Generative Interpretable Visual Design

Application to Visual Conjoint and Market Structure Mapping

Vineet Kumar

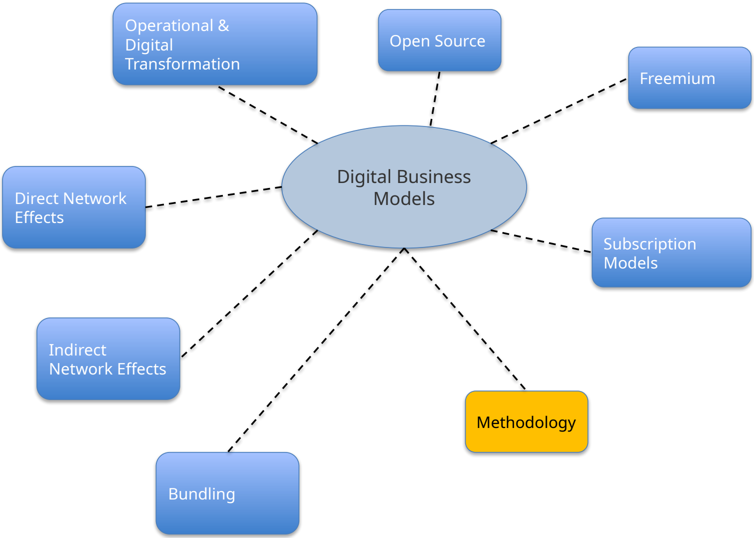
Yale School of Management

Presenting at:

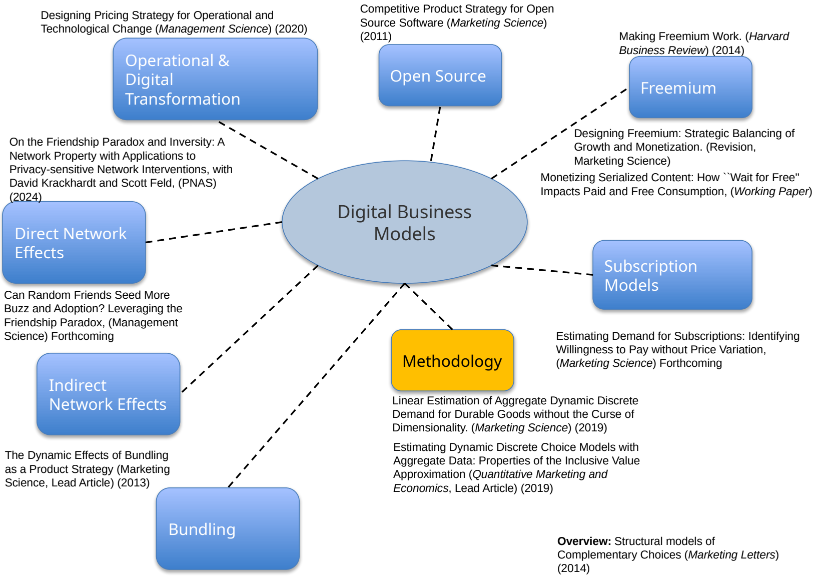
University of Notre Dame Marketing Seminar

May 2025

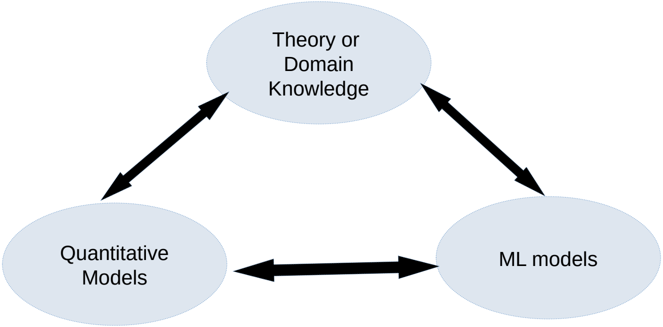
Research Overview



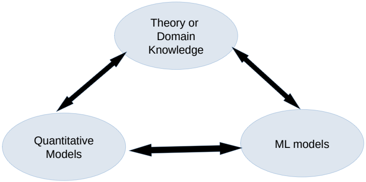
Research Overview



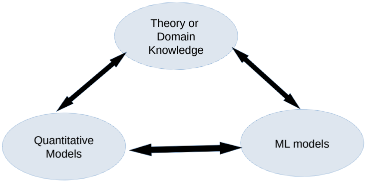
Research Overview



Research Overview

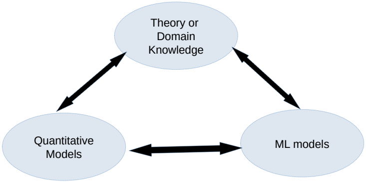


Research Overview



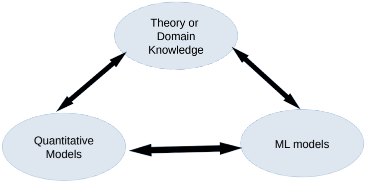
- Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. Forthcoming at (Marketing Science)

Research Overview



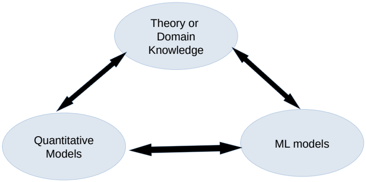
- Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. Forthcoming at (Marketing Science)
- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. Forthcoming at (Marketing Science)

Research Overview



- Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. Forthcoming at (Marketing Science)
- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. Forthcoming at (Marketing Science)
- Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. Forthcoming at (Journal of Marketing Research)

Research Overview



- Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. Forthcoming at (Marketing Science)
- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. Forthcoming at (Marketing Science)
- Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. Forthcoming at (Journal of Marketing Research)
- Market Structure Mapping with Visual Characteristics. (Research in progress)

Generative Interpretable Visual Design

Application to Visual Conjoint and Market Structure Mapping

Ankit Sisodia, Alex Burnap and Vineet Kumar

Yale School of Management

Presenting at:

University of Notre Dame Marketing Seminar

May 2025

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

Cars

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

Cars

Fashion

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

Cars

Fashion

Furniture

. . . even for mundane categories like yogurt

> *"We worked hard to get the packaging right . . . American yogurt has always been sold in containers with relatively narrow openings. In Europe yogurt containers are wider and squatter, and that's what I wanted for Chobani."*
>
> *—Hamdi Ulukaya, Founder & CEO, Chobani*

Consumer Preferences for Visual Design

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

What about preferences in visual space?

Cannot do this because characteristics for visual design are unknown!

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\Longrightarrow$ Preferences over characteristics makes this problem feasible.

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

What about preferences in visual space?

Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products $\implies$ Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

What about preferences in visual space?

Cannot do this because characteristics for visual design are unknown!

What this research seeks to do

Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products

What this research seeks to do

Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics

What this research seeks to do

Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics
- *generate* visual designs that span the space of visual characteristics

What this research seeks to do

Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics
- *generate* visual designs that span the space of visual characteristics

What this research seeks to do

Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics
- *generate* visual designs that span the space of visual characteristics

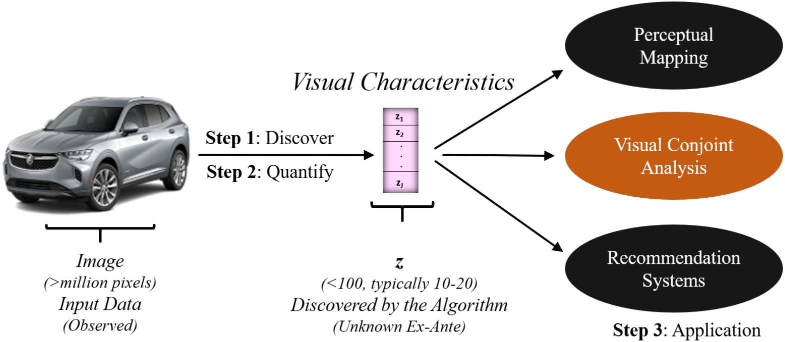
Hyundai: $(3, 8, 5, 9)$ compared to BMW: $(1, 3, 10, 1)$

Research Goals

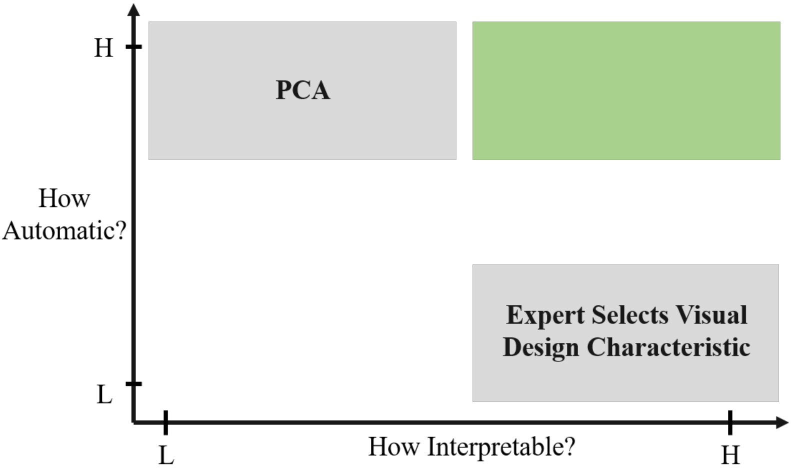
Hyundai: $(3, 8, 5, 9)$ compared to BMW: $(1, 3, 10, 1)$
Several questions come to mind:

- What does the first number represent? Does 3 mean something different from 1?
- Can humans interpret these numeric values?
- What domain knowledge does the model need to have?

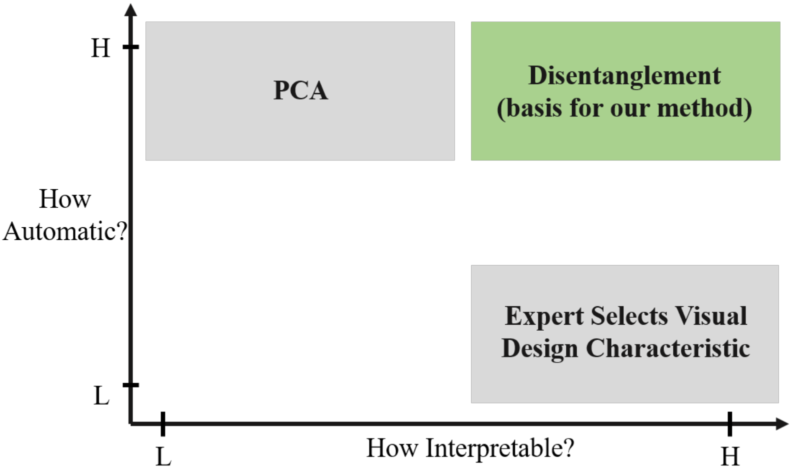
Why Visual Characteristics?



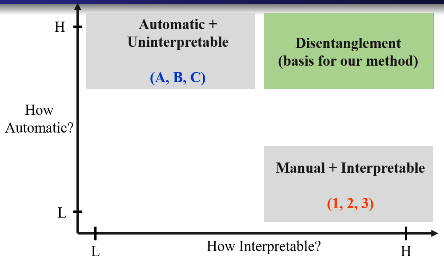
Modeling Visual Characteristics: A comparison of methods



Modeling Visual Characteristics: A comparison of methods



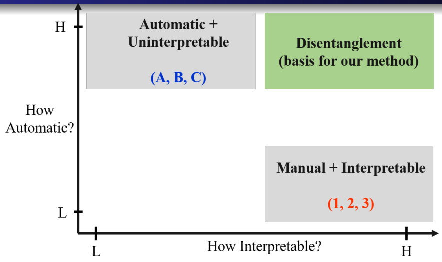
Modeling Visual Characteristics: A comparison of methods



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"*

What is disentanglement?

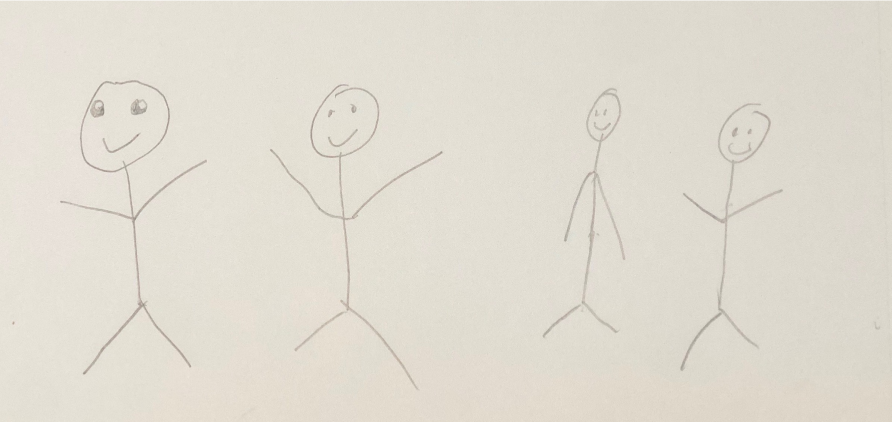
Bengio et al (2013)

"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"*

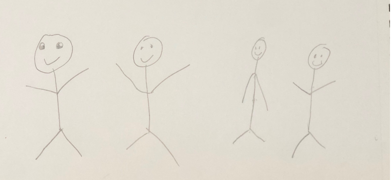
- Latent Units ($\mathbf{z}$): Dimensions in the model's latent space
- Generative factors ($\mathbf{c}$): Human-interpretable true characteristics

Idea: Reality or Data generating process is compositional based on generative factors.

What is disentanglement?
Stick



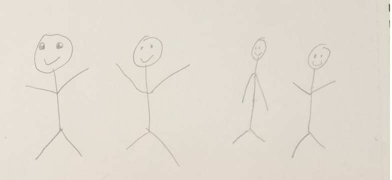
What is disentanglement?



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"*

What is disentanglement?

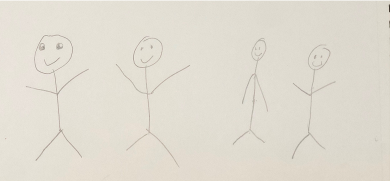


Bengio et al (2013)

"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 ($\mathbf{z}$): What algorithm discovers – dimensions in the model's latent space
- Generative factors ($\mathbf{c}$): Human-interpretable characteristics

What is disentanglement?



Bengio et al (2013)

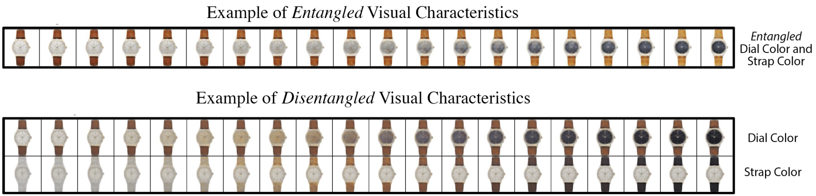
> *"A disentangled representation can be defined as one where* **single latent units** *are sensitive to changes in single generative factors, while being relatively invariant to changes in other factors"*

- Latent Units ($\mathbf{z}$): What algorithm discovers – dimensions in the model's latent space
- Generative factors ($\mathbf{c}$): Human-interpretable characteristics

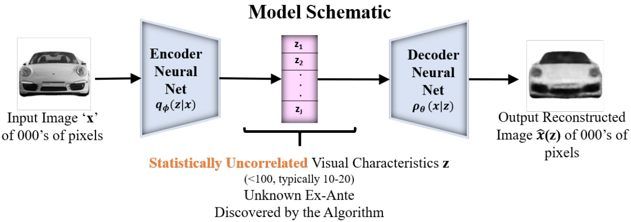
Goal: One to one mapping between $z \leftrightarrow c$

Product Images and Parts of Watch

Disentangled and Entangled Representations

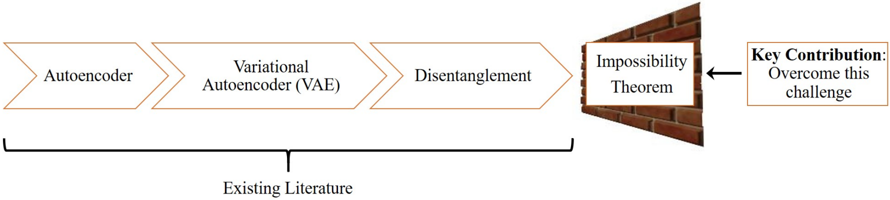


Models in Existing Literature



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

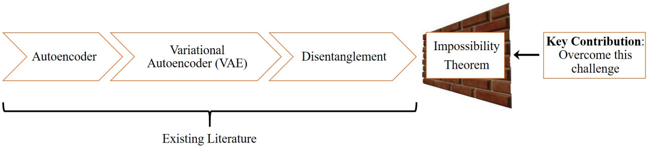
Roadmap of Our Approach



Contribution

We aim to overcome this impossibility theorem with a simple approach of using structured product characteristics.

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.

ML Approach to Impossibility Theorem

Impossibility: Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility

ML Approach to Impossibility Theorem

Impossibility: Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal

ML Approach to Impossibility Theorem

Impossibility: Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- ML methods assume that ground truth is known by researchers

ML Approach to Impossibility Theorem

Impossibility: Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- ML methods assume that ground truth is known by researchers
    - How?

ML Approach to Impossibility Theorem

Impossibility: Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

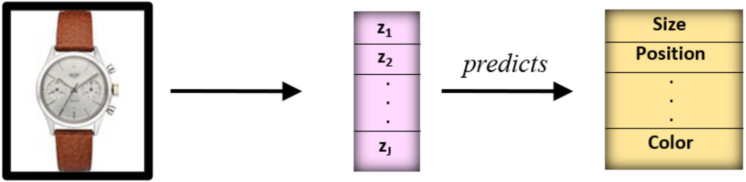
- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- ML methods assume that ground truth is known by researchers
  - How?
- Human labeling

ML Approach to Impossibility Theorem

Impossibility: Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- ML methods assume that ground truth is known by researchers
    - How?
- Human labeling
- Can we use this approach to discover visual characteristics?

Impossibility Theorem – Implications

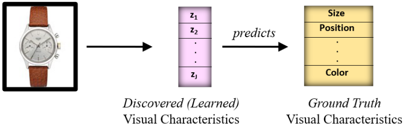


Discovered (Learned)
Visual Characteristics

Ground Truth
Visual Characteristics

Impossibility Theorem – Implications

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



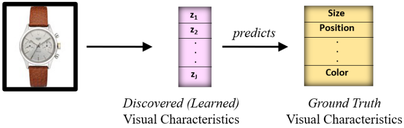
What's the Problem?

- 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.

Impossibility Theorem – Implications

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



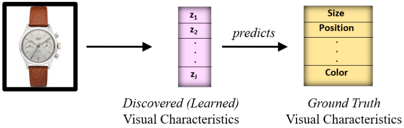
What's the Problem?

- Ground truth on visual characteristics is *unknown*. In fact, these are precisely what we want to find.
- Researcher needs to determine what are the *true characteristics* to focus on $\implies$ not Automatic

[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]

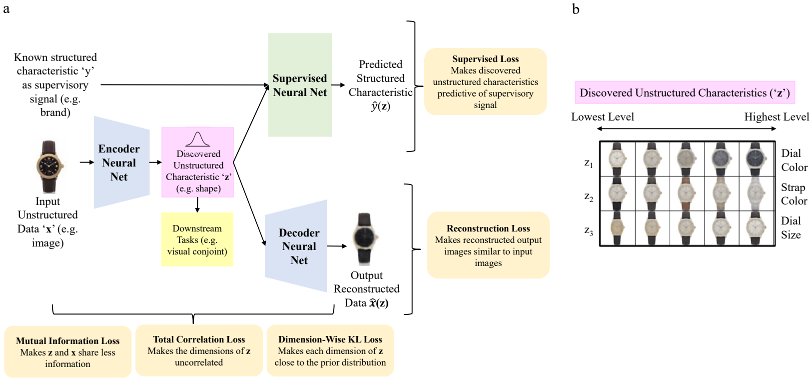


What's the Problem?

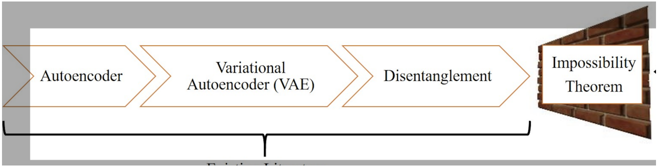
- Ground truth on visual characteristics is *unknown*. In fact, these are precisely what we want to find.
- Researcher needs to determine what are the *true characteristics* to focus on $\Longrightarrow$ not Automatic
- Need to ensure humans understand what these labels are and *how to quantify them* for each image

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

Schematic of Proposed Approach



Contribution



- **Solution** without ground truth on visual characteristics:
- Leverage **structured product characteristics** to provide a supervisory signal for disentanglement

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

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

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

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

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

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

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

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.

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.

Idea to Overcome Impossibility Theorem

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

Why might brand aid the disentanglement model?

Brand as a Supervisory Signal

Idea: Brands have a specific "look" that can be correlated with visual appearance (and therefore visual characteristics)

Why might brand aid the disentanglement model?

Brand Perception

- . . . Cartier has many case shapes from round and oval to cushion-shaped, tonneau, and of course, the many square-shaped or rectangular-shaped Tank watches.[a]

[a] https://www.prestigetime.com/blog/rolex-vs-cartier.html

[b] https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html

[c] https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html

Why might brand aid the disentanglement model?

Brand Perception

- . . . Cartier has many case shapes from round and oval to cushion-shaped, tonneau, and of course, the many square-shaped or rectangular-shaped Tank watches.[a]
- Patek (Philippe) is more conservative and classic on the design front, which is great since most people looking for a Patek (Philippe) are looking for a dress watch.[b]

[a] https://www.prestigetime.com/blog/rolex-vs-cartier.html
[b] https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html
[c] https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html

Why might brand aid the disentanglement model?

Brand Perception

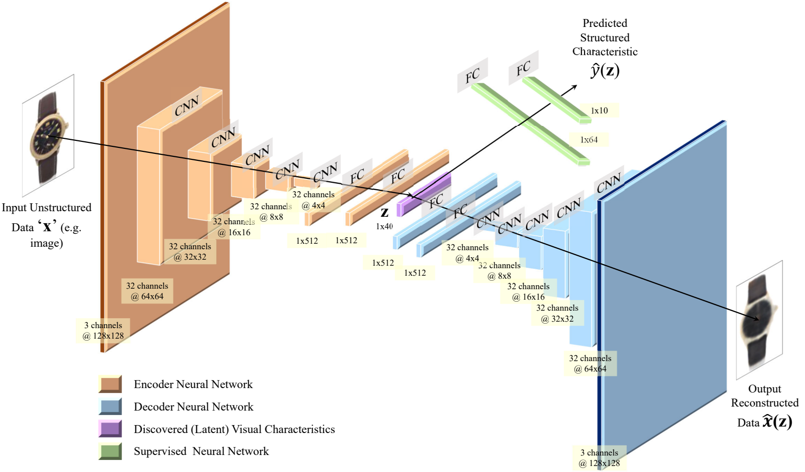
- . . . Cartier has many case shapes from round and oval to cushion-shaped, tonneau, and of course, the many square-shaped or rectangular-shaped Tank watches.[a]
- Patek (Philippe) is more conservative and classic on the design front, which is great since most people looking for a Patek (Philippe) are looking for a dress watch.[b]
- Rolex is much more well-known for its highly-functional and iconically-designed sports and tool watches . . . [c]

[a] https://www.prestigetime.com/blog/rolex-vs-cartier.html

[b] https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html

[c] https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html

Model Architecture



Evaluating Visual Characteristics

Human Interpretable Characteristics?

- UDR indicates disentanglement, but are these visual characteristics human interpretable?
- Without any domain knowledge about the product category?

Human Interpretable Characteristics?

- UDR indicates disentanglement, but are these 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 most** 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: Interpretability?

Do humans agree with the model's quantification?

- Show two pairs of visual designs: (A, B) and (C, D)
- If the model says pair (A, B) are more similar than pair (C, D), do humans agree?

Visual Characteristics: Quantification?

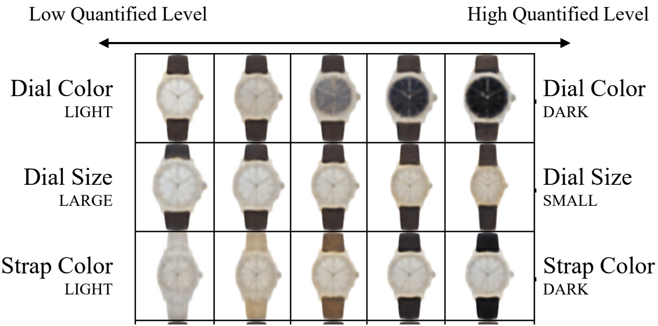
Interpretability and Quantification

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%

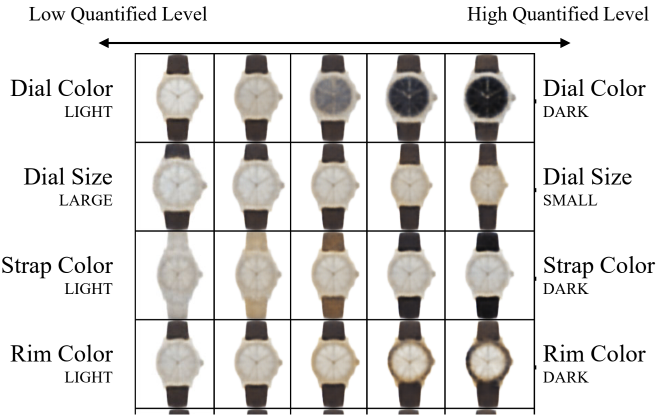
Discovered Visual characteristics

Discovered Visual characteristics

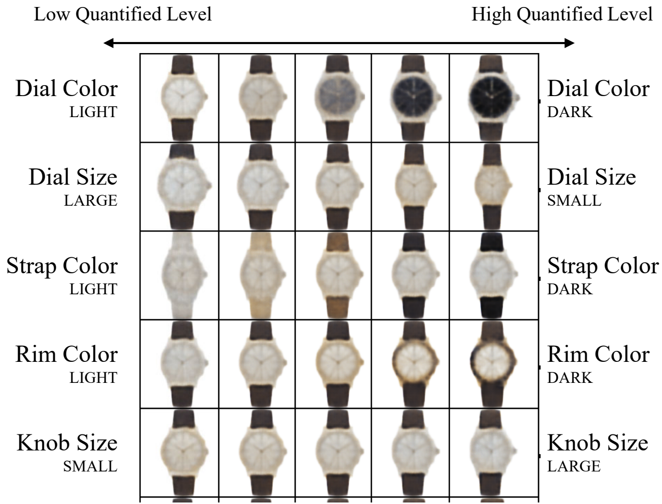
Discovered Visual characteristics



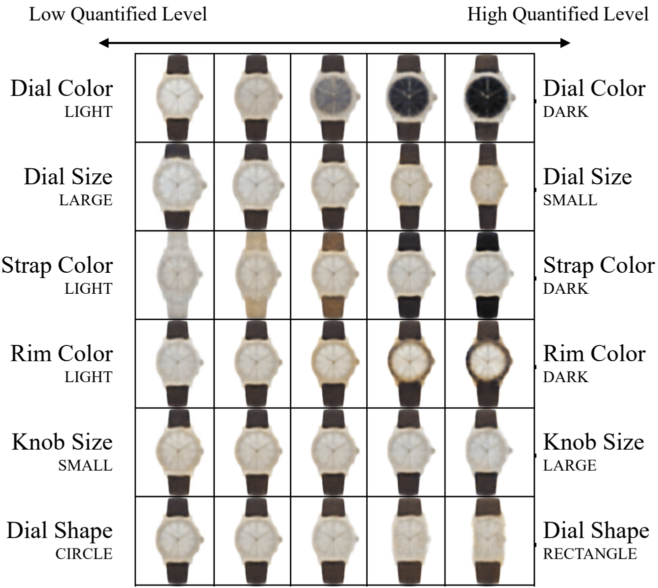
Discovered Visual characteristics



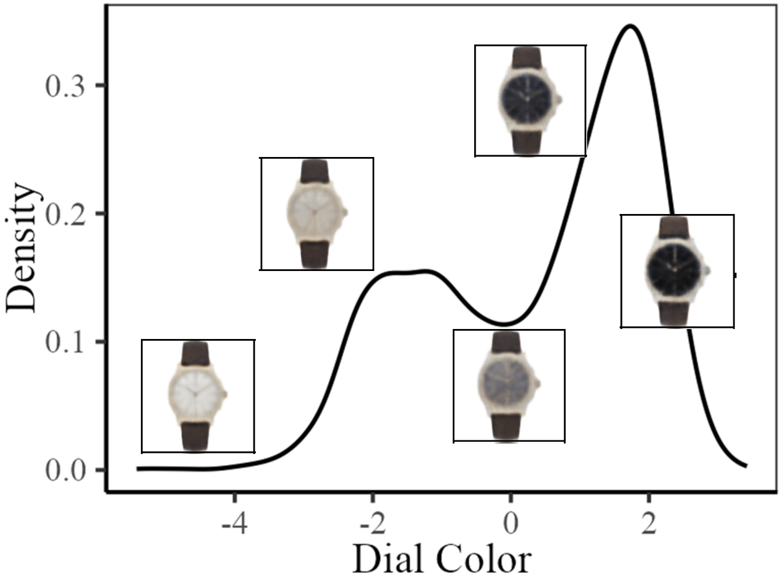
Discovered Visual characteristics



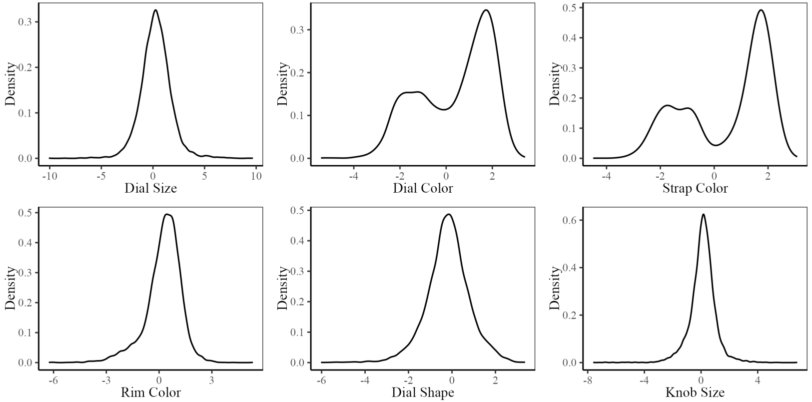
Discovered Visual characteristics



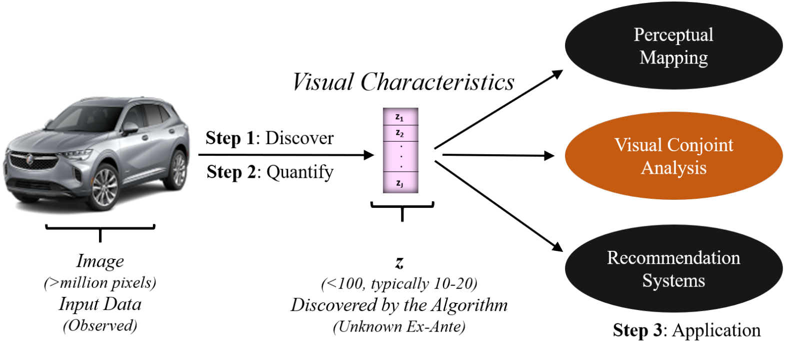
Density of Discovered Visual characteristics

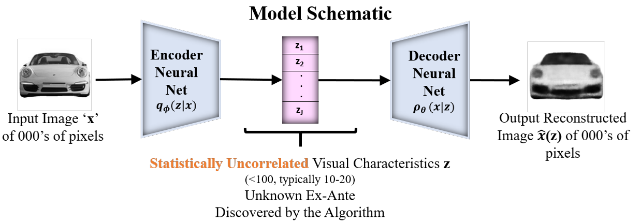


Density of Discovered Visual characteristics



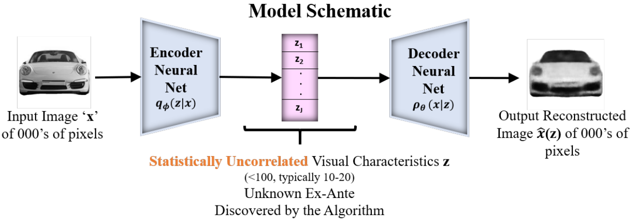
Research Goals





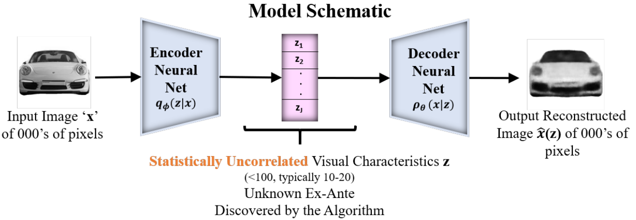
- Visual conjoint has been challenging to do because elements of visual space are correlated

Visual Conjoint Analysis: Background



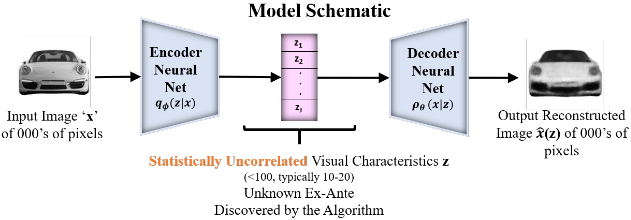
- Visual conjoint has been challenging to do because elements of visual space are correlated
- Designs have always been manually generated by product designers (prototypes)

Visual Conjoint Analysis: Background



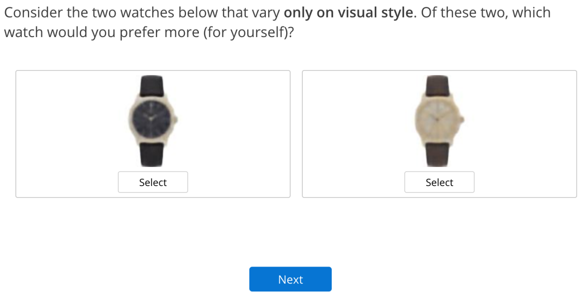
- Visual conjoint has been challenging to do because elements of visual space are correlated
- Designs have always been manually generated by product designers (prototypes)
- Our approach generates new never-seen visual designs (counterfactual)

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.



Utility: Hierarchical Bayesian Model

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

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

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^2_Θ, $\boldsymbol{\Lambda}_\Theta$, 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

Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover* (extract) characteristics

Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

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

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

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

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