

Layout Parser

A Unified Toolkit for Deep Learning Based Document Image Analysis

Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, Weining Li



Motivation

Demo

Design & Implementation

Future Work

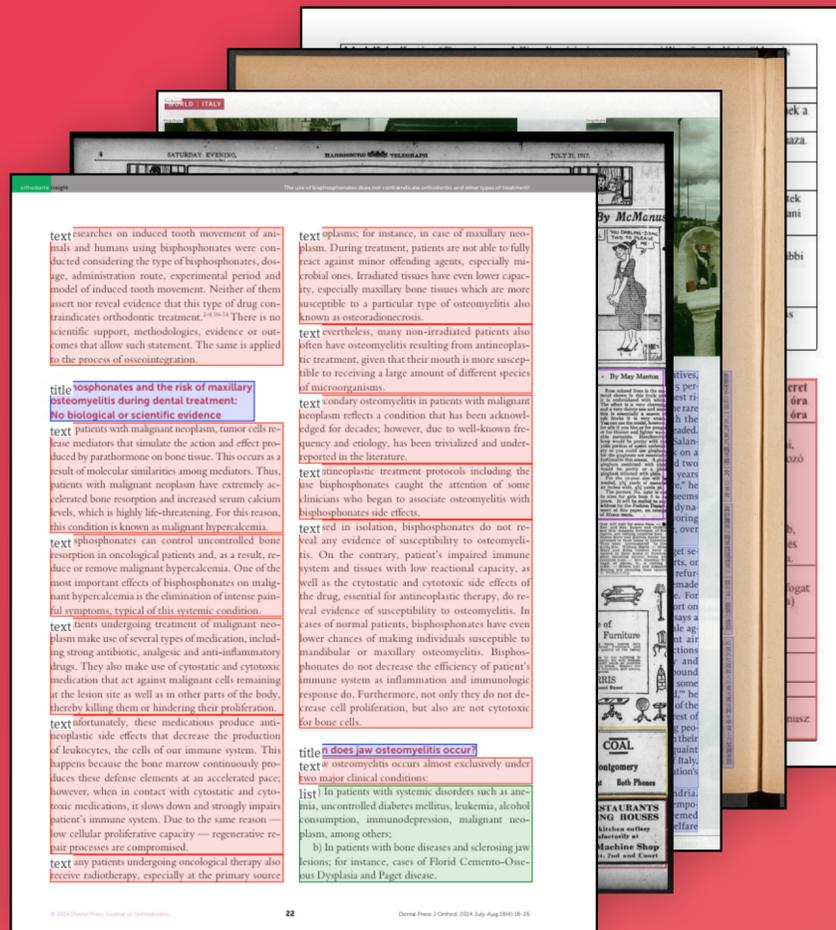
Community

Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges



Input: Doc Images

```
{  
  "title": "Construction of  
the Literature Graph in  
Semantic Scholar",  
  "authors": "Waleed Ammar et.  
al.",  
  "abstract": "We describe a  
deployed scalable system for  
organizing published  
scientific literature into a  
heterogeneous graph to  
facilitate algorithmic  
manipulation and ...",  
  "sections": ["..."]  
}
```

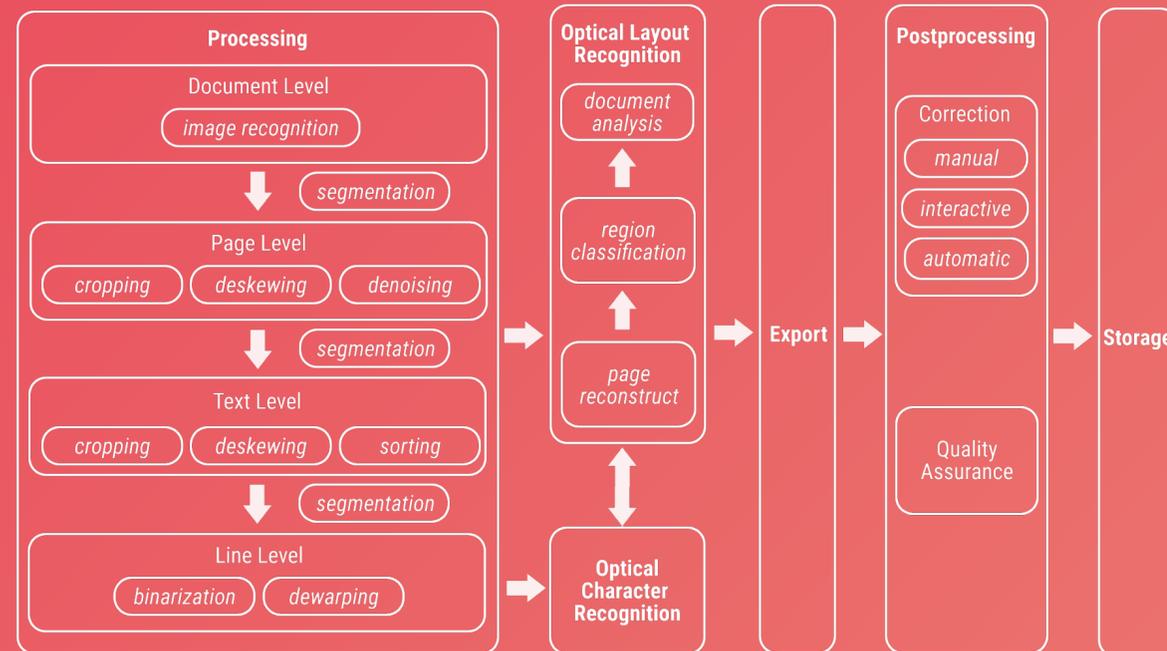
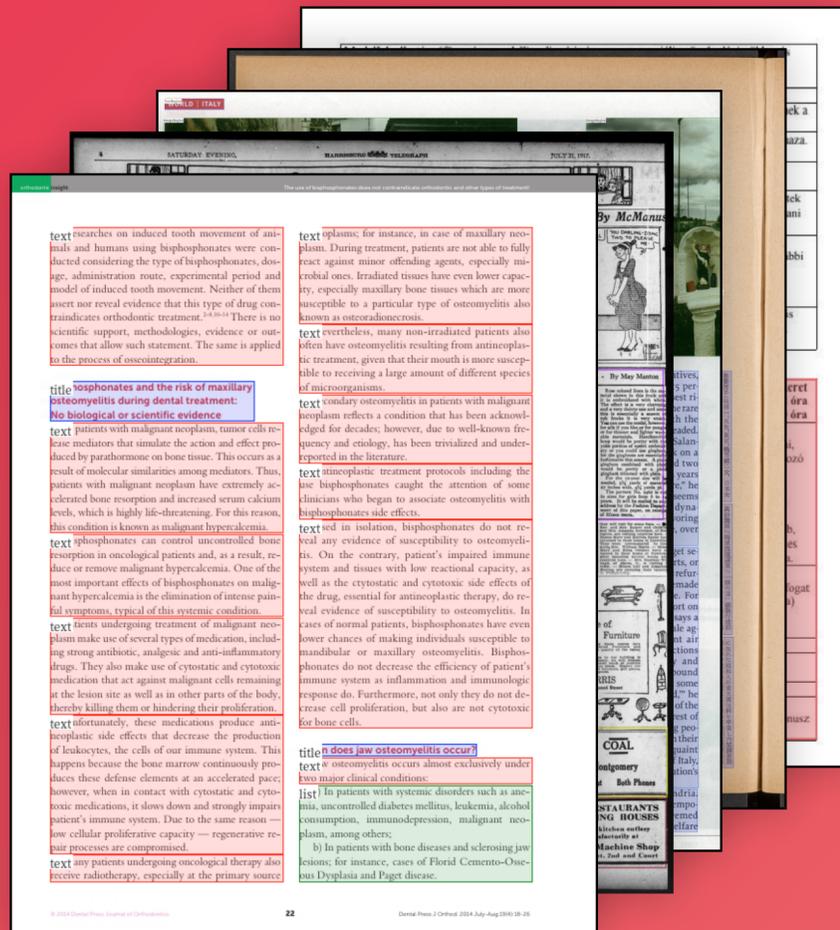
Output: Layout/Text

Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges



```
{
  "title": "Construction of
the Literature Graph in
Semantic Scholar",
  "authors": "Waleed Ammar et.
al.",
  "abstract": "We describe a
deployed scalable system for
organizing published
scientific literature into a
heterogeneous graph to
facilitate algorithmic
manipulation and ...",
  "sections": ["..."]
}
```

Input: Doc Images

DIA Pipeline

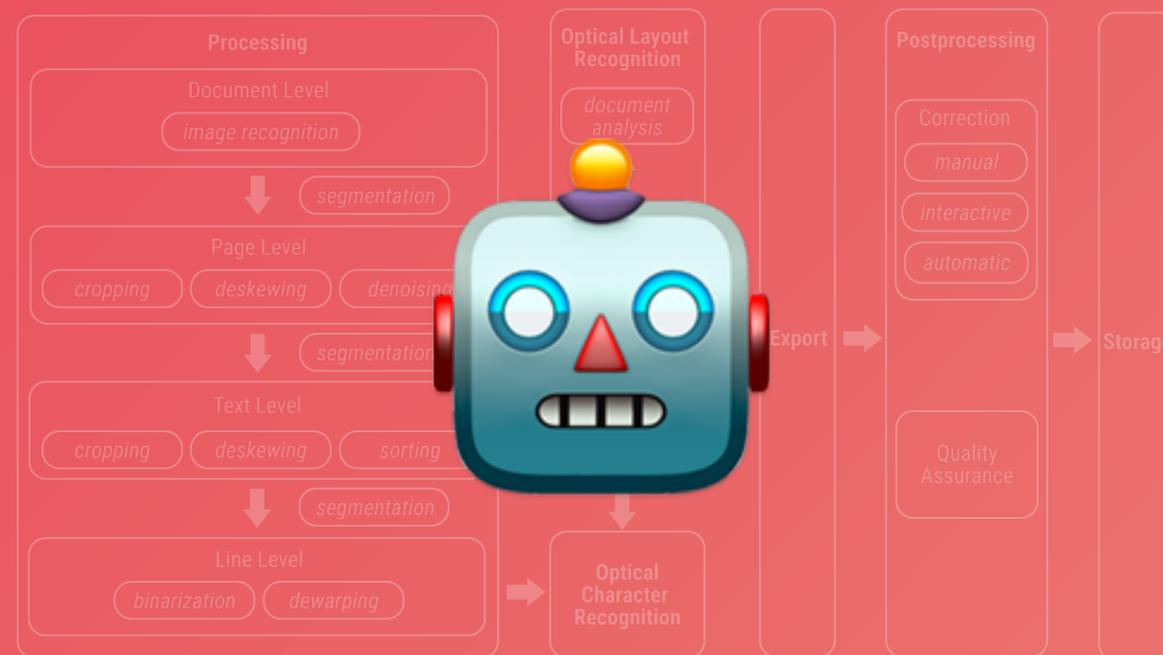
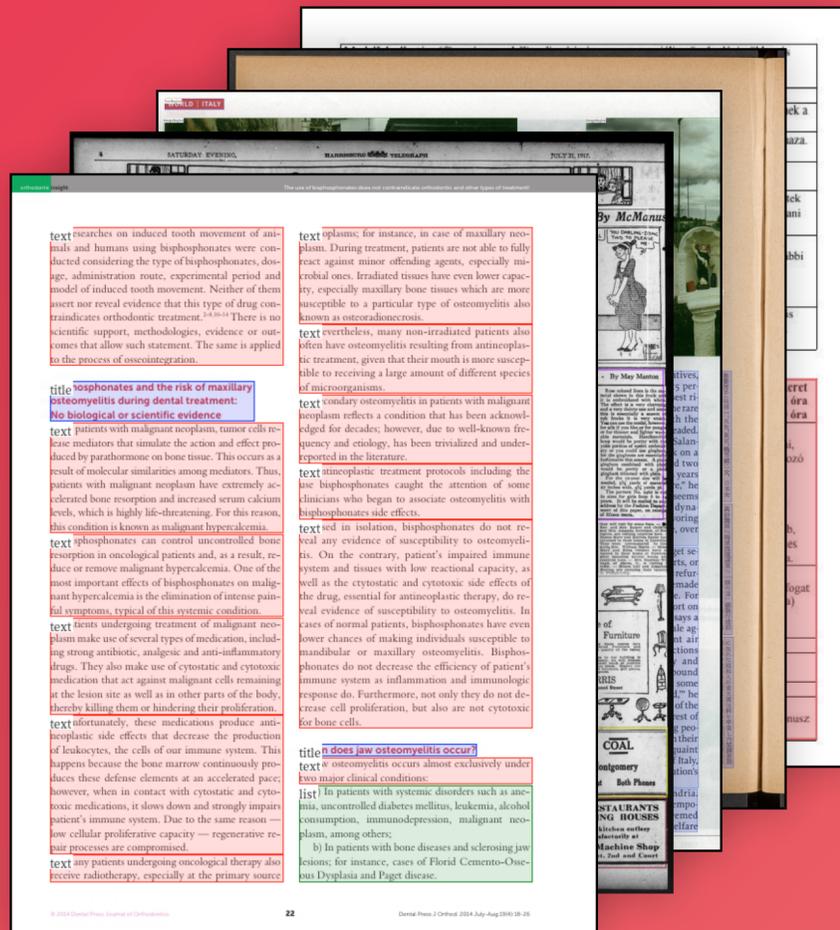
Output: Layout/Text

Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges



```
{  
  "title": "Construction of  
the Literature Graph in  
Semantic Scholar",  
  "authors": "Waleed Ammar et.  
al.",  
  "abstract": "We describe a  
deployed scalable system for  
organizing published  
scientific literature into a  
heterogeneous graph to  
facilitate algorithmic  
manipulation and ...",  
  "sections": ["..."]  
}
```

Input: Doc Images

Deep Learning Models

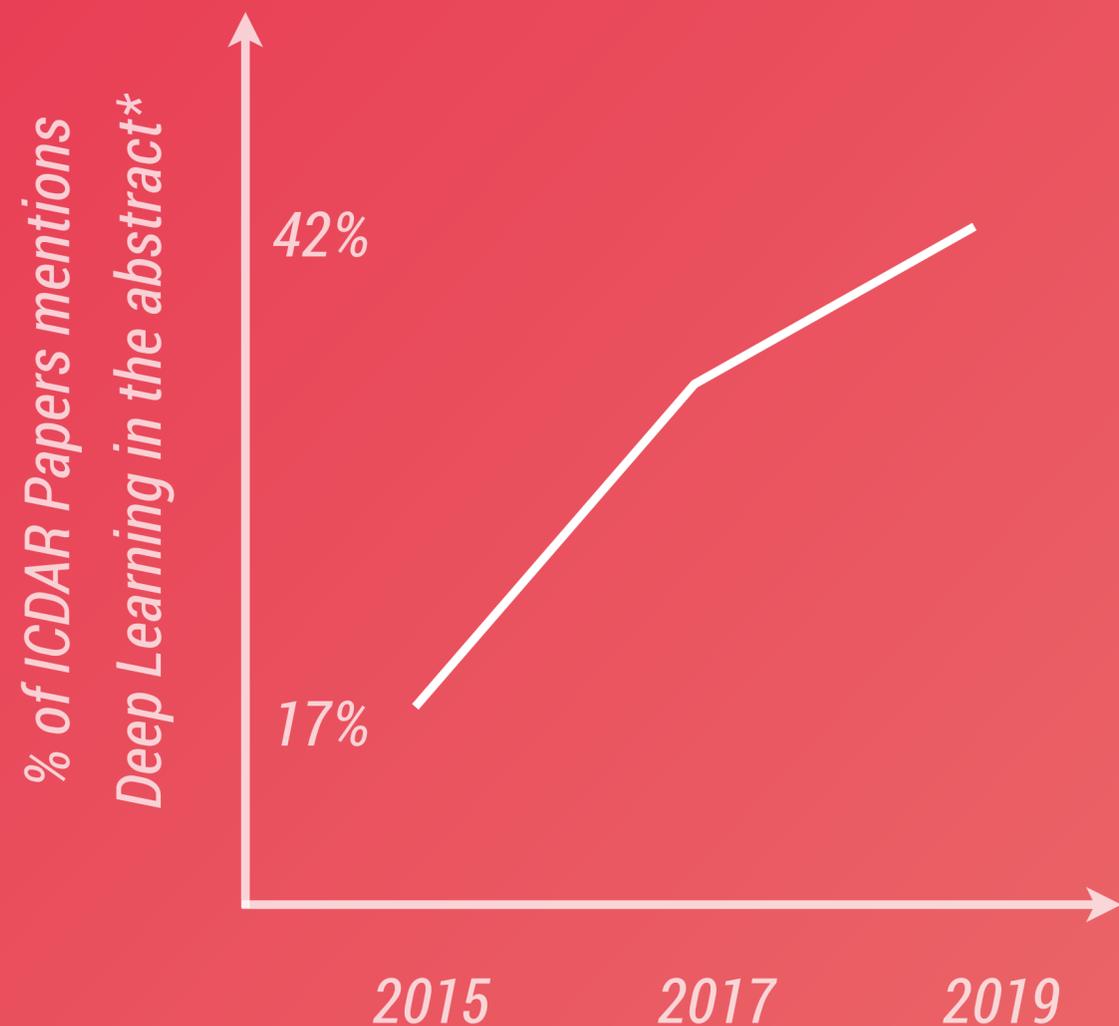
Output: Layout/Text

Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges



Large datasets

Better computation infrastructure

More work focus on DL, pushing SOTA

* We count papers with either Deep Learning, DNN, or Neural Network appeared in the abstract. Source data is from Semantic Scholar.

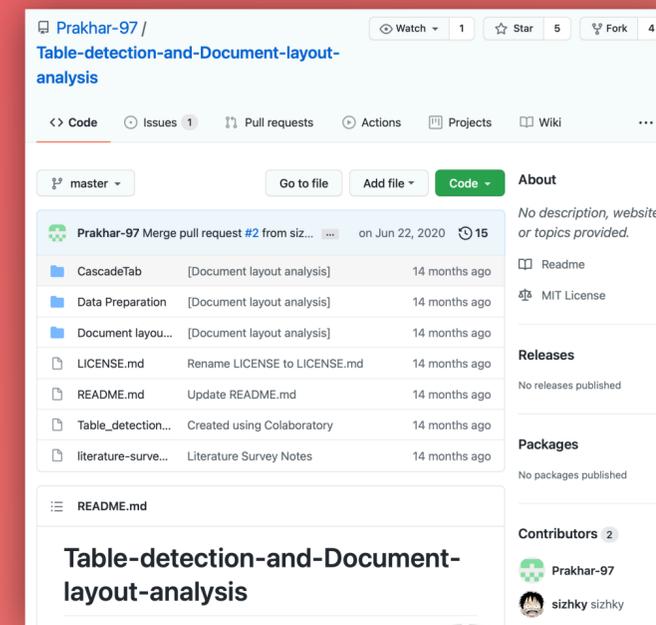
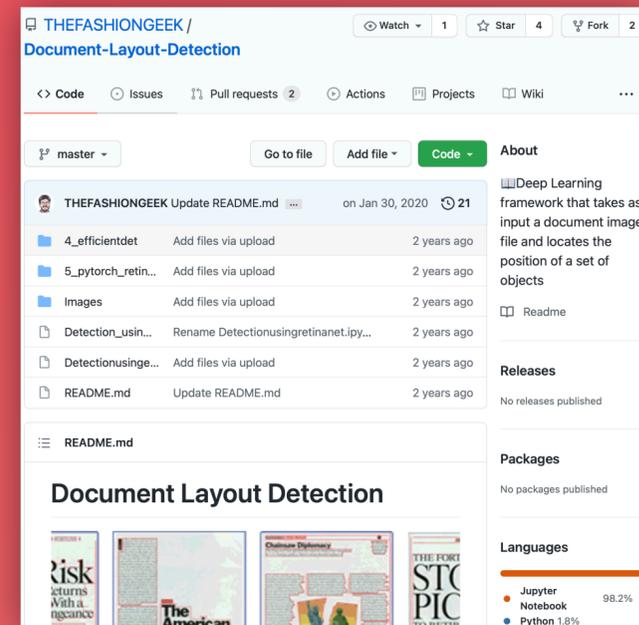
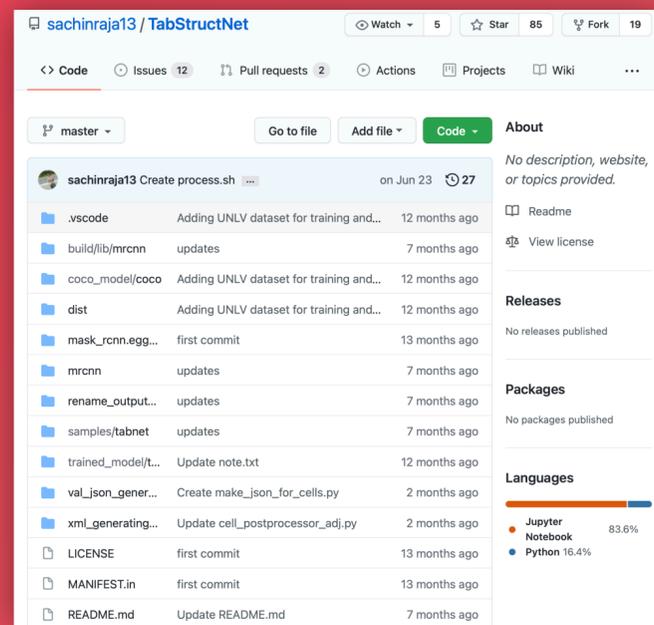
Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Model code exists in different GitHub repos,
using inconsistent DL backends & APIs



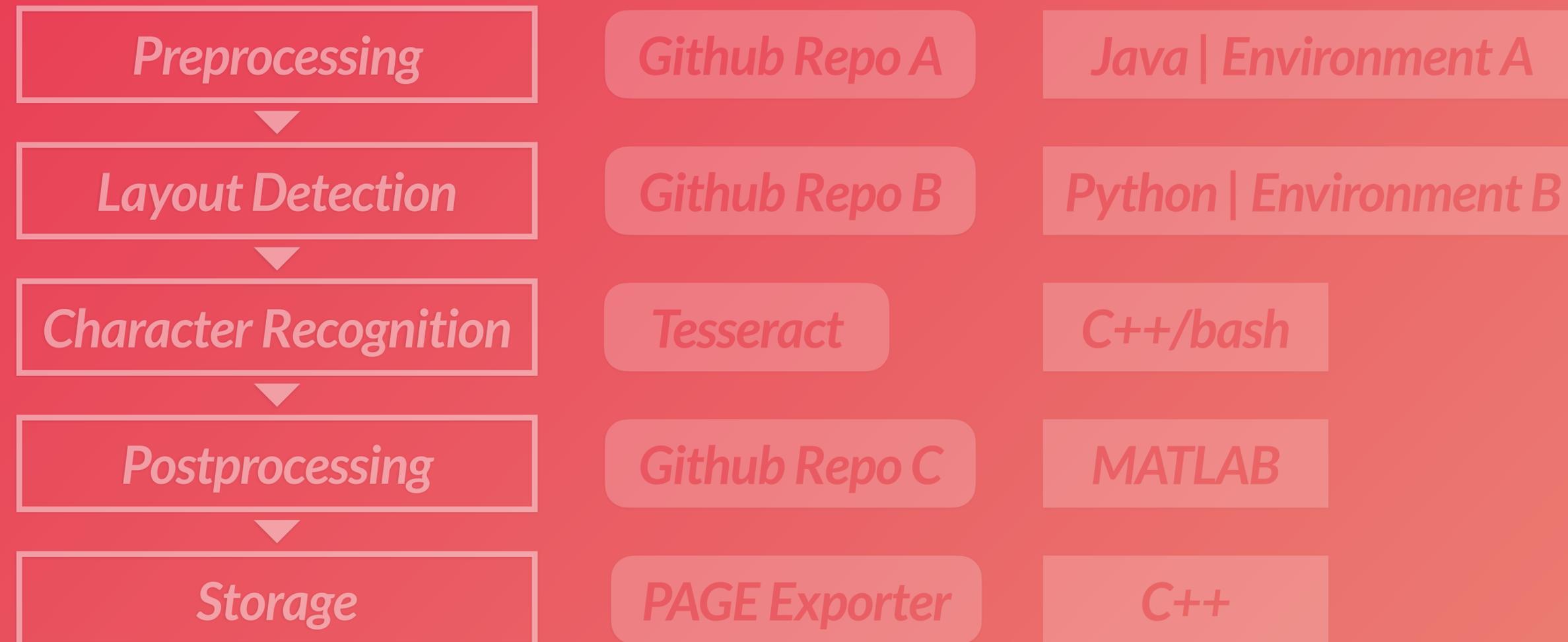
Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Hard to be incorporated to existing DIA pipelines



Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Hard to be incorporated to existing DIA pipelines



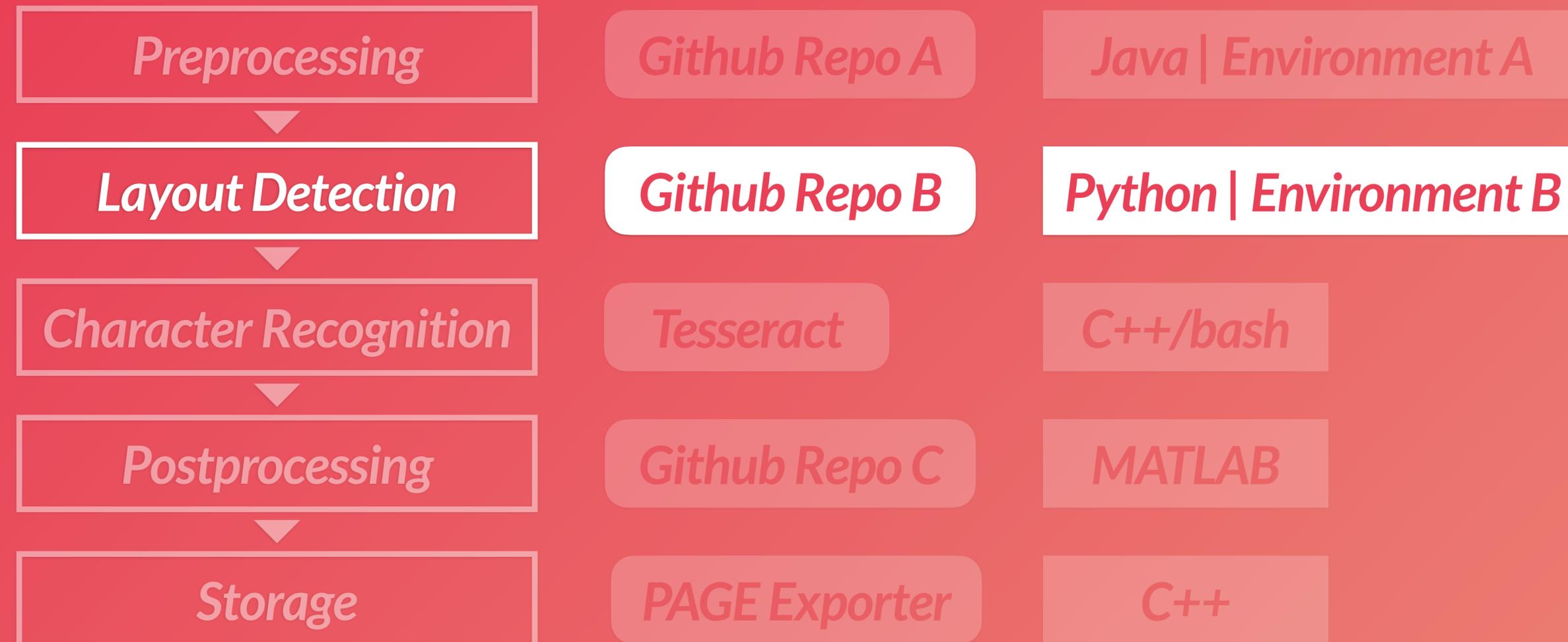
Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Hard to be incorporated to existing DIA pipelines



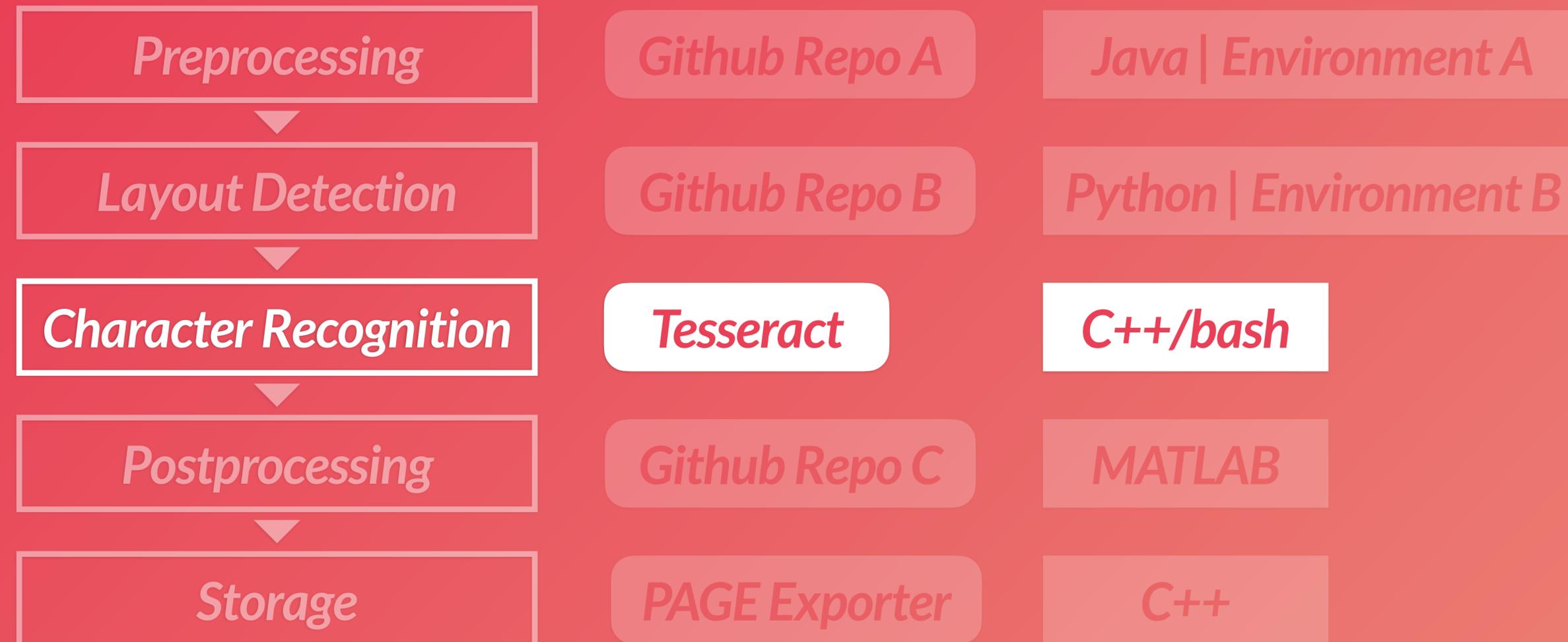
Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Hard to be incorporated to existing DIA pipelines



Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Hard to be incorporated to existing DIA pipelines



Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

Hard to be incorporated to existing DIA pipelines



Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges

The research advances becomes less accessible



for DIA researchers

hard to reproduce the results & improve models



for end users

who might come from non-technical backgrounds

What should an ideal toolkit be?

Simple

Comprehensive

Customizable

Extensible

Open Platform

Layout Parser

Simple

Comprehensive

Customizable

Extensible

Open Platform

Motivation

Demo

Design & Implementation

Future Work

Community

Layout Parser usage example

Installation

Layout Detection

Optical Character Recognition

```
$ pip install layoutparser
```

```
$ python
```

```
>>> import layoutparser as lp
```

```
>>> # Ready to go!
```

Layout Parser usage example

Installation

Layout Detection

Optical Character

```
>>> model =  
lp.Detectron2LayoutModel()  
>>> image = load_image()  
>>> layout =  
model.detect(image)
```

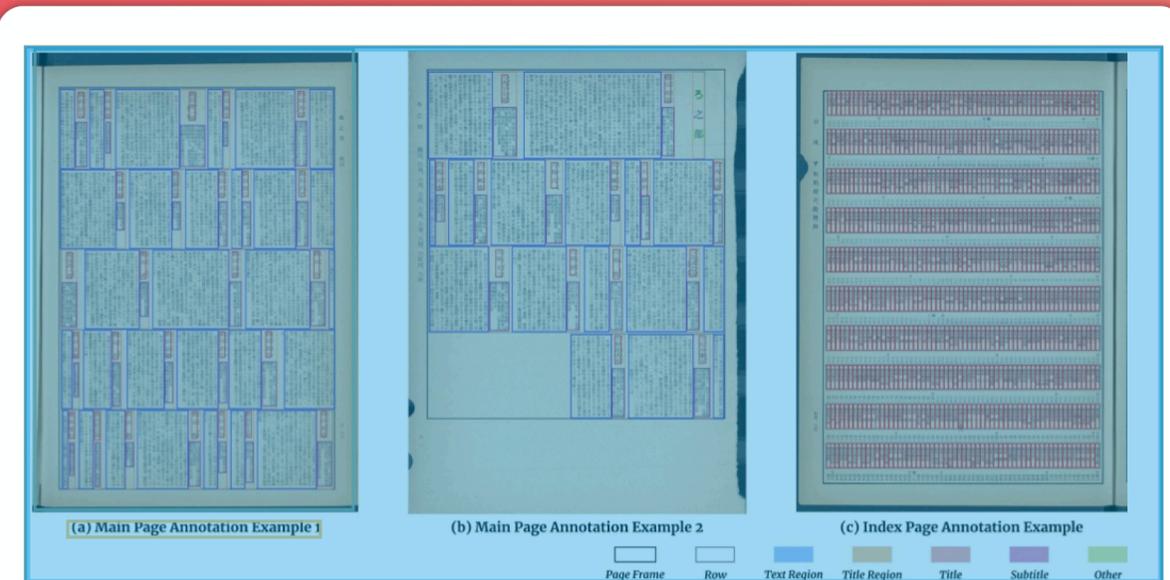


Figure 7: **Annotation Examples in HJDataset.** (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

tion over union (IOU) level $[0.50:0.95]^2$, on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

²This is a core metric developed for the COCO competition [12] for evaluating the object detection quality.

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	Faster R-CNN	Mask R-CNN ^a	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

^aFor training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.

Layout Parser usage example

Installation

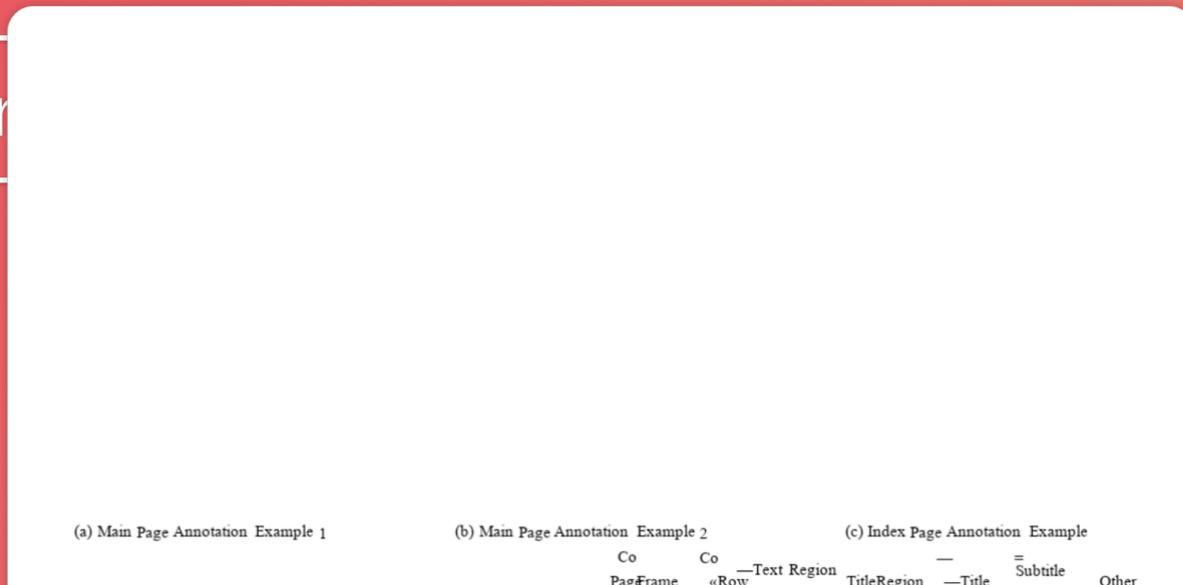
Layout Detection

Optical Character

```
>>> ocr_model =  
lp.TesseractAgent()
```

```
>>> ocr_text =  
ocr_model.detect(image)
```

```
>>> ocr_text.to_json()
```



(a) Main Page Annotation Example 1 (b) Main Page Annotation Example 2 (c) Index Page Annotation Example

Co Co —Text Region — TitleRegion — Subtitle Other
PagFrame «Row

Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

tion over union (IOU) level [0.50:0.95], on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	FasterR-CNN	Mask R-CNN*	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

*For training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.

²This is a core metric developed for the COCO competition [1] for evaluating the object detection quality.

Motivation

Demo

Design & Implementation

Future Work

Community

Deep Layout Models

Rich Repository of Pre-trained Models

Simple Model Usage

Layout Data Structure

Layout Visualization

OCR Engine Support

Data Import and Export

Open-the-box Usage

Design & Implementation

Modularized Design

Layout Data Annotation

DL Models Training

Multi-backend support

Tutorials & Examples

Sharing Platform

Community Support

Layout Data Annotation

DL Models Training

Multi-backend support

Deep Layout Models

Simple Model Usage

Layout Model Zoo

Sharing Platform

Tutorials & Examples

Community Support

Layout Data Structure

Layout Visualization

OCR Engine Support

Data Import and Export

Layout Data Annotation

**Model
Customization**

*DL Models Training
Multi-backend support*

*Deep Layout Models
Simple Model Usage*

**Deep Learning
Models for
Layout Detection**

Layout Model Zoo

*Sharing Platform
Tutorials & Examples*

**Layout Parser
Open Platform**

Community Support

Layout Data Structure

Infrastructure APIs

OCR Engine Support

Layout Visualization

Data Import and Export

Layout Data Annotation
Deep Learning
DL Models Training
Model
Multi-backend support
Customization

Deep Layout Models
Deep Learning
Simple Model Usage
Models for
Layout Detection
Layout Model Eval

Sharing Platform
Layout Parser
Tutorials & Examples
Open Platform
Community Support

Layout Visualization

Layout Data Structure
Core APIs
OCR Engine Support

Data Import and Export

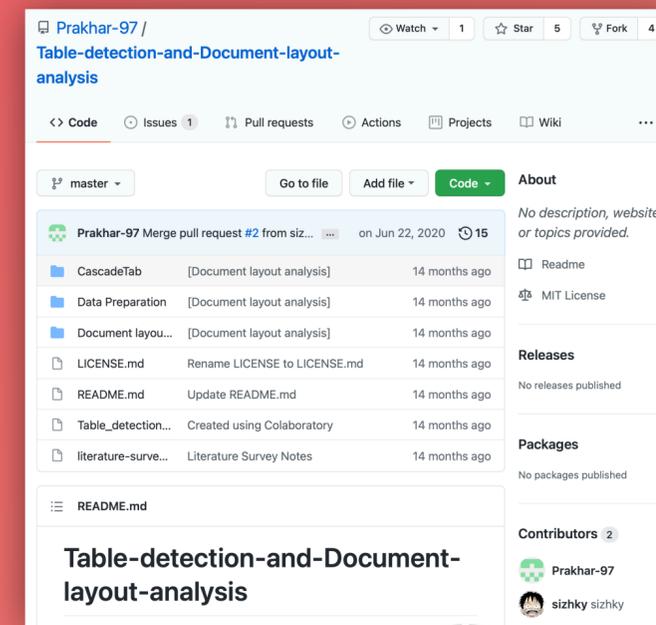
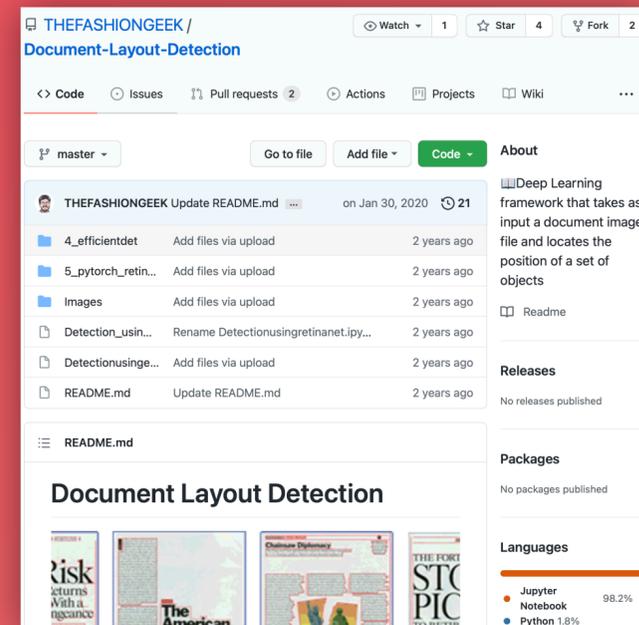
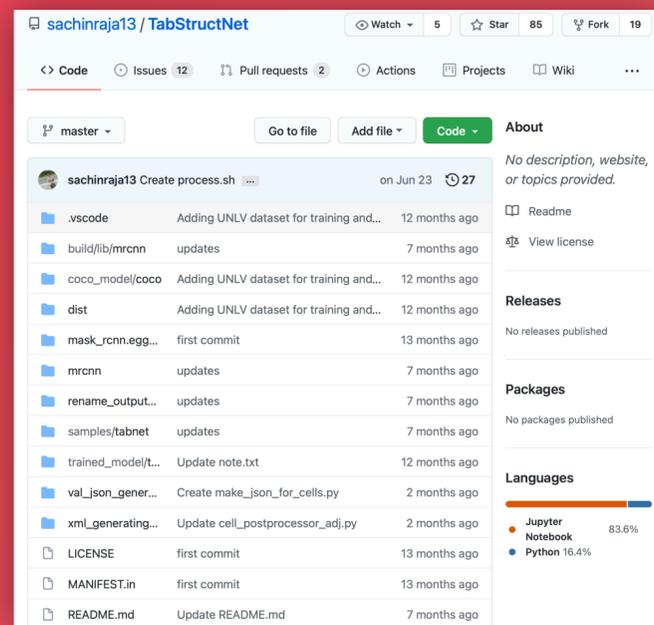
Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

No standard way for sharing and re-using existing models



Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

▼ *Specify the model configuration*

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

▼ *Training Dataset Name*

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

▼ *Model Architecture Name*

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
```

```
>>> model = lp.Detectron2LayoutModel(config)
```

```
>>> layout = model.detect(image) ▲ Standardized Model Initialization
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
```

```
>>> model = lp.Detectron2LayoutModel(config)
```

```
>>> layout = model.detect(image) ▲ The Deep Learning Backend Name
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

▲ *Standardized Model Detection API*

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

What if we want to make some changes?

```
>>> config = "lp://PubLayNet/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

Switch to a different model architecture?

```
>>> config = "lp://PubLayNet/faster_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

Switch to models trained on another dataset?

```
>>> config = "lp://PrimaLayout/mask_rcnn_R_50_FPN_3x/config"
>>> model = lp.Detectron2LayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

Switch to a different model architecture from another DL backend?

```
>>> config = "lp://PubLayNet/ppyo1ov2_r50vd_dcn_365e/config"
>>> model = lp.PaddleDetectionLayoutModel(config)
>>> layout = model.detect(image)
```

Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

Even simpler!

```
>>> model = lp.AutoLayoutModel("lp://detectron2/publaynet")
```

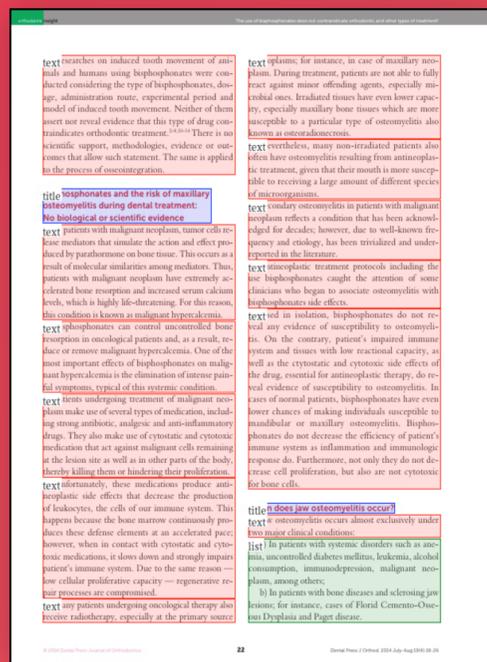
Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

Layout Parser has pre-trained models on 6 datasets, including:



PubLayNet



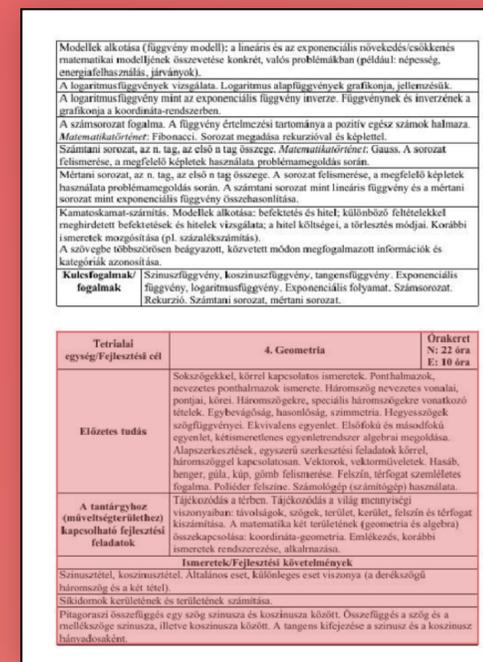
Newspaper Navigator



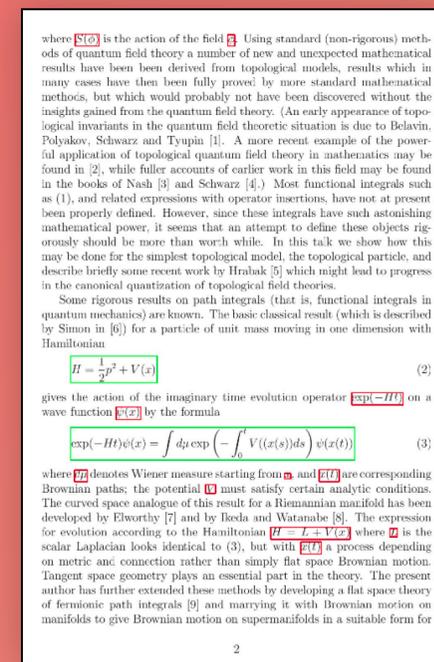
PRImA Layout



HJDataset



TableBank



Math Formula Detection

What if we need to post-process model outputs?

Layout Data Annotation
Deep Learning
DL Models Training
Model
Multi-backend support
Customization

Deep Layout Models
Deep Learning
Simple Model Usage
Models for
Layout Detection
Layout Model Zoo

Sharing Platform
Layout Parser
Tutorials & Examples
Open Platform
Community Support



Layout Data Structure
Infrastructure APIs
OCR Engine Support

Layout Visualization

Data Import and Export

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout Visualization

Load & Storage

Convenient APIs that simplify postprocessing

```
>>> layout = model.detect(image)
```

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout Visualization

Select layout regions as input for postprocessing

```
>>> layout = model.detect(image)
```

```
>>> width, height = image.size
```

```
>>> left_column = \
lp.Interval(0, width/2, axis="x")
```

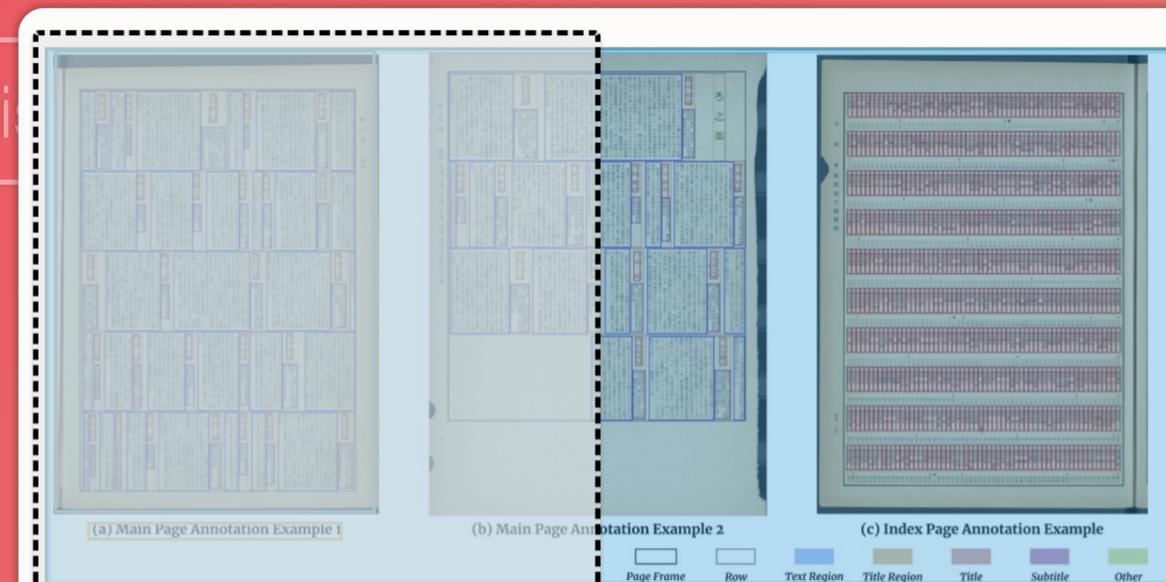


Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

tion over union (IOU) level [0.50:0.95]², on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

²This is a core metric developed for the COCO competition [12] for evaluating the object detection quality.

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	Faster R-CNN	Mask R-CNN ^a	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

^aFor training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout Vi

Select layout regions in the left column:

```
>>> selected = layout.filter_by(  
    left_column, center=True)
```

```
>>> selected.sort()
```

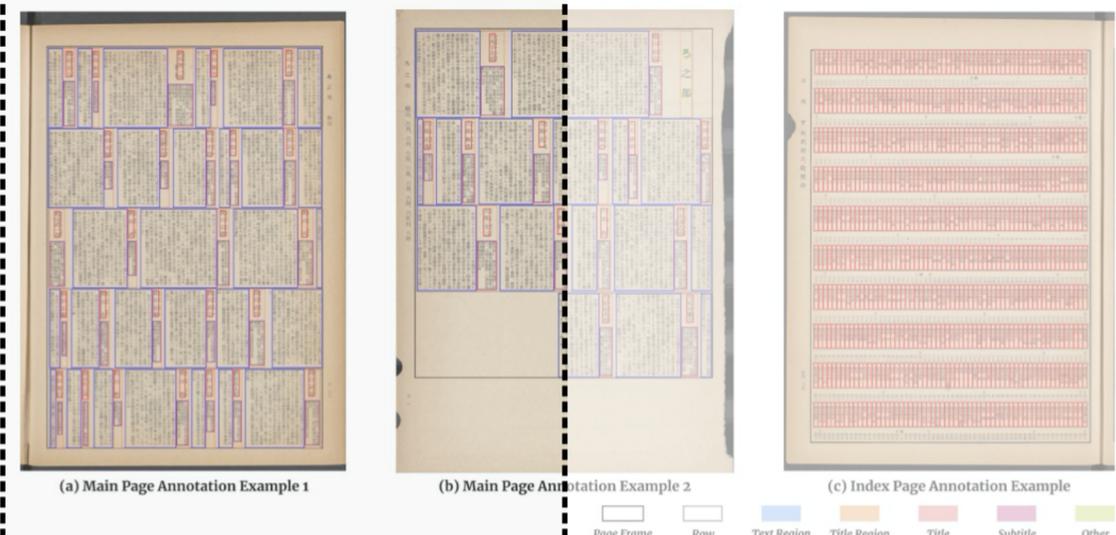


Figure 7: **Annotation Examples in HJDataset.** (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

Figure 8: (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

²This is a core metric developed for the COCO competition [12] for evaluating the object detection quality.

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	Faster R-CNN	Mask R-CNN ^a	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

^a For training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout Visualization

Load & Storage

Interaction between layout regions and OCR

```
>>> layout = model.detect(image)
```

```
>>> ocr_text = ocr_model.detect(image)
```

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout Visualization

Load & Storage

Interaction Between each API and OCR

```
>>> layout = model.detect(image)
>>> for block in layout:
    segment = block.crop_image(segment)
    block.text = ocr_agent.detect(segment)
```

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout V

Highly configurable visualization

```
>>>> draw_box(image, layout,  
             show_element_type=True,  
             show_element_id=True,  
             box_width=4,  
             color_map={...})  
  
>>>> draw_text(image, ocr_text,  
              with_box_on_text=True, ...)
```

The screenshot displays a document page with various layout elements highlighted in different colors. A legend at the top right identifies the categories: PageFrame (blue), Text Region (orange), Title Region (green), Subtitle (red), and Other (purple). Below the legend, there are three examples of annotations: (a) Main Page Annotation Example 1, (b) Main Page Annotation Example 2, and (c) Index Page Annotation Example. The main text of the page is annotated with bounding boxes, and a table at the bottom right shows the detection mAP for different categories and models.

Category	FasterR-CNN	Mask R-CNN*	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

*For training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.

Layout Parser Infrastructure APIs

Layout Data Structures

OCR Engine Support

Layout Visualization

Load & Storage

Exporting

```
>>> layout.to_csv()
```

```
>>> layout.to_json()
```

Loading

```
>>> layout = lp.load_csv()
```

```
>>> layout = lp.load_json()
```

Currently supports

CSV

JSON

More formats will be added:

PAGES

METS/ALTO

hOCR

...

What if we want better models?

Layout Data Annotation

Model Customization

DL Models Training

Multi-backend support



Deep Layout Models

Deep Learning Models for Layout Detection

Simple Model Usage

Layout Model Zoo

Sharing Platform

Layout Parser

Tutorials & Examples

Open Platform

Community Support



Layout Data Structure

Infrastructure APIs

Layout Visualization

OCR Engine Support

Data Import and Export

Deep Learning Model Customization

Why

How different is the target data?



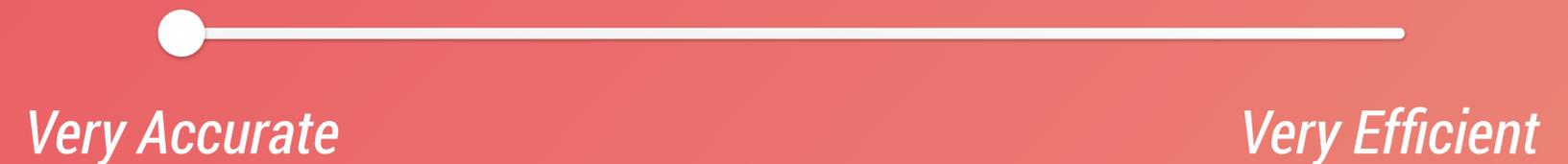
Deep Learning Model Customization

Why

How different is the target data?



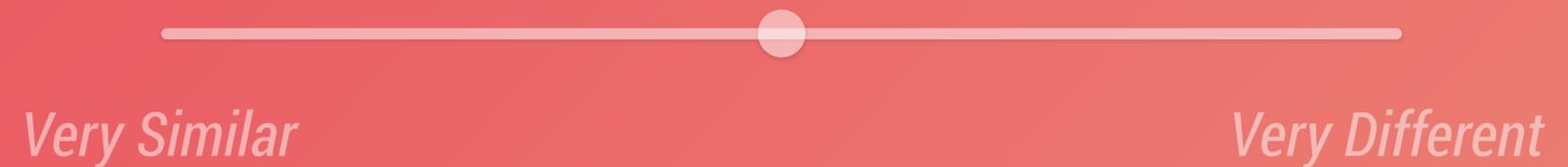
Accuracy/efficiency requirements?



Deep Learning Model Customization

Why

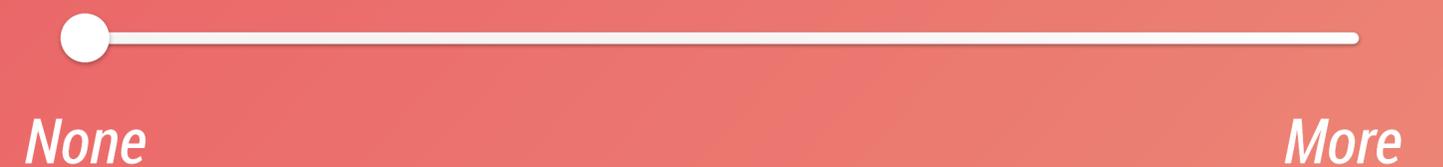
How different is the target data?



Accuracy/efficiency requirements?



How much training data is available?



Deep Learning Model Customization

Target Data Difference



Accuracy/efficiency trade-off



Available training data



Deep Learning Model Customization

Target Data Difference



Similar

Different

Multi-backend Support



Detectron2

Accurate

Accuracy/efficiency trade-off



Accurate

Efficient

EfficientDet

Efficient

Available training data



None

More



PaddleDetection

And more in the future

Deep Learning Model Customization

Target Data Difference



Accuracy/efficiency trade-off



Available training data



Deep Learning Model Customization

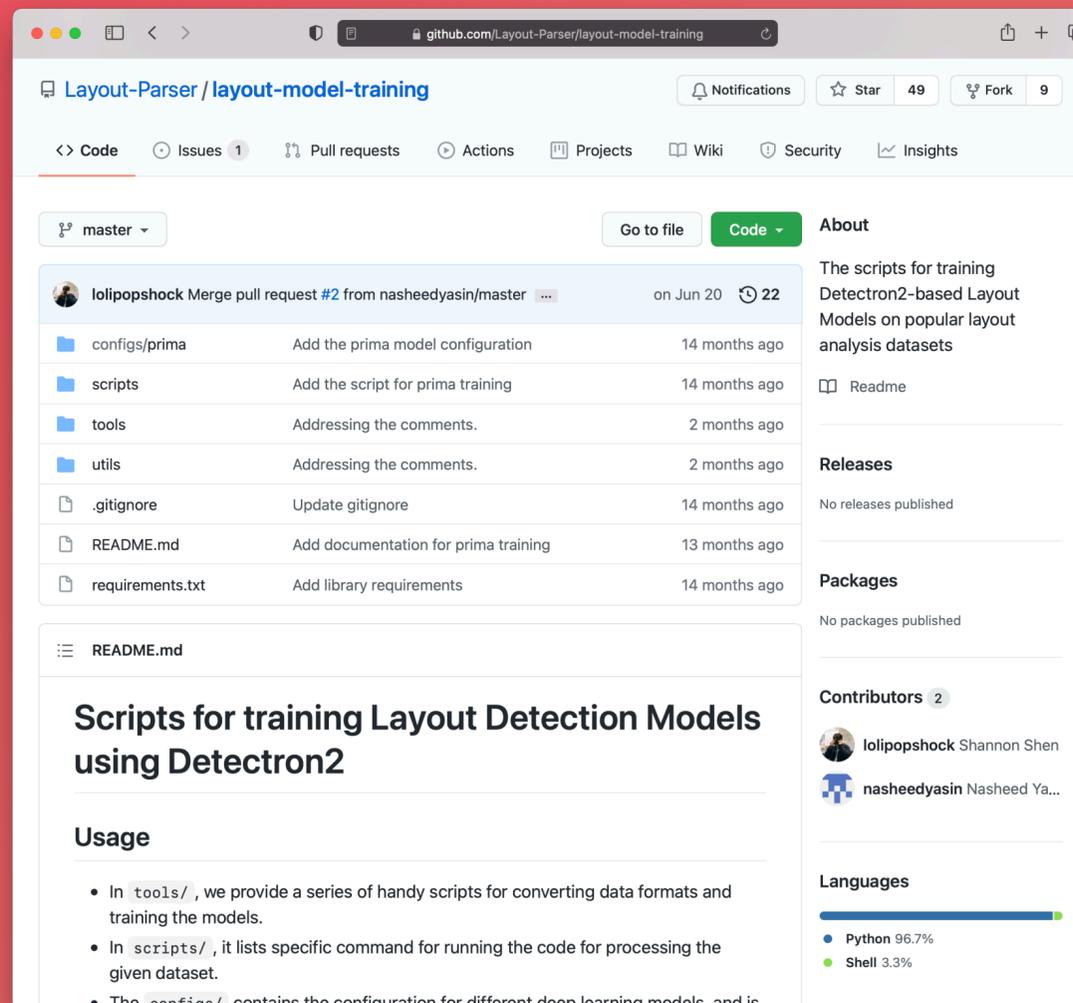
Target Data Difference



Similar

Different

Model Fine-tuning



Accuracy/efficiency trade-off



Accurate

Efficient

Available training data



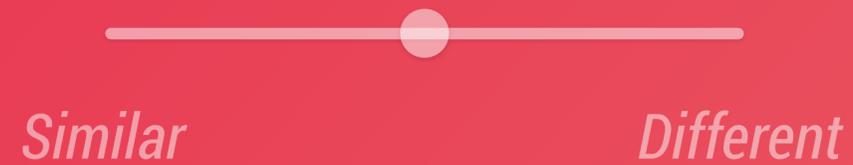
None

More

▲ *Layout Parser comes with the script that fine-tunes existing models to new datasets.*

Deep Learning Model Customization

Target Data Difference



Accuracy/efficiency trade-off



Available training data



Deep Learning Model Customization

Target Data Difference

Similar

Different

Annotation & Model Retraining

Accuracy/efficiency trade-off

Accurate

Efficient

Available training data

None

More



③ Prediction Confidence Slider
Filter boxes that are within the confidence range

② Class Selector
Annotators can pick the class for the text region

④ Switch between current labels and model predictions

① The Labeling Canvas
Where different box colors are for different classes

▲ Layout Parser incorporates labeling toolkits from existing resources to streamline the labeling and improve efficiency.

How about sharing your work with the community?

Layout Data Annotation
Deep Learning Model Training
DL Model Training
Customization
Multi-backend support



Deep Layout Models
Models for Layout Detection
Simple Model Usage
Layout Model Zoo



Sharing Platform
Layout Parser Open Platform
Tutorials & Examples
Community Support

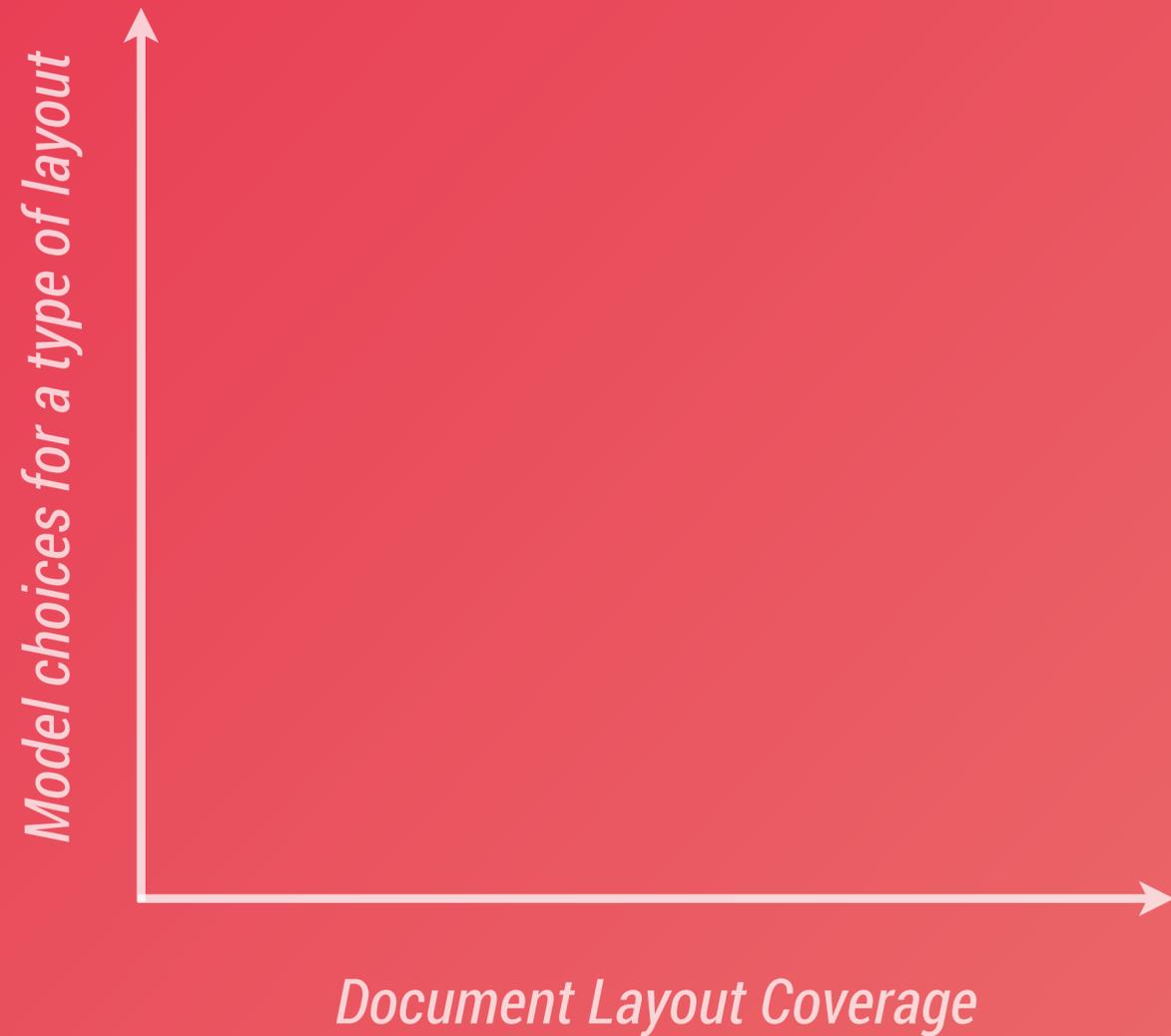


Layout Data Structure
Infrastructure APIs
Layout Visualization *OCR Engine Support* *Data Import and Export*

Layout Parser Open Platform

Share Layout Models

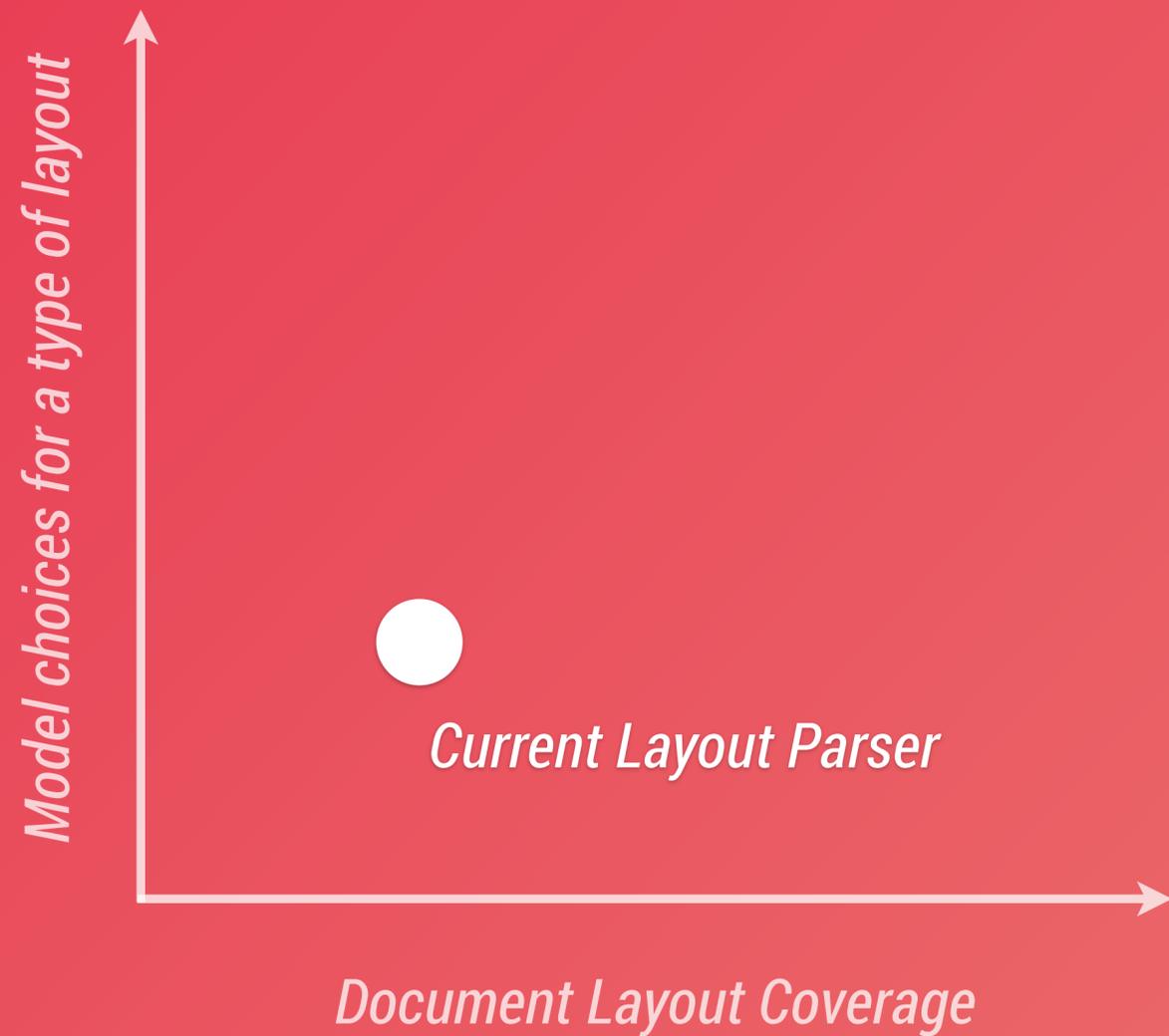
Share DIA Pipelines



Layout Parser Open Platform

Share Layout Models

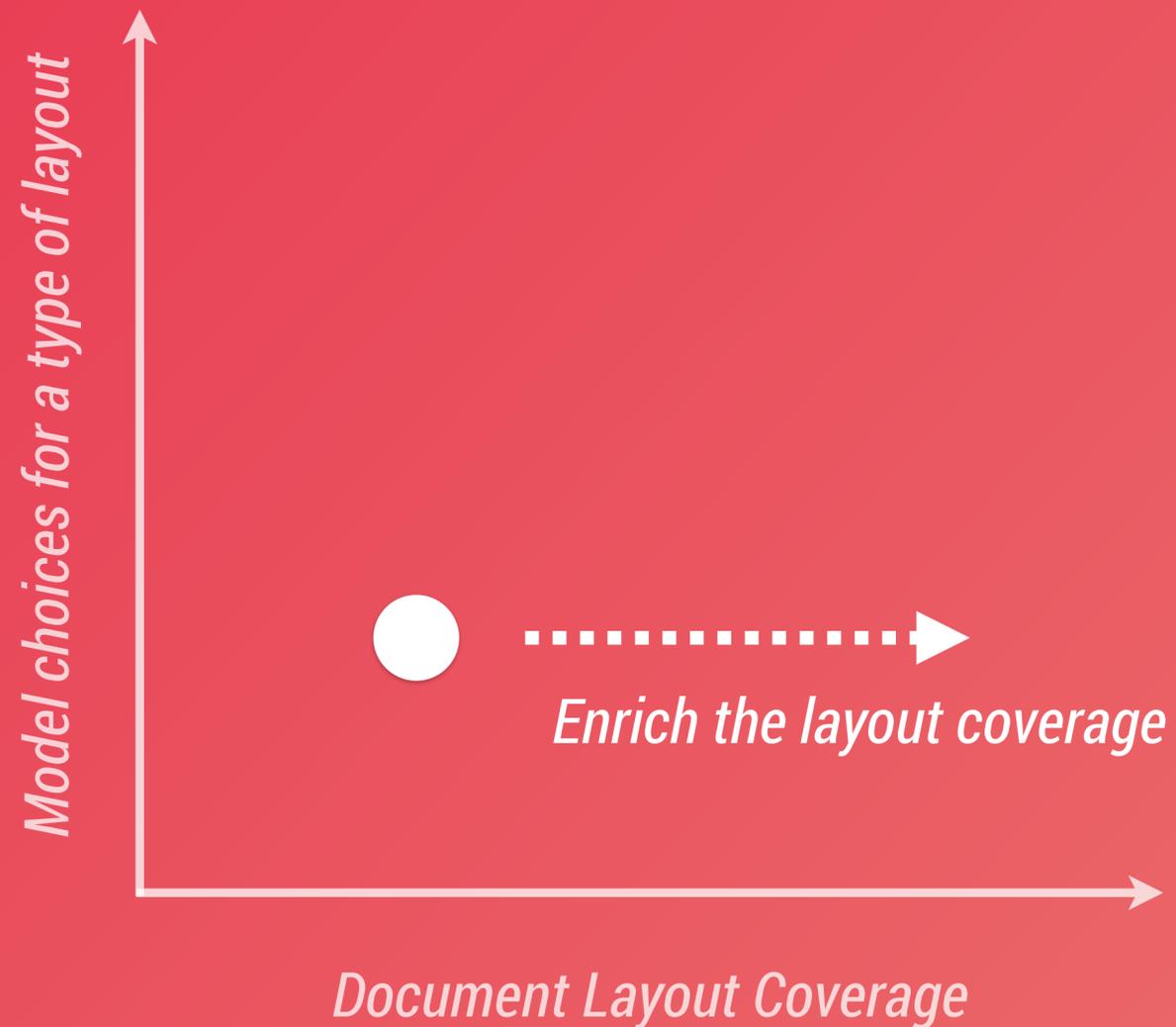
Share DIA Pipelines



Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines



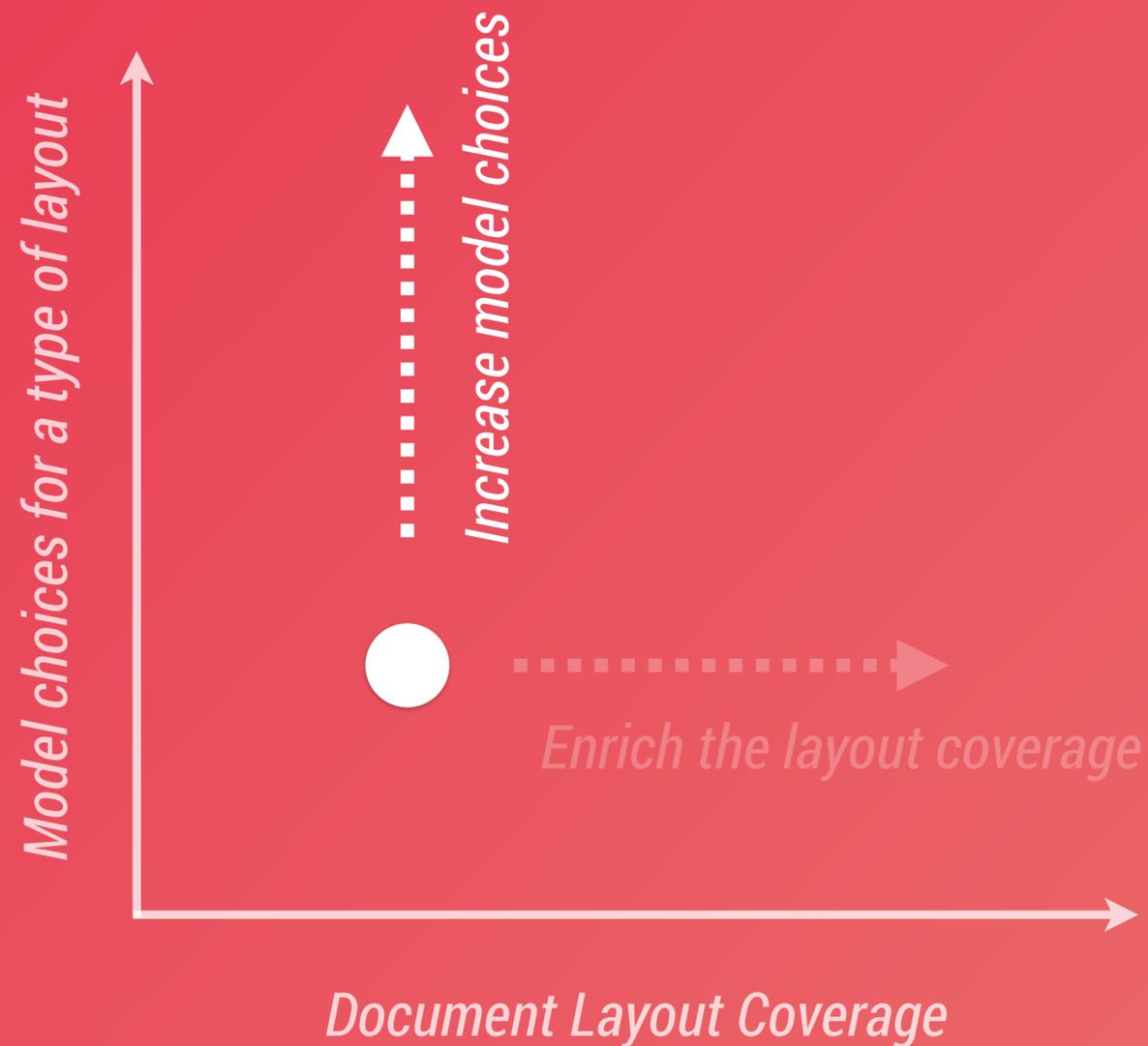
If we share:

Models trained on different datasets

Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines



If we share:

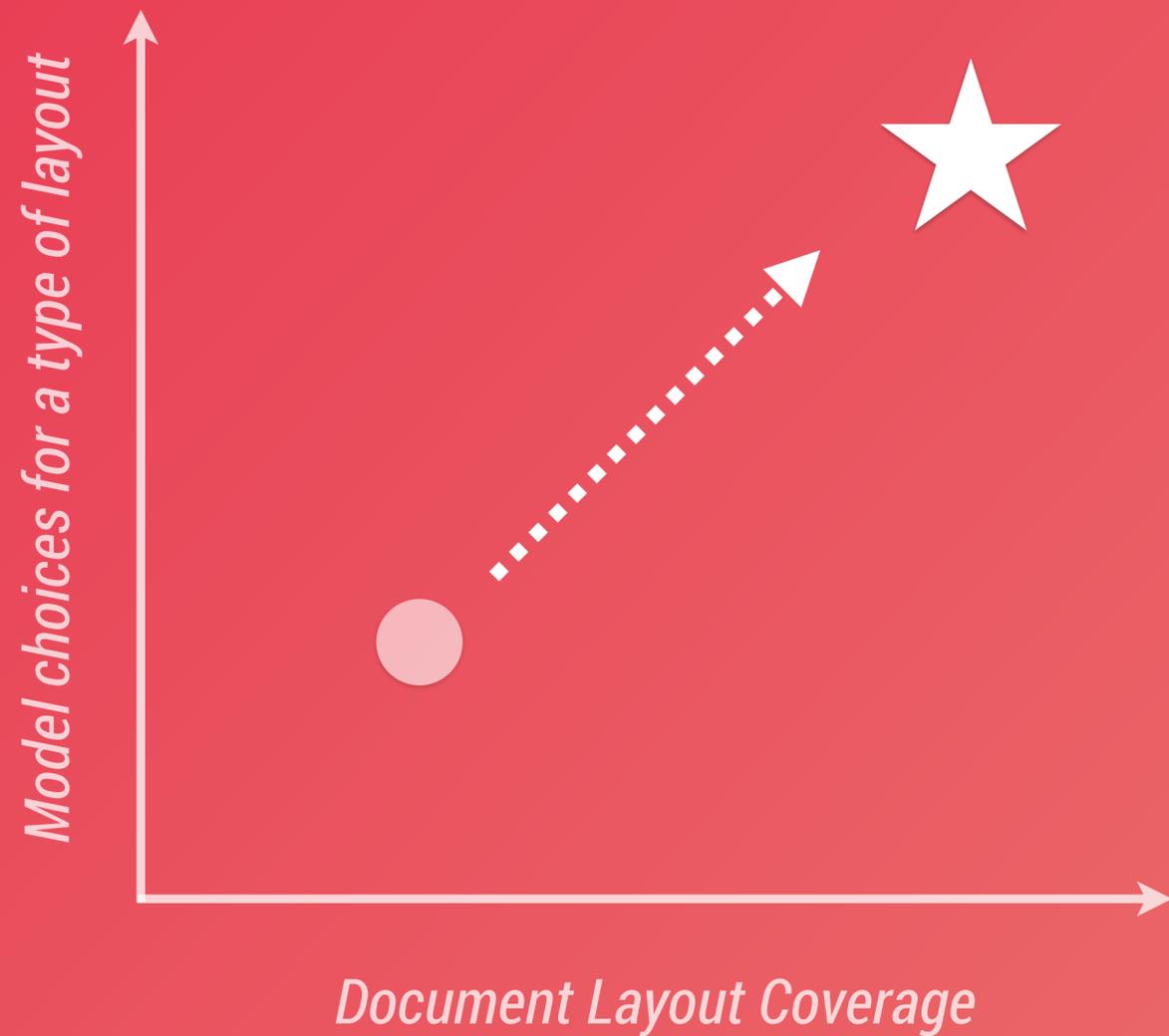
Models trained on different datasets

Models of different architecture/backend

Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines



Ultimately:

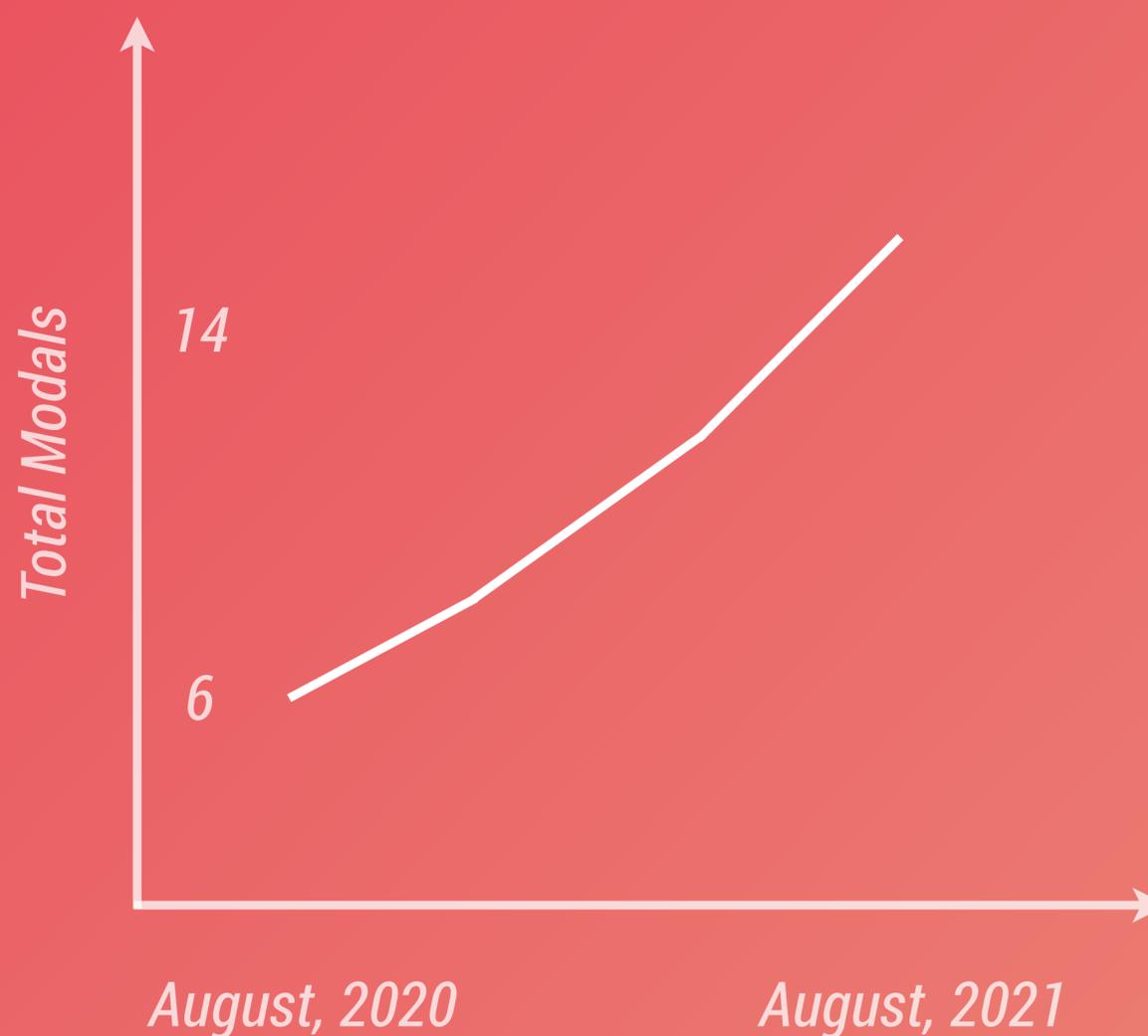
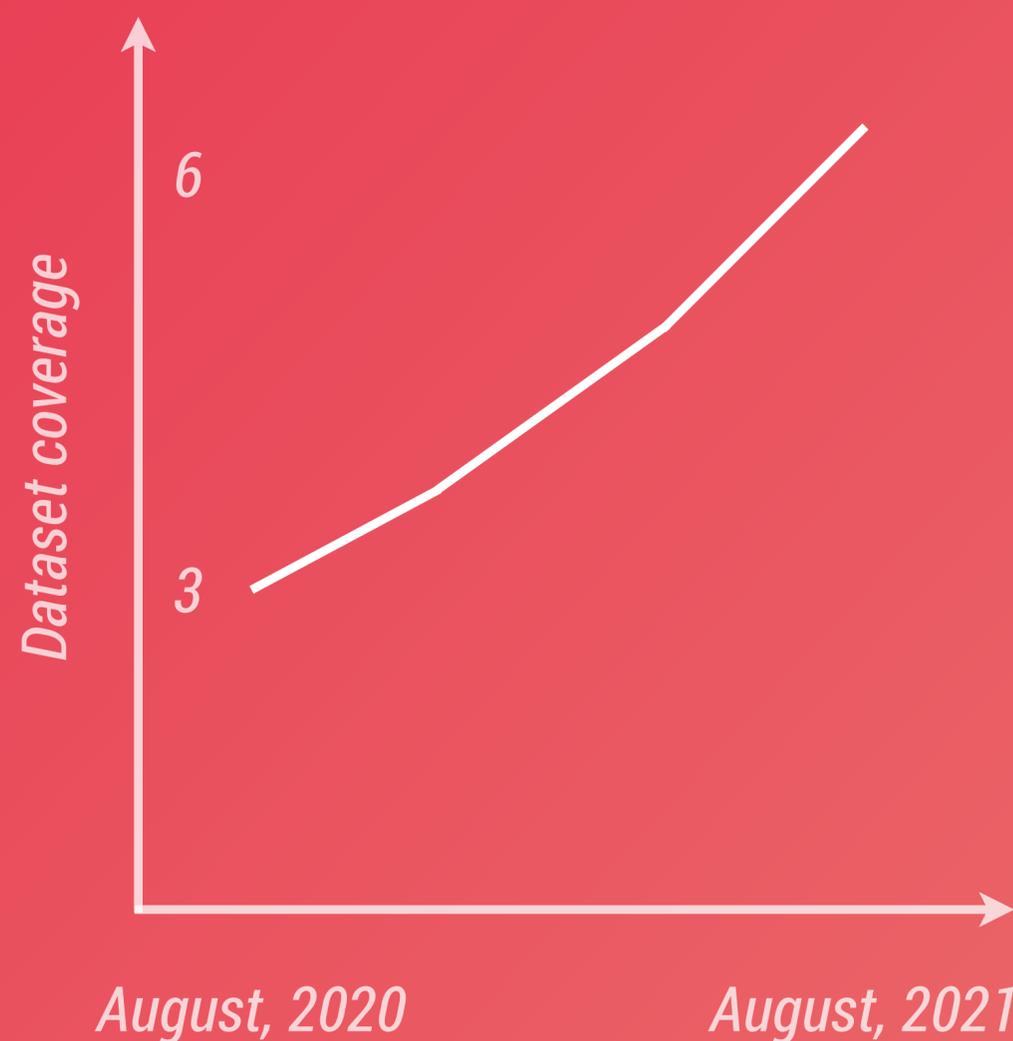
Make it easier to find the ideal model

Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

14 models for 6 datasets, 2x in the past year, with the help of community



Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

DIA pipelines have multiple steps:

Preprocessing



Layout Detection



Character Recognition



Postprocessing



Storage

Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

DIA pipelines have multiple steps:

Preprocessing

Layout Detection

Character Recognition

Postprocessing

Storage

Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

DIA pipelines have multiple steps:

Can we share them as a whole?

Preprocessing



Layout Detection



Character Recognition



Postprocessing



Storage

Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

Examples:

Table Extraction

The first screenshot shows a legal docket entry with a table of docket text. The second screenshot shows a transaction receipt table with columns for PACER, Date, Case, and Description. The third screenshot shows a PACER service center table with columns for PACER, Date, Case, and Description.

Scientific Document Parsing

The diagram illustrates the scientific document parsing process. It shows a screenshot of a PDF document (a) with colored boxes indicating the text blocks. A segment with complex layout structures (b) is shown with text lines in rectangles with black outlines. Text inputs with layout indicators [BLK] (c) are shown with text highlighted according to their semantic categories. The H-VILA: Visual Layout-guided Hierarchical Model (d) is shown as a flowchart with Group Encoder, Page Encoder, and MLP Classifier components.

▲ Layout parser has been used for extraction tables and other layout elements in legal documents.

▲ Layout parser is used in a research project that studies how layout information can improve scientific document analysis.

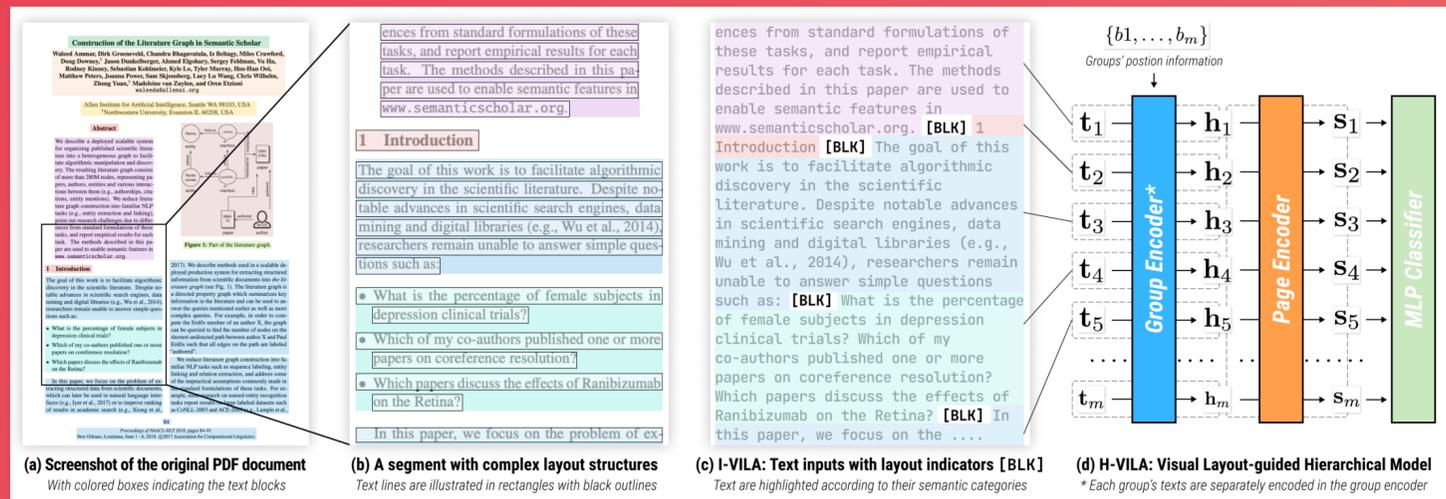
Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

Examples:

Scientific Document Parsing



Historical Document Analysis



▲ Layout parser is used in a research project that studies how layout information can improve scientific document analysis.

▲ Layout parser is also used for large-scale digitization of Historical Japanese Documents

Layout Data Annotation

DL Models Training

Multi-backend support

Deep Layout Models

Simple Model Usage

Layout Model Zoo

Sharing Platform

Tutorials & Examples

Community Support

Layout Data Structure

Layout Visualization

OCR Engine Support

Data Import and Export

**Deep Learning
Models for
Layout Detection**

Layout Data Annotation

DL Models Training

Multi-backend support

Sharing Platform

Tutorials & Examples

Community Support

Layout Data Structure

Layout Visualization

OCR Engine Support

Data Import and Export

Deep Learning Models
Simple Model Usage
Layout Model Test

Deep Learning Models for Layout Detection



Layout Data Structure

Infrastructure APIs

OCR Engine Support

Layout Data Annotation

DL Models Training

Multi-backend support

Sharing Platform

Tutorials & Examples

Community Support

Layout Visualization

Data Import and Export

Layout Data Annotation
Model Customization
DL Models Training
Multi-backend support



Deep Layout Models
Deep Learning Models for Layout Detection
Simple Model Usage
Layout Model Zoo

Sharing Platform
Tutorials & Examples
Community Support



Layout Data Structure
Infrastructure APIs
Layout Visualization *OCR Engine Support* *Data Import and Export*

Layout Data Annotation

Model Customization

DL Models Training

Multi-backend support



Deep Layout Models

Deep Learning Models for Layout Detection

Simple Model Usage

Layout Model Zoo



Sharing Platform

Layout Parser Open Platform

Tutorials & Examples

Community Support



Layout Data Structure

Infrastructure APIs

Layout Visualization *OCR Engine Support* *Data Import and Export*

Motivation

Demo

Design & Implementation

Future Work

Community

Future Work

Generalized Models

Multimodal Modeling

Future Work

Generalized Models

Multimodal Modeling

text teaches on induced tooth movement of animals and humans using bisphosphonates were conducted considering the type of bisphosphonates, dosage, administration route, experimental period and model of induced tooth movement. Neither of them assert nor reveal evidence that this type of drug counteracts orthodontic treatment.^{10,11,12} There is no scientific support, methodology, evidence or outcomes that allow such statement. The same is applied to the process of osseointegration.

title Bisphosphonates and the risk of maxillary osteomyelitis during dental treatment
No biological or scientific evidence

text Patients with malignant neoplasm, tumor cells release mediators that stimulate the action and effect produced by parathyroid hormone on bone tissue. This occurs as a result of molecular similarities among mediators. Thus, patients with malignant neoplasm have extremely accelerated bone resorption and increased serum calcium levels, which is highly life-threatening. For this reason, this condition is known as malignant hypercalcemia.

text Bisphosphonates can control uncontrolled bone resorption in oncological patients and, as a result, reduce or remove malignant hypercalcemia. One of the most important effects of bisphosphonates on malignant hypercalcemia is the elimination of intense painful symptoms, typical of this systemic condition.

text Treats underlying treatment of malignant neoplasm make use of several types of medication, including strong antibiotic, analgesic and anti-inflammatory drugs. They also make use of cytotoxic and cytostatic medication that act against malignant cells remaining at the lesion site as well as in other parts of the body, thereby killing them or hindering their proliferation.

text Unfortunately, these medications produce anti-osteogenic side effects that decrease the production of leukocytes, the cells of our immune system. This happens because the bone marrow continuously produces these defense elements at an accelerated pace; however, when in contact with cytotoxic and cytostatic medications, it slows down and strongly impairs patient's immune system. Due to the same reason — low cellular proliferative capacity — regenerative repair processes are compromised.

text Any patient undergoing oncological therapy also receive radiotherapy, especially at the primary source

text explains, for instance, in case of maxillary neoplasm. During treatment, patients are not able to fully react against tumor offending agents, especially microbial ones. Irradiated tissues have even lower capacity, especially maxillary bone tissues which are most susceptible to a particular type of osteomyelitis also known as osteoradionecrosis.

text Nevertheless, many non-irradiated patients also often have osteomyelitis resulting from antimetastatic treatment, given that their mouth is more susceptible to receiving a large amount of different species of microorganisms.

text Today osteomyelitis in patients with malignant neoplasm reflects a condition that has been acknowledged for decades; however, due to well-known frequency and etiology, has been trivialized and under-reported in the literature.

text Antimetastatic treatment protocols including the use of bisphosphonates caught the attention of some clinicians who began to associate osteomyelitis with bisphosphonates side effects.

text Bisphosphonates do not reveal any evidence of susceptibility to osteomyelitis. On the contrary, patient's impaired immune system and tissues with low resectional capacity, as well as the cytostatic and cytotoxic side effects of the drug, essential for antimetastatic therapy, do reveal evidence of susceptibility to osteomyelitis. In cases of normal patients, bisphosphonates have even lower chances of making individuals susceptible to mandibular or maxillary osteomyelitis. Bisphosphonates do not decrease the efficiency of patient's immune system as inflammation and immunologic response do. Furthermore, not only they do not decrease cell proliferation, but also are not cytotoxic for bone cells.

title Does jaw osteomyelitis occur?

text Osteomyelitis occurs almost exclusively under two major clinical conditions:

1) In patients with systemic disorders such as anorexia, uncontrolled diabetes mellitus, leukemia, alcohol consumption, immunodeficiency, malignant neoplasm, among others.

2) In patients with bone diseases and sclerotic jaw lesions; for instance, cases of Foresti Cemento-Ductus Dysplasia and Paget disease.

SATURDAY EVENING

Reading for Women and all the Family

Copyright, 1911, International News Service

By McManus

The Daredevil

All's Well That Ends Well

MECHANICS TRUST COMPANY

PEA COAL

HOTELS, RESTAURANTS and BOARDING HOUSES

3% PAID

ORIENTAL CRAM

Yuletide

THE FEDERAL MATCH SHED

formation of a caretaker grand coalition. It is a sad day indeed for democracy when smart people start pulling for both sides to lose.

...and there's another bad omen. Calabria and the south are conspicuously absent from the national agenda. Only in passing does the region feature in campaign speeches, and there are few premium spots for southerners on the political parties' parliamentary candidate lists. True commitment to solving the problems of the "Mezzogiorno"—an Italy's eight southernmost regions are known—is clearly not considered a vote getter. Yet for reasons that transcend geography, turning around the south ought to be Italy's most pressing national priority. Youth unemployment in the Mezzogiorno is a staggering 36% and between 1991 and 2009, according to one recent study, the Interior Ministry dissolved 154 local city councils in the area because of Mafia infiltration. These conditions have caused a steady exodus of the region's most promising youth to points north and abroad.

...so often, Italian politicians have addressed the south as an isolated regional problem: some say too much public money is frittered away there, while others say the 14 million southerners among Italy's population of 59 million need more support. That's all beside the point, says Domenico Cersosimo, an economics professor at the University of Calabria. "We shouldn't see this as a country divided in two," he says. "The malaise of the Mezzogiorno are the malaise of Italy. It's just a question of degree: what is gray in Italy is black in the south." Indeed, entrenched nationwide ills like tax evasion, cumbersome bureaucracy and a self-serving political class are

...piece with the south's Mafia—crime and blatant corruption. Neither the public nor private sectors have been modernized in Italy, as they have been elsewhere in Europe, explains Fabrizio Barca, a senior Italian Economy Ministry official. "The north has found ways to compensate for this, and can be competitive in spite of the state of country," he says. "It is the north that is the anomaly, not the south. Rome and its ministers operate like the south. Fixing the south means fixing Italy."

...able Man

...SPOCK offers a TINY BLIP OF hope on the otherwise bleak map of Calabria. Current local leaders have pushed to maximize the town's tourist potential and improve living conditions for residents. Beginning in 2001, Mayor Mario Meli, a former union leader, implemented a municipal program under the grand slogan "Amendola wants to be in Italy in Europe, in peace." Funded by 1.5 million a year in local property taxes and 100,000 in revenue from traffic tickets—plus additional grants from Rome and Brussels—the town has offered financial incentives and improved infrastructure to attract private businesses. The mayor's program lured the town's first local bank and four-star hotel, promoted the uncovering of pre-Roman archaeological treasures, and led to the establishment of scuba and sailing schools. Thanks to local efforts, Amendola has managed to renovate the historic city center, open a state-of-the-art physiology center, and keep up environmental efforts like recycling.

...In the strength of those initiatives, Meli was re-elected in 2006 by a 15 percentage point margin over his closest rival. Talk to the locals and you hear the rare sound of southerners pleased with the direction in which their town is headed. One morning in March, Pasquale Salamita, taking a break from his work on a city clean-up crew, gestures toward two new seaside cafes and a disco. "Ten years ago there was almost nothing here," he says. Indeed, the town of 1,000 now seems to strike a nice balance between dynamism and coastal pleasantness, favoring local sports facilities, for instance, over outsize tourist lodgings.

...All, it would be a mistake to get so fixed entirely by the mayor's efforts, or by the good life charms of the town's refurbished 17th century chapel, homemade salami and Mediterranean breeze. For Amendola's residents are still short on opportunities. Meli himself says a lack of industry, large-scale agriculture and sufficient air and highway connections means that poverty and unemployment are bound to persist. "We're not some kind of 'happy island,'" he says. "We've got many of the same problems as the rest of Calabria. Too many young people are packing their bags with their college diploma inside." Indeed, the quiet face that Amendola, like much of Italy, puts on for visitors often hides the nation's great plague: wasted potential.

...ake the street cleaner, Salandria. He has spent the last decade on temporary public works contracts—dreaded "socially useful" jobs by a state welfare

Handwritten mathematical notes on a grid background, covering various topics in mathematics and physics, including calculus, geometry, and trigonometry.

Modellek alkotásai (függvénymodell): a lineáris és az exponenciális növekedés/csökkenés matematikai modelljének összevetése konkrét, valós problémákban (például: népség, energiafelhasználás, járványok).

A logaritmusfüggvények vizsgálata. Logaritmus alapfüggvények grafikonja, jellemzői. A logaritmus függvény mint az exponenciális függvény inverze. Függvénynek és inverzének a grafikonja a koordináta-rendszerben.

A számsorozat fogalma. A függvény értelmezési tartománya a pozitív egész számok halmaza. Matematikáról: Fibonacci. Sorozat megadása rekurzíval és képlettel.

Számtani sorozat, az n. tag, az első n tag összege. Matematikáról: Gauss. A sorozat felismerése, a megfelelő képletek használata problémamegoldás során.

Mértani sorozat, az n. tag, az első n tag összege. A sorozat felismerése, a megfelelő képletek használata problémamegoldás során. A számtani sorozat mint lineáris függvény és a mértani sorozat mint exponenciális függvény összehasonlítása.

Kamatok-számítás. Modelllek alkotása: befektetés és hitel, különböző feltételekkel megírdetett befektetések és hitelek vizsgálata; a hitel költségei, a törlesztés módjai. Korábbi ismeretek mozgósítása (pl. százalékszámítás). A szövegbe többszörösen begyórtott, közvetett módon megfogalmazott információk és kategóriák szomszólása.

Kulcsfogalmak: Szinuszfüggvény, Koszinuszfüggvény, tangensfüggvény, Exponenciális függvény, logaritmusfüggvény, Exponenciális folyamat, Számsorozat, Rekurzió, Számtani sorozat, mértani sorozat.

Triali egység/Felvezetés cél	4. Geometria	Órakeret N: 22 óra E: 10 óra
Eldozetes tudás	Sokszögekkel, körrel kapcsolatos ismeretek. Pont-halmazok, nevezetes pont-halmazok ismerete. Háromszög nevezetes vonalait, pontjait, körét. Háromszög-ek, speciális háromszögekre vonatkozó tételek. Egybevágóság, hasonlóság, szimmetria. Hegyesszögek szögfüggvényei. Ekvivalens egyenlet. Elsőfokú és másodfokú egyenlet, kétszemélyes egyenletrendszer algebrai megoldása. Alapvető trigonometriai, egyenlet- és egyenlőségi feladatok körrel, háromszöggel kapcsolatosan. Vektorok, vektorműveletek. Hasáb, henger, gúla, kúp, gömb felismerése. Felzár, térfogat szemléletes fogalma. Poláris felzár. Számológép (számitógép) használata.	
A tantárgyhoz (műveltségterülethez) kapcsolható fejlesztési feladatok	Tájékozódás a térben. Tájékozódás a világ természeti viszonyában: távolodások, szögek, terület, kerület, felszín és térfogat kiszámítása. A matematika két területének (geometria és algebra) összekapcsolása: koordináta-geometria. Emlékeztés, korábbi ismeretek rendszeresítése, alkalmazása.	
Sziszteréti, koszinusztréti. Alkalmazás eset, különleges eset viszonya (a derékszögű háromszög és a két tétel). Szikidomok kerületének és területének számítása.	Tanulmányozás: Pitagorasz összefüggés egy szög szinusza és koszinusza között. Összefüggés a szög és a mellékszög szinusza, illetve koszinusza között. A tangens kifejezése a szinusz és a koszinusz hányadosaként.	

where $\mathcal{L}[v]$ is the action of the field \mathcal{L} . Using standard (non-rigorous) methods of quantum field theory a number of new and unexpected mathematical results have been derived from topological models, results which in many cases have then been fully proved by more standard mathematical methods, but which would probably not have been discovered without the insights gained from the quantum field theory. (An early appearance of topological invariants in the quantum field theoretic situation is due to Belavin, Polyakov, Schwarz and Tyutin [1]. A more recent example of the powerful application of topological quantum field theory in mathematics may be found in [2], while fuller accounts of earlier work in this field may be found in the books of Nash [3] and Schwarz [4].) Most functional integrals such as (1), and related expressions with operator insertions, have not at present been properly defined. However, since these integrals have such astonishing mathematical power, it seems that an attempt to define these objects rigorously should be more than worth while. In this talk we show how this may be done for the simplest topological model, the topological particle, and describe briefly some recent work by Hrabak [5] which might lead to progress in the canonical quantization of topological field theories.

Some rigorous results on path integrals (that is, functional integrals in quantum mechanics) are known. The basic classical result (which is described by Simon in [6]) for a particle of unit mass moving in one dimension with Hamiltonian

$$H = \frac{1}{2}p^2 + V(x) \quad (2)$$

gives the action of the imaginary time evolution operator $\exp(-HT)$ on a wave function $\psi(x)$ by the formula

$$\exp(-HT)\psi(x) = \int dx \exp\left(-\int_0^T V(x(s))ds\right) \psi(x(t)) \quad (3)$$

where W denotes Wiener measure starting from x_0 and $\mathcal{L}(T)$ are corresponding Brownian paths; the potential V must satisfy certain analytic conditions. The curved space analogue of this result for a Riemannian manifold has been developed by Elworthy [7] and by Ikeda and Watanabe [8]. The expression for evolution according to the Hamiltonian $H = T + V(\mathcal{L})$ where \mathcal{L} is the scalar Laplacian looks identical to (3), but with $\mathcal{L}(T)$ a process depending on metric and connection rather than simply flat space Brownian motion. Tangent space geometry plays an essential part in the theory. The present author has further extended these methods by developing a flat space theory of fermionic path integrals [9] and marrying it with Brownian motion on manifolds to give Brownian motion on supermanifolds in a suitable form for

2



Future Work

Generalized Models

Multimodal Modeling

Future Work

Generalized Models

Multimodal Modeling

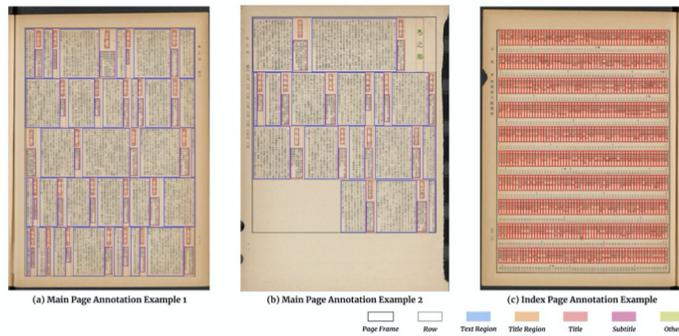


Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

tion over union (IOU) level [0.50:0.95]², on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

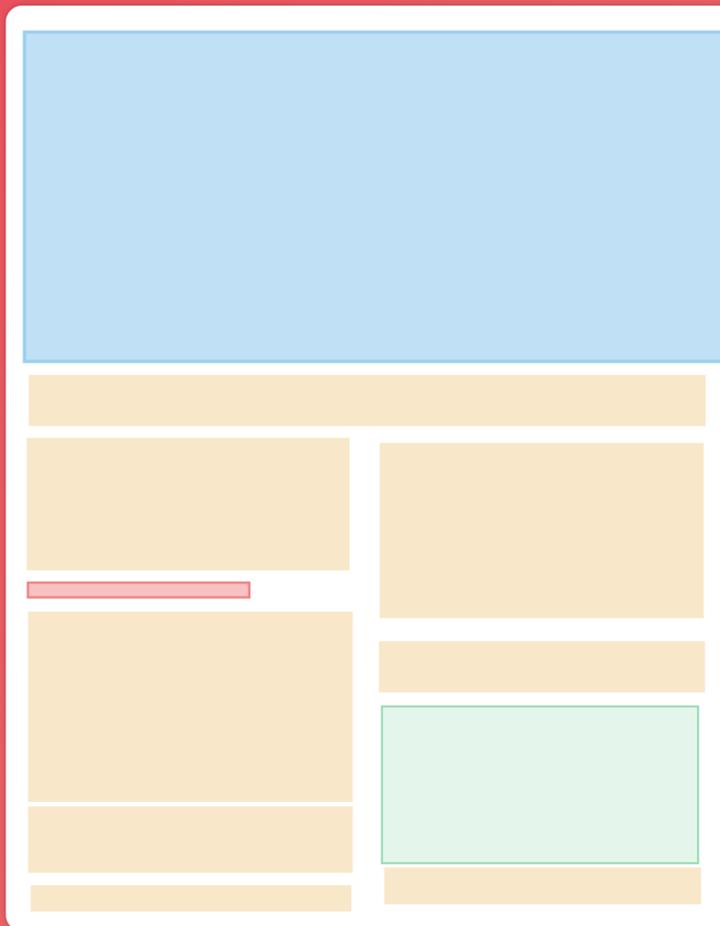
²This is a core metric developed for the COCO competition [12] for evaluating the object detection quality.

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	Faster R-CNN	Mask R-CNN*	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

*For training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.



Layout



Text

Image

Future Work

Generalized Models

Multimodal Modeling

Can we design better ways that model doc image, layout, and text together?



Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

non over union (IOU) level [0.50:0.95]², on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

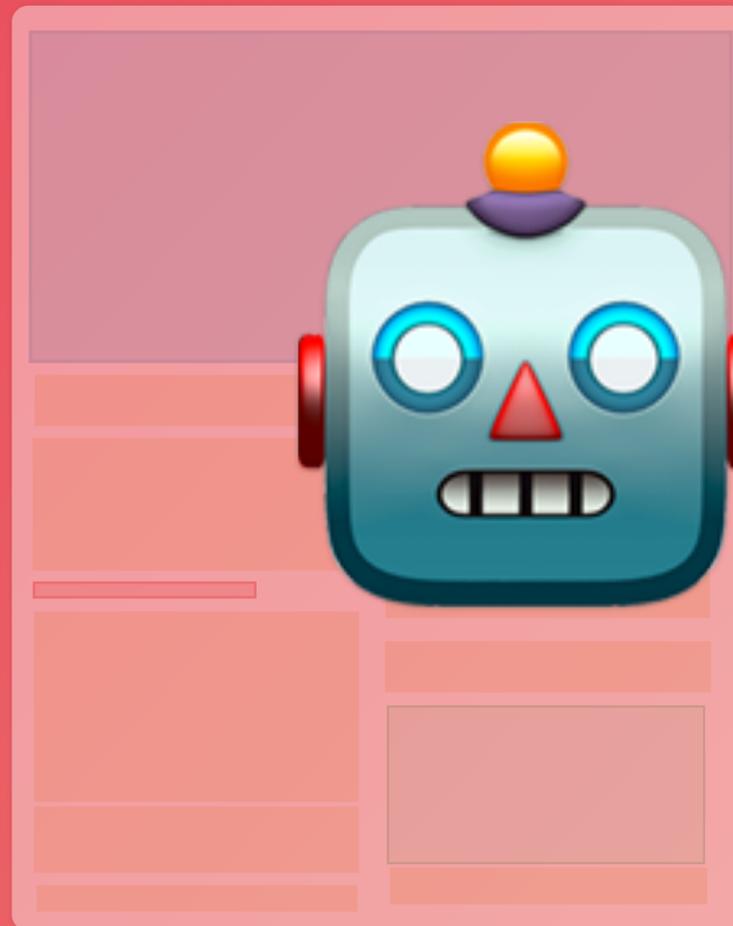
²This is a core metric developed for the COCO competition [12] for evaluating the object detection quality.

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	Faster R-CNN	Mask R-CNN*	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

*For training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.



Layout



Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

non over union (IOU) level [0.50:0.95]², on the test data. In general, the high mAP values indicate accurate detection of the layout elements. The Faster R-CNN and Mask R-CNN achieve comparable results, better than RetinaNet. Noticeably, the detections for small blocks like title are less precise, and the accuracy drops sharply for the title category. In Figure 8, (a) and (b) illustrate the accurate prediction results of the Faster R-CNN model.

5.2. Pre-training for other datasets

We also examine how our dataset can help with a real-world document digitization application. When digitizing new publications, researchers usually do not generate large scale ground truth data to train their layout analysis models. If they are able to adapt our dataset, or models trained on our dataset, to develop models on their data, they can build their pipelines more efficiently and develop more accurate models. To this end, we conduct two experiments. First we examine how layout analysis models trained on the main pages can be used for understanding index pages. Moreover, we study how the pre-trained models perform on other historical Japanese documents.

Table 4 compares the performance of five Faster R-CNN models that are trained differently on index pages. If the model loads pre-trained weights from HJDataset, it includes information learned from main pages. Models trained over

²This is a core metric developed for the COCO competition [12] for evaluating the object detection quality.

all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models trained on more data perform better than those with fewer samples. We also directly use the model trained on main to predict index pages without fine-tuning. The low zero-shot prediction accuracy indicates the dissimilarity between index and main pages. The large increase in mAP from 0.344 to 0.471 after the model is

Table 3: Detection mAP @ IOU [0.50:0.95] of different models for each category on the test set. All values are given as percentages.

Category	Faster R-CNN	Mask R-CNN*	RetinaNet
Page Frame	99.046	99.097	99.038
Row	98.831	98.482	95.067
Title Region	87.571	89.483	69.593
Text Region	94.463	86.798	89.531
Title	65.908	71.517	72.566
Subtitle	84.093	84.174	85.865
Other	44.023	39.849	14.371
mAP	81.991	81.343	75.223

*For training Mask R-CNN, the segmentation masks are the quadrilateral regions for each block. Compared to the rectangular bounding boxes, they delineate the text region more accurately.

Image

Text

Future Work

Generalized Models

Multimodal Modeling



MMDA

Multi Model Document Analysis

Our Contributions



Layout Parser



Our Contributions **A unified DLA toolkit**

 *Layout Parser*

A unified DfIA toolkit

Our Contributions

Open the box usage

 *Layout Parser*

A unified DfA toolkit

Open the box usage

Our Contributions

Deep learning integration

 *Layout Parser*

A unified DfA toolkit

Open the box usage

Deep learning integration

Our Contributions **Simple APIs + customization**

 *Layout Parser*

A unified DfA toolkit

Open the box usage

Deep learning integration

Simple APIs + customization

Our Contributions

Open platform & community

LP *Layout Parser*

Community & Discussion

The screenshot shows the homepage of the Layout Parser website. At the top, it says "Layout Parser" and "Layout Analysis - in 4 Lines of Code". Below that, it says "Transform document image analysis pipelines with the full power of Deep Learning." There is a code block showing "pip install layoutparser". The main content area is titled "What is Layout Parser?" and "A Unified Toolkit for Deep Learning Based Document Image Analysis". It lists features like "Accurate Layout Detection with a Simple and Clean Interface" and "With the help of state-of-the-art deep learning models, Layout Parser enables extracting complicated document structures using only several lines of code." There are also images of document layouts and a code snippet.

The screenshot shows the GitHub repository page for "Layout-Parser / layout-parser". It displays the repository statistics: 2.4k stars, 204 forks, and 20 issues. The "Issues" tab is selected, showing a list of 20 open issues. The issues listed include "Paragraph and titles with bigger line spacing", "Layout Parser text boxes not properly aligned causing incorrect sorting of text boxes", "Multi modal approach to LP's Deep Layout Parsing capability", "group_blocks_by_distance example addition", "lp.draw_box() is not able to displays the result in .py file", "enforce_cpu not working", and "lp.Detectron2LayoutModel often hangs on download step (config/model)".

The screenshot shows the Twitter profile for @layoutparser. It displays 11 tweets. The most recent tweet is from April 12, 2020, and says: "(2/6) We incorporate a new model from the Tablebank dataset and now the model can detect table regions on various documents." It includes three images showing table detection results. Another tweet from April 12, 2020, says: "(1/n) Layout Parser v0.2 is out! New models, better API support, and much more!" It includes a list of highlights: "Add support for loading and saving with JSON and CSV", "New shape operations between blocks (union and intersection) are available", and "Table detection models are up for grabs!".

The screenshot shows the Slack channel #general for Layout-Parser. It displays a list of messages. The most recent message is from Zejiang (Shannon) Shen, dated Thursday, April 8th, 2020, and says: "Hi everyone! Welcome to join the Layout-Parser slack channel! Please feel free share your thoughts about this library or just discuss document digitization in general! In case you encounter a bug, let us know in #bug-report channel or just submit a Github issue. Please stay tuned for more incoming updates in the near future!" There are also messages from other users, including one from a user named "KBLabb" who says: "This looks very promising and timing could not be better for us! Awesome! /KBLabb (National Library of Sweden)".



Website
layout-parser.github.io



Github
@layout-parser



Twitter
@layoutparser



Slack
layout-parser.slack.com

* No "-" for twitter account name, as it disallows...

Open-the-box Usage

Modularized Design

Open Sharing Platform

Layout Visualization

Deep Learning Integration

DL Models Customization

OCR Engine Support

Data Import and Export

Layout Data Annotation

Multi-backend support

Modeling Tutorials

*Commandline Tools**

Layout Parser



Website

layout-parser.github.io



Github

[@layout-parser](https://github.com/layout-parser)



Twitter

[@layoutparser](https://twitter.com/layoutparser)



Slack

layout-parser.slack.com



Zejiang Shen
[@ZejiangS](https://twitter.com/ZejiangS)



Ruochen Zhang
[@ruochenz_](https://twitter.com/ruochenz)



Melissa Dell
[@MelissaLDell](https://twitter.com/MelissaLDell)



Benjamin Lee
[@lee_bcg](https://twitter.com/lee_bcg)



Jacob Carlson
[@J_S_Carlson](https://twitter.com/J_S_Carlson)



Weining Li
[@WeiningLiCA](https://twitter.com/WeiningLiCA)