AI & ML in Marketing

Vineet Kumar

Yale School of Management

Presenting at: AIM-AMA Sheth Foundation Doctoral Consortium 2023

Generative Interpretable Visual Design

Ankit Sisodia (Purdue), Alex Burnap (Yale) and Vineet Kumar (Yale)

Visual (or aesthetic) design matters across many product categories . . .

Cars

Fashion

Furniture

Visual design matters

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

—JD Power Avoider Study 2015

What this paper seeks to do

Research Goals

Our research aims to obtain **interpretable** visual characteristics (not surprising / outlier) directly from unstructured product images

- *automatically discover* (extract) characteristics

What this paper seeks to do

Research Goals

Our research aims to obtain **interpretable** visual characteristics (not surprising / outlier) directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics

What this paper seeks to do

Research Goals

Our research aims to obtain **interpretable** visual characteristics (not surprising / outlier) directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate* visual design that span the space of visual characteristics

What this paper seeks to do

Research Goals

Our research aims to obtain **interpretable** visual characteristics (not surprising / outlier) directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate* visual design that span the space of visual characteristics

What this paper seeks to do

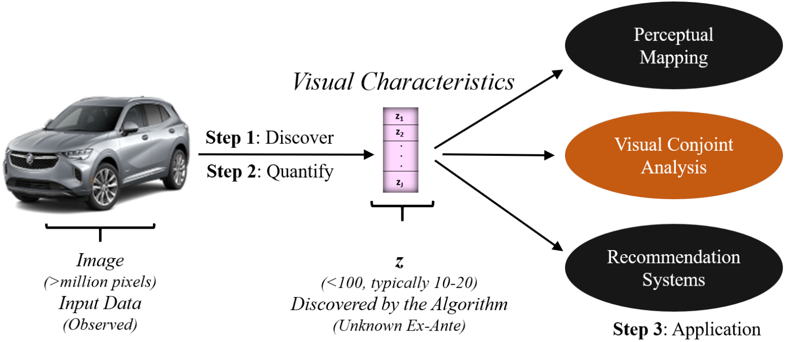
Research Goals

Our research aims to obtain **interpretable** visual characteristics (not surprising / outlier) directly from unstructured product images

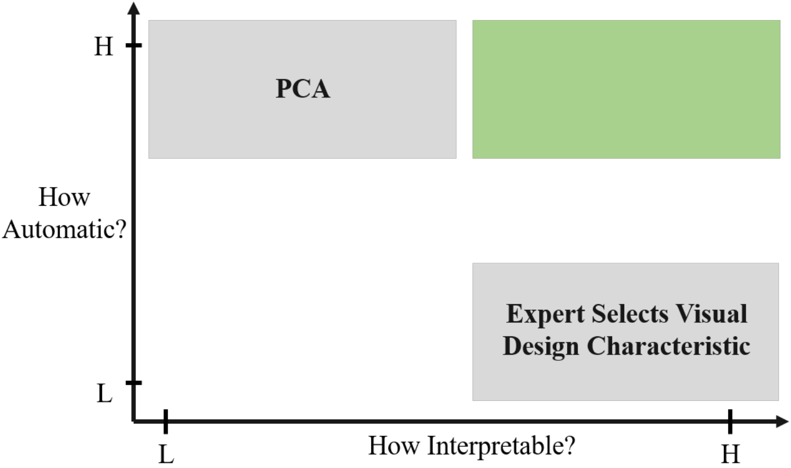
- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate* visual design that span the space of visual characteristics

Hyundai: $(3, 8, 5, 9)$ compared to BMW: $(1, 3, 10, 1)$

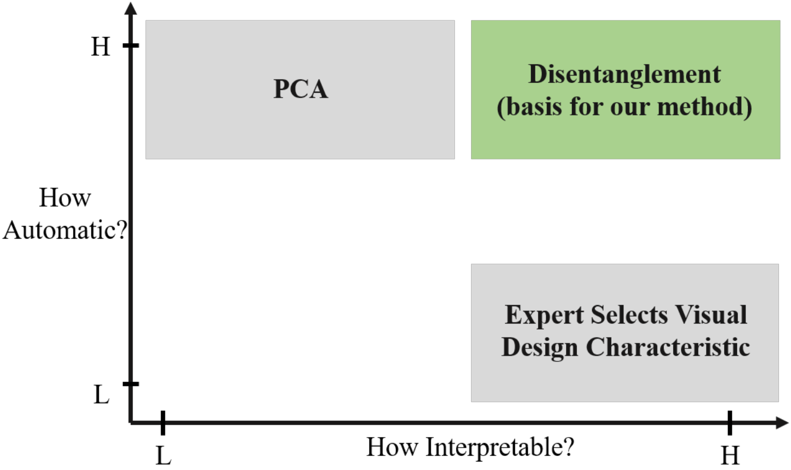
Why Visual Characteristics?



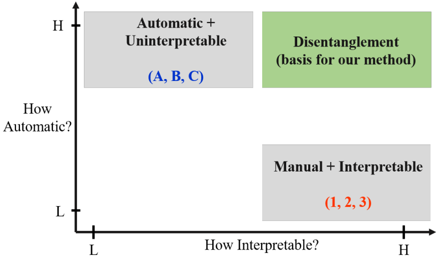
Modeling Visual Characteristics: A comparison of methods



Modeling Visual Characteristics: A comparison of methods



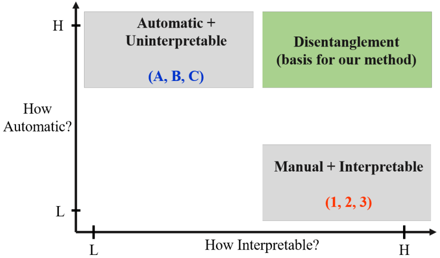
Modeling Visual Characteristics: A comparison of methods



Automatic + Uninterpretable

- A - Bajari, P. L. et al. (2021) : Hedonic prices and quality adjusted price indices powered by AI, *CENMAP working paper*
- B - Law, S., et al. (2019) : Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*
- C - Aubry, S., et al. (2019) : Machine learning, human experts, and the valuation of real assets. *CFS Working Paper Series*

Modeling Visual Characteristics: A comparison of methods



Automatic + Uninterpretable

- A - Bajari, P. L. et al. (2021) : Hedonic prices and quality adjusted price indices powered by AI, *CENMAP working paper*
- B - Law, S., et al. (2019) : Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*
- C - Aubry, S., et al. (2019) : Machine learning, human experts, and the valuation of real assets. *CFS Working Paper Series*

Manual + Interpretable

- 1 - Zhang, M. et al. (2022) : Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from yelp. *Management Science*
- 2 - Liu, Y., et al. (2017) : The effects of products' aesthetic design on demand and marketing-mix effectiveness: The role of segment prototypicality and brand consistency. *Journal of Marketing*
- 3 - Zhang, S., et al. (2021) : What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Science*

What is disentanglement?

Bengio et al (2013)

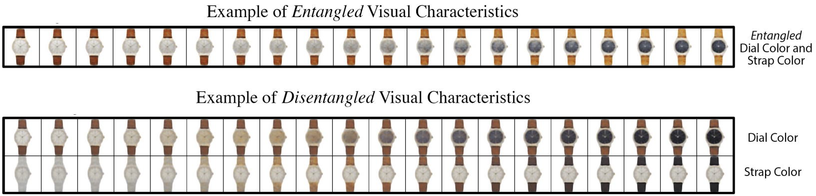
"A disentangled representation can be defined as one where **single latent units** *are sensitive to changes in single generative factors, while being relatively invariant to changes in other factors"*

- Latent Units ($\mathbf{z}$): Dimensions in the model's latent space
- Generative factors ($\mathbf{c}$): Human-interpretable true characteristics

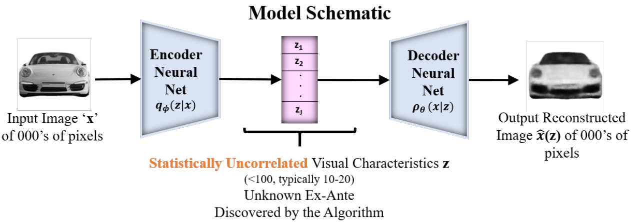
Goal of machine learning process:
Recover latent space and make one-to-one correspondence $\mathbf{c} \longleftrightarrow \mathbf{z}$

Disentangled and Entangled Representations

Goal of machine learning process:
Recover latent space and make one-to-one correspondence $\mathbf{c} \longleftrightarrow \mathbf{z}$

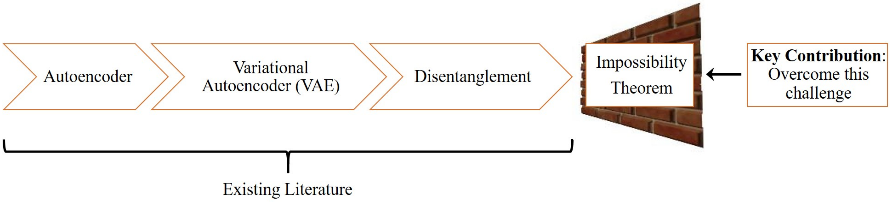


Models in Existing Literature

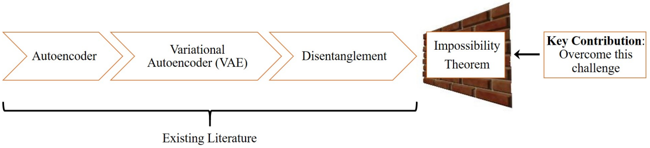


Model	Goal
Autoencoder (AE)	Reconstruction accuracy
Variational Autoencoder (VAE)	...+ structured latent space
Disentanglement	...+ ...+ statistically independent latent space

Roadmap of Our Approach



Impossibility Theorem



Impossibility Theorem

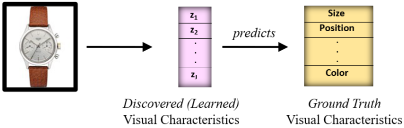
Unsupervised (*i.e. only images*) learning of disentangled representations is *fundamentally impossible* except under certain restrictive conditions.[a]

[a] Locatello, Francesco, et al. "Challenging common assumptions in the unsupervised learning of disentangled representations." ICML. PMLR, 2019.

Implication: Every disentangled representation can have other *infinite* equivalent entangled representations.

Impossibility Theorem – Implications

Common approach to ground truth in ML is to get humans to label[1]



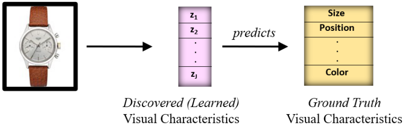
What's the Problem?

- Ground truth on visual characteristics is *unknown*.

[1] Locatello, Francesco, et al. "Disentangling factors of variation using few labels." ICLR. 2020.

Impossibility Theorem – Implications

Common approach to ground truth in ML is to get humans to label[1]

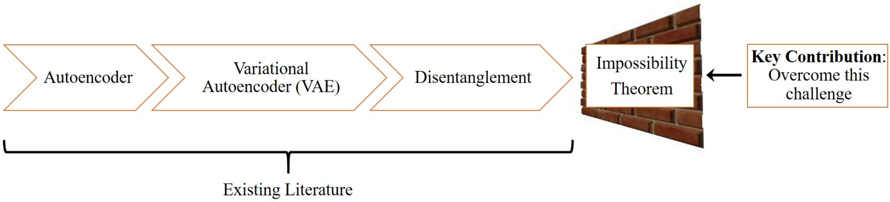


What's the Problem?

- Ground truth on visual characteristics is *unknown*.
- In fact, these are precisely what we want to find.

[1] Locatello, Francesco, et al. "Disentangling factors of variation using few labels." ICLR. 2020.

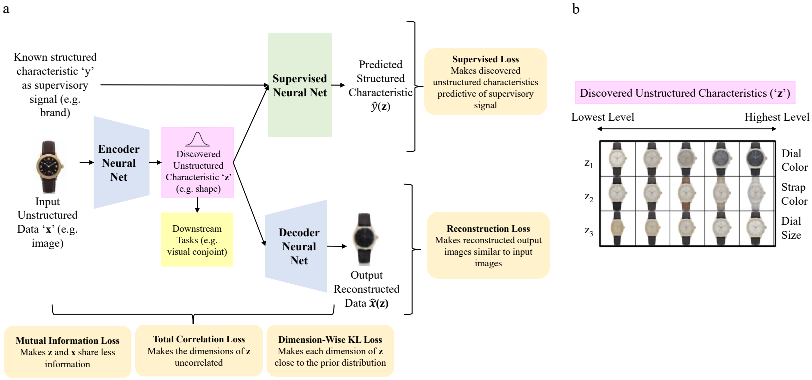
Roadmap of Our Approach



Contribution

We aim to overcome this impossibility theorem without ground truth by using structured product characteristics.

Schematic of Proposed Approach



Model

- Learn model parameters by minimizing loss $L(\theta, \phi; \mathbf{x}, \mathbf{z})$ of integrated model
- θ and ϕ are encoder and decoder parameters; $\mathbf{x}$ are images

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

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

Model

- Learn model parameters by minimizing loss $L(\theta, \phi; \mathbf{x}, \mathbf{z})$ of integrated model
- θ and ϕ are encoder and decoder parameters; $\mathbf{x}$ are images

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

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

Loss Term	Why is this term included?
Reconstruction	Promotes accurate reconstruction of images
Mutual Information	Minimizes redundant information
Total Correlation	Promotes statistical independence between visual characteristics
Dimension-Wise KL	Penalizes deviations from a prior
Supervised	Provides a signal to address the impossibility theorem

Model – Role of Supervised Loss

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

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

- Supervised Loss is used to predict signal from latent representation z: $s = f(z)$

Model – Role of Supervised Loss

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

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

- Supervised Loss is used to predict signal from latent representation z: $s = f(z)$
- Can use structured product characteristics as signals: *brand, price, material etc.*

Model – Role of Supervised Loss

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

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

- Supervised Loss is used to predict signal from latent representation z: $s = f(z)$
- Can use structured product characteristics as signals: *brand, price, material etc.*

Model – Role of Supervised Loss

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

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

- Supervised Loss is used to predict signal from latent representation z: $s = f(z)$
- Can use structured product characteristics as signals: *brand, price, material etc.*

Key Idea

If the supervisory signal is sufficiently correlated with visual characteristics, then it can help obtain the unique (true) disentangled representation

Human Interpretable Characteristics?

- Are these discoveredt visual characteristics human interpretable?
    - Without any domain knowledge about the product category?

Human Interpretable Characteristics?

- Are these discoveredt visual characteristics human interpretable?
  - Without any domain knowledge about the product category?

Starting from the image on the left, **what part of the watch changes the <u>most</u>** as you go from left to right? Carefully check both large and small visual aspects. Go through each part of the watch one by one before selecting any option. Refer to the above image to see parts of the watch.

Note: Images are low-quality on purpose

- ○ Bezel
- ○ Hands
- ○ Crown
- ○ Hour Marker
- ○ Date Window
- ○ Lug
- ○ Dial
- ○ Strap

How is that part of the watch changing?

Visual Characteristics: Quantification?

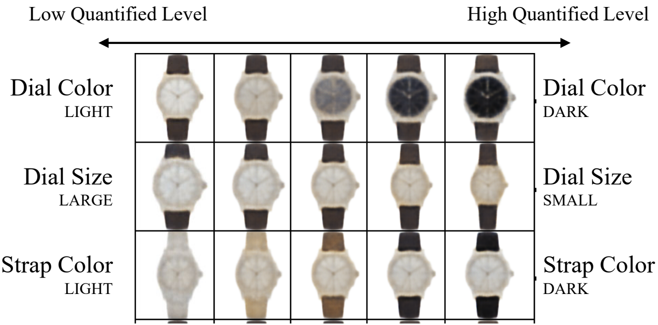
Interpretability and Quantification

Visual characteristic	Interpretability Survey	Quantification Survey
Dial Size	76%	83%
Dial Color	80%	92%
Strap Color	88%	92%
Rim (Bezel) Color	79%	88%
Dial Shape	87%	68%
Knob (Crown) Size	70%	85%

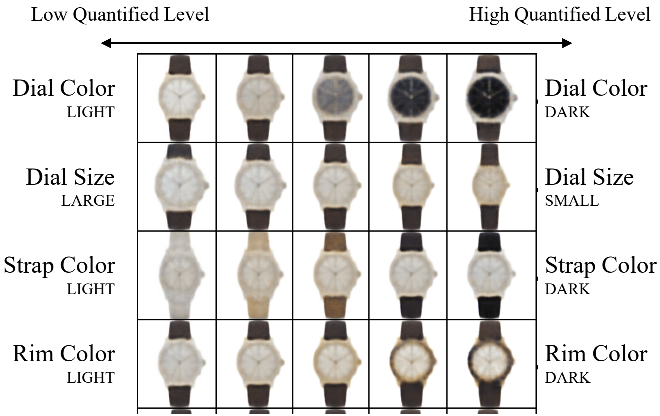
Discovered Visual characteristics

Discovered Visual characteristics

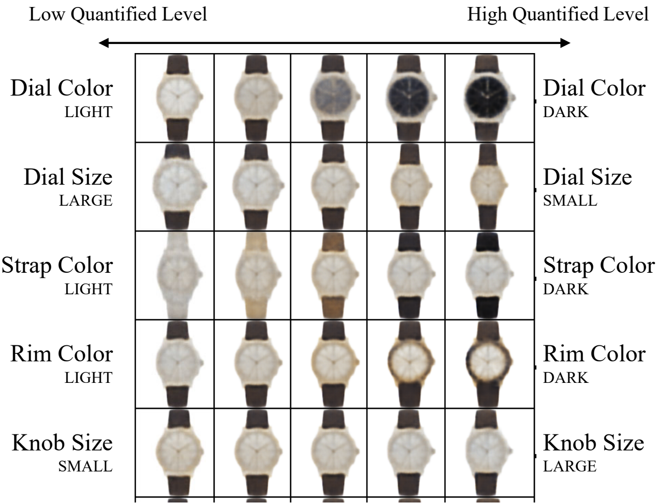
Discovered Visual characteristics



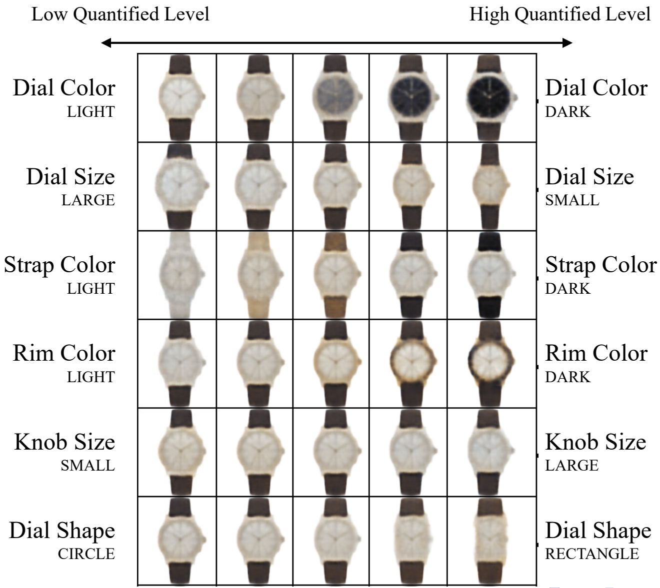
Discovered Visual characteristics



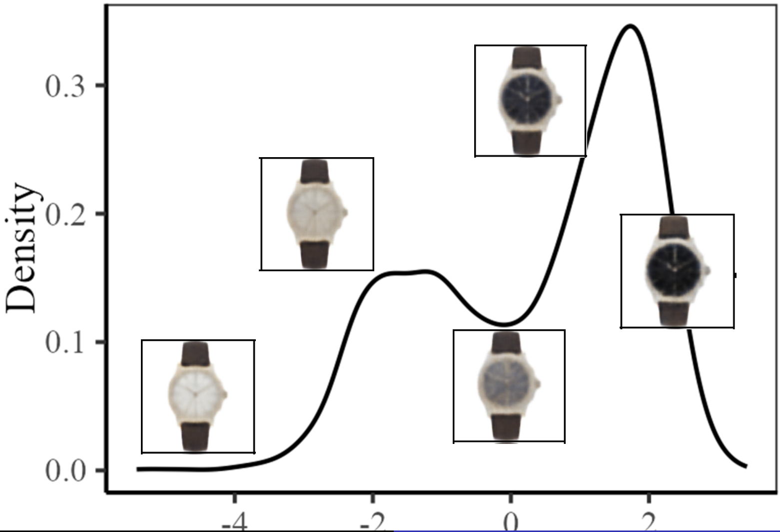
Discovered Visual characteristics



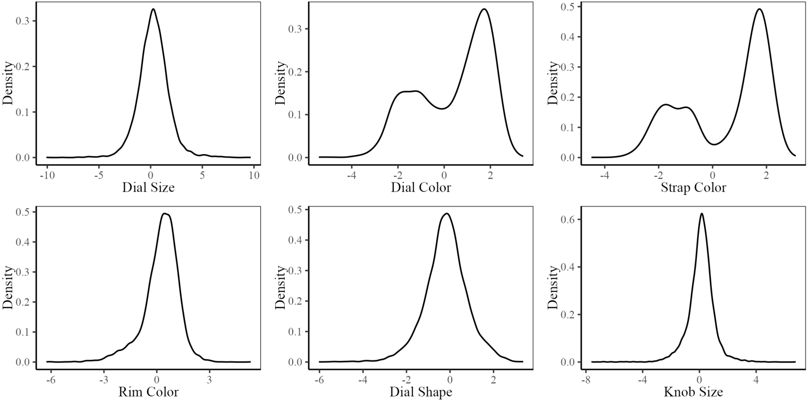
Discovered Visual characteristics



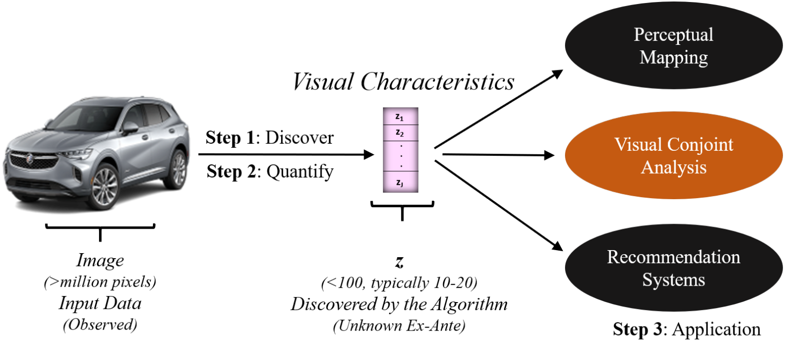
Density of Discovered Visual characteristics (from 'Brand+Material' Signal)



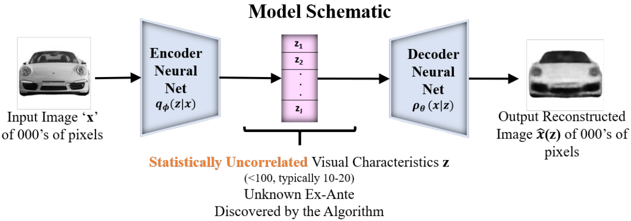
Density of Discovered Visual characteristics (from 'Brand+Material' Signal)



Research Goals

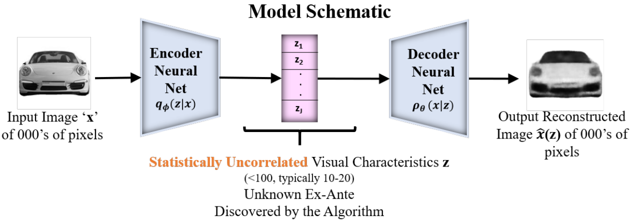


Visual Conjoint Analysis: Background



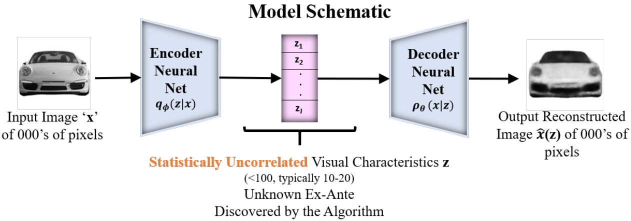
- Visual conjoint has been challenging to do because elements of visual space are correlated

Visual Conjoint Analysis: Background



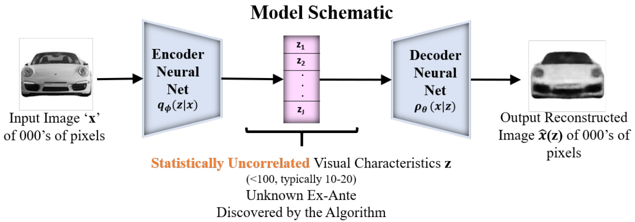
- Visual conjoint has been challenging to do because elements of visual space are correlated
- Designs have always been manually generated by product designers (prototypes)

Visual Conjoint Analysis: Background



- Visual conjoint has been challenging to do because elements of visual space are correlated
- Designs have always been manually generated by product designers (prototypes)
- Our approach generates new never-seen visual designs (counterfactual)

Visual Conjoint Analysis: Background



- Visual conjoint has been challenging to do because elements of visual space are correlated
- Designs have always been manually generated by product designers (prototypes)
- Our approach generates new never-seen visual designs (counterfactual)
- Can span the entire space of visual designs *without being bound by the correlations in the data*.

Example choice-based conjoint (CBC) question in conjoint survey.

Consider the two watches below that vary **only on visual style**. Of these two, which watch would you prefer more (for yourself)?

Select

Select

Next

Utility: Hierarchical Bayesian Model

$$u(\mathbf{z}; \beta_i) = \beta_1 z_1 + \ldots + \beta_K z_K$$

$$\begin{aligned}
\mu_\Theta &\sim \mathcal{N}(\mathbf{0}, \sigma_\Theta^2) \\
\boldsymbol{\Theta} &\sim \mathcal{N}(\mu_\Theta, \boldsymbol{\Lambda}_\Theta) \\
\boldsymbol{\Omega}_\beta &\sim \mathrm{LKJ}(\eta) \\
\boldsymbol{\Lambda}_\beta &= \mathbf{D}(\sigma_\beta)\boldsymbol{\Omega}_\beta \mathbf{D}(\sigma_\beta) \\
\beta_i &\sim \mathcal{N}(\boldsymbol{\Theta}^T \mathbf{r}_i, \boldsymbol{\Lambda}_\beta) \\
u_i^j &= z_j \beta_i + \epsilon_{ij} \\
y_i^{j,j'} &\sim \mathrm{Bernoulli}(\omega_i(j, j')) \\
\text{where} \quad \omega_i(j, j') &= \frac{\exp(u_i^j)}{\exp(u_i^j) + \exp(u_i^{j'})}
\end{aligned}$$

where $\mathrm{LKJ}(\eta)$ is a Cholesky factorization of the correlation matrix $\boldsymbol{\Omega}_\beta$ of the individual "part-worth" preference vector over visual characteristics. $\mathbf{D}(\cdot)$ denotes a diagonal matrix, $\mathbf{r}_i$ are consumer covariates, u_i^j is the utility customer i gets from watch design j, and ϵ_{ij} is a Gumbel random variable. The Bernoulli probability parameter $\omega_i(j, j')$ is specified by the logit function, and $\{j, j'\}_i$ denotes the set of all pairwise choice comparisons for watches $j, j' \in J$ that customer i chose over in the conjoint survey. Note that σ_Θ^2, $\boldsymbol{\Lambda}_\Theta$, η are researcher-defined hyperparameters chosen via model selection using prediction accuracy on the validation data split as the evaluation metric.

Conjoint Model Accuracy

Generated Watches

Model	Out-of-Sample Hit Rate (SD)
Disentangled Embedding + Logit Model (–)	63.16% (2.34%)
Disentangled Embedding + Neural Net (–)	65.81% (2.22%)
Pretrained Deep Learning Model Embedding (O)	68.31% (1.54%)
Disentangled Embedding + Neural Net (O)	67.52% (0.92%)
Disentangled Embedding + Random Forest (O)	68.77% (0.90%)
Disentangled Embedding + XGBoost (O)	69.10% (0.41%)
Disentangled Embedding + HB Model (O + U)	**71.61% (1.87%)**
Disentangled Embedding + HB Model + Interactions (O + U)	70.68% (1.35%)

- Pretrained Deep learning model is trained on *millions of images*, and has millions of parameters
- Our Hierarchical Bayes (HB) model has a small number parameters, and all predictions are based on only 6 visual characteristics
- With 6 visual characteristics, our HB model outperforms the pretrained deep neural net

Ideal Point

- Marketing Literature has conceptualized the notion of ideal point (DeSarbo, Ramaswamy, and Cohen 1995).
- Optimal positioning of a product in the space of characteristics
  - In this study, visual characteristic space
- Can also do this across researcher-defined segments

Generated Ideal Point Watches for Two Segments

Ideal Point: Optimal positioning of a product in characteristic space based on preferences of a selected consumer segment.

Segment 1:
"Ideal Point" Watch Design

Segment 2:
"Ideal Point" Watch Design

Segment 1	Young moderately-affluent females
Segment 2	Older males

Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate* visual design that span the space of visual characteristics

Applications

We then used the model to:

- generate new counterfactual designs to obtain consumer preferences over visual characteristics.
- obtain ideal point visual designs corresponding to different consumer segments