Supplementary material for
MetaHTR: Towards *Writer-Adaptive* Handwritten Text Recognition

Ayan Kumar Bhunia[1] Shuvozit Ghose* Amandeep Kumar* Pinaki Nath Chowdhury[1,2]
Aneeshan Sain [1,2] Yi-Zhe Song[1,2]
[1]SketchX, CVSSP, University of Surrey, United Kingdom.
[2]iFlyTek-Surrey Joint Research Centre on Artificial Intelligence.
{a.bhunia, p.chowdhury, a.sain, y.song}@surrey.ac.uk.
{shuvozit.ghose, kumar.amandeep015}@gmail.com.

A. More clarity on learnable instance specific weights and novelty behind it

First, it is important to note that the instance-specific weighting mechanism is an *integral* part of the meta-learning framework. It *can not* be applied to learning of baseline models without meta-learning, as no fixed weight labels exist to supervise the learning process. As such, we let the model meta-learn instance-specific weights for the character-wise cross-entropy loss during adaptation. That is, the model itself needs to implicitly discover the weights, and adaptively assign weights to different characters so to prioritise learning from more discrepant characters.

B. Clarification on marginal improvement via fine-tuning

This is not a surprise – fine-tuning [2] requires significantly more data and hundreds of gradient updates, yet gradient-based meta-learning (e.g,. MAML) does it with instantaneously with one single gradient update and with minimal data (≤ 16 for ours). The fact that our meta-learning approach can achieve superior performance despite the disadvantages is in fact a good reflection on the effectiveness of our method.

C. More details on experimental setup and analysis:

(i) W_S/W_T denotes writer-set, while D_S/D_T denotes the set of writer specific word-images paired with annotations.

(ii) As calculating gradients with respect to all the model's parameters is quite cumbersome, we calculate gradient of t-th character specific cross-entropy loss with respect to final classification layer (parameter ϕ) as $\nabla_{\phi}\mathcal{L}_{ce}^{t}(\theta)$. This statement refers to the inner loop update that involves calculating character specific weights. In our formulation, inside the inner loop (Eq. 5-7), gradient needs to be calculated T times where T represents the character sequence length of word image. Hence, considering the gradient of the whole network would have been rather cumbersome. Instead, the gradient only w.r.t. final classification layer seems to give reasonable results at a much better latency.

(iii) As we consider k=16 in the adaptation set, we want to ensure that evaluation is done on writers having at least 16 word images to reduce the randomness.

(iv) As adaptation set is not defined in the dataset, we randomly sample 16 images for adaptation set and rest are used for evaluation. We ensured this split to remain constant w.r.t. every writer, for every competitor for fair evaluation. Because of this evaluation protocol, our numbers for ASTER are a bit different from those reported in [1].

References

[1] Ayan Kumar Bhunia, Abhirup Das, Ankan Kumar Bhunia, Perla Sai Raj Kishore, and Partha Pratim Roy. Handwriting recognition in low-resource scripts using adversarial learning. In *CVPR*, 2019. 1

*Interned with SketchX

[2] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. In *ICLR*, 2020. 1