Visual Search for Fashion

Divyansh Agarwal
Prateek Goel
Contents

- Problem Statement
- Motivation for Deep Learning
- Previous Work
- System Architecture
- Preliminary Results
- Improvements
- Future Work
What are they wearing?

Detect, classify, and describe clothes appearing in natural scenes with a focus on upper body clothing.
Motivation

- **E-commerce**
  - Automatic labelling of apparels inventory
  - Image based vertical search engines
- **Online Advertising**
  - Clothes reflect social status, lifestyle, age, gender
- **Image description**
  - ‘Man with gray sweater’, ‘Man with striped gray shirt’
- **Surveillance**
Why deep learning?

- Hand crafted features do not generalize well
  - Poor performance with HOG, SURF, SIFT (< 45%)
  - Need to bridge “semantic gap”

- Success of image-net
  - More robust features giving 90%+ accuracy on classification

- Availability of massive apparel centric data
  - Via Amazon, eBay, Instagram and Facebook
Relevant Work

- Apparel Classification with Style, Bossard et al.
  - Released dataset: Apparel Classification with Style (ACS)
  - Feature Extraction
    - HOG, SURF, LBP, and color information
  - Multiclass classification
    - One Vs All SVM, Random Forests & Transfer Forests
    - Accuracies 35.03% 38.29% and 41.36% respectively
Core Tasks

- Apparel Type Classification
- Apparel Attribute Classification
- Clothing Retrieval
- Object Detection
System Architecture

Query Image (User Input)

Apparel Type Classification

Apparel Attribute Classification

Cloth Type Tags

Feature Vectors

Feature Vectors

Cloth Attribute Tags

Clothing Retrieval

Top - 10 Results
Apparel Type Classification
Dataset for Apparel Type

- Apparel Classification with Style (ACS)
  - Apparel Classification with Style, Bossard et al., ACCV 2012
- 145,718 total images
- 89,484 cropped images
  - Training: ~71k  Testing: ~18k
  - 15 class labels
## Class distribution

<table>
<thead>
<tr>
<th>Category</th>
<th>Images</th>
<th>Boxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long dress</td>
<td>22,372</td>
<td>12,622</td>
</tr>
<tr>
<td>Coat</td>
<td>18,782</td>
<td>11,338</td>
</tr>
<tr>
<td>Jacket</td>
<td>17,848</td>
<td>11,719</td>
</tr>
<tr>
<td>Cloak</td>
<td>15,444</td>
<td>9,371</td>
</tr>
<tr>
<td>Robe</td>
<td>13,327</td>
<td>7,262</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Images</th>
<th>Boxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suit</td>
<td>12,971</td>
<td>7,573</td>
</tr>
<tr>
<td>Undergarment</td>
<td>10,881</td>
<td>6,927</td>
</tr>
<tr>
<td>Uniform</td>
<td>8,830</td>
<td>4,194</td>
</tr>
<tr>
<td>Sweater</td>
<td>8,393</td>
<td>6,515</td>
</tr>
<tr>
<td>Short dress</td>
<td>7,547</td>
<td>5,360</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Images</th>
<th>Boxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shirt</td>
<td>3,140</td>
<td>1,784</td>
</tr>
<tr>
<td>T-shirt</td>
<td>2,339</td>
<td>1,784</td>
</tr>
<tr>
<td>Blouses</td>
<td>1,344</td>
<td>1,121</td>
</tr>
<tr>
<td>Vest</td>
<td>1,261</td>
<td>938</td>
</tr>
<tr>
<td>Polo shirt</td>
<td>1,239</td>
<td>976</td>
</tr>
</tbody>
</table>

| Total        | 145,718 | 89,484 |

Architecture

- **Feature Extraction**
  - From pre-trained AlexNet

- **Multi-class Classifier**
  - SVM classifier for the 15 type classes
  - One vs Rest
Experiments

● Baseline SVM
  ○ Goal: Bossard et al.: 35% using hand-crafted features
  ○ Three sets of size 10k, 35k and 71k (randomly sampled).

● Confusion matrix
  ○ “Diagonal” indicates good differentiability between classes

● Drop classes
Baseline SVM

- Goal: 35%
- Three sets of size 10k, 35k and 71k
- 5-fold cross validation accuracy:
  - 10k: 31.8%
  - 35k: 30.1%
  - 71k: 27.3%
Drop Classes

- Reduce number of classes to 7:
  - drop blouse, cloak, robe, undergarment, uniform and vest
  - merge long and short dress, tshirt and polo shirt
- Cross validation accuracy: 36.7%
Clothing Attribute Classification
Clothing Attribute (CA) Dataset

- 26 attributes containing 1856 images
- Attributes having binary labels
  - Clothing pattern - Solid, Floral, Spotted etc.
  - Colors - Red, Blue, Green, Multi-color, etc.
  - Gender
  - Neckline Shape - V-shape, Round
  - Sleeve length
  - Collar Presence
  - And some more..
Clothing Attribute (CA) Dataset
Architecture

- **Feature Extraction**
  - From pre-trained Alex-Net

- **Multi-label Classifier**
  - SVM classifier for each of the 26 attribute
One Vs All SVM

- **Impressive results:**
  - Colors
  - Necktie
  - Patterns

- **Poor Results:**
  - Gender
  - Category
  - Neckline
  - Scarf
Clothing Retrieval
Clothing Retrieval

● Feature Extraction
  ○ Two set of features extracted from second fully connected layer (fc-7) in pretrained AlexNet

● Image Retrieval
  ○ KNN image retrieval algorithm
Improvements

- Poor results on Apparel Type Classification
  - Two phase fine-tuning of Alex-Net
- Fine tuning approach
  - Freeze all previous layers and retrain only the last inner product layer with 15 (or 7) classes.
  - Unfreeze all layers and retrain.
- Augment dataset using Amazon API to remove skew.
Future Work

- Incorporate object localization and detection task
- Expand dataset by leveraging e-commerce website APIs:
  - Imagenet - 1.2 million images
  - ACS : 89 K images
  - CA : 1.8 K images
Thank You!