HANDWRITTEN WRITER IDENTIFICATION

MARTIN KUBIČKA | SAMUEL ŠIMÚN | DANIEL ŽÁRSKÝ

INTRODUCTION

*The goal is to identify a writer from a digital image of their handwriting. Starting with a scanned or photographed text sample, our system extracts features like stroke width, spacing, slant and pressure patterns to form a unique signature. We employ **CNNs** to learn local textures and **Transformers** to capture broader **stylistic cues**. The result is either the author's name (if known) or a new identifier tied to that handwriting profile.*

APPROACHES

*To address the identification task, our objective was to train a model capable of producing **informative feature representations**. We utilized a classification neural network framework, training it with a standard classification loss like **AdaFace**, **CosFace** or **ArcFace**. Crucially, the network was designed such that the terminal classification layer could be detached, thereby allowing the preceding layers to function as a feature extractor yielding discriminative vectors. In the case of the **hybrid approach**, the **ViT encoder** produces the **embeddings**, so there is no need to detach any layers.*

CONCLUSION

Our system determines the writer of a handwritten image by extracting feature embeddings. We observed that larger embedding dimensions boost identification accuracy but also increase training time and memory usage.
*Trained from scratch on **406,000** samples, our hybrid model reached **75.73 % Rank-1 accuracy** on a **529-identity** gallery. Future work will investigate larger text regions or full-color (RGB) inputs and test alternative backbones beyond ResNet-18.*

VIT

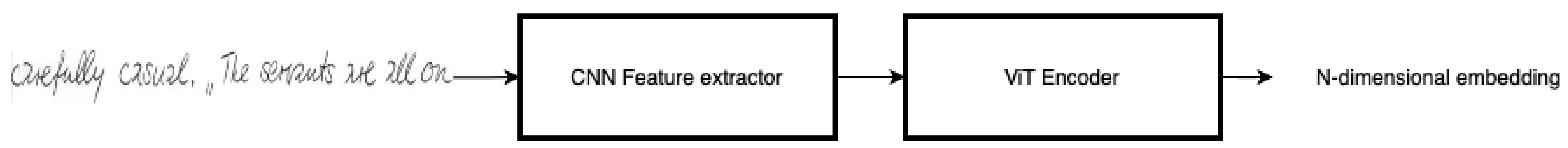
We used a custom Vision Transformer (ViT)–based model (base, patch size 8, resolution 224) adapted to our dataset. Training adds two fully connected layers: the penultimate yields embeddings for inference, and the final layer handles classification during training only.

HYBRID

This approach consists of two main steps. In general, the input is an image containing handwritten text, which is then passed into a CNN feature extractor. The output of the CNN is fed into a ViT encoder, which generates an N-dimensional embedding representing the author.

CNN

We developed a CNN-based approach centered on MobileNetV3 architecture, selected for its efficiency and strong performance on lightweight vision tasks. . Incorporating MobileNet into our framework enables a broader evaluation of existing methods. and comparison for previously suggested solutions.



RESULTS FOR HYBRID

We plotted ROC curves (TPR vs FPR) on pair-similarity scores and used the AUC-ROC to quantify discrimination. We visualized genuine versus impostor score distributions to gauge embedding separability. Finally, closed-set ID was measured via top-k (Rank-1) accuracy as the gallery size N varied.