Supplementary Material: Learning Affinity via Spatial Propagation Networks

Anonymous Author(s)
Affiliation
Address
email

1 Proof of Theorem 3

Theorem 1. *Let* $\left\{p_{t,k}^K\right\}_{k\in\mathbb{N}}$ *be the weights in* w_t*, the model can be stabilized if* $\sum_{k\in\mathbb{N}}\left|p_{t,k}^K\right| \leq 1$.

Proof. Let λ be the eigenvalue of matrix w_t and λ_{max} be the largest one. According to Gershgorin′s Theorem [2], where every eigenvalue of a square matrix w_t satisfies:

$$|\lambda - p_{t,t}| \leq \sum_{k=1,k\neq t}^{n} |p_{k,t}|, \quad t \in [1, n] \tag{1}$$

then $|\lambda - p_{t,t}| + |p_{t,t}| \leq \sum_{k=1}^{n} |p_{k,t}|$. According to the triangular inequality, and since $\sum_{k=1,t\neq k}^{n} |p_{k,t}| \leq 1$, we have

$$\lambda_{max} \leq |\lambda - p_{t,t}| + |p_{t,t}| \leq \sum_{k=1}^{n} |p_{k,t}| \leq 1 \tag{2}$$

which satisfies the model stability condition. □

Theorem 1 (*i.e.*, Theorem 3 in the paper) shows that the stability of a linear propagation model can be maintained by regularizing all the weights of each pixel in the hidden layer such the summation of their absolute values is less than one. For the one-way connection, Chen *et al.* [1] maintain each scalar output p to be within $(0, 1)$. Liu *et al.* [4] extend the range to $(-1, 1)$, where the negative weights show preferable effects for learning image enhancers. This indicates that the affinity matrix is not necessarily restricted to be positive/semi-positive definite. (*e.g.*, this setting is also used for a pre-defined affinity matrix in [3].) For the three-way connection, we simply regularize the three weights (the output of a deep CNN) according to Theorem 1 without any positive/semi-positive definite restriction.

2 Parsing results on the HELEN dataset

In this section, we show more parsing results on the HELEN dataset. The detailed regions are cropped from the high resolution results. Figure 1 shows the effectiveness of the proposed spatial propagation network (SPN).

3 Semantic segmentation results on the PASCAL dataset

In this section, we show more semantic segmentation results (left) and object probability (*i.e.*, $1 - P_b$, where P_b denotes the probability of the background region) on the Pascal VOC 2012 dataset (Figure 2).

Submitted to 31st Conference on Neural Information Processing Systems (NIPS 2017). Do not distribute.

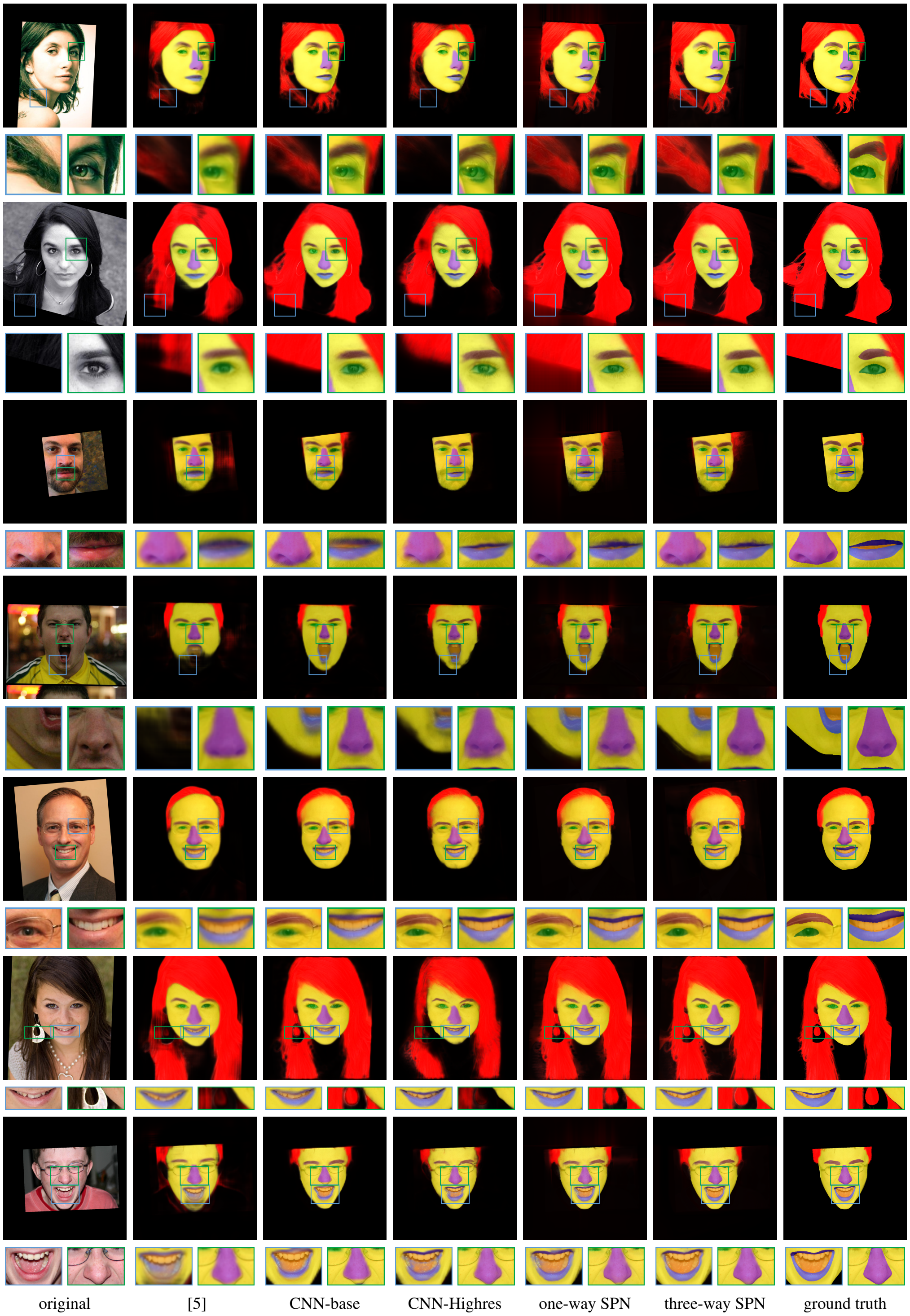

Figure 1: Parsing result on the HELEN dataset with detailed regions cropped from the high resolution results.

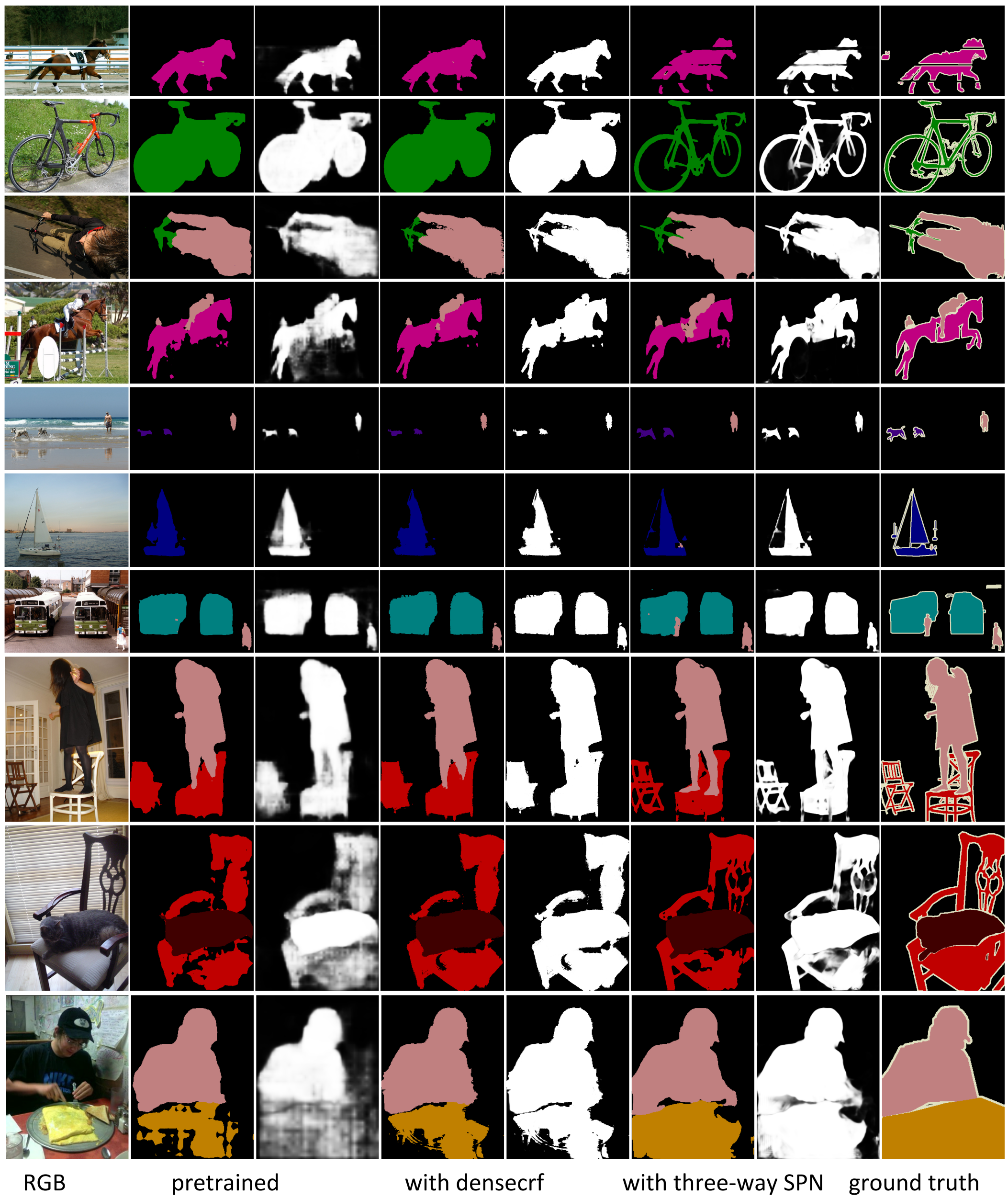

Figure 2: Visualization of Pascal VOC segmentation results (left) and object probability (by $1 - P_b$, where P_b denotes the probability of the background region). The results provided by the proposed three-way SPN framework are marked in the red rectangle.

References

[1] L. Chen, J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille. Semantic image segmentation with task-specific edge detection using cnns and a discriminatively trained domain transform. *arXiv preprint arXiv:1511.03328*, 2015.

[2] S. Geršgorin. Uber die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 1931.

[3] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228–242, 2008.

[4] S. Liu, J. Pan, and M.-H. Yang. Learning recursive filters for low-level vision via a hybrid neural network. In *European Conference on Computer Vision*, 2016.

[5] S. Liu, J. Yang, C. Huang, and M.-H. Yang. Multi-objective convolutional learning for face labeling. In *CVPR*, 2015.