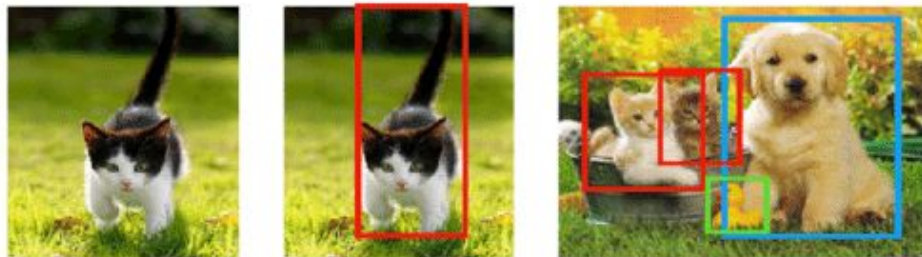


Modeling for Computer Vision

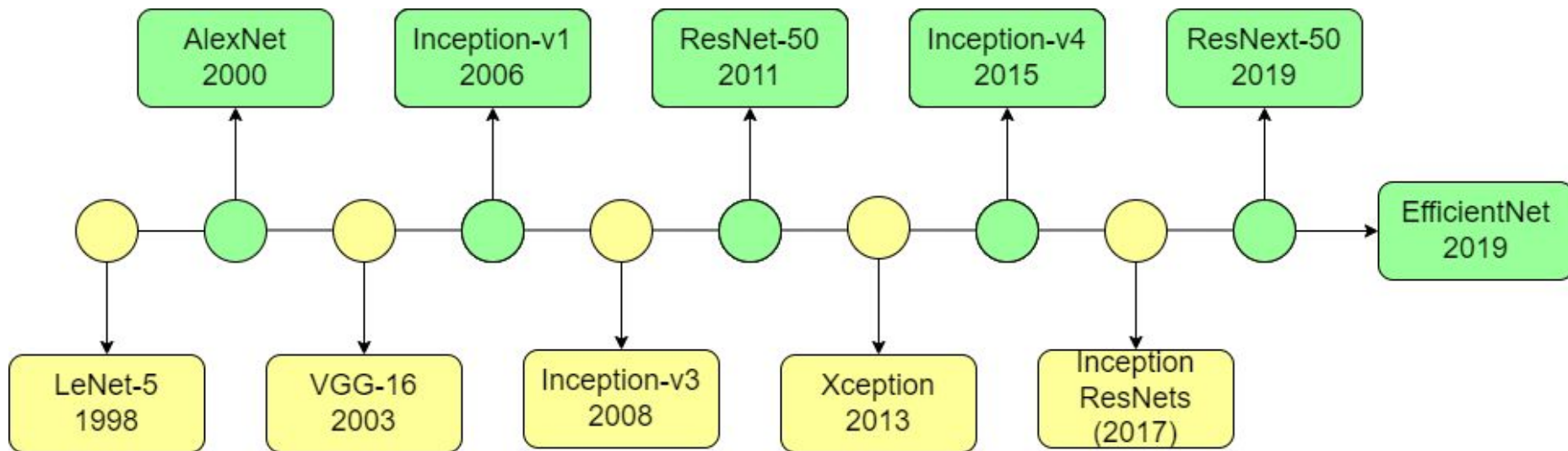


Optical
Optical
Character
Recognition
Recognition

Nội dung

1. Classification
2. Object Detection
3. Optical Character Recognition

1 - Classification



1 - Classification

Timm - [pytorch-image-models](#)

The screenshot shows the GitHub repository page for `pytorch-image-models` by `wrightman`. The repository is public and has 25.8k stars, 4.2k forks, and 298 watches. It is a Python project with 55 branches and 46 tags. The repository contains a variety of pre-trained models and scripts for image classification.

Repository Information:

- Repository: `pytorch-image-models` (Public)
- Stars: 25.8k
- Forks: 4.2k
- Watches: 298
- Sponsor: [Sponsor](#)

Repository Contents:

File/Folder	Description	Last Commit
<code>.github</code>	[doc build] Use secrets	2 weeks ago
<code>convert</code>	Move aggregation (convpool) for nest into NestLevel, cleanup and enab...	2 years ago
<code>docs</code>	Rotate changelogs, add redirects to mkdocs -> equivalent HF docs pages	5 months ago
<code>hfdocs</code>	Update installation.mdx	5 months ago
<code>results</code>	Add 0.9 imagenet and ood test set results files	last month
<code>tests</code>	fix unit test for samvit	last month
<code>timm</code>	Merge pull request #1850 from huggingface/effnet_improve_features_o...	yesterday
<code>.gitattributes</code>	Add .gitattributes	4 years ago
<code>.gitignore</code>	Add FlexiViT models and weights, refactoring, push more weights	6 months ago
<code>CONTRIBUTING.md</code>	More models w/ multi-weight support, moving to HF hub. Removing in...	2 months ago
<code>LICENSE</code>	Add Apache LICENSE file	4 years ago
<code>MANIFEST.in</code>	Move ImageNet metadata (aka info) files to timm/data/_info. Add helpe...	4 months ago

About:

PyTorch image models, scripts, pretrained weights -- ResNet, ResNeXT, EfficientNet, EfficientNetV2, NFNet, Vision Transformer, MixNet, MobileNet-V3/V2, RegNet, DPN, CSPNet, and more

huggingface.co/docs/timm

Tags:

- `pytorch`
- `resnet`
- `pretrained-models`
- `mixnet`
- `pretrained-weights`
- `imagenet-classifier`
- `distributed-training`
- `dual-path-networks`
- `cnn-classification`
- `mobilenet-v2`
- `mnasnet`
- `mobile-deep-learning`
- `mobilenetv3`
- `efficientnet`
- `augmix`
- `randaugment`
- `efficientnet-training`
- `nfnet`
- `normalization-free-training`
- `vision-transformer-models`

Readme:

Apache-2.0 license

1 - Classification

PaddleClas - PaddleClas

类别	场景	示例图	场景简介	Recall@1
动植物	花		104种花细分类识别	0.9788



人体属性分类

性别：

年龄：

朝向：

帽子：

眼镜：

手提包：

挎包：

背包：

上衣：

下衣：

：



1 - Classification

mmclassification - [mmpretrain](#)



OpenMMLab website *HOT*

OpenMMLab platform *TRY IT OUT*

[pypi](#) [v1.0.0rc8](#) [docs](#) [latest](#) [build](#) [failing](#) [codecov](#) [85%](#) [license](#) [Apache-2.0](#) [open issues](#) [7%](#) [issue resolution](#) [5 d](#)

[Documentation](#) | [Installation](#) | [Model Zoo](#) | [Update News](#) | [Reporting Issues](#)

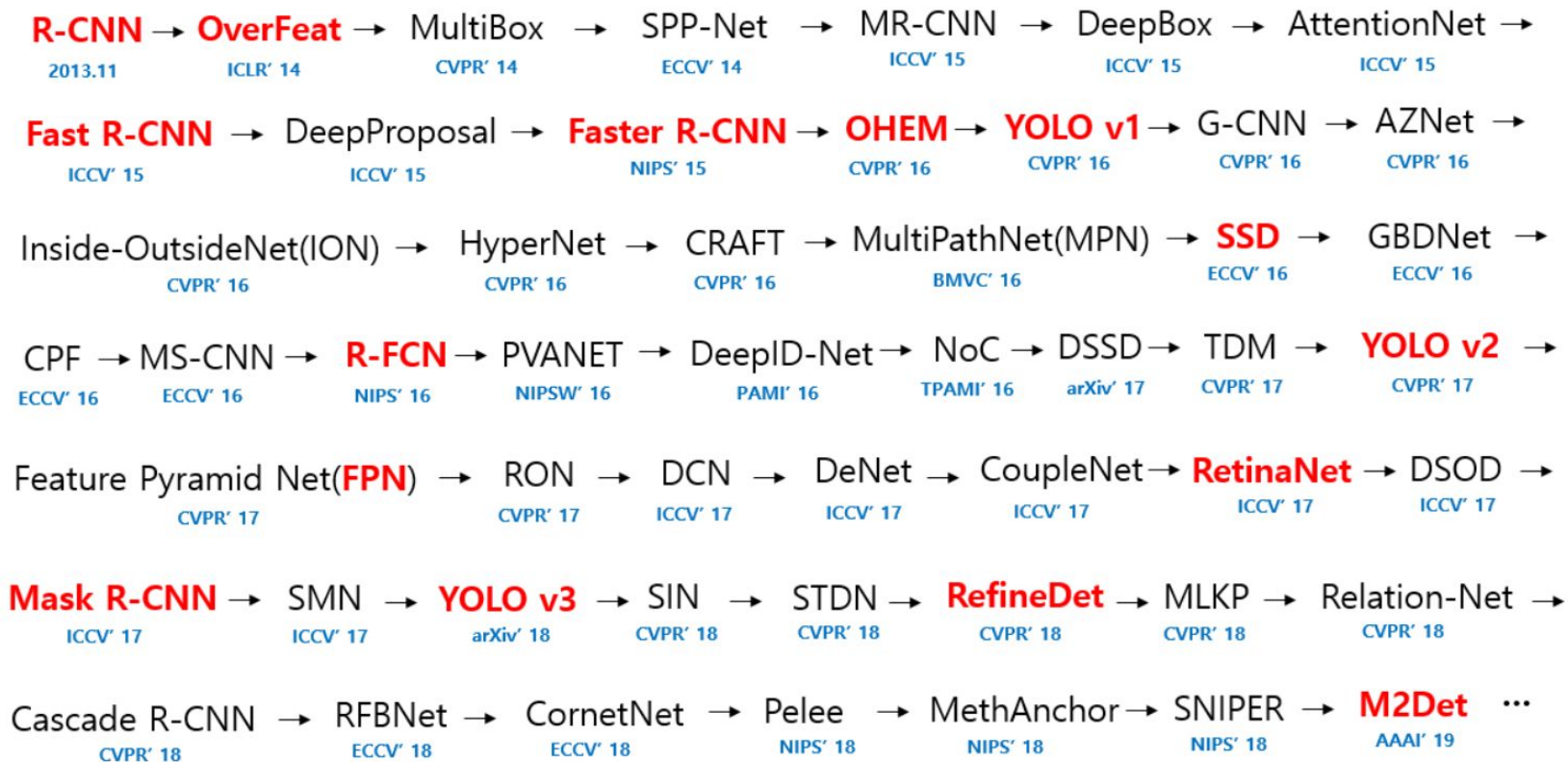
Originates from



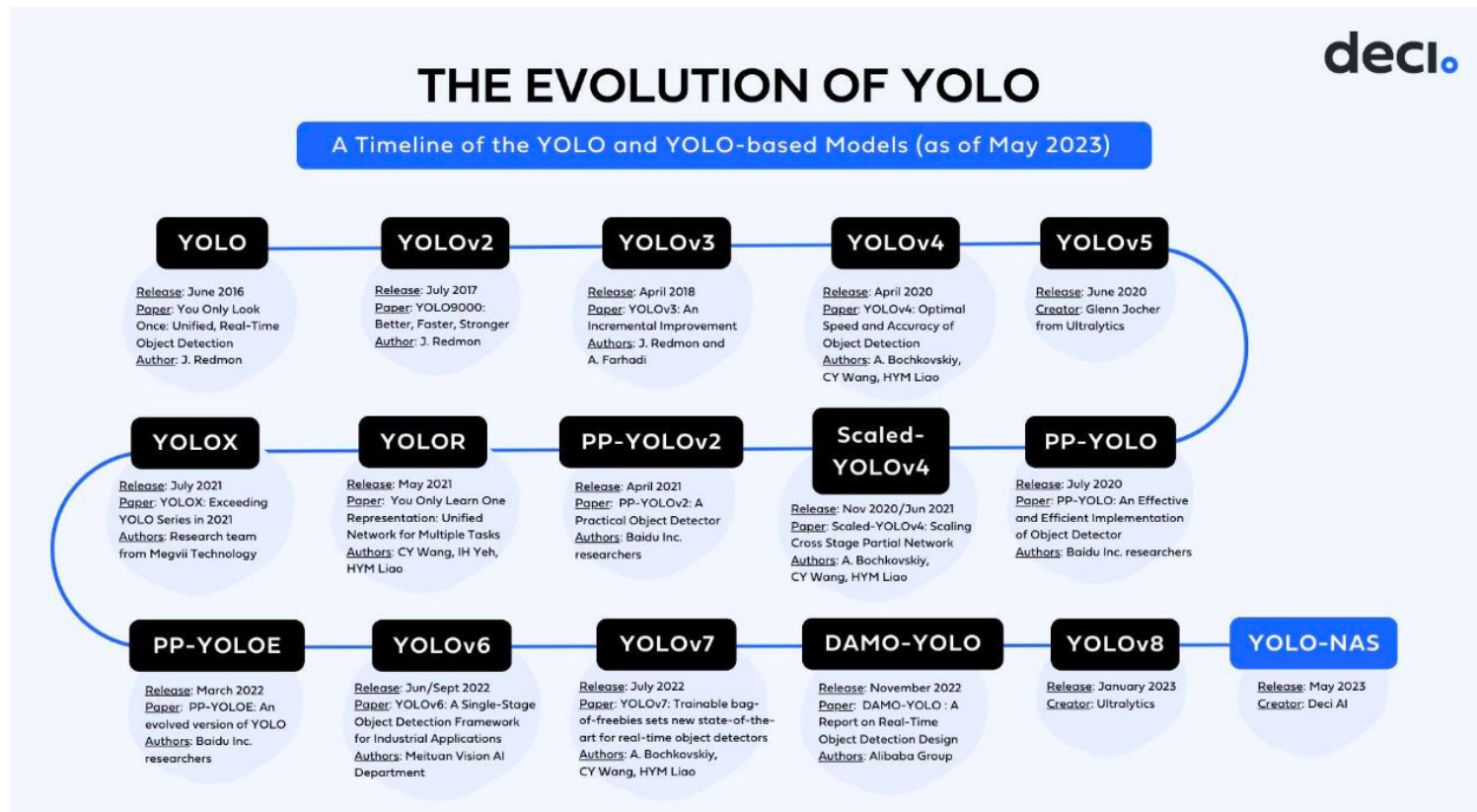
English | 简体中文



2 - Object Detection



2 - Object Detection



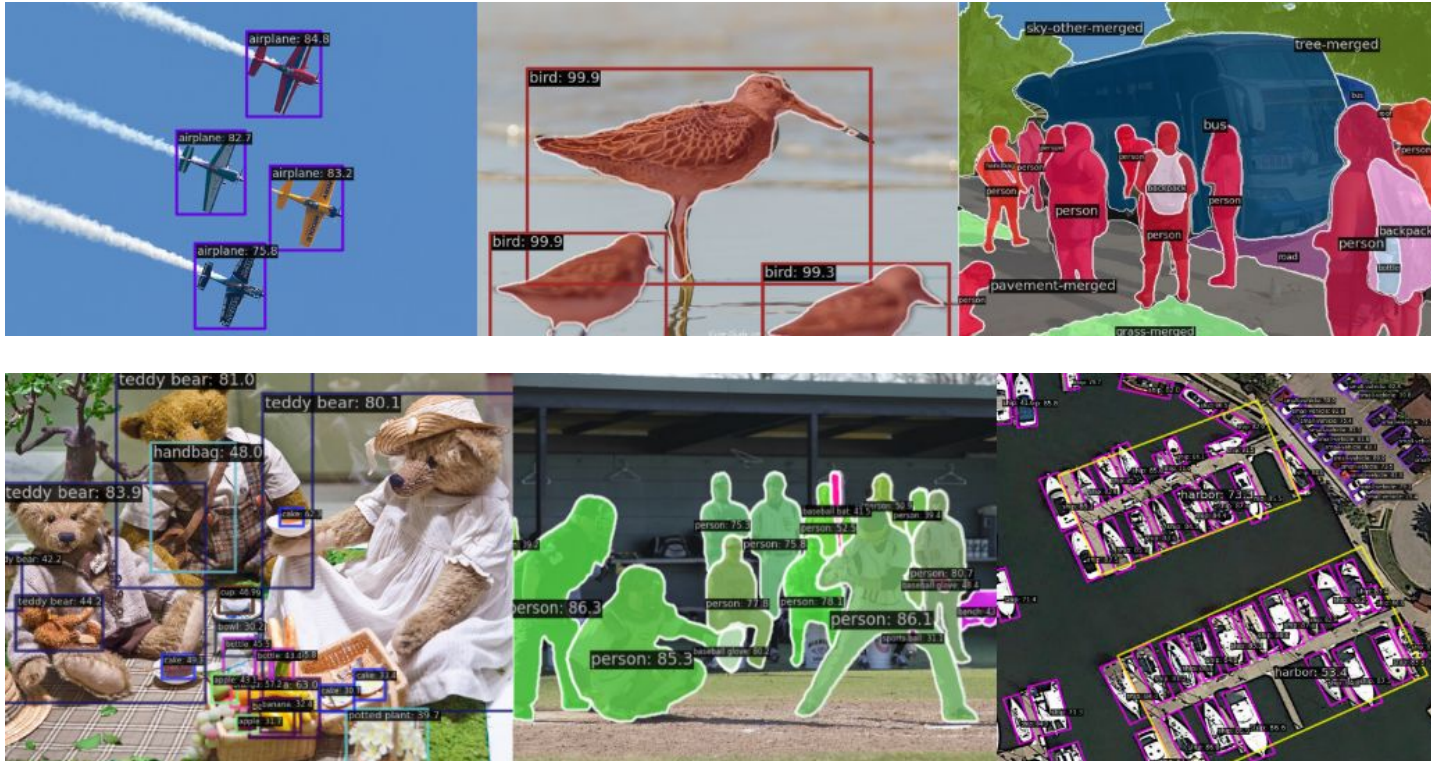
2 - Object Detection

detectron2 - [detectron2](#)



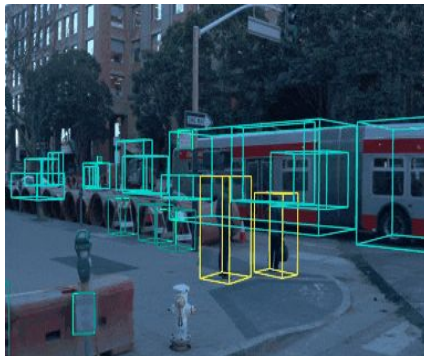
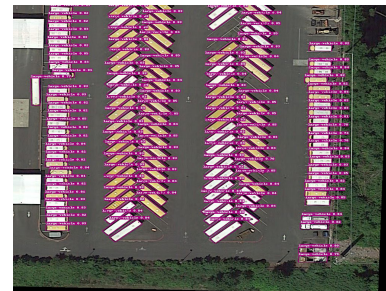
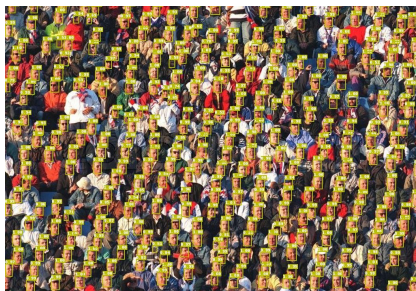
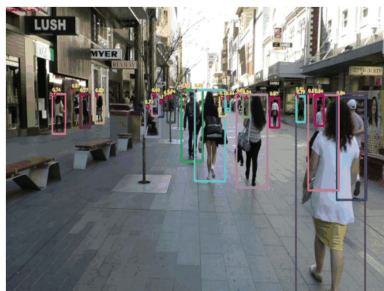
2 - Object Detection

mmdetection - mmdetection



2 - Object Detection

PaddleDetection - PaddleDetection



3 - Optical Character Recognition

Text Detection

Model	Backbone	Precision	Recall	Hmean	Download link
EAST	ResNet50_vd	88.71%	81.36%	84.88%	trained model
EAST	MobileNetV3	78.20%	79.10%	78.65%	trained model
DB	ResNet50_vd	86.41%	78.72%	82.38%	trained model
DB	MobileNetV3	77.29%	73.08%	75.12%	trained model
SAST	ResNet50_vd	91.39%	83.77%	87.42%	trained model
PSE	ResNet50_vd	85.81%	79.53%	82.55%	trianed model
PSE	MobileNetV3	82.20%	70.48%	75.89%	trianed model
DB++	ResNet50	90.89%	82.66%	86.58%	pretrained model/trained model

3 - Optical Character Recognition

Text Recognition

Model	Backbone	Avg Accuracy	Module combination	Download link
Rosetta	Resnet34_vd	79.11%	rec_r34_vd_none_none_ctc	trained model
Rosetta	MobileNetV3	75.80%	rec_mv3_none_none_ctc	trained model
CRNN	Resnet34_vd	81.04%	rec_r34_vd_none_bilstm_ctc	trained model
CRNN	MobileNetV3	77.95%	rec_mv3_none_bilstm_ctc	trained model
StarNet	Resnet34_vd	82.85%	rec_r34_vd_tps_bilstm_ctc	trained model
StarNet	MobileNetV3	79.28%	rec_mv3_tps_bilstm_ctc	trained model
RARE	Resnet34_vd	83.98%	rec_r34_vd_tps_bilstm_att	trained model
RARE	MobileNetV3	81.76%	rec_mv3_tps_bilstm_att	trained model
SRN	Resnet50_vd_fpn	86.31%	rec_r50fpn_vd_none_srn	trained model
NRTR	NRTR_MTB	84.21%	rec_mtb_nrtr	trained model
SAR	Resnet31	87.20%	rec_r31_sar	trained model

3 - Optical Character Recognition

Text Recognition

SEED	Aster_Resnet	85.35%	rec_resnet_stn_bilstm_att	trained model
SVTR	SVTR-Tiny	89.25%	rec_svtr_tiny_none_ctc_en	trained model
ViTSTR	ViTSTR	79.82%	rec_vitstr_none_ce	trained model
ABINet	Resnet45	90.75%	rec_r45_abinet	trained model
VisionLAN	Resnet45	90.30%	rec_r45_visionlan	trained model
SPIN	ResNet32	90.00%	rec_r32_gaspin_bilstm_att	trained model
RobustScanner	ResNet31	87.77%	rec_r31_robustscanner	trained model
RFL	ResNetRFL	88.63%	rec_resnet_rfl_att	trained model

3 - Optical Character Recognition

Text Super-Resolution Algorithms

Model	Backbone	PSNR_Avg	SSIM_Avg	Config	Download link
Text Gestalt	tsrn	19.28	0.6560	configs/sr/sr_tsrn_transformer_strock.yml	trained model
Text Telescope	tbsrn	21.56	0.7411	configs/sr/sr_telescope.yml	trained model

3 - Optical Character Recognition

Key Information Extraction Algorithms

Model	Backbone	Task	Config	Hmean	Download link
VI-LayoutXLM	VI-LayoutXLM-base	SER	ser_vi_layoutxlm_xfund_zh_udml.yml	93.19%	trained model
LayoutXLM	LayoutXLM-base	SER	ser_layoutxlm_xfund_zh.yml	90.38%	trained model
LayoutLM	LayoutLM-base	SER	ser_layoutlm_xfund_zh.yml	77.31%	trained model
LayoutLMv2	LayoutLMv2-base	SER	ser_layoutlmv2_xfund_zh.yml	85.44%	trained model
VI-LayoutXLM	VI-LayoutXLM-base	RE	re_vi_layoutxlm_xfund_zh_udml.yml	83.92%	trained model
LayoutXLM	LayoutXLM-base	RE	re_layoutxlm_xfund_zh.yml	74.83%	trained model
LayoutLMv2	LayoutLMv2-base	RE	re_layoutlmv2_xfund_zh.yml	67.77%	trained model

3 - Optical Character Recognition

PaddleOCR - PaddleOCR



Figure 8.8. Visual experimental results. The blue contours are boundary proposals, and the green contours are final detection boundaries.

Text 6. Experimental results on CTW-1500

Table Methods	Ext	R	P	F	FPS
TextSnake [18]	Syn	85.3	67.9	75.6	-
CSE [17]	MLT	76.1	78.7	77.4	0.38
LOMO [40]	Syn	76.5	85.7	80.8	4.4
ATTR [35]	Sy-	80.2	80.1	80.1	-
SegLink++ [28]	Syn	79.8	82.8	81.3	-
TextField [37]	Syn	79.8	83.0	81.4	6.0
MSR [38]	Syn	79.0	84.1	81.5	4.3
PSNet-1s [33]	MLT	79.7	84.8	82.2	3.9
DB [12]	Syn	80.2	86.9	83.4	22.0
CRAFT [2]	Syn	81.1	86.0	83.5	-
TextDragon [5]	MLT+	82.8	84.5	83.6	-
PAN [34]	Syn	81.2	86.4	83.7	39.8
ContourNet [36]	-	84.1	83.7	83.9	4.5
DRRG [41]	MLT	83.02	85.93	84.45	-
TextPerception [23]	Syn	81.9	87.5	84.6	-
Ours	-	80.57	87.66	83.97	12.08
Ours	Syn	81.45	87.81	84.51	12.15
Ours	MLT	83.60	86.45	85.00	12.21

Text 6. In testing, the threshold th_s is set to 0.8. Representative visible results are shown in Fig. 8 (c) and (d), which indicate our method precisely detects boundaries of long curved text with line-level. The quantitative results are listed in Tab. 6. Compared with the previous state-of-the-art methods [12, 34, 36], our approach achieves

Text 7. Experimental results on MSRA-TD500

Table Methods	R	P	F	FPS
SegLink [26]	70.0	86.0	77.0	8.9
PixelLink [4]	73.2	83.0	77.8	-
TextSnake [18]	73.9	83.2	78.3	1.1
TextField [37]	75.9	87.4	81.3	5.2
MSR [38]	76.7	87.4	81.7	-
FTSN [3]	77.1	87.6	82.0	-
LSE [30]	81.7	84.2	82.9	-
CRAFT [2]	78.2	88.2	82.9	8.6
MCN [16]	79	88	83	-
ATTR [35]	82.1	85.2	83.6	-
PAN [34]	83.8	84.4	84.1	30.2
DB [12]	79.2	91.5	84.9	32.0
DRRG [41]	82.30	88.05	85.08	-
Ours (SynText)	80.68	85.40	82.97	12.68
Ours (MLT-17)	84.54	86.62	85.57	12.31

Conclusion

Text this paper, we propose a novel adaptive boundary proposal network for arbitrary shape text detection which adopt an boundary proposal model to generate coarse boundary proposals, and then adopt an adaptive boundary deformation model combined with GCN and RNN to ner

Layout Analysis:

- 1.Text
- 2.Table
- 3.Title
- 4.Figure