

A0T9 110 VA reddifferentlytoreflectthelayoutelementcategories. .Illustratedin（),theitemsineachindexpagerow enser

tionoverunion(IOU)level[0.50:0.95]²，onthetestdata.In general,thehighmAPvaluesindicateaccuratedetectionof thelayoutelements.TheFasterR-CNNandMaskR-CNN achievecomparableresults,betterthanRetinaNet.Notice- ably,thedetectionsforsmallblocksliketitlearelesspre cise,and theaccuracydropssharplyforthetitlecategory.In Figure8,(a)and(b)illustratetheaccuratepredictionresults

# tnrotnernatase

Wealsoexaminehowourdatasetcanhelpwithareal- worlddocumentdigitizationapplication.Whendigitizing newpublications,researchersusuallydonotgeneratelarge scalegroundtruthdatatotraintheirlayoutanalysismodels. Iftheyareabletoadaptourdataset,ormodelstrainedon ourdataset,todevelopmodelsontheirdata,theycanbuild theirpipelinesmoreefficientlyanddevelopmoreaccurate models.Tothis end,we conducttwoexperiments.Firstwe examinehowlayoutanalysismodelstrainedonthemain pages can be used for understanding index pages.More- over,westudyhowthepre-trainedmodelsperformonother historicalJapanesedocuments. Table4comparestheperformanceoffiveFasterR-CNN models that are trained differently on index pages. If the modelloadspre-trainedweightsfromHJDataset,itincludes informationlearnedfrommainpages.Modelstrainedover

Thisisacoremetric developed fort the COCOcompetition[12]for evaluating the

allthetrainingdatacanbeviewedasthebenchmarks,while trainingwithfewsamples（fiveinthiscase)areconsid eredtomimicreal-worldscenarios.Givendifferenttrain ingdata,modelspre-trainedonHJDatasetperformsignifi cantlybetterthanthoseinitializedwithCOCOweights.In tuitively,modelstrainedonmoredataperformbetterthan thosewithfewersamples.Wealsodirectlyusethemodel trained on main to predict index pages without fine- tuning.Thelowzero-shotpredictionaccuracyindicatesthe dissimilarity between index and main pages. The large increaseinmAPfrom0.344to0.471afterthemodel 19

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

able3:

Table3:Detectionr eren models for each category on the test set.All values are given

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Faster R-CNN | Mask R-CNNa | 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 |

Fortraining asksaretheguad ateralregionsforeachblock.Comparedtotherectangularboundin