

# *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***

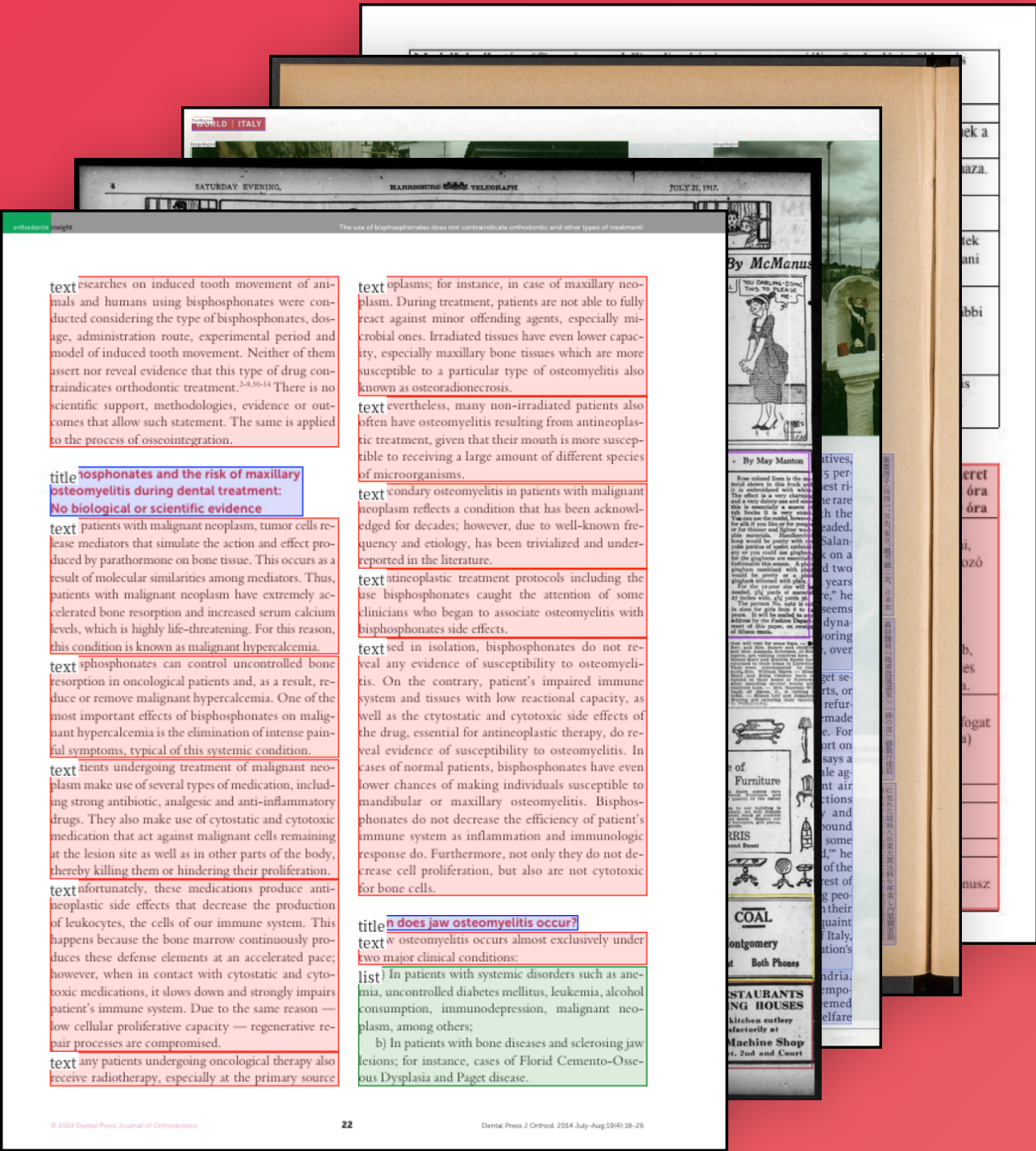


# Deep Learning for Document Image Analysis

The Task

Exciting Progress

Challenges



```
{  
  
  "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

Output: Layout/Text

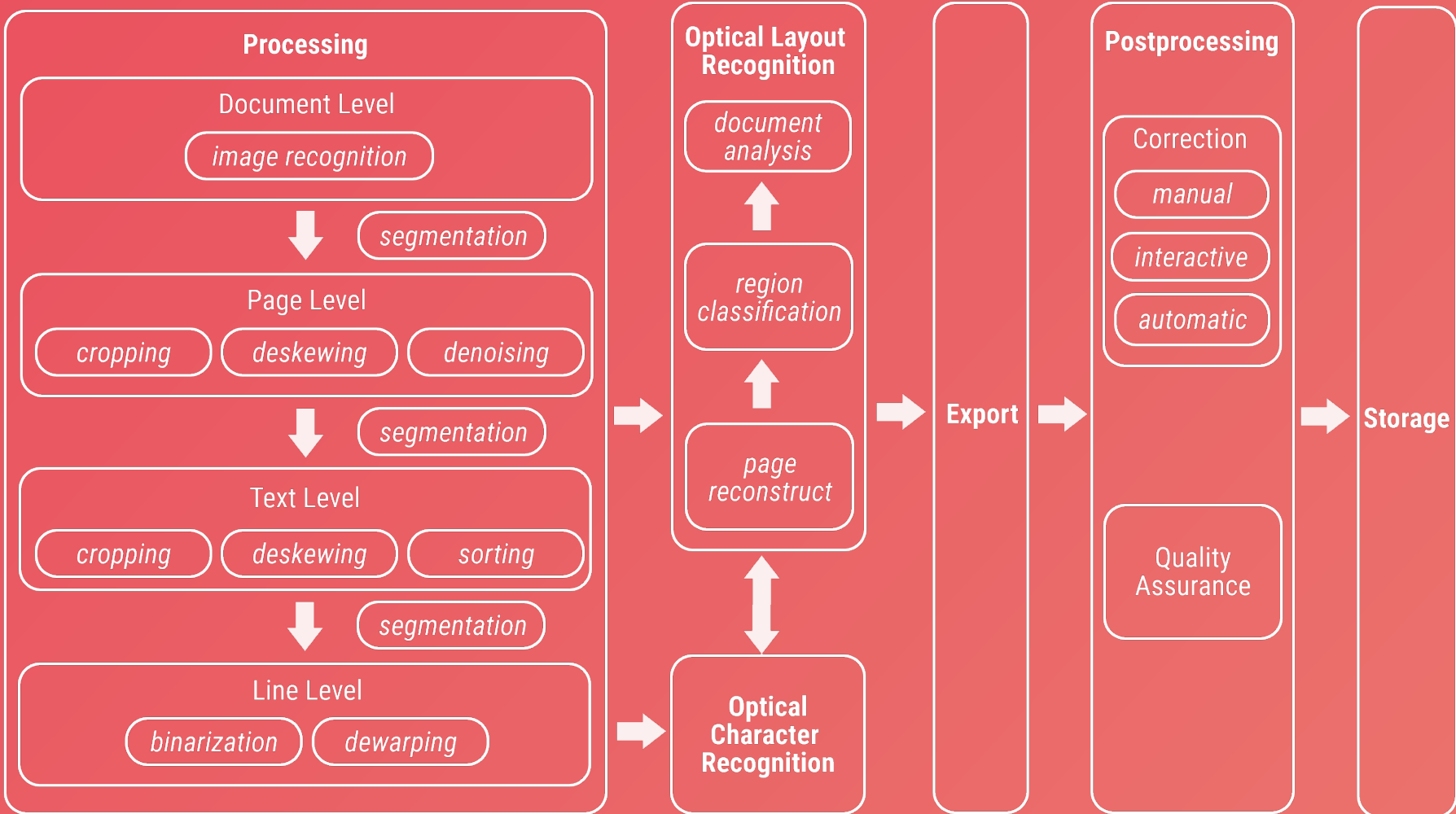
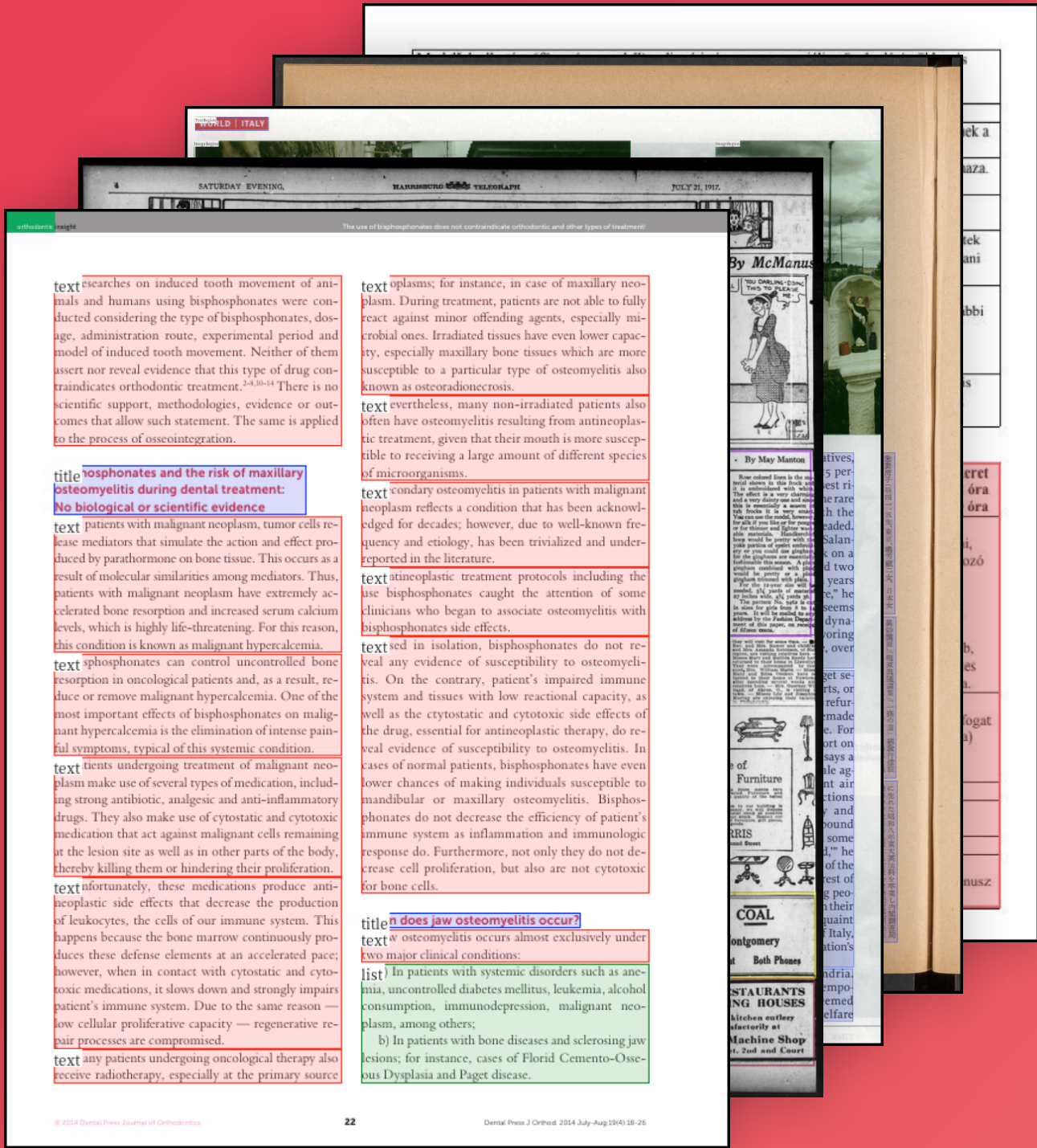


# Deep Learning for Document Image Analysis

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

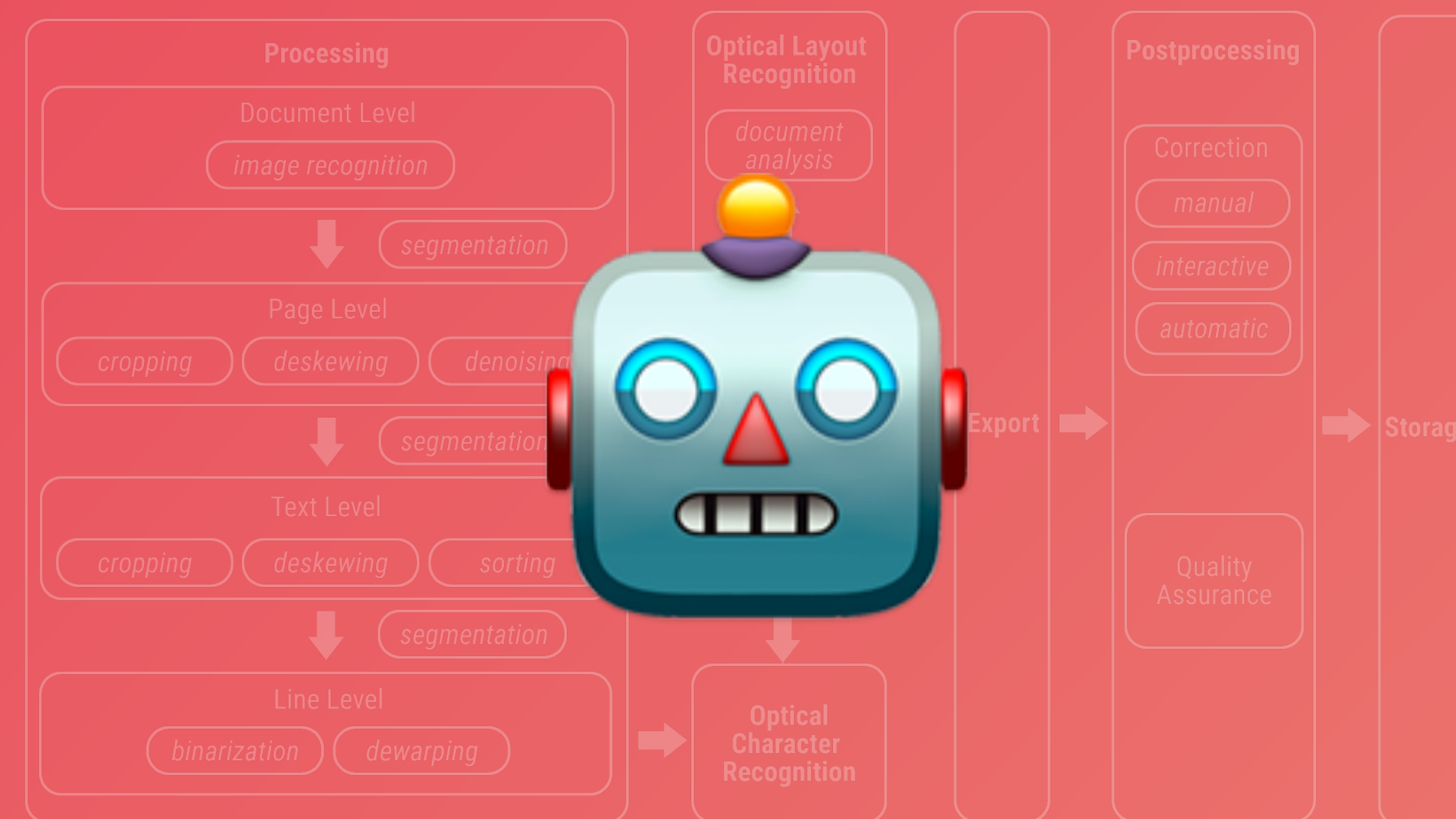
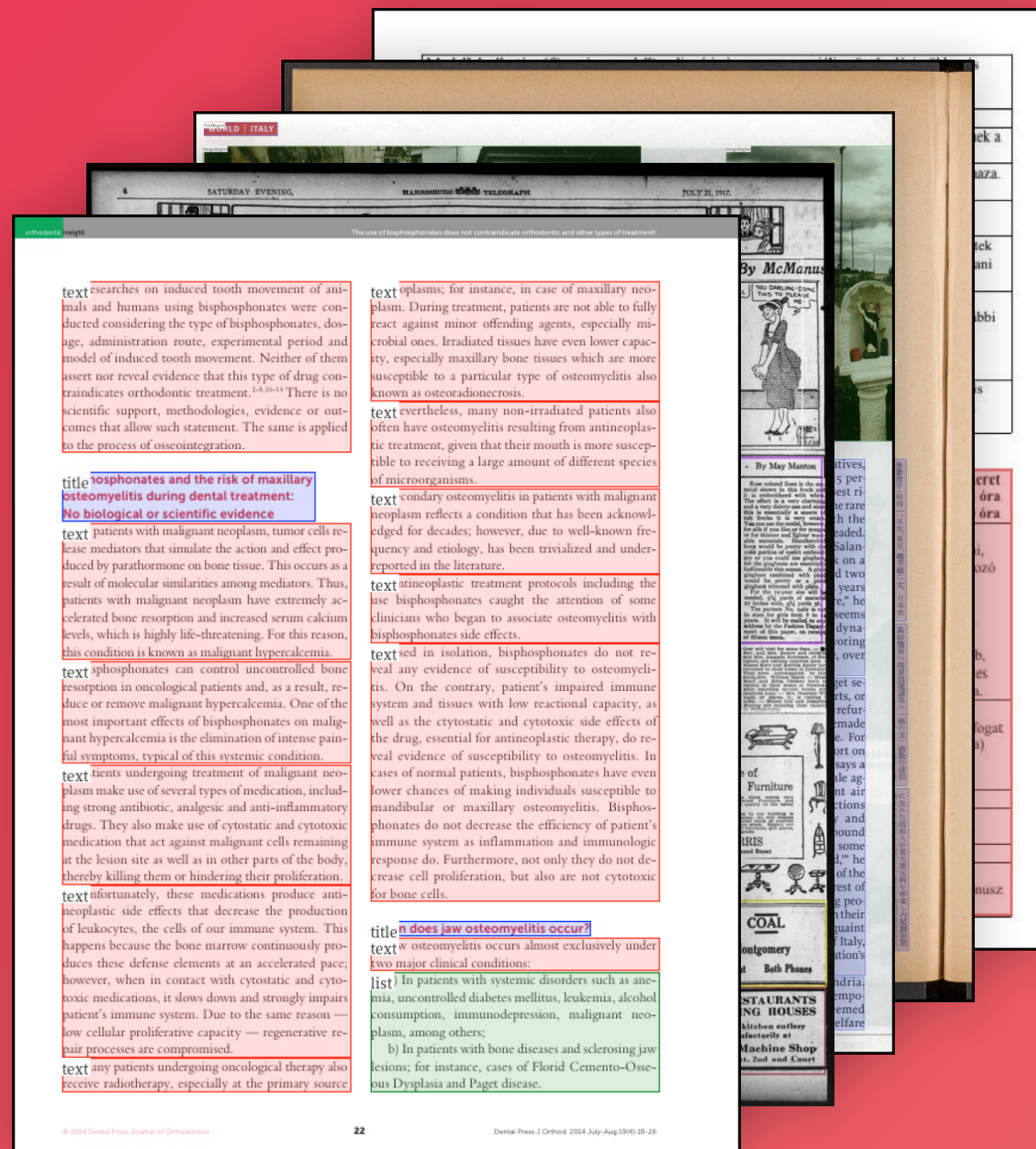


# Deep Learning for Document Image Analysis

# 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

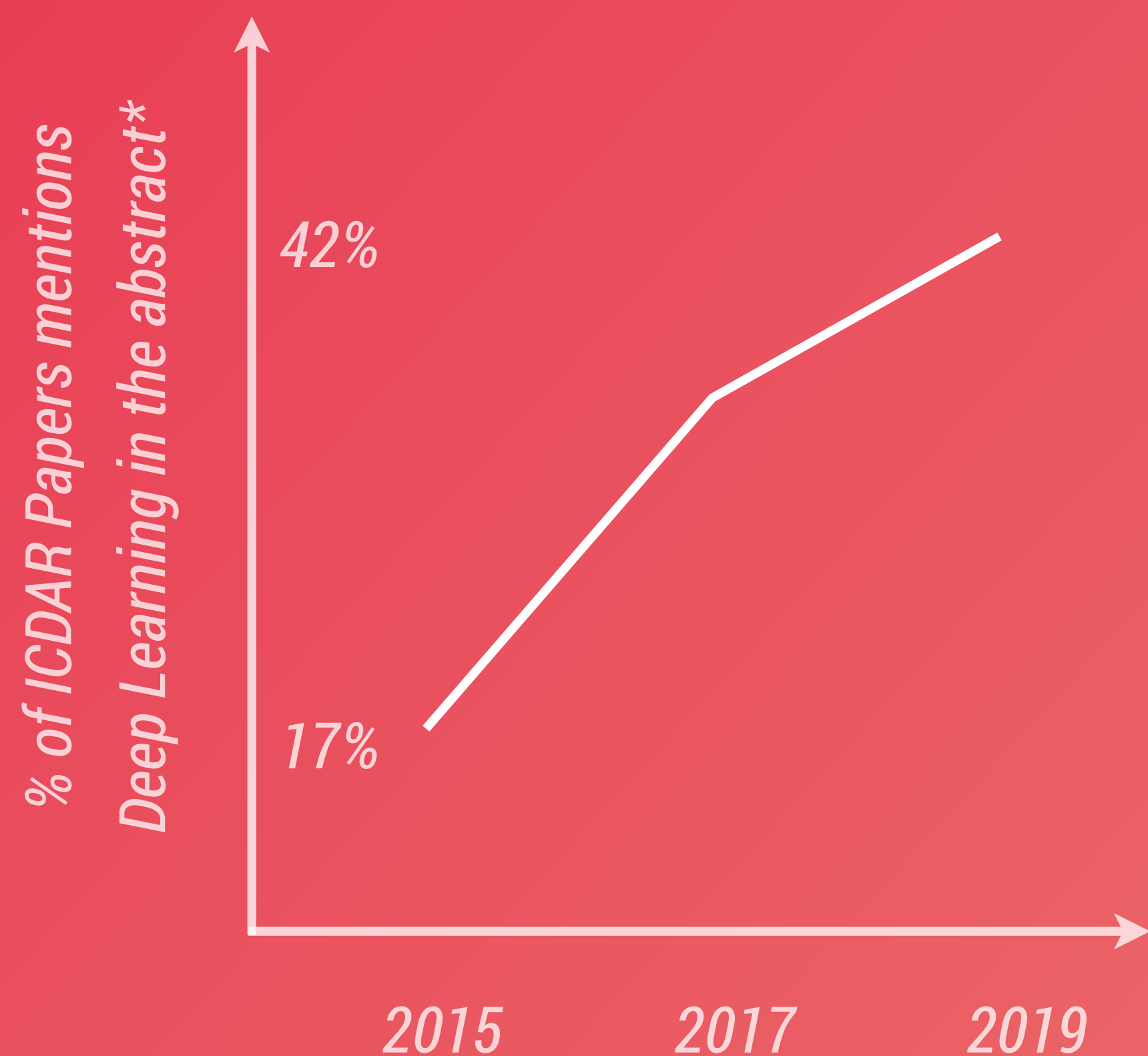


# Deep Learning for Document Image Analysis

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.*



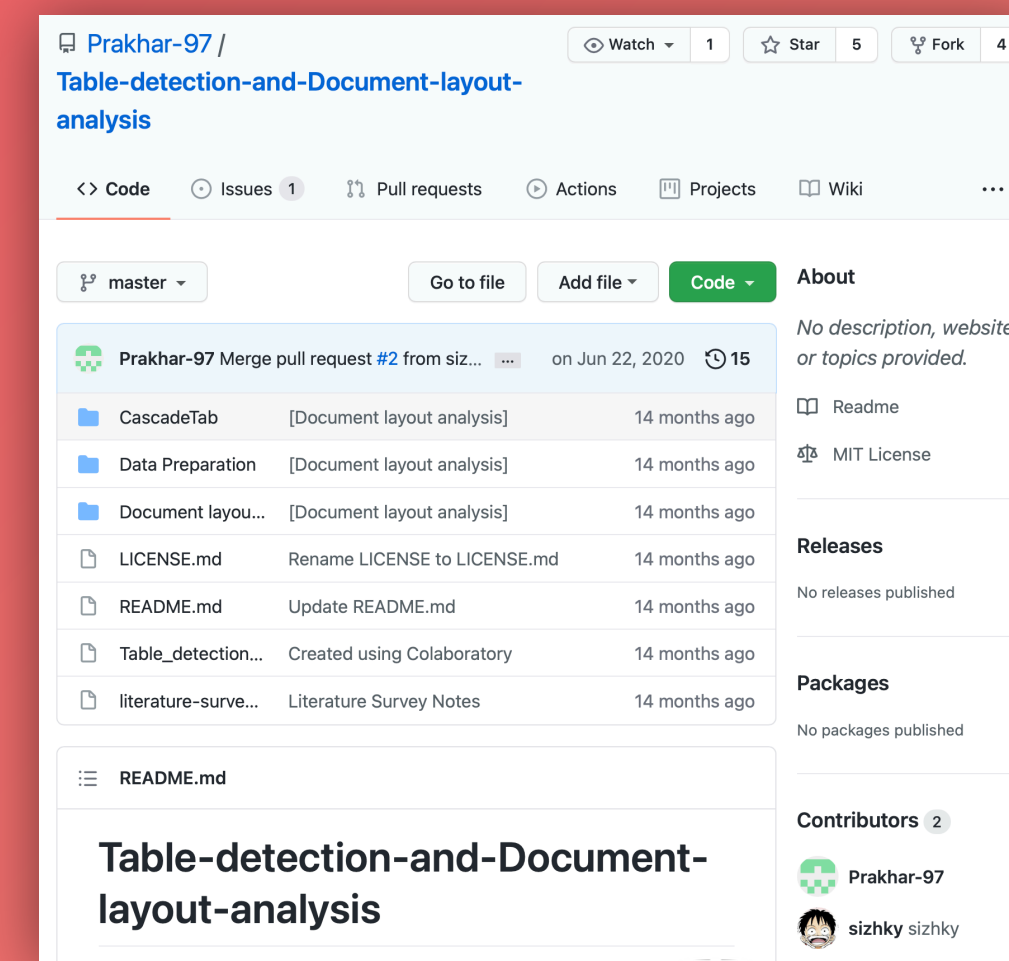
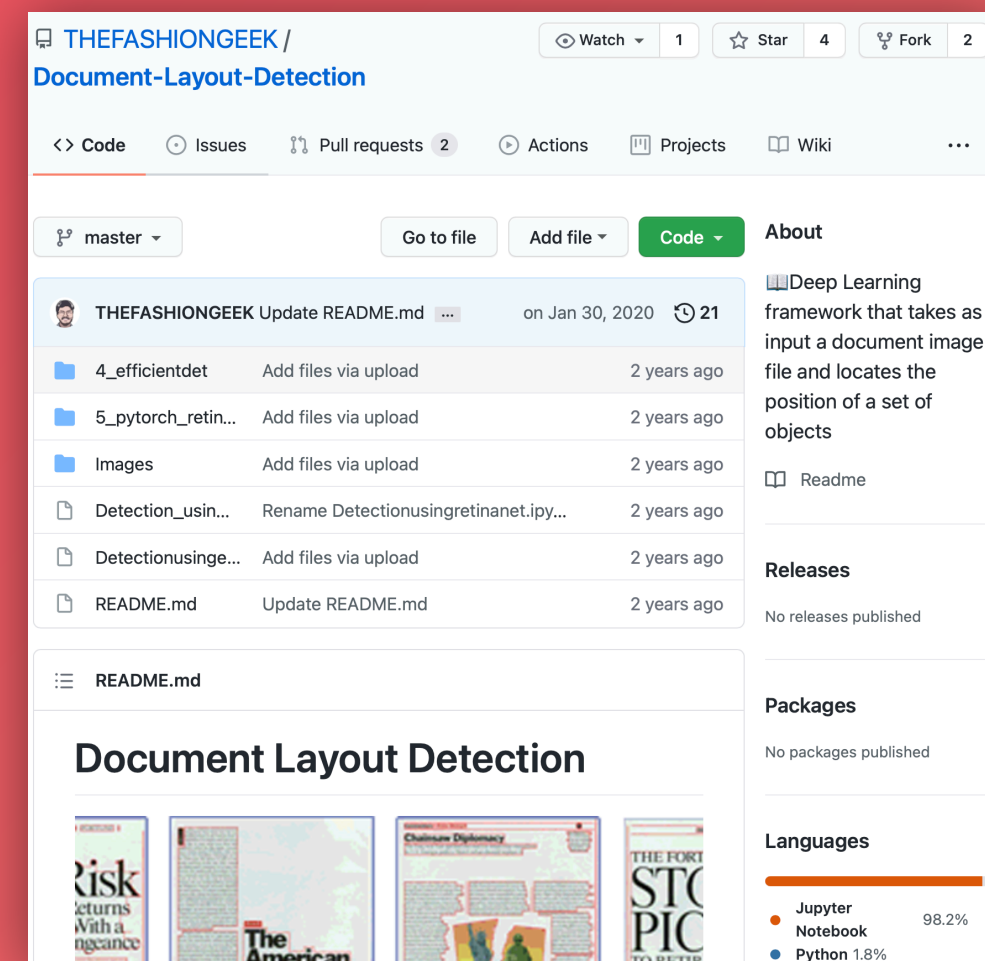
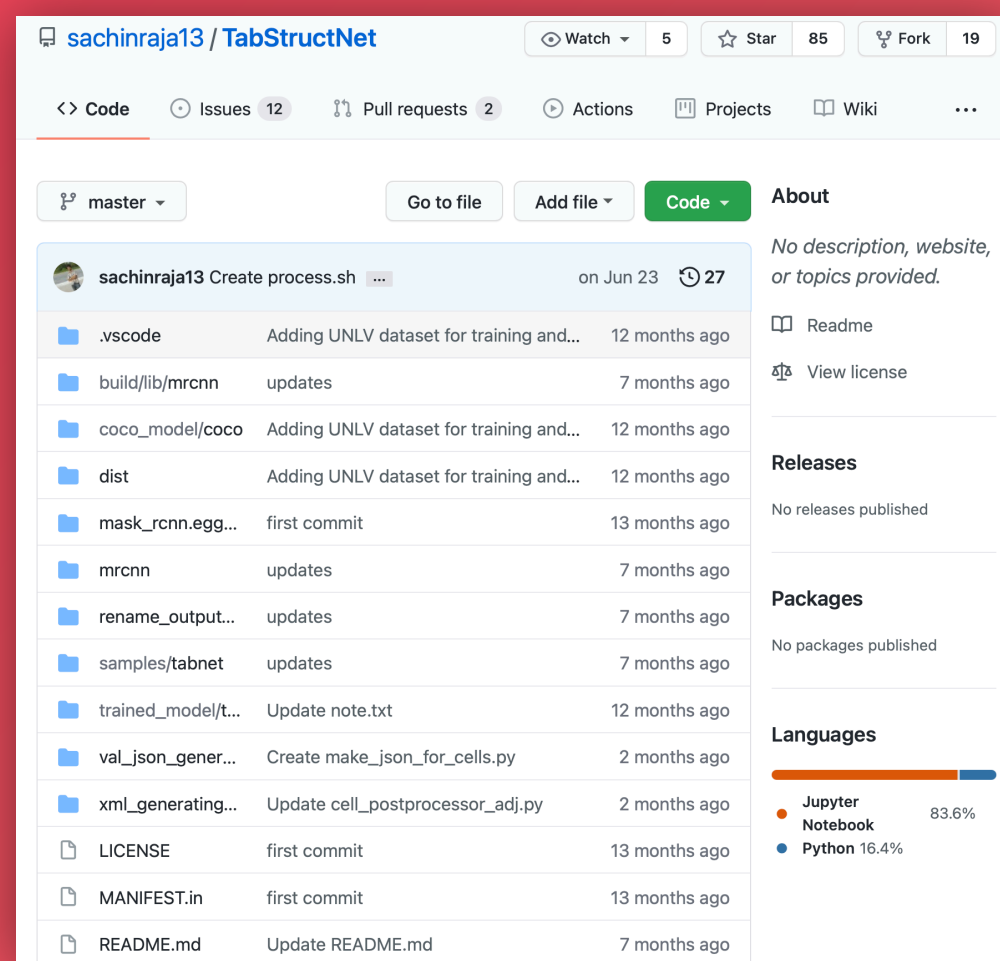
# Deep Learning for Document Image Analysis

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

*Preprocessing*

*Github Repo A*

*Java | Environment A*

*Layout Detection*

*Github Repo B*

*Python | Environment B*

*Character Recognition*

*Tesseract*

*C++/bash*

*Postprocessing*

*Github Repo C*

*MATLAB*

*Storage*

*PAGE Exporter*

*C++*



# 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

*Preprocessing*

*Github Repo A*

*Java | Environment A*

*Layout Detection*

*Github Repo B*

*Python | Environment B*

*Character Recognition*

*Tesseract*

*C++/bash*

*Postprocessing*

*Github Repo C*

*MATLAB*

*Storage*

*PAGE Exporter*

*C++*



# 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)
```

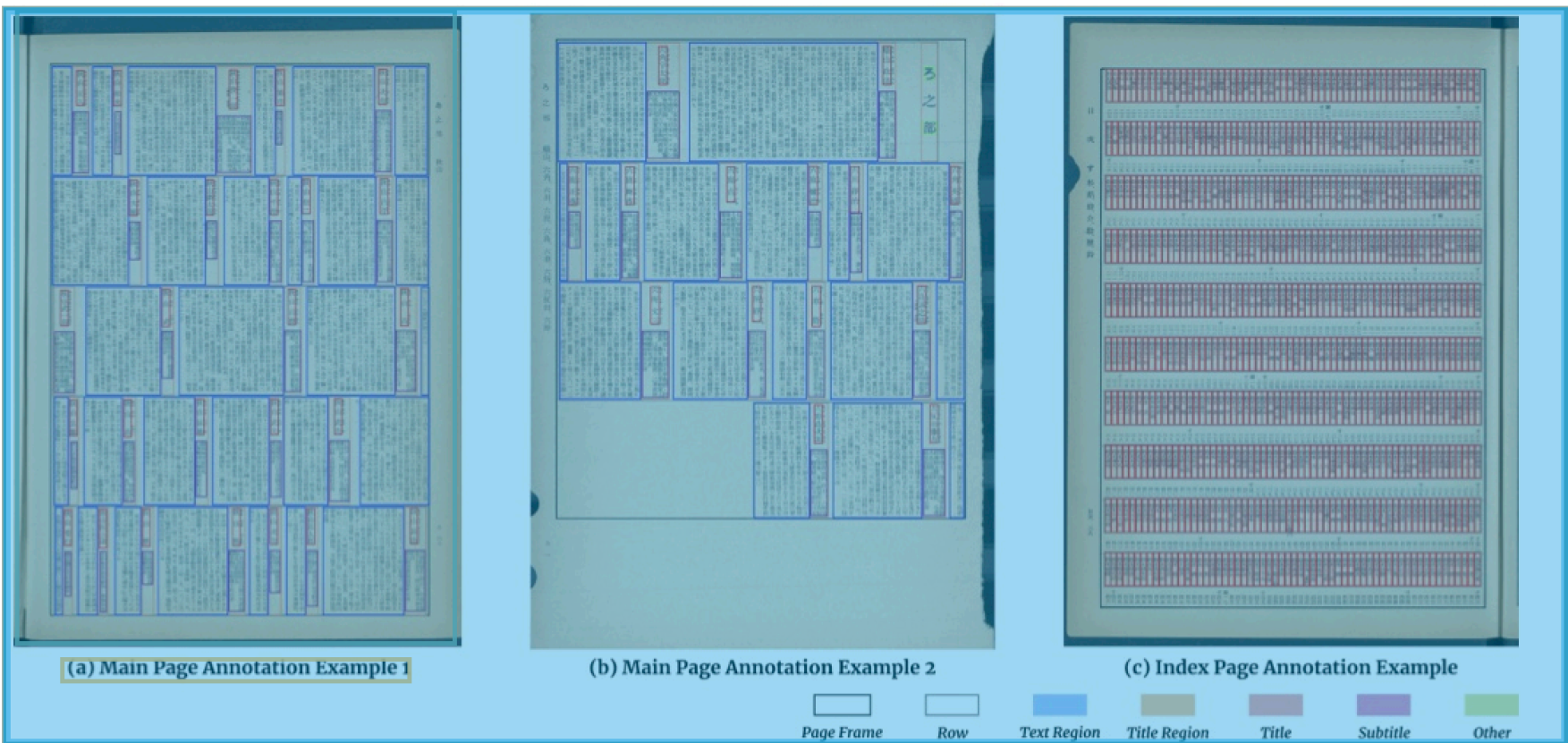


Figure 7: **Annotation Examples in HJDataset.** (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

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

<sup>2</sup>This 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 <sup>a</sup>	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

<sup>a</sup> 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 usage example

Installation

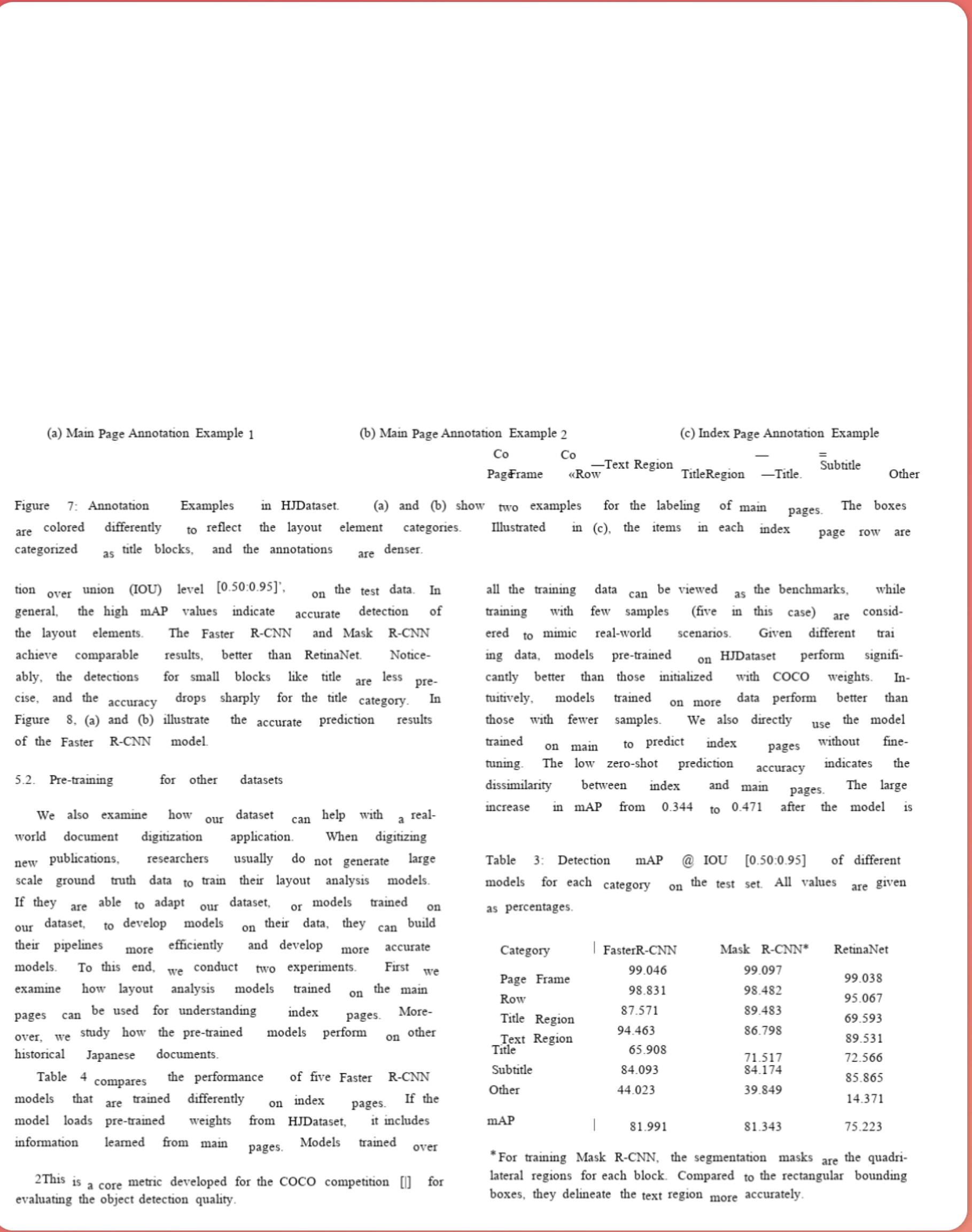
Layout Detection

Optical Character

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

>>> ocr_text =
ocr_model.detect(image)

>>> ocr_text.to_json()
```



*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*  
**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**  
*OCR Engine Support*

*Layout Visualization*

*Data Import and Export*



*Layout Data Annotation*  
**Deep Learning**  
**Model**  
*DL Models Training*  
**Customization**  
*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 Visualization*

*Layout Data Structure*  
**Core APIs**  
*OCR Engine Support*

*Data Import and Export*

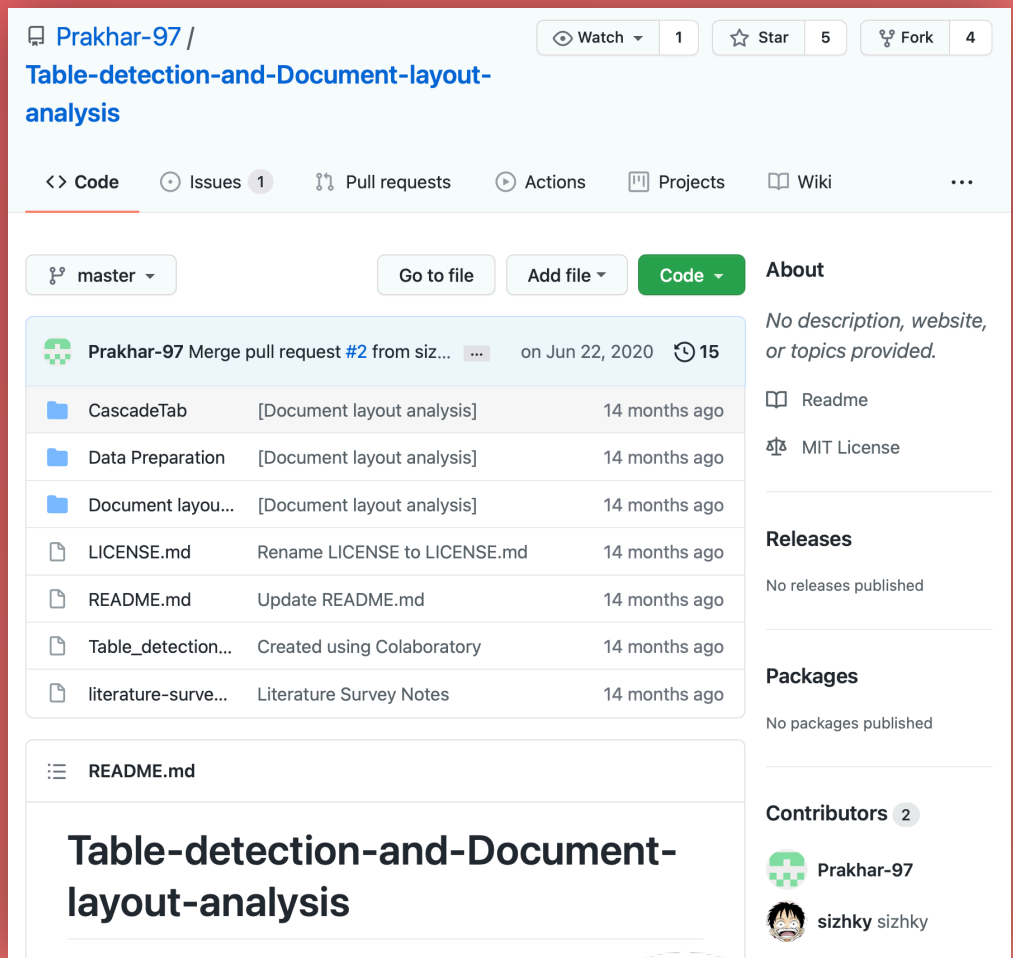
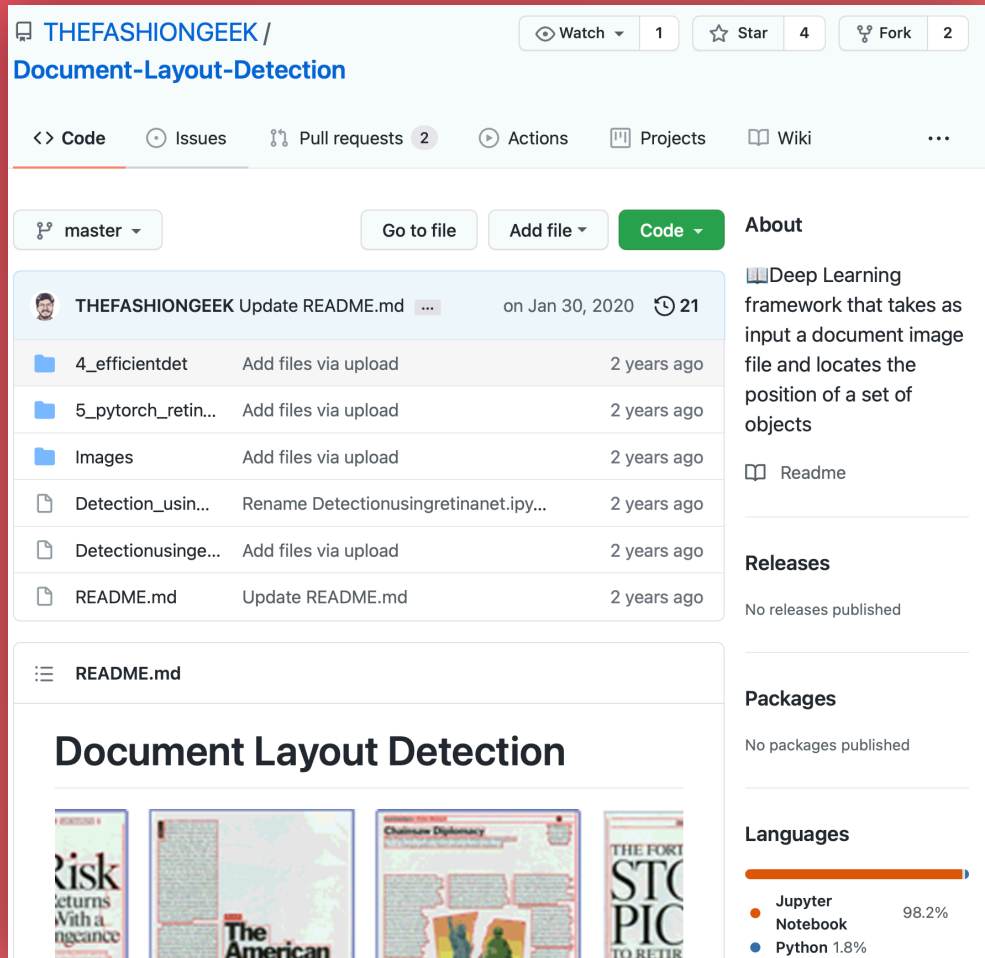
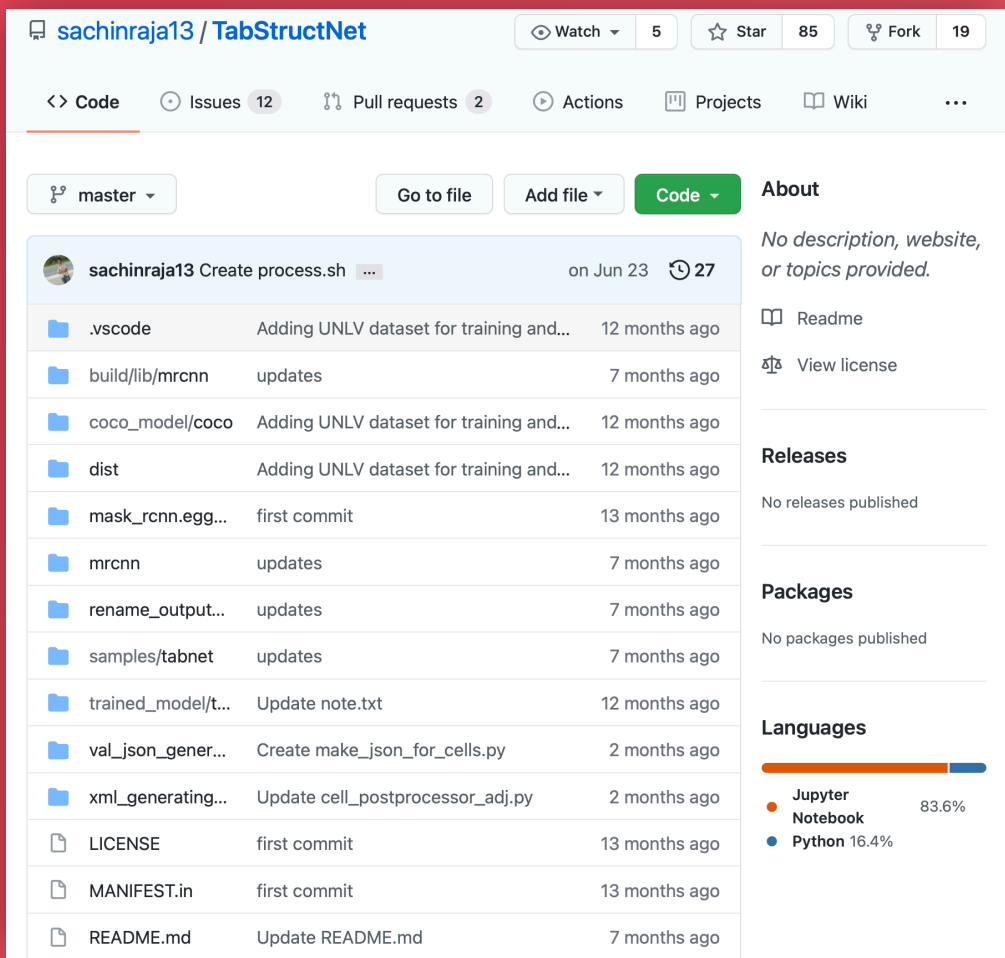
# Deep Learning Models for Layout Detection

Challenges

Standardized Modeling API

Model Zoo

No standard way for sharing and re-using existing models



PyTorch

TensorFlow

mxnet



# 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/pppyolov2_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")
```



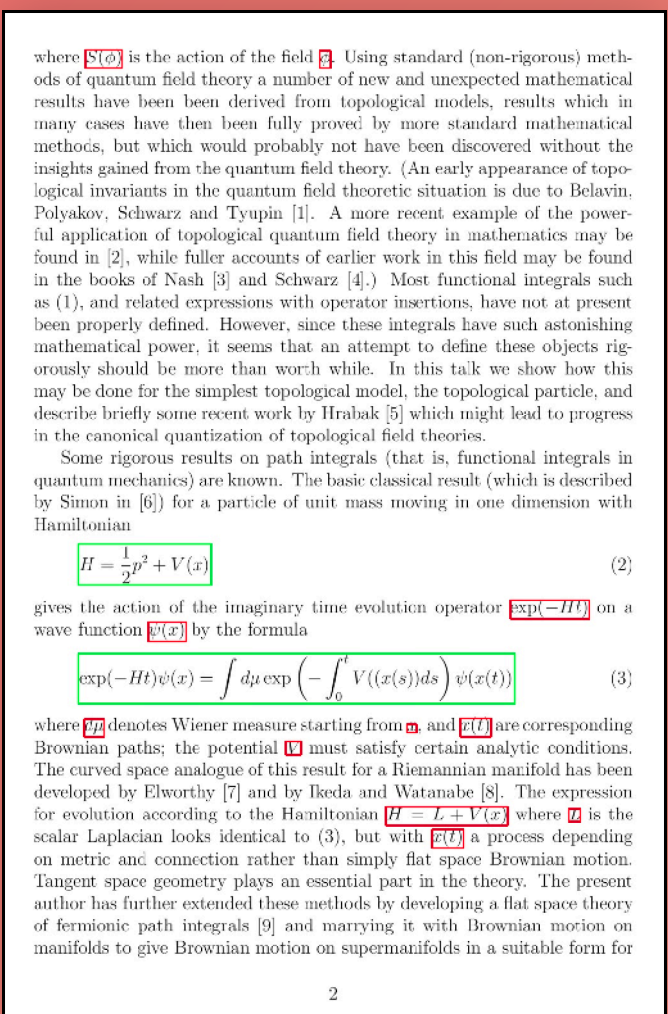
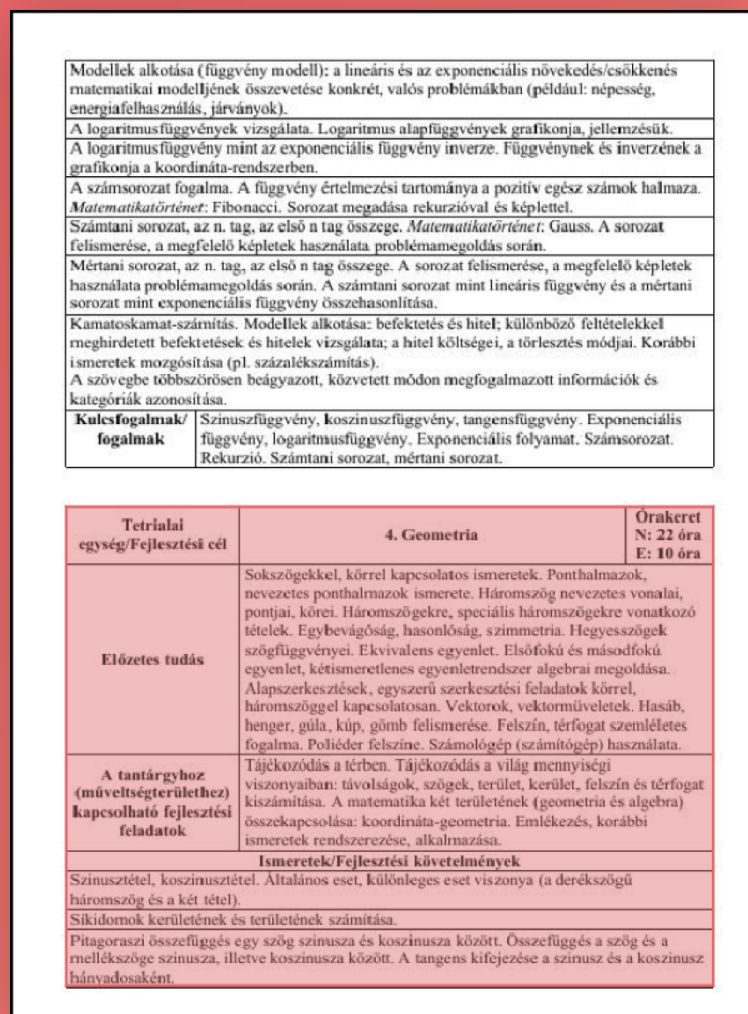
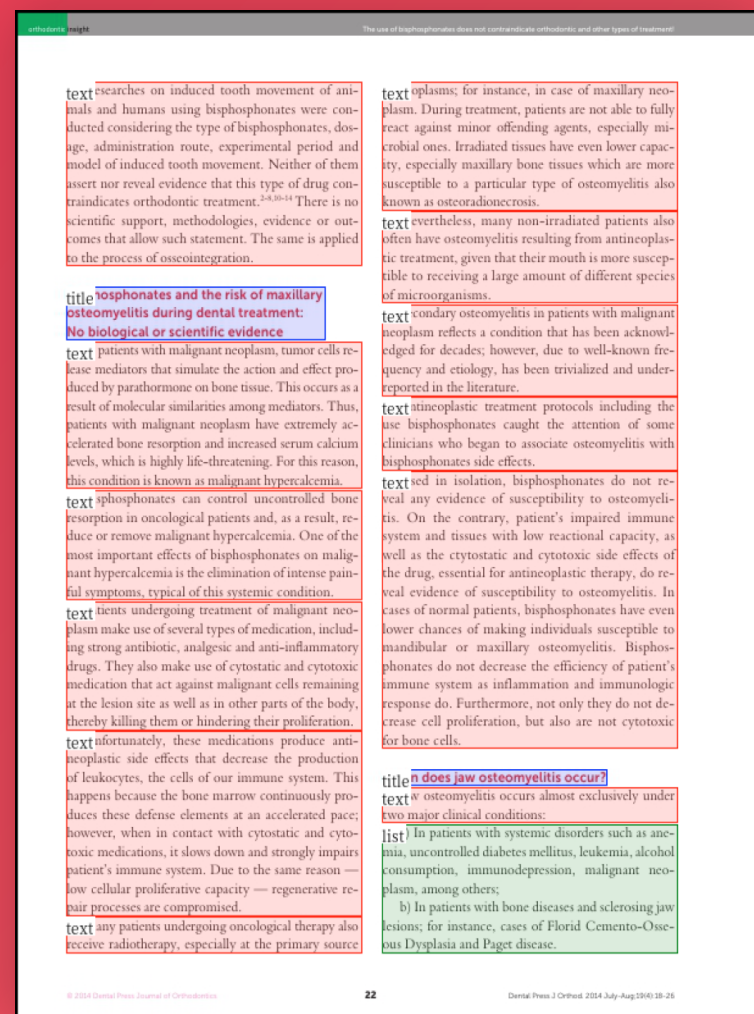
# Deep Learning Models for Layout Detection

# 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**  
**Model**  
*DL Models Training*  
**Customization**  
*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*



# 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")
```

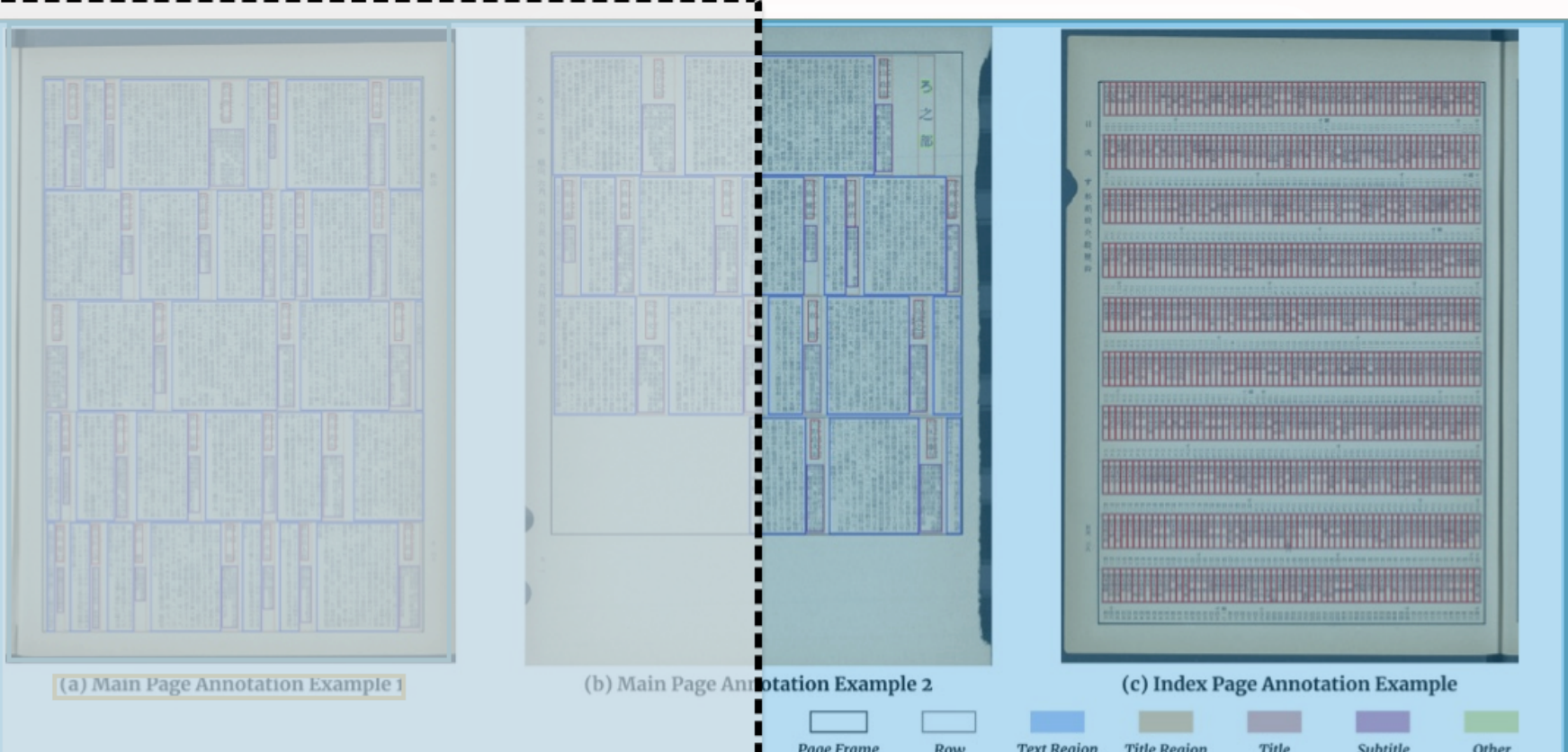


Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

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

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 <sup>a</sup>	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

<sup>a</sup> 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

Select layout regions in the left column:

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

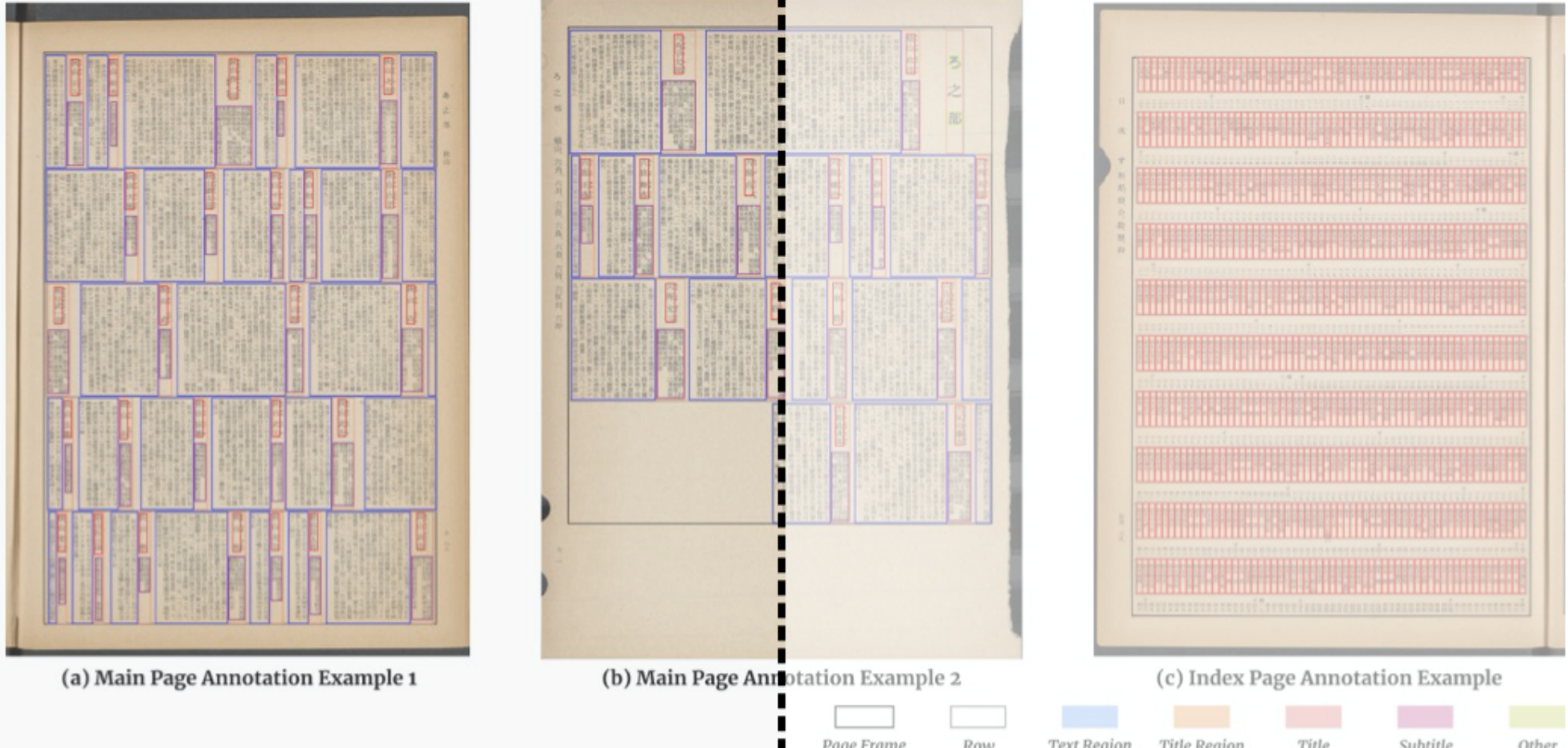


Figure 7: **Annotation Examples in HJDataset.** (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

tion over union (IOU) level [0.50:0.95]<sup>2</sup>, 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

<sup>2</sup>This 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 <sup>a</sup>	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

<sup>a</sup> 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

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

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



# Layout Parser Infrastructure APIs

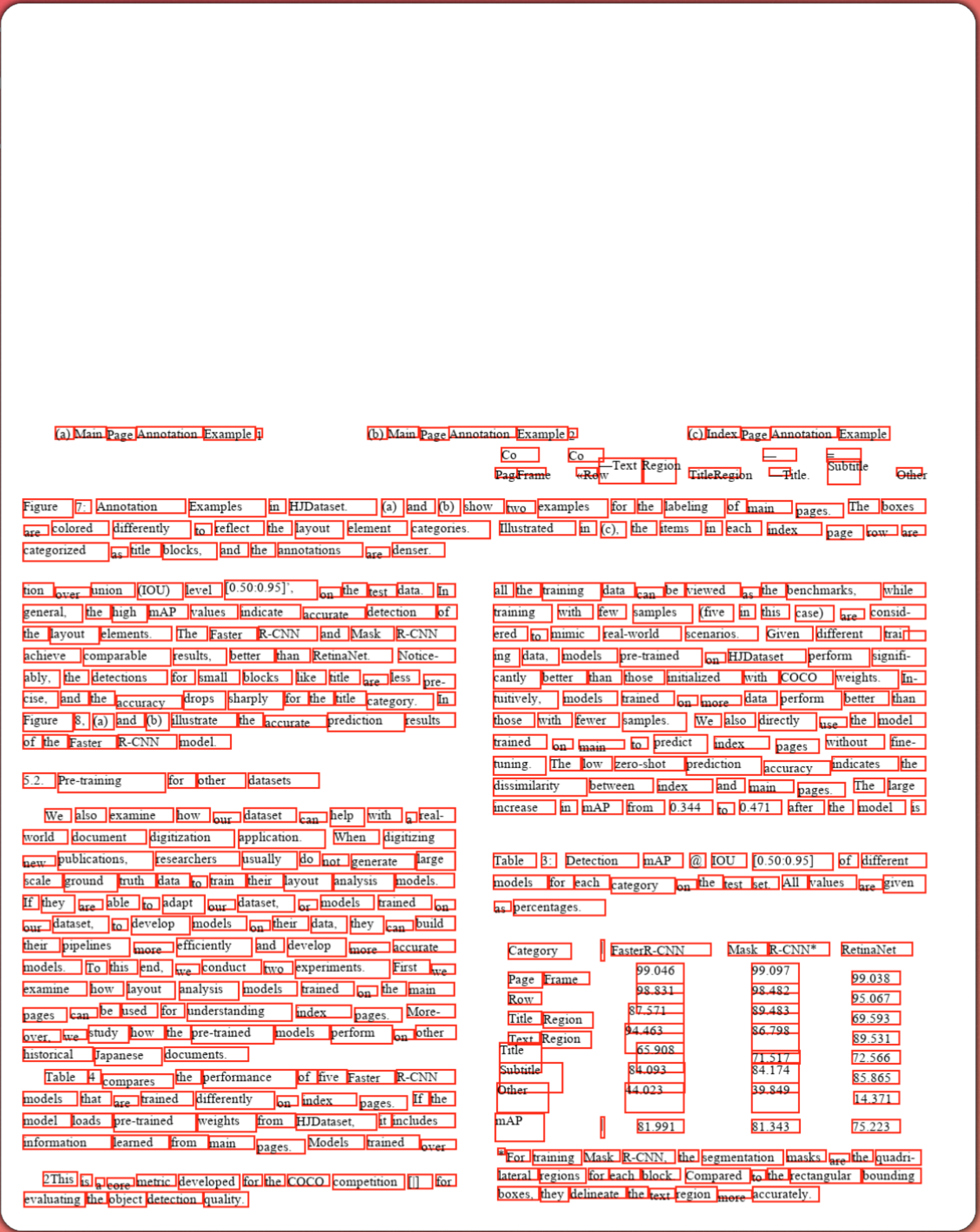
Layout Data Structures

OCR Engine Support

Layout Visualization

## 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, ...)
```



# 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

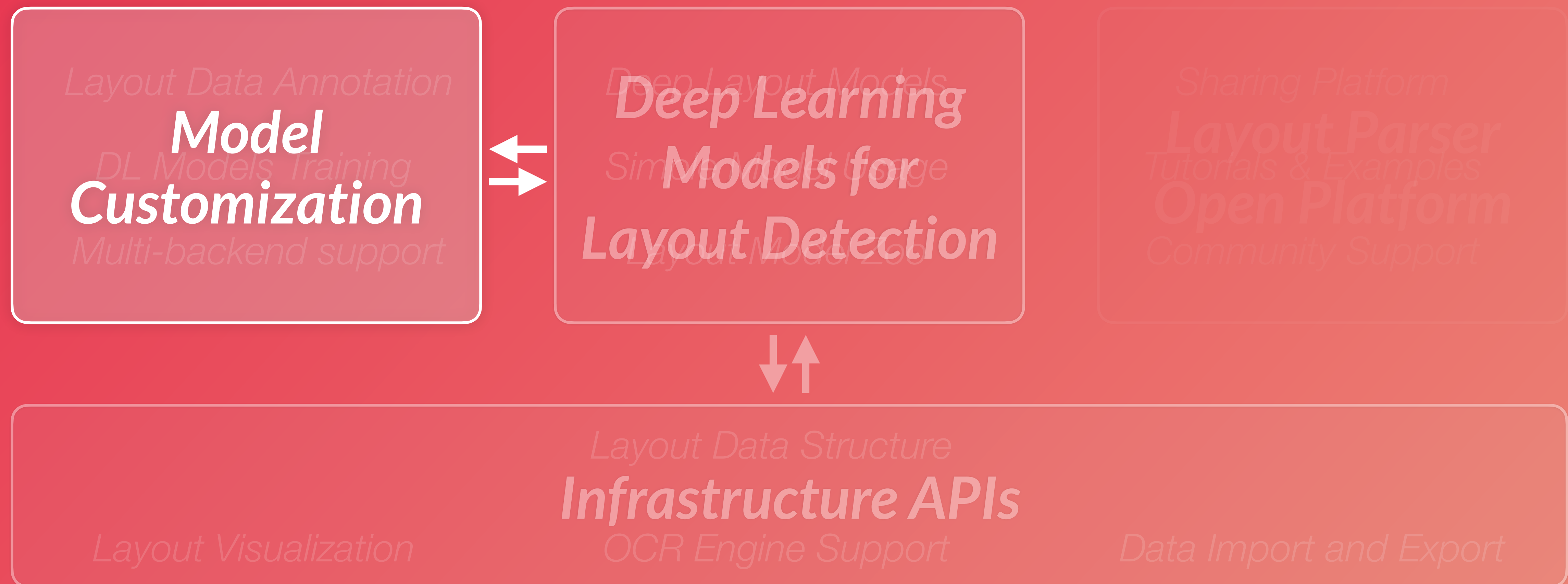
**CSV**      **JSON**

## More formats will be added:

*PAGES*    *METS/ALTO*    *hOCR*    ...



*What if we want better models?*





# Deep Learning Model Customization

Why

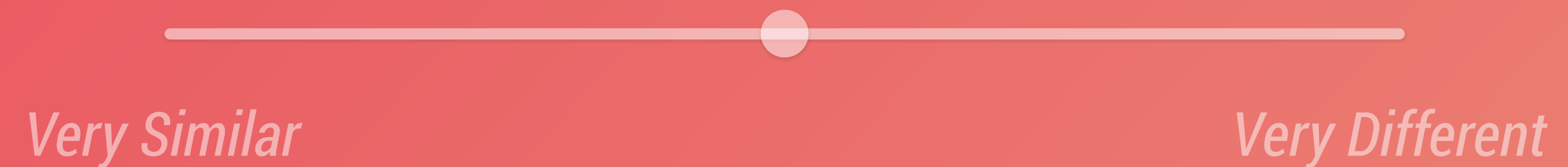
How different is the target data?



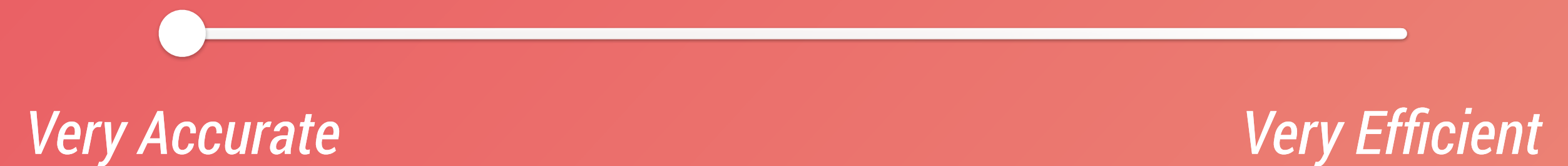
# Deep Learning Model Customization

Why

How different is the target data?



Accuracy/efficiency requirements?

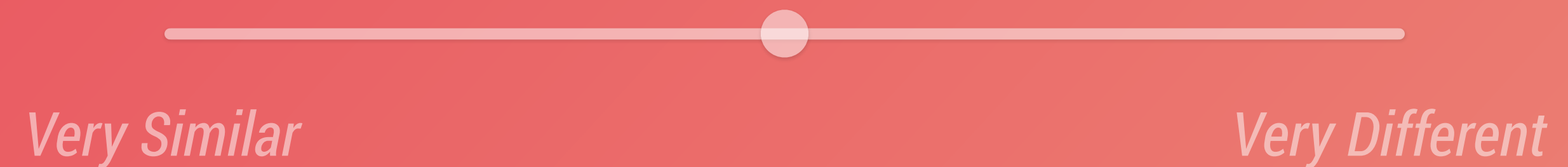




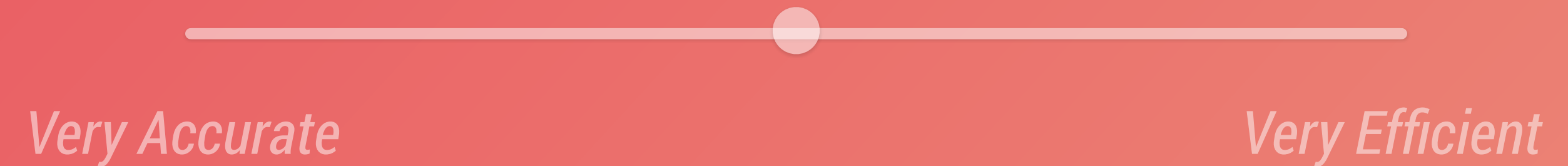
# Deep Learning Model Customization

Why

How different is the target data?



Accuracy/efficiency requirements?

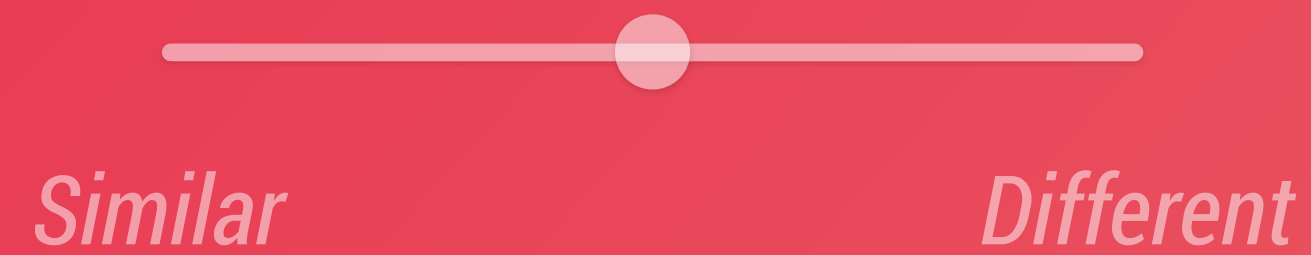


How much training data is available?

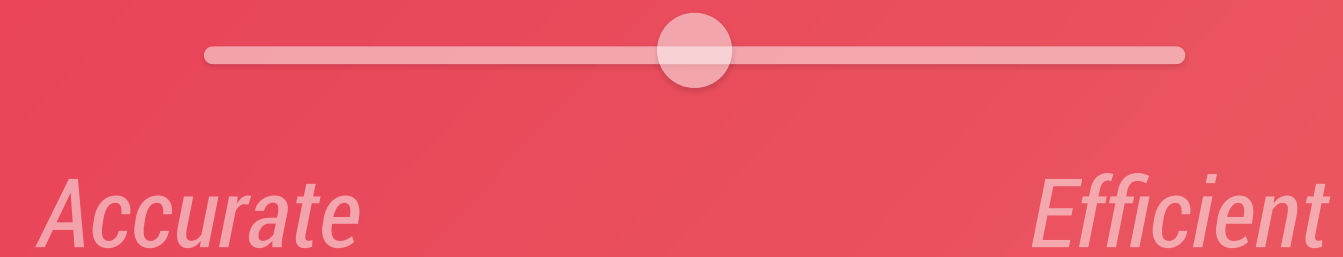


# Deep Learning Model Customization

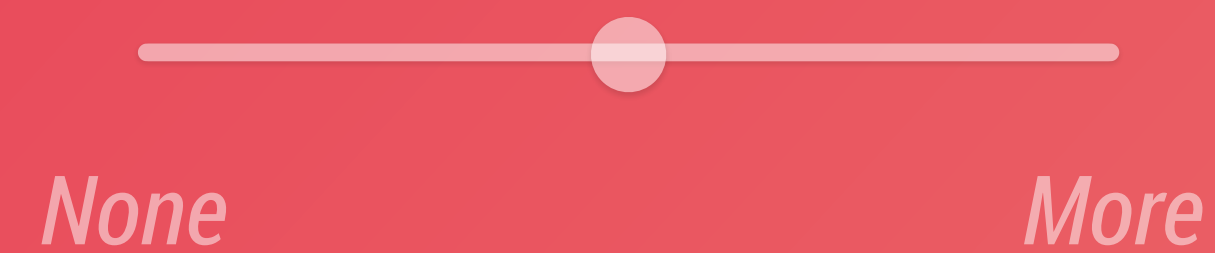
Target Data Difference



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*

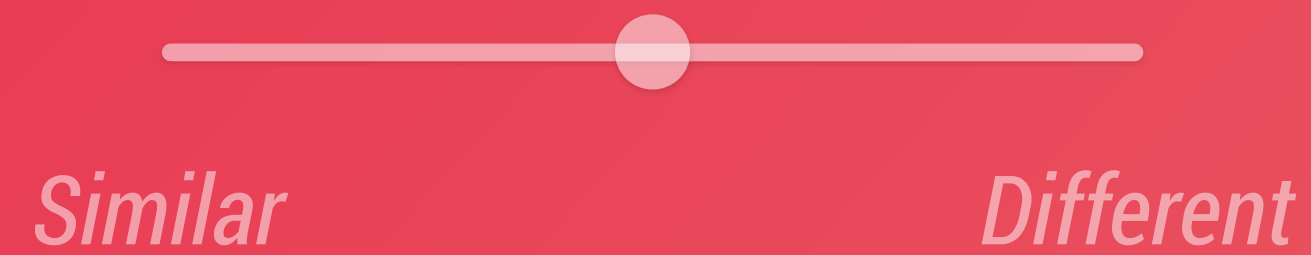
*More*

 **PaddleDetection**

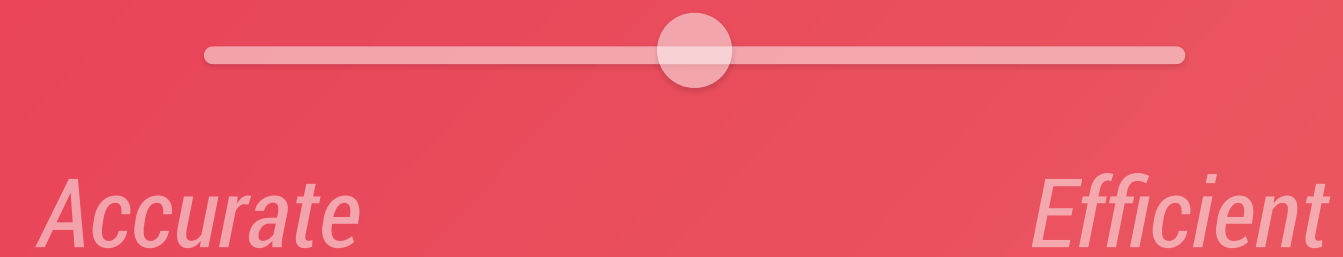
*And more in the future*

# Deep Learning Model Customization

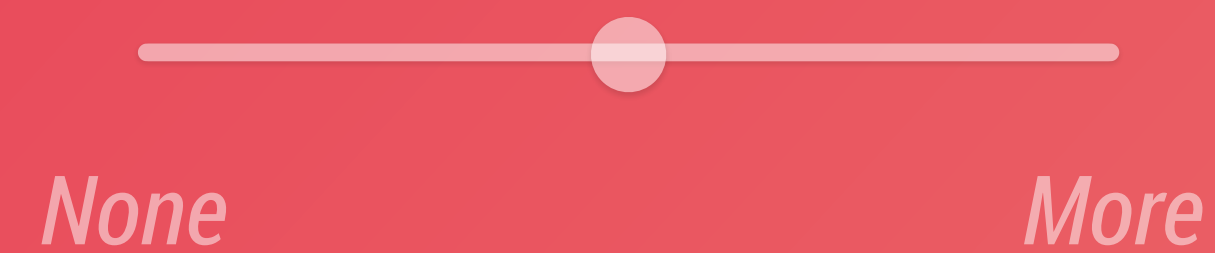
Target Data Difference



Accuracy/efficiency trade-off



Available training data





# Deep Learning Model Customization

Target Data Difference



*Similar*

*Different*

Model Fine-tuning

Accuracy/efficiency trade-off



*Accurate*

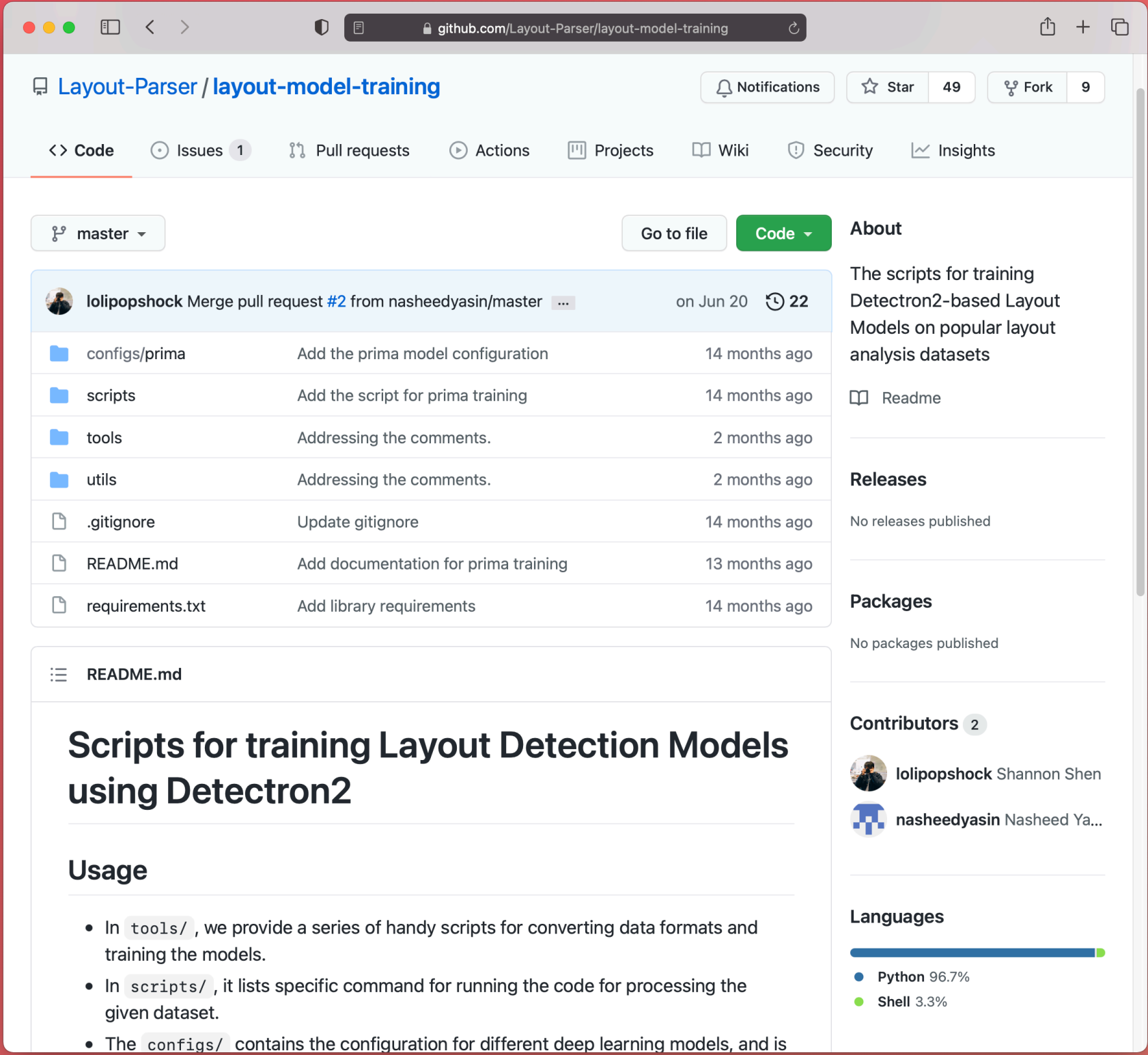
*Efficient*

Available training data



*None*

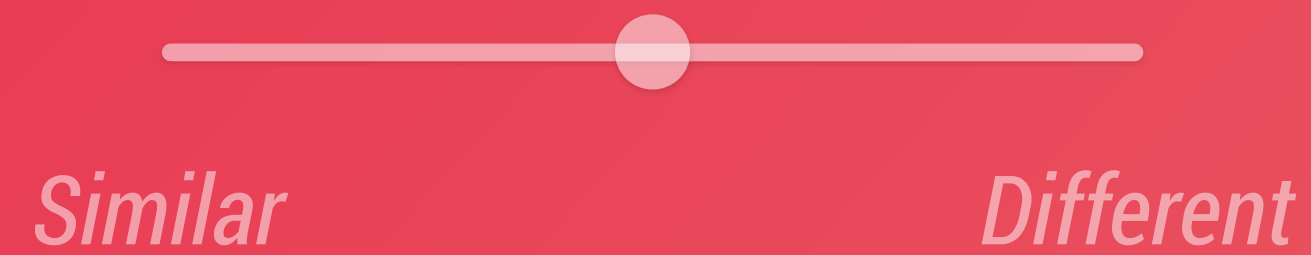
*More*



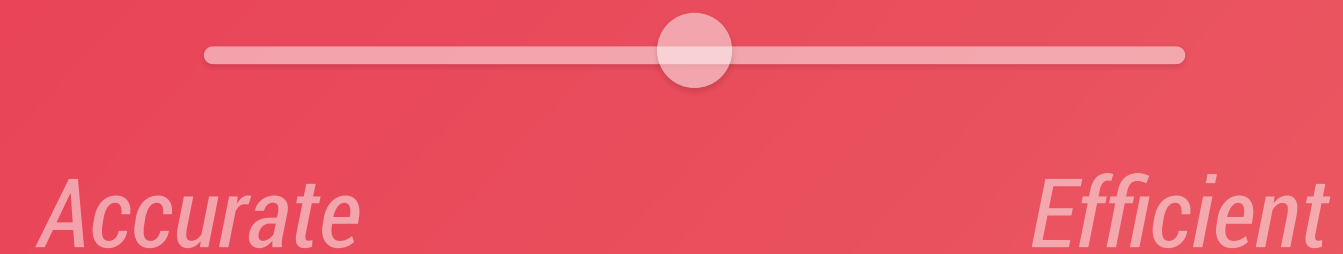
▲ *Layout Parser comes with the script that fine-tunes existing models to new datasets.*

# Deep Learning Model Customization

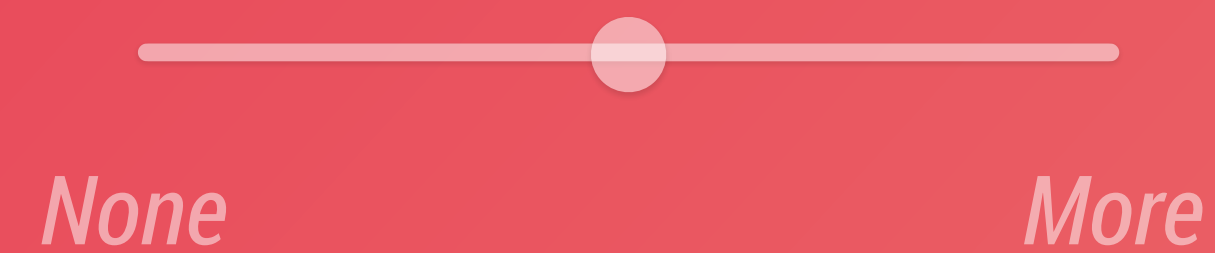
Target Data Difference



Accuracy/efficiency trade-off



Available training data





# Deep Learning Model Customization

# Target Data Difference

## Similar

## Different

# Accuracy/efficiency trade-off

## Accurate

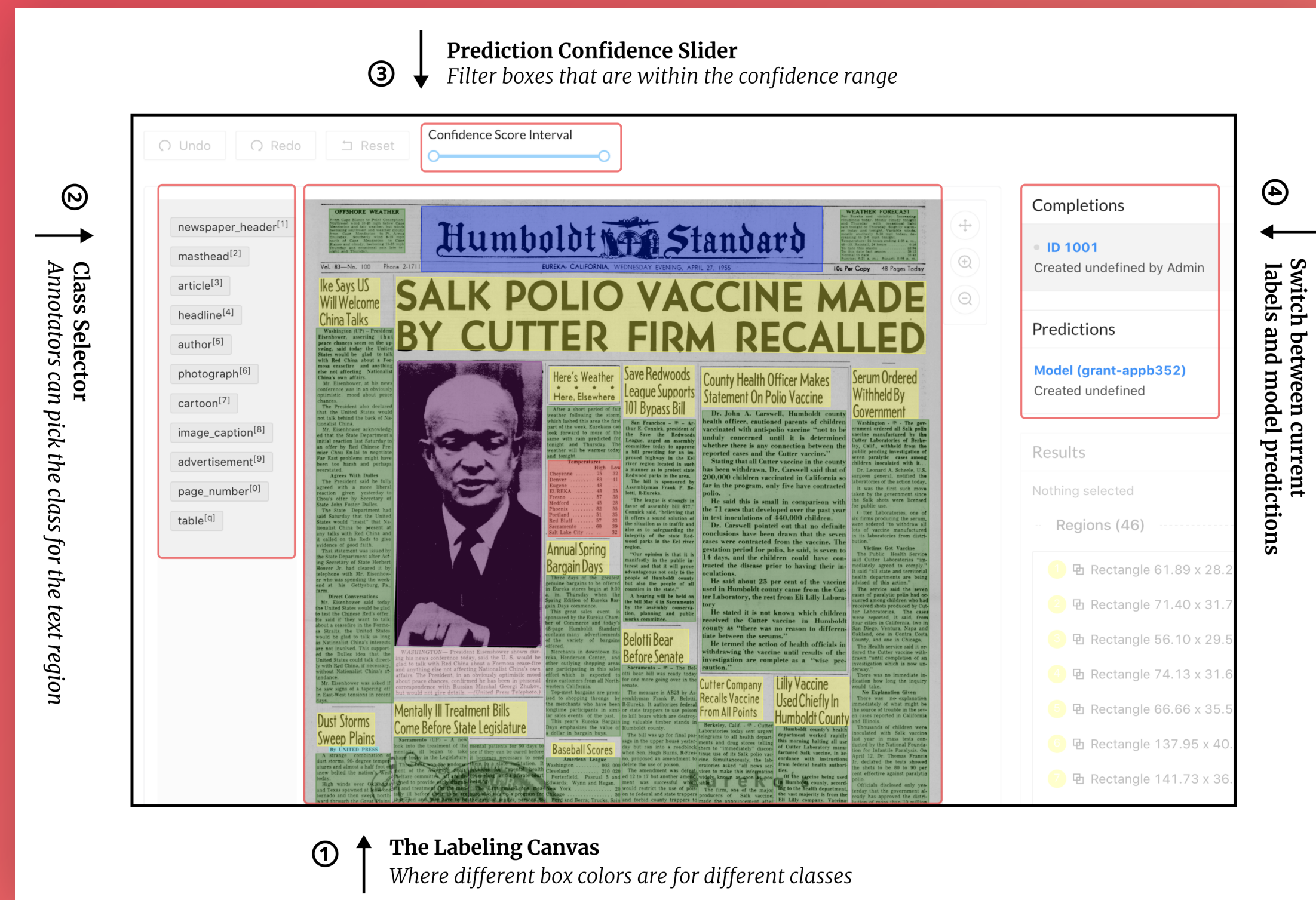
## Efficient

# Available training data

***None***

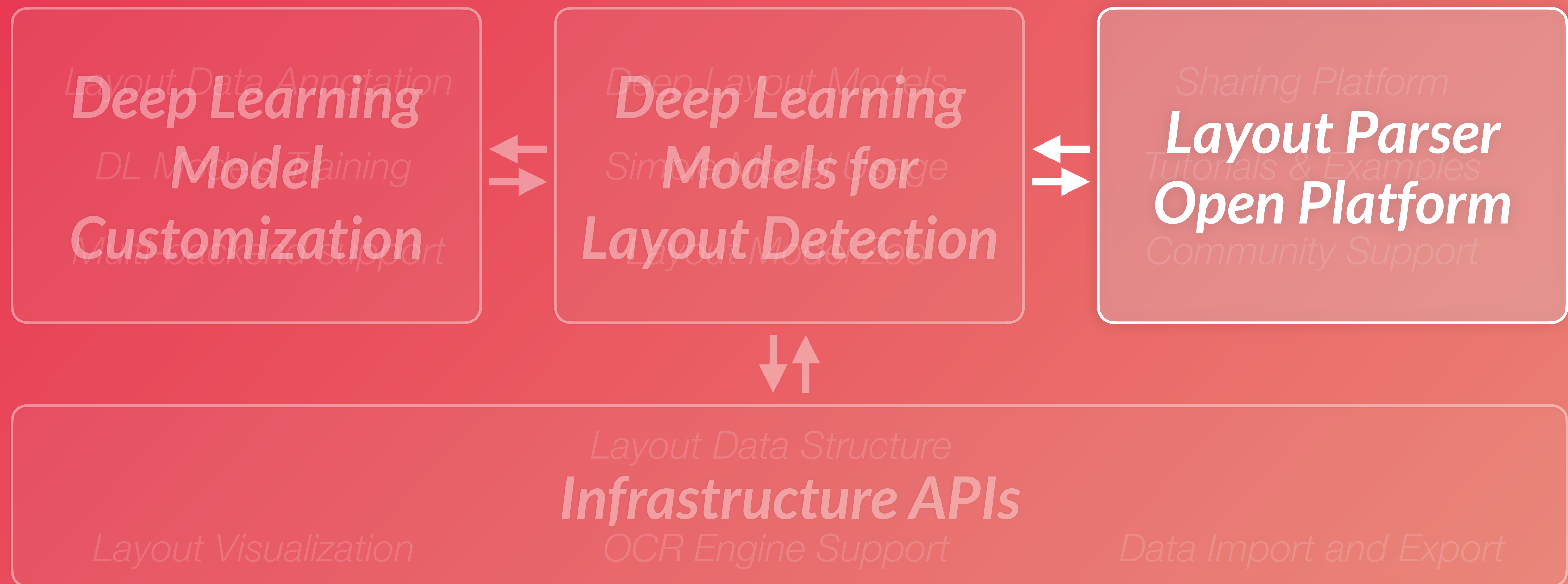
*More*

# Annotation & Model Retraining



*How about sharing your work with the community?*

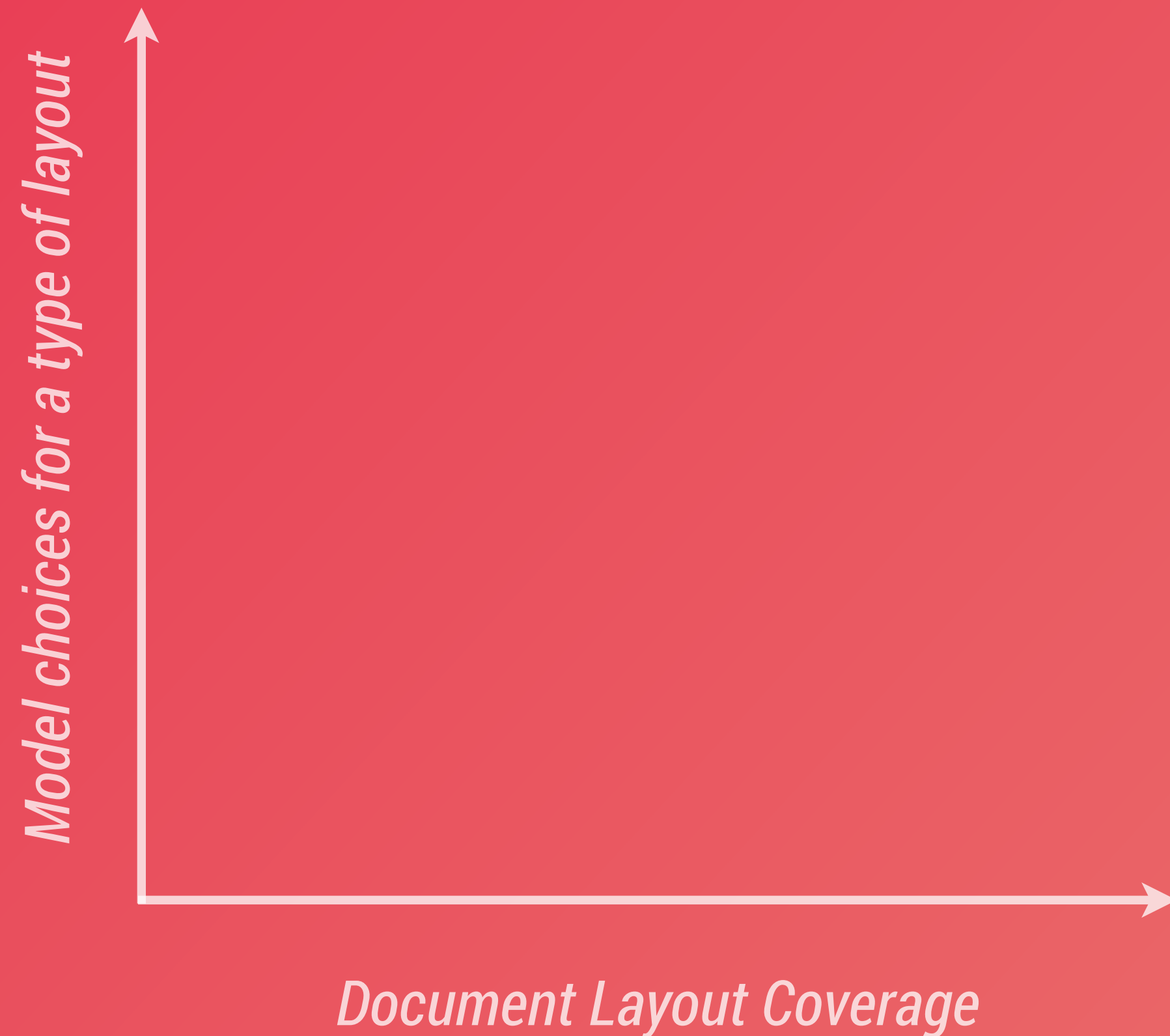




# Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines

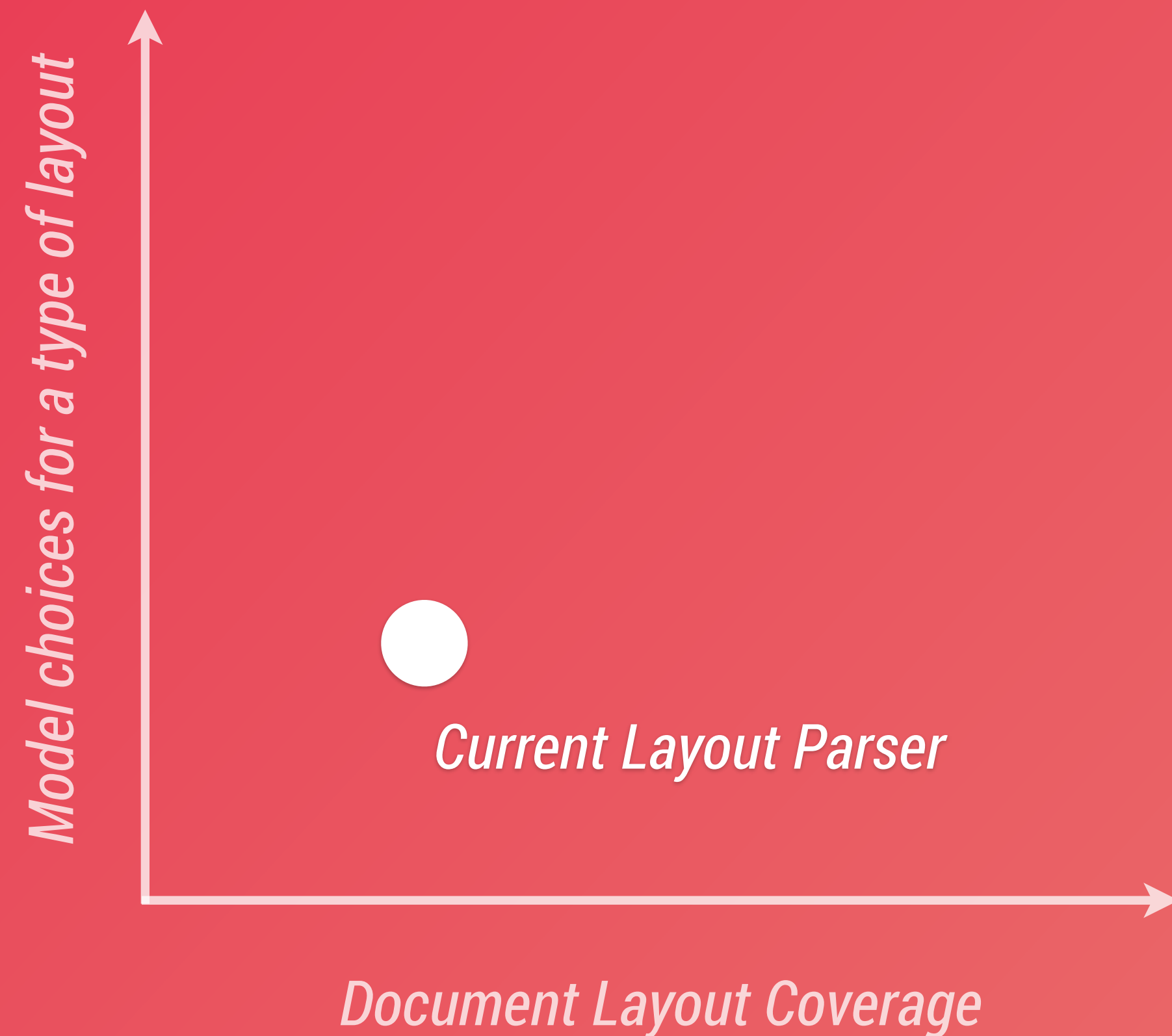




# Layout Parser Open Platform

Share Layout Models

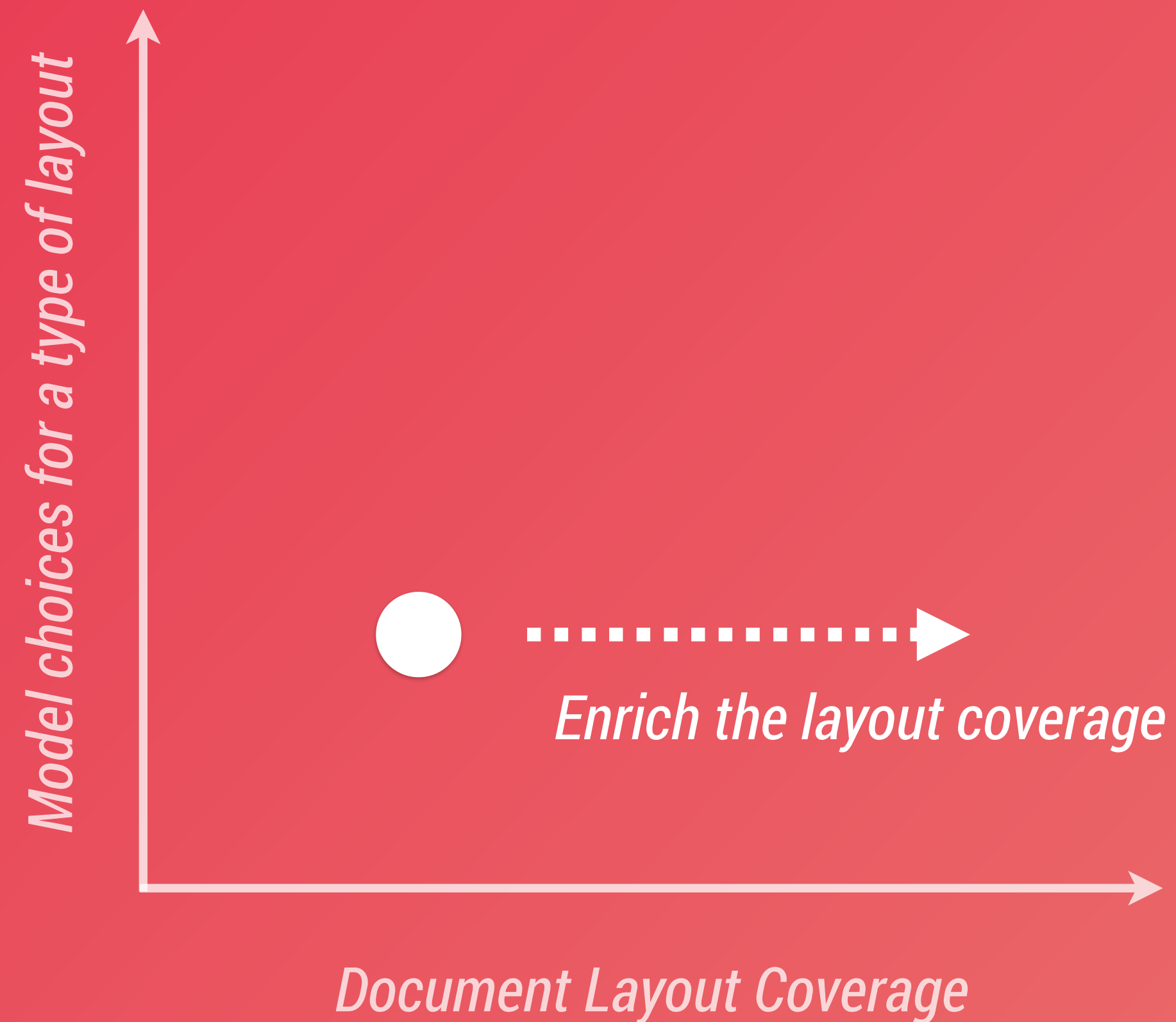
Share DIA Pipelines



# Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines



If we share:

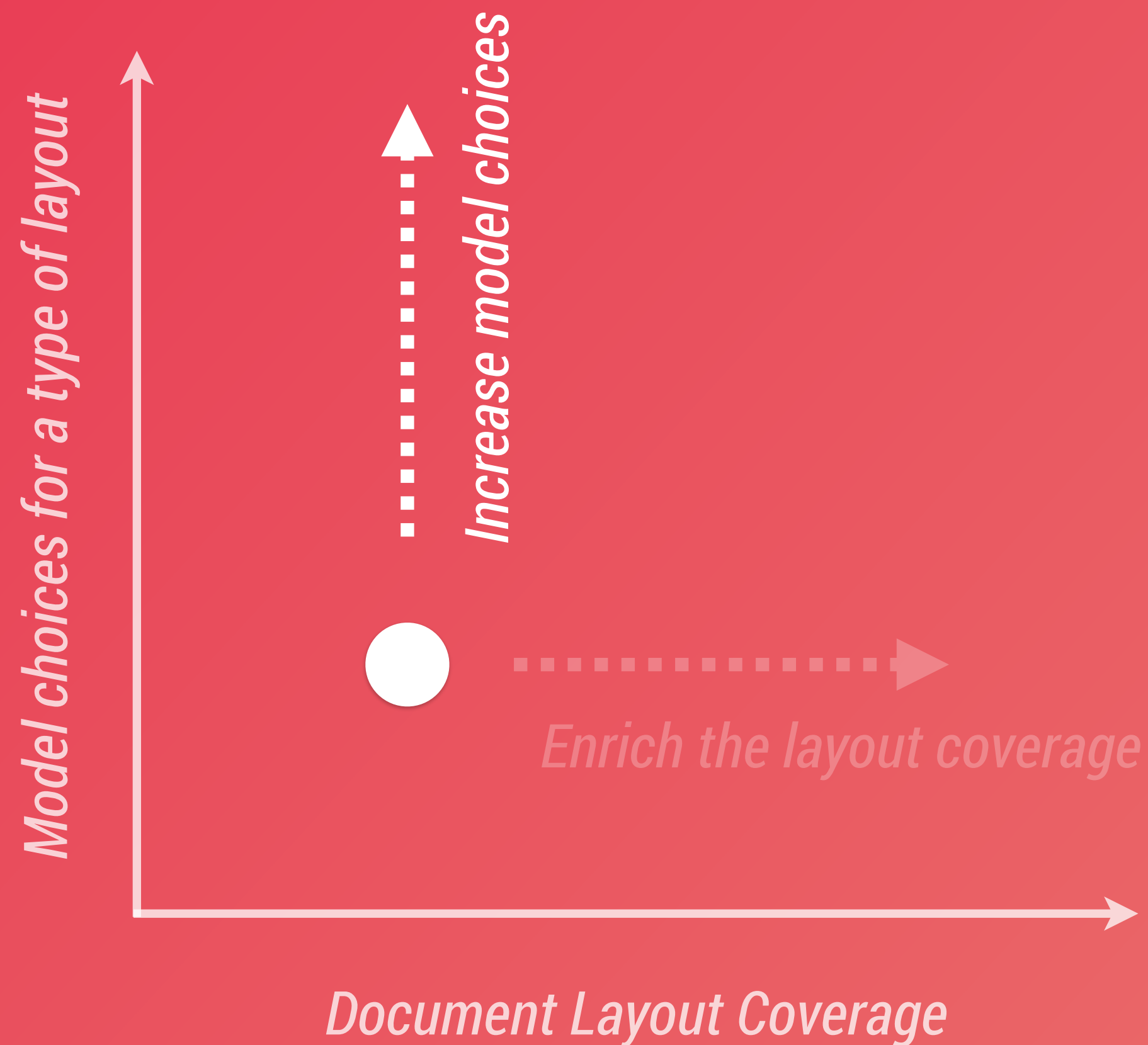
Models trained on different datasets



# Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines



If we share:

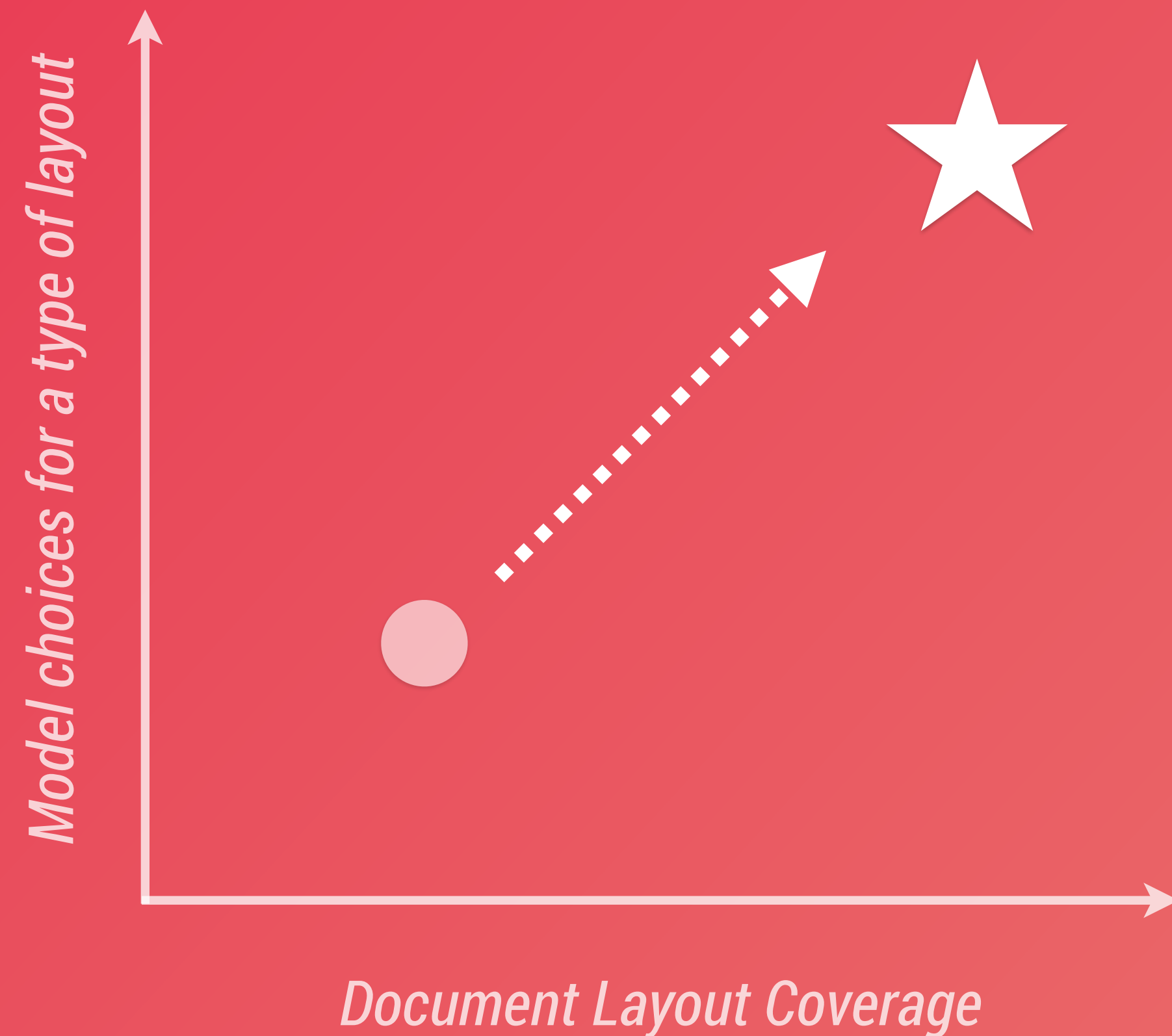
Models trained on different datasets

Models of different architecture/backend

# Layout Parser Open Platform

Share Layout Models

Share DIA Pipelines



Ultimately:

Make it easier to find the ideal model

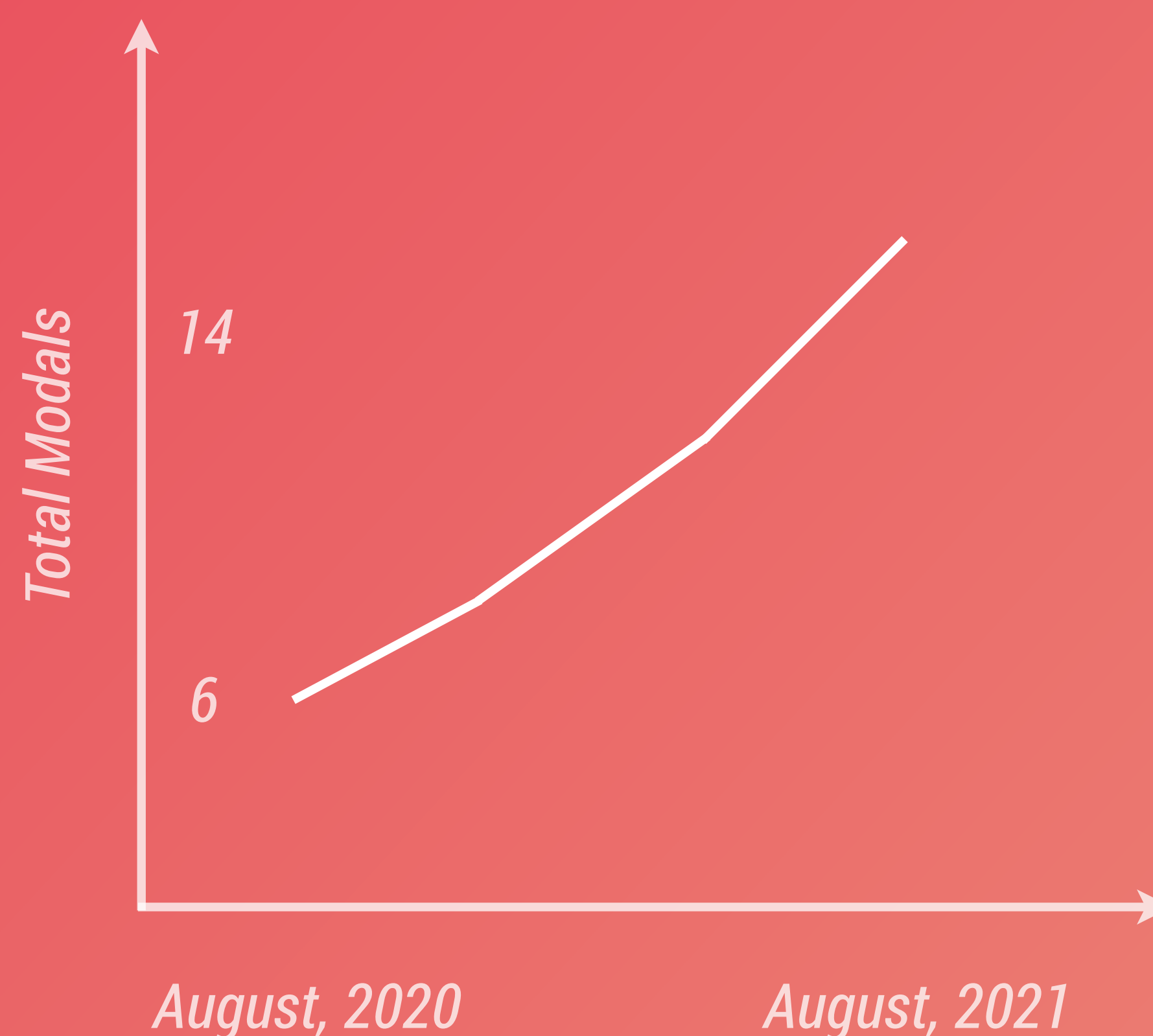
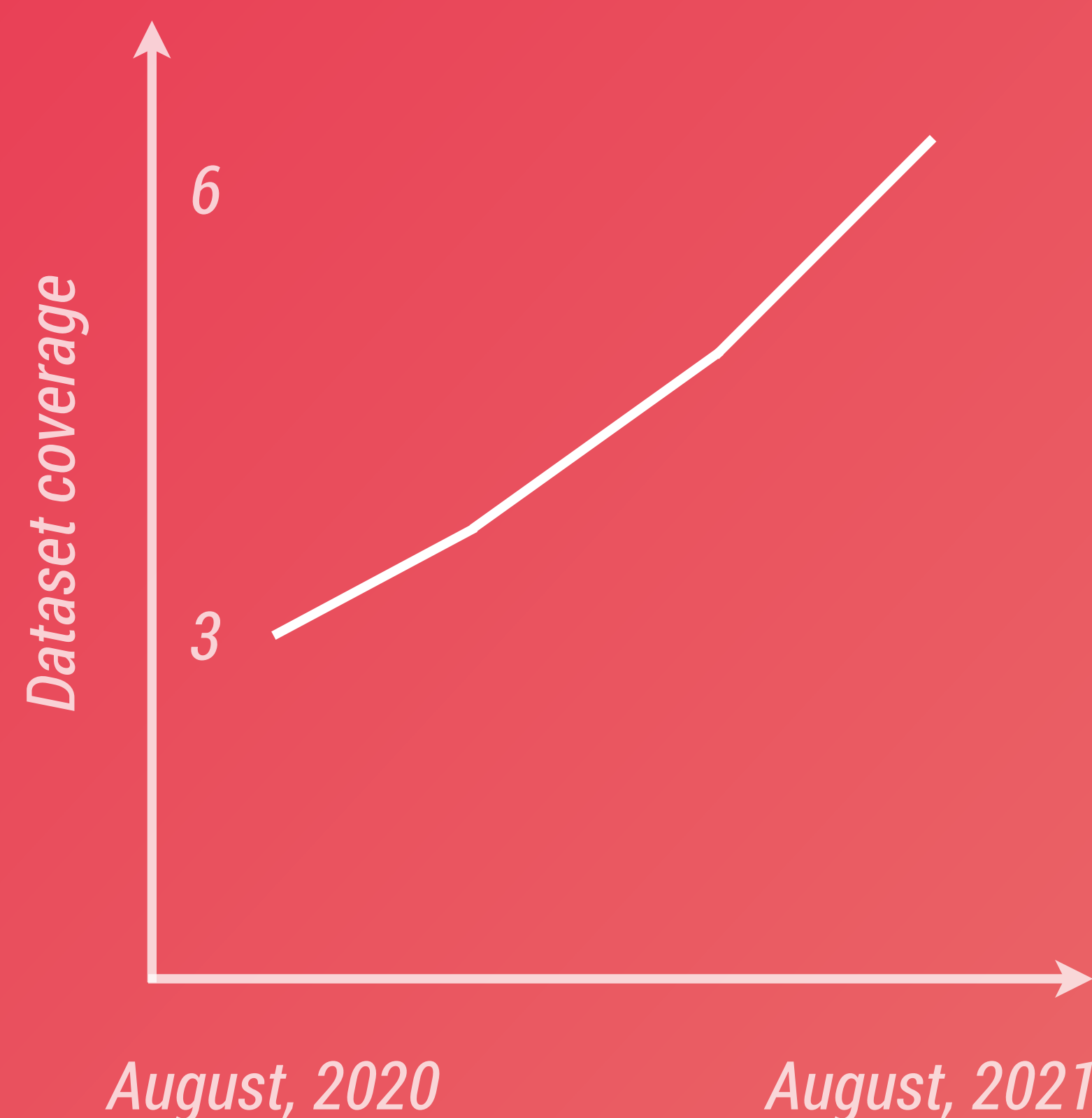


# Layout Parser Open Platform

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 DLA Pipelines

## Examples:

# Table Extraction

CMCECF - District of Minnesota - Live - Docket Report

https://ecm.court.gov/docm.aspx?caseNo=1043693773371072534\_s\_4

ATTORNEY TO BE NOTICED

V.

Defendant:

Minnesota Beef Industries, Inc.

represented by **William J Egan**  
Oppenheimer, Wolff & Donnelly  
Suite 1300, Flaro VII  
45 South Seventh Street  
Minneapolis, MN 55402  
612-467-2409  
Fax: 612-467-1100  
Email: [wegan@oppenheimer.com](mailto:wegan@oppenheimer.com)  
**LEAD ATTORNEY**  
**ATTORNEY TO BE NOTICED**

Interpleader:

Sheila Katz

represented by **Coleen E Culberts**  
Culberts & Lienemann, LLP  
414 Cedar St Ste 1650  
St Paul, MN 55101  
612-296-9300  
Fax: 612-296-9305  
Email: [cculberts@clawayers.com](mailto:cculberts@clawayers.com)  
**LEAD ATTORNEY**  
**ATTORNEY TO BE NOTICED**

Leslie L Lienemann

Culberts & Lienemann, LLP  
414 Cedar St Ste 1650  
St Paul, MN 55101  
612-296-9300  
Fax: 612-296-9305  
Email: [llienemann@clawayers.com](mailto:llienemann@clawayers.com)  
**LEAD ATTORNEY**  
**ATTORNEY TO BE NOTICED**

Date Filed	#	Docket Text
4/17/2002	1	COMPLAINT - Summary issued. Assigned to Senior Judge David S. Dwyer for Initial Rights Test and referred to Magistrate Judge Susan R. Nelson. (Spjco) (Entered: 04/17/2002)
4/17/2002		NOTICE given to any addressees clerk & Jean K Piron on behalf of the plaintiff. (Entered: 04/17/2002)

CMCECF - District of Minnesota - Live - Docket Report

https://ecm.court.gov/docm.aspx?caseNo=1043693773371072534\_s\_4

04/24/2002

2

PETITION AND ORDER for admission pro hac vice of government atty. : Clerk Richard D. Shellen on behalf of plaintiff by Rosemary J. Fox (Spjco) (MD) (Amended: 04/24/2002) (Entered: 04/24/2002)

04/03/2002

1

Summary - RETURN OF SERVICE executed upon defendant Minnesota Beef Industries on 5/10/2002 (MD) (Entered: 05/09/2002)

05/14/2002

4

ANSWER by defendant (MD) (Entered: 05/17/2002)

05/14/2002

5

NOTICE OF PRETRIAL CONFERENCE (Magistrate Judge Susan R. Nelson : 5/20/02), pretrial conference set for 2:30 p.m.12; Rule report did not set for 04/02 (Spjco) (cc all counsel) (MD) (Entered: 05/23/2002)

05/29/2002

6

Amended Notice of Pretrial Conference (Magistrate Judge Susan R. Nelson : 5/29/02), pretrial conference set for 1:30 p.m.6/18/02; Rule report did not set for 04/02 (Spjco) (cc all counsel) (MD) (Entered: 05/30/2002)

06/04/2002

7

REPORT OF RULE 36(b) MEETING by plaintiff, defendant (Spjco) (MD) (Entered: 06/05/2002)

06/20/2002

8

Pretrial Settlement Order (Magistrate Judge Susan R. Nelson : 6/19/02), and complaint set for 10/12; discovery set for 3/7/03; non-dispositive motions set for 4/15/03; dispositive motions set for 6/10/03; ready for trial set for 8/1/03 (cc all counsel) (MD) (Entered: 06/25/2002)

08/02/2002

9

MOTION by movant Sheila Katz to leave to intervene as plaintiff to (Magistrate Judge Susan R. Nelson : 1/20) (PTN) (Entered: 08/09/2002)

08/02/2002

10

DECLARATION of Coleen E. Culberts in motion for leave to intervene as plaintiff [P-1] (PTN) (MD) (Entered: 08/09/2002)

08/19/2002

11

Amended Notice by movant Sheila Katz of hearing setting hearing for motion for leave to intervene as plaintiff to (Magistrate Judge Susan R. Nelson : 8/19/02) (PTN) (MD) (Entered: 08/21/2002)

08/21/2002

12

RESPONSE by plaintiff to Sheila Katz's motion to intervene [P-1] (Spjco) (MD) (Entered: 08/27/2002)

08/18/2002

13

STIPULATION AND ORDER (Magistrate Judge Susan R. Nelson : granting motion that Sheila Katz may intervene as plaintiff movement [P-1] (Spjco) (cc all counsel) (MD) (Entered: 08/26/2002)

09/26/2002

14

AMENDED COMPLAINT [P-1] (Spjco) plaintiff, deny (MD) (VEM) (Entered: 09/26/2002)

09/26/2002

15

SUMMONS issued to as Minnesota Beef Ltd (VEM) (Entered: 09/26/2002)

09/27/2002

16

OFFER OF RECKENMENT filed by def. (Spjco) (DDBM) (Entered: 09/26/2002)

10/09/2002

17

AMENDMENT by defendant to offer of settlement [P-1] to negotiate all jointly by pbf and interview for total amount of \$40,000, including costs, attorneys, attorney fees \$20,000 (MD) (Entered: 10/23/2002)

10/09/2002

17

AMENDMENT by defendant to offer of settlement [P-1] to negotiate all jointly by pbf and interview for total amount of \$40,000, including costs, attorneys, attorney fees \$20,000 (MD) (Entered: 10/23/2002)

12/20/2002

18

RULE 17 - DISCLOSURE STATEMENT by Minnesota Beef Ltd that none exist (MD) (DPL) (Entered: 01/03/2003)

CMCECF - District of Minnesota - Live - Docket Report

https://ecm.court.gov/docm.aspx?caseNo=1043693773371072534\_s\_4

03/12/2004

52

almschmidt@judd.net 10/29/2004 09:00 (MD) (Entered: 12/11/2003)

03/12/2004

52

NOTICE OF ADDITIONAL CASES FOR TRIAL (Senior Judge David S. Dwyer 11/20/04 jury trial set for 9:00 a.m. 4/20/04 - 3 pgs) (cc all counsel) (HL) (Entered: 03/12/2004)

03/12/2004

53

NOTICE OF FINAL SETTLEMENT CONFERENCE (Magistrate Judge Susan R. Nelson : 1/13/04) final settlement conference set for 9:30 a.m. on 4/30/04 - 2 pgs (cc all counsel) (HL) Additional additional add to 10/20/04 (akb) (Entered: 03/12/2004)

02/26/2004

54

TRIAL NOTICE: Jury trial set for 4/26/04 9:00 AM at Minneapolis - Courthouse 149 before Senior Judge David S. Dwyer. (PTN) (Entered: 02/26/2004)

04/08/2004

55

Minute Entry for proceedings held before Mag. Judge Susan R. Nelson - Final Settlement Conference held on 4/8/2004. No Settlement Reached. (HL) (Entered: 04/08/2004)

04/14/2004

56

Minute Entry for proceedings held before Mag. Judge Susan R. Nelson - Telephone conference is videotaped discovery of initial packing plant site take place at 10:00 a.m. on 4/20/04 held at 4/20/04. (HL) (Entered: 04/15/2004)

05/17/2004

57

STIPULATION AND ORDER for Default Trial With Prejudice. Signed by Senior Judge David S. Dwyer on 5/17/04. (HL) (Entered: 05/17/2004)

05/17/2004

58

Consent Decree Signed by Senior Judge David S. Dwyer on 5/17/04. (HL) (Entered: 05/17/2004)

PACER Service Center

Transaction Receipt

Case

Doc#

Date

ba128

Search

06/01/2007 13:02:08

Document

Search

06/01/2007 13:02:08

ba128

Search

06/01/2007 13:02:08

Page

Count

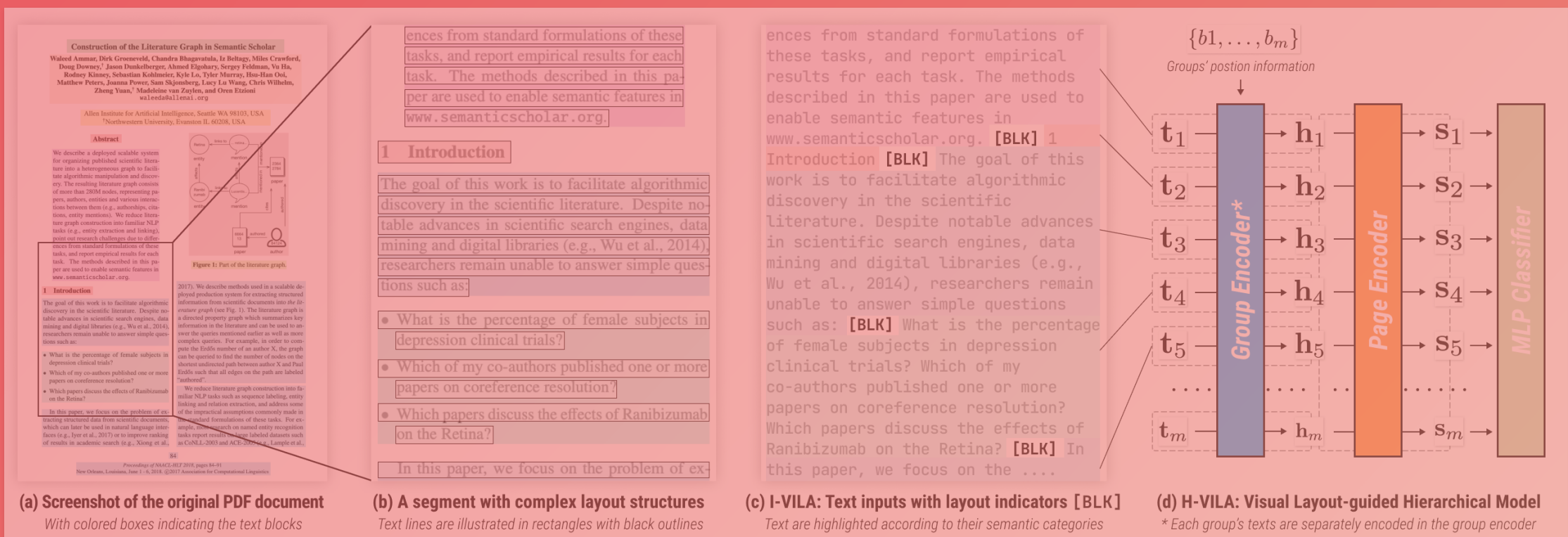
4

Count

0.32

</

# Scientific Document Parsing





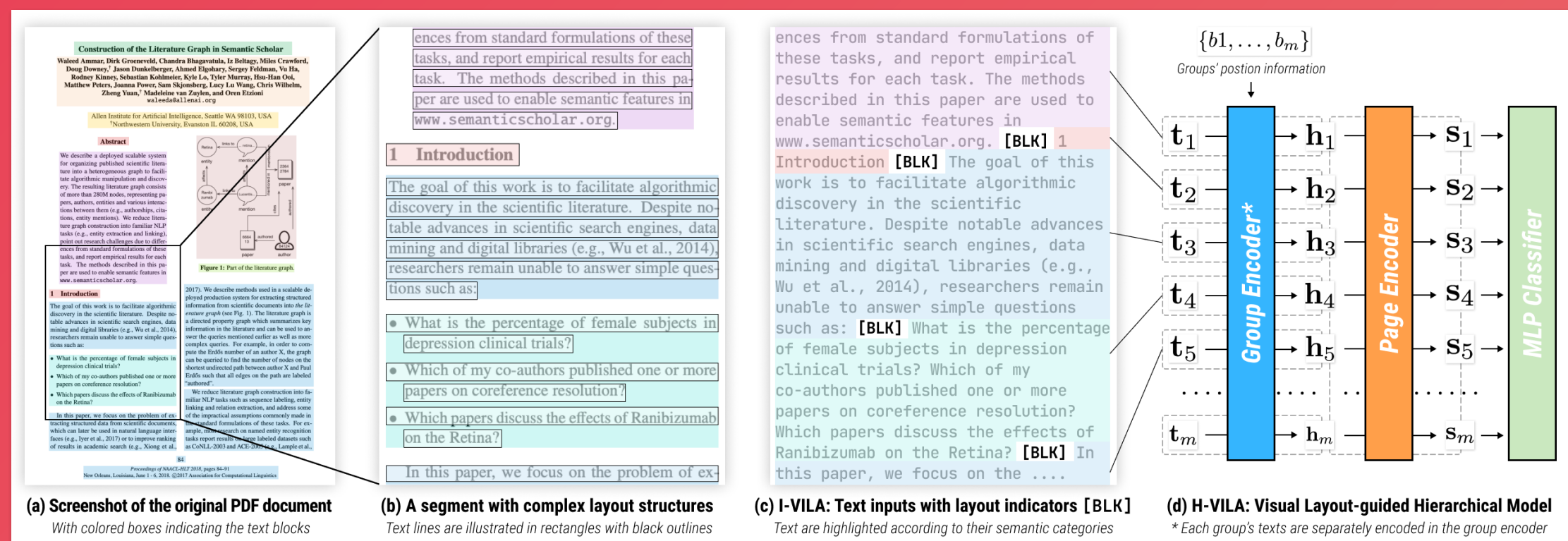
# Layout Parser Open Platform

# Share Layout Models

# Share DLA Pipelines

## Examples:

# Scientific Document Parsing



# Historical Document Analysis



# Layout Parser Open Platform

# Share Layout Models

# Share DIA Pipelines

## Examples:

# Historical Document Analysis

<b>江島</b>	織物有	寺西入電四局二八三二	設立 昭和廿五年九月	目的 西陣織物製造販賣	資本金 百二十萬圓(二千二百口)	決算期 八月	配當無	代江島喜一郎 次郎	監江島吟次郎	大口出資者 (出資者數一〇)	江島喜一郎	從業員 一五	年商內高 四千二百萬圓內外	取引銀行 三和(西陣)	<b>永和化成工業株</b>	南區吉祥院落合町六四 電五局六三三七	設立 昭和卅年十一月	目的 化學藥品製造	資本額 百萬圓(二千株)	決算期 五月	配當無	代西村永治郎 重雄	野々村儀一郎 吉田	大株主 (株主數一一)	西村永治郎 ECC株	從業員 二四	年商內高 四千六百萬圓內外	取引銀行 滋賀(九條)	設備 敷地四〇〇坪、建物三〇〇坪、脫水機三、其他六	<b>英昌</b>	織物株	竹野郡彌榮町和田野 電溝谷七七	設立 昭和廿八年七月	目的 絹人絹織物製造販賣	資本額 百萬圓(二千株)	決算期 六月	配當無	代天下倉保一郎 綿大下倉留藏	天下倉知惠 監八畑市郎	天下倉八重	大株主 (株主數七)	天下倉保一郎 700株	從業員 一七	年商內高 二千八百萬圓內外	取引銀行 京都(峰山)丹後	中央信金(彌榮)	<b>株榮</b>	光商店	中京區西ノ京左馬寮町 東部一〇電公局三六九	設立 昭和廿八年四月	目的 纖維服飾品製造販賣	資本額 百二十萬圓(二千四百株)	決算期 三月	配當一・二割	代光村和雄 壽男	夜野白郎 監荒川正太郎	山中 錦造
-----------	-----	------------	---------------	----------------	---------------------	-----------	-----	--------------	--------	-------------------	-------	-----------	------------------	----------------	----------------	-----------------------	---------------	--------------	-----------------	-----------	-----	--------------	--------------	----------------	---------------	-----------	------------------	----------------	------------------------------	-----------	-----	--------------------	---------------	-----------------	-----------------	-----------	-----	-------------------	----------------	-------	---------------	----------------	-----------	------------------	------------------	----------	-----------	-----	--------------------------	---------------	-----------------	---------------------	-----------	--------	-------------	----------------	----------



*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*

*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 Layout Models*  
*Simple Model Usage*  
*Layout Model Zoo*

# Deep Learning Models for Layout Detection



*Layout Data Structure*  
*OCR Engine Support*

# Infrastructure APIs

*Data Import and Export*

*Layout Data Annotation*

*DL Models Training*

*Multi-backend support*

*Sharing Platform*

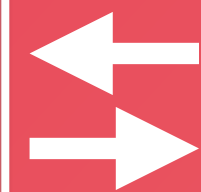
*Tutorials & Examples*

*Community Support*

*Layout Visualization*



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

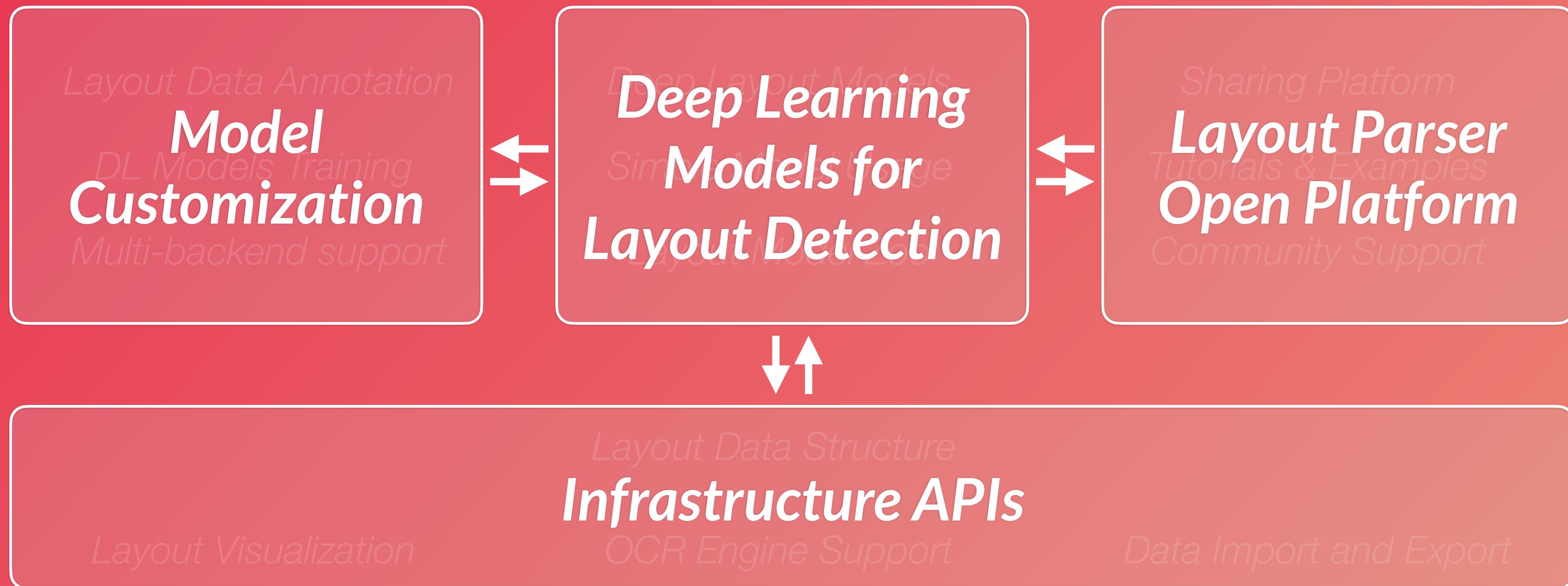


*Deep Layout Models*  
**Deep Learning  
Models for  
Layout Detection**  
*Simple Model Usage*  
*Layout Model Zoo*

*Sharing Platform*  
*Tutorials & Examples*  
*Community Support*



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





*Motivation*

*Demo*

*Design & Implementation*

***Future Work***

*Community*

# Future Work

Generalized Models

Multimodal Modeling



## Generalized Models

## Multimodal Modeling





# Future Work

Generalized Models

Multimodal Modeling

## Can we have a single model for multiple layouts?

text searches on induced tooth movement of animals and human 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 osteoclastic treatment.<sup>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup> There is no scientific support, methodologies, evidence or outcomes that allow such statement. The same is applied to the process of osteoneogenesis.

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 uncomfortable 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 parathyroid hormone, typical of this extreme 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 cytotoxic 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-bisphosphonate 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 cytotoxic 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

© 2012 Elsevier Inc. All rights reserved.

12

Journal Pre-proof

SATURDAY EVENING

Reading for Women and all the Family

The Daredevil

By McManus

All's Well That Ends Well

By May Martin

PEA COAL

J. B. Montgomery

Hotels, Restaurants and Boarding Houses

The Federal Machine Shop

3% PAID ON SAVINGS ACCOUNTS

© 2012 Elsevier Inc. All rights reserved.

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"—as 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 34% and between 1991 and 2002, according to one recent study, the Interior Ministry dissolved 134 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.

...too often, Italian politicians have allowed 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 well and good, but the problem is, Cersosimo, an economics professor at the University of Calabria, "We shouldn't see this as a country divided in two," he says. "The malaises of the Mezzogiorno are the malaises of Italy. It's just a question of degree: what is gray in Italy is black in the south." Indeed, entrenched nationwide Mafia tax evasion, cumbersome bureaucracy and a self-serving political class are

...piece with the south's bright and blatant corruption. Nobody in the private sector has a leg in Italy, as they have in Europe, explains Fabrizio Italian Economy Minister north has found ways to fix this, and can be competitive state of country," he says that is the anomaly, not and its ministers operate. Fixing the south means

...durable Man

...SOUTHERN OVERSIGHT WORKS on the other side of the abris. Current local leaders have to maximize the town's tourist and improve living conditions for. Beginning in 2001, Mayor Mario former union leader, implement municipal program under the grant "Annunziata wants to be in Italy 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, Annunziata has managed to renovate the historic city center, open a state-of-the-art physiotherapy center, and keep up environmental efforts like recycling.

...work. They've got to overcome 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 Annunziata, 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—demanded "socially useful" jobs by a state welfare

Table with 10 columns and 10 rows of text, likely a data table or index.

Modellek alkotása (függvény modell): a lineáris és az exponenciális növekedés/csökkenés matematikai modelljének összevetése konkrét, valós problémákban (például: népesség, energiateljesítmény, járványok).

A logaritmusfüggvények vizsgálata. Logaritmus alapfüggvények grafikonja, jellemzőik.

A logaritmusfüggvény mint az exponenciális függvény inverze. Függvénynek és inverzenek a grafikonja a koordináta-rendszerben.

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

Számtani sorozat, az n. tag, az első n tag összege. Matematikailag: 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. Modelltek alkotása: befektetés és hitel, kölcsönök feltevésekkel meghatározott befektetések és hitelek vizsgálata; a hitelezés módjai. Korábbi ismeretek mozgósítása (pl. százalékszámítás). A szövegbe többszörösen bevezetett, közvetett módon megfogalmazott információk és kategóriák szomszósága.

Kulcsfogalmak/ fogalmak

Számzsfüggvény, koszinuszfüggvény, tangensfüggvény. Exponenciális függvény, logaritmusfüggvény. Exponenciális folyamat. Számsorozat. Rekurzió. Számtani sorozat, mértani sorozat.

Órakeret N: 22 óra E: 10 óra

4. Geometria

Sokszögökkel, körrel kapcsolatos ismeretek. Pont-halmazok, nevezetesen pont-halmazok, ismeretek. Háromszög nevezetesen vonalak, pontjai, körét. Háromszögek, speciális háromszögek vonalakon, tételek. Egybevágóság, hasonlóság, szimmetria. Hegyesszögű szögfüggvények. Ekvivalens egyenlet. Elsőfokú és másodfokú egyenlet, kétszemélyes egyenletrendszer algebrai megoldása. Alapvető tételek, egyenlet rendezései feladatok köré, háromszöggel kapcsolatosan. Vektorok, vektorműveletek. Hasab, henger, gúla, kúp, gömb felismerése. Felzár, térfogat számítás fogalma. Poláris felzár. Számológép (számológép) használata.

Tájkörzés a térben. Tájkörzés a világ felismerése viszonyban: távolok, szögök, terület, terület, felzár és térfogat kiszámítása. A matematika két területének (geometria és algebra) összekapcsolása: koordináta-geometria. Értékek, korábbi ismeretek rendszeresítése, alkalmazása.

Számológép, koszinusz, tangens, koszinusz és koszinusz (a derékszögű háromszög és a két tért).

Síkdomok területének és területének számítása.

Pitagorasz összefüggés egy szög szinusz és koszinusz között. Összefüggés a szög és a mellékelt szinusz, illetve koszinusz között. A tangens kifejezése a szinusz és a koszinusz hányadosaként.

where  $\int \mathcal{D}\mathbf{x}$  is the action of the field  $\mathbf{x}$ . 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  $\int_0^T$  denotes Wiener measure starting from  $\mathbf{x}$ , and  $\psi(\mathbf{x})$  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  $H = \frac{1}{2} \nabla^2 + V(\mathbf{x})$  where  $\nabla^2$  is the scalar Laplacian looks identical to (3), but with  $\int_0^T$  a process depending on metric area 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

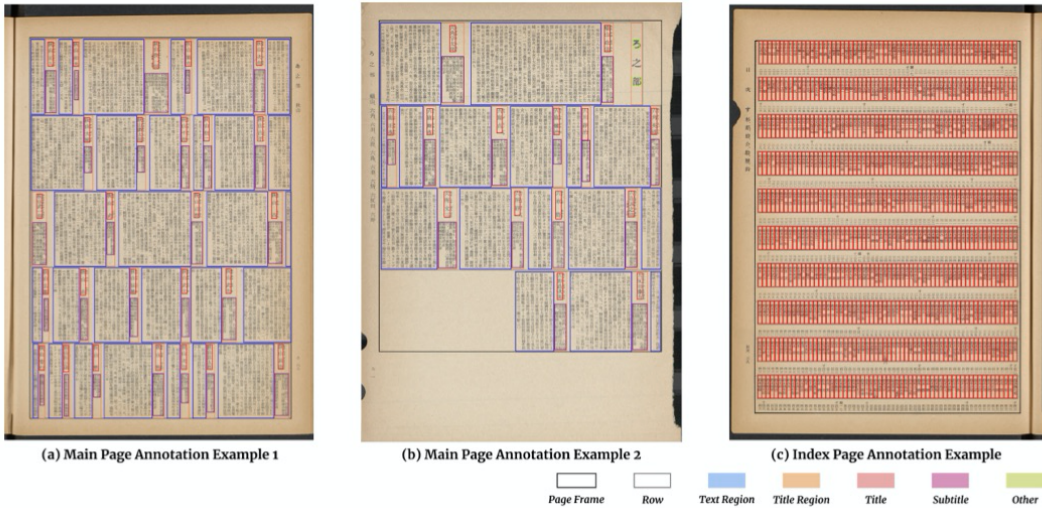


Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

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

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.

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

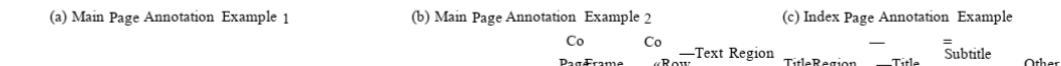
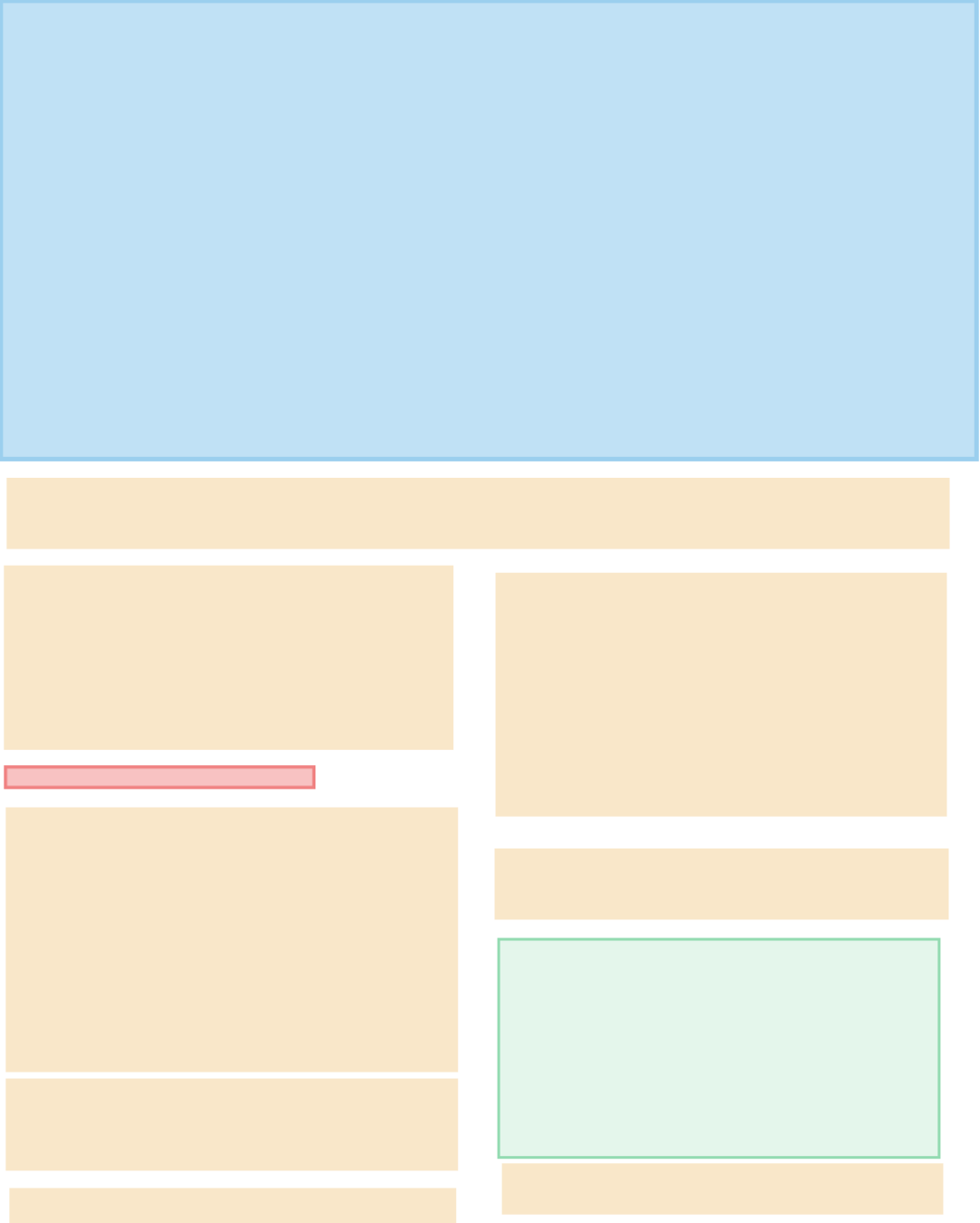


Figure 7: Annotation Examples in HJDataset. (a) and (b) show two examples for the labeling of main pages. The boxes are colored differently to reflect the layout element categories. Illustrated in (c), the items in each index page row are categorized as title blocks, and the annotations are denser.

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

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.

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

Image

Layout

Text



# 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.

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

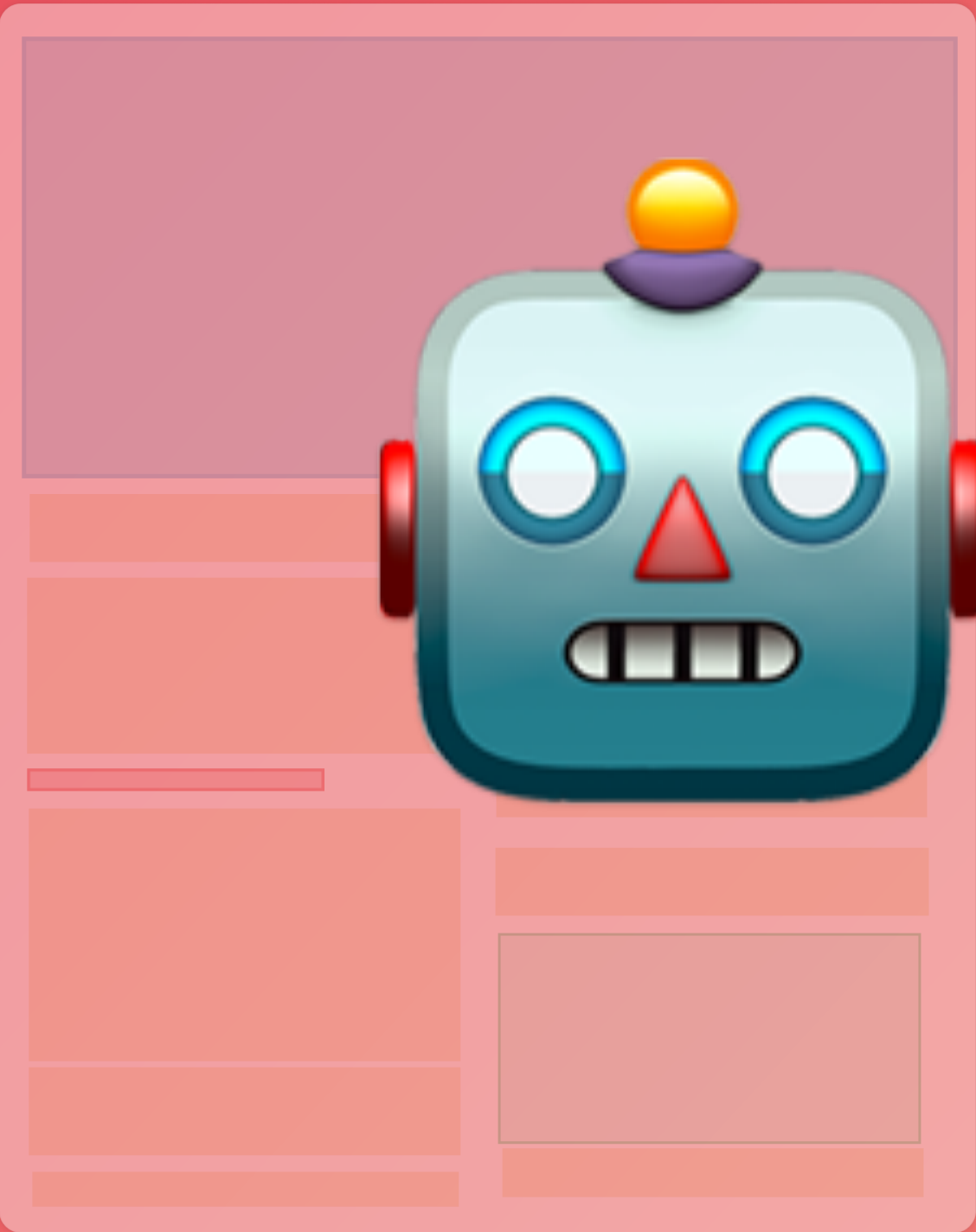
all the training data can be viewed as the benchmarks, while training with few samples (five in this case) are considered to mimic real-world scenarios. Given different training data, models pre-trained on HJDataset perform significantly better than those initialized with COCO weights. Intuitively, models 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

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.

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

Image



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.

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

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

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.

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

Text



# Future Work

Generalized Models

Multimodal Modeling



**MMDA**

*Multi Model Document Analysis*



Our Contributions  *Layout Parser*



**Our Contributions    A unified DLA toolkit**





A unified DLA toolkit

Our Contributions    Open the box usage



A unified DfA toolkit

Open the box usage

**Our Contributions    Deep learning integration**





A unified DfA toolkit

Open the box usage

Deep learning integration

**Our Contributions    Simple APIs + customization**



A unified DLA toolkit

Open the box usage

Deep learning integration

Simple APIs + customization

**Our Contributions**    **Open platform & community**

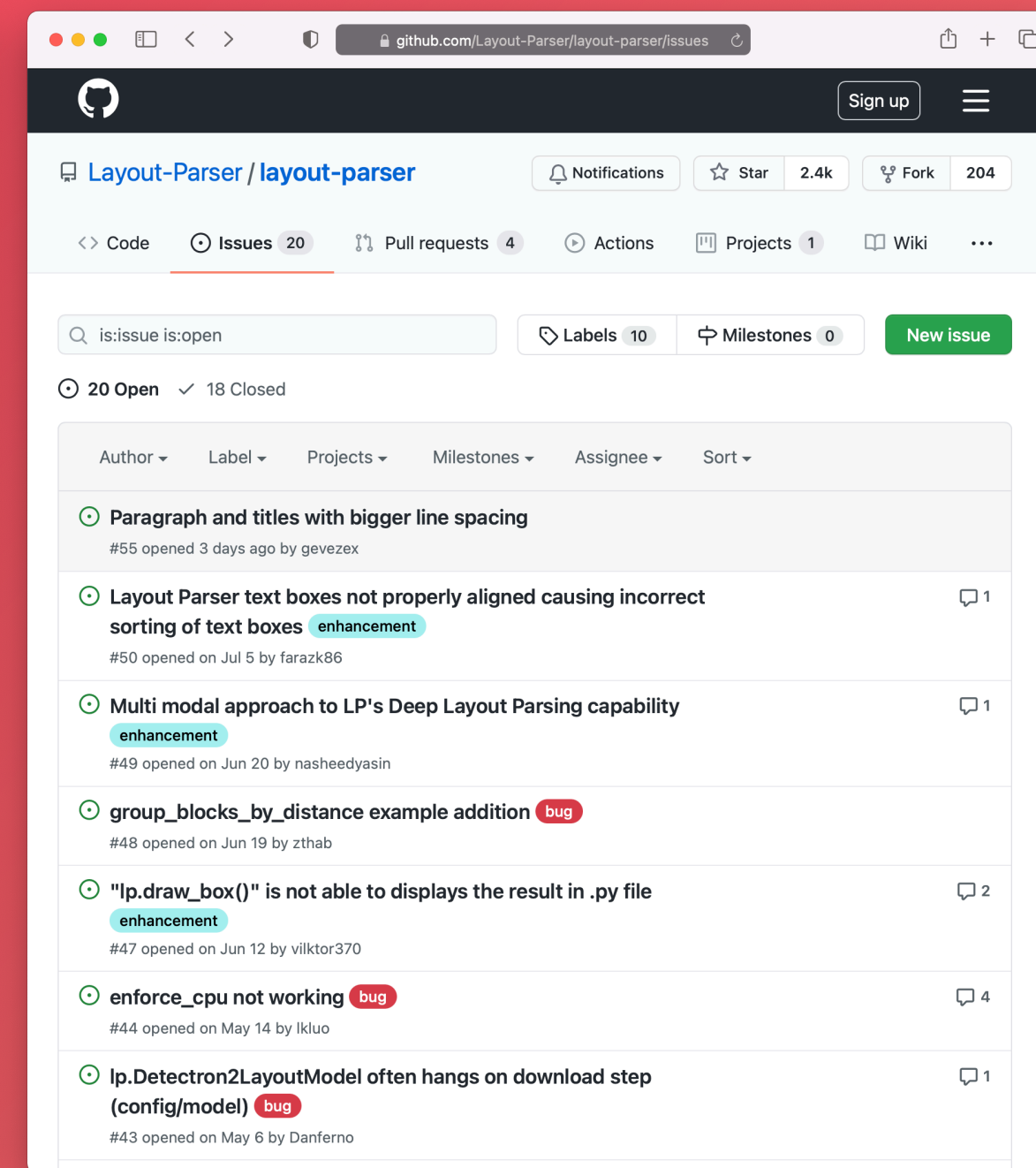


# LP Layout Parser

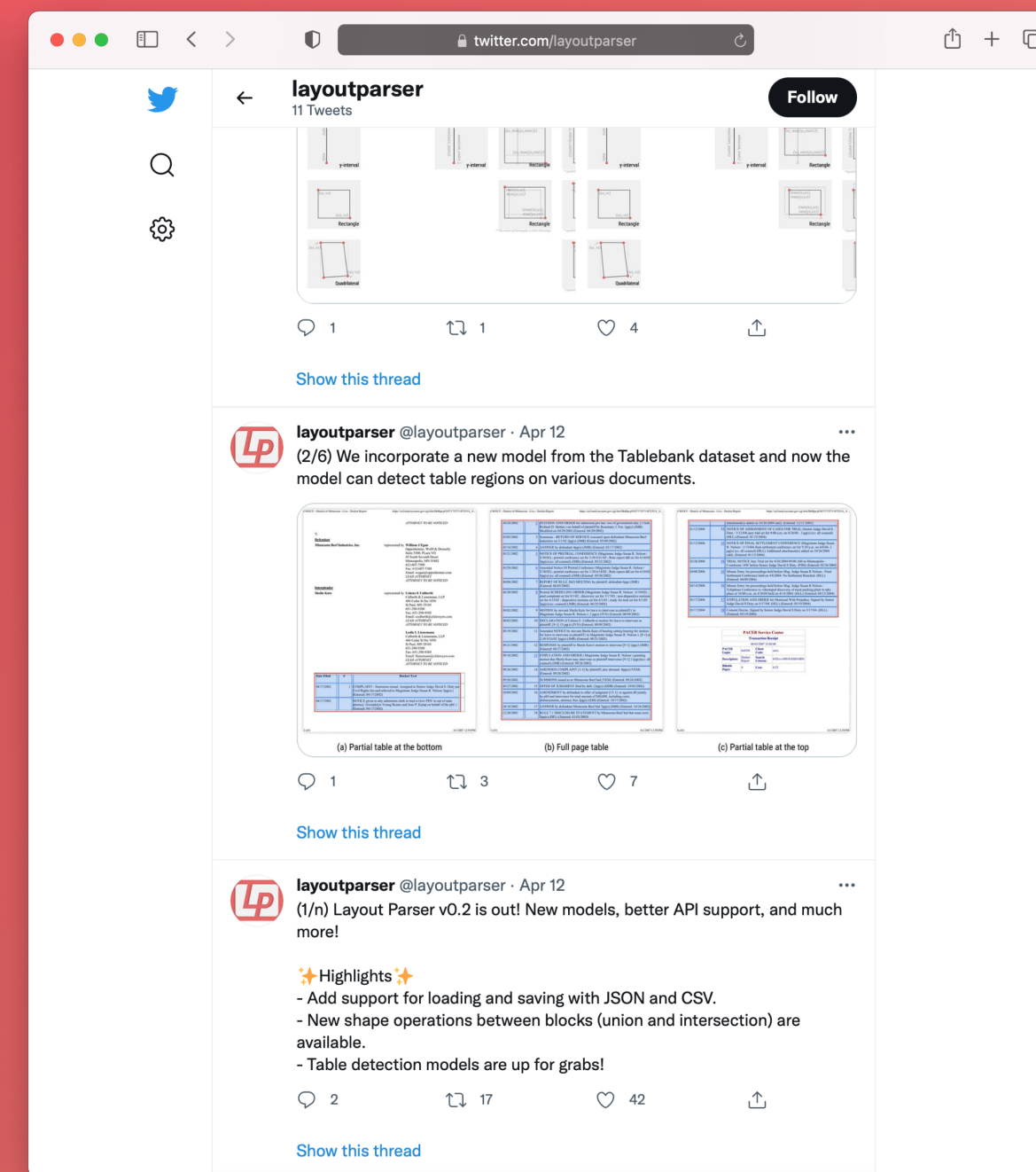
Community & Discussion



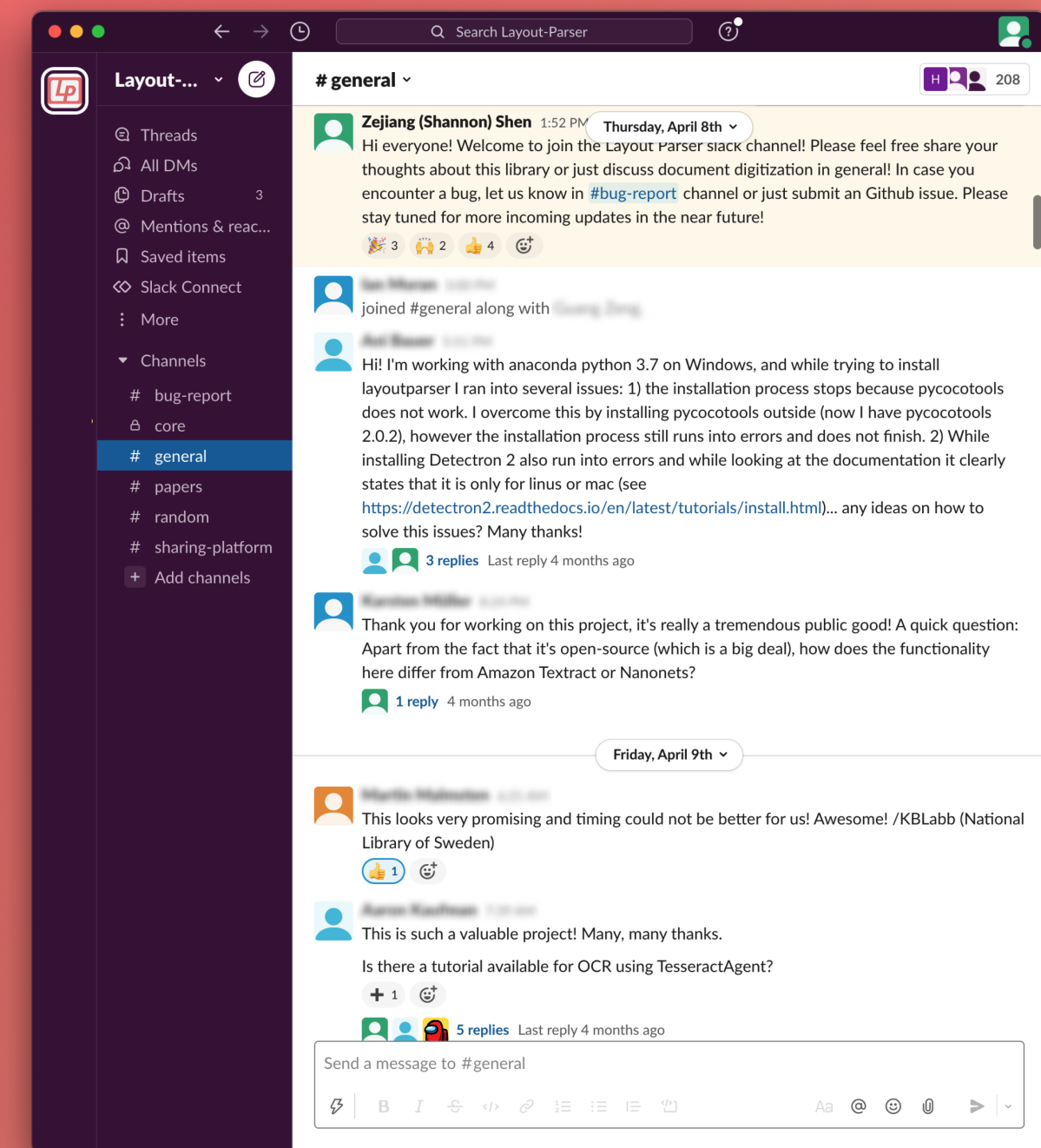
 Website  
layout-parser.github.io



 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*

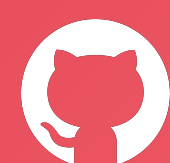
*Commandline Tools\**

# **Layout Parser**



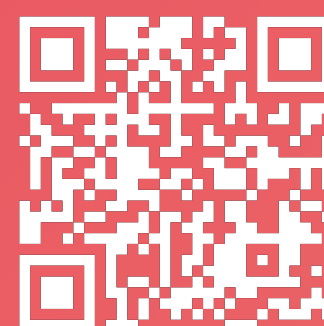
**Website**

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



**Github**

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



**Twitter**

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



**Slack**

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