Learning the Latent Look: Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images

Motivation
How to define similarity for fashion images?
- Query: CNN feature, manual style labels, stylistically similar?
- CNN feature: low diversity, inconsistent
- Manual style labels: consistent, diverse

Proposed Method
I. Discover styles from unlabeled images
II. Style-coherent representation

Approach
Localized attribute recognition
- Segmentation: DeepLab-DenseCRF
- Classification: Fine-tune ResNet50

Style modeling by Polylingual Latent Dirichlet Allocation (PolyLDA)
- attribute distribution: \( \phi_k \sim \text{Dir}(\beta) \)
- style composition: \( \theta_i \sim \text{Dir}(\alpha) \)
- style assignment: \( \omega_{ij} \sim \text{Multinomial}(\theta_i) \)
- observed attribute: \( x_{ij} \sim \text{Multinomial}(\phi_k) \)

Overview
1. Discovered styles
- tribal, chiffon, ...
- t-shirt, denim, ...

2. Style composition
- PolyLDA: decouples different regions
- LDA: decouples whole body

Experiments
Baselines:
- StyleNet [Simo-Serra et al. 16]: clothing feature trained on Fashion144K
- ResNet [He et al. 16]: last layer of (ImageNet pretrained) ResNet
- Attr-ResNet: ResNet-50 fine-tuned on our attributes
- Attribute: Indicator vectors

Discovered styles align with human defined style labels
- HipsterWars: NMI: 0.21
- DeepFashion: NMI: 0.012

Applications
Traversing styles:
- Bohemian
- Hipster

Summarizing styles:
- Given a gallery of photos
- Summarize by dominant styles