Fine-Tuning TrOCR for Handwriting Recognition

1. Dataset & Model

Model: TrOCR (Transformer-based OCR) from Hugging Face, fine-tuned in TensorFlow.

Datasets: IAM Handwriting (line-level), Imgur5K (real-world handwritten words).

2. Preprocessing

Images resized to 384x384, normalized, converted to grayscale.

Text labels tokenized using TrOCR tokenizer (max length: 128 tokens).

3. Training Setup

- Optimizer: Adam, Learning Rate: 5e-5
- Epochs: 10, Batch Size: 4 (due to memory limits)
- Mixed Precision: Enabled for efficiency
- Validation Split: 10%
- Dataset loaded using tf.data.Dataset

4. Evaluation Metrics

- Character Error Rate (CER): ~5.8%
- Word Error Rate (WER): ~12.5%

Evaluated using `jiwer` library over test set.

5. Challenges

- Memory limits required small batch sizes.
- Dataset variance (style, quality) required strong preprocessing.
- No official TF TrOCR model - converted from PyTorch using `from_pt=True`.

6. Future Improvements

- Add layout-aware OCR models (e.g., DocTR).
- Integrate post-processing with language models.
- Increase dataset diversity with augmentation.
- Fine-tune with larger batch sizes and LR schedulers.