Deep Learning in Fashion Industry

Hudson’s Bay Company | Gilt.com | Pau Carré Cardona | pcardona at gilt dot com
Introduction

- **Product Filtering Based on Automated Faceting**
  - Dataset Quality
  - Supervised Learning

- **Recommendations Based on Image Similarity**
  - Unsupervised Learning
  - Approximate Nearest Neighbours
Automated Product Faceting: Dataset Quality
Automated Faceting: Dataset Quality

Dataset wrongly labeled
Automated Faceting: Dataset Quality

Example of wrongly labeled sample:
Dataset: Long dress
Reality: Short dress
Automated Faceting: Dataset Quality

Dataset wrongly labeled → Train Network
Automated Faceting: Dataset Quality

- Dataset wrongly labeled
- Train Network
- Extract worst predictions
Automated Faceting: Dataset Quality

Prediction different from dataset label:
Dataset: **Long dress**
Network prediction:
- **Short dress**: 99%
- **Long dress**: 1%
Automated Faceting: Dataset Quality

Dataset wrongly labeled → Train Network → Extract worst predictions

Crowdsourced human-in-the-loop
- Qualification tests
- Gold standards
- Consensus voting
# Automated Faceting: Dataset Quality

<table>
<thead>
<tr>
<th>Long</th>
<th>Short</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
</tbody>
</table>
### Automated Faceting: Dataset Quality

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Dress" /></td>
<td><img src="image2" alt="Dress" /></td>
<td><img src="image3" alt="Dress" /></td>
<td><img src="image4" alt="Dress" /></td>
<td><img src="image5" alt="Dress" /></td>
</tr>
<tr>
<td>Long</td>
<td>Long</td>
<td>Long</td>
<td>Long</td>
<td>Long</td>
</tr>
<tr>
<td>Short</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

- **First Dress**: Long ( ✔️ )
- **Second Dress**: Long ( ✔️ )
- **Third Dress**: Long ( ✔️ )
- **Fourth Dress**: Long ( ✔️ )
- **Fifth Dress**: Long ( ✔️ )

*Note: The images represent different dresses with facets marked for long and short options.*
Automated Faceting: Dataset Quality

Gold Standard
We know it's long. Only adding to evaluate quality of worker's solution.

<table>
<thead>
<tr>
<th></th>
<th>Long</th>
<th>Short</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td></td>
<td>✗</td>
</tr>
<tr>
<td>Gold Standard</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>✗</td>
</tr>
</tbody>
</table>
Automated Faceting: Dataset Quality

Worker A

- Long:
- Short: 

Worker B

- Long:
- Short: 

Worker C

- Long: 
- Short: 

Automated Faceting: Dataset Quality

Worker A
- Long: 
- Short: 

Worker B
- Long: 
- Short: 

Worker C
- Long: 
- Short: 
Automated Faceting: Dataset Quality

- Dataset wrongly labeled
- Train Network
- Extract worst predictions
- Human-in-the-loop retags
  - Qualification tests
  - Gold standards
  - Consensus voting
Automated Product Faceting: Deep Learning Model
Automated Faceting: Learning to See

Deep Convolutional Neural Network as starting point

- We are using **Residual Networks**, but any state-of-the-art network will do.
- **Residual networks are based on bypassing layer’s input** to prevent information loss caused by layers themselves.
Automated Faceting: Learning to Localize

Facet information is sometimes located in specific parts of the image

- Neckline ~10% improvement using crops
Automated Faceting: Learning to Localize

Spatial transformer to locate features in the product image

- **No need for location dataset**, it learns to locate directly from classification error backpropagation

Spatial Transformer Networks - Google DeepMind - 2015
Automated Faceting: Learning to Read

- **Product’s** databases contain free text (description, name, brand, reviews …)
- **Recurrent neural networks** have issues
  - **Slow**: each word has to be fed into the network one at a time
  - **Difficult to train**: training is very unstable

Any Alternative? **dilated convolutions**
- Results comparable to images, sometimes better (handbags)
Automated Faceting: Learning to Read

- Convolutions only detect patterns between words close to each other

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.
Dilated Convolutions can detect patterns between words distant from each other.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.
Automated Faceting: Learning to Read

- **Dilated Convolutions** can detect patterns between words **distant** from each other

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

Silk chiffon poppy print u-neck dress with cap sleeves and gathered shoulder detail.

*Fast and Accurate Sequence Labeling with Iterated Dilated Convolutions - University of Massachusetts Amherst - 2017*
Automated Faceting: Learning to Learn

Each module should be able to be trained independently
Automated Faceting: Learning to Learn

Each module should be able to be trained independently

- Classifier
- ResNet
- Dilated Convolutional Neural Network
- Spatial Transformer
- Product Image
- Product Description
Automated Faceting: Learning to Learn

Each module should be able to be trained independently
Automated Faceting: Learning to Learn

Each module should be able to be trained independently

- ResNet
- Dilated Convolutional Neural Network
- Spatial Transformer
- Product Image
- Product Description

Classifier
Automated Faceting: Learning to Learn

Each module should be able to be trained independently
Product Image Similarity
**Similar Items**
Product Image Similarity: Unsupervised Similarity

Trained Deep Neural Network

Low level features: edges, lines, corners, curves...

High level features: shapes, textures...

Probability(\textit{flower}) = 99.9%
Probability(\textit{flower}) = 95.01%
...
Probability(\textit{flower}) = 0.05%

Network features can be generic enough to extract similarity between images
Product Image Similarity: Unsupervised Similarity

Image mapped into an n-dimensional vector using the output of an intermediate layer
Product Image Similarity: Unsupervised Similarity

Embeddings distance as dissimilarity metric
Product Image Similarity: Scaling Similarity

Dress embedding query

Dresses to compare to...
Product Image Similarity: Scaling Similarity

Dress embedding query

Dresses to compare to...

**Brute force** is **not efficient**: compare one embedding with all the other embeddings in stock
Product Image Similarity: Scaling Similarity

Approximate Nearest Neighbours

Index for Embeddings
N-dimensional Space Partitioning data structure
Leaf nodes contain clusters of embeddings close together
Product Image Similarity: Scaling Similarity

Approximate Nearest Neighbours

Index for Embeddings
N-dimensional Space Partitioning data structure
Leaf nodes contain clusters of embeddings close together

Search
Walk through nodes in the data structure until it reaches a leaf
Find embedding cluster
Compare embeddings within cluster
Product Image Similarity: Scaling Similarity

Approximate Nearest Neighbours

Index for Embeddings
N-dimensional Space Partitioning data structure
Leaf nodes contain clusters of embeddings close together

Search
Walk through nodes in the data structure until it reaches a leaf
Find embedding cluster
Compare embeddings within cluster

Fine tuning
Cluster size
Search in nearby leafs
thank you!

Twitter @hbcdigital

e-mail pcardona at gilt dot com

Join Us! https://www.gilt.com/careers/open-positions/
Links

- Spatial Transformer Networks
- Fast and Accurate Sequence Labeling with Iterated Dilated Convolutions
- Learning Fine-grained Image Similarity with Deep Ranking
  - [https://users.eecs.northwestern.edu/~jwa368/pdfs/deep_ranking.pdf](https://users.eecs.northwestern.edu/~jwa368/pdfs/deep_ranking.pdf)
- Approximate Nearest Neighbors Oh Yeah
  - [https://github.com/spotify/annoy](https://github.com/spotify/annoy)