NEURAL INFORMATION PROCESSING SYSTEMS

Paris, France · 9 – 13 December 2026

Handwritten Text Recognition Lives in the High-Pixel Variance Subspace

Carlos Garrido-Munoz · Jorge Calvo-Zaragoza

University of Alicante, Spain {carlos.garrido,jorge.calvo}@ua.es

Read the paper

Universitat d'Alacant
Universidad de Alicante

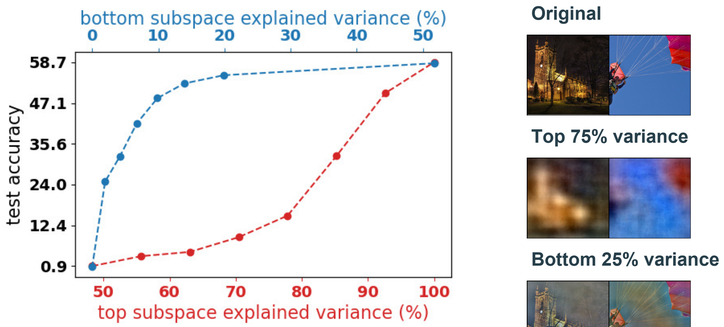
prAIg

The useful pixel subspace depends on the recognition task

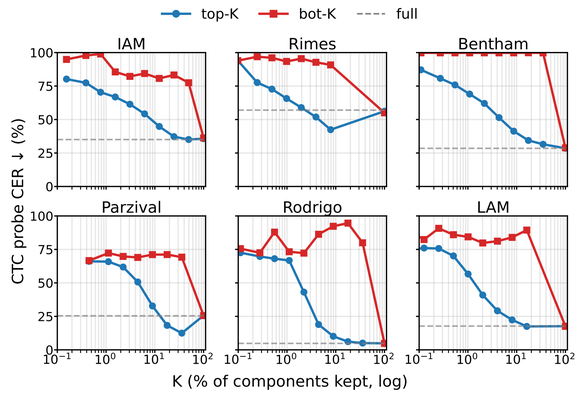
Handwriting recognition relies on high-variance pixel directions, reversing the pattern observed in natural images

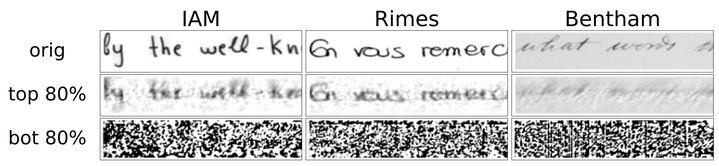
The recognition signal reverses

Natural-image classification (Balestriero & LeCun, 2024)



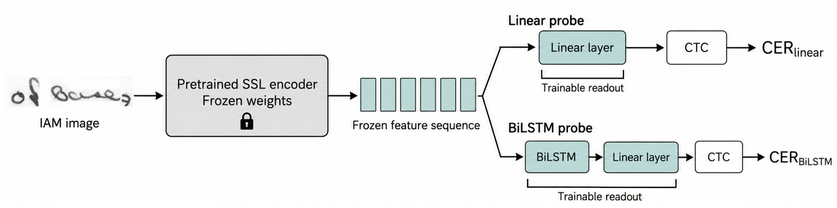
Handwritten text recognition





Why test handwriting separately?

- On natural images, pixel-grounded SSL has been argued to waste capacity on irrelevant detail (Balestriero & LeCun, 2024)
- For handwriting, that visual detail may carry the recognition signal
- We test whether preserving high pixel variance aids transcription

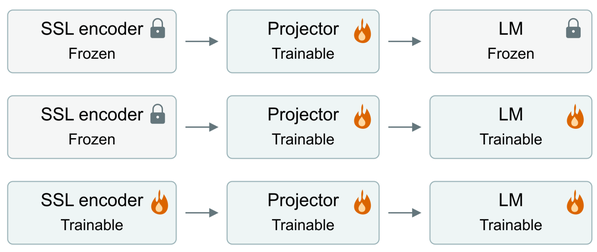


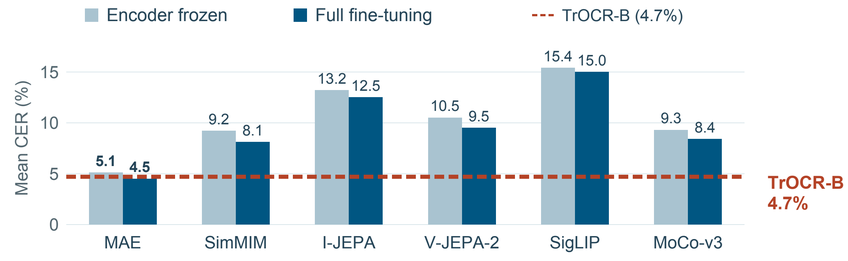
Recognition with a VLM

Stage 1
Alignment
Synthetic pairs

Stage 2
Encoder frozen
Real pairs

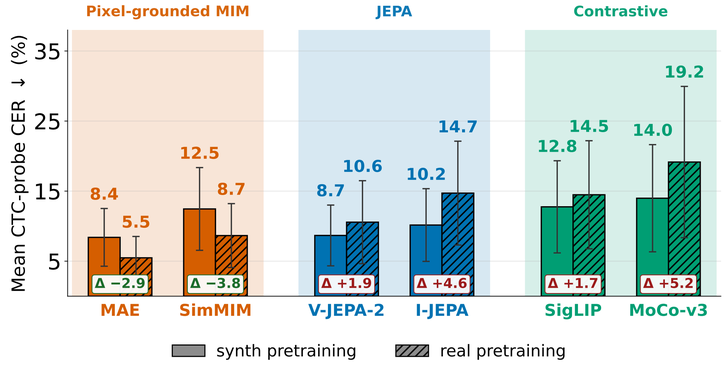
Stage 3
Full fine-tuning
Real pairs



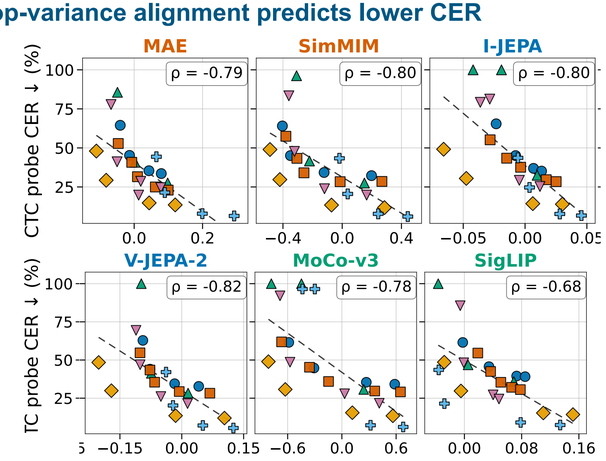


Recognition results

Frozen ViT-B + BiLSTM–CTC



Top-variance alignment predicts lower CER



R² gap: larger values indicate top-variance alignment

Conclusions

- HTR signal lies in high-variance pixel directions, reversing the natural-image pattern
- Stronger top-variance alignment predicts lower CER within every tested SSL method
- Full fine-tuning of the MAE-based VLM achieves results competitive with state-of-the-art HTR systems

Acknowledgments and Disclosure of Funding

This research was supported by the Spanish Ministry of Science and Innovation through the LEMUR research project (PID2023-148259NB-I00), funded by MCIU/AEI/10.13039/501100011033/FEDER, EU, and the European Social Fund Plus (FSE+). The first author was supported by grant CIACIF/2021/465 from the Programa I+D+i de la Generalitat Valenciana.

Reference

Balestriero, R., & LeCun, Y. (2024). How Learning by Reconstruction Produces Uninformative Features For Perception. ICML, PMLR 235, 2566–2585.