

# A Calibration Study for Local OCR

Ada Researcher, Boris Scientist

## Abstract

We present a calibration study for local optical character recognition. Our method converts academic documents into Markdown using a vision model running entirely on device. We report layout grounding accuracy and discuss coordinate-frame conventions used by the decoder.

## 1. Introduction

Document understanding has progressed rapidly. Converting PDFs to clean text is a prerequisite for downstream language-model pipelines. We focus on the offline setting, where no cloud service is involved.

We summarize related work and motivate the need for figure grounding. Existing tools transcribe text well but rarely localize figures, which is essential when figures must be replaced by generated descriptions.

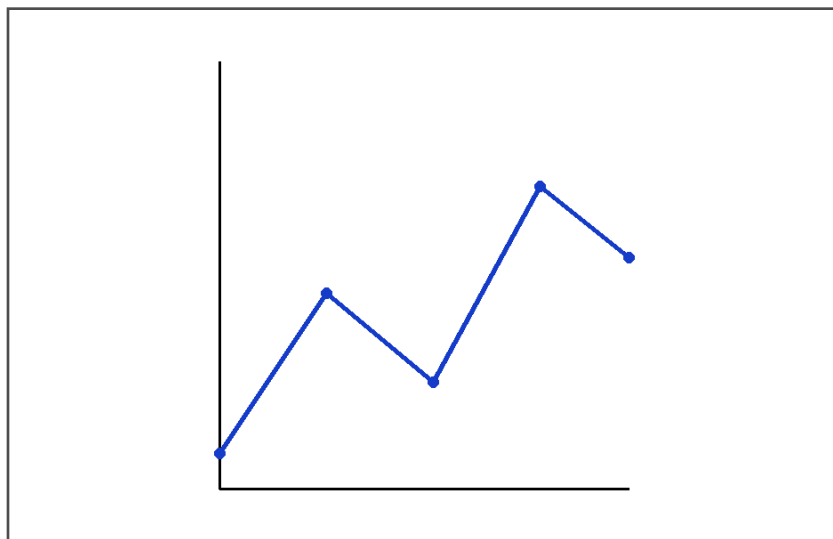


Figure 1: Calibration curve for the proposed method.

## 2. Method

Let  $x$  denote the input page and  $y$  the predicted markdown. We model  $p(y | x)$  with a decoder conditioned on encoder features. The loss is  $L = -\log p(y | x)$ .