LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis

Zejiang Shen[1] (✉), Ruochen Zhang[2], Melissa Dell[3], Benjamin Charles Germain Lee[4], Jacob Carlson[3], and Weining Li[5]

[1] Allen Institute for AI
shannons@allenai.org
[2] Brown University
ruochen_zhang@brown.edu
[3] Harvard University
{melissadell,jacob_carlson}@fas.harvard.edu
[4] University of Washington
bcgl@cs.washington.edu
[5] University of Waterloo
w422li@uwaterloo.ca

Abstract. Recent advances in document image analysis (DIA) have been primarily driven by the application of neural networks. Ideally, research outcomes could be easily deployed in production and extended for further investigation. However, various factors like loosely organized codebases and sophisticated model configurations complicate the easy reuse of important innovations by a wide audience. Though there have been on-going efforts to improve reusability and simplify deep learning (DL) model development in disciplines like natural language processing and computer vision, none of them are optimized for challenges in the domain of DIA. This represents a major gap in the existing toolkit, as DIA is central to academic research across a wide range of disciplines in the social sciences and humanities. This paper introduces `LayoutParser`, an open-source library for streamlining the usage of DL in DIA research and applications. The core `LayoutParser` library comes with a set of simple and intuitive interfaces for applying and customizing DL models for layout detection, character recognition, and many other document processing tasks. To promote extensibility, `LayoutParser` also incorporates a community platform for sharing both pre-trained models and full document digitization pipelines. We demonstrate that `LayoutParser` is helpful for both lightweight and large-scale digitization pipelines in real-word use cases. The library is publicly available at https://layout-parser.github.io.

Keywords: Document Image Analysis · Deep Learning · Layout Analysis · Character Recognition · Open Source library · Toolkit.

1 Introduction

Deep Learning(DL)-based approaches are the state-of-the-art for a wide range of document image analysis (DIA) tasks including document image classification [11,

[37], layout detection [38, 22], table detection [26], and scene text detection [4]. A generalized learning-based framework dramatically reduces the need for the manual specification of complicated rules, which is the status quo with traditional methods. DL has the potential to transform DIA pipelines and benefit a broad spectrum of large-scale document digitization projects.

However, there are several practical difficulties for taking advantages of recent advances in DL-based methods: 1) DL models are notoriously convoluted for reuse and extension. Existing models are developed using distinct frameworks like TensorFlow [1] or PyTorch [24], and the high-level parameters can be obfuscated by implementation details [8]. It can be a time-consuming and frustrating experience to debug, reproduce, and adapt existing models for DIA, and *many researchers who would benefit the most from using these methods lack the technical background to implement them from scratch.* 2) Document images contain diverse and disparate patterns across domains, and customized training is often required to achieve a desirable detection accuracy. Currently *there is no full-fledged infrastructure for easily curating the target document image datasets and fine-tuning or re-training the models.* 3) DIA usually requires a sequence of models and other processing to obtain the final outputs. Often research teams use DL models and then perform further document analyses in separate processes, and these pipelines are not documented in any central location (and often not documented at all). This makes it *difficult for research teams to learn about how full pipelines are implemented* and *leads them to invest significant resources in reinventing the DIA wheel.*

`LayoutParser` provides a unified toolkit to support DL-based document image analysis and processing. To address the aforementioned challenges, `LayoutParser` is built with the following components:

1. An off-the-shelf toolkit for applying DL models for layout detection, character recognition, and other DIA tasks (Section 3)
2. A rich repository of pre-trained neural network models (Model Zoo) that underlies the off-the-shelf usage
3. Comprehensive tools for efficient document image data annotation and model tuning to support different levels of customization
4. A DL model hub and community platform for the easy sharing, distribution, and discussion of DIA models and pipelines, to promote reusability, reproducibility, and extensibility (Section 4)

The library implements simple and intuitive Python APIs without sacrificing generalizability and versatility, and can be easily installed via pip. Its convenient functions for handling document image data can be seamlessly integrated with existing DIA pipelines. With detailed documentations and carefully curated tutorials, we hope this tool will benefit a variety of end-users, and will lead to advances in applications in both industry and academic research.

`LayoutParser` is well aligned with recent efforts for improving DL model reusability in other disciplines like natural language processing [8, 34] and computer vision [35], but with a focus on unique challenges in DIA. We show `LayoutParser` can be applied in sophisticated and large-scale digitization projects