

# Automatic Discovery and Generation of Visual Design Characteristics

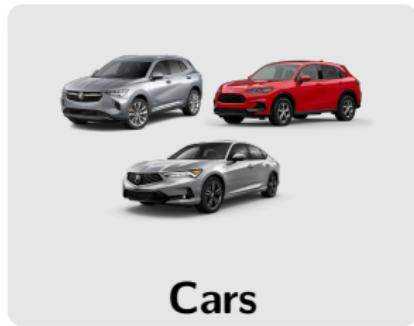
## Application to Visual Conjoint

Ankit Sisodia, Alex Burnap and Vineet Kumar

Yale School of Management

Presenting at: Indian School of Business  
August 2023

# 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 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 directly from unstructured product images

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

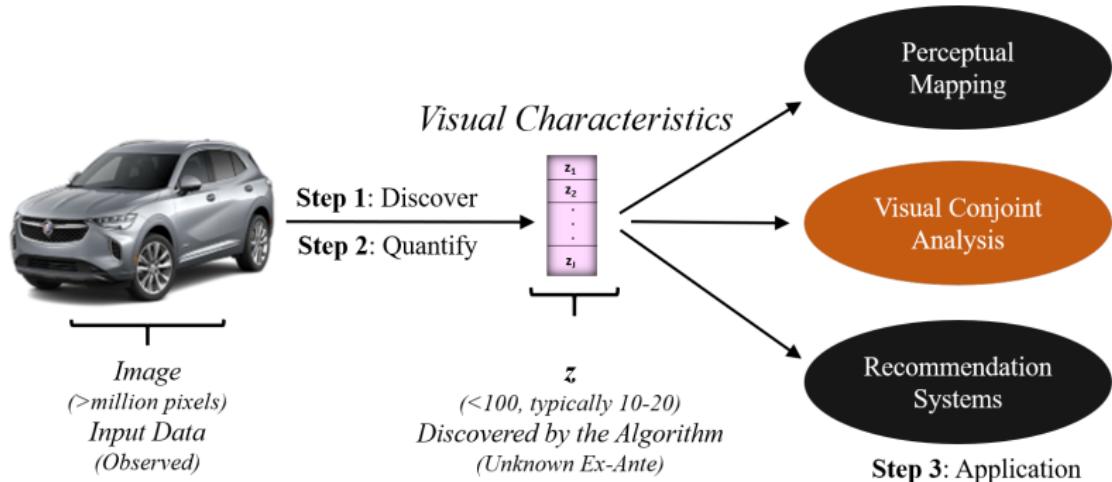
# What this paper seeks to do

## Research Goals

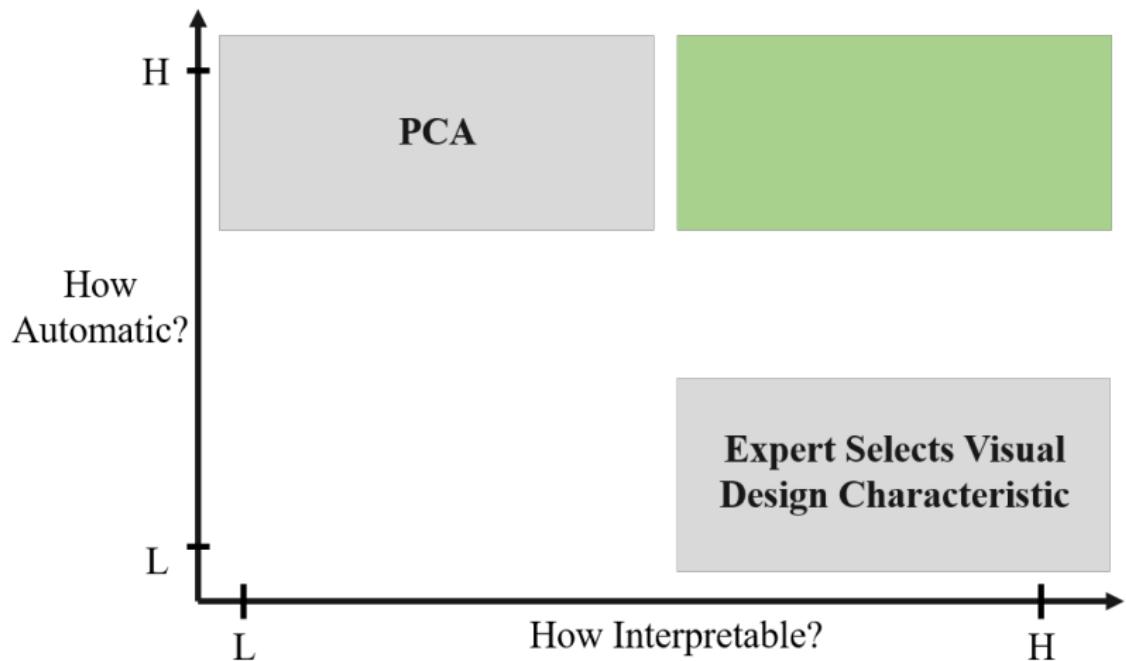
Our research aims to obtain **interpretable** visual characteristics directly from unstructured product images

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

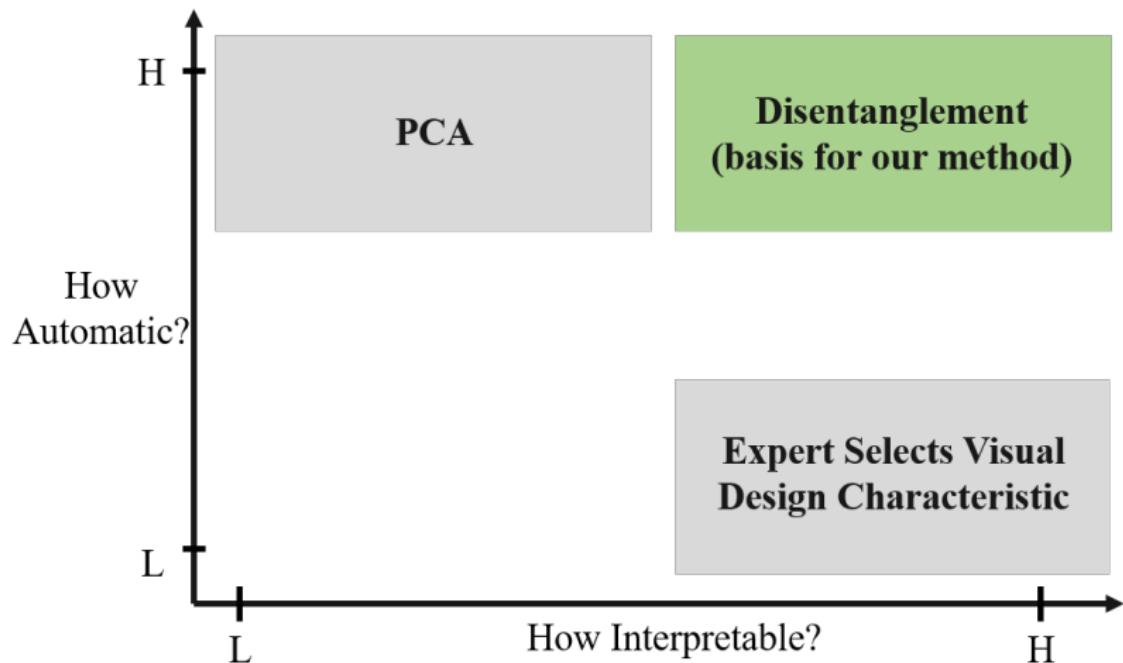
# Research Goals



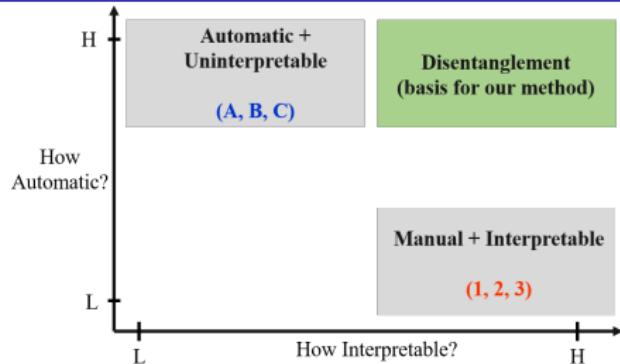
# Modeling Visual Characteristics: A comparison of methods



# Modeling Visual Characteristics: A comparison of methods



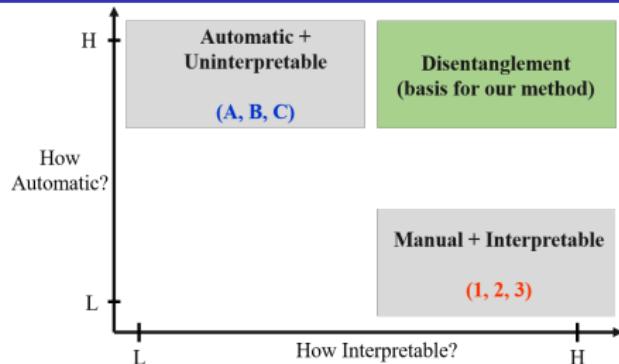
# Modeling Visual Characteristics: A comparison of methods



## 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

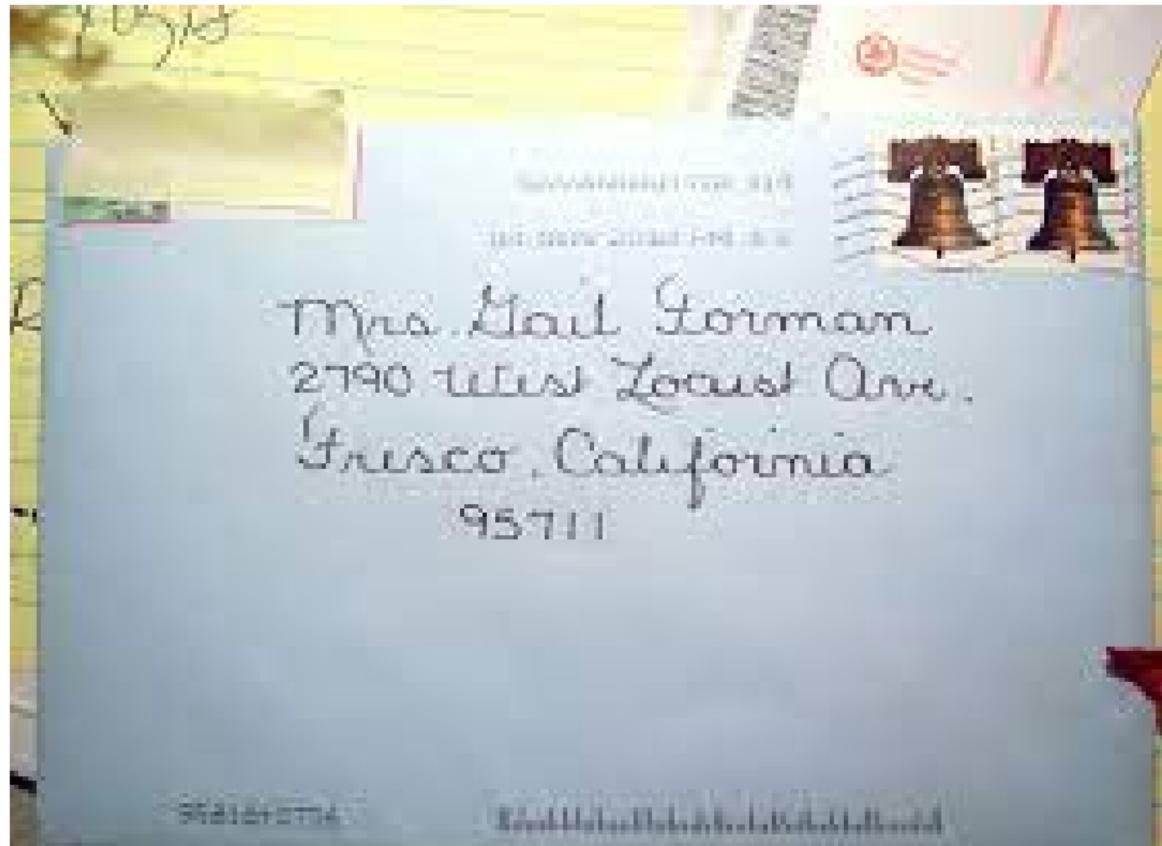
- A - Bajari, P. L. et al. (2021) : Hedonic prices and quality adjusted price indices powered by AI, *CENMAP working paper*
- B - Law, S., et al. (2019) : Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*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

## Manual + Interpretable

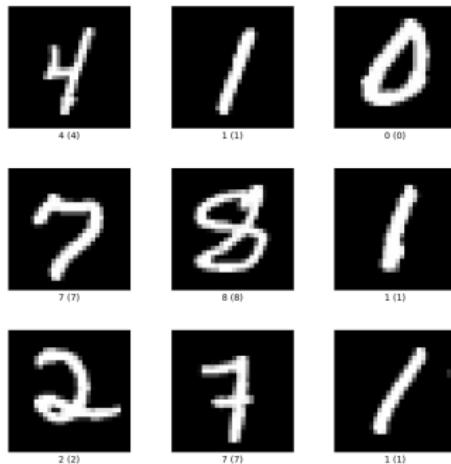
- 1 - Zhang, M. et al. (2022) : Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from yelp. *Management Science*
- 2 - Liu, Y., et al. (2017) : The effects of products' aesthetic design on demand and marketing-mix effectiveness: The role of segment prototypicality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

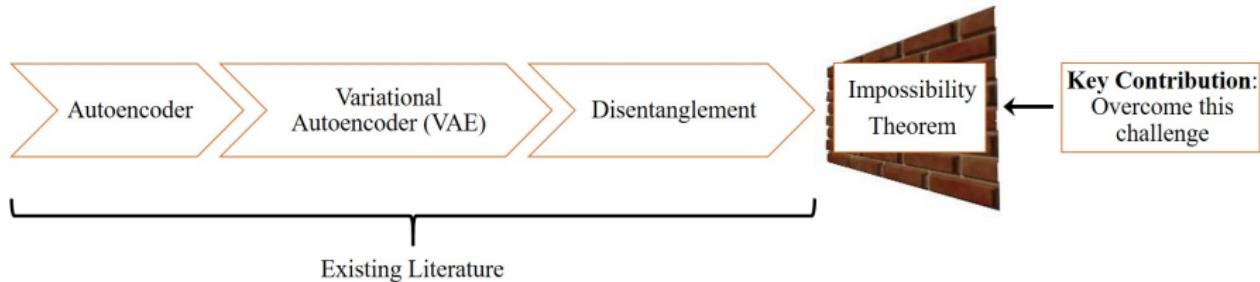
# What is disentanglement?

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 (**v**): Dimensions in the model's latent space
- Generative factors (**c**): Human-interpretable true characteristics

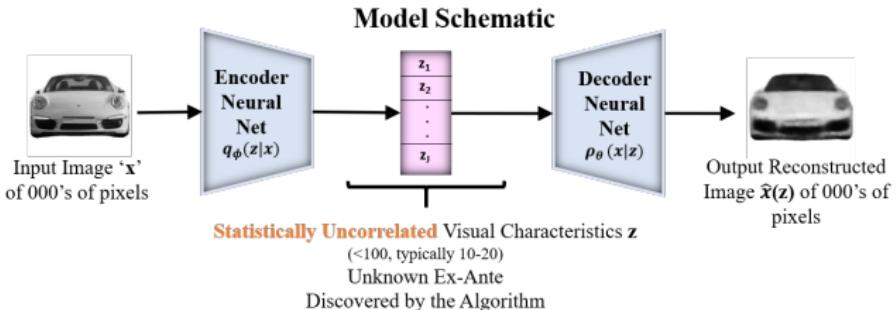
# Roadmap of Our Approach



## Contribution

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

# Models in Existing Literature



---

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

---

# 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"*

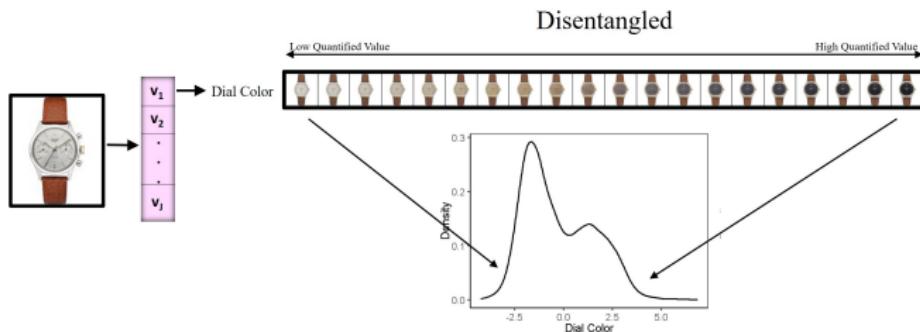
# What is disentanglement?

Bengio et al (2013)

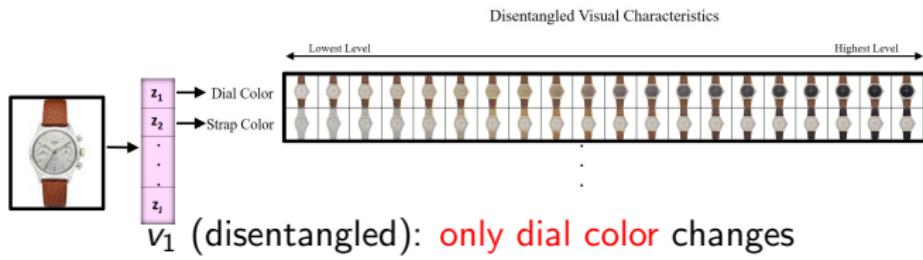
*"A disentangled representation can be defined as one where **single latent units** are sensitive to changes in **single generative factors**, while being relatively invariant to changes in other factors"*

- Latent Units (**v**): Dimensions in the model's latent space
- Generative factors (**c**): Human-interpretable true characteristics

# Disentangled v Entangled Representation

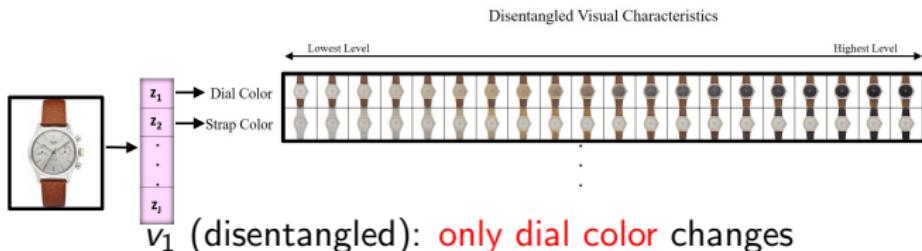


# Disentangled v Entangled Representation

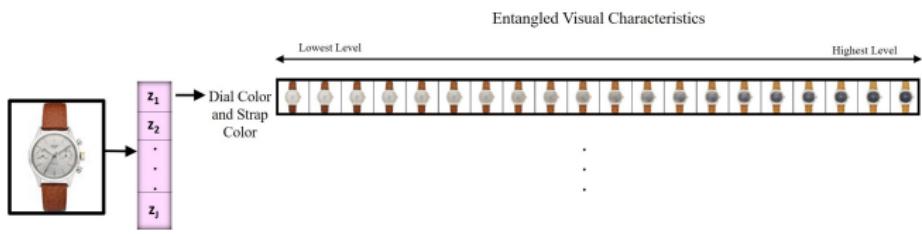


$v_2$  (disentangled): only strap color changes

# Disentangled v Entangled Representation



$v_2$  (disentangled): only strap color changes



$v_1$  (entangled): both dial color and strap color changes

# 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"*

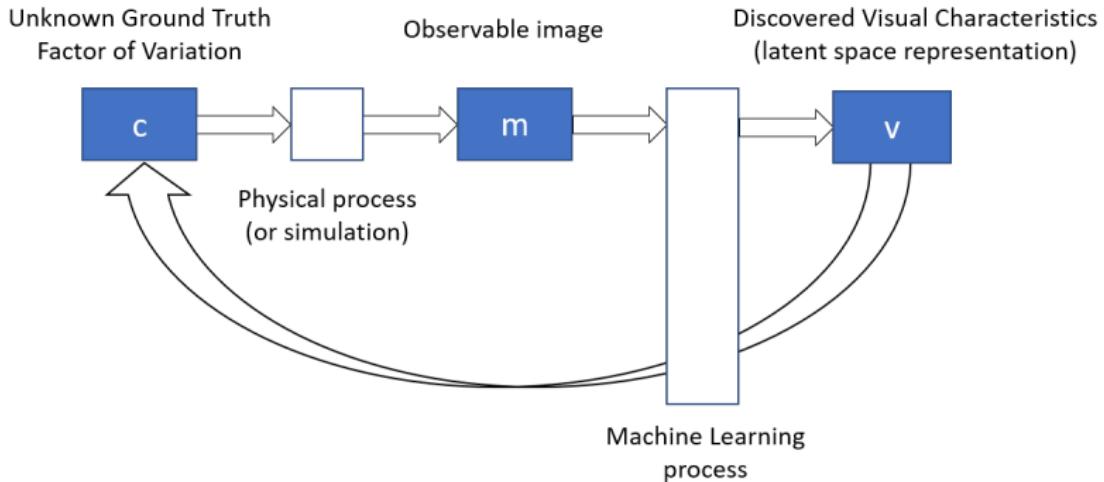
# What is disentanglement?

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 (**v**): Dimensions in the model's latent space
- Generative factors (**c**): Human-interpretable true characteristics

# What is disentanglement?

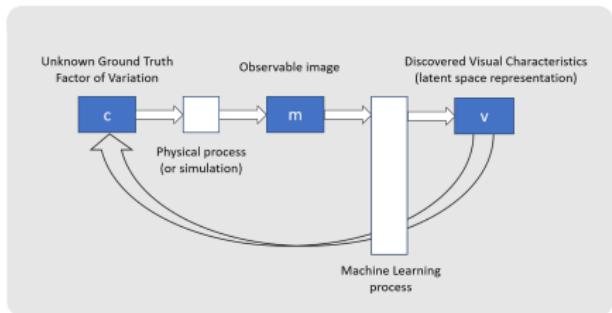


Goal of machine learning process: Recover latent space  $v(m(c))$  and make correspondence  $c \longleftrightarrow v$

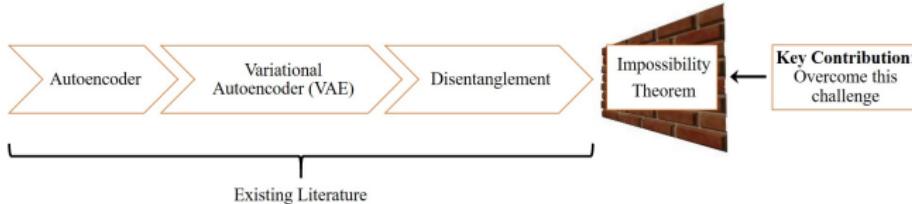
# What is disentanglement?

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"*



# 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.

# Overcoming Impossibility Theorem



$z_1$
$z_2$
.
.
$z_j$

*predicts*

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

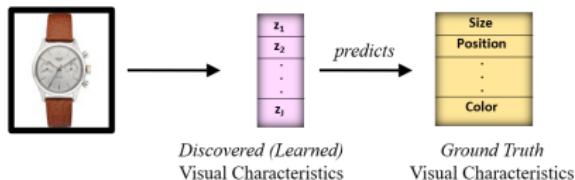
Size
Position
.
.
Color

*Discovered (Learned)*  
Visual Characteristics

*Ground Truth*  
Visual Characteristics

# Overcoming Impossibility Theorem

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



## What's the Problem?

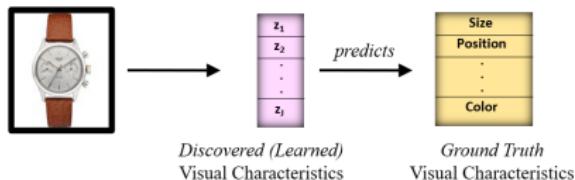
- Ground truth on visual characteristics is unknown. In fact, these are precisely what we want to find.

<sup>1</sup>

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

# Overcoming Impossibility Theorem

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



## What's the Problem?

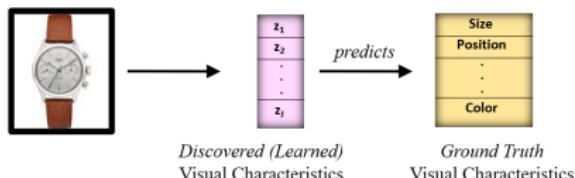
- Ground truth on visual characteristics is **unknown**. In fact, these are precisely what we want to find.
- Researcher needs to determine what are the true characteristics to focus on

<sup>1</sup>

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

# Overcoming Impossibility Theorem

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

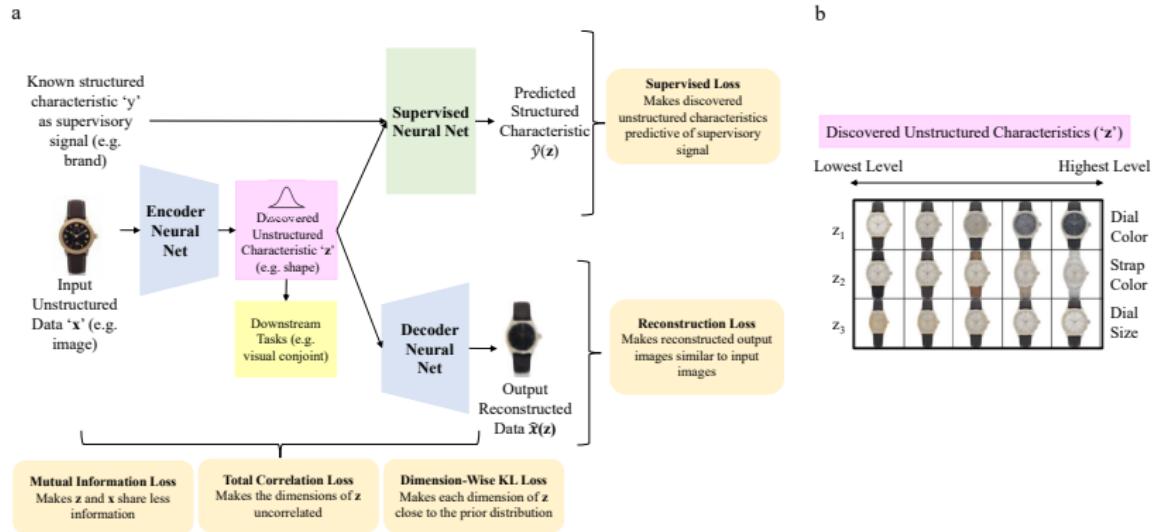


## What's the Problem?

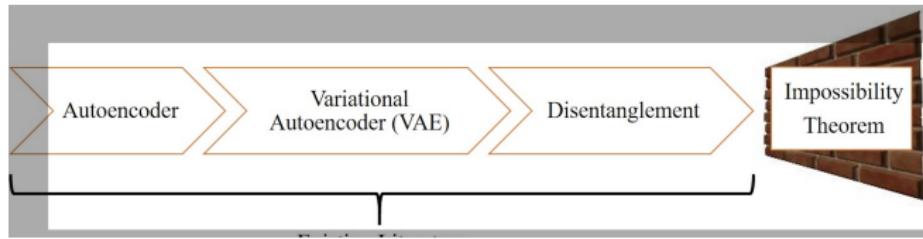
- Ground truth on visual characteristics is **unknown**. In fact, these are precisely what we want to find.
- Researcher needs to determine what are the **true characteristics** to focus on
- Need to ensure humans understand what these labels are and how to quantify them for each image

<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

# 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

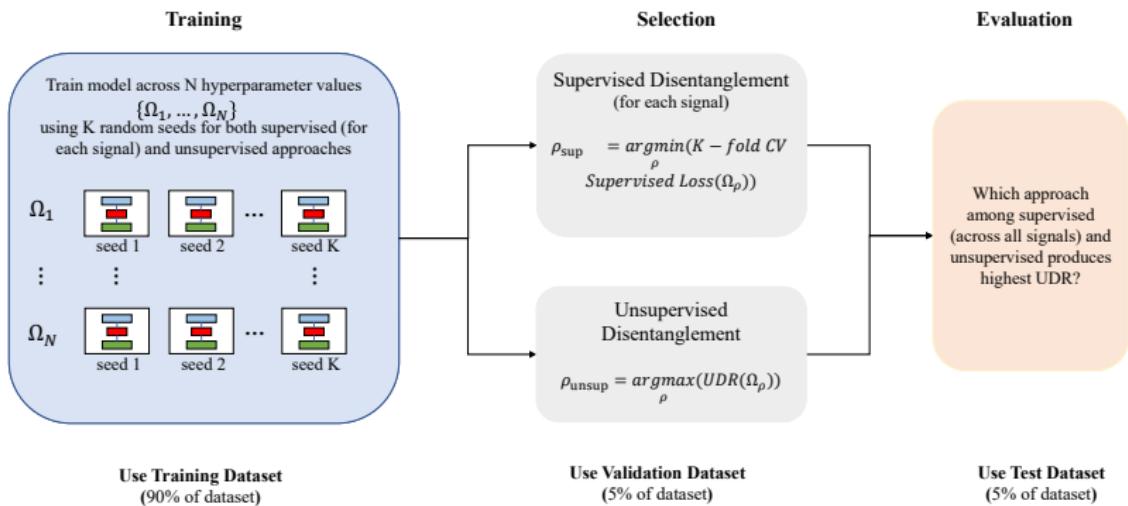
# Model Estimation

- 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

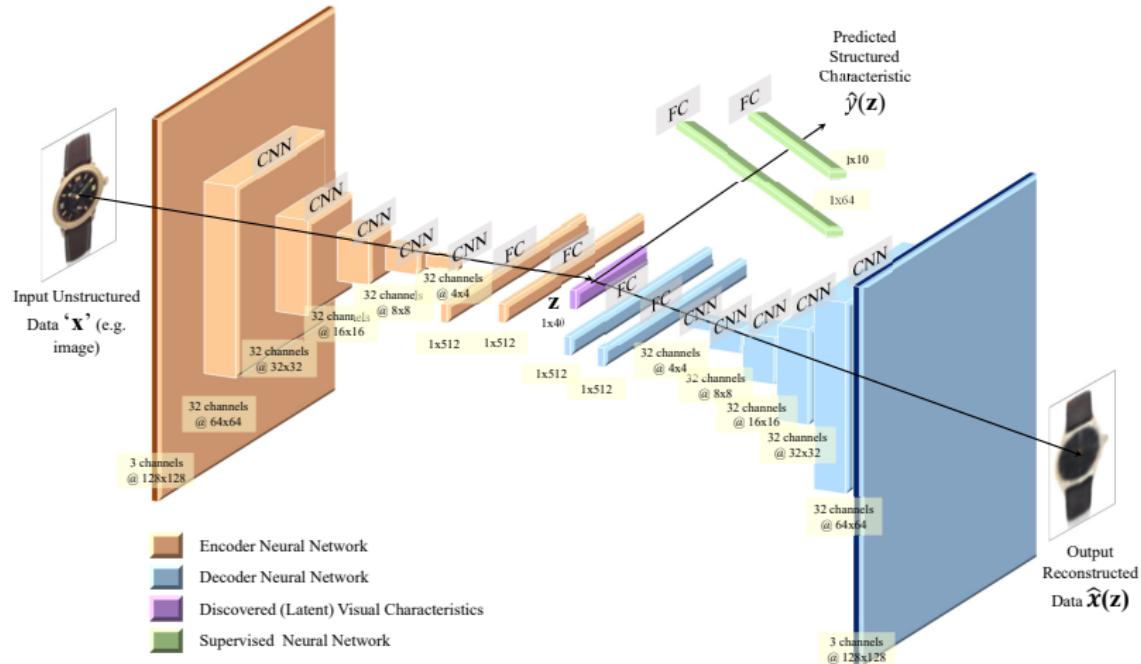
$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta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

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 Training, Selection, & Evaluation



# Model Architecture



# 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.*

# Disentanglement Evaluation Metric

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>
- **Key Idea**: “Models that disentangle well are more likely to be similar to each other than the ones that do not disentangle”<sup>3</sup>

---

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

<sup>3</sup> Estermann, B., Marks, M., & Yanik, M. F. (2020). Robust Disentanglement of a Few Factors at a Time using rPVAE. Advances in Neural Information Processing Systems, 33, 13387-13398.

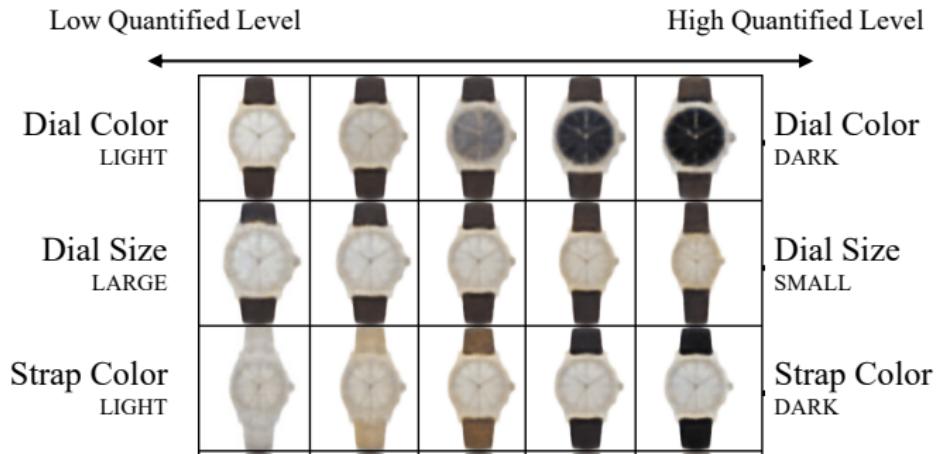
# Discovered Visual characteristics



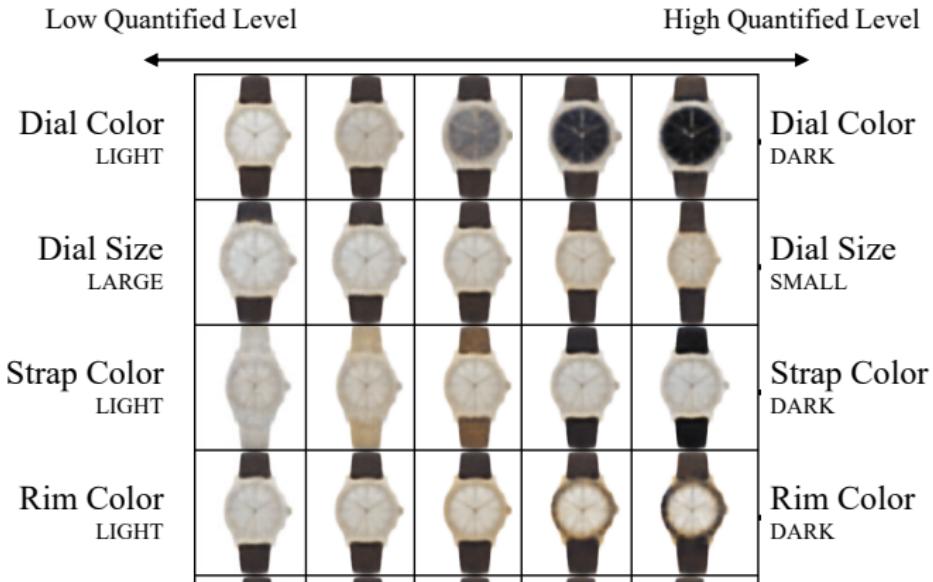
# Discovered Visual characteristics



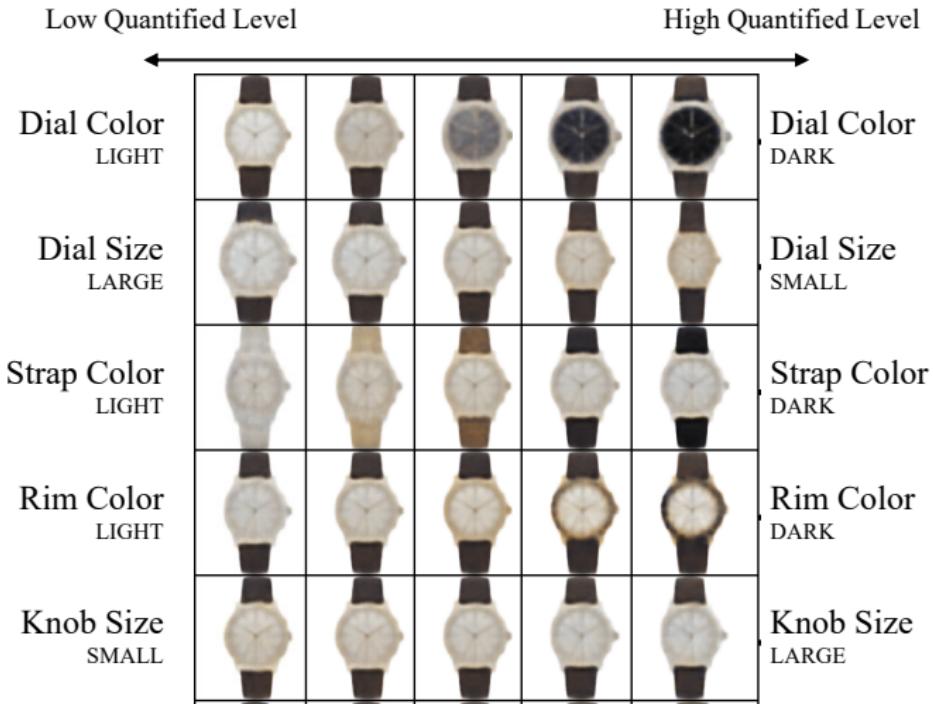
# Discovered Visual characteristics



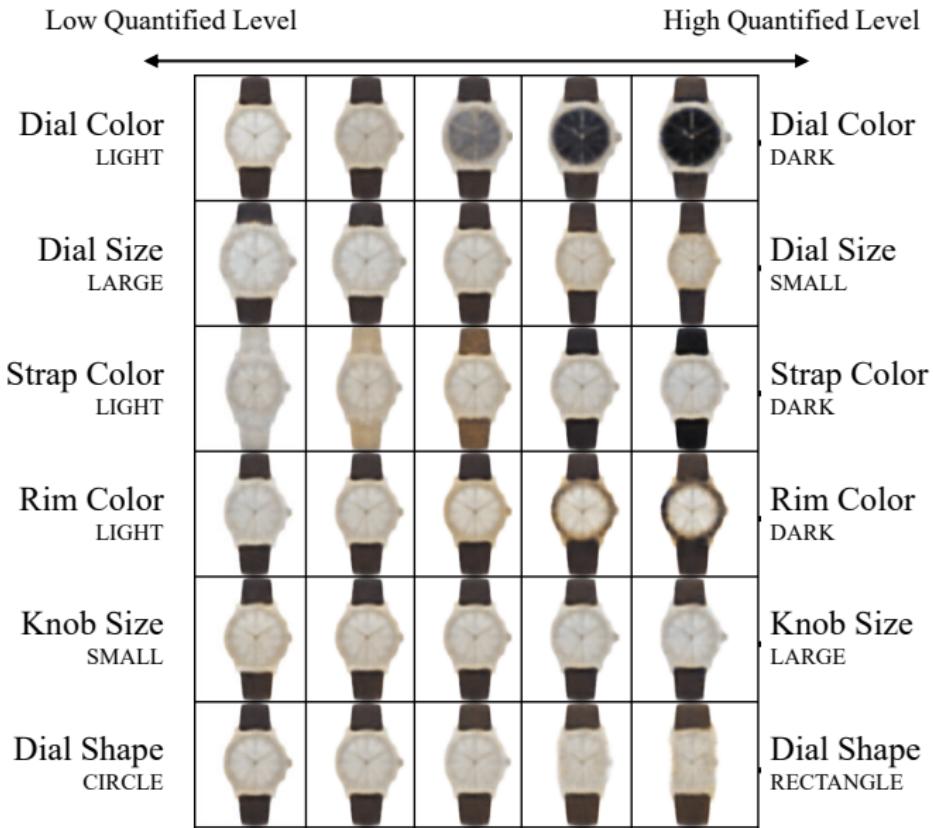
# Discovered Visual characteristics



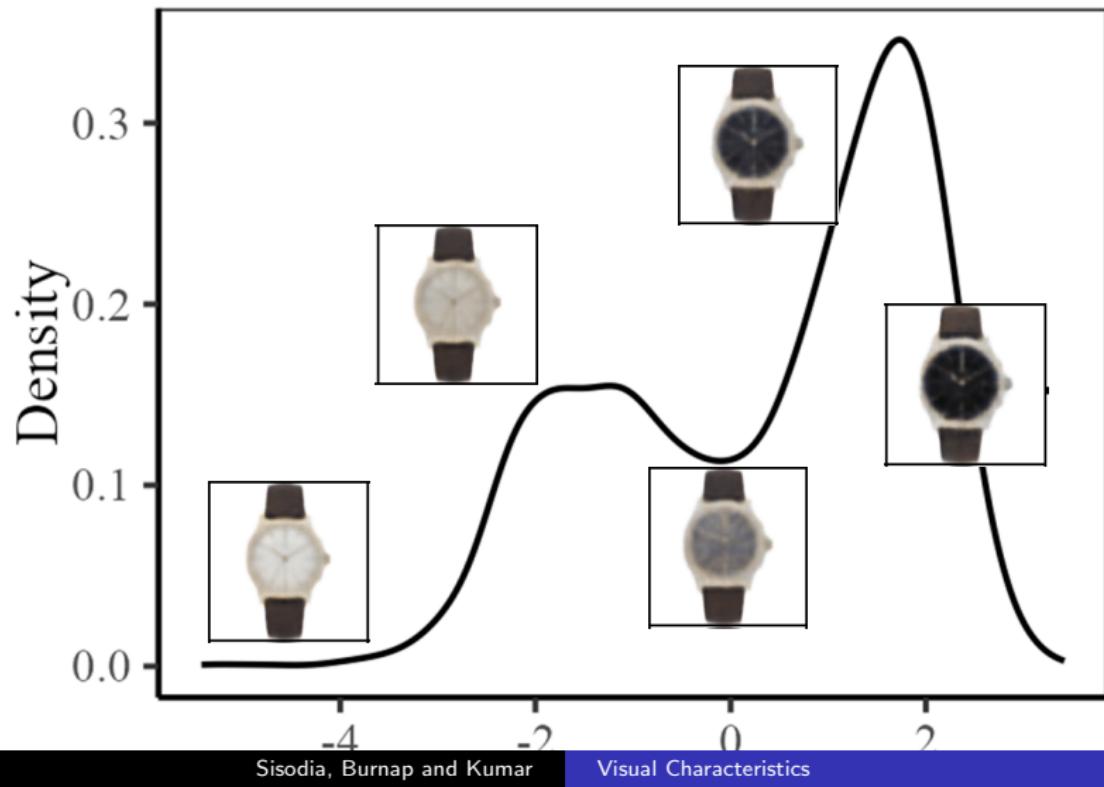
## Discovered Visual characteristics



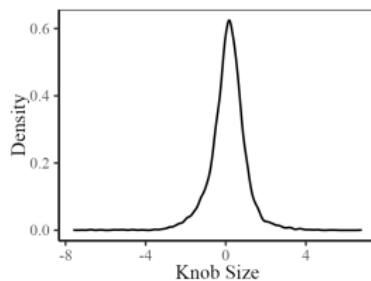
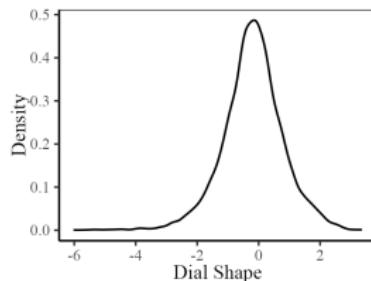
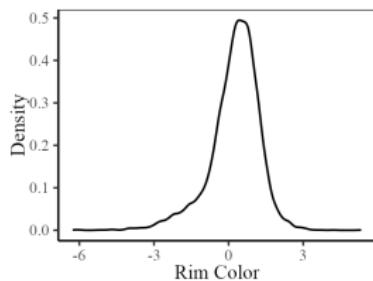
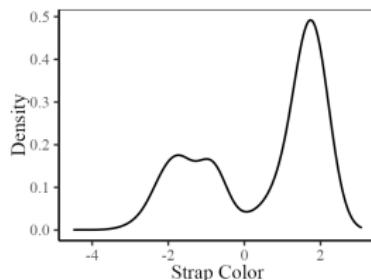
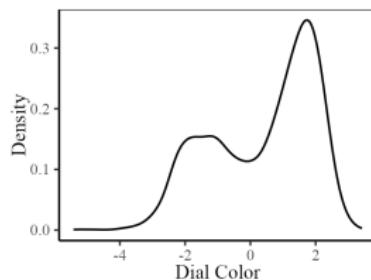
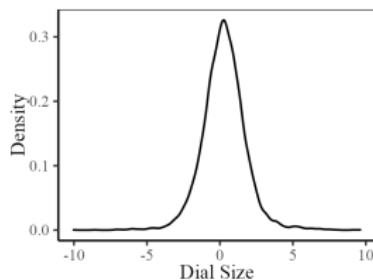
# Discovered Visual characteristics



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



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



# 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

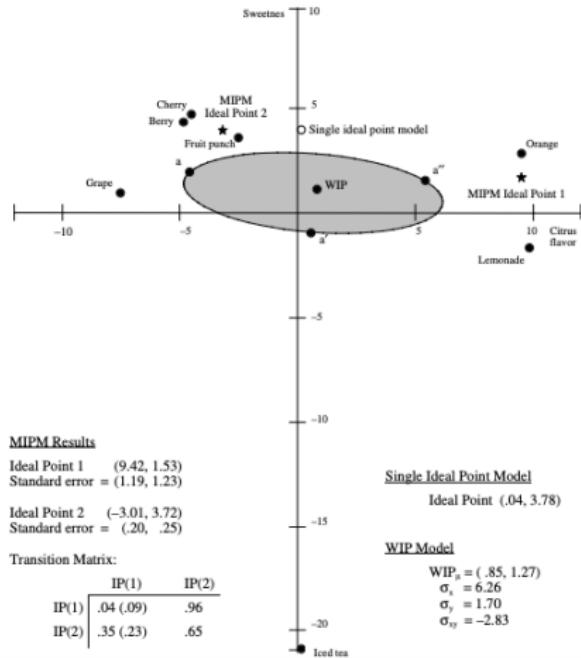
Model	Hit Rate	(Std. Dev.)
Disentangled Embedding + Logit Model (Homogeneous)	63.16%	(2.34%)
Disentangled Embedding + Neural Net (Homogeneous)	65.81%	(2.22%)
Disentangled Embedding + Neural Net (Observable Heterogeneity)	67.52%	(0.92%)
Pretrained Deep Learning Model (Observable Heterogeneity)	68.31%	(1.54%)
Disentangled Embedding + HB Model (Observable + Unobservable Heterogeneity)	71.61%	(1.87%)

- Pretrained Deep learning model is trained on millions of images, and has millions of parameters
- Our HB model also has a small number parameters, and all predictions are based on only 6 visual characteristics

# Ideal Point

## Fruit Punch: Example from Lee, Sudhir and Steckel (2002)

Figure 2  
MIPM RESULTS FOR HOUSEHOLD 057



# 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*

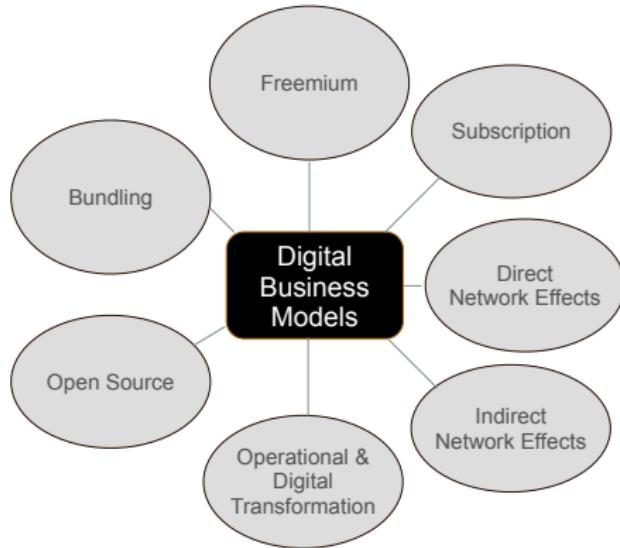
# Conclusion

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*

# Research Overview

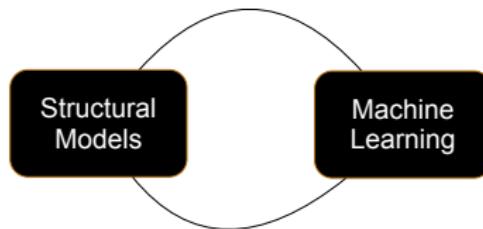
# Research Overview – Substantive



# Research Overview – Substantive

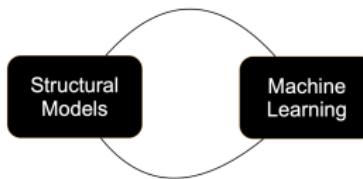


# Research Overview – Methodological



- Structural Models:
  - Linear Estimation of Aggregate Dynamic Discrete Demand for Durable Goods without the Curse of Dimensionality, with C. Chou and T. Derdenger  
**Marketing Science**
  - Estimating Dynamic Discrete Choice Models with Aggregate Data: Properties of the Inclusive Value Approximation, with T. Derdenger, **Quantitative Marketing and Economics**

# Research Overview – Methodological



- Machine Learning:
  - Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve, with I. Weaver, *Major Revision at Marketing Science*
  - A Theory-Based Interpretable Deep Learning Architecture for Music Emotion, with H. Fong and K. Sudhir, *Major Revision at Marketing Science*
  - Automatically Discovering Unknown Product Attributes Impacting Consumer Preferences, with A. Sisodia and A. Burnap, *Revision at Journal of Marketing Research*

# The End

vineet.kumar@yale.edu