CRFace: Confidence Ranker for Model-Agnostic Face Detection Refinement
(Supplementary Material)

Noranart Vesdapunt
Microsoft Cloud&AI
noves@microsoft.com

Baoyuan Wang
Xiaobing.AI
zjuwby@gmail.com

We propose a confidence ranking network with a pairwise ranking loss to re-rank the predicted confidences locally within the same image to improve average precision (AP) across multiple face detectors on WiderFace [5] and our 8K resolution test set. In this supplementary, we provide performance curves for section 5.1 and 5.4, and additional visualization results from our 8K resolution test set.

1. Performance Curves

Ablation Study. We show performance curves for both of our ablation studies in section 5.1 on WiderFace validation set [5]. Our baseline is a re-implementation of HAMBox [2] and we replace ResNet-50 backbone with HRNet [4, 3]. We show the importance of ranking loss in Fig. 2 where the regression losses are struggling to preserve the confidence order and result in an even worse AP than the baseline, while our ranking loss improves AP by 1-1.2%. The importance of pair selection is shown in Fig. 3 where 10-pairs provides the best trade-off between the easy and hard pairs, thus achieving the highest AP.

8K Resolution. We show the performance curves of the single-scale, multi-scale, and our method on 8K resolution test set in Fig. 1. Our confidence ranker is model-agnostic, and it can improve AP across all three detectors (RetinaFace [1], HAMBox, HRNet), and allows us to train indirectly on 8K resolution to further boost AP without any speed decrease during test time.

2. Visualization of 8K Test Set

We show more visual comparisons between the single-scale, multi-scale, and our method on 8K resolution test set in Fig.4,5,6. In general, single-scale usually fails to predict large face due to the maximum anchor size is only 512 and the training data only has an image width of 1024. Multi-scale forwards the images 6 times on 3 resolutions and flip on each scale, so the smaller resolution performs a better job at predicting large face, but it also generates more false positive. Our method improves on top of single-scale, so we predict less false positive than multi-scale, and achieves the best result while running 1.7-2.1x faster than multi-scale.
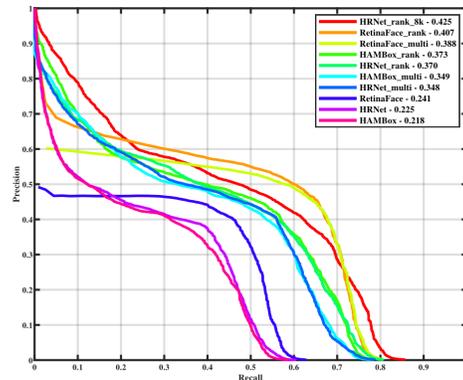
Figure 1. Performance curves on 8K test set. Our confidence ranker (rank) on a single-scale (scale) has higher AP than multi-scale (multi) on all backbones and runs 1.7-2.1x faster.

References

[1] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[2] Yang Liu, Xu Tang, Junyu Han, Jingtuo Liu, Dinger Rui, and Xiang Wu. Hambox: Delving into mining high-quality anchors on face detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[3] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *CVPR*, 2019. 1

[4] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *TPAMI*, 2019. 1

[5] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *CVPR*, pages 5525–5533, 2016. 1
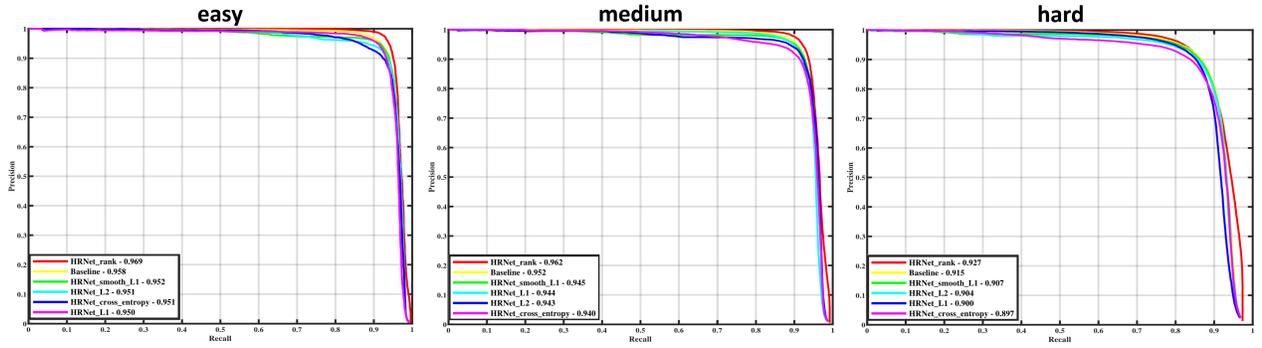
Figure 2. Performance curves for each loss function. Regression losses decrease AP due to the focus on confidence magnitude rather than confidence order. Our ranking loss improves AP Hard by +1.2%.
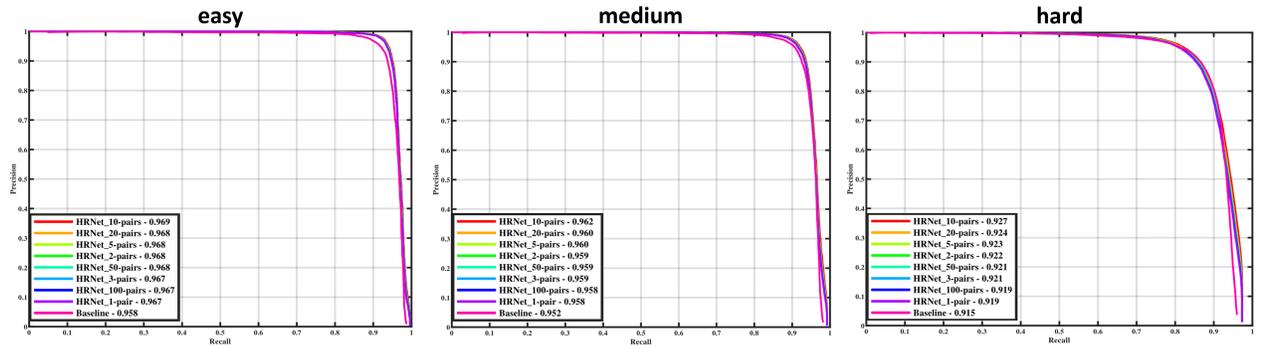
Figure 3. Performance curves for each number of pairs for pairwise ranking loss. Choosing a single pair will omit easy pairs, and choosing 100-pairs will disregard hard pairs. 10-pairs provides the best trade-off between easy and hard pairs.
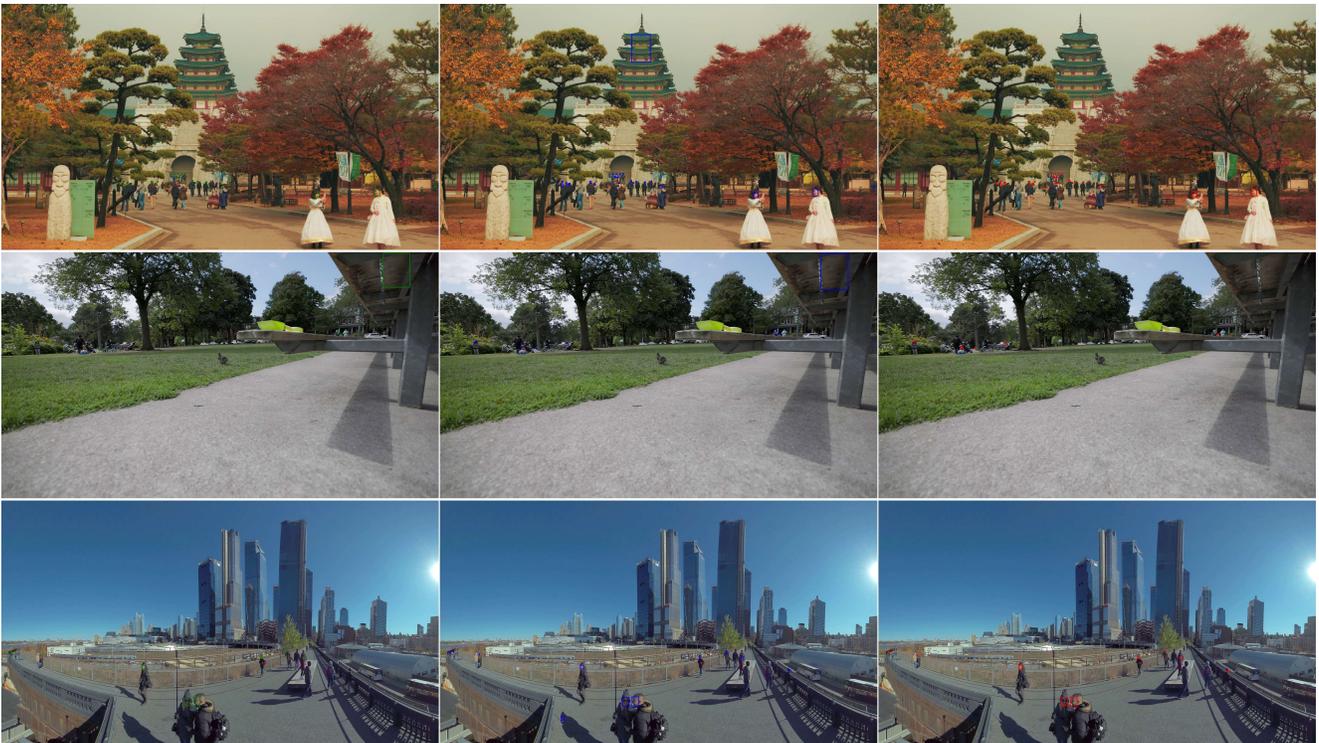
Figure 4. (Best viewed electronically) Prediction examples on 8K test set. We show the predictions from HRNet with confidences >0.5 on the single-scale (left), multi-scale (mid), and confidence ranker trained on 8K (right).
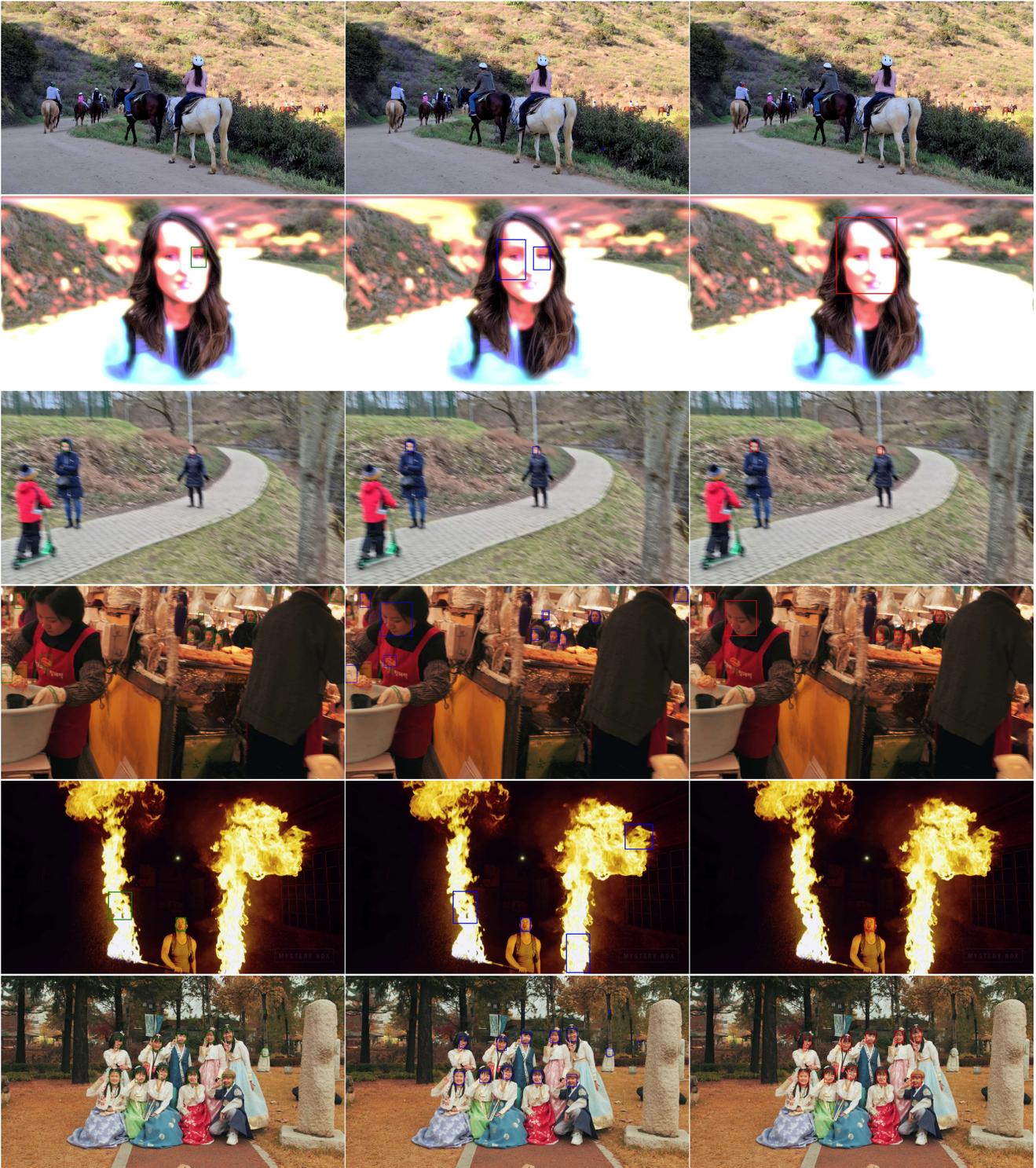
Figure 5. (Best viewed electronically) Prediction examples on 8K test set. We show the predictions from HRNet with confidences >0.5 on the single-scale (left), multi-scale (mid), and confidence ranker trained on 8K (right).
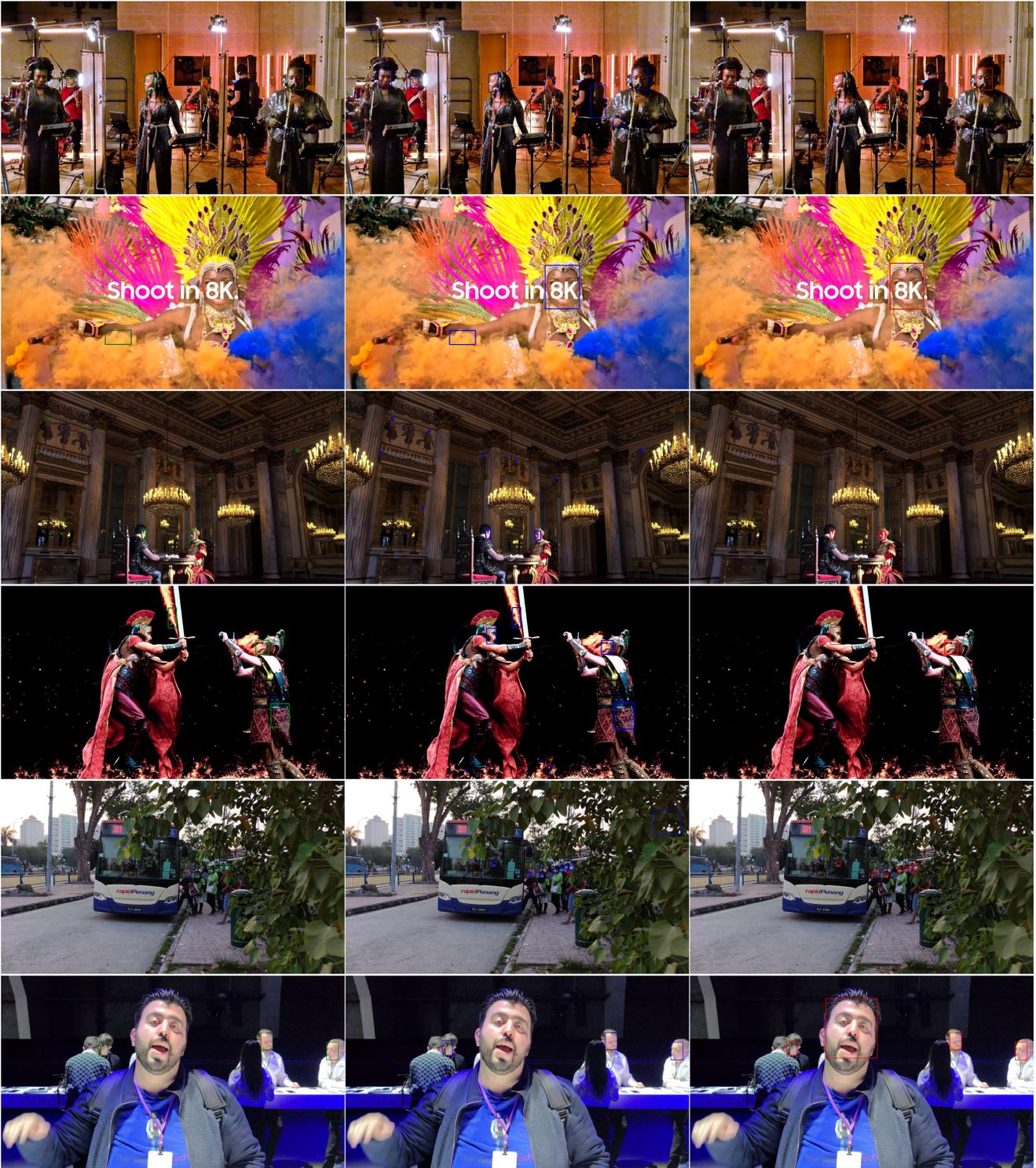
Figure 6. (Best viewed electronically) Prediction examples on 8K test set. We show the predictions from HRNet with confidences >0.5 on the single-scale (left), multi-scale (mid), and confidence ranker trained on 8K (right).

4