A High Performance CRF Model for Clothes Parsing

Edgar Simo-Serra[1] Sanja Fidler[2] Francesc Moreno-Noguer[1] Raquel Urtasun[2]

[1]IRI (CSIC-UPC)
Barcelona, Spain

[2]University of Toronto
Toronto, Canada

Problem:

- Semantic segmentation of clothing garments
- Large inter-class variability
- Fine-grained recognition task

Contributions:

- 30% over State-of-the-Art Performance
- Novel potentials that exploit the task
- Efficient model that dresses the person

CRF Model

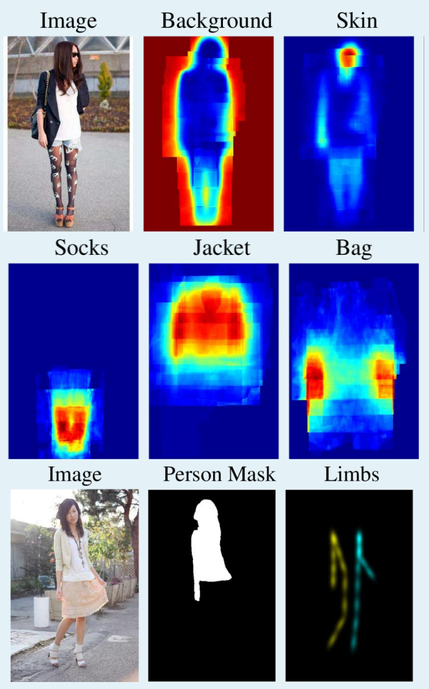
Number of Classes

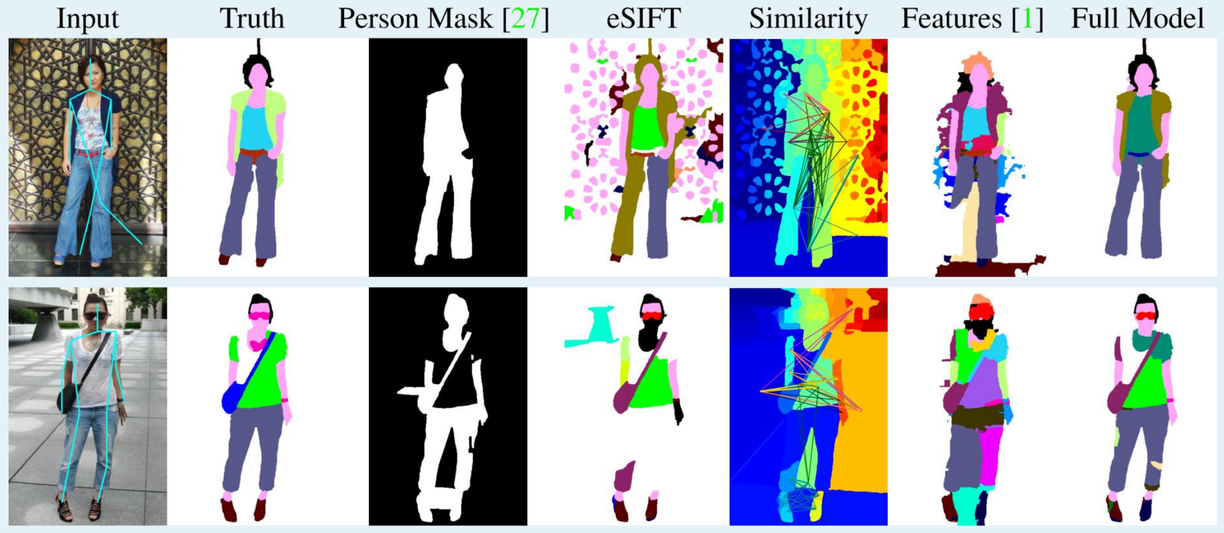
Superpixel labels: $y_i \in \{1,\dots,C\}$ Used to segment the image
Limb segment labels: $l_p \in \{1,\dots,C\}$ Correspond to each 2D limb

CRF Energy $E(y,l) = E_{unary}(y) + E_{similarity}(y) + E_{limbs}(y,l)$

Unary Potentials

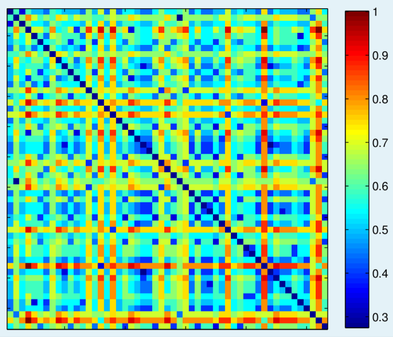
- *Simple Features [1]*
  - Logistic regression trained for one-vs-all
  - Color histograms (RGB and CIE L*a*b)
  - Texture histograms (Gabor filters)
  - Location features (relative position to 2D pose)
- *Person Mask*
  - Foreground/background segmentation by [3]
  - Avoids false positives on background
- *Clothelets*
  - Pose-conditioned garment likelihood
  - Masks for all garments are averaged based on 2D pose bounding boxes
- *Shape Features*
  - Logistic regression trained for one-vs-all
  - Enriched SIFT descriptors with second-order pooling [3]





Pairwise Potentials

- *Similarity (*Long-range connection between similar superpixels)
  - Shape, color and texture similarity [4] used with classifier
  - Minimum spanning tree to lower computational cost
- *Limbs* (Edges between limb nodes and superpixels)
  - Symmetric parts (e.g. both shins) are connected
  - Connection strength based on overlap



Loss Function

- Wordnet-based similarity between synsets of the different labels weighted by class occurence
- Learnt using maximum likelihood

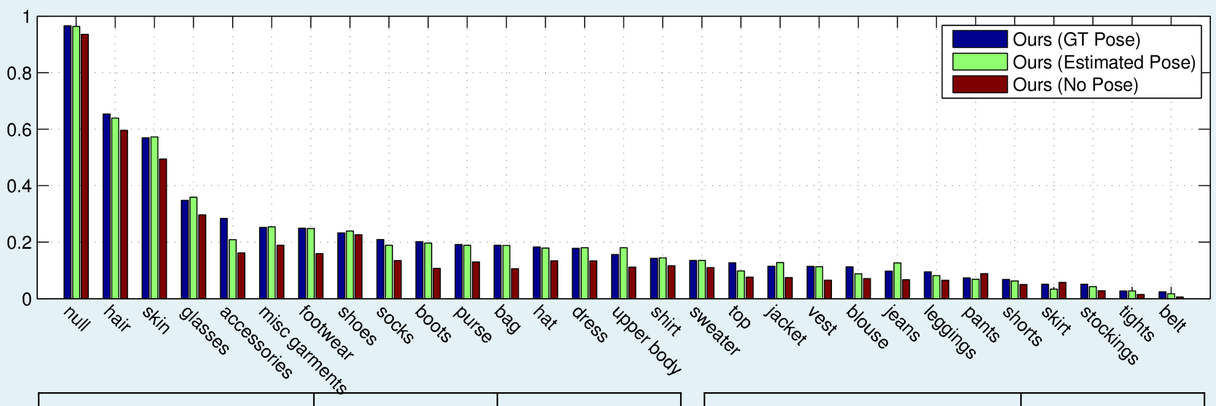
Results

	29 Classes		56 Classes		
Method	[1]	Ours	[1]	[2]	Ours
Jaccard index	12.32	**20.52**	7.22	9.22	**12.28**

Evaluation on Fashionista dataset [1]

Jaccard index (intersection over union) as metric

1. Fashionista v0.2 with 56 Classes using a 67-33 split as done in [2]
2. Fashionista v0.3 with 29 Classes using a 50-50 split



Method	29 Classes	56 Classes
Simple Features	13.80	7.93
Clothelets	8.91	3.02
eSIFT	16.65	9.29
eMSIFT	13.65	7.80

Method	56 Classes
Full Model	12.28
No Similarity	11.64
No Limb Segments	12.24
No Simple Features	10.07
No Clothelets	11.94
No Object Mask	10,02
No eSIFT	10.70
No eMSIFT	12.25

Contribution of each potential in the 56 class scenario. All potentials increase performance.

References

[1] Yamaguchi, K., Kiapour, M.H., Ortiz, L.E., Berg, T.L.: Parsing clothing in fashion photographs. In: CVPR. (2012)
[2] Yamaguchi, K., Kiapour, M.H., Berg, T.L.: Paper doll parsing: Retrieving similar styles to parse clothing items. In: ICCV. (2013)
[3] Carreira, J., Caseiroa, R., Batista, J., Sminchisescu, C.: Semantic segmentation with second-order pooling. In: ECCV. (2012)
[4] Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M.: Selective search for object recognition. IJCV 104 (2013) 154–171

Source Code at http://www.iri.upc.edu/people/esimo/research/fashion/