Composite Scoring of Speed and Accuracy for Face Detection Models

Yuanjun Xiong **Shuo Yang**

Amazon Rekognition and Video

Abstract

We propose a way to give a single score to a face detection model considering both its detection accuracy and runtime speed. This method will be used in the face detection track of the WIDER 2019 challenge. The major idea is that a better model than the reference one should achive Pareto improvement in terms of speed-accuracy trade off. We use the challenge reference model as the reference point and compare the method with the one proposed in [1]. The

1 Speed and Accuracy of Face Detection Models

Face detection is the foundation of modern computer vision based face analysis pipeline. With more and more applications in real-time analysis systems, a low runtime cost of the face detection models has become as important as high detection accuracy. In WIDER challenge 2019, we propose a new track for evaluating the face detection model in both aspects: speed and accuracy. The major idea is that given a reference or baseline model, a better face detection model should achieve Pareto improvement, *i.e.*, it should be better than the baseline model no matter how the baseline model trade-off one aspect for the other. To evaluate this, we propose a composite score for the accuracy and speed of face detection models.

2 Method

Score Formulation. The composite score is computed as

$$\mathcal{C}(g|\theta) = \Phi(g)\Psi(g), \tag{1}$$

where $g \in \mathcal{G}$ represents a function that performs face detection on one input image. The function $\Phi : \mathcal{G} \rightarrow \mathrm{R}$ scores the face detection model in terms of detection accuracy, while the function $\Psi : \mathcal{G} \rightarrow \mathrm{R}$ scores the model in terms of runtime speed. A general choice of Φ is the averaged average precision (AAP) function [], which maps the face detection model to a score in $[0, 1]$ which higher score the better. It is used in the WIDER Challenge 2018 and 2019 to assess the face deteciton models' accuracy. However, for the function Ψ, there is no widely accepted form. In WIDER 2019 challenge, we adopt the form of

$$\Psi(g) = \sigma(\mathrm{FPS}(g, \mathcal{D})), \tag{2}$$

where $\sigma(x) = 1/[1 + \exp -\alpha(x - \beta)]$. The function FPS is measured on a dataset $\mathcal{D}$ with N images as

$$\mathrm{FPS}(g, \mathcal{D}) = \frac{N}{\sum_{i=1}^{N} t(g, d_i)}, \tag{3}$$

where $t(g, d_i)$ is the runtime cost of model g on the i-th image of the dataset, d_i.

Parameter Identification. Given the form of the composite score, there are two parameters α and β that need to be set before we use the score to assess the submissions. In terms of Pareto improvement, our composite scores should be similar for a same model at its different operating points trading-off speed and accuracy. So we use a baseline face detection model g_0 and measure its speed and accuracy at different operating points $\{g_0^0, \ldots, g_0^K\}$. Given the data points, we minimize the difference between $\mathcal{C}(g_0^i)$, to get the parameters. This optimization can be approximated by a linear regression problem to the following form

$$\log(\frac{\Phi(g) - C}{C}) = AFPS(g) + B, \tag{4}$$

where C can be an arbitrary real number (we chose $C = 0.2$), $A =$, and $B = -\alpha * \beta$. According to this method, we identified the parameters in the scoring function as

$$\alpha = 0.05, \beta = 20. \tag{5}$$

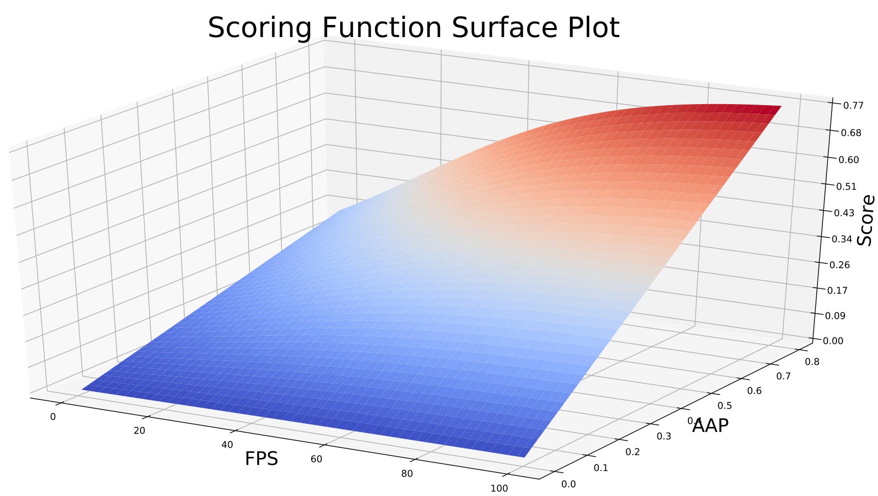

Figure 1: Value surface of the scoring function.

3 Results Analysis

The scoring curve of the baseline model can be found in Fig. 2.

To show more results, we provide visualization of the scoring function's value surface in 1.

References

[1] M. Tan, B. Chen, R. Pang, V. Vasudevan, and Q. V. Le, "Mnasnet: Platform-aware neural architecture search for mobile," *arXiv preprint arXiv:1807.11626*, 2018.

Figure 2: **Blue**: AAP vs. speed for the reference model. **Orange**: Composite score vs. speed for the reference model.