Printed/Handwritten Texts and Graphics Separation in Complex Documents using Conditional Random Fields

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*—In this paper we propose a structured prediction based system for text/non-text classification and printed/handwritten texts separation at connected component (C-C) level in complex documents. We formulate the separation of different elements as joint classification problems and use conditional random fields (CRFs) to integrate both local and contextual information for improving the classification accuracy. Both our unary and pairwise potentials are formulated as neural networks for better exploiting contextual information. Considering the different properties in text/non-text classification and printed/handwritten texts separation, we use multilayer perception (MLP) and convolutional neural network (CNN) for potentials, respectively. To evaluate the performance of the proposed method, we provide a test paper document database named TestPaper1.0, which can be used for many other tasks as well. Our method achieve impressive results for both tasks on TestPaper1.0 dataset. Moreover, even with very shallow CNNs as potentials, our method achieves state-of-the-art performance for writing type (printed/handwritten) separation on the highly heterogeneous Maurdor dataset, surpassing Maurdor2013 and Maurdor2014 campaign winners. This demonstrates the effectiveness and superiority of our method.**

***Keywords*-text/non-text, printed/handwritten, document understanding, structured prediction**

I. Introduction

Automatic analysis and recognition of test paper documents finds important applications in modern education. It poses a challenge due to the complex document layout mixing texts and non-texts (graphics, table forms, stains, etc.), printed texts and handwritten annotations, different scripts and languages. There is no single algorithm that can handle all the contents simultaneously. A practical strategy is to separate the contents into different classes and use corresponding recognition algorithms in further process. A test paper document sample is shown in Fig. 1.

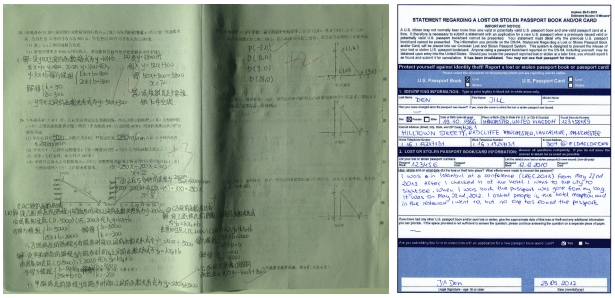


Figure 1: Test paper (left) and Maurdor (right) samples

Text/non-text classification [1] and printed/handwritten texts separation [2] are two crucially important tasks in document analysis field. Although many methods [1-15] have been proposed, they are still not solved satisfactorily because of the large variation of different documents. As shown in Fig. 1, some text/non-text strokes or printed/handwritten texts can be extremely similar to each other. Since local features on connected components are not discriminative enough, contextual information such as spatial relationship and temporal relationship become crucially important for the classification tasks.

In this paper, we handle these two tasks within a single framework by using conditional random fields (CRFs), whose unary and pairwise potentials are both formulated as neural networks to better exploit contextual information. Given a test paper document, after binarization and connected components (CCs) extraction, the foreground CCs are first classified into three classes: text, graphics and table; then all the texts are classified into printed and handwritten. For potentials in text/non-text classification and printed/handwritten texts separation, we use multilayer perception (MLP) and convolutional neural network (CNN), respectively, to adapt to the different property of difficulty in two problems. To achieve high separation performance, we also design and validate a feature set which can be used both for text/non-text classification and printed/handwritten separation. To evaluate the performance of our proposed method, we provide a test paper document database named TestPaper1.0, which can be used for many document analysis tasks and will be made public for research purpose.

The impressive results achieved on the TestPaper1.0 database demonstrate the effectiveness and superiority of our system. Moreover, even use very shallow CNNs as potentials, our method achieves state-of-the-art results for writing type separation on the highly heterogeneous Maurdor dataset, surpassing Maurdor2013 and Maurdor2014 campaign winners. Specifically, we achieved 99.95% correct for text/non-text separation on TestPaper1.0, 99.33% and 99.12% correct for printed/handwritten separation on TestPaper1.0 and Maurdor, respectively.

The rest of this paper is organized as follows. Section 2 briefly reviews related works. Section 3 gives details of the proposed method. Section 4 presents the experimental results, and Section 5 draws concluding remarks.

978-1-5386-3346-5/18 $31.00 © 2018 IEEE
DOI 10.1109/DAS.2018.19

II. RELATED WORKS

For text/non-text classification in online handwritten documents, Jain et al. [3] and Indermuhle et al. [4] adopted isolated stroke classification strategy using only local features. To integrate more contextual information, Zhou et al. [5] proposed a Markov random field (MRF) based method. Delaye et al. [1] and Ye et al. [6] utilize contexts better by using CRF with multiple interactions between strokes or joint training of CRF and neural network. Indermuhle et al. [7] and Van et al. [8] use bidirectional long short-term memory (BLSTM) to exploit global and local contexts, and by using ensemble classifiers [8], new state-of-the-art accuracy 98.30% was achieved on the IAMonDo database.

Text/non-text classification in offline documents are more challenging since the lack of temporal information. In the work of Vidya et al. [9], document images are separated into text and non-text regions by using Simplified Fuzzy ARTMAP (SFAM) classifier. Ahmed et al. [10] localize and distinguish text components touching with graphics using Speeded Up Robust Features (SURF). In the field of text localization in natural scene images, Pan et al. [11] proposed a CRF based method to filter out non-text components from text ones.

Printed/handwritten texts separation can be done at multiple levels such as block level, text line level, word level, CC level and pixel level. The technique in [12] uses morphological and geometrical analysis to separate printed/handwritten text blocks, but it cannot distinguish single words in the same block. Haboubi et al. [13] and Saidani et al. [2] use some carefully designed features to separate Arabic/Latin and printed/handwritten texts at word level on a combined database of IAM and IFNENIT which contain 1,320 words. Peng et al. [14] use a MRF based two-step method to classify printed/handwritten texts at two levels: patch level and then pixel level, and they reported patch-level accuracy 95.52% and pixel-level accuracy 86.82%. Seuret et al. [15] treat printed/handwritten texts separation as a segmentation problem and handle it at pixel level using MLP classifier on extracted features. A pixel-level accuracy of 96.10% was achieved with their method.

In recent years, the convolutional neural network (CCN) has been widely used in fields of pattern recognition and computer vision, and the CRF is often used to exploit contextual information in many tasks, e.g. semantic image segmentation [16] [17]. In work [16], Zheng et al. formulate CRF as RNN and integrate it with CNN so that it can be trained end-to-end. Lin et al. [17] formulate both unary and pairwise potentials as CNNs and jointly train the parameters of CNN and CRF with piecewise training.

III. PROPOSED METHOD

A. System Overview

In our work, text/non-text classification and printed/handwritten texts separation are handled within a single framework by classification at connected component (CC) level. After binarization and CC extraction, the processing flow contains two major stages: text/non-text classification and printed/handwritten separation. In the first stage, we classify all the CCs into three classes: text, graphics and table; and in the second stage, we separate the texts into printed and handwritten. After that, we use a k-nearest neighbor (kNN) classifier to process the noise CCs which were separated from the others before classification. The framework of our method is illustrated in Fig. 2.

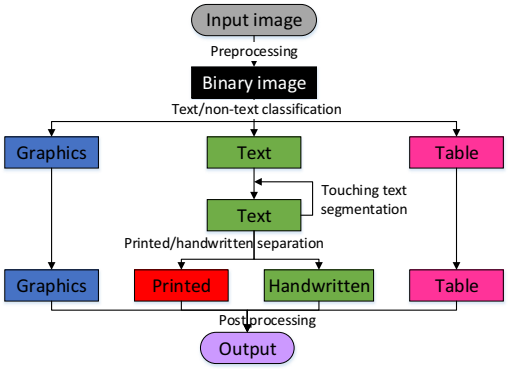


Figure 2: Framework of our proposed method

B. Preprocessing

To sepatate foreground pixels in low-quality document images, we designed a robust yet simple binarization algorithm based on contour extraction and local binarization. Our binarization algorithm can be divided into three steps. In the first step, we use gray level transformation [18] based on local intensity information to eliminate the influence of bad illumination. Then the contours of foreground objects are extracted with the contour extraction methods proposed in [19], the noise contours are excluded with some heuristic rules. After that, we use local OTSU binarization method to extract foreground pixels inside local windows centered at each contour pixel. Note that each pixel can be covered by multiple local windows, thus it will be binarized multiple times, so we use voting strategy to get its final binarization result. After binarization, CCs are extracted with the method proposed in [20].

C. Classification Model

1) Problem Definition: Given a set of labeled documents $S = \{(x^{(i)}, y^{(i)})|i = 1, ..., N\}$, in which each document is represented by a set of connected components $x^{(i)} = \{x_c^{(i)}|c = 1, ..., C_i\}$ and a set of labels $y^{(i)} = \{y_c^{(i)}|c = 1, ..., C_i\}$. Each $x_c^{(i)}$ has one associated label $y_c^{(i)}$, which is one of K classes. The task is to learn a model from the training data set S that can predict test set with good performance.

2) Model Formulation: CRF is a powerful undirected discriminative probabilistic graphical model which has been widely used for various structural prediction tasks such as semantic image segmentation [16] and sequence data labeling [21]. A second order CRF model can be defined as:

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

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 nodes set for potential U and w_U is 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 parameters of V. Each node is connected with its k nearest neighbors (kNN). Due to the discriminative nature of CRF, U and V can take any form of functions. In our work, both unary and pairwise potentials are formulated as neural networks named as Unary-Net and Pairwise-Net, respectively. This leads to the following formulation of U and V:

Unary Potentials: We formulate unary potential function as follows:

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

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 weight coefficient of $z_{p,k}$. Here K is the classes number and the output number of Unary-Net whose input is based on each single CC.

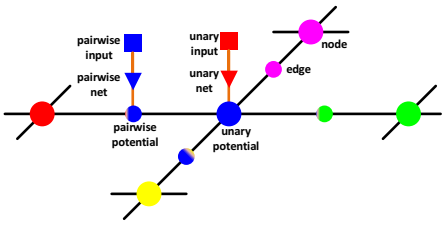
Pairwise Potentials: We formulate pairwise potential function as follows:

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

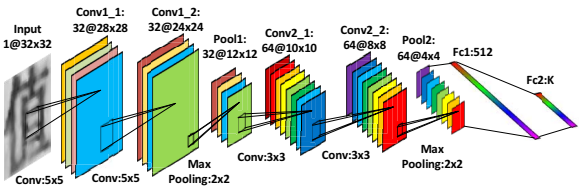
where z_{p,k_p,q,k_q} is the Pairwise-Net output which corresponds to the node pair (p,q) when they take the label pair (k_p, k_q). It measures the compatibility of the label pair (y_p, y_q) given the input pairwise features. λ_{k_p,k_q} is the weight coefficient of z_{p,k_p,q,k_q}. The output number of Pairwise-Net is K^2, where K is the number of classes. The input of Pairwise-Net is based on each CC pair. The largest difference between our neural networks based pairwise potentials with Potts model potentials is that ours can formulate not only neighborhood compatibility but also neighborhood non-compatibility, thus it can avoid excessive smoothness. The structure of our CRF model is illustrated in Fig.3(a). Without loss of generality, we only show 4 neighbors of the central CC, in fact, it can have any number of neighbors if needed.

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

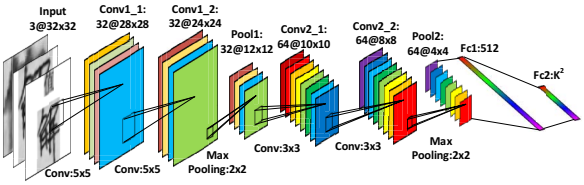
$$\begin{aligned} y^* &= arg\max_{y} P(y|x;w) \\ &= arg\max_{y} \frac{1}{Z(x;w)} exp[-E(y,x;w)] \end{aligned} \tag{6}$$



(a)



(b)



(c)

Figure 3: Structure of our CRF and its Unary-Net CNN and Pairwise-Net CNN.

To calculate $Z(x;w)$ we need to know the distribution of y for each given x. Usually the configuration number of y is exponential to the nodes number in CRF. Since there are usually hundreds even thousands of CCs in a document image and the structure of our CRF doesn't form a tree, hence the exact inference is impossible. To address this problem we apply a widely used approximate inference method named loopy belief propagation [22] which is a well known message passing algorithm and can be referenced in plenty of literatures so we don't give its details in this paper.

4) Learning: the purpose of learning is to find the best parameters of CRF from the given training set. Since the structure of our CRF doesn't form a tree, the exact MAP learning is impossible. Instead, we need to learn it by approximate learning. The parameters of our CRF 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 [23].

D. Text/non-text Classification

For Text/non-text Classification, the Unary-Net and Pairwise-Net are both formulated as MLP. The input of Unary-Net are unary features extracted from each single CC (Table I) and the input of Pairwise-Net are the concatenations of unary features of each connected CC pair. These unary features include some previous frequently used features by other works and some newly designed features by ourselves. Each feature is normalized to mean 0 and standard deviation 1 on the training set of CCs. The

Table I: Unary features extracted from each CC.

#	description	dim
1	Height, width, area, and aspect ratio of the bonding box of CC	4
2	Pixel number of CC	1
3	Duty factor of CC (Pixel number / area)	1
4	Hole number of CC	1
5	Mean run length of CC	1
6	Variance of run length of CC	1
7	Perimeter: Pixel number of CC's contour	1
8	Length of straight contour	1
9	Straight contour length / perimeter	1
10	Harris corner point number of CC's image	1
11	Gabor feature: 8 direction, each direction's mean and standard deviation of filtered gray level	16
12	Horizontal and vertical run length histogram, each direction has 6 bins	12
13	Two scale hog feature of CC's normalized image (32x32), 32x32 cell or 16x16 cell, nine direction	45
14	Horizontal and vertical binary level co-occurrence feature, the distance between pixel pairs is 1,2,3,4,5,6	12
15	Contour curvature's histogram, mean and standard deviation; bin size set to 10, contour segment length set to 10,20,30,40,50	48

output numbers of Unary-Net and Pairwise-Net are K and K^2, respectively, where K is class number.

E. Printed/handwritten Texts Separation

Different from MLP based potentials in text/non-text classification, we use CNN to formulate unary and pairwise potentials in printed/handwritten separation. The structures of our Unary-Net CNN and Pairwise-Net CNN are shown in Fig.3(b)(c). The input of our Unary-Net are normalized one channel gray level CC images whose sizes are 32×32 pixels, while the input of Pairwise-Net are normalized three channel CC image pairs whose third channel shows the relative position of the two CCs.

Before the separation of printed/handwritten texts, we detect those touching CCs which contain both printed pixels and handwritten pixels (Fig. 4) using the same CRF model. Then the touching CCs are split into smaller CCs and each small CC contains only one category of pixels presumably. We add those small CCs back into the text CCs set and use our CNN based CRF model to separate all the text CCs into two classes: printed and handwritten.

Figure 4: Printed-handwritten touching CCs.

F. Post Processing

There are numerous dot-like CCs which are so small that they cannot provide appropriate features and cannot be correctly classified by our MLP or CNN based CRF classifiers. We call them noise CCs. We separated those noise CCs from other CCs and don't let them participate in the previous text/non-text and printed/handwritten classification. After the other CCs are classified and used as reference CCs, a kNN classifier is used to classify those noise CCs. In our work, the number of neighbors take the value of 9. Some text/non-text and printed/handwritten texts classification results are shown in Fig. 5.

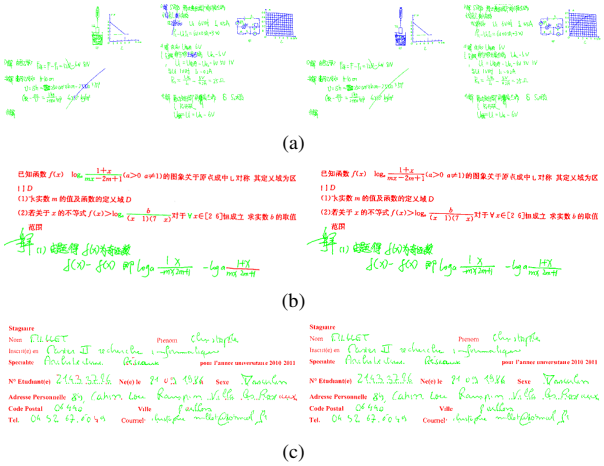


Figure 5: Separation results (left: Unary-Net only, right: CRF). (a) Text/non-text on TestPaper1.0; (b) Writing type on TestPaper1.0; (c) Writing type on Maurdor.

IV. Experiments

A. Dataset

We conduct our experiments on two datasets of complex documents. The first one is Testpaper1.0 dataset, which is collected and annotated by ourselves. The other is the public available Maurdor dataset [24], which is more heterogeneous and challenging. TestPaper1.0 dataset consists of 400 test paper documents written by different person in Chinese and English, out of which 300 documents are used as training set and the rest are used as test set.

The Maurdor dataset is multi-lingual (French, English, Arabic) with both handwritten and printed documents. This dataset is composed of 8129 pages with 176293 text zones, out of which 6129 pages with 130508 text zones are used as training set, 1000 pages with 23306 text zones are used as development set and 1000 pages with 22479 text zones are used as test set. A test paper document sample and a Maurdor sample can be found in Fig. 1.

B. Experiment Setting

1) MLP Structure: Unary-Net MLP: 4 layers, each layer has 146, 32, 16 and 3 (text/non-text) or 2 (printed/handwritten) nodes. Pairwise-Net MLP: 4 layers, each layer has 294, 32, 16 and 9 (text/non-text) or 4 (printed/handwritten) nodes.

The activation functions of hidden layer and output layer are sigmoid function and soft max function, respectively. All the parameters of MLPs are randomly initialized within $\pm$ 0.05 using uniform distribution.

2) CNN Structure: Structure of our Unary-Net CNN and Pairwise-Net CNN can be found in Fig. 4. We use the open source library *Caffe* for implementation.

3) CRF Structure: The unary potentials and pairwise potentials of CRF are formulated with Unary-Net and Pairwise-Net, respectively. Each CC in the document corresponds to a node in CRF, and each CC is connected with its all kNN. In our experiments, k=4. For the belief propagation inference of CRF, we adopt the open source library *OpenGM* for implementation.

Table II: CC-level text/non-text classification results on TestPaper1.0 dataset

Method	Text			Graphics			Table			GP
	P	R	F-m	P	R	F-m	p	R	F-m	
MLP	99.96	99.97	99.96	91.53	87.66	89.55	91.91	93.28	92.59	99.91
CRF_MLP	**99.96**	**99.99**	**99.98**	**98.59**	**90.91**	**94.59**	**96.15**	**93.28**	**94.70**	**99.95**
CNN	99.96	99.92	99.94	81.41	89.61	85.31	93.99	93.28	93.63	99.87
CRF_CNN	99.95	99.98	99.97	94.48	88.96	91.64	95.38	92.54	93.94	99.93

Table V: Region-level writing type separation results on Maurdor dataset

System	Printed			Handwritten			GP	SR
	P	R	F-m	P	R	F-m		
Maurdor2013-S2	92.43	95.61	93.99	83.07	73.33	77.90	90.55	6.56
Maurdor2013-S5	93.96	92.59	93.27	78.88	82.30	80.56	90.00	0.02
Maurdor2014-S2	94.93	96.23	95.57	88.10	84.46	86.24	93.30	0.15
Maurdor2014-S5	96.92	98.09	97.50	93.18	89.35	91.23	96.11	11.12
CRF_CNN_Vote*	98.18	97.24	97.71	91.84	94.52	93.16	96.57	0
CRF_CNN_Vote	98.18	97.26	97.72	91.89	94.51	93.18	96.58	0.02
CRF_CNN_Vote	98.18	97.35	97.76	92.13	95.50	93.30	96.65	0.15
CRF_CNN_Vote	98.61	98.89	98.75	**96.25**	95.35	95.80	98.07	6.56
CRF_CNN_Vote	**98.63**	**98.87**	**98.75**	96.22	**95.42**	**95.82**	**98.08**	11.12

4) Hardware Requirements: Our method is implemented by C++ and all the experiments are performed on a computer with an Intel Core i7-4790 CPU (3.60GHz) expect that our CNNs are trained on a GPU server with Titan GTX 980.

C. Experimental Results

1) Evaluation Metrics: We evaluate our experimental results with the metrics of precision (P), recall (R) and F-measure (F-m) of each class and the global precision (GP) of all classes at CC level (TestPaper1.0 and Maurdor) or region level (Maurdor). We train our models with training sets and report our experimental results on test sets.

2) Results on TestPaper1.0: Table II shows P, R, F-m and GP of each class for text/non-text classification. MLP based potentials beats CNN based potentials with nearly 3 percent gain in F-m of graphics and nearly 1 percent gain in F-m of table. The reason for this is that CNN takes images of normalized size as input, thus some details of original CCs which are extremely important for text/non-text classification such as scale information will be lost after the normalization procedure. What's more, graphics CCs and table CCs are much fewer (under 2k) than text CCs, which make it harder to train complex CNNs than simple MLPs. This makes MLP a better choice than CNN. It's worthy noting that all the CRF models (CRF_MLP and CRF_CNN) perform much better than their baselines (MLP and CNN), which confirms the importance of context information for classification tasks.

Table III: CC-level writing type separation results on TestPaper1.0 dataset

Method	Printed			Handwritten			GP
	P	R	F-m	P	R	F-m	
MLP	95.23	95.76	95.49	91.98	91.03	91.51	94.11
CRF_MLP	98.53	99.09	98.81	98.29	97.23	97.76	98.45
CNN	97.86	97.18	97.52	94.81	96.03	95.42	96.78
CRF_CNN	**99.70**	**99.27**	**99.49**	**98.64**	**99.45**	**99.05**	**99.33**

Table IV: CC-level writing type separation results on Maurdor dataset

Method	Printed			Handwritten			GP
	P	R	F-m	P	R	F-m	
CNN	98.45	98.32	98.38	87.83	88.66	88.24	97.15
CRF_CNN	**99.42**	**99.58**	**99.50**	**96.90**	**95.78**	**96.34**	**99.12**

Table III shows results for writing type separation. CNN potential based CRF model achieves the best result. Compared with artificial designed features based MLP, CNN has more strength in feature extraction especially in texture, margin and curvature extraction which are particularly suitable for printed/handwritten texts classification.

3) Results on Maurdor: We also do experiments of writing type separation with our CNN potential based CRF model on Maurdor dataset. The CC-level classification results can be found in table IV. Our CRF_CNN model achieves an impressive global precision of 99.12%, which is about 2 percent higher than the baseline CNN classifier.

To compare our model with existing methods, we also report region-level results in table V. Maurdor2013-S2 and Maurdor2013-S5 are two systems from Maurdor2013 competition [25], Maurdor2014-S2 and Maurdor2014-S5 are two systems from Maurdor2014 competition. The "SR" column stands for silence ratio which means the ratio of regions that are rejected to give result labels with corresponding systems. Our classification are conducted at CC level, so we need a post processing procedure to get each region's label. We do this by voting. In other words, we take the label with most CCs inside a region as that region's label. This leads to the region-level results of 96.57% GP (CRF_CNN_Vote*) in table V, which outperform the best existing results of 96.11% GP. But this comparison is somewhat unfair to our system since our silence ratio is zero. So we also list our results with same SR with existing systems. Table V shows that with the same SR (11.12%), our method (98.08%) outperform the best existing result (96.11%) by nearly 2 percent GP.

V. Conclusion

In this paper we propose a CRF based method to classify text/non-text and printed/handwritten texts in complex documents. Both our unary and pairwise potentials are formulated as neural networks for better exploiting spacial context information in documents. Interestingly we show that in some scenarios if carefully designed, artificial

features based MLP can outperform CNN which may lose some important information for classification. Our experiment results on TestPaper1.0 dataset and Maurdor dataset are impressive. Specifically, even use very shallow CNNs as potentials, our method achieves state-of-the-art results for writing type separation on the highly heterogeneous Maurdor dataset, which confirmed the superiority and effectiveness of the proposed method.

In the future, we will further investigate the region segmentation and classification, script and language identification, text line extraction and recognition, reading order identification tasks on complex documents.

ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation of China (NSFC) Grants 61411136002, 61573355, 61733007, 61773376 and 61721004.

REFERENCES

[1] A. Delaye and C.-L. Liu, "Contextual text/non-text stroke classification in online handwritten notes with conditional random fields," *Pattern Recognition*, vol. 47, no. 3, pp. 959–968, 2014.

[2] A. Saïdani, A. K. Echi, and A. Belaid, "Identification of machine-printed and handwritten words in arabic and latin scripts," in *Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013)*. IEEE, 2013, pp. 798–802.

[3] K. Jain, A. M. Namboodiri, and J. Subrahmonia, "Structure in on-line documents," in *Proceedings of the 6th International Conference on Document Analysis and Recognition (ICDAR 2001)*. IEEE, 2001, pp. 844–848.

[4] E. Indermühle, M. Liwicki, and H. Bunke, "Iamondo-database: an online handwritten document database with non-uniform contents," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems (DAS 2010)*. ACM, 2010, pp. 97–104.

[5] X.-D. Zhou and C.-L. Liu, "Text/non-text ink stroke classification in japanese handwriting based on markov random fields," in *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 1. IEEE, 2007, pp. 377–381.

[6] J.-Y. Ye, Y.-M. Zhang, and C.-L. Liu, "Joint training of conditional random fields and neural networks for stroke classification in online handwritten documents," in press.

[7] E. Indermühle, V. Frinken, and H. Bunke, "Mode detection in online handwritten documents using blstm neural networks," in *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition (ICFHR 2012)*. IEEE, 2012, pp. 302–307.

[8] T. Van Phan and M. Nakagawa, "Combination of global and local contexts for text/non-text classification in heterogeneous online handwritten documents," *Pattern Recognition*, vol. 51, pp. 112–124, 2016.

[9] V. Vidya, T. Indhu, and V. Bhadran, "Classification of handwritten document image into text and non-text regions," in *Proceedings of the 4th International Conference on Signal and Image Processing 2012 (ICSIP 2012)*. Springer, 2013, pp. 103–112.

[10] S. Ahmed, M. Liwicki, and A. Dengel, "Extraction of text touching graphics using surf," in *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS 2012)*. IEEE, 2012, pp. 349–353.

[11] Y.-F. Pan, X. Hou, and C.-L. Liu, "Text localization in natural scene images based on conditional random field," in *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR 2009)*. IEEE, 2009, pp. 6–10.

[12] S. Kanoun, I. Moalla, A. Ennaji, and A. M. Alimi, "Script identification for arabic and latin printed and handwritten documents," in *Proceedings of the 4th IAPR International Workshop on Document Analysis Systems (DAS 2000)*, 2000, pp. 159–165.

[13] S. Haboubi, S. S. Maddouri, and H. Amiri, "Discrimination between arabic and latin from bilingual documents," in *Proceedings of the International Conference on Communications, Computing and Control Applications (CCCA 2011)*. IEEE, 2011, pp. 1–6.

[14] X. Peng, S. Setlur, V. Govindaraju, and R. Sitaram, "Handwritten text separation from annotated machine printed documents using markov random fields," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no. 1, pp. 1–16, 2013.

[15] M. Seuret, M. Liwicki, and R. Ingold, "Pixel level handwritten and printed content discrimination in scanned documents," in *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition (ICFHR 2014)*. IEEE, 2014, pp. 423–428.

[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 (ICCV 2015)*, 2015, pp. 1529–1537.

[17] G. Lin, C. Shen, A. van den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2016, pp. 3194–3203.

[18] M. Valizadeh, N. Armanfard, M. Komeili, and E. Kabir, "A novel hybrid algorithm for binarization of badly illuminated document images," in *Computer Conference, 2009. CSICC 2009. 14th International CSI*. IEEE, 2009, pp. 121–126.

[19] B. Su, S. Lu, and C. L. Tan, "Robust document image binarization technique for degraded document images," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1408–1417, 2013.

[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] 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.

[22] 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 (ICCV 2003)*, vol. 2, 2003, pp. 900–906.

[23] 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 (ICCV 2003)*. IEEE, 2003, pp. 1150–1157.

[24] S. Brunessaux, P. Giroux, B. Grilheres, M. Manta, M. Bodin, K. Choukri, O. Galibert, and J. Kahn, "The maurdor project: Improving automatic processing of digital documents," in *Iapr International Workshop on Document Analysis Systems*, 2014, pp. 349–354.

[25] I. Oparin, J. Kahn, and O. Galibert, "First maurdor 2013 evaluation campaign in scanned document image processing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 5090–5094.