InkSight: Offline-to-Online Handwriting Conversion by Learning to Read and Write

Blagoj Mitrevski*, Arina Rak*, Julian Schnitzler*, Chengkun Li*, Andrii Maksai, Jesse Berent, Claudiu Musat
(*first authors, random order generated by AEA tool)

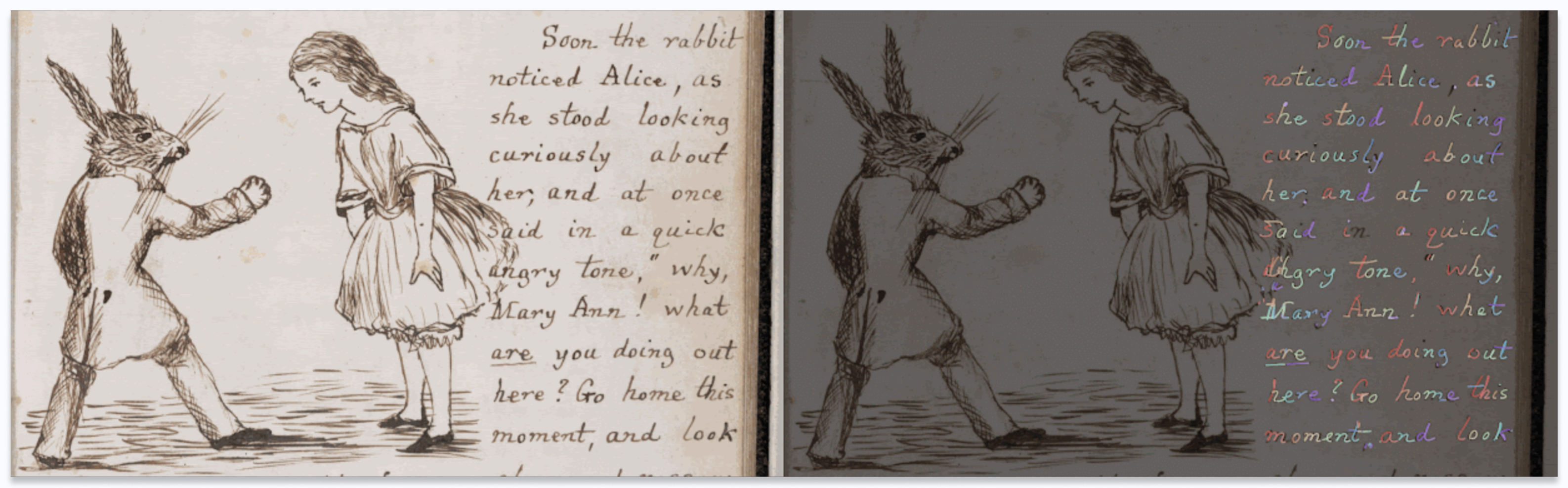
EPFL Google DeepMind

Google Blog Post Project Website Paper

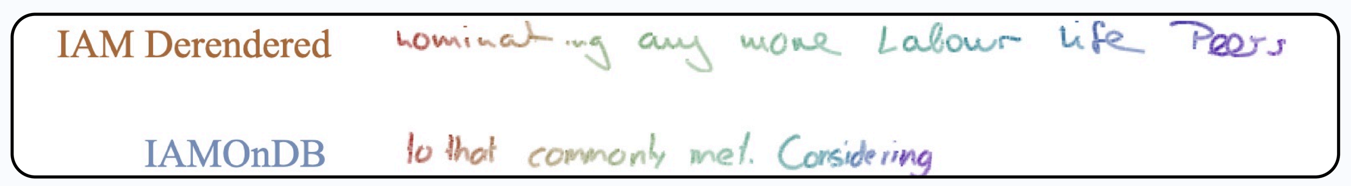
Introduction

✍️ Handwritten notes offer personal expression but lack digital conveniences

InkSight converts **offline** handwriting to **online** handwriting (we call it derendering), bridging the gap between the two



An example of offline to online conversion, photo from *Alice's Adventures in Wonderland*



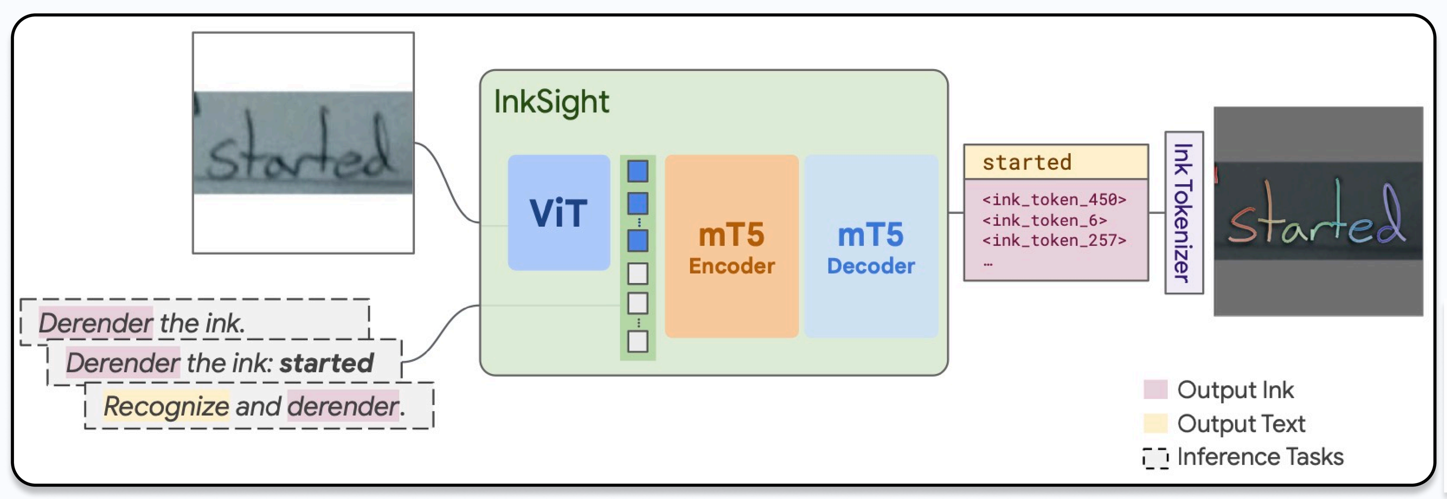
Online handwriting data produced by InkSight (Top)

Scan QR code to visualize in animation

Overview (TL; DR)

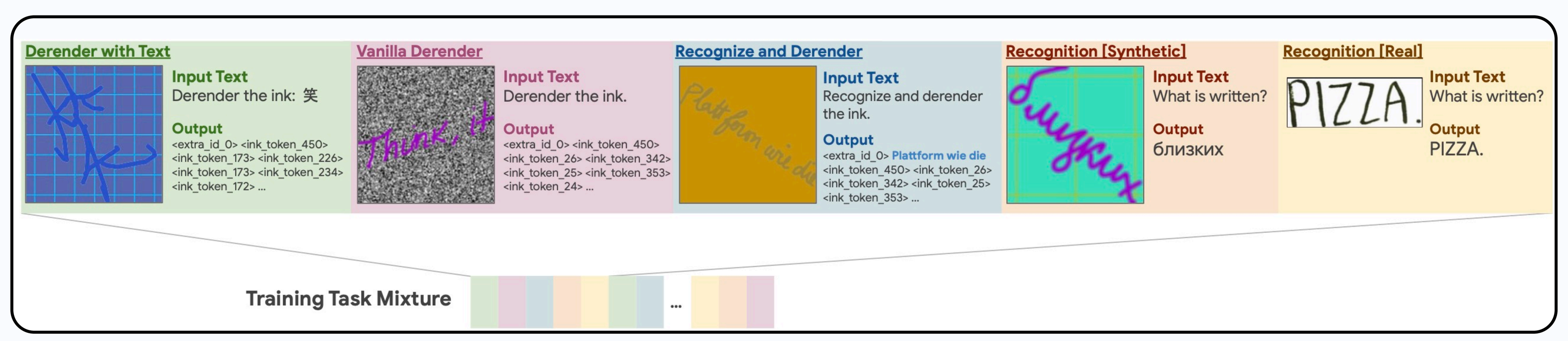
InkSight uses ***reading*** and ***writing*** priors to interpret and recreate handwritings

The model combines a **Vision Transformer** (ViT) encoder with an **mT5 encoder-decoder Transformer**



InkSight model inference for single word (scan QR code to visualize)

It is trained using a **multi-task setup** with real and synthetic data



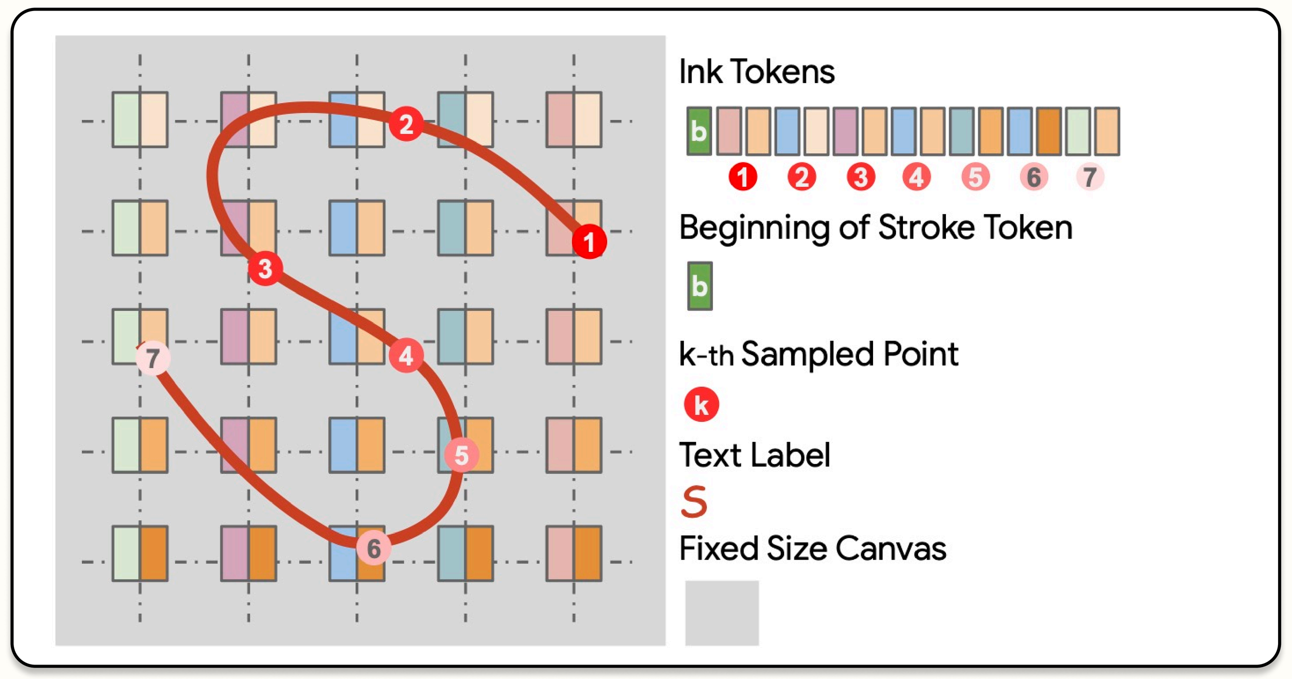
Datasets mixture with 5 types of tasks

Digital Ink Tokenization

Digital ink is represented as a sequence of **strokes**, each stroke consists of **coordinate-time triplets**

$$I = \{s_1, s_2, \cdots, s_n\} \qquad s_i = \{(x_i, y_i, t_i)\}_{i=1}^{m_i}$$

A novel ink tokenizer converts ink strokes into discrete tokens optimized for VLMs

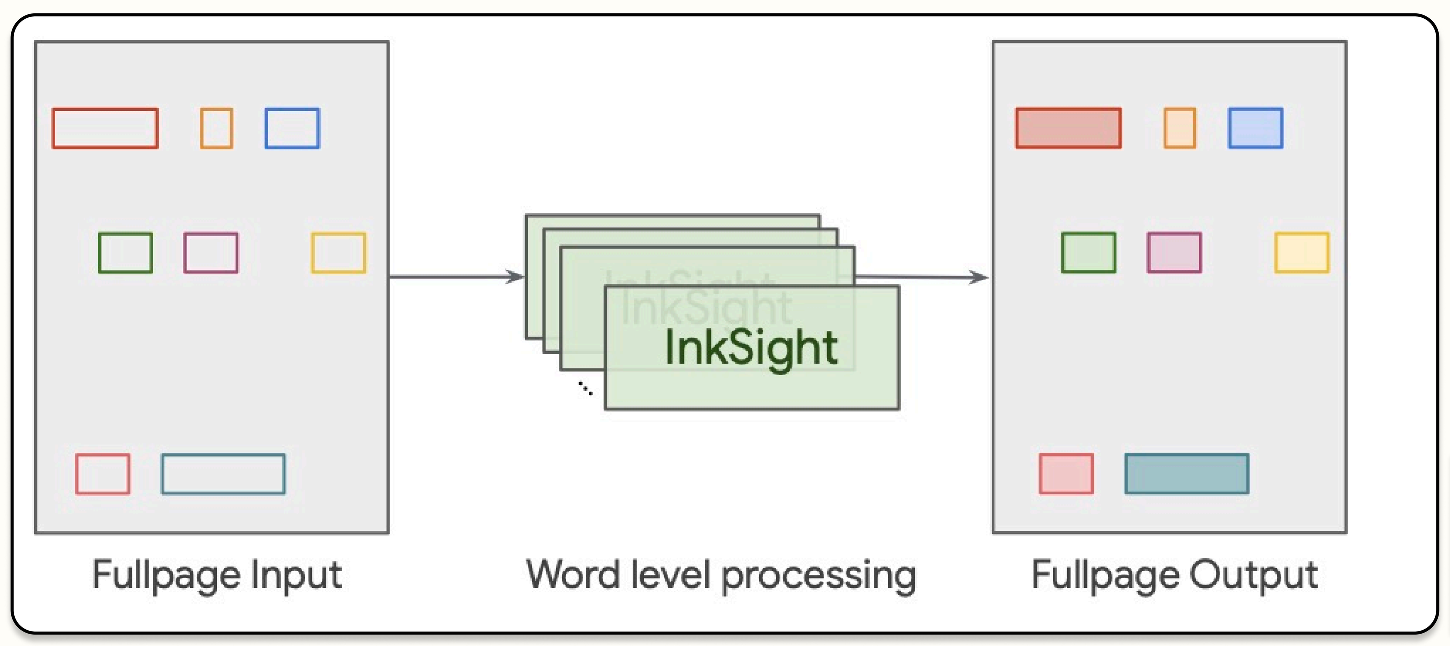


Digital ink tokenizer (scan QR code to visualize)

Each digital ink stroke is normalized by resampling it at a fixed rate, applied with the Ramer-Douglas-Peucker algorithm, and centering it on a fixed-size canvas

Full-Page Derendering

InkSight handles entire pages of handwritten notes by identifying and derendering each word individually and process in batches

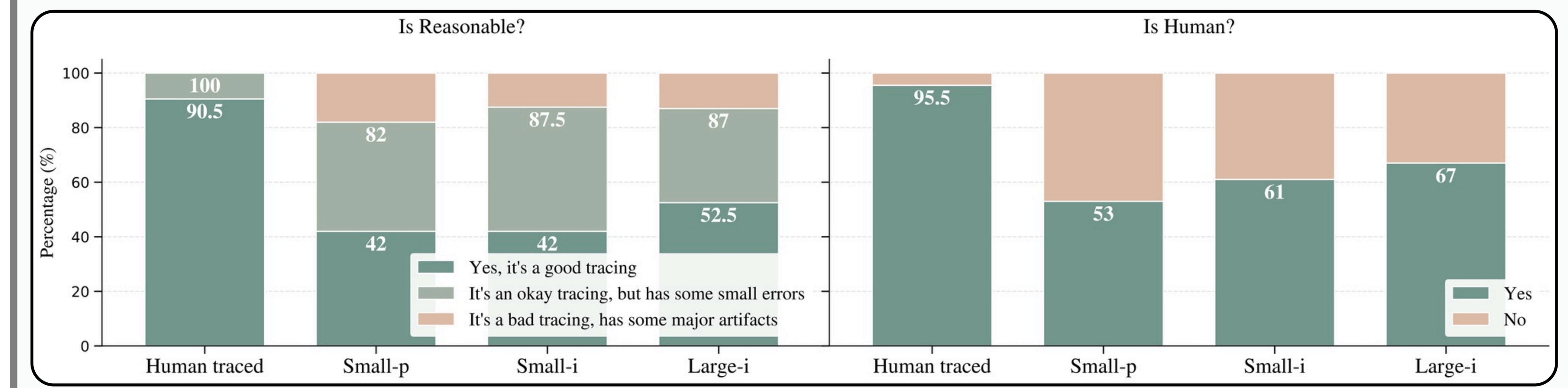


Full-Page pipeline with InkSight model (scan QR code to visualize)

Three inference modes can be selected flexibly, depending on the requirement for understanding the semantics

Highlighted Findings

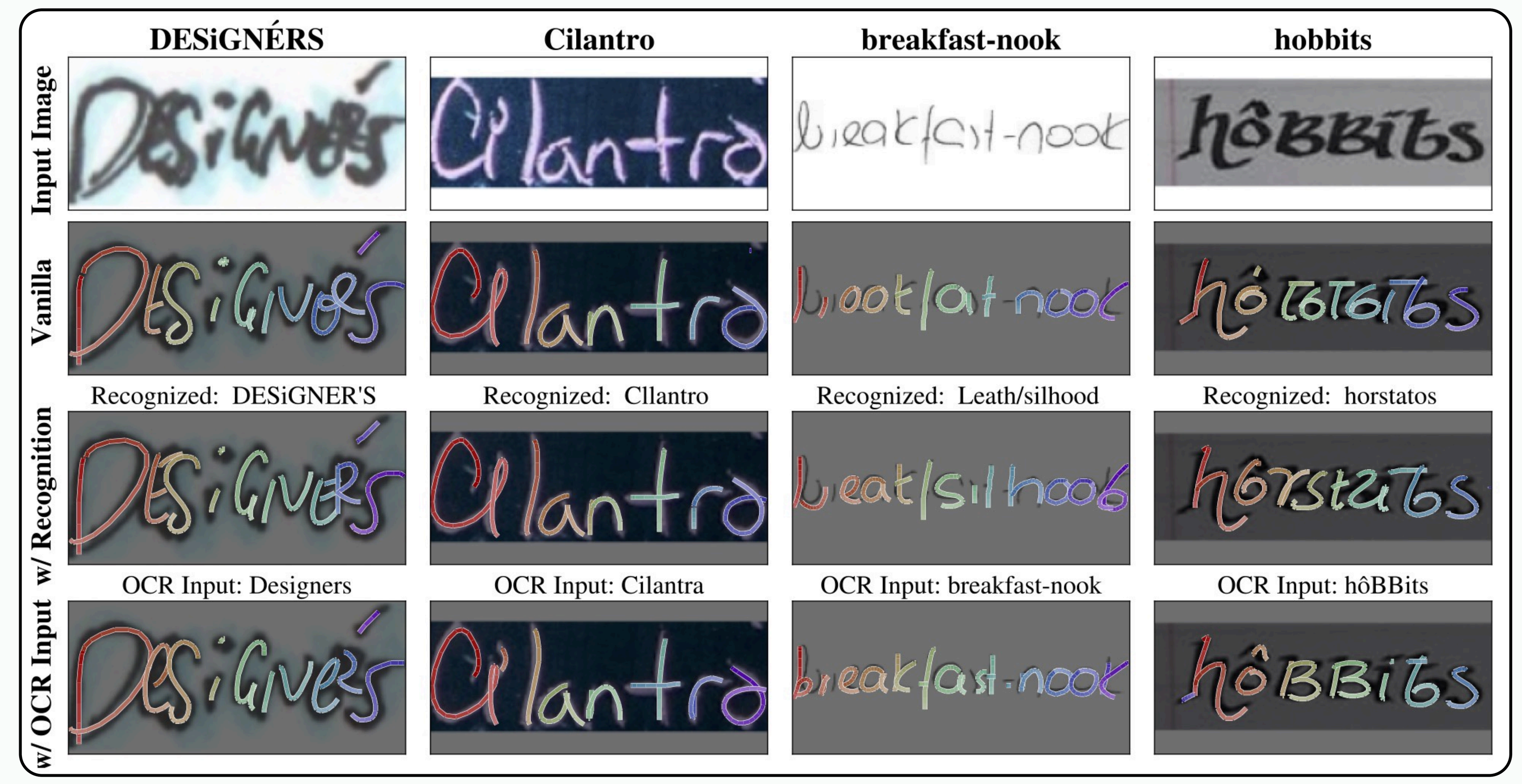
Human Evaluation: 87% of InkSight's outputs were judged as valid tracings, and 67% were deemed indistinguishable from human-generated digital ink



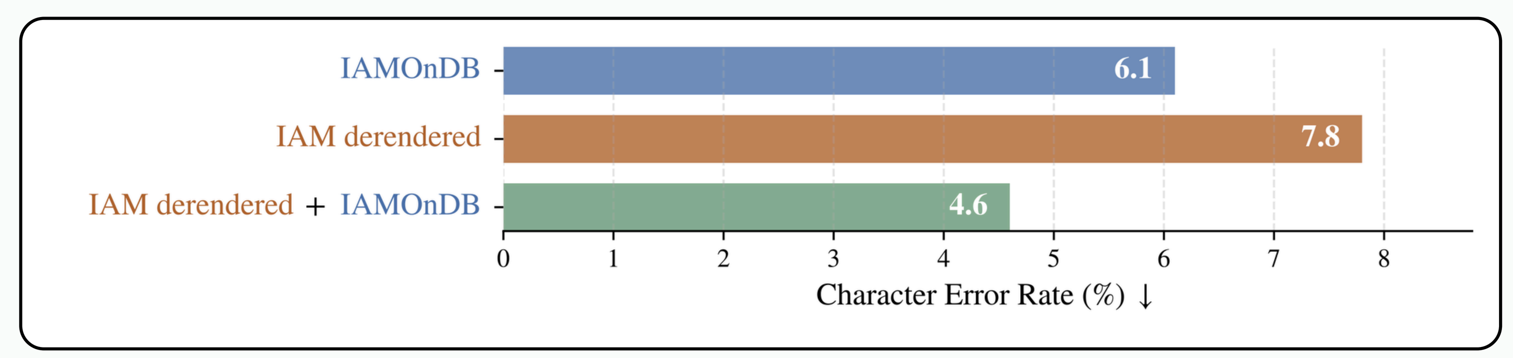
Role of Recognition: Recognition training is key in producing semantically consistent writing

Setup	IAM		IMGUR5K		HierText	
	F1	Acc.	F1	Acc.	F1	Acc.
Small-i†	0.66±0.07	0.59±0.01	0.51±0.09	0.33±0.02	0.61±0.07	0.45±0.01
Vanilla	0.61↓	0.53↓	0.44↓	0.28↓	0.53↓	0.35↓
R+D	0.62↓	0.57↓	0.46↓	0.32	0.57↓	0.42↓
Remove						
data aug*	0.42↓	0.33↓	0.21↓	0.09↓	0.23↓	0.13↓
syn rec	0.58↓	0.50↓	0.50↓	0.25↓	0.56↓	0.38↓
real rec	0.64↓	0.50↓	0.55↑	0.19↓	0.59↓	0.36↓
all rec	0.65	0.51↓	0.55↑	0.22↓	0.61	0.38↓
frozen ViT†	0.65±0.13	0.53±0.06↓	0.43±0.24↓	0.31±0.06	0.59±0.15	0.41±0.05↓

Handling Ambiguity: Different inference strategies yield distinct interpretations of ambiguous handwriting



Data Source: Derendered ink can serve as valuable complementary data for training recognition systems



Performance of recognizer trained with different data source