Deep Learning Face Attributes in the Wild

Supplementary Material

Ziwei Liu[1] Ping Luo[1] Xiaogang Wang[2] Xiaoou Tang[1]
[1]Department of Information Engineering, The Chinese University of Hong Kong
[2]Department of Electronic Engineering, The Chinese University of Hong Kong
{lz013,pluo,xtang}@ie.cuhk.edu.hk, xgwang@ee.cuhk.edu.hk

1. Network Structures

Table. 1 shows network structures for LNet and ANet. Detailed information are listed: filter number $\times$ filter size (*e.g.* 96×11^2), filter stride (*e.g.* str:4), pooling window size (*e.g.* pool:3^2), pooling stride (*e.g.* pool str:2). LRN represents local response normalization and (l) represents locally shared filters.

	C1	C2	C3	C4	C5
LNet	96×11^2 str:4,LRN pool:3^2 pool str:2	256×5^2 str:1,LRN pool:3^2 pool str:2	384×3^2 str:1	384×3^2 str:1	256×3^2 str:1 pool:3^2 pool str:2
ANet	20×4^2 str:1 pool:2^2 pool str:2	40×3^2 str:1 pool:2^2 pool str:2	60×3^2 str:1 (l) pool:2^2 pool str:2	80×2^2 str:1 (l)	-

Table 1. Network structures of LNet and ANet.

2. Effectiveness of LNet

Attribute-specific regions discovery Different attributes capture information from different regions of face. We show that LNet automatically learns to discover these regions. Given an attribute, by converting fully connected layers of LNet into fully convolutional layers following [2], we can locate important region of this attribute. Fig.1 shows some examples. The important regions of some attributes are locally distributed, such as 'Bags Under Eyes', 'Straight Hair' and 'Wearing Necklace', but some are globally distributed, such as 'Young', 'Male' and 'Attractive'.

More examples of LNet response maps Fig.4 shows more examples of LNet response maps on full images under different circumstances (lighting, pose, occlusion, image resolution, background clutter *etc*.).

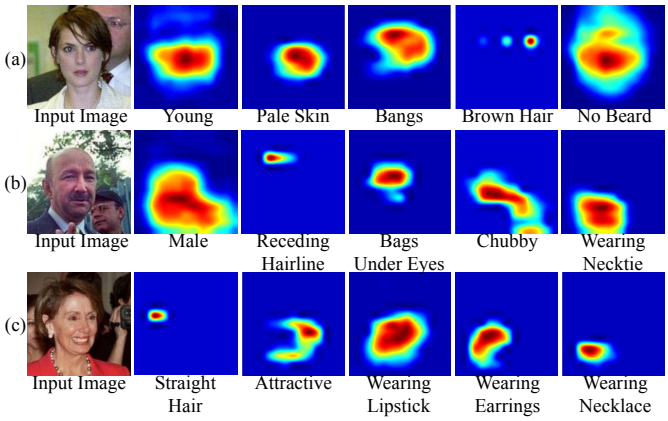

Figure 1. Attribute-specific regions discovery.

3. Effectiveness of ANet

Attribute Grouping Here we show that the weight matrix at the FC layer of ANet can implicitly capture relations between attributes. Each column vector of the weight matrix can be viewed as a decision hyperplane to partition the negatives and positive samples of an attribute. By simply applying k-means to these vectors, the clusters show clear grouping patterns, which can be interpreted semantically. As shown in Fig.2, Group #1, Group #2 and Group #4 demonstrate co-occurrence relationship between attributes, *e.g.* 'Attractive' and 'Heavy Makeup' have high correlation. Attributes in Group #3 share similar color descriptors, while attributes in Group #6 correspond to certain texture and appearance traits.

Semantic Concepts Emerging In the video we illustrate semantic concepts emerge w.r.t. the training iterations. Both pre-training and fine-tuning process are visualized.

4. Attribute Prediction

Performance on LFWA+ This experiment shows that the proposed approach can be generalized to attributes

	Asian	Indian	White	Black	Baby	Child	Middle Aged	Senior	No Eyewear	Frowning	Harsh Light.	Flash	Soft Light.	Outdoor	Curly Hair	F. V. Forehead	P. V. Forehead	Obs. Forehead	Eyes Open	Mouth Closed
FaceTracer [1]	86	87	74	91	97	83	77	78	84	78	73	82	69	73	66	75	80	77	80	74
PANDA-w [3]	80	84	71	88	97	81	74	76	78	77	69	80	61	72	68	73	83	80	74	69
PANDA-l [3]	86	92	82	93	97	83	76	78	82	89	**78**	87	**74**	79	75	79	81	82	70	82
LNets+ANet	**95**	**93**	**87**	**97**	**98**	**86**	**80**	**82**	**89**	**91**	77	**89**	71	**82**	**76**	**83**	**88**	**87**	**76**	**84**

	Teeth N. V.	Round Jaw	Square Face	Round Face	Color Photo	Posed Photo	Shiny Skin	Strong N. Lines	Flushed Face	Brown Eyes		**Average**
FaceTracer [1]	74	80	82	80	81	66	81	73	75	67		77
PANDA-w [3]	71	77	80	78	77	64	80	73	69	63		75
PANDA-l [3]	86	80	80	77	90	72	80	85	79	71		82
LNets+ANet	**87**	**82**	**83**	**84**	**91**	**75**	**84**	**87**	**83**	**75**		**85**

Table 2. Performance comparison of FaceTracer [1], PANDA-w [3], PANDA-l [3] and LNets+ANet on LFWA+.

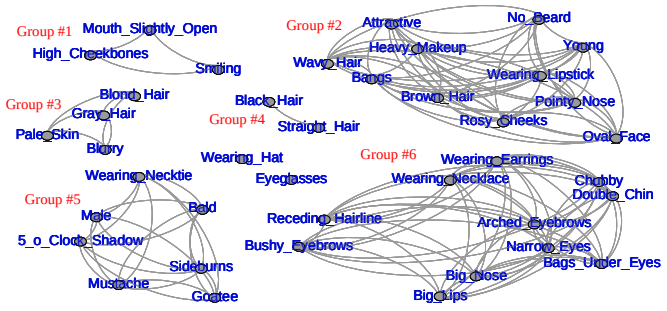

Figure 2. Attribute grouping.

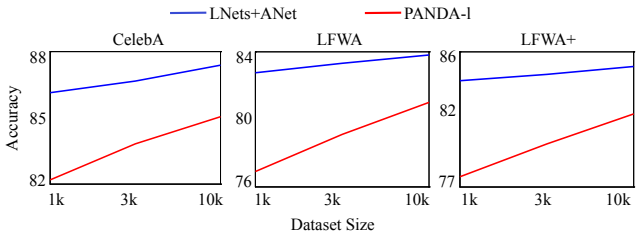

Figure 3. Performances of different sizes of training dataset.

which are not presented in the training stage. We manually label another 30 attributes on LFW and denote this extended dataset as LFWA+. Table.2 reports the attribute prediction results. LNets+ANet outperforms the other three approaches (FaceTracer [1], PANDA-w [1] and PANDA-l [1]) by 8, 10 and 3 percent on average, respectively. It demonstrates that our method learns discriminative face representations and has good generalization ability.

Size of Training Dataset We compare the attribute prediction accuracy of the proposed method with the accuracy of PANDA-l, regarding different sizes of training datasets. Fig.3 demonstrates that our method performs well when dataset size is small, but the performance of PANDA-l drops significantly.

References

[1] N. Kumar, P. Belhumeur, and S. Nayar. Facetracer: A search engine for large collections of images with faces. In *ECCV*, pages 340–353. 2008. 2

[2] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015. 1

[3] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev. Panda: Pose aligned networks for deep attribute modeling. In *CVPR*, 2014. 2

Figure 4. More examples of LNet response maps.