2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Page Object Detection from PDF Document Images by Deep Structured Prediction and Supervised Clustering

Xiao-Hui Li[1,2], Fei Yin[1,2], Cheng-Lin Liu[1,2,3]
[1]National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, P.R. China
[2]University of Chinese Academy of Sciences, Beijing, P.R. China
[3]CAS Center for Excellence of Brain Science and Intelligence Technology, Beijing, P.R. China
Email: {xiaohui.li, fyin, liucl}@nlpr.ia.ac.cn

***Abstract*—Page object detection in document images remains a challenge because the page objects are diverse in scale and aspect ratio, and an object may contain largely apart components. In this paper, we propose a hybrid method combining deep structured prediction and supervised clustering to detect formulas, tables and figures in PDF document images within a unified framework. The primitive region proposals extracted from each column region are classified and clustered with conditional random field (CRF) based graphical models which can integrate both local and contextual information. Both the unary and pairwise potentials of CRFs are formulated as convolutional neural networks (CNNs) to better exploit spatial contextual information. The CRF for clustering predicts the linked/cut label of between-region links. After CRF inference, the line regions of same class within a cluster are grouped into a page object. The state-of-the-art performance obtained on the public available ICDAR2017 POD competition dataset demonstrates the effectiveness and superiority of the proposed method.**

***Index Terms*—page object detection, supervised clustering, deep learning, structured prediction**

I. Introduction

Page Object Detection (POD) plays an important role in document image understanding and information extraction. It is aimed at locating logical objects which have high level semantics such as text lines, formulas, tables and figures in document pages. These detected objects can support many applications or be further processed by other procedures such as table reconstruction, figure classification, formula recognition and text line transcription.

However, POD suffers from some new difficulties that can not be solved by traditional layout analysis or existing deep learning based detection methods such as Faster R-CNN [1], SSD [2] and YOLO [3]. First, as shown in Fig. 1, document images are very different from natural scene images in the richness of color and texture, and the objects in document images are more likely logical clusters of small elements that are shared between different classes of objects. It makes the texture based region proposal generation methods, such as selective search and region proposal network (RPN), not appropriate for page object detection task. Second, page objects are more diverse in scale and aspect ratio than natural scene objects, while generic object detection methods are usually designed for nearly square objects. What's more, the background spaces inside page objects maybe much larger than those between objects. This makes the grouping of object elements difficult.

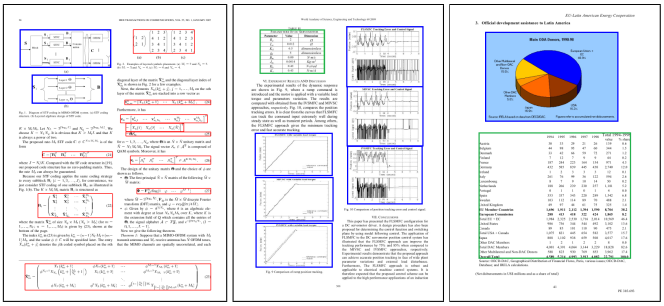


Fig. 1: Examples of PDF document images.(red: formula, green: table, blue: figure)

To overcome the above difficulties, this paper proposes a hybrid method combining deep structured prediction and supervised clustering to detect formulas, tables and figures in PDF document images within a unified framework. Before object detection, document images are segmented into column regions and then into line regions. The line regions are primitive region proposals, which are grouped into objects by classification and clustering. For incorporating spatial contexts in classification, we use conditional random fields [4](CRFs) with unary and pairwise potentials formulated as convolutional neural networks (CNNs). The CRF model is also used to learn the distance metric for clustering. After classifying and clustering primitive regions, the regions of the same classes within a cluster are merged to get page objects. Super figure regions which contain multiple figures are split with some simple heuristic rules. The proper design of the POD framework, the models of deep structured prediction and supervised clustering, and the demonstrated superior performance on the public available ICDAR2017 POD competition dataset, are the main contributions of this paper.

The rest of this paper is organized as follows. Section 2 briefly reviews related works. Section 3 introduces the

framework of the proposed method. Section 4 gives details of our CRF models for classification and clustering. Section 5 presents experimental results and section 6 draws concluding remarks.

II. Related Works

POD of different classes are usually treated as independent tasks such as table detection and formula detection.

For table detection, Hao et al. [5] use a CNN classifier to determine labels of the areas which are selected with some heuristic rules. Schreiber et al. [6] and Gilani et al. [7] use Faster R-CNN as their base detectors, and use many techniques such as data augmentation, transfer learning and domain adaption to guarantee the performance. He et al. [8] use a multi-scale, multi-task fully convolutional neural network (FCN) to predict region class probabilities at each pixel. Then tables and figures are detected with some heuristic rules and a verification network.

For formula detection, the existing methods can be divided into two categories: rule-based and learning-based. Most of the rule-based methods [9] [10] first identify special mathematical symbols, such as "+", "$\sum$", etc., and then attempt to extend formula area according to the operator domains of these symbols. The rule-based methods usually contain a set of artificial parameters, which can hardly be adaptive to various documents and formulas. By learning based methods [11] [12], document pages are first split into primitive regions such as text lines, which are classified and grouped using a classification model. Recently, Gao et al. [13] combine Convolutional Neural Network and Recurrent Neural Network model to detect formulas from PDF files, but the metadata information of PDF is also needed.

There are some works that detect multiple types of page objects in a single framework. Yi et al. [14] redesign the region proposal method and network structure on the basis of traditional CNN based object detection methods to detect page objects. In ICDAR2017 competition on page object detection [15], some submitted methods detect multiple types of page objects based on deep learning, but do not perform satisfactorily. This is mostly due to the different characteristics between document images and natural scene images.

Different from the existing methods, we detect page objects by deep structured prediction and supervised clustering. Deep structured prediction has been widely used for pixel-level segmentation tasks [16], but has not been applied to region-level prediction in document images as far as we know. Region-level prediction consumes much less computation than pixel-level prediction but relies on good segmentation of region proposals. Supervised clustering takes advantage of the association labels of the training set items [17]. A key issue in supervised clustering is to learn a distance metric to model the similarities and dissimilarities between items appropriately. The previous works of graph-based text line extraction [18] and scene text detection [19] are examples of applications of supervised learning.

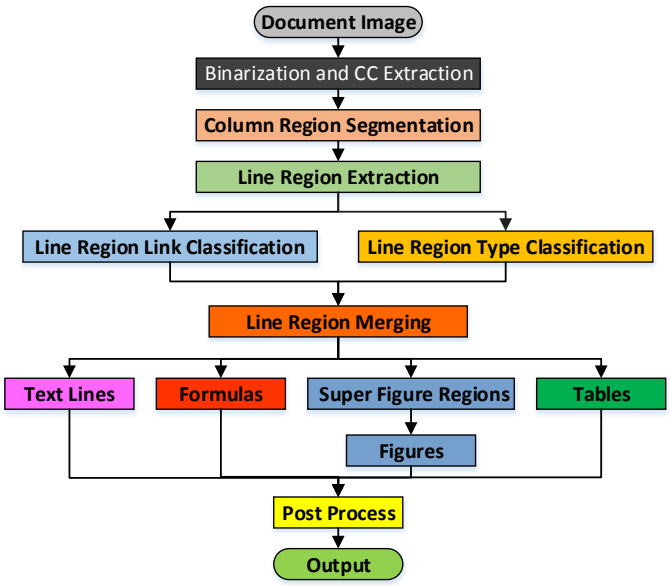


Fig. 2: Diagram of the proposed method.

III. Proposed Method

A. System Overview

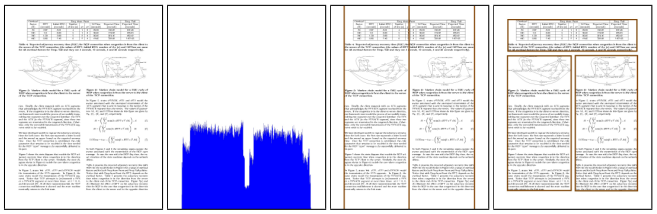
We aim to detect multiple types of page objects within a single unified framework. After binarization and connected component (CC) extraction [20], we first segment page images into column regions and then into line regions. The line regions are then classified into four class (text, formula, table and figure) using a CRF classification model. This CRF model is also used to predict whether two line regions should be linked or not, or in other words, whether they belong to the same cluster or not. After that, regions belonging to the same class and the same cluster are merged to get page objects. Super figure regions which contain multiple figures are split using some simple heuristic rules. The framework of our method is shown in Fig. 2.

B. Column Region and Line Region Segmentation

As shown in Fig. 1, distance between characters within line regions, especially formula and table regions, could be extremely large. Thus it's necessary to segment the page images into column regions before line region extraction to avoid interference from different columns.

For PDF documents, we use simple projection-based techniques for column and line region segmentation. For more complex documents such as handwritten or severely distorted documents, sophisticated region segmentation techniques are needed, but that is not the focus of this paper.

To extract column regions from document images, we first project all the CCs along vertical direction to get the projection profiles. Then we can get column separators using simple thresholding and rules. To get column regions that span across multiple columns (Fig. 3(d)), we first segment each column into line regions, then line regions which are near to each



(a) (b) (c) (d)

Fig. 3: Column region segmentation. (a) Original image; (b) Projection profile; (c) Column separators; (d) Column regions.

other and to the column separators are merged. An example of column region segmentation procedure is shown in Fig. 3.

After column region segmentation, line region extraction becomes trivial. We use a bottom-up method to extract line regions within each column region. To be specific, if two CCs in the same column region are horizontally overlapping, then they are merged into a super component. This procedure is iterated until no CCs could be merged . Then, the bounding boxes of super components correspond to line regions.

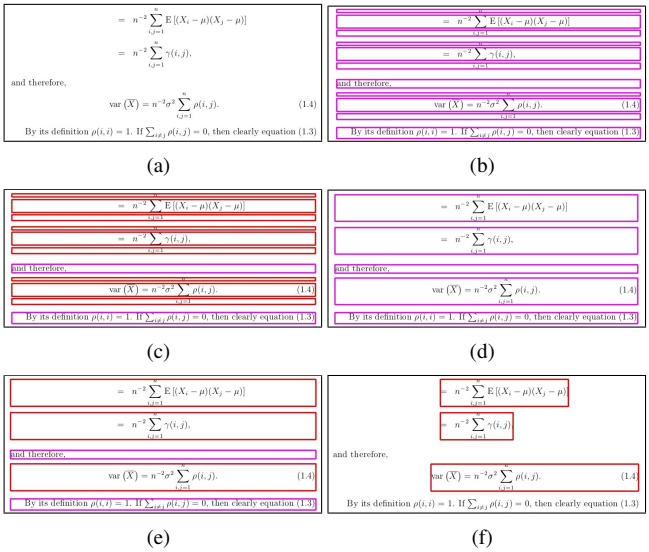
C. Line Region Type Classification

The line regions extracted are used as primitive region proposals for page objects which need to be classified and clustered. Since line regions of different classes could be extremely similar to each other, local information of each single line region is not enough for accurate classification. Thus we use a CRF based method to jointly predict the categories (text, formula, table, figure) of all line regions. In the CRF model, both unary potentials and pairwise potentials are formulated as CNNs to better exploit spacial context information. Specifically, line regions in each column region are linked as a graph, then the unary CNN takes single line region images as input and predicts the probabilities of each object class, and the pairwise CNN takes line region images and region pair images as input and predicts the probabilities of class pairs. Then an inference algorithm is carried out on the graph to jointly predict all regions' labels.

D. Line Region Link Classification

Since each page object may contain multiple proposals, it's necessary to determine which regions belong to the same object. We formulate this as a clustering problem and use CRF based supervised learning to guide the clustering. The unary and pairwise potentials of the CRF are also formulated as CNNs. However, different from that for line region type classification, the nodes and edges in the constructed graph for line region link classification are links between regions and adjacent link pairs, respectively. The unary network takes line region images and pair images as input and predicts the probability of link class ("linked" or "cut"). And the pairwise network takes line region pair images and triple images as input and predicts the probabilities of link pair class (combination of "linked" and "cut"). Then a inference algorithm is carried out to jointly predict all links' labels.

After line region type and link classification, the regions belong to the same class and the same cluster are merged to get



(a) (b) (c) (d) (e) (f)

Fig. 4: Example of line region type and link classification. (a) Original image; (b) Line regions; (c) Type classification; (d) Link classification; (e) Region merging; (f) Page objects.

page object regions (text lines, formulas, tables and figures). An example of line region type and link classification is shown in Fig. 4.

E. Super Figure Region Spliting

It's worthy noting that some figure regions may consist of multiple horizontally arranged figures (Fig. 8(3)(5)), we call these figure regions *super figure regions*, and they should be split into smaller figures in vertical direction. For that purpose, we first split the super figure regions into sub figures through vertical projection, then the sub figures are merged horizontally. Although the merging of sub figure regions can also be treated as supervised clustering problem, we find some simple heuristic rules can handle this problem just as well, such as two sub figures' height ratio, width ratio, horizonal overlap and dist, etc.

F. Object Class Verification

Since for classification, we normalize line region images to a fixed size of 32x640 pixels, some line regions whose original heights are extremely large may not be classified accurately. To solve this problem, we use a verification CNN to reclassify those large objects detected (height larger than 100 pixels). The structure of our verification CNN is illustrated in Fig. 5. The input of the network are normalized object images with size of 64x64 pixels and the output are probabilities of them belonging to each class (text, figure, table and figure, K=4).

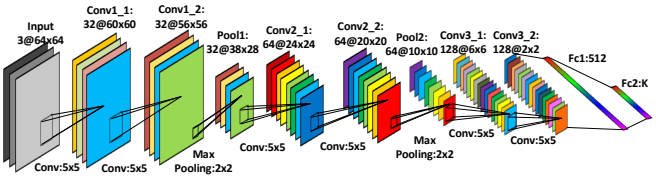


Fig. 5: Structure of the verification CNN

IV. Classification Model

A. CRFs Formulation

In this section, we give details about our CRFs classification model. A second order CRF model can be formulated as:

$$P(y|x;w) = \frac{1}{Z(x;w)} exp[-E(y,x;w)], \tag{1}$$

where

$$Z(x;w) = \sum_{y} exp[-E(y,x;w)] \tag{2}$$

is the partition function, and

$$E(y,x;w) = \sum_{p \in N_U} U(y_p, x_p, ; w_U) + \sum_{(p,q) \in S_V} V(y_p, y_q, x_{pq}; w_V) \tag{3}$$

is the energy function. U is unary potential function which represents the cost that node p takes the label y_p, N_U is the set of nodes for potential U and w_U is the set of parameters of the U. Likewise, V is pairwise potential function which represents the cost that node p takes the label y_p and node q takes the label y_q simultaneously, S_V is the set of edges for the potential V and w_V is the set of parameters of V. In our work, both unary and pairwise potentials are formulated as CNNs named as Unary-Net and Pairwise-Net, respectively. This leads to the following formulation of U and V:

Unary Potentials:

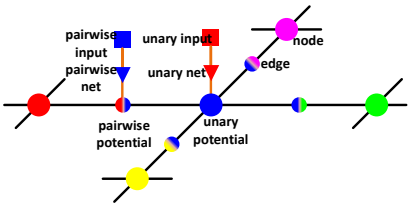
$$U(y_p, x_p; w_U) = \sum_{k=1}^{K} -\lambda_k \delta(k = y_p) z_{p,k}(x; w_U), \tag{4}$$

where $z_{p,k}$ is the output value of Unary-Net which corresponds to the p-th node and the k-th class. λ_k is the coefficient of $z_{p,k}$. Here K is the class number and the output number of Unary-Net.

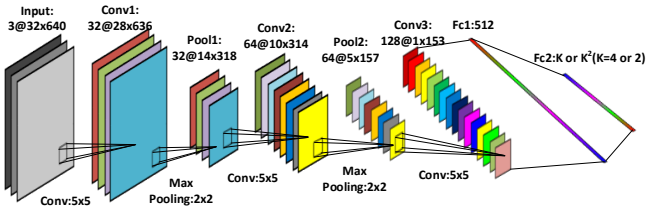
Pairwise Potentials:

$$V(y_p, y_q, x_{p,q}; w_V) = \sum_{k_p=1}^{K} \sum_{k_q=1}^{K} -\lambda_{k_p,k_q} \delta(k_p = y_p) \cdot \delta(k_q = y_q) z_{p,k_p,q,k_q}(x; w_v), \tag{5}$$

where z_{p,k_p,q,k_q} is the Pairwise-Net output which corresponds to the node pair (p,q) when they take the label pair (k_p, k_q). It measures the compatibility of the label pair (y_p, y_q) given the input pairwise features. λ_{k_p,k_q} is the coefficient of z_{p,k_p,q,k_q}. The output number of Pairwise-Net is K^2, where K is the number of classes. The structure of our CRF model is illustrated in Fig. 6(a). Without loss of generality, we show 4 neighbors of the central node, in fact, it can have any number of neighbors if needed.

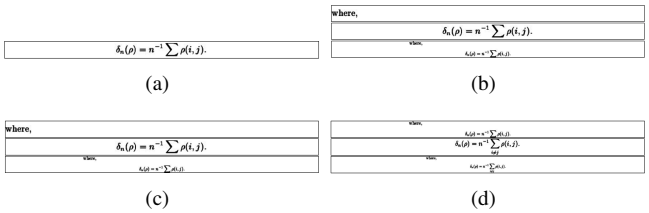


(a)



(b)

Fig. 6: Structure of our CRF and its potential CNN.



(a) (b) (c) (d)

Fig. 7: Input of CNNs. (a)(b) Input of Unary-Net and Pairwise-Net for line region type classification; (c)(d) Input of Unary-Net and Pairwise-Net for line region link classification

B. Unary and Pairwise Net

Fig. 6(b) shows the base structure of our Unary-Net CNN and Pairwise-Net CNN for line region type classification and line region link classification. The only difference is that their input layer and output layer are different which has been introduced in last section. Examples of input for each CNN are shown in Fig. 7.

C. Inference and Learning

We adopt the maximum a posteriori (MAP) strategy to predict the labels of line regions given a new document. It is to find the most likely labels of the line regions given their features. MAP inference of CRFs can be formulated as the following optimization problem:

$$y^* = arg \max_{y} P(y|x;w) \tag{6}$$

To solve this problem, we apply a widely used approximate inference method named loopy belief propagation [21].

The purpose of learning is to find the best parameters of CRFs from the given training set. The parameters of our CRFs include Unary-Net's weights w_U and Pairwise-Net's weights w_V and a combination coefficient vector λ (dimention: $K + K^2$) of U and V. w_U and w_V are learned using the SGD method. Then they are fixed and λ is learned using the Pseudo Likelihood method [22].

V. Experimental Results

A. Dataset

We conducted experiments on the ICDAR2017 POD competition dataset [15] (abbreviated as POD2017 dataset), which is a public available dataset for page object detection. This dataset consist of 2,417 English document page images selected from 1,500 scientific papers of CiteSeer, in which 1,600 pages are used as training set and 817 pages are used as test set. It exhibits a considerable variety in both page layout styles and object styles, including single-cloumn pages, two-column pages, multi-column pages and various kinds of formulas, tables, graphics and figures [15]. Some images of POD2017 dataset are shown in Fig. 1.

TABLE I: Page object detection result of our method on ICDAR2017 POD dataset.

Method	Class	IOU=0.8				IOU=0.6			
		Precision	Recall	F1	AP	Precision	Recall	F1	AP
NLPR-PAL	Formula	0.888	0.916	0.902	0.816	0.901	0.929	0.915	0.839
	Table	0.958	0.943	0.951	0.911	0.968	0.953	0.960	0.933
	Figure	0.892	0.904	0.898	0.805	0.920	0.933	0.927	0.849
Proposed	Formula	0.921	0.944	0.932	0.863	0.930	0.953	0.942	0.878
	Table	0.965	0.953	0.959	0.923	0.974	0.962	0.968	0.946
	Figure	0.921	0.913	0.917	0.854	0.948	0.940	0.944	0.896

TABLE II: Comparison of our methods with the state-of-the-art (IOU=0.8).

Method	F1-m (IOU=0.8)				AP (IOU=0.8)			
	Formula	Table	Figure	Average	Formula	Table	Figure	Aveage (mAPs)
NLPR-PAL	0.902	0.951	0.898	0.917	0.816	0.911	0.805	0.844
icstpku	0.841	0.763	0.708	0.770	0.815	0.697	0.597	0.703
FastDetectors	0.636	0.896	0.616	0.717	0.427	0.884	0.365	0.559
Vislnt	0.241	0.826	0.643	0.570	0.117	0.795	0.565	0.492
SOS	0.218	0.796	0.656	0.557	0.109	0.737	0.518	0.455
UITVN	0.200	0.635	0.619	0.485	0.061	0.695	0.554	0.437
Matiai-ee	0.065	0.776	0.357	0.399	0.005	0.626	0.134	0.255
HustVision	0.042	0.115	0.132	0.096	0.293	0.796	0.656	0.582
Proposed	**0.932**	**0.959**	**0.917**	**0.936**	**0.863**	**0.923**	**0.854**	**0.880**

TABLE III: Comparison of our methods with the state-of-the-art (IOU=0.6).

Method	F1-m (IOU=0.6)				AP (IOU=0.6)			
	Formula	Table	Figure	Average	Formula	Table	Figure	Aveage (mAPs)
NLPR-PAL	0.915	0.960	0.927	0.934	0.839	0.933	0.849	0.874
icstpku	0.859	0.813	0.763	0.811	0.849	0.753	0.679	0.760
FastDetectors	0.675	0.921	0.638	0.745	0.474	0.925	0.392	0.597
Vislnt	0.605	0.921	0.824	0.783	0.524	0.914	0.781	0.740
SOS	0.604	0.937	0.847	0.796	0.537	0.931	0.785	0.751
UITVN	0.377	0.782	0.775	0.645	0.193	0.924	0.786	0.634
Matiai-ee	0.337	0.865	0.562	0.588	0.116	0.781	0.325	0.407
HustVision	0.078	0.132	0.164	0.124	0.854	0.938	0.853	0.882
Proposed	**0.942**	**0.968**	**0.944**	**0.951**	**0.878**	**0.946**	**0.896**	**0.907**

B. Evaluation

Our experimental results are measured with the precision, recall, F1 metric, Average Precision (AP) for each class of object (formula, table, figure), average F1 metric and mean of AP (mAPs) for all classes with two Intersection Over Union (IOU) threshold 0.6 and 0.8. These are the standard evaluation metrics adopted by the ICDAR2017 POD competition.

C. Training Details

All of our classification models are trained with the training set of POD2017 dataset and we don't use any other dataset for pre-training. To overcome the problem of serious class unbalance, we flip the line region images of formula, table and figure horizontally and vertically for data enhancement.

Our method is implemented in C++ and all the experiments are performed on a computer with an Intel Core i7-4790 CPU (3.60GHz) except that our CNNs are trained on a GPU server with Titan GTX 980. For CNNs and CRFs used in our models, we use the open source library *Caffe* and *OpenGM* for implementation, respectively.

D. Results and Analysis

Our baseline system NLPR-PAL submitted to the ICDAR2017 POD Competition [15] which is based on CC classification, rule based region proposal generation, CNN based region image classification has already won the first place on most of the metrics in ICDAR2017 POD competition. In this paper, we modify the detection of page objects as a supervised clustering problem and use machine learning to guide the generation of object regions. Compared with our baseline system NLPA-PAL which use plenty of elaborate heuristic rules, the proposed method in this paper is more compact, efficient and accurate.

Table I shows the comparison of the proposed method and NLPR-PAL system. As we can see, the proposed method achieves better results on all the metrics for all object classes, especially for formulas. This is mostly because formulas are usually composed of several line regions such as superscripts, subscripts, main bodies, numerators and denominators. It's hard to design simple rules that can handle all these conditions. On the contrary, CRFs based line region link classification can automatically learn which line regions belong to the same objects, thus increasing the detection accuracy. The same situation can be found for figures composed of several sub line regions and sparse line tables which have no complete frame lines. Some POD result examples are shown in Fig. 8.

Comparison of our method with the state-of-the-art methods from the ICDAR2017 POD competition are shown in Table II and Table III with IOUs take value of 0.8 and 0.6, respectively. The proposed method surpasses all existing methods on all metrics. Most of the methods in ICDAR2017 POD competition use deep learning techniques but their results are not very satisfactory. This is possibly because they didn't handle the

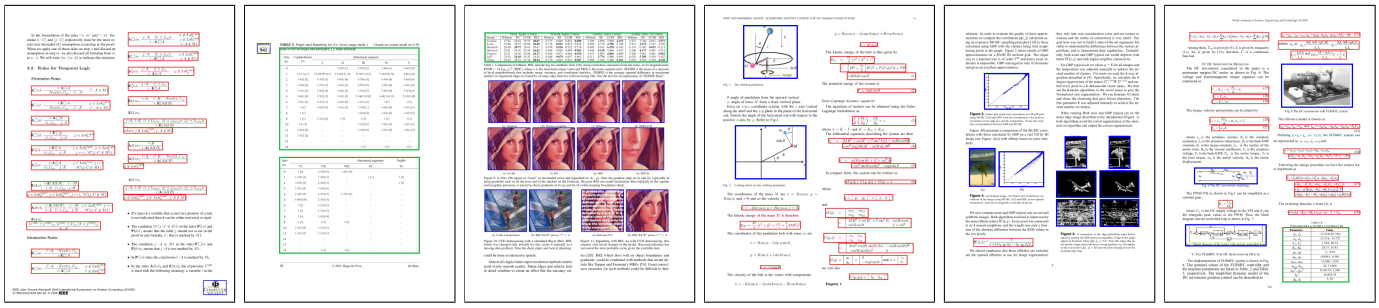


Fig. 8: Page object detection result samples.(red: formula, green: table, blue:figure)

significant difference between document images and natural scene images appropriately, and directly using the detection frameworks designed for natural scene images to detect page objects may lead to degraded performance. How to adapt those detection frameworks to page object detection is still a problem worthy of study.

VI. Conclusion

In this paper, we propose a hybrid method combining deep structured prediction and supervised clustering for page object detection in PDF document images. The primitive region proposals, line regions, are classified and clustered with CRF based classification models whose unary and binary potentials are both formulated as CNNs to better exploit spatial context information. Our method achieves state-of-the-art performance on the ICDAR2017 POD competition dataset on all the evaluation metrics for all object classes. This demonstrates the effectiveness and superiority of our method.

The proposed framework can be applied to more complex and distorted documents other than PDF documents if taking different primitive regions (such as CCs) instead of line regions for classification and clustering.

Acknowledgment

This work has been supported by the National Natural Science Foundation of China (NSFC) Grants 61721004, 61411136002, 61573355 and 61733007.

References

[1] S.-Q. Ren, K.-M. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European Conference on Computer Vision*. Springer, 2016, pp. 21–37.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[4] J. Lafferty, A. McCallum, F. Pereira *et al.*, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the 18th International Conference on Machine Learning, ICML*, vol. 1, 2001, pp. 282–289.

[5] L. Hao, L. Gao, X. Yi, and Z. Tang, "A table detection method for pdf documents based on convolutional neural networks," in *Proceedings of the 12th IAPR Workshop on Document Analysis Systems*, 2016, pp. 287–292.

[6] S. Schreiber, S. Agne, I. Wolf, A. Dengel, and S. Ahmed, "Deepdesrt: Deep learning for detection and structure recognition of tables in document images," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[7] A. Gilani, S.-R. Qasim, I. Malik, and F. Shafait, "Table detection using deep learning," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[8] D.-F. He, S. Cohen, B. Price, D. Kifer, and C. L. Giles, "Multi-scale multi-task fcn for semantic page segmentation and table detection," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[9] U. Garain, "Identification of mathematical expressions in document images," in *Proceedings of the 10th International Conference on Document Analysis and Recognition*, 2009, pp. 1340–1344.

[10] N. Liu, D.-X. Zhang, X. Xu, L. Guo, L.-J. Chen, W.-J. Liu, and D.-F. Ke, "Robust math formula recognition in degraded chinese document images," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[11] S. Chowdhury, S. Mandal, A. K. Das, and B. Chanda, "Automated segmentation of math-zones from document images," in *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003, pp. 755–759.

[12] Y. Liu, K. Bai, and L. Gao, "An efficient pre-processing method to identify logical components from pdf documents," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2011, pp. 500–511.

[13] L.-C. Gao, X.-H. Yi, Y. Liao, Z.-R. Jiang, Z.-Y. Yan, and Z. Tang, "A deep learning-based formula detection method for pdf documents," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[14] X.-H. Yi, L.-C. Gao, Y. Liao, X. Zhang, R.-T. Liu, and Z.-R. Jiang, "Cnn based page object detection in document images," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[15] L.-C. Gao, X.-H. Yi, Z.-R. Jiang, L.-P. Hao, and Z. Tang, "Icdar2017 competition on page object detection," in *Proceedings of the 14th International Conference on Document Analysis and Recognition*, 2017.

[16] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[17] T. Finley and T. Joachims, "Supervised k-means clustering," *Faculty of Computing & Information Science*, 2008.

[18] F. Yin and C.-L. Liu, "Handwritten chinese text line segmentation by clustering with distance metric learning," *Pattern Recognition*, vol. 42, no. 12, pp. 3146–3157, 2009.

[19] Y.-F. Pan, X.-W. Hou, and C.-L. Liu, "A hybrid approach to detect and localize texts in natural scene images," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 800–813, 2011.

[20] F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer Vision & Image Understanding*, vol. 93, no. 2, pp. 206–220, 2004.

[21] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters." in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, 2003, pp. 900–906.

[22] S. Kumar and M.Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2003, pp. 1150–1157.