

# Generative Interpretable Visual Design

## Application to Visual Conjoint

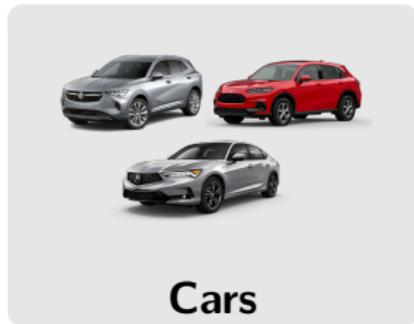
Ankit Sisodia<sup>1</sup>, Alex Burnap<sup>2</sup> and Vineet Kumar<sup>2</sup>

<sup>1</sup>Purdue University

<sup>2</sup>Yale School of Management

Presenting at:  
*Santa Clara University*  
March 2024

# 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*

# Visual design matters



# 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)

# Research Goals

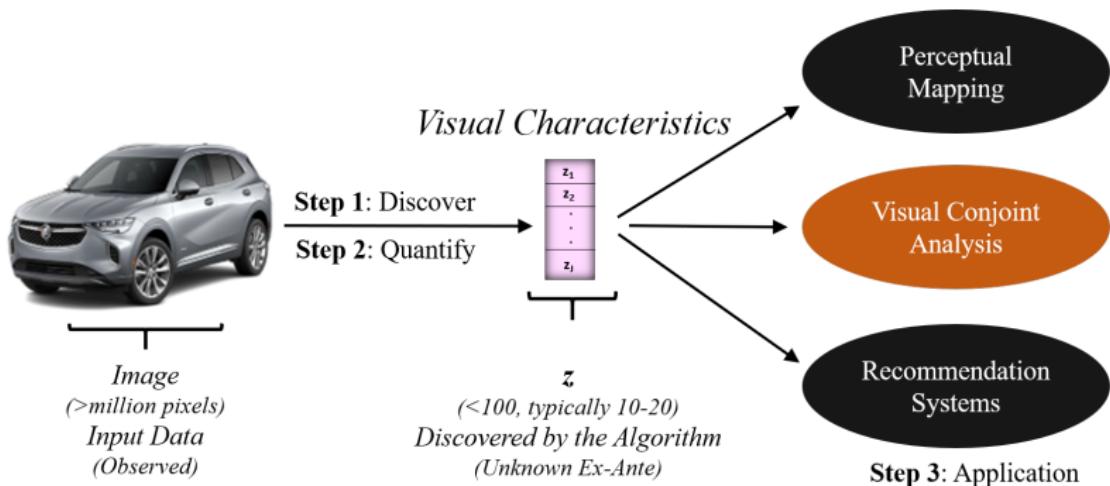


Hyundai: (3, 8, 5, 9) compared to BMW: (1, 3, 10, 1)

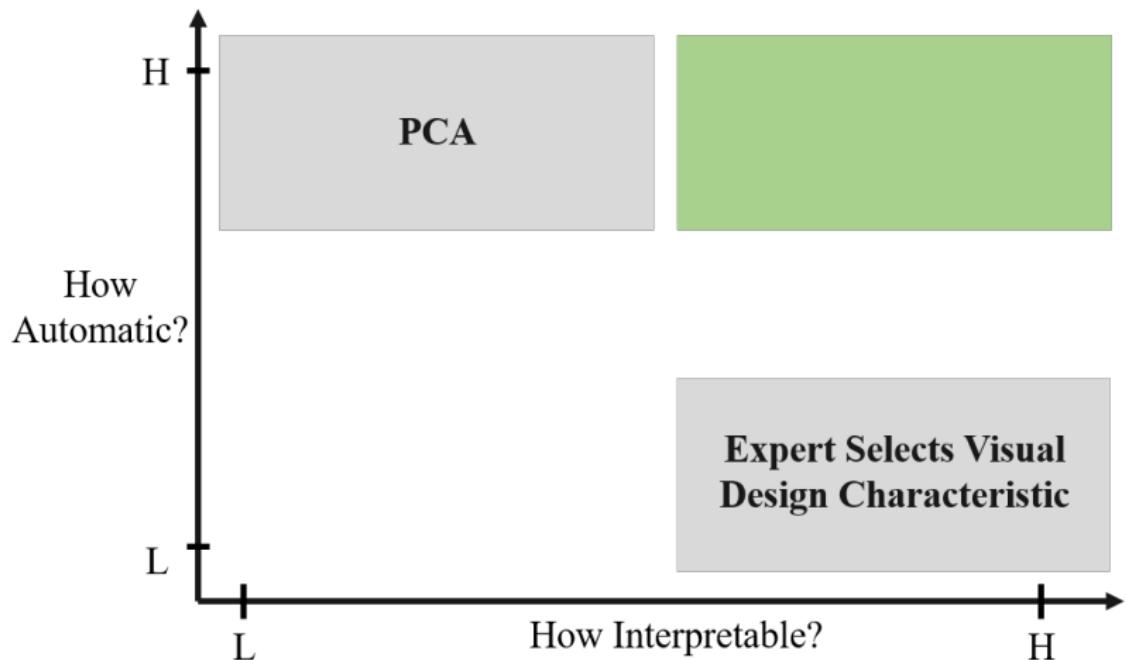
Several questions come to mind:

- 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?

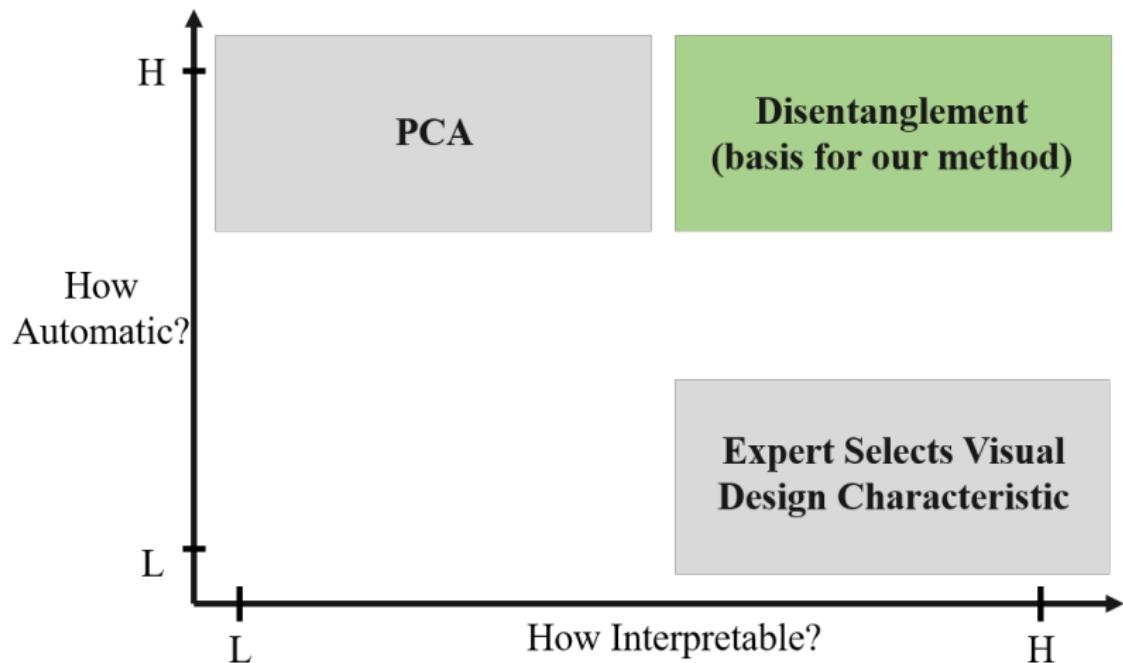
# Why Visual Characteristics?



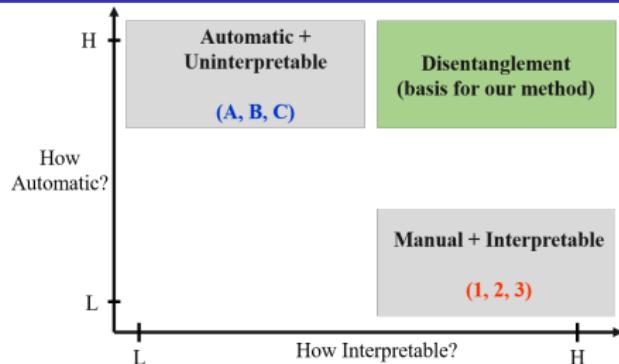
# Modeling Visual Characteristics: A comparison of methods



# Modeling Visual Characteristics: A comparison of methods



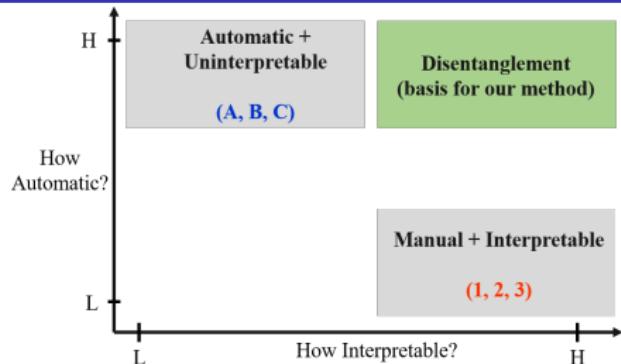
# Modeling Visual Characteristics: A comparison of methods



## 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

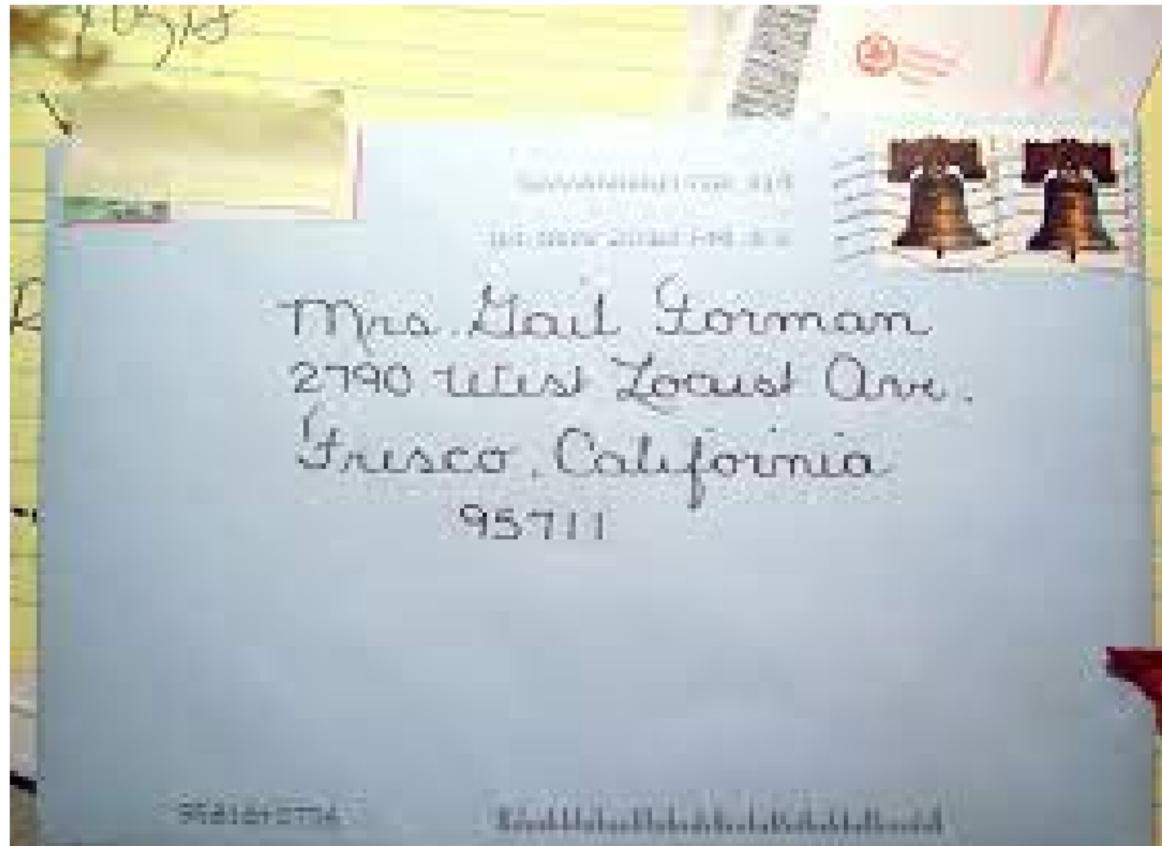
- A - Bajari, P. L. et al. (2021) : Hedonic prices and quality adjusted price indices powered by AI, *CENMAP working paper*
- B - Law, S., et al. (2019) : Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*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

## Manual + Interpretable

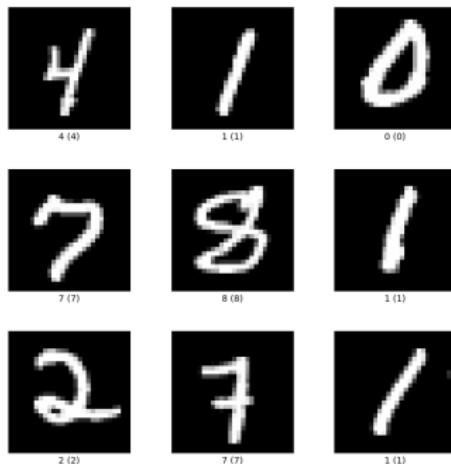
- 1 - Zhang, M. et al. (2022) : Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from yelp. *Management Science*
- 2 - Liu, Y., et al. (2017) : The effects of products' aesthetic design on demand and marketing-mix effectiveness: The role of segment prototypicality 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*



# Is Human Interpretability always necessary?



# Is Human Interpretability always necessary?



# 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

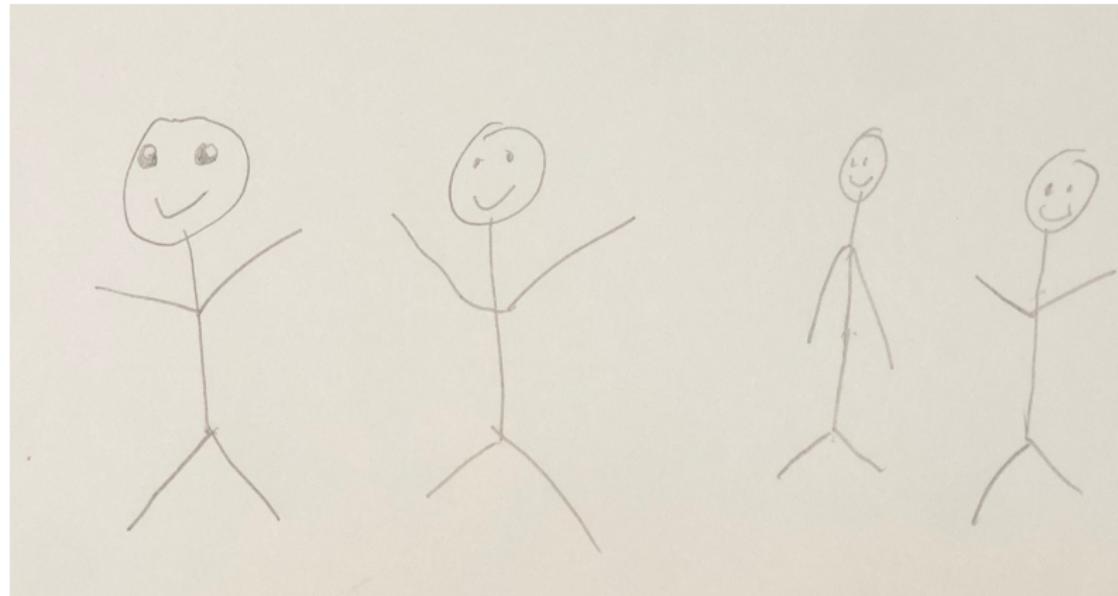
Idea: Reality or Data generating process is compositional based on generative factors.

## What is disentanglement?

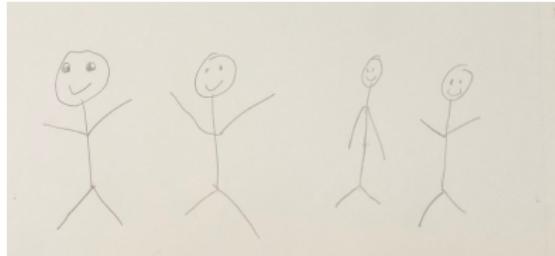
## Stick

# What is disentanglement?

Stick



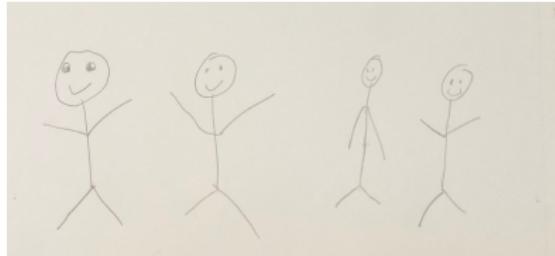
# What is disentanglement?



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?

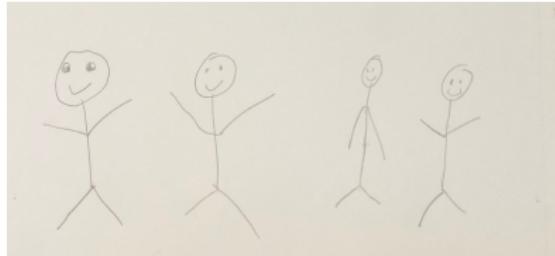


Bengio et al (2013)

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

# What is disentanglement?



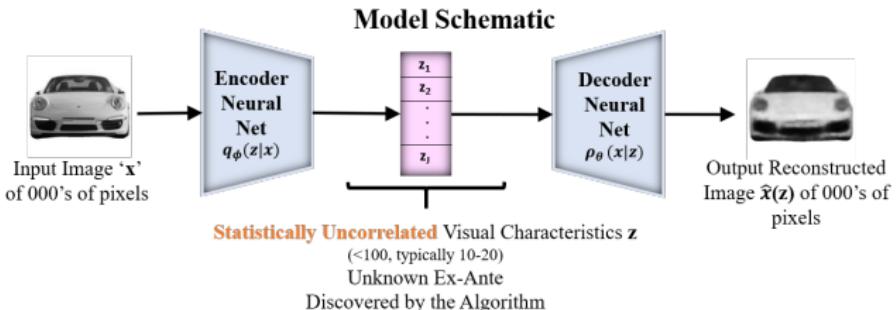
Bengio et al (2013)

*"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 ( $z$ ): Dimensions in the model's latent space
- Generative factors ( $c$ ): Human-interpretable characteristics

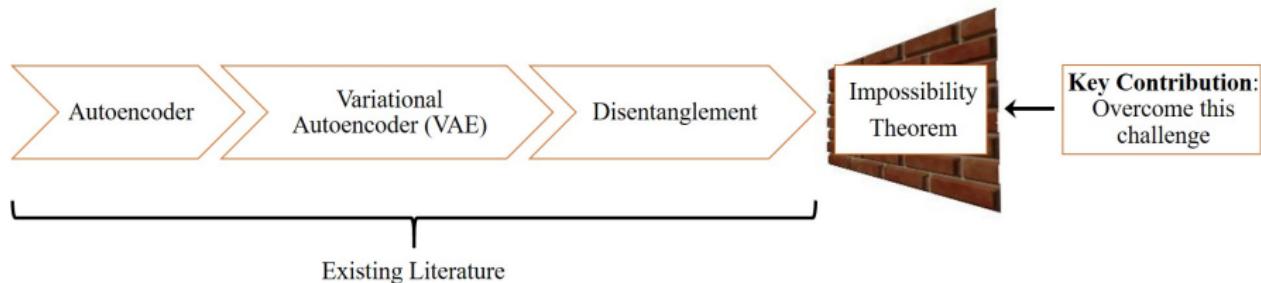
Goal: One to one mapping between  $z \Leftrightarrow c$

# 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.

# Disentangled and Entangled Representations

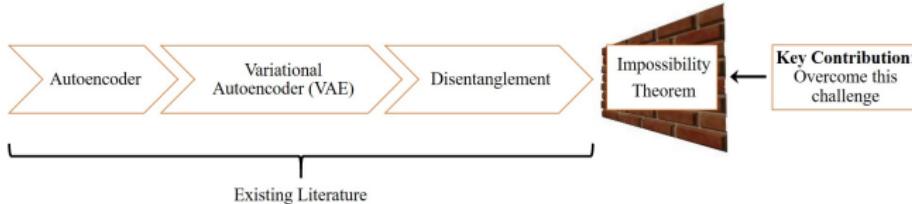
Example of *Entangled* Visual Characteristics



Example of *Disentangled* Visual Characteristics



# Impossibility Theorem



## Impossibility Theorem

Unsupervised (*i.e. only images*) learning of disentangled representations is *fundamentally impossible* except under certain restrictive conditions.<sup>a</sup>

<sup>a</sup>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
- 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
- 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
- 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
- Assume that ground truth is known by researchers
  - How?
- Human labeling
- Can we use this approach to discover visual characteristics?

# Impossibility Theorem – Implications



$z_1$
$z_2$
.
.
.
$z_j$

*predicts*

A horizontal arrow pointing from the learned characteristics list to the ground truth list.

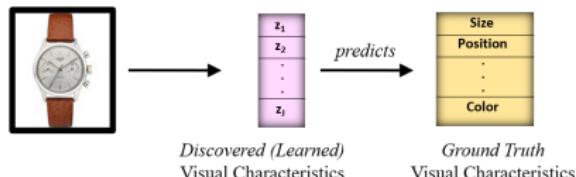
Size
Position
.
.
.
Color

*Discovered (Learned)*  
Visual Characteristics

*Ground Truth*  
Visual Characteristics

# Impossibility Theorem – Implications

Common approach to ground truth in ML is to get humans to label<sup>1</sup>



## What's the Problem?

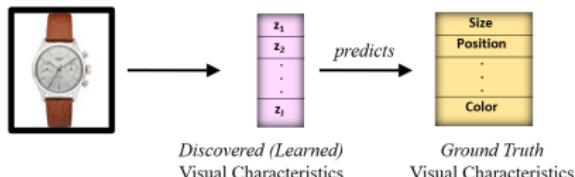
- Ground truth on visual characteristics is *unknown*. In fact, these are precisely what we want to find.

<sup>1</sup>

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<sup>1</sup>



## What's the Problem?

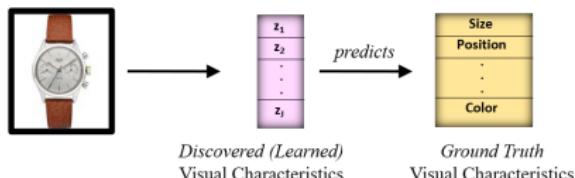
- Ground truth on visual characteristics is *unknown*. In fact, these are precisely what we want to find.
- Researcher needs to determine what are the *true characteristics* to focus on  $\Rightarrow$  not Automatic

<sup>1</sup>

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<sup>1</sup>

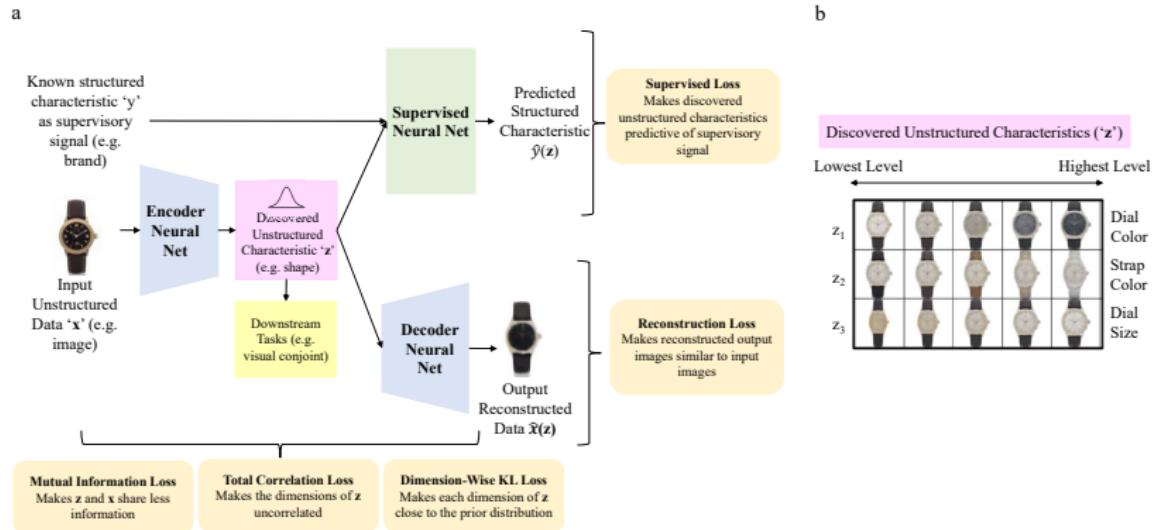


## What's the Problem?

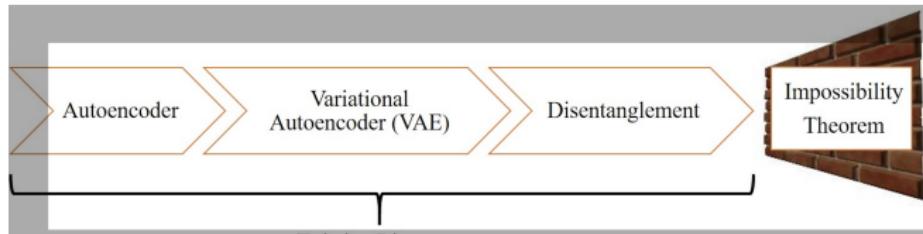
- Ground truth on visual characteristics is *unknown*. In fact, these are precisely what we want to find.
- Researcher needs to determine what are the *true characteristics* to focus on  $\implies$  not Automatic
- Need to ensure humans understand what these labels are and *how to quantify them* for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

<sup>1</sup>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
- $\theta$  and  $\phi$  are encoder and decoder parameters;  $\mathbf{x}$  are images

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

# Model

- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of integrated model
- $\theta$  and  $\phi$  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
<b>Total Correlation</b>	Promotes statistical independence between visual characteristics
Dimension-Wise KL	Penalizes deviations from a prior
<b>Supervised</b>	Provides a signal to address the impossibility theorem

# Model – Role of Supervised Loss

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

- 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

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

- 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.

## 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.<sup>a</sup>

---

<sup>a</sup> <https://www.prestigetime.com/blog/rolex-vs-cartier.html>

<sup>b</sup> <https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html>

<sup>c</sup> <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.<sup>a</sup>
- 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.<sup>b</sup>

---

<sup>a</sup><https://www.prestigetime.com/blog/rolex-vs-cartier.html>

<sup>b</sup><https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html>

<sup>c</sup><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.<sup>a</sup>
- 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.<sup>b</sup>
- Rolex is much more well-known for its highly-functional and iconically-designed sports and tool watches ...<sup>c</sup>

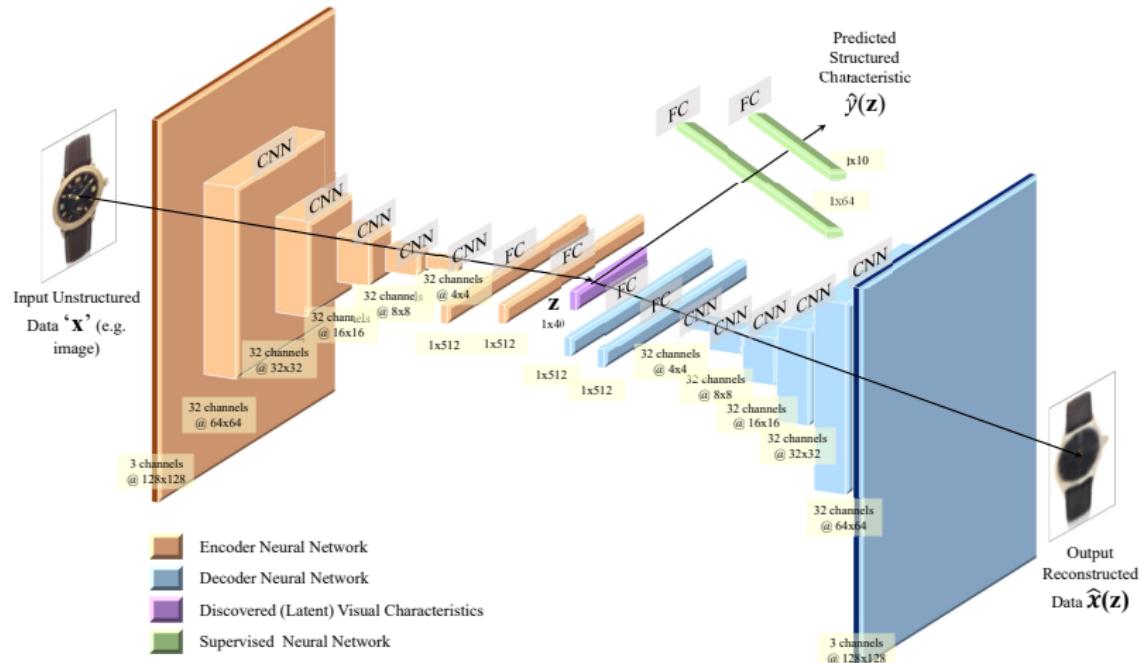
---

<sup>a</sup><https://www.prestigetime.com/blog/rolex-vs-cartier.html>

<sup>b</sup><https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html>

<sup>c</sup><https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html>

# Model Architecture



# Disentanglement Evaluation Metric

How to evaluate?

Unsupervised Disentanglement Ranking (UDR) measures disentanglement

- **Why UDR?:** “There are no labels available for many real-life applications and for some data, generative factors of interest are hard or impossible for humans to annotate.”<sup>2</sup>

<sup>2</sup> Estermann, B., Marks, M., & Yanik, M. F. (2020). Robust Disentanglement of a Few Factors at a Time using rPU-VAE. *Advances in Neural Information Processing Systems*, 33, 13387-13398.

# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- All good disentangled representations are similar to one another

# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- All good disentangled representations are similar to one another
  - Entangled representations are more likely to be different from one another. Why?

# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- All good disentangled representations are similar to one another
  - Entangled representations are more likely to be different from one another. Why?
  - There are many possible ways to get entangled representations.

# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- All good disentangled representations are similar to one another
  - Entangled representations are more likely to be different from one another. Why?
  - There are many possible ways to get entangled representations.
  - Disentangled representation is generated from ground truth factors

# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- All good disentangled representations are similar to one another
  - Entangled representations are more likely to be different from one another. Why?
  - There are many possible ways to get entangled representations.
  - Disentangled representation is generated from ground truth factors

# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- All good disentangled representations are similar to one another
  - Entangled representations are more likely to be different from one another. Why?
  - There are many possible ways to get entangled representations.
  - Disentangled representation is generated from ground truth factors

### Key Idea

"Models that disentangle well are more likely to be similar to each other than the ones that do not disentangle"<sup>a</sup>



# Disentanglement Evaluation Metric

## Unsupervised Disentanglement Ranking (UDR)

UDR measures disentanglement, ranges from 0 to 1  
(higher is better)

- Perturbed Models (different seed)  $i$  and  $j$
- Define UDR at the pairwise level as  $UDR_{ij}$
- Average across all possible combinations of  $i$  and  $j$

$$UDR_{ij} = \frac{1}{d_a + d_b} \left[ \sum_b \frac{r_a^2}{\sum_a R(a, b)} I_{KL}(b) + \sum_a \frac{r_b^2}{\sum_b R(a, b)} I_{KL}(a) \right]$$

where  $a$  is a visual characteristic in model  $i$  and  $b$  in model  $j$

# Disentanglement Evaluation Metric

UDR

$$UDR_{ij} = \frac{1}{d_a + d_b} \left[ \sum_b \frac{r_a^2}{\sum_a R(a, b)} I_{KL}(b) + \sum_a \frac{r_b^2}{\sum_b R(a, b)} I_{KL}(a) \right]$$

- ➊  $\frac{r_a^2}{\sum_a R(a, b)}$ : ratio of the (squared) correlation of the visual characteristic  $a$  in model  $i$  that is most similar to  $b$ , to the sum of the correlations across *all the visual characteristics* in model  $i$
- ➋ It will be higher if there is a one-to-one mapping between one visual characteristic in model  $i$  and another in model  $j$ .
- ➌ Add across all informative visual characteristics  $b$  of model  $j$ , which are represented by  $I_{KL}(b)$  using a threshold for KL divergence between the characteristic's posterior and the prior.
- ➍  $\sum_a \frac{r_b^2}{\sum_b R(a, b)} I_{KL}(a)$ : Counterpart to the first term

# 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

- |                                   |                                   |
|-----------------------------------|-----------------------------------|
| <input type="radio"/> Bezel       | <input type="radio"/> Hands       |
| <input type="radio"/> Crown       | <input type="radio"/> Hour Marker |
| <input type="radio"/> Date Window | <input type="radio"/> Lug         |
| <input type="radio"/> Dial        | <input type="radio"/> 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?

Which pair of watches in your judgment are more similar in terms of dial color than the other pair? (ignore all the other features of the watches)



Left



Right

# Visual Characteristics: Quantification?

## Interpretability and Quantification

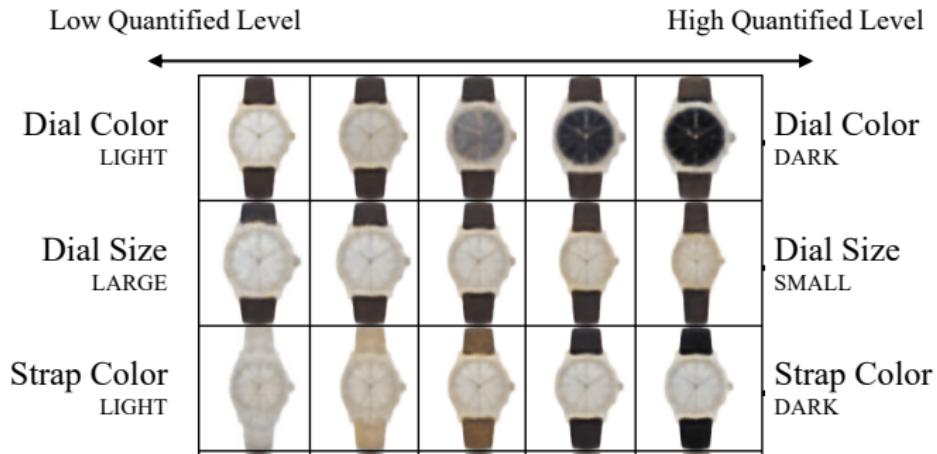
Visual characteristic	Interpretability Survey	Quantification Survey
Dial Size	76%	83%
Dial Color	80%	92%
Strap Color	88%	92%
Rim (Bezel) Color	79%	88%
Dial Shape	87%	68%
Knob (Crown) Size	70%	85%

# Discovered Visual characteristics

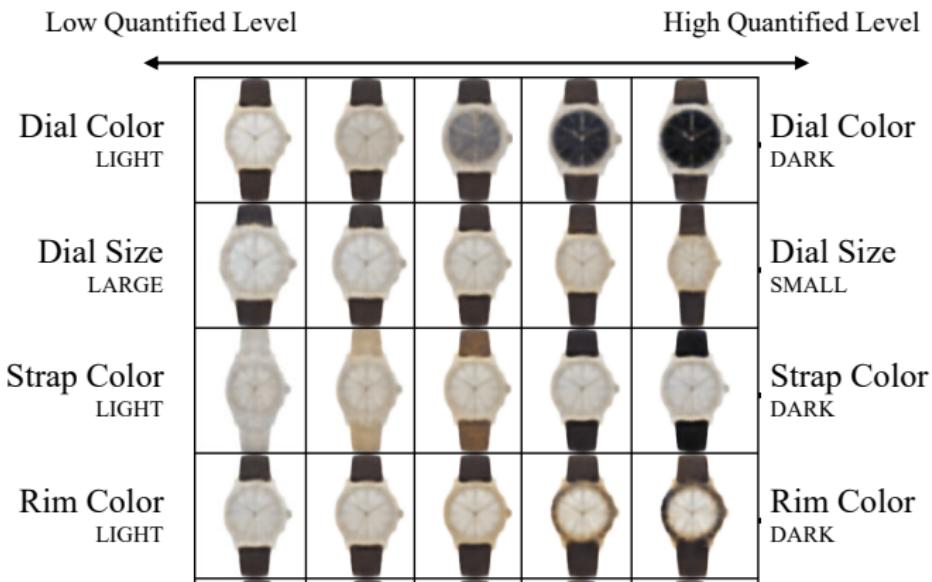
# Discovered Visual characteristics



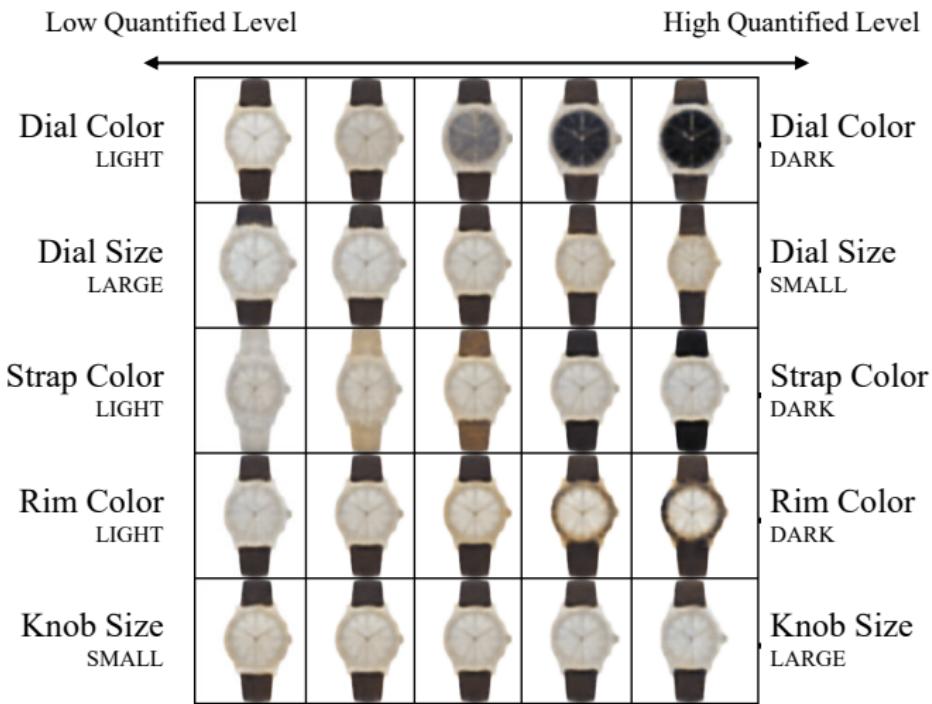
# Discovered Visual characteristics



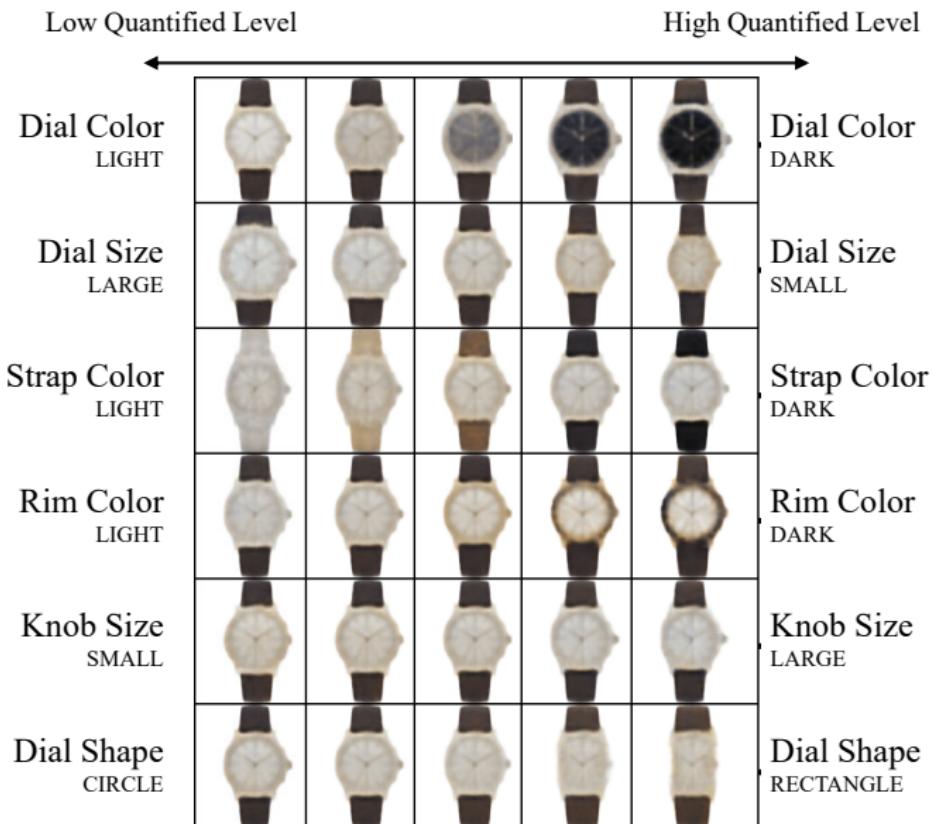
# Discovered Visual characteristics



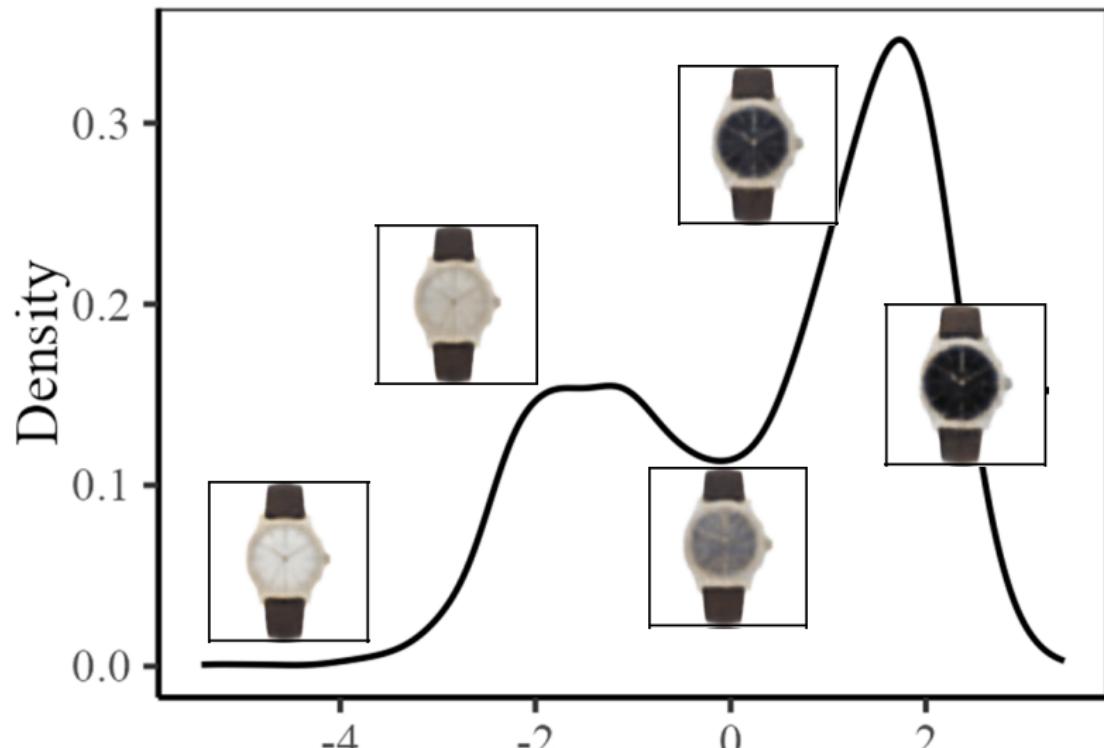
# Discovered Visual characteristics



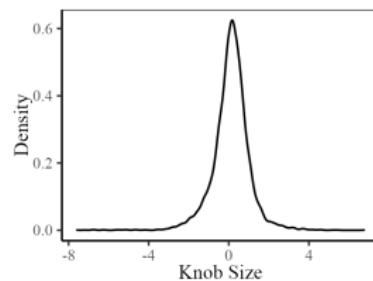
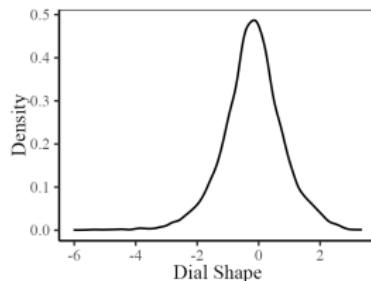
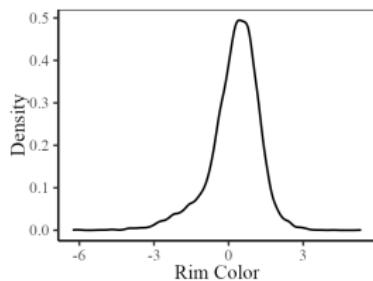
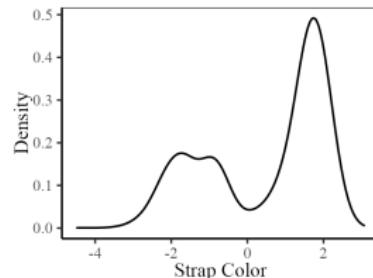
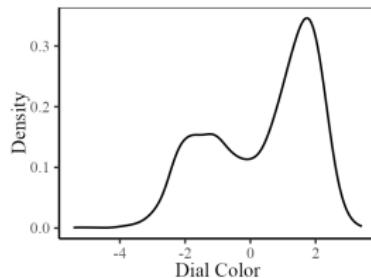
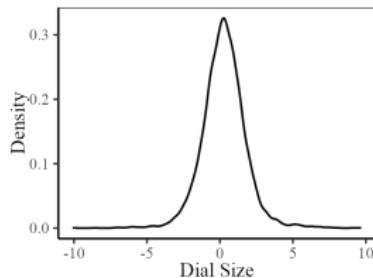
# Discovered Visual characteristics



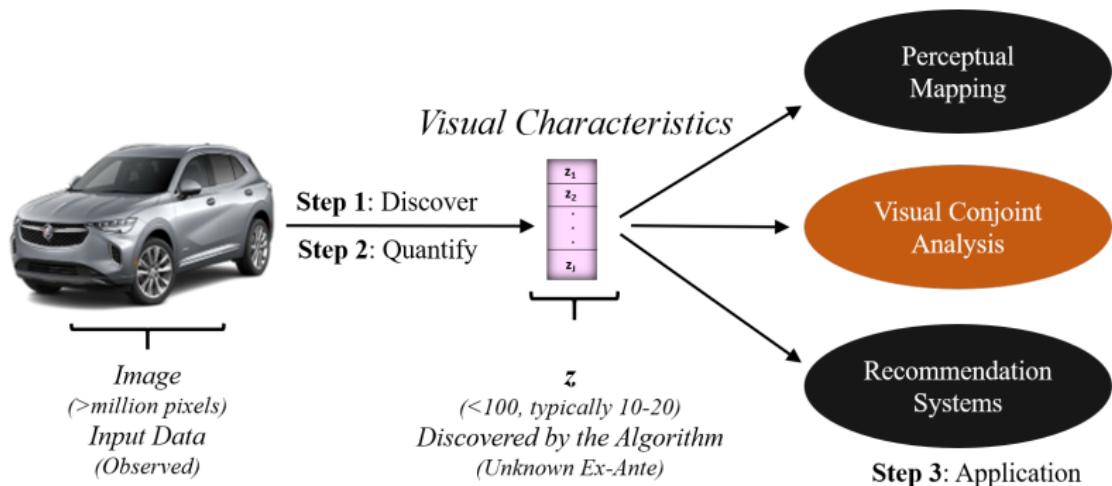
# Density of Discovered Visual characteristics (from 'Brand+Material' Signal)



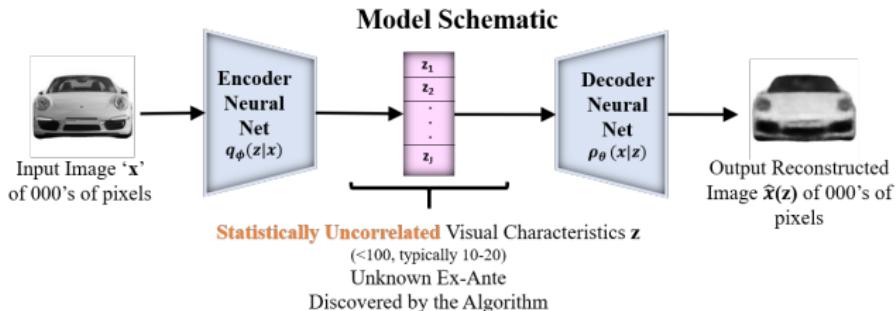
# Density of Discovered Visual characteristics (from 'Brand+Material' Signal)



# Research Goals

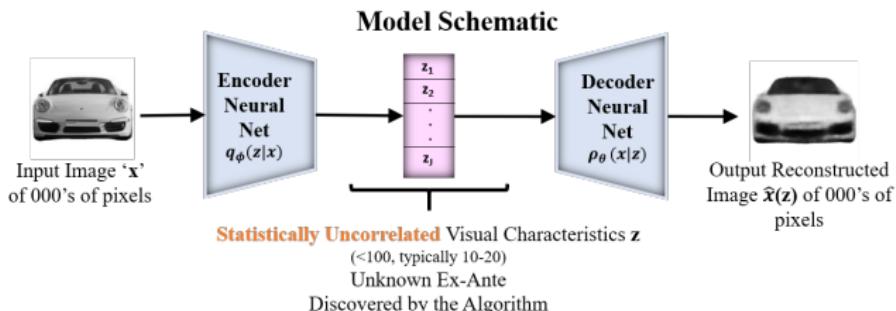


# Visual Conjoint Analysis: Background



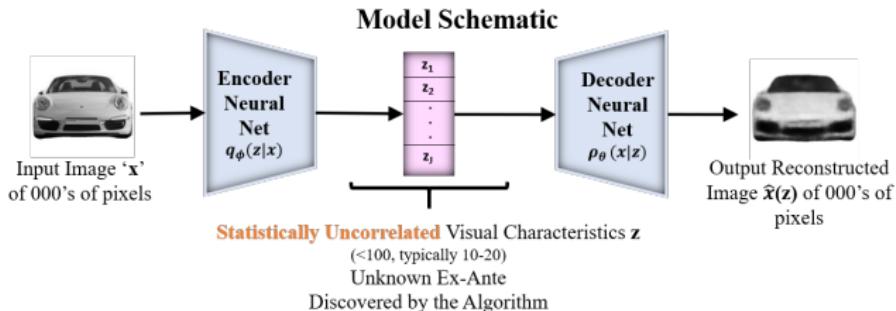
- Visual conjoint has been challenging to do because elements of visual space are correlated

# Visual Conjoint Analysis: Background



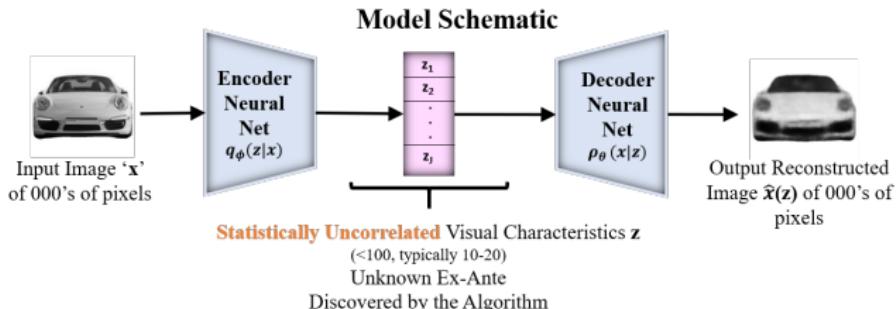
- Visual conjoint has been challenging to do because elements of visual space are correlated
- Designs have always been manually generated by product designers (prototypes)

# 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

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

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

where  $\text{LKJ}(\eta)$  is a Cholesky factorization of the correlation matrix  $\Omega_\beta$  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  $\epsilon_{ij}$  is a Gumbel random variable. The Bernoulli probability parameter  $\omega_i(j, j')$  is specified by the logit function, and  $\{j, j'\}$  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  $\sigma_\Theta^2, \Lambda_\Theta, \eta$  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%)
<b>Disentangled Embedding + HB Model (O + U)</b>	<b>71.61% (1.87%)</b>
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

---

<b>Segment 1</b>	Young moderately-affluent females
<b>Segment 2</b>	Older males

---

# 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

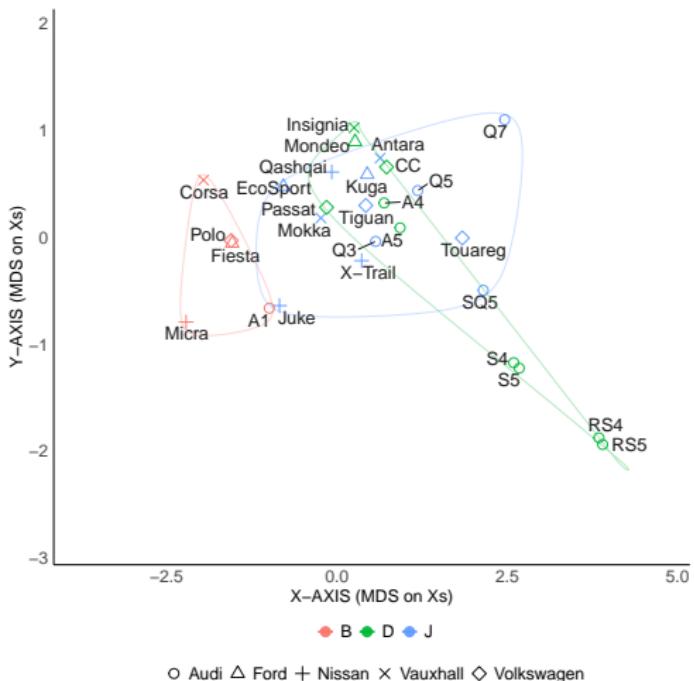
# Taking Disentanglement to the Automobile Market

Our interest is in figuring out whether competitive structure in **visual space** is different from competition in **structured space**.

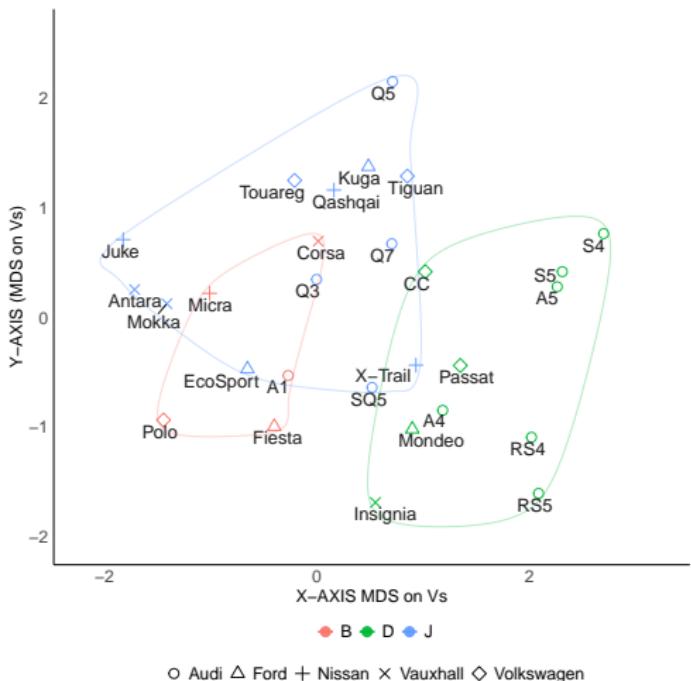
- We have typically focused on structured product characteristics: HP, mpg, range, trunk capacity etc.
- Market structure maps have been created to display structure of competition
  - Which car models compete most with my focal model/
  - Which brands are similar?
  - How does the competitive structure vary over time?

**Methodology:** We can get a market signal from market outcomes (which cars have a larger unobserved term?).

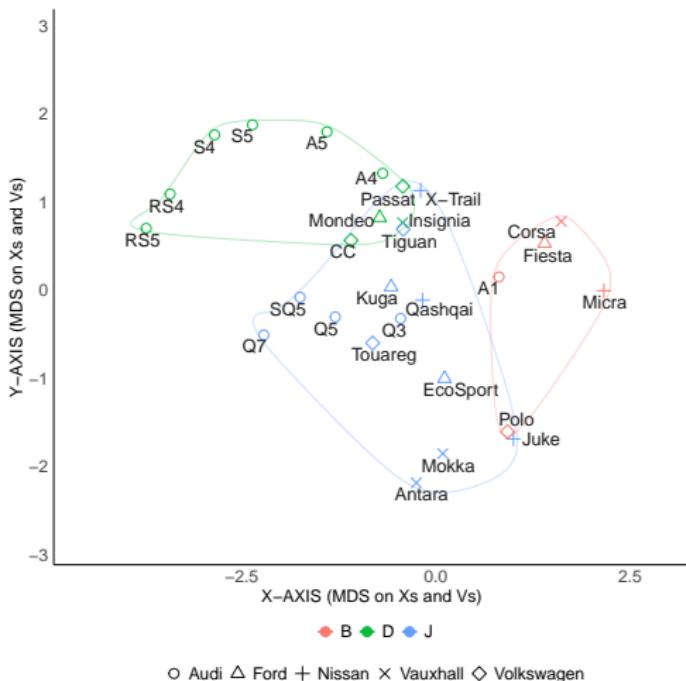
# Market Structure Map (Structured)



# Market Structure Map (Visual)



# Market Structure Map (Both)



# Big Picture

- Potential for domain knowledge (what we as researchers know) to help in machine learning

# Big Picture

- Potential for domain knowledge (what we as researchers know) to help in machine learning
- ML has typically been quite atheoretical (more general?)

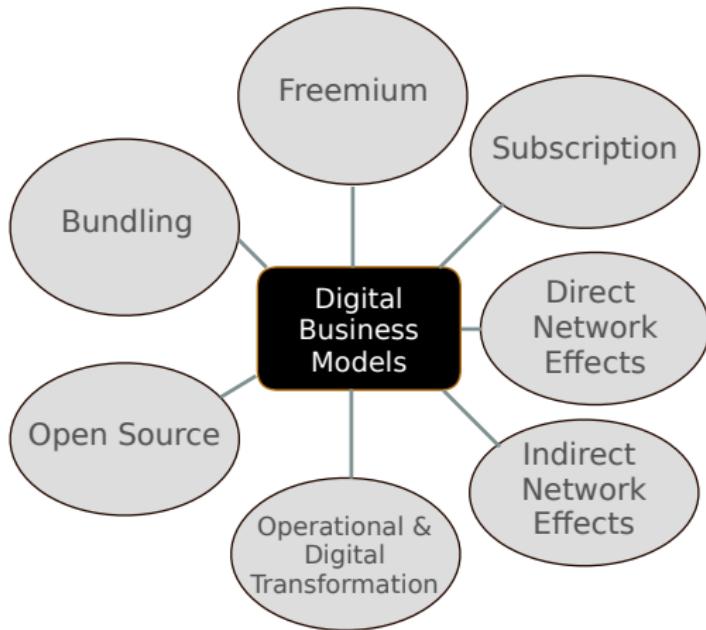
# Big Picture

- Potential for domain knowledge (what we as researchers know) to help in machine learning
- ML has typically been quite atheoretical (more general?)
- Worth asking *where* and *how* domain knowledge can be incorporated in ML models

# Big Picture

- Potential for domain knowledge (what we as researchers know) to help in machine learning
- ML has typically been quite atheoretical (more general?)
- Worth asking *where* and *how* domain knowledge can be incorporated in ML models
- Could be lots of complementarities

# Research Overview



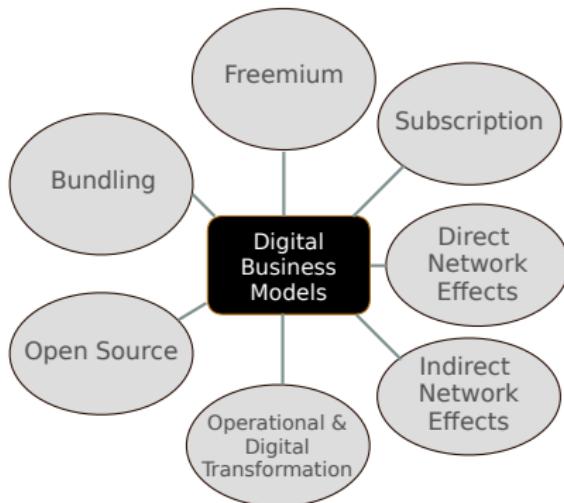
# Research Overview

The Dynamic Effects of  
Bundling as a Product  
Strategy, with T. Derdenger

**Marketing Science, 32(6)**

Competitive Product  
Strategy for Open Source  
Software, with B. Gordon  
and K. Srinivasan

**Marketing Science, 30(6)**



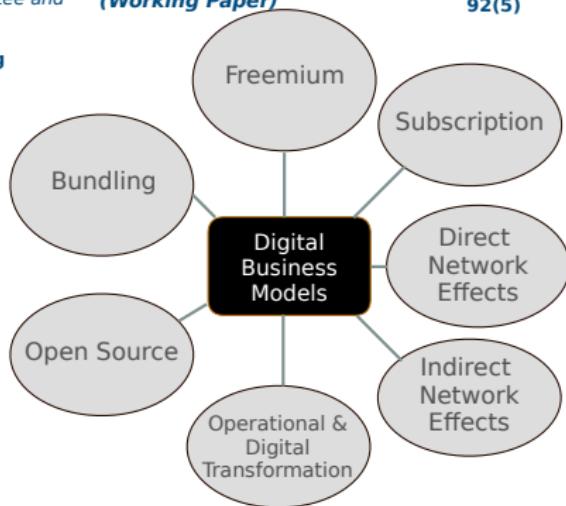
# Research Overview

Designing Freemium: Strategic Balancing of Growth and Monetization, with C. Lee and S. Gupta

Revision at **Marketing Science**

Designing Plans on Digital Platforms, with I. Weaver and S. Jonnalagedda (**Working Paper**)

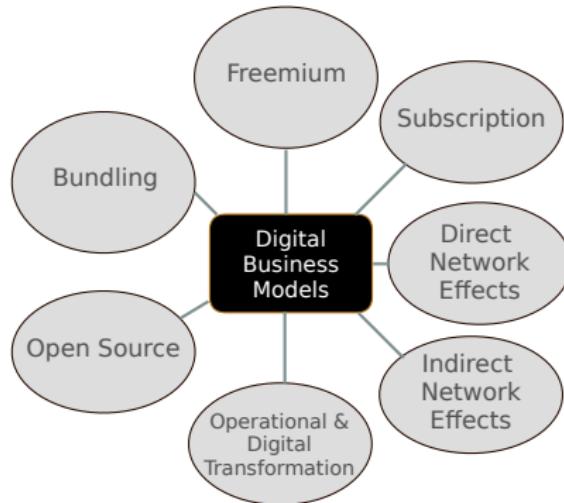
Making "Freemium" Work.  
**Harvard Business Review**, 92(5)



# Research Overview

The Dynamic Effects of Bundling as a Product Strategy, with T. Derdenger

Marketing Science, 32(6)



Demand Estimation for Subscriptions: Identification of WTP, with C. Chou

Marketing Science (Forthcoming)

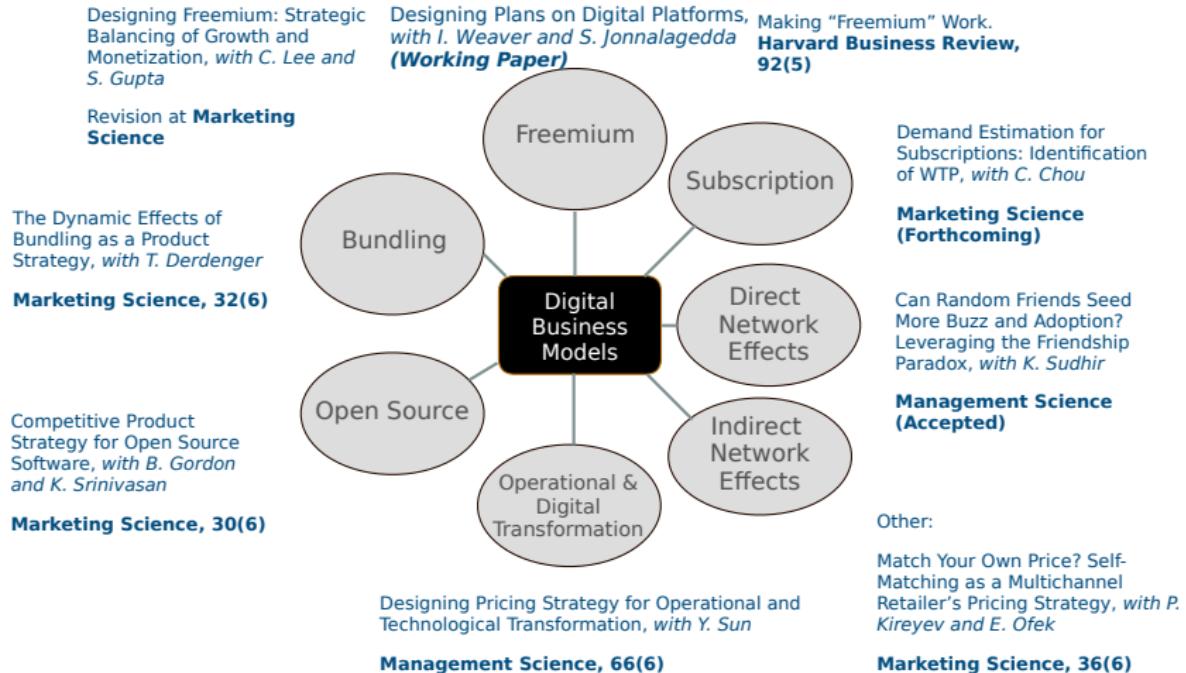
Can Random Friends Seed More Buzz and Adoption? Leveraging the Friendship Paradox, with K. Sudhir

Management Science (Accepted)

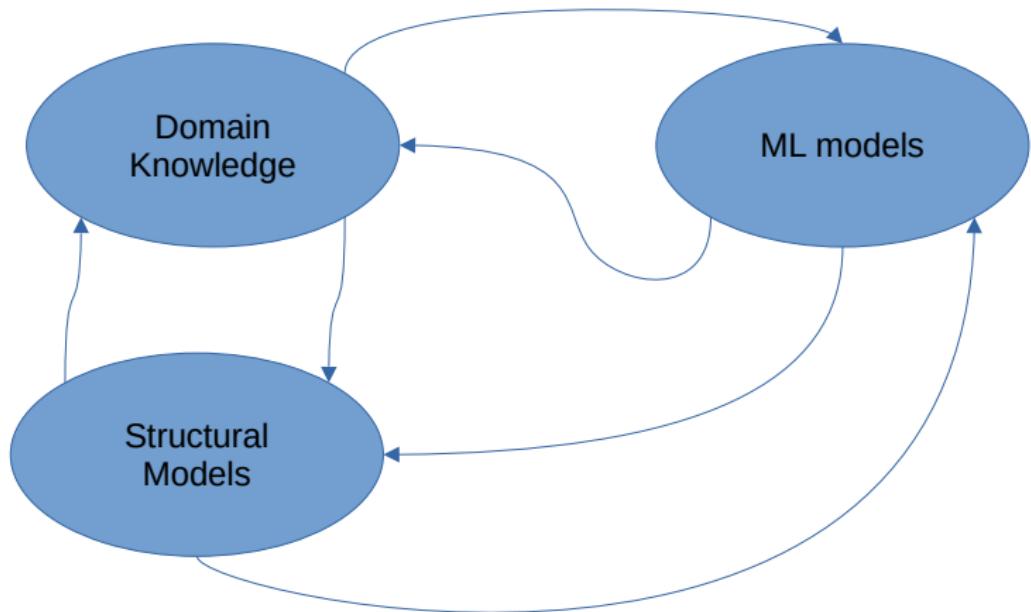
Designing Pricing Strategy for Operational and Technological Transformation, with Y. Sun

Management Science, 66(6)

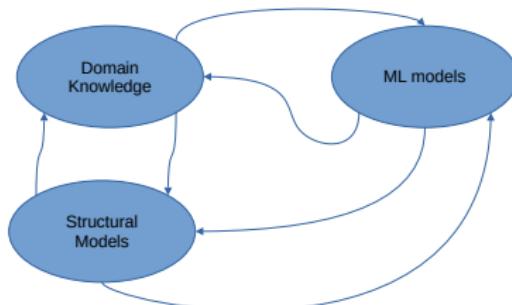
# Research Overview



# Research Overview



# Research Overview



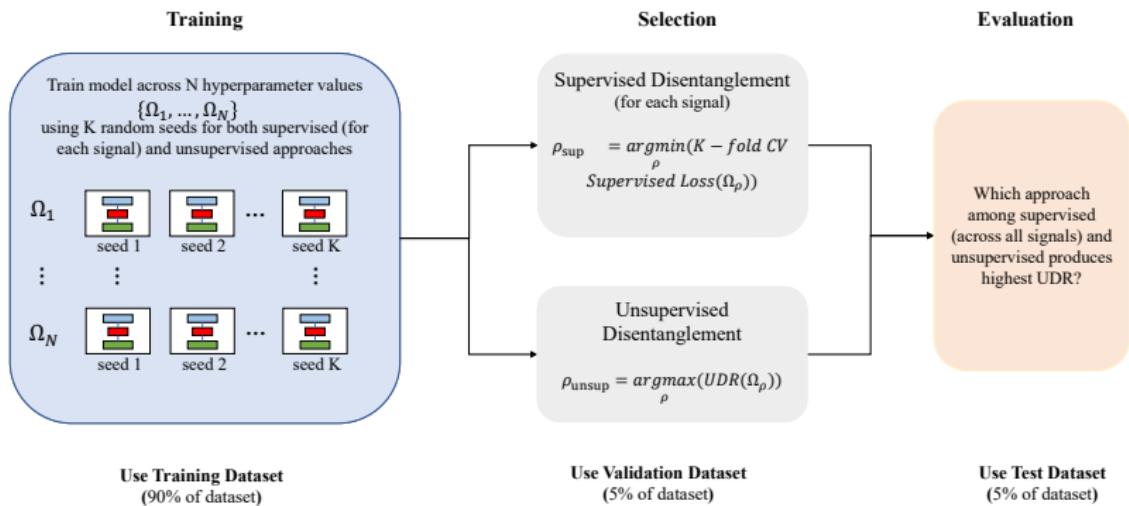
- *Market Structure Mapping with Visual Characteristics*, with Ankit Sisodia and Alex Burnap
- *A Theory-Based Interpretable Deep Learning Architecture for Music Emotion*, with Hortense Fong and K. Sudhir
- *Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve*, with Ian Weaver

# The End

vineet.kumar@yale.edu

# Additional Slides

# Model Training, Selection, & Evaluation



# Disentanglement Evaluation Metric

Esterman details the value of UDR, which we quote below:

*“There are no labels available for many real-life applications and for some data, generative factors of interest are hard or impossible for humans to annotate.*

# Table of Notation

Symbol	Category	Meaning
$\mathbf{x}$	Input Data	Product image
$\mathbf{y}$	Input Data	Supervisory signal(s)
$\hat{\mathbf{x}}$	Output Data	Reconstructed image
$\hat{\mathbf{y}}$	Output Data	Predicted Supervisory Signal(s)
$\mathbf{z}$	Latent Space	Visual characteristic vector
$\mathbf{z}_{\text{inf}}$	Subset of Latent Space	Informative visual characteristic
$p(\mathbf{z})$	Model	Prior distribution
$p_{\theta}(\mathbf{x} \mathbf{z})$	Decoder Neural Net	Conditional Probability of Generating Image Data given Latent Space
$q_{\phi}(\mathbf{z} \mathbf{x})$	Encoder Neural Net	Conditional Probability of Latent Space given Image Data
$p_w(\mathbf{y} \mathbf{z})$	Supervisory Neural Net	Conditional Probability of Supervisory Signal given Latent Space
$\theta$	Weights of Neural Net	Decoder's parameters
$\phi$	Weights of Neural Net	Encoder's parameters
$w$	Weights of Neural Net	Supervisory Net's parameters
$\mathbf{E}_{q_{\phi}(\mathbf{z} \mathbf{x})} [\log p_{\theta}(\mathbf{x} \mathbf{z})]$	Loss Function	Reconstruction Loss
$I_q(\mathbf{z}, \mathbf{x})$	Loss Function	Mutual Information Loss
$KL \left[ q(\mathbf{z})    \prod_{j=1}^J q(z_j) \right]$	Loss Function	Total Correlation Loss
$\sum_{j=1}^J KL [q(z_j)    p(z_j)]$	Loss Function	Dimension KL Divergence Loss
$P(\hat{y}(\mathbf{z}), y)$	Loss Function	Supervised Loss
$\mathcal{L}(\theta, \phi, \beta; \mathbf{x}, \mathbf{z})$	Loss Function	Total Loss
$J$	Hyperparameter	Dimensionality of latent space
$\alpha$	Hyperparameter	Weight on Mutual Information Loss
$\beta$	Hyperparameter	Weight on Total Correlation Loss
$\gamma$	Hyperparameter	Weight on Dimension KL Divergence Loss
$\delta$	Hyperparameter	Weight on Supervised Loss

# Why does brand aid the disentanglement model?



**Cartier**



**Patek Philippe**



**Rolex**

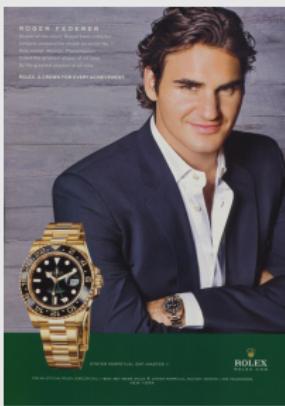
# Why does brand aid the disentanglement model?



**Cartier**



**Patek Philippe**



**Rolex**

# Why does 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.<sup>a</sup>

---

<sup>a</sup> <https://www.prestigetime.com/blog/rolex-vs-cartier.html>

<sup>b</sup> <https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html>

<sup>c</sup> <https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html>

# Why does 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.<sup>a</sup>
- 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.<sup>b</sup>

---

<sup>a</sup><https://www.prestigetime.com/blog/rolex-vs-cartier.html>

<sup>b</sup><https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html>

<sup>c</sup><https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html>

# Why does 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.<sup>a</sup>
- 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.<sup>b</sup>
- Rolex is much more well-known for its highly-functional and iconically-designed sports and tool watches ...<sup>c</sup>

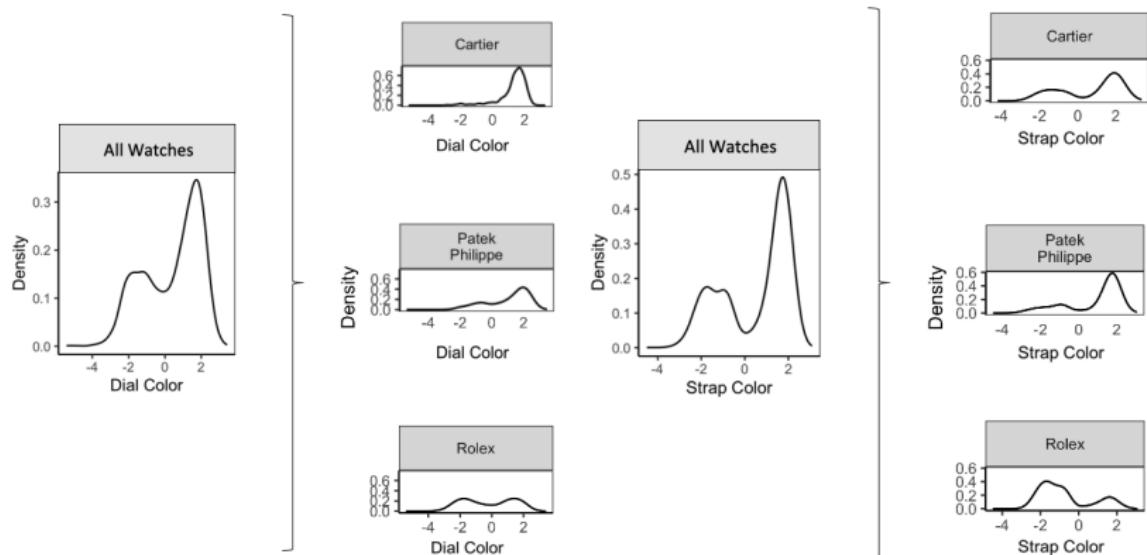
---

<sup>a</sup><https://www.prestigetime.com/blog/rolex-vs-cartier.html>

<sup>b</sup><https://www.prestigetime.com/blog/audemars-piguet-vs-patek-philippe.html>

<sup>c</sup><https://www.bobswatches.com/rolex-blog/watch-101/patek-philippe-better-rolex.html>

# Why does brand aid the disentanglement model?



Our method is able to quantify these differences across brands from the visual product images.

# Why does brand aid the disentanglement model?

Average of pairwise JS Divergence Of a visual characteristic across different brands is much higher than across different price points.

Visual Characteristic	Brand	2 Discrete Prices	3 Discrete Prices	5 Discrete Prices	10 Discrete Prices
Dial Color	6.28	1.62	1.77	1.91	1.91
Dial Size	9.71	2.13	3.24	4.13	4.86
Strap Color	6.16	0.56	1.76	1.85	2.08
Rim Color	9.07	3.61	6.56	6.57	8.34
Dial Shape	5.97	0.29	0.66	0.75	3.16
Knob Size	8.20	6.02	5.76	13.08	13.46
Across All 6 Char	7.56	2.37	3.29	4.71	5.63