Exploiting Texture Cues for Clothing Parsing in Fashion Images

Tarasha Khurana[1]; Kushagra Mahajan[1]; Chetan Arora[1]; Atul Rai[2]
[1]IIIT Delhi, [2]Staqu Technologies

Abstract

We focus on the problem of parsing fashion images for detecting various types of clothing and style. The current state-of-the-art techniques formulate the problem as segmentation and typically rely on geometrical shapes and position to segment the image. However, specifically for fashion images, each clothing item is made of specific type of materials with characteristic visual texture patterns. Exploiting the texture for recognizing the clothing type is an important cue which has been ignored so far by the state-of-the-art.

In this paper, we propose a two-stream deep neural network architecture for fashion image parsing. While the first stream uses the regular fully convolutional network segmentation architecture to give accurate spatial segments, the second stream provides texture features learned from handcrafted Gabor feature maps, and helps in determining the clothing type. Our approach achieves state-of-the-art results on the standard benchmark datasets, such as Fashionista and CFPD.

Motivation

We show how the additional supervision from texture descriptors improves garment labelling. While Outfit Encoder by Tangseng *etal.* [2] mispredicts a portion of `top' as `sweater' in the first image and `stockings' as `skin' in the second, characteristic textures of these clothing items help our model to disambiguate between them.

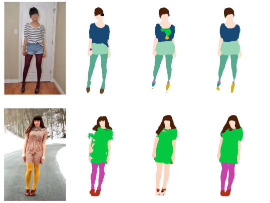

Fig 1. Motivation

Texture Descriptors

Gabor Features: Gabor feature responses are extracted corresponding to different wavelengths, orientations and phases. The combined set of these feature maps is used. Configurations used → wavelength: [3, 8] pixels, orientation: {0˚, 45˚, 90˚, 135˚} and 5 phase values spaced uniformly from 0 to the wavelength (λ).

Local Binary Patterns Features: LBP features are extracted over a sliding window of size 11×11. The number of neighbours is set to 8. For multiresolution feature extraction, we use feature maps comprising of 2, 3, 4 radius values. 2% increase in accuracy is achieved in case of multiresolution analysis.

Our Approach

Early Fusion: Texture feature maps are directly merged with the segmentation feature maps (experimented with merging at various layers). The merging was followed by a 5×5 convolutional layer to learn local context in the fused information.

Late Fusion: The two streams generate score maps for the clothing labels independently. These two score maps are concatenated and a 1×1 convolutional layer is applied to obtain the final category maps for each label.

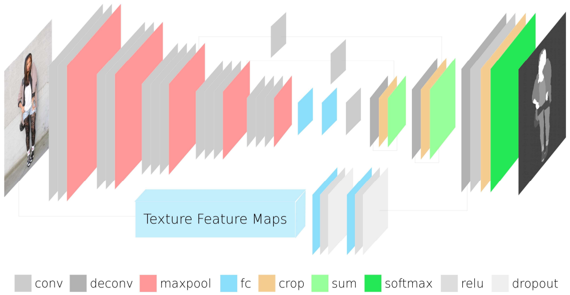

Fig 2. Proposed two-stream architecture for clothing parsing

Experimental Results

Table 1. Results of various configurations of proposed model on benchmark datasets.

	Fashionista		CFPD	
	Gabor	LBP	Gabor	LBP
scorefr	87.7	88.1	91.7	91.8
upscore4	88.9	88.3	92.3	92.5
upscore8	89.4	88.7	92.8	91.9
upscore8 + 1 conv	91.1	89.8	93.5	92.9
upscore8 + 2 conv	90.4	90.0	93.0	92.3

Table 2. Comparison with the state-of-the-art in terms of overall accuracy and overall IoU

	Fashionista		CFPD	
	Accuracy	IoU	Accuracy	IoU
OE	88.6	38.0	92.3	54.7
PaperDoll	84.7	-	87.1	-
SSL	84.8	33.2	88.5	49.1
CCP	90.2	-	-	-
DLV2 (Resnet)	86.6	36.8	89.9	48.3
DLV2 (VGG)	86.2	35.4	89.2	47.2
FCN-8s	87.5	33.8	91.6	51.2
Ours	91.1	42.1	93.5	58.7

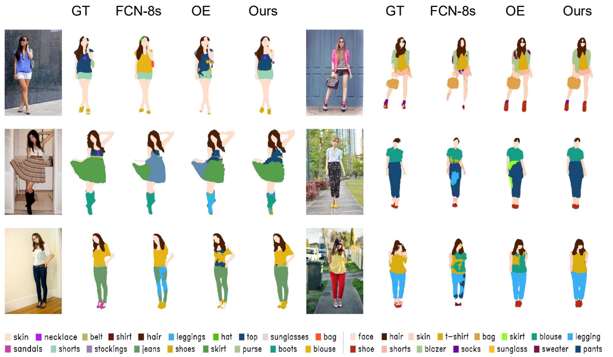

Fig 3. Comparison of results of proposed model on Fashionista (left) & CFPD (right).

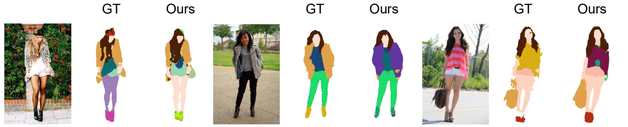

Fig 4. Examples of failure cases of proposed model

Conclusion

We show in our experiments that the proposed two-stream model helps in disambiguating similarly shaped but different textured clothing items, and achieves state-of-the-art performance on the various benchmark datasets.

References

[1] Shelhamer *et al.*, "*Fully convolutional networks for semantic segmentation*," TPAMI 2017.

[2] Tangseng *et al*. "*Looking at outfit to parse clothing*," CoRR 2017.

[3] Chen *et al*., "*Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*," TPAMI 2017.

[4] Yamaguchi *et al*. "*Paper doll parsing: Retrieving similar styles to parse clothing items*," ICCV 2013.