D-StaR: A Generic Method for Stamp Segmentation from Document Images

Junaid Younas*†, Muhammad Zeshan Afzal*, Muhammad Imran Malik‡, Faisal Shafait‡, Paul Lukowicz*†, Sheraz Ahmed†
*Kaiserslautern University of Technology, Germany
†German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
Email: firstname.lastname@dfki.de
‡National University of Sciences and Technology (NUST), Islamabad H-12, Pakistan
Email: firstname.lastname@seecs.edu.pk

***Abstract*—B This paper presents a novel approach, named D-StaR, for stamp segmentation from scanned document images. The presented approach is generic (applicable to stamps of any color, shape, size, and orientation) and based on deep learning. In particular, it uses Fully Convolutional networks for semantic analysis of documents to extract stamps. The presented approach is evaluated on a publicly available stamp dataset. Evaluation results show that the presented approach outperforms the state-of-the-art methods for stamp segmentation and achieves pixel based precision and recall of 87% and 84%, respectively. Deeper analysis of the evaluation reveals that the presented approach can segment both overlapping and non-overlapping stamps, which was always a problem for existing systems in the literature.**

I. INTRODUCTION

Stamps are considered as mark of authenticity and originality of documents. Stamps are used to mark the documents with information relating to creation, distribution, and storage. Thousands of documents (including legal, financial, security documents, bank receipts, checks, and utility bills) are received and sent by large organizations on daily basis. These documents have single or multiple stamps on them. Every stamp on a document highlights the significance and purpose of the document. Stamps largely vary in shape, size, and color from organization to organization as well as with in departments of an organization. In general, stamps appear in textual, graphical, regular (official), and irregular (fun purposes) shapes, as shown in Figure 1.

Stamp segmentation is an important part of automated classification and verification of documents. Stamps, however, may overlap text, logos, and/or other information present in documents. Furthermore, orientation of stamps varies from document to document and different scanning environments add their overhead as well. Hence, proper and correct stamps segmentation from scanned documents is a challenging problem.

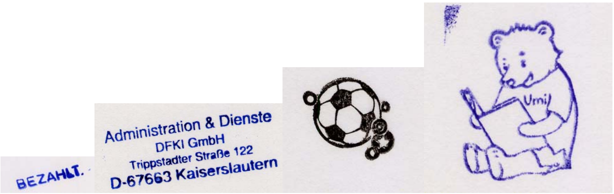


Figure 1: Textual, graphical, official and fun purpose stamps

In past, various approaches have been presented for stamps detection [1], [2]. Most of these approaches used different sets of stamp features including color [1], [3], shape [4], and local keypoint descriptors [2], [5], [6] to separate stamps from logos, text, and other information present in scanned document images. These approaches may serve for specific organizations which use predefined set of stamps based on color, shape, or features. However, none of these approaches can be applied to develop a generic system for stamp detection and segmentation applicable to a vast majority of stamps having different colors, shapes and textures, and especially for overlapping stamp segmentation.

Deep learning has been successfully applied for object classification and detection [7]–[9]. While deep learning has seen success with a breakthrough paper by Krizhevsky et. al. [8], the history of successful deep learning methods for handwriting recognition [10] is rather old. For pixel level image labeling, deep learning methods have been applied for binarization and layout analysis [11]–[13]. However, Fully Convolutional Networks (FCNs) remain unexplored in this context. A closely related work [14] uses FCNs for detecting text in natural scenes. Nevertheless, FCNs have never been explored earlier for information extraction from document images.

In this paper we present a generic approach, based on FCNs, to segment stamps from scanned document images. The presented approach, named D-StaR, can detect unseen stamps of any shape, color, size, and orientation. Moreover, D-StaR is also capable of detecting overlapping stamps. This is the first method to use deep learning for stamp detection.

DOI 10.1109/ICDAR.2017.49

We used a Fully Convolutional Network to segment stamp masks from scanned document images. Contour refinement is applied to the predicted masks for pixel based evaluation and reforming the original stamps. The proposed method is evaluated on a publicly available stamp detection and verification dataset [1] where it yields pixel-based precision and recall of 87% and 84%, respectively.

The rest of the paper is structured as follows: section II summarizes the previous work done for stamps detection and verification. Section III elaborates the D-StaR approach for stamp segmentation and detection. Section IV presents the evaluation methodology and finally Section V concludes the paper and provides a brief outlook of our future work planned in the direction of stamp segmentation.

II. Related Work

Different methods are proposed to detect the stamps from document images in the past. Most of these methods used heuristics based approaches to detect and classify stamps. Some approaches used color-based features to segment stamps, while other approaches used geometric features with keypoint descriptors to extract the stamps from scanned document images.

One of the earliest methods to detect seal imprints and signatures from checks of Japanese banks was proposed in Ueda et al. [3], assuming signatures, seal imprints and background to be different from each other. It uses the color information (RGB 3D) to detect stamps and signatures from images. The proposed technique fails when any of three clusters is not monochromatic.

An automatic segmentation and verification system is presented by Micenkova et al. [1] to detect and verify the stamps from scanned document images. This approach is based on color segmentation of documents in $\mathrm{Y}C_bC_r$ color space. Candidate solutions are extracted using XY-cut algorithm [15]. Candidate solutions are further processed using geometrical and color-based features to extract the stamp region. This approach doesn't address the black stamps detection. It also fails to detect the stamp when stamp background matches with the color of the stamp. The extended model with stamp verification is presented in Micenkova et al. [16] .

Ahmed et al. [2] presented a part-based feature extraction method. It uses two-step approach to classify stamps from non-stamp regions using geometrical features. First, it computes the key-points and then descriptors from these key-points. Their presented approach outperforms other methods in the detection of black stamps, while the results reported for colored images in [2] are on lower side.

An outliers-based approach has been presented to detect the stamps and logos from scanned document images by Dey et al. [6]. In this approach it is assumed that the documents only consist of text, stamps and logos. Considering stamps and logos as outliers, it divides the document into foreground and background using PCA and color information, treating both of them as separate images. These images are then individually processed further for pixel level evaluation.

Another shape specific segmentation approach is presented in Forczmaski et al. [5]. This approach uses color space transformation to look for potential color stamps, followed by different object detection algorithms to compute the shape descriptors. Isolated regions are extracted from scanned documents, which are classified using computed shape descriptors. This approach potentially addresses the detection of well-defined shapes (official stamps) regardless of stamp color.

Recently, a shape specific stamp segmenting approach using examplar features is proposed by Bhalgat et al. [4]. This approach uses unsupervised learning methods to extract the dictionary items for stamp shapes. Feature vectors are extracted by using single Convolution layer with 4×4 quadrant maxpooling. Dictionary ranking item scheme is used for recognition of stamps. This approach produces excellent results for only oval shaped stamps.

Note that Dey et al. [6] also presented an outliers based approach to detect stamps and logos from scanned document images. It is highly fragile and prone to error approach as it is explicitly based on a large set of experimentally computed parameters are generated from the whole dataset. These computed parameters are then applied to the very same dataset for evaluation purposes. Hence, comparison doesn't stand valid with this approach.

The discussed methods use heuristic approaches and have their constraints in particular scenarios. Deep learning approach hasn't been use to detect the stamps from the images in any case till time up to the best of author's knowledge.

III. D-StaR: The Presented System

This section provides details on the presented approach (D-StaR) for stamp segmentation in document images. Figure 2 shows architecture of D-StaR, which uses a Fully Convolutional Neural Network to generate semantic segmentation of input images. The generated segments are pixel level maps of stamp location in the original scanned document images. The FCN's generated stamp maps are then post processed using connected component analysis to detect the exact stamp pixels from input image. Usually, deep learning based approaches require a lot of training data. However, the publicly available "Stamp Detection and Verification" dataset we used contains only 400 scanned document images. Therefore, to resolve this problem we used the concept of domain adaptation and transfer learning to train our fully convolutional network.

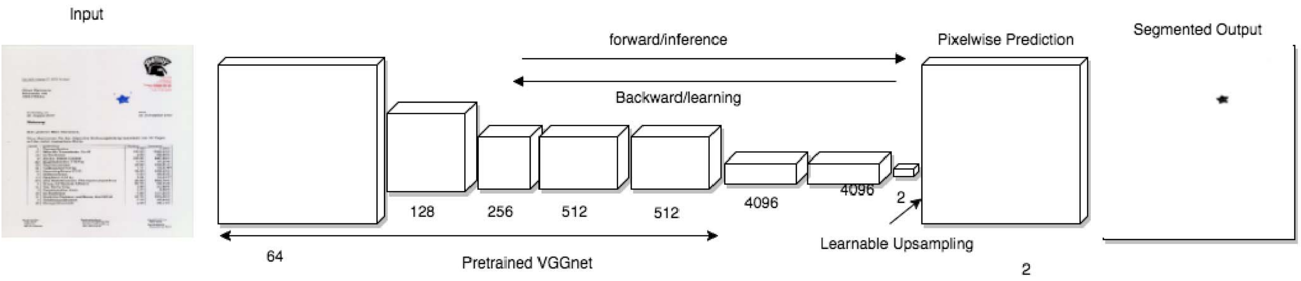


Figure 2: D-StaR architecture with input image and pixel-level segmentation result of FCN

A. *Domain Adaptation and Transfer Learning*

In this paper, we adapted the domain of general-purpose object detection and segmentation from natural scene images, to segmentation of stamps in document images. Both of these domains are totally different. Furthermore, to compensate for the problem of non-availability of large dataset, we used the concept of transfer learning. Transfer learning is defined as transfer of knowledge from a learned task to a new task [17]. In Convolution Neural Networks (CNNs), transfer learning refers to use of learned features from a pretrained network for a new task. Using pretrained network is particularly useful when we have very little data available to train a new network. In D-StaR, we used a pretrained VGG-Net16 [9] for transfer learning. The VGG-Net16 was trained on PASCAL-VOC-2011 dataset [18]. VGG-Net16 was preferred as backbone because it produced better segmentation results despite inference time on higher side [4].

B. *VGG-Net*

VGG-Net is runner up for ImageNet Large Scale Visual Recognition Challenge 2014 (ILSVRC). It's an advance CNN architecture, which takes the fixed size input of 224×224 RGB images. The filter size used in VGG-Net convolutional layers is 3×3, the smallest possible receptive field size to capture the features. The convolution stride is fixed to 1 pixel. After convolution, spatial pooling is carried out by 5 maxpooling layers with a 2×2 pixel window, with stride of 2. Input images are processed through stack of these convolution layers followed by three Fully Connected (FC) layers. For detailed information we refer our readers to [9].

To adapt the pretrained VGG-Net to the problem of stamp segmentation and to improve the performance of FCN, we removed the FC layers and used the output of 5th maxpool layer for fully convolution processing. Figure 2 provides an overview of the architecture of pretrained VGG-Net used in D-StaR. By removing FC layers from VGG-Net, we are now able to process input images of arbitrary size.

C. *Fully Convolutional Networks*

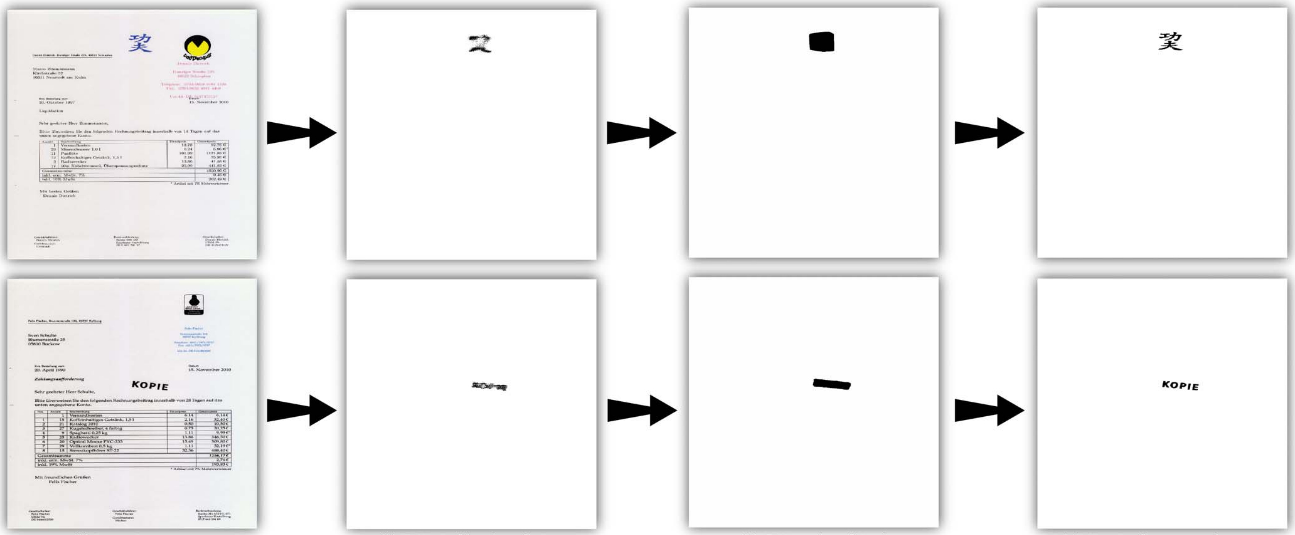
Fully Convolutional Networks are able to process arbitrary sized images due to removal of fully connected layers in the network at the end requiring fixed size input. FCNs are mainly used for semantic segmentation of images in which pixel level output is generated by combining context from higher layers and information from lower layers while retaining spatial information. Deconvolution layers are used for decoding the embeddings generated by the encoder. No learning is needed for deconvolution layers as these are initialized as bilinear up-sampling layers. FCNs are given priority over conventional CNNs because of following advantages [19]:

- Highly computational efficient networks. It takes about 100 milliseconds to process 1000×1000 image in training phase.
- FCNs generate pixel level masks for every corresponding class because of their ability to perform segmentation at higher levels.
- FCN can be build on any state-of-the-art CNN e.g Alexent, ResNet, ImageNet, Google-Inception models, rendering vast scope of adaptability. Using pretrained networks significantly boost and fine-tune the performance of FCNs, helping in very fast convergence even on very small datasets.

As FCN processing is on pixel-level, it needs per pixel annotations for training. Our focus in the scanned document images is stamp regions. So, we used the annotations containing only the pixel level stamp masks, resulting into binary classification task for FCN.

We used convolution layers from pretrained VGG-Net. Three fully convolution layers on top of VGG-Net were added to perform FCN functionality. The kernel size for the first, second and third FC layers are 7×7, 1×1, and 1×1 respectively. The output of size 512 generated by fifth maxpool layer of VGG-Net served as input of first fully convolution layer. The output of first fully convolution layer of size 4096 is input of second fully convolution layer, which generates the output of same size as of input. The output of second fully convolution layer is then processed in last fully convolution layer for pixel level prediction of every class.

As we used the FCN8-s, it means 3 levels up scaling or deconvolution are done to produce per pixel classification at stride 8. FCN8-s uses predictions from maxpooling layer $5, 4$ and 3 respectively, of stride at 8 to generate the pixel level predictions as shown in figure 2.



(b). Input Image (b). FCN pixel level prediction (c). FCN predicted Mask (d). Binarized segmented output

Figure 3: D-StaR overview, (a) Two different input images one containing graphical and colored stamp, while the other input image with black and textual stamp on it, (b) shows the results for both scanned document images for pixel level predictions, (c) shows the processed predicted masks from FCN, and (d) shows the detected and segmented binary stamps.

We tried images of different sizes as input of FCN for optimal performance, because number of scanned document images are very less in reference to deep learning. We used RGB images of size 1000×1000 as input to our FCN. FCN was trained with pixel level binary masks, marking stamps only as point of interest. Training was done for 10 epochs. We used the batch size of 2 in training phase. To fine tune the network parameters and optimize the performance, learning rate of .0001 was used.

FCN generates the pixel level predictions for stamp regions. Figure 3 shows the complete work-flow of the D-StaR with intermediate results on each step. The pixel level predictions are post processed for connected component analysis to generate FCN predicted masks. These predicted masks are used to extract stamp pixels from input image. The stamp-segmented images are then converted to binary images for evaluation purposes.

IV. Evaluation

A. Dataset

For evaluation of D-StaR, we used a publicly available stamp detection and verification dataset[1] [4]. This dataset contains 400 document images scanned at 200, 300, and 600 dpi resolution. The scanned document images contain printed text, stamps (textual and non-textual), logos, and signatures. The dataset contains stamps of varying sizes, shapes, and colors. 341 (out of 400) scanned document images contain single or multiple stamps and remaining 59 images have no stamps. Out of 341 scanned document images, 80 contain black stamps, and remaining 241 contain colored stamps. In 55 scanned document images stamps are overlapped with text, logos, and/ or signatures. For every scanned image, there are two ground truth images available; one containing the pixel level information and the other containing the bounding box information for each stamp. So, this dataset can be used for both region classification and pixel level evaluation.

B. Evaluation Protocol

For Evaluation of D-StaR, the dataset has been split into train and test set with different configurations. We used document images scanned at 200 dpi resolution. The train sets containing 90% of scanned documents and remaining 10% are in the test sets. As we evaluate D-StaR for three different categories of scanned document images, we customized the test sets with only colored stamps, black stamps, and overlapping stamps, respectively. Furthermore, we evaluated our presented approach with system generated random test set and with class balanced (equally distributed samples from every category) for overall performance evaluation.

We evaluated the D-StaR for pixel level detection of stamps and therefore the most relevant evaluation terminologies are Precision and Recall [20]. Precision is the intuitive ability of a classifier to distinguish a negative sample from positive one. It is computed as:

$$Precision = \frac{tp}{tp + fp} \tag{1}$$

Recall is the ability of a classifier to classify all the positive samples. It is computed as:

$$Recall = \frac{tp}{tp + fn} \tag{2}$$

[1] The dataset is available at: vhttp://madm.dfki.de/downloads-ds-staver

In equations 1 & 2, tp, fp, and fn denote the true positives, false positives, and false negatives, respectively. True positives refer to the predicted number of pixels actually belonging to stamps. False positives refer to the number of pixels that are predicted as stamps but they don't actually belong to stamps. False negatives specify the number of stamp pixels the system fails to predict.

C. Results and Discussion

We present a detailed comparison of our presented approach, considering different aspects (segmentation of colored, monochrome, and overlapping stamps) with the state-of-the-art approaches. Our presented approach is independent of shape, size, color, and orientation of stamps with regard to text, or logos present in scanned document images. Results used for the comparisons are computed when only the mentioned stamp category was present in the test set, with the rest of scanned document images in the training set.

Table I: Performance Evaluation of D-StaR on overlapping stamps

Approach	Precision(%)	Recall(%)
D-StaR	74	77
Micenkova et al. [1]	68	69
Ahmed et al. [2]	Not Reported	

Segmentation of overlapping information is considered as a very difficult task in information segmentation and classification [21] [22]. Table I shows the results when D-StaR is tested on overlapping stamps. These stamps overlap with the background text and/or logos at different positions. D-StaR outperforms the state-of-the-art in segmenting overlapping stamps by a large margin. It correctly segments the overlapping stamps with pixel level precision of 74% and recall of 77%. Figure 4 shows some overlapped stamp and their binary segmented results by D-StaR. Micenkova et al. [1] reported the pixel level recall and precision of 69% and 68%, respectively for overlapping stamps. Ahmed et al. [2] didn't present results for overlapping stamps in the dataset, but mentioned their approach fails to detect the severely overlapping stamps.

Furthermore, we also evaluate D-StaR from another dimension i.e., colored stamps detection and segmentation. Table II provides results of colored stamp detection. For colored stamps, the presented approach reports the pixel level precision and recall of 92.7% and 84.3%, respectively. Micenkova et al. [1] reported pixel level precision and recall of 82.7% and 82.8%, respectively. Their approach fails to detect the stamps when its color matches with background, whereas D-StaR can be used successfully in such scenarios as well. Ahmed et al [2] approach is although independent of color and shape of stamps, as it uses part-based key points and feature descriptors for stamp detection, logos are also misclassified as stamps [2]. Therefore, their results are on lower side with pixel level precision and recall of 62% and 57%, respectively.

Table IV: Overall performance of D-StaR on randomly generated test set

Approach	Precision	Recall
D-StaR	87	84
Micenkova et al. [1]	Not Reported	
Ahmed et al. [2]	Not Reported	

Table II: D-StaR in comparison with the state-of-the-art approaches for colored stamps

Approach	Precision(%)	Recall(%)
D-StaR	92.7	84.3
Ahmed et al. [2]	62	57
Micenkova et al. [1]	82.7	82.8

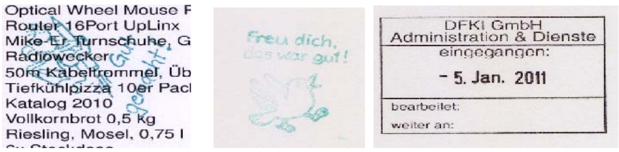


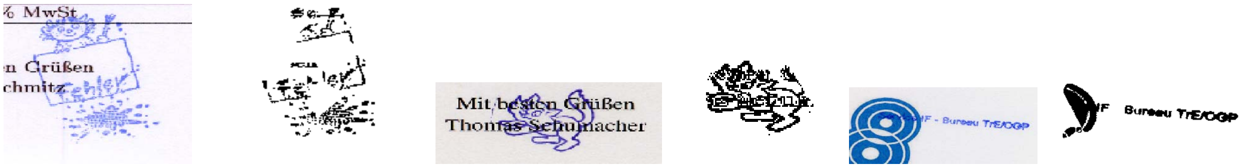
Figure 5: Stamps D-StaR failed to segment out.

Table III: D-StaR's Evaluation for black stamps w.r.t the state-of-the-art approaches

Approach	Precision(%)	Recall(%)
D-StaR	93.75	50.2
Ahmed et al.	83	73
Micenkova et al.	Not Applicable	

Table III elaborates the precision and recall comparison of D-StaR with the existing state-of-the-art approaches for black stamps in scanned document images. Micenkova et al. [1] approach assumes the stamps as colored objects only by processing the stamps document images for YC_bC_r color clusters. These color clusters are used for segmentation and detection of stamps. When it comes to black stamps, this approach does not stand valid (applicable). Ahmed et al. [2] approach reports the pixel level precision and recall of 83% and 73%, respectively in comparison to the D-StaR's 93.75% and 50.2%, for black stamps.

Table IV reports the overall stamp segmentation results of D-StaR. The state-of-the-art approaches do not report their overall results. D-StaR, however, achieve pixel level precision and recall of 87% and 84%, respectively. Note that the test and training set division for evaluations (presented in Table IV) have been made the system randomly and autonomously without any human intervention.



(a) Cropped input (b) Segmented binary output (c) Cropped input (d) Segmented binary output (e) cropped input (f) Segmented binary output

Figure 4: Overlapping stamps detected by D-StaR successfully (Overlapping images with predicted binary outputs)

V. Conclusions

This paper presents a novel and generic approach (D-StaR) to detect stamps using deep learning for the very first time. D-StaR is capable of detecting unseen stamps of any shape, size, and color. The major advantage D-StaR owes over previously presented approaches is that it can detect and segment the overlapping stamps from other information in scanned document images. It also outperforms the state-of-the-art approaches by successfully differentiating between stamps and logos, despite of huge similarity between the two. We, however, note that deep neural network's performance mainly depends on learning the spatial features of the data used for training. As the text on every image is in tabular form and we have lower number of tabular stamps in our dataset, D-StaR faced difficulty in classifying background tabular text and tabular stamps efficiently.

In the future, we plan to produce multi-class labels for segmenting logos, text, and stamps separately using FCN to improve the overall performance since the scanned document images also contain text and logo information. The extracted information can be further processed to classify documents into different classes depending on use-case.

VI. Acknowledgment

Authors would like to thank Shoaib Ahmed Siddiqui for extended discussions and support. This work is partially funded by Higher Education Commission (HEC), Pakistan.

References

[1] B. Micenkov and J. van Beusekom, "Stamp detection in color document images," in *ICDAR*, 2011, pp. 1125–1129.

[2] S. Ahmed, F. Shafait, M. Liwicki, and A. Dengel, "A generic method for stamp segmentation using part-based features," in *ICDAR*, 2013, pp. 708–712.

[3] K. Ueda, "Extraction of signature and seal imprint from bankchecks by using color information," in *ICDAR*, vol. 2, 1995, pp. 665–668.

[4] Y. Bhalgat, M. Kulkarni, S. Karande, and S. Lodha, "Stamp processing with examplar features," *arXiv preprint arXiv:1609.05001*, 2016.

[5] P. Forczmański and A. Markiewicz, "Low-level image features for stamps detection and classification," in *CORES 2013*, 2013, pp. 383–392.

[6] S. Dey, J. Mukherjee, and S. Sural, "Stamp and logo detection from document images by finding outliers," in *NCVPRIPG*, 2015, pp. 1–4.

[7] G. Zhu and D. Doermann, "Automatic document logo detection," in *ICDAR*, 2007, pp. 864–868.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition." *PAMI*, pp. 855–68, 2009.

[11] J. Pastor-Pellicer, M. Z. Afzal, M. Liwicki, and M. J. Castro-Bleda, "Complete system for text line extraction using convolutional neural networks and watershed transform," in *DAS*, 2016, pp. 30–35.

[12] M. Seuret, M. Alberti, R. Ingold, and M. Liwicki, "Pca-initialized deep neural networks applied to document image analysis," *arXiv preprint arXiv:1702.00177*, 2017.

[13] K. Chen, C.-L. Liu, M. Seuret, M. Liwicki, J. Hennebert, and R. Ingold, "Page segmentation for historical document images based on superpixel classification with unsupervised feature learning," in *DAS*, 2016, pp. 299–304.

[14] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, and X. Bai, "Multi-oriented text detection with fully convolutional networks," in *ICVPR*, 2016, pp. 4159–4167.

[15] M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan, "Syntactic segmentation and labeling of digitized pages from technical journals," *PAMI*, vol. 15, no. 7, pp. 737–747, 1993.

[16] B. Micenková, J. van Beusekom, and F. Shafait, "Stamp verification for automated document authentication," in *Computational Forensics*, 2015, pp. 117–129.

[17] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[18] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *IJCV*, vol. 111, no. 1, pp. 98–136, 2015.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *ICVPR*, 2015, pp. 3431–3440.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, pp. 2825–2830, 2011.

[21] M. I. Malik, S. Ahmed, F. Shafait, A. S. Mian, C. Nansen, A. Dengel, and M. Liwicki, "Hyper-spectral analysis for automatic signature extraction," in *BCIGS*, 2015.

[22] S. Ahmed, M. I. Malik, M. Liwicki, and A. Dengel, "Signature segmentation from document images," in *ICFHR*, 2012, pp. 425–429.