Dynamic Facial Analysis: From Bayesian Filtering to Recurrent Neural Network (Supplementary Material)

Jinwei Gu Xiaodong Yang Shalini De Mello Jan Kautz
NVIDIA
{jinweig,xiaodongy,shalinig,jkautz}@nvidia.com

Abstract

The supplementary material includes two parts. The first part is a powerpoint presentation that includes the video results of head pose estimation and facial landmark localization as well as an example video of the SynHead dataset. The second part is this document where we provide more evaluation results for facial landmark localization on the 300-VW dataset [1].

1. Evaluation of RNN Variants

We first evaluate different types of RNN for facial landmark localization. In addition to the Post-RNN and FC-RNN used in the paper, we also implement and compare to the standard RNN and LSTM for facial landmark estimation. The results on the Split1 of 300-VW dataset are shown in Figure 1 and in Table 1, where all the variants of RNN are more accurate than the per-frame estimations. However, the FC-RNN outperforms all the other variants of RNN. For Post-RNN, since it uses the per-frame estimates as the input, it does not exploit mid-level features and thus, only slightly, improves the performance. Compared with the standard RNN and LSTM, as we explained in the main paper, FC-RNN maintains the structure of a pre-trained CNN to as much as possible and introduces fewer parameters, and thus is more robust and effective.

2. Comparison with 300-VW Challenge

We use the FC-RNN architecture in all the remaining experiments for the end-to-end learning. Figures 2, 3, and 4 are the plots of the cumulative error distributions for Splits1/2/3 of the 300-VW dataset. These plots are associated with Table 5 in the main paper. In the three splits, we used 80% (91) videos for training, and 20% (23) videos for testing. As we can see, the end-to-end learning with FC-RNN achieves the top performance in these splits. It is also significantly better than the recently proposed HyperFace [3] which employs CNN for per-frame estimation with a multi-tasking network.

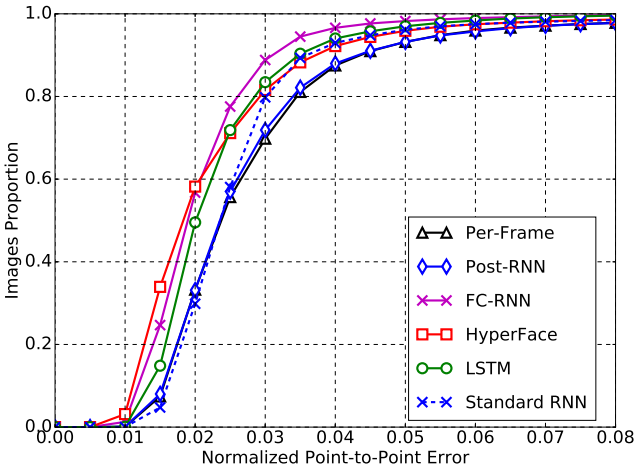

Figure 1: Comparison of the variants of RNN on Split1 of the 300-VW dataset.

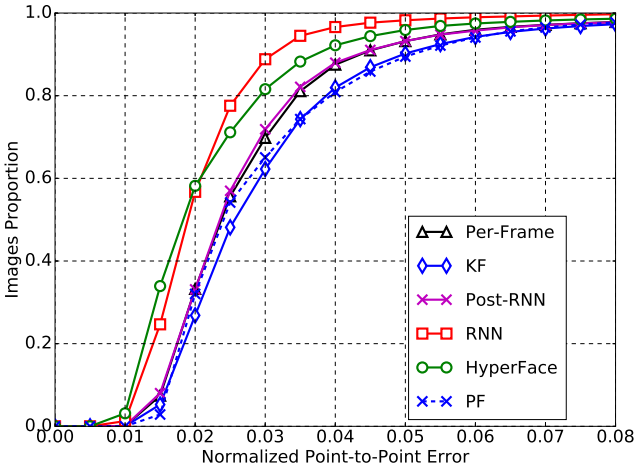

Figure 2: The cumulative error distributions for the variants of our methods for Split1 of the 300-VW dataset. We also compared with HyperFace [3].

Table 1: Comparison of the variants of RNN on Split1 of the 300-VW dataset. The areas under the curves (AUC) and failure rates (FR) are reported.

	Per-Frame	Post-RNN	Standard RNN	LSTM	FC-RNN	HyperFace [3]
AUC	0.66	0.66	0.68	0.72	**0.74**	0.73
FR	2.12	2.16	1.48	0.38	**0.28**	1.34

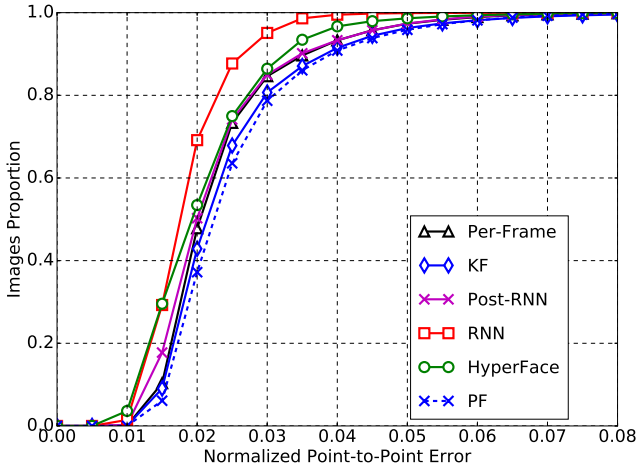

Figure 3: The cumulative error distributions for the variants of our methods for Split2 of the 300-VW dataset. We also compared with HyperFace [3].

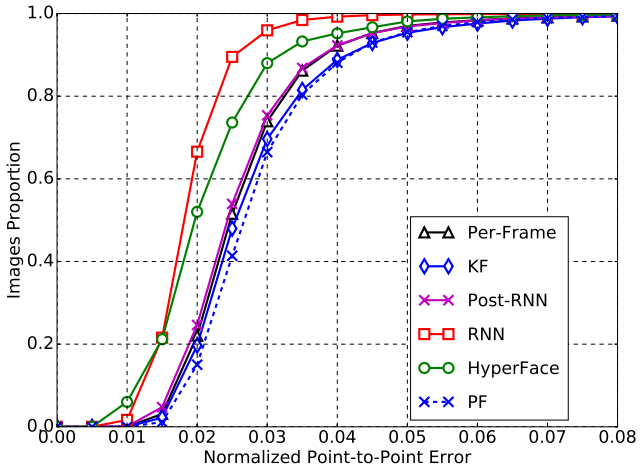

Figure 4: The cumulative error distributions for the variants of our methods for Split3 of the 300-VW dataset. We also compared with HyperFace [3].

References

[1] J.Shen, S.Zafeiriou, G. S. Chrysos, J.Kossaifi, G.Tzimiropoulos, and M. Pantic. The first facial landmark tracking in-the-wild challenge: Benchmark and results. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2015. 1

[2] G. Rajamanoharan and T. Cootes. Multi-view constrained local mmodel for large head angle face tracking. In *IEEE Proceedings of International Conference on Computer Vision, 300 Videos in the Wild (300-VW): Facial Landmark Tracking in-the-Wild Challenge & Workshop (ICCV-W)*, 2015.

[3] R. Ranjan, V. M. Patel, and R. Chellappa. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *arXiv preprint arXiv:1603.01249*, 2016. 1, 2

[4] S. Xiao, S. Yan, and A. Kassim. Facial landmark detection via progressive intialization. In *IEEE Proceedings of International Conference on Computer Vision, 300 Videos in the Wild (300-VW): Facial Landmark Tracking in-the-Wild Challenge & Workshop (ICCV-W)*, 2015.

[5] J. Yang, J. Deng, K. Zhang, and Q. Liu. Facial shape tracking via spatial-temporal cascade shape regression. In *IEEE Proceedings of International Conference on Computer Vision, 300 Videos in the Wild (300-VW): Facial Landmark Tracking in-the-Wild Challenge & Workshop (ICCV-W)*, 2015.