Computer Vision in E-commerce

How Walmart Uses Machine Learning in Computer Vision to Protect and Serve Customers

Alessandro Magnani amagnani@walmartlabs.com
Feng Liu feng.liu@walmartlabs.com

Walmart

Product Images Impact Our Customers

- Provide customers with instant information
- Are easily sharable between readers
- Allow us to quickly scan an article
- Are search engine optimization (SEO) friendly and tailorable to a user's needs

Our Two Missions: To Serve & Protect

- **Serve the Customer**
  - Product type classification
  - Visual Attribute Extraction
    - Furniture style
  - Optimal Image Selection for a Product
  - Visual Search
    - Near duplicate detection
- **Protect the Customer**
  - Non-Compliant Images: Logos, Badges, Nudity, Sexually Explicit, Racially Inappropriate

MISSION : TO SERVE

Product type classification

https://arxiv.org/abs/1611.09534

Improve product type classification with images

Correct product type is key

- Easier to navigate catalog
- High quality search results

Challenges

- Very Large set of product types
- Text data can be noisy
- Evolving taxonomy

Title model is the best single model

[Women's Woven Plaid Shirt]

→ Women Shirt

SAMSUNG 65" Class 4K UHD 2160p LED Smart TV

→ TV

Improve product type classification with images

”tank”

“skirt”

Technical approach

- Multi task fusion model https://arxiv.org/abs/1611.09534
- Models ensemble
- K-NN using visual embedding

Television Stands Accuracy ⬆ 5.3%

Walmart

Visual Attribute extraction

http://www.vision.ee.ethz.ch/webvision/2019/files/scandinavian_or_mid-century_modern.pdf

Visual Attribute Extraction

- Clean set of attributes is essential
  - Better compare products
  - Quickly navigate to desire product
- Hundreds of attributes and thousands of values
  - Material, shape, color, …

- Image can be used to extract values
  - Furniture: décor style, leg type, back style
  - Fashion: neck style, sleeve style

Visual Attribute Extraction

Challenges:

Small differences between values

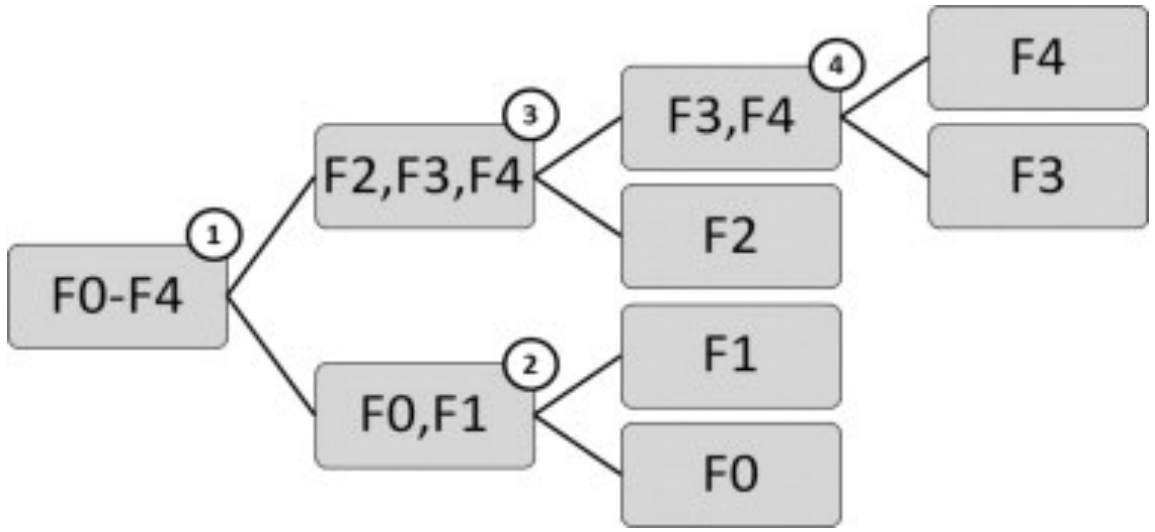
Ruffle Sleeve

Bishop Sleeve

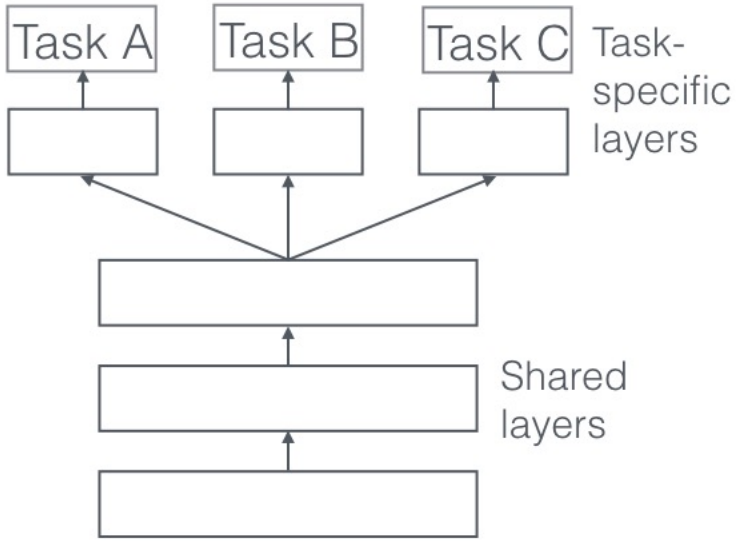
Quality of image/background

Technical Learnings

- Data is key
- Good tools for tagging can have big impact
- Multitask learning is often helpful
- Multi-stage classification with fine grain differences



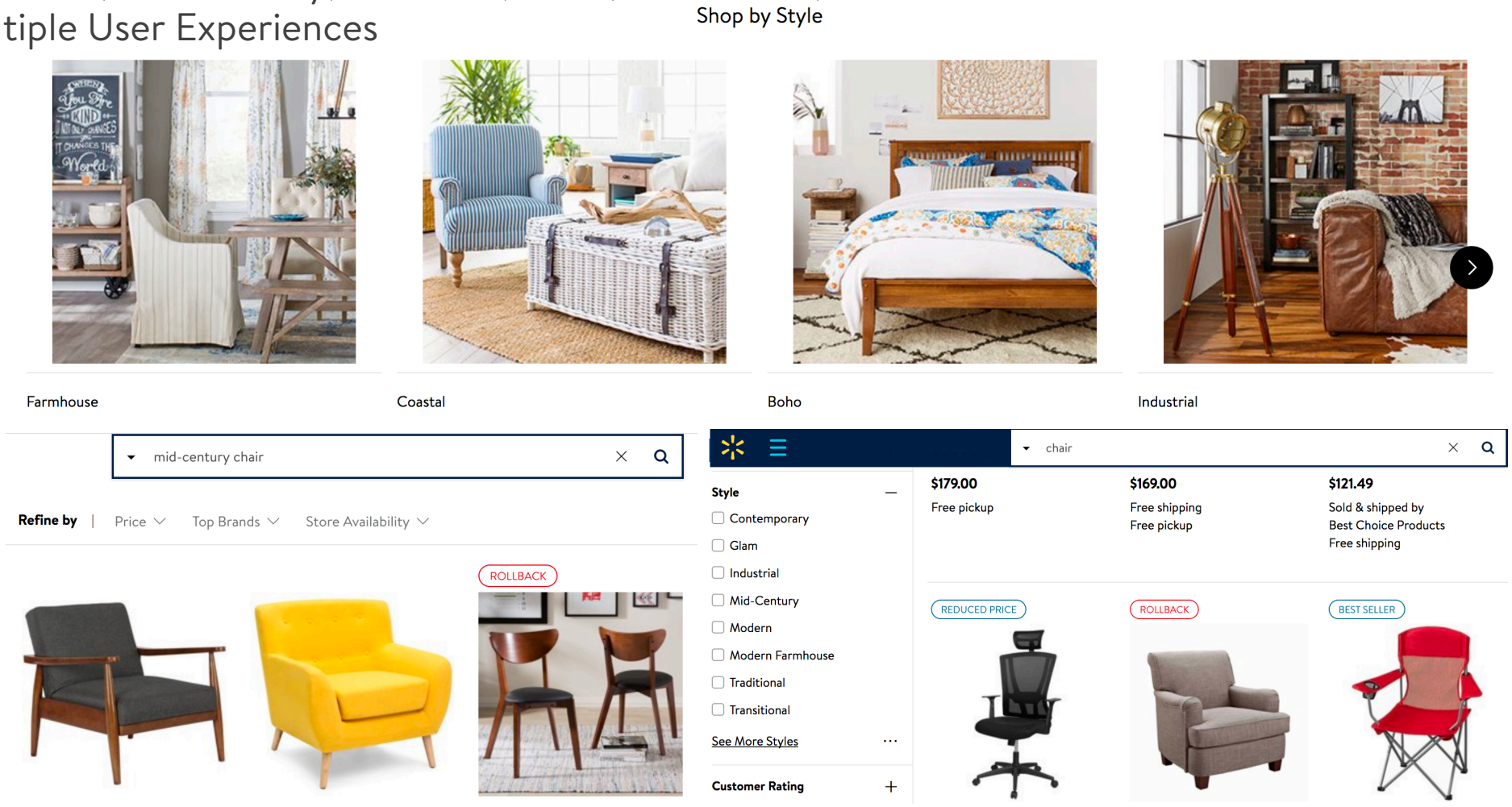
Multi-stage Classification



Multi Task Classification

Furniture Style Understanding

- Understanding style is key when selling furniture
- 16 Styles (Coastal, Mid-century, Industrial, Boho, Farmhouse, …)
- Powers Multiple User Experiences



Furniture Style Understanding

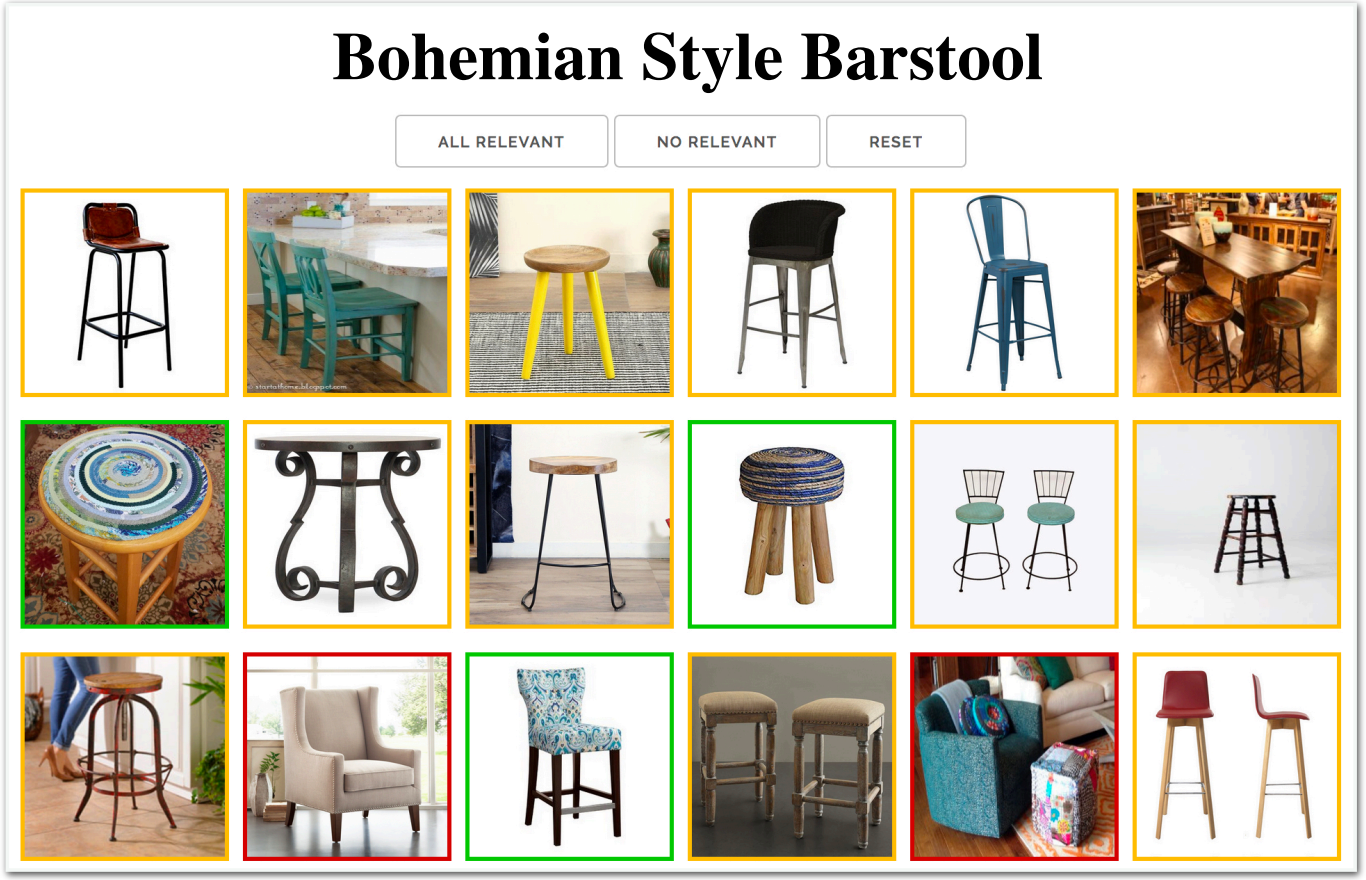
- Some categories are easily mis-classified, such as
  - Contemporary vs Modern

Furniture Style Understanding

- Some categories are easily mis-classified, such as
  - Scandinavian vs Mid-century

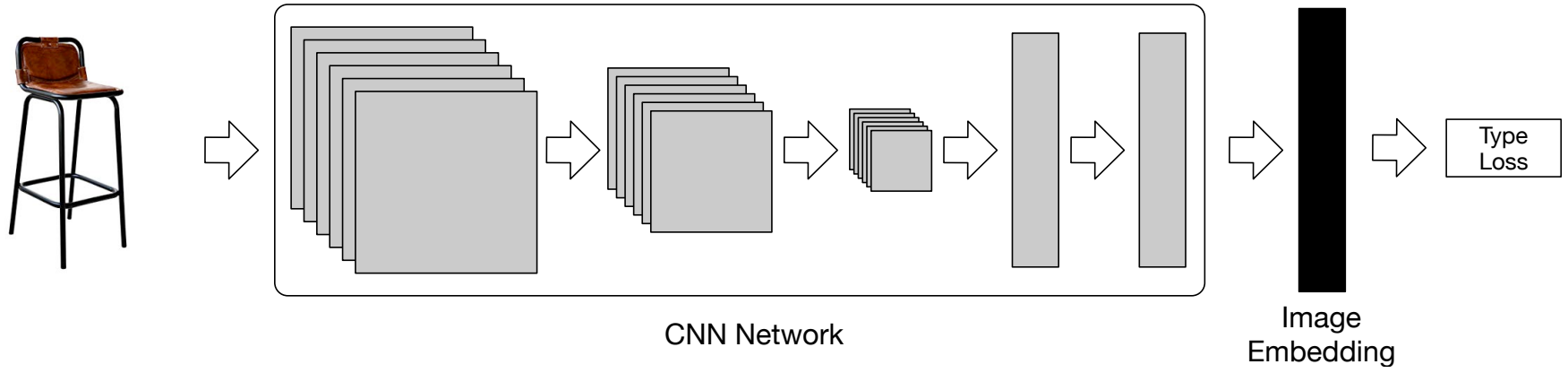
Editor tools to improve data collection

- Reduce context switching for editors
  - Focus on one style and one PT
- Allow editor to skip evaluation
- Show multiple images at once



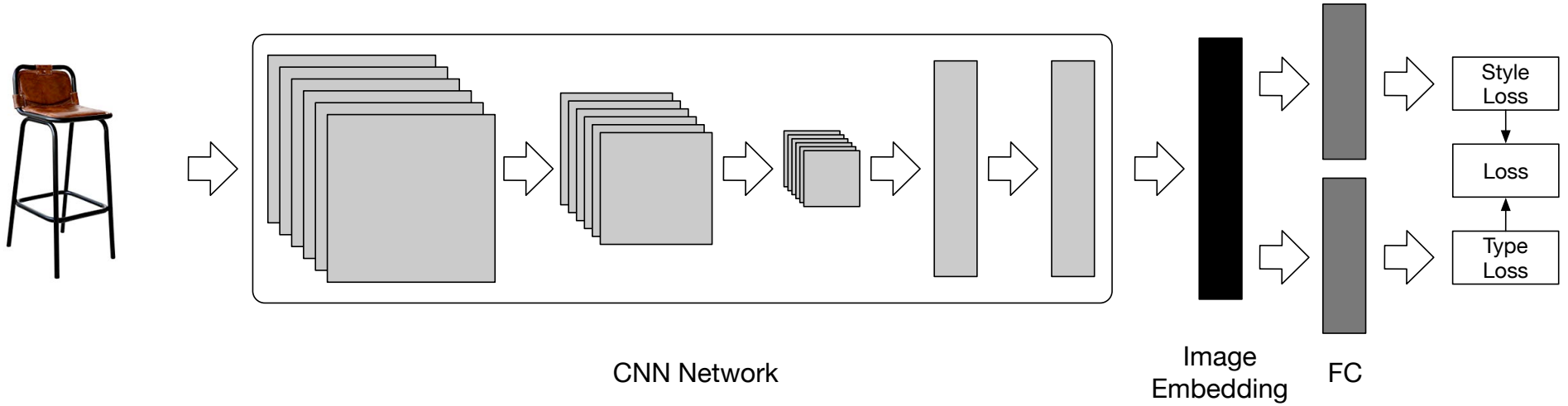
Multi-task Learning

- Two labels to learn: category and style
- Pre-trained networks are used to get embedding
- From the same embedding, build two classifiers
- Combine the objective functions into one



Multi-task Learning

- Two labels to learn: category and style
- Pre-trained networks are used to get embedding
- From the same embedding, build two classifiers
- Combine the objective functions into one



Multi-task Learning

- Best style accuracy comes from model that learns both category and style
- Learning to differentiate categories helps to learn style

	Top-1 Accuracy	Top-5 Accuracy
style-only loss	0.5850	0.9227
0.8 x style + 0.2 x category loss	0.6090	0.9312

CRACKING STYLE OF FURNITURE: AN E-COMMERCE PERSPECTIVE

http://www.vision.ee.ethz.ch/webvision/2019/files/scandinavian_or_mid-century_modern.pdf

A SMART SYSTEM FOR SELECTION OF OPTIMAL PRODUCT IMAGES IN E-COMMERCE

https://arxiv.org/abs/1811.07996

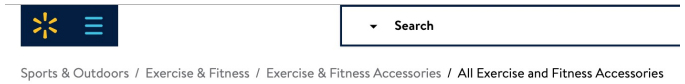
What Could Images Provide for E-Commerce?

Views

Features and details

Experience

In an Imperfect World …



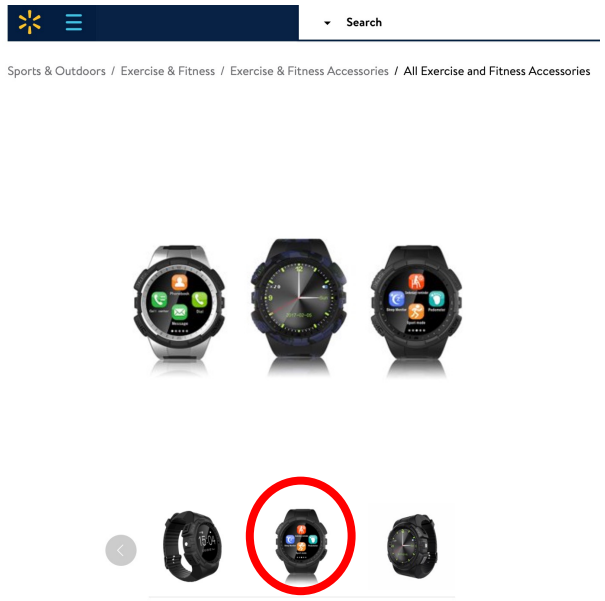
There aren’t enough pictures.

Primary image

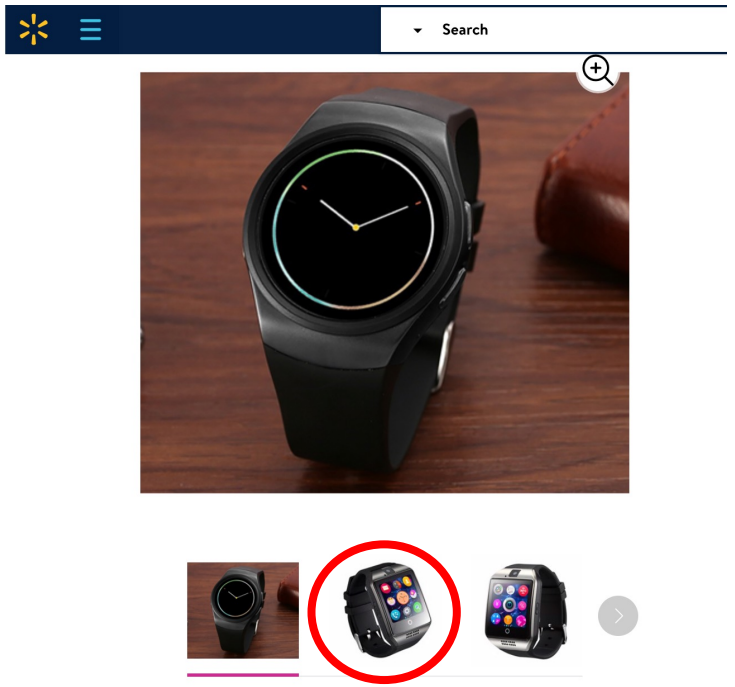
Additional images

Too many duplicate images.

More Examples



The primary image shows multiple products



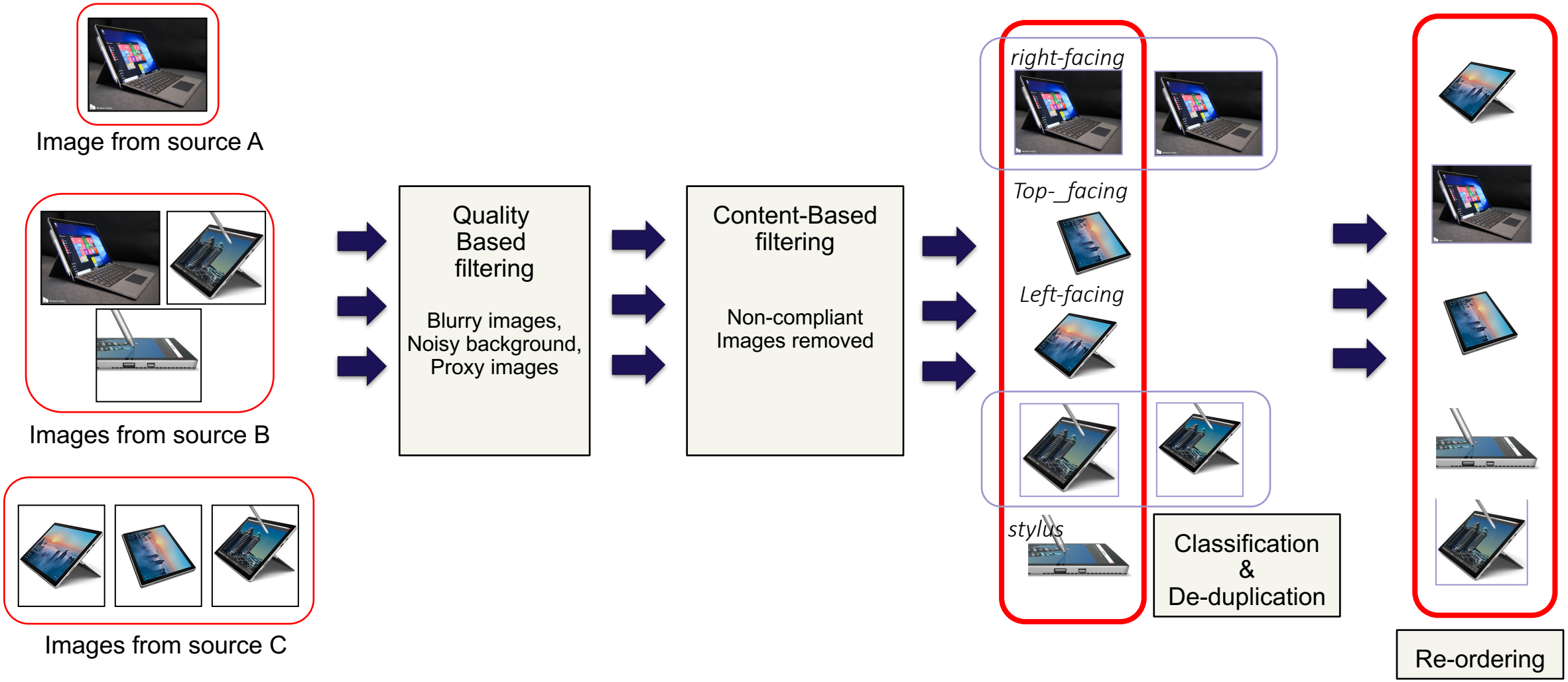
The primary image does not have a clean background

We circled better primary image options available in each product listing.

Proposed Data-Driven Solution

- Aggregate images from all sources
- Use computer vision and machine learning to understand images
- Remove images that don't meet quality and compliance standards
- Remove duplicate images from product listings
- Re-order images to quickly convey information efficiently

Proposed Framework

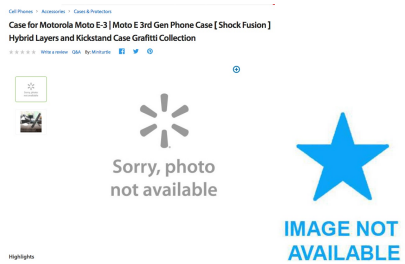


Goal: The final set of images is better than the images the original sources

Quality-Based Filtering

Image Quality Issues

Duplicate/multiple objects



Proxy or placeholder images

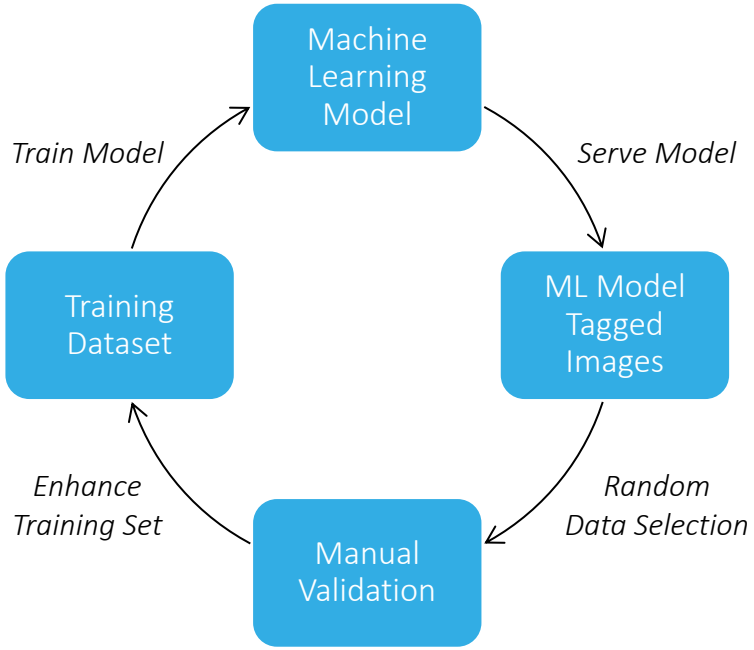
Blurry images

Images with excessive background space

and many more…

Outline of Technical Solution

- Treat each issue as a classification task
- Application-driven data augmentation
  - Artificially blur images
  - Programmatically add white spaces around the main object
- Train shallow classifiers on top of pre-trained deep learning models
- Continuous re-training until training set is big enough and model accuracy is high enough



Improving the Product Image Set

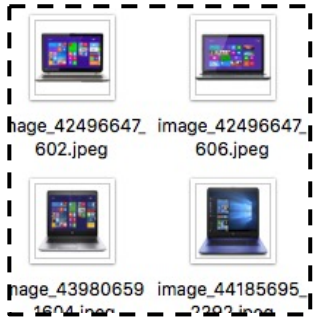
Classification and deduplication

Image Classification by Type

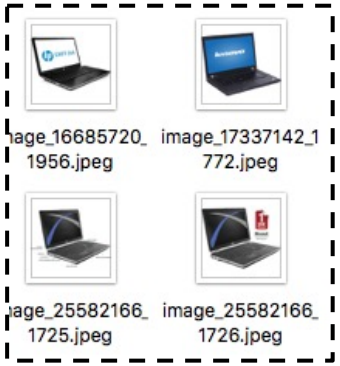
Figure out the important types of images for a category of items

Collect tagged image data

Train and evaluate a deep learning classifier



front_with_screen_on



right_facing

front_with_screen_on

left_facing

right_facing

...

full_left_facing

lifestyle

Model Details: Pre-trained networks fine-tuned with in-house data (a few thousand images per category)

Image Classification Model

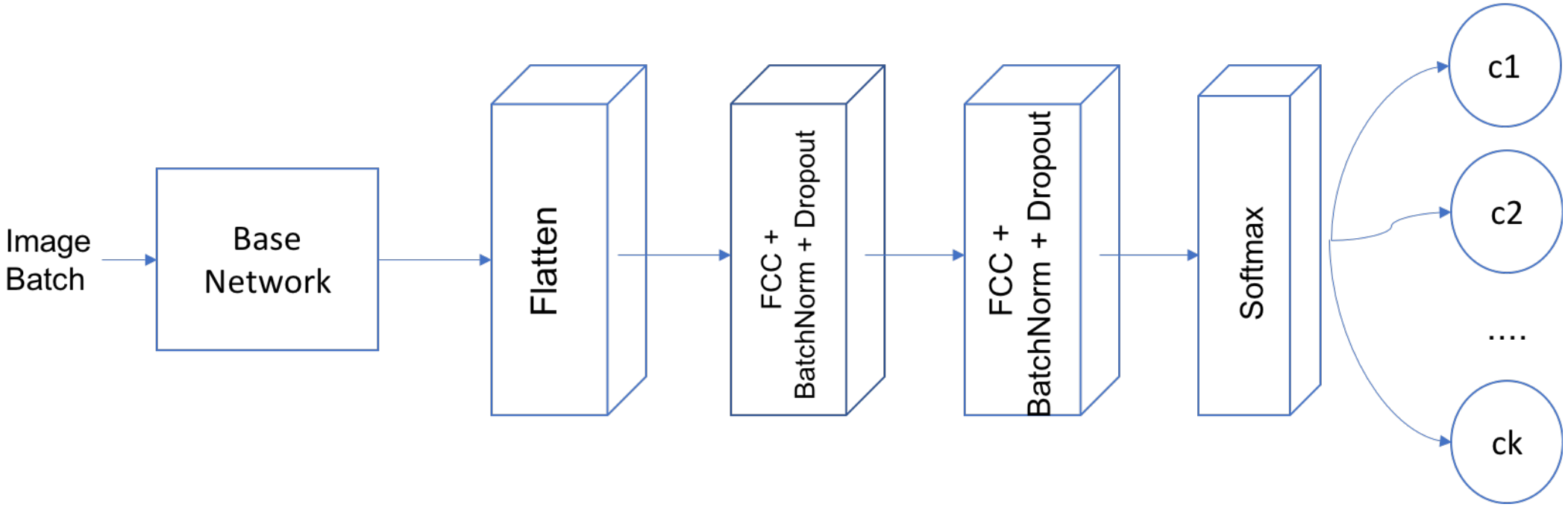


Image Classification Architecture

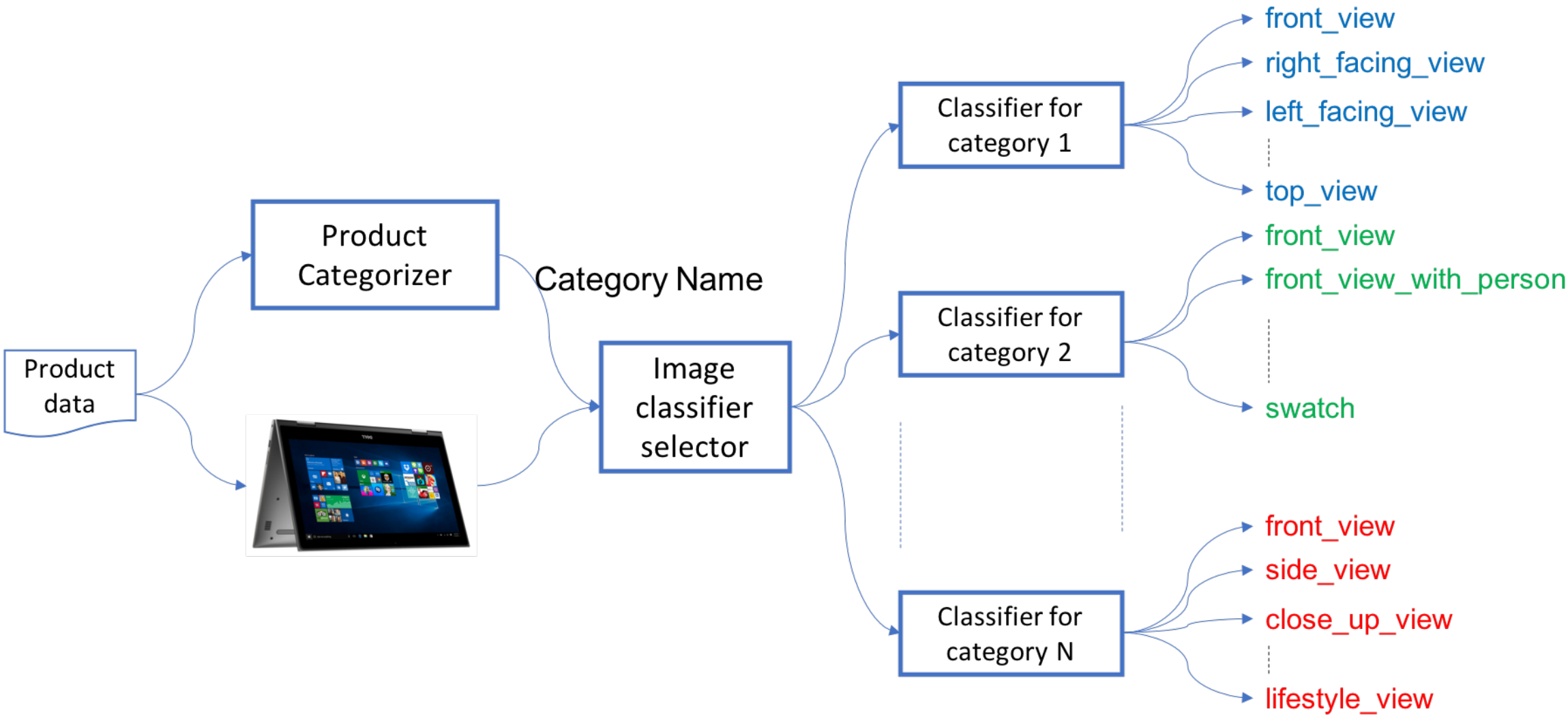
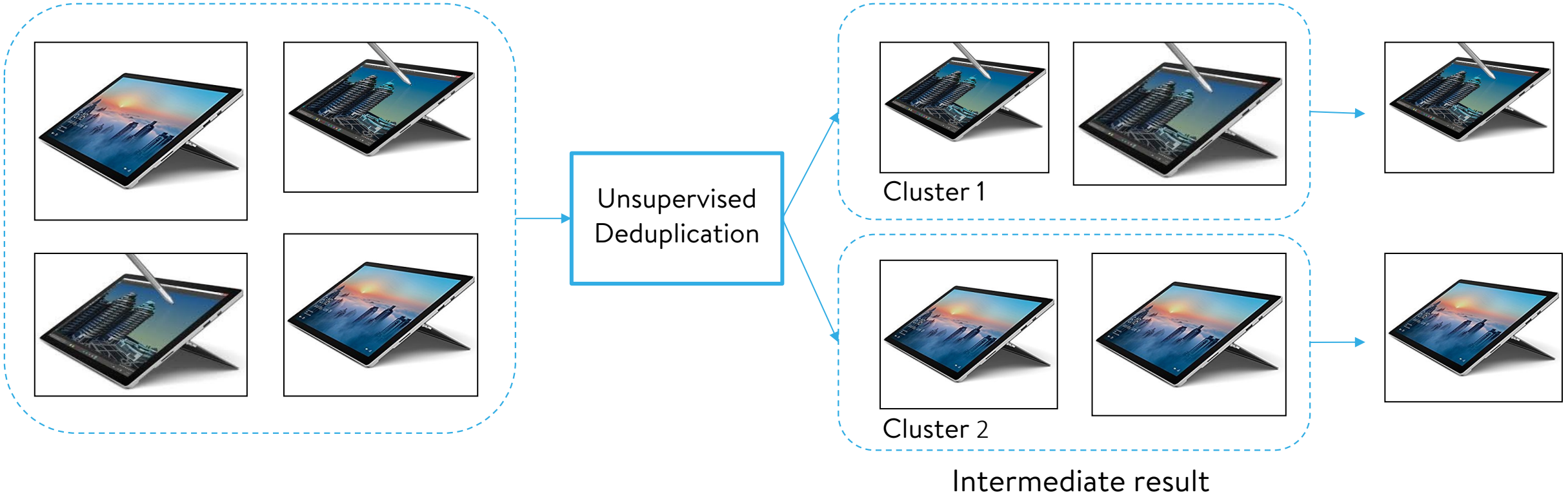


Image Classification Performance

Technique	Precision (%)	Recall (%)	F1-score (%)
Resnet50 + shallow	-0.01	-0.0	-0.0
Inception + shallow	-0.05	-0.04	-0.04
Resnet50 retrained	+0.0	+0.01	-0.0
Inception retrained	+0.06	-0.05	-0.05
VGG19 + shallow	x	x	x

- The numbers are based on the classification of Tablet Computers
- Smaller networks may perform better than larger ones by using smaller datasets for individual categories

Image Deduplication within a Type



Technical Challenges

Image Pairs considered duplicate

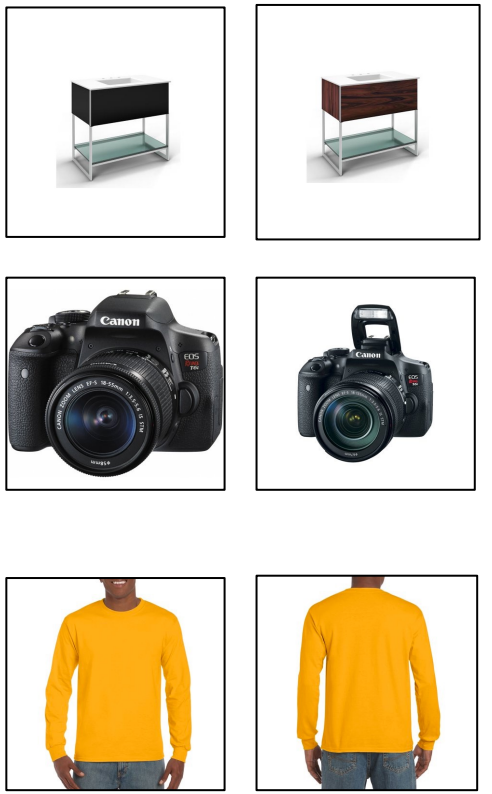
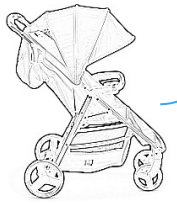
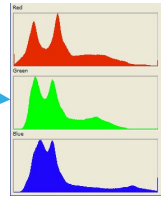
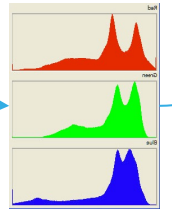


Image Pairs considered different

Difference hash

Difference hash

Classifier 1

Perception hash

Perception hash

Classifier 2

Color histogram

Color histogram

Classifier 3

Aggregator

Final Decision

Image Deduplication Performance

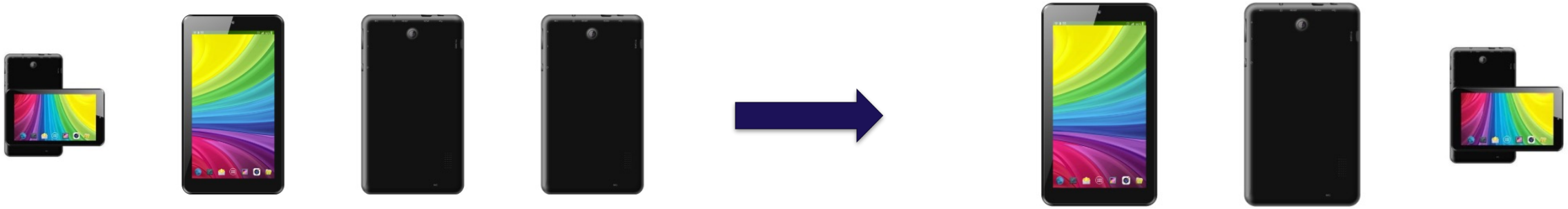
Technique	Precision (%)	Recall (%)	F1-score (%)
Cosine SImilarity	+0	-0.18	-0.1
Average Hash	-0.07	-0.05	-0.01
Perception Hash	-0.07	+0.0	-0.04
Wavelet Hash	-0.09	+0.01	-0.03
VGG19 + Cosine	-0.09	-0.15	-0.12
Inception + Cosine	-0.23	-0.2	-0.21
Resnet50 + Cosine	-0.10	-0.21	-0.15
Proposed Method	x	x	x

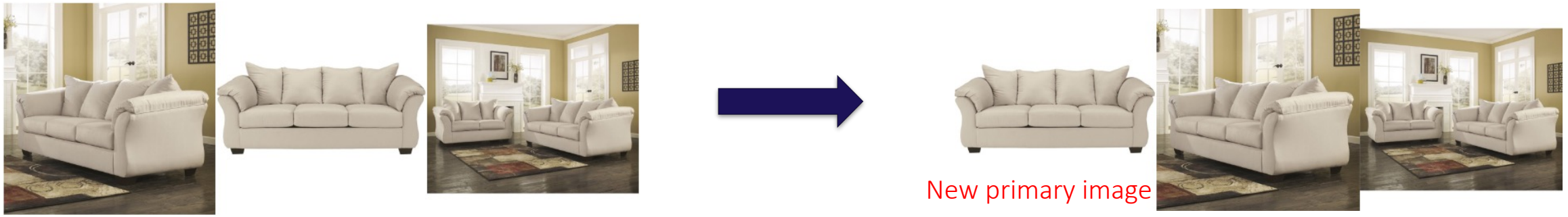
- The numbers are based on a benchmark dataset created out of a few thousand pairs
- Pre-trained deep learning classifiers are optimized to ignore subtle differences
- We chose a hash-based method over deep learning methods in favor of speed, interpretability, and lower cost of dataset creation

Image Reordering

- Once it has identified the image type, images of a product can be re-ordered based on:
  - Optimal order pre-defined by business
    - Example: side view should precede a front view for tires
  - Optimal order derived from customer data

Results





More Results

New primary image

New image

New feature view

Causal Impact Analysis

Category	Relative Effect on ATC (%)	Probability (%) of Causal Effect
Laptops	-1.6	64
Tablets	24	99
Televisions	9	93
Monitors	27	99

Category	Relative Effect on Conversion (%)	Probability (%) of Causal Effect
Laptops	5.6	79
Tablets	34	99
Televisions	20	99
Monitors	39	99

VISUAL SEARCH

Transform retail experiences through image understanding

How do customers form a query?

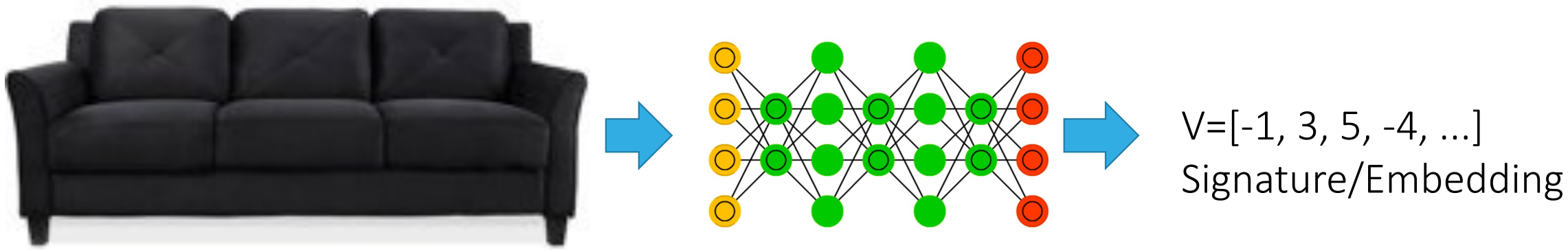
- Sofa or Loveseat?
- Blue or Teal?
- Chenille vs Cotton Fabric?
- Mid Century or Modern Style?

- Customers have detailed preferences in categories like clothing, furniture, home décor, etc.
- Hard-to-form text query expressing detailed preferences for color, style, material, etc.

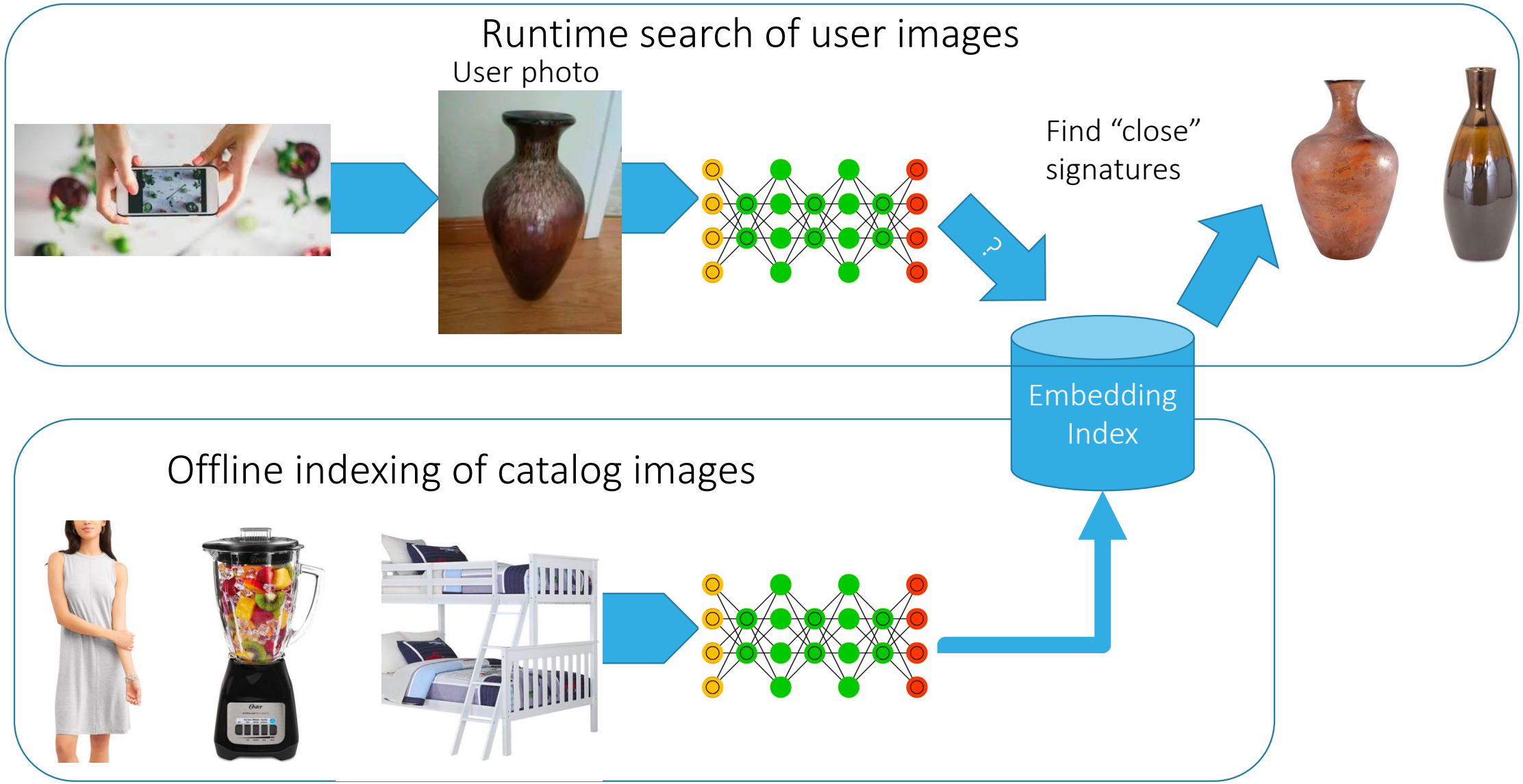
Architecture

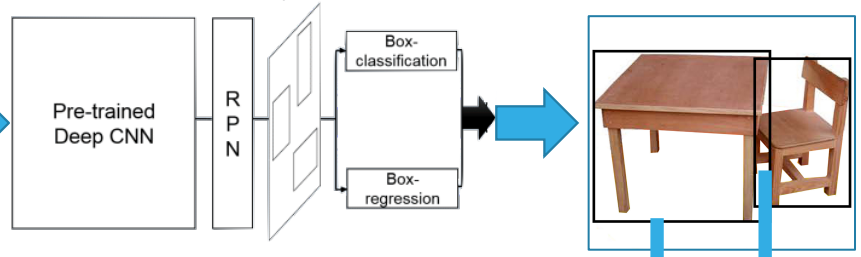
Given Product Images:

- Extract signature (embedding) using a neural network
- Signature captures product “essence”
- **“similar”** products should have **close signatures** (close in Euclidean norm)
- Images of the **same product** should have **identical signatures**

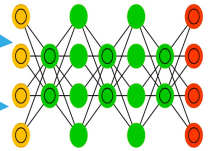


Architecture





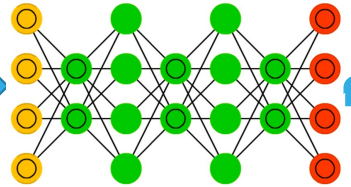
Find “close” signatures



?

Approximate Nearest Neighbors

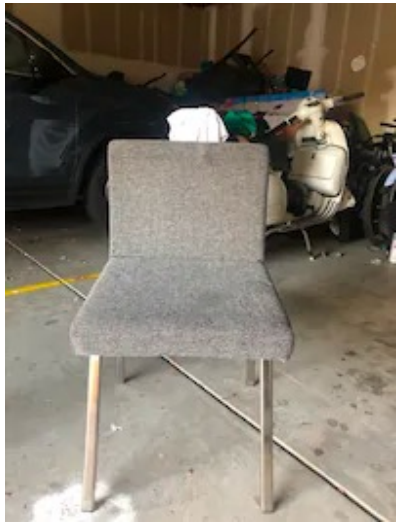
Offline indexing of catalog images



Embedding Index

Visual Search – Technical Challenges

Quality of user taken images



Lots of near duplicates

Scaling

- Very large Set of products/images
- Scaling ANN search

Dealing with user images

- More user taken data
- Heavy augmentation on product images

Near duplicate detection

Shallow Hashes

Elastic Search NN

https://arxiv.org/pdf/1902.08498.pdf

Walmart

Learnings & Future Work

- Data is the key
- Trade-off between deep learning and other methods should be evaluated
- Customer, data-driven personalized image ordering can lead to higher business metrics

But More Importantly

1. Be responsible
2. Be aware of bias

...

...3. Join Walmart Labs!

Questions?

amagnani@walmartlabs.com
Feng.liu@walmarlabs.com