PP-OCRv5: A Specialized 5M-Parameter Model Rivaling Billion-Parameter Vision-Language Models on OCR Tasks

Supplementary Material

I. Performance Benchmarks and Resource Utilization

This section provides comprehensive performance metrics for the PP-OCRv5 model, including detailed analysis of inference speed, computational resource consumption, and memory utilization under both GPU and CPU configurations.

The PP-OCRv5 recognition module incorporates an expanded character dictionary to enhance recognition accuracy across diverse text scenarios. This design choice results in increased computational requirements, including higher memory (RAM) and video memory (VRAM) consumption, alongside marginally extended per-image inference times.

I.1. Experimental Setup

Performance benchmarks were conducted under controlled conditions with the following specifications:

- **GPU**: NVIDIA Tesla V100 (32GB VRAM)
- **CPU**: Intel Xeon Gold 6271C (2.60GHz)
- **Framework**: PaddlePaddle 3.0.0
- **Test Dataset**: 200 diverse images

Note: Pre-loading images into memory can reduce average processing time by approximately 25ms per image.

I.2. Performance Analysis Results

I.2.1. GPU Performance Metrics

Table S1 presents comprehensive GPU performance characteristics, demonstrating PP-OCRv5's efficiency in GPU-accelerated environments.

I.2.2. CPU Performance Metrics

Table S2 demonstrates PP-OCRv5's CPU-only performance, highlighting its suitability for edge deployment scenarios without GPU acceleration.

II. Comprehensive Visual Analysis

This section provides extensive visual demonstrations of PP-OCRv5's capabilities in end-to-end OCR processing. These examples showcase the system's robustness across diverse linguistic, stylistic, and layout challenges.

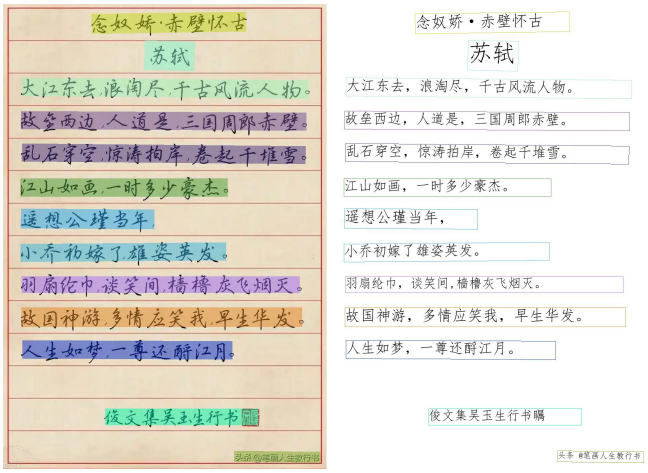
The complete PP-OCRv5 pipeline integrates detection and recognition modules to deliver comprehensive OCR solutions. Figures S1 and S2 present end-to-end results across diverse real-world scenarios, demonstrating the system's practical applicability.

Table S1. Comprehensive GPU Performance Analysis for PP-OCRv5 on NVIDIA Tesla V100

Metric	Value
Model: PP-OCRv5	
Performance:	
Avg. Time per Image	**0.62 s**
Throughput	**1054.23 chars/sec**
CPU Resources:	
CPU Utilization	106.35 %
Peak RAM	1829.36 MB
Avg. RAM	1521.92 MB
GPU Resources:	
GPU Utilization	17.42 %
Peak VRAM	4190.00 MB
Avg. VRAM	3114.02 MB

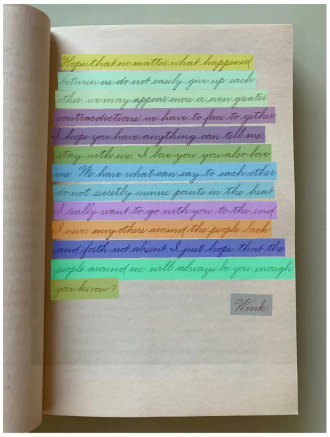
Table S2. CPU Performance Analysis for PP-OCRv5 on Intel Xeon Gold 6271C

Metric	Value
Model: PP-OCRv5	
Performance:	
Avg. Time per Image	**1.75 s**
Throughput	**371.82 chars/sec**
CPU Resources:	
CPU Utilization	965.89 %
Peak RAM	2219.98 MB
Avg. RAM	1830.97 MB

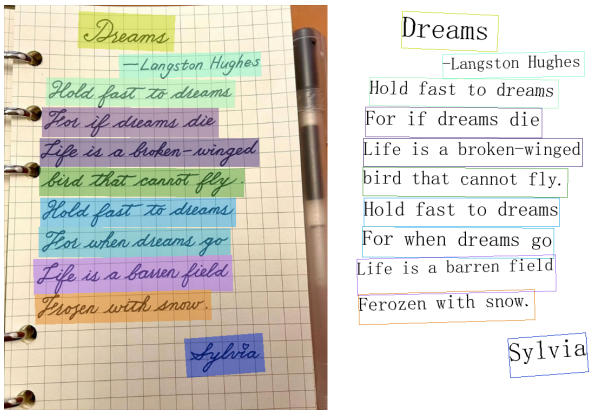


(a) Handwritten Chinese documents

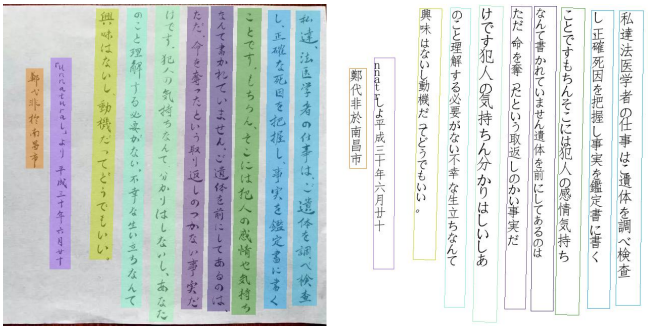
(b) Traditional Chinese typography



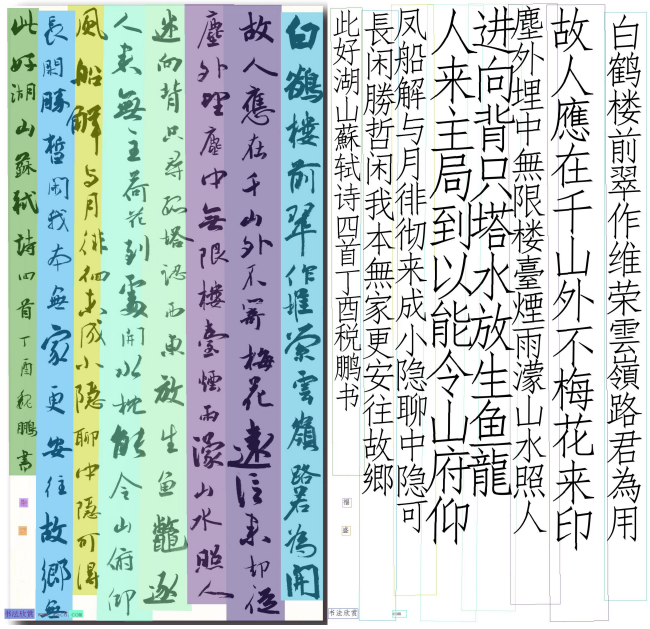
(c) Casual handwritten English



(d) Formal handwritten English

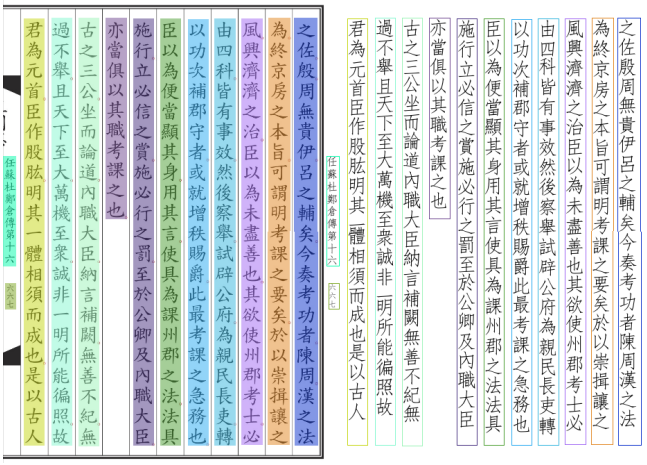


(e) Handwritten Japanese mixed scripts

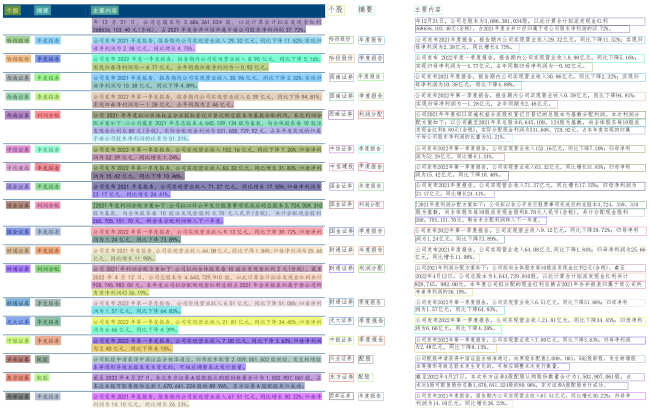


(f) Vertical ancient Chinese texts

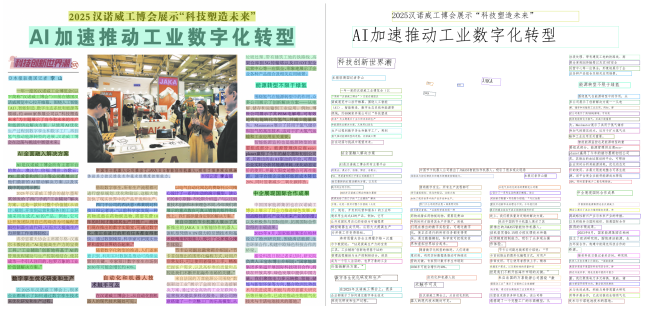
Figure S1. End-to-end OCR results (Part I): PP-OCRv5 processing diverse handwritten and traditional documents with high accuracy across multiple languages and writing systems.



(a) Complex vertical Chinese layouts

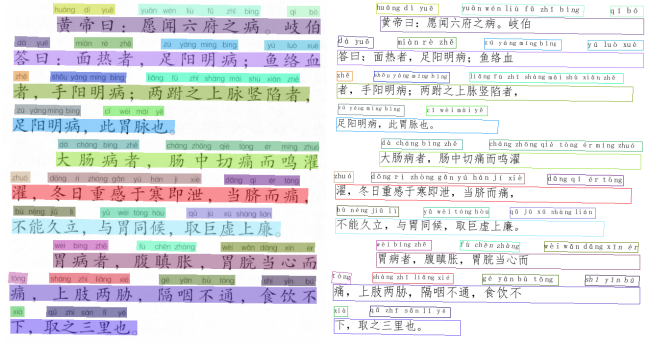


(b) Dense Chinese text documents

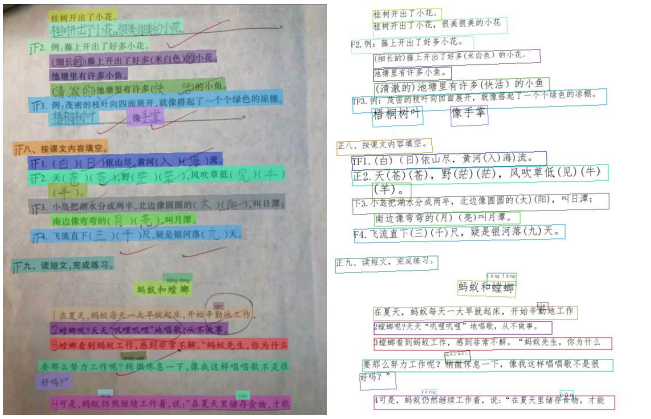


(c) Multi-column Chinese layouts

(d) Dense English technical documents



(e) Mixed Pinyin and Chinese characters



(f) Hybrid printed-handwritten documents

Figure S2. End-to-end OCR results (Part II): PP-OCRv5 handling complex document layouts, dense text regions, and mixed content types with consistent performance across challenging scenarios.