AttentionHTR: Handwritten Text Recognition based on Attention Encoder-Decoder Networks

Introduction:

Handwritten text possesses high variability due to different writing styles, languages, and scripts. Training an accurate and robust system calls for data-efficient approaches due to the unavailability of enough annotated multi-writer text.

Methods:

1. **Four-stage framework**:
   - Transformation: TPS
   - Feature extraction: ResNet
   - Sequence modeling: BiLSTM
   - Prediction: LSTM with attention
2. **Transfer learning** from Scene Text Recognition (STR) to Handwritten Text Recognition (HTR).
3. **Novel multi-writer dataset Imgur5K** previously not used for HTR.
4. **Error analysis.**

Dmitrijs Kass: dmitrijs.kass@it.uu.se
Department of IT, Uppsala University, Sweden

Ekta Vats: ekta.vats@abm.uu.se
Centre for Digital Humanities Uppsala,
Department of ALM, Uppsala University, Sweden

Overcome training data scarcity on Handwritten Text Recognition tasks and **train a general-purpose model**

Transfer learning

14.4M of **synthetic** training data.

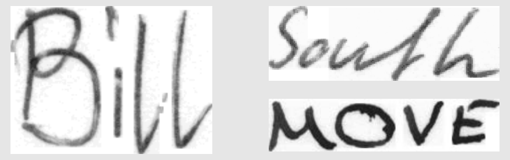
230K of **real** handwritten training data for fine-tuning.

MJSynth

Imgur5K

SynthText

IAM



Results

1. Error rates are comparable with the state-of-the-art.
2. The final model is trained on handwriting from thousands of authors, with varying image conditions, in order to aid generalisation in the real-world.
3. Model accuracy can be further strengthened by adding more datasets to the pipeline.

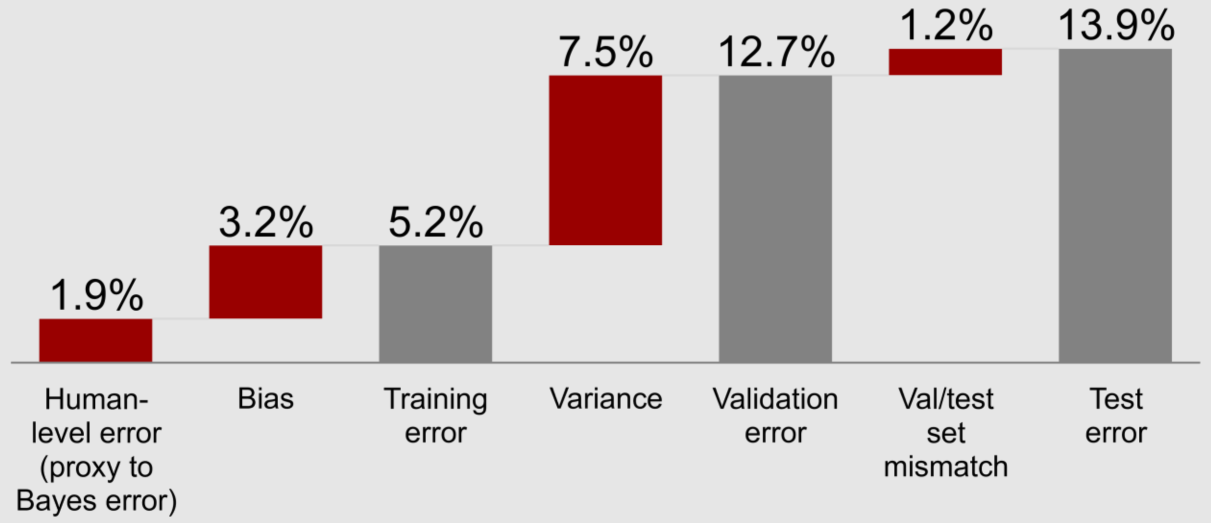
Character set	Test-CER			Test-WER		
	Imgur5K	IAM	Both	Imgur5K	IAM	Both
Ours, case-insensitive	6.46[a]	4.30[b]	5.96[c]	13.58[a]	12.82[b]	13.89[c]
Ours, case-sensitive	9.47[a]	6.50[b]	8.59[c]	20.45[a]	**15.40**[b]	18.97[c]
Kang *et al.* [12]	-	6.88	-	-	17.45	-
Bluche *et al.* [3]	-	12.60	-	-	-	-
Sueiras *et al.* [20]	-	8.80	-	-	23.80	-
Chowdhury *et al.* [5]	-	8.1	-	-	-	-
Johannes *et al.* [15]	-	**5.24**	-	-	-	-
Kang *et al.* [11][d]	-	5.79	-	-	15.91	-

Fine-tuning approaches: [a] IAM→Imgur5K; [b] Imgur5K→IAM; [c] Imgur5K+IAM.

Design a **prioritized test error reduction strategy** through error analysis

Error analysis

1. Test error decomposition ⇒ focus on variance



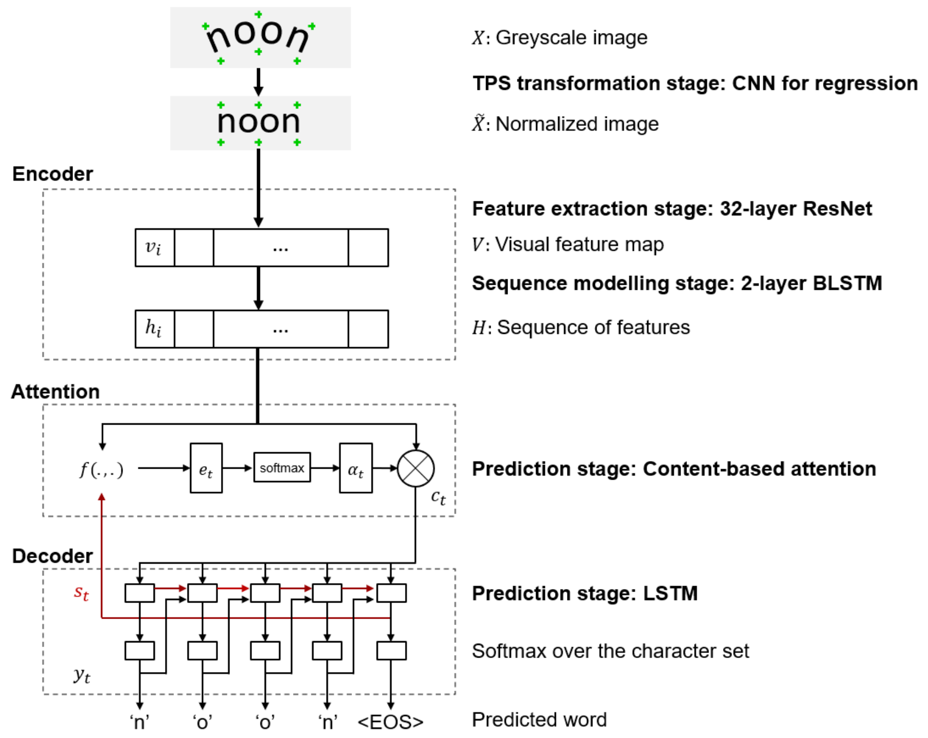
2. Character-level error analysis ⇒ tailor-made representation of characters.
3. Visual analysis of errors by a human ⇒ tailor-made data augmentation

Prioritized strategy:

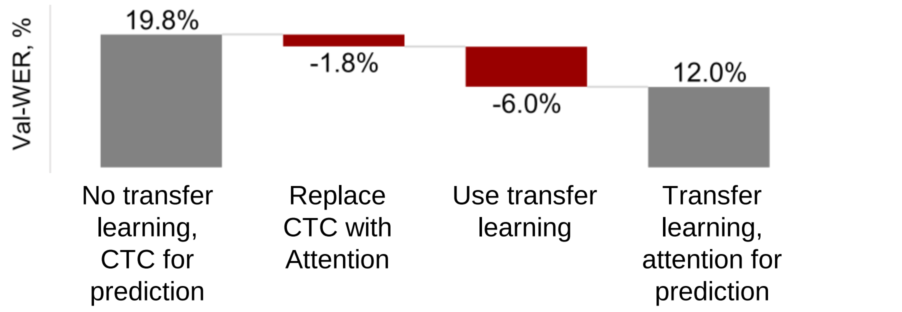
1. Variance reduction:
   - Replace early stopping with norm penalty or dropout regularization.
   - Data augmentation tailored to the level of visual effects and characters.
   - Other annotated multi-writer datasets.
   - Language modeling.
2. Bias reduction:
   - Run training longer.
3. Validation/testing set mismatch.

Code and pretrained models:

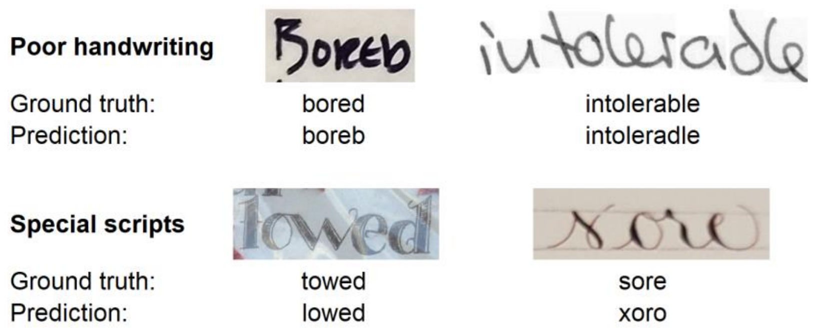
Model architecture



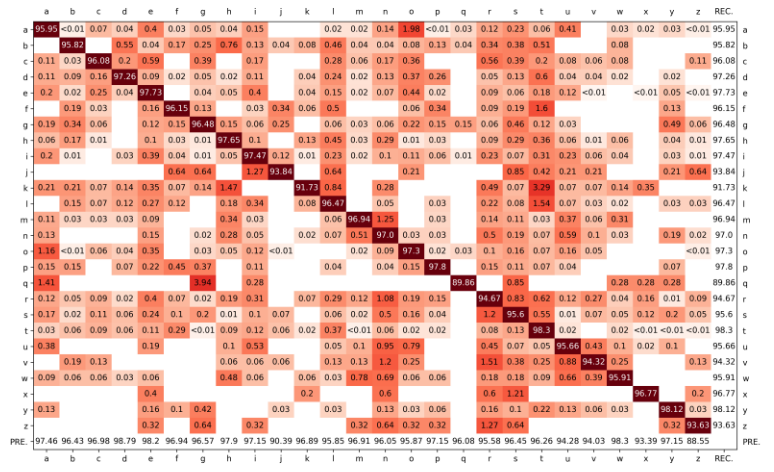
Ablation study on the IAM dataset



Visual analysis of errors



Character-level confusion matrix



Character-level F1 scores