SizeNet: Weakly Supervised Learning of Visual Size and Fit in Fashion Images

Nour Karessli  Romain Guigourès  Reza Shirvany
Zalando SE - Berlin, Germany
{nour.karessli, romain.guigoures, reza.shirvany}@zalando.de

Motivation
Finding clothes that fit is the biggest problem for customers shopping fashion online and offline. Supporting customers on their size and fit purchase decision in e-commerce context is particularly challenging:
- Thousands of new articles get activated everyday with short lifetime
- Return process takes from few days to few weeks resulting in zero or a few sales and returns data points for new articles

Contributions
In this paper we introduce a novel teacher-student approach on fashion images to:
- Investigate and demonstrates the rich value of fashion images in inferring size characteristics of fashion apparel
- Effectively tackle the challenging cold start problem of providing size advice for new articles with zero/low return data
- Generate large scale confidence-weighted weak annotations from crowds subjective feedback- enabling us to control the influence of weak annotations on the final model

Related Work
Teacher-Student Transfer Learning
Transferring knowledge from privileged information space to decision space [1]:
- Teacher leverages privileged historical weakly annotated data of sales and returns
- Student uses this knowledge to learn from images in decision space

SizeNet: Learning Visual Size Cues

Evaluation

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Baseline</th>
<th>With sample weighting</th>
<th>Without sample weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes-AUC</td>
<td>0.78</td>
<td>0.75</td>
<td>0.70</td>
</tr>
<tr>
<td>Attributes-AP</td>
<td>0.81</td>
<td>0.77</td>
<td>0.74</td>
</tr>
<tr>
<td>ResNet-AP</td>
<td>0.74</td>
<td>0.74</td>
<td>0.70</td>
</tr>
</tbody>
</table>

CNN Backbone Feature Extractor
Transfer learning using bottleneck features of pre-trained network

MPL: Multi-Layer Perceptron
4 fully connected layers with nonlinear activations. Use weighted binary cross entropy loss based on estimator confidence score $\sigma$:
- $\log (1 + e^{-\sigma})$: the logarithmic transformation reduces the score skewness

Teacher Statistical Model
- $L$ = $\binom{n}{k} p^k (1-p)^{n-k}$, $\binomial$: binomial likelihood, $p$: expected size return rate of article category, $k$: number of article size returns, $n$: number of article sales. The estimator score is define as $s = \log(L)$

Dataset
Women textile including 12 categories such as dresses, blouses, jeans, skirts, etc.
- Class 
- #Articles # Images
- Size issue 66,892 69,064
- no size issue 58,152 58,321
- Total 127,044 127,485

Attributes Baseline
Replace article images with sparse k-hot encoding of human annotated binary fashion attributes $\rightarrow$ fashion images achieve comparable results

Weights Importance
Exploiting weights offers better generalization capacity

Student Prediction vs. Teacher Confidence
BOTTOM RIGHT almost no samples are misclassified by SizeNet when Teacher is certain of no size issue; TOP LEFT high density of correctly predicted samples by SizeNet where Teacher is unsure; TOP RIGHT samples show that SizeNet has learned accurately from Teacher; BOTTOM LEFT SizeNet misclassifies fewer samples where Teacher is unsure

Size Issue Explanations using RISE [5]
True Positives show localized heatmaps where False Positives are affected by article design

{(1) Image and Fashion. Learning using privileged information: rating fully annotated knowledge transfer. In: JMLR
(3) Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In: CVPR
(5) Petsiuk, V.; Das, A.; Saenko, K. RISE: Randomized input sampling for explanation of black-box models. In: BMVC
(7) Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In: CVPR
(9) Petsiuk, V.; Das, A.; Saenko, K. RISE: Randomized input sampling for explanation of black-box models. In: BMVC
(10) Image and Fashion. Learning using privileged information: rating fully annotated knowledge transfer. In: JMLR
(13) Image and Fashion. Learning using privileged information: rating fully annotated knowledge transfer. In: JMLR
(14) Image and Fashion. Learning using privileged information: rating fully annotated knowledge transfer. In: JMLR
}