Low-Rank Adaptation vs. Fine-Tuning for Handwritten Text Recognition Supplementary Material

Lukas Hüttner, Martin Mayr, Thomas Gorges, Fei Wu,
Mathias Seuret, Andreas Maier and Vincent Christlein
Pattern Recognition Lab, Friedrich-Alexander Universität Erlangen-Nürnberg
91058 Erlangen, Germany
{lukas.huettner, martin.mayr, thomas.gorges, river.wu,
mathias.seuret, andreas.maier, vincent.christlein}@fau.de

We aim to provide detailed information here to ensure the reproducibility of our results.

1. Configuration-specific hyperparameters

The optimal learning rate and the percentage of time allocated for the one-cycle warm-up phase, specific to each method configuration, is summarized in Tab. 1.

2. Augmentation

Our augmentation pipeline includes vertical and horizontal padding, image squeezing and stretching, morphological operations like dilation and erosion, subtle warping, as well as contrast and brightness adjustments. Each technique was applied independently with a probability of 20 %.

The results with applied augmentation for the PEFT experiments are shown in Tab. 2 and for data-scaled fine-tuning in Tab. 3.

While all experiments demonstrate notable improvement over the case with no augmentation, changes in the trend include both PEFT methods narrowing the gap to fine-tuning further, especially at higher ranks. In particular, DoRA at rank 256 nearly matches the performance of fine-tuning on the IAM Handwriting database.

In line with the results observed with no augmentation, DoRA consistently outperforms LoRA at rank 1, most likely attributed to the extra trainable parameter.

For data-scaled fine-tuning, significant improvements can be observed for both datasets.

Interestingly, using only 10 samples reaches a CER of 7.67% on the IAM Handwriting database, while on the READ 2016 dataset 2000 samples are needed to roughly match the same performance, underlining the different domain gaps.

Despite using augmentation, the fine-tuning performance using the whole IAM Handwriting database (CER 3.96%) remains lower than the original TrOCR result (CER 2.89%). This discrepancy is presumably due to the missing task-specific pre-training step (stage 2) from the original TrOCR approach, for which both the model weights and the dataset were not made public.

Table 1. Best performing hyperparameters for each configuration.

Dataset	Method	Rank	Learning Rate	Warm-up Phase
IAM	LoRA/DoRA	1	2.2×10^{-6}	15%
		4	2.2×10^{-6}	15%
		16	6.2×10^{-7}	15%
		64	6.5×10^{-7}	20%
		256	5.7×10^{-7}	20%
	FT	-	1.6×10^{-6}	30%
READ	LoRA/DoRA	1	8.0×10^{-6}	15%
		4	6.5×10^{-6}	15%
		16	6.5×10^{-6}	15%
		64	3.5×10^{-6}	20%
		256	2.0×10^{-6}	20%
	FT	-	5.5×10^{-6}	10%

Table 2. Character-Error Rate (%) of LoRA and DoRA with applied augmentation.

Dataset	Method	Rank				
		1	4	16	64	256
READ 2016	LoRA	13.02	6.6	5.11	4.73	4.45
	DoRA	8.52	6.21	5.07	4.66	4.48
IAM	LoRA	5.8	4.67	4.3	4.12	4.03
	DoRA	5.04	4.63	4.25	4.1	3.98

Table 3. Character-Error Rate (%) of data-scaled fine-tuning with applied augmentation.

Dataset	Samples							Proportion of dataset			
	10	50	100	500	1000	2000	4000	70%	80%	90%	100%
READ 2016	64.68	42	29.7	13.27	9.82	7.66	5.99	4.95	4.81	4.51	4.38
IAM	7.67	7.28	6.38	5.44	5.1	4.8	4.42	4.29	4.21	4.14	3.96