Singleframe grille, distinctive LED lighting, and muscular body contours, have become synonymous with the brand’s identity. The consistent pursuit of aesthetic perfection has made Audi vehicles instantly recognizable and admired worldwide.”
>
> https://thebrandhopper.com/2023/05/23/crafting-dreams-on-wheels-the-audi-story-of-luxury/
>
> “When I introduced the singleframe grille, the clinic’s results showed that consumers liked everything about the car but the grille. Nevertheless, the singleframe grille has become one of the most important design features of the Audi brand, and our customers love it now. Audi has received a distinctive face that everybody recognizes and remembers – something very important for a premium brand. . . . As an example, think of Audi 30 years ago. Audi has always produced very good cars, which was reflected in the claim “Vorsprung durch Technik.” The cars were based on great engineering skills, but they were too rational: made only for the head but not for the heart. It was Audi’s new focus on design that added an emotional component to the brand. The combination of strong rational aspects, such as technology, mechanics, ergonomics and functionality, with the emotional power of design makes the brand so successful now.”

– Direct Quote by Walter de Silva, former head of design for Audi (and later Volkswagen Group) https://www.nim.org/fileadmin/PUBLIC/12_NIM_MIR_Issues/MIR_Marketing_und_Produktdesign/MIR_Marketing_und_Produktdesign_EN/interview_desilva_vol_7_no_2_engl.pdf

"Wolfgang Egger is currently a designer at Chinese brand BYD, but previously acted as the head of Audi design. In his view, Audi's 2000s-era Singleframe (the brand has gone so far as to trademark this word) grille design is the natural progression and culmination of the decades of evolution and iteration before it. Egger claims that with Audi designs from the 1980s and 90s introducing and then gradually giving greater prominence to the lower front grille, it only makes sense to eventually join it with the upper grille, as a way to complete the design and create a uniform whole. Perhaps just as notably is the precedent that the Singleframe design has set for the rest of the industry. The approach has given the confidence for several other brands, such as BMW, to undertake their own interpretation of having a larger grille. Perhaps even more directly inspired by the precedent set by Audi is Lexus, which has developed a somewhat similar, but slightly different take that it calls its 'spindle' grille design. Audi continues to use an evolved version of the Singleframe design today. Unlike the slightly squarish, beard-like original Singleframe designs (as featured on 2000s era Audi A4, A6 and A8 models) where the vertical height of the grille was emphasised, the contemporary interpretation has an angular, aggressive approach. Vaguely resembling a hexagon, the latest Singleframe designs as featured on recent Audi models aims to emphasise the width of the car."

https://www.carexpert.com.au/car-news/from-subtle-to-singleframe-audis-grille-revolution

"The world may be moving toward electrification but Audi is determined to keep its brand identity woven tightly into the thread of its car design. . . . Audi designers have stressed the brand's iconic grille will remain part of the design language despite electric drivetrains making them largely unnecessary. . . . Audi said its grille would even sit prominently on its future electric cars and even the yet to be revealed 'Sphere' concept vehicles, which it says showcases how fully autonomous cars can be turned into a more usable and personal space. . . . In a recent round table discussion, Audi spoke of how this commitment to a large grille would separate its design language from competitors as the industry moved towards electrification, according to US publication Motor Trend. . . . Designers did not reveal how these grilles will be integrated, but Audi's current EV such as the e-tron large SUV uses a largely blocked out grille finished in gloss black to appear similar to its combustion counterpart in the Q7."

https://www.chasingcars.com.au/news/car-technology/audis-future-evs-will-wear-the-iconic-grille-for-form-not-function