ON NOISE REDUCTION FOR HANDWRITTEN WRITER IDENTIFICATION

Karl Ni, Patrick Callier, Bradley Hatch, Jonathan Mastarone, James Cline

In-Q-Tel, Lab41, Menlo Park, USA

ABSTRACT

Academic work in identifying writers of handwritten documents has previously focused on clean benchmark datasets: plain white documents with uniform writing instruments. Solutions on this type of data have achieved hit-in-top-10 accuracy rates reaching upwards of 98%. Unfortunately, transferring competitive techniques to handwritten documents with noise is nontrivial, where performance drops by two-thirds. Noise in the context of handwritten documents can manifest itself in many ways, from irrelevant structured additions, e.g., graph paper, to unstructured partial occlusion, e.g. coffee stains and stamps. Additional issues that confound algorithmic writer identification solutions include the use of different writing implement, age, and writing state of mind. The proposed work explores training denoising neural networks to aid in identifying authors of handwritten documents. Our algorithms are trained on existing clean datasets artificially augmented with noise, and we evaluate them on a commissioned dataset, which features a diverse but balanced set of writers, writing implements, and writing substrates (incorporating various types of noise). Using the proposed denoising algorithm, we exceed the state of the art in writer identification of noisy handwritten documents by a significant margin.

Index Terms**—** Deep learning, Document Analysis, Biometrics, Offline Handwriting, Denoising, Image Enhancement

1. INTRODUCTION

The area of document forensics has special applications in law enforcement, banking, and historical document analysis. Unfortunately, even for a medium as pervasive as the literal written word, there are few forensic subject matter experts in writer identification in handwritten documents, and manually doing so documents is laborious and often unfeasible. Automating the process with recent advances in machine learning would reduce workloads and enable the analysis of documents at scale.

Such automation has been in demand in both commercial and academic spaces leading to some innovation in the field. In 2015, Sciometrics/Gannon released its product FLASH ID [1], now in use by several government agencies (we compare against FLASH ID's performance below). Academic work in writer identification in handwritten documents is dense as well, and includes both hand-engineered features [2, 3] as well as deep networks [4, 5, 6]. Such work has large representation in special workshops at conferences like ICDAR 2011,'13, and '15 [7, 8, 9], though in all reported cases, benchmarking and evaluation datasets are noiseless, uniform, and well-culled.

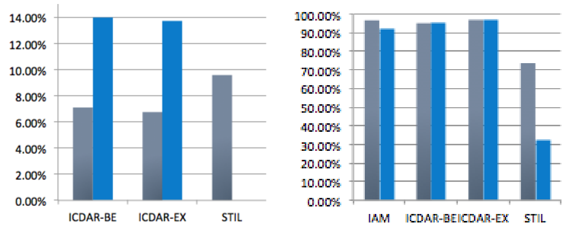
By evaluating on the STIL Handwriting Dataset [10] (STIL-HD), this work concentrates on both structured and unstructured noise in the form of background graph paper, lines, stamps, coffee stains, in addition to traditional noise and artifacts produced by image capture. In addition, the writing implement is also a controlled variable in the STIL-HD, where our experiments consider a controlled distribution of colored pens, pencils, and markers. All forms of variability are specifically designed to simulate current challenges in handwriting that subject matter expects may face. As an illustration of how important it is to consider such factors, consider that in the noiseless and uniform environment of the ICDAR 2013 evaluation set, state of the art algorithms achieve upwards of 96.5% accuracy in some cases, while the same algorithms perform up to 60% worse when applied to noisy data (i.e., the STIL-HD).

For this reason, we propose a denoising binary CNN, while comparing against traditional methods. In doing so, we will also assess the generalizability and extensibility of the algorithms for future use, where types of noise are less predictable. Our work is meant to couple with any feature extraction process, and possibly for other applications. We present results show the potential of denoising networks over more traditional signal processing efforts in writer ID on a dataset with limited documents per author.

2. BACKGROUND

Identifying writers in handwritten documents has traditionally [2, 3, 4, 11, 5] involve some combination of handwriting isolation, segmentation, feature extraction, feature aggregation, and then classification. In contrast to conventional computer vision datasets [12, 13], where examples per category number in the thousands, the number of documents per author is typically less than eight. Making matters worse, for a single writer, it would be prohibitively expensive to collect more data with conventional crowdsourcing methods. This has led research to focus heavily on feature extraction rather than classification methodology. Finally, fidelity in representing handwriting requires high-resolution and dots per inch imagery, which makes fitting entire documents into GPU memory over large batch sizes a considerable engineering effort.

While the deep learning literature revival arose from handwriting, its computer vision applications require an underlying assumption that the data is sufficiently "big". Based on direct application of both manual [2] and deep learning features [4], CNN feature extraction is effective without noise, but suffers drastically with its introduction, even with data augmentation [4] as seen hard and soft top k evaluation results [1] in Fig. 1. More innovative architectures may evolve to address the data issues for an individual writer necessary for deep learning, but development and application of denoising algorithm is clearly beneficial in either approach..



(a) Hard score, top 3

(b) Soft score, top 10

Fig. 1: Comparing convolutional neural network features (in blue) [4] to manual features [14] (in gray), using overall soft and hard criteria for clean (ICDAR-BE and ICDAR-EX) and noisy (AMA) datasets.

Specific to denoising handwritten documents, as recently as a few months ago, Kaggle completed its series on denoising dirty documents [15], with 161 teams participating; the dataset is hosted at the UCI repository [16]. Another front in which the field has made progress is via the problem of historical documents [17]. In most cases, the focus has been on OCR, but additive unstructured noise (coffee stains, rips, holes, folds, illumination errors) encountered in historical document analysis is typical of challenges that forensic document analysts also face.

Further relevant research [18, 19, 20] on structured noise (lined and graph paper) is centered around procured datasets organized by Kumar, Doermann, and Abd-Almageed [21]. Such heuristics-based methods tend to rectify a single problem rather than a broad set of them, addressing, for example, horizontal rule lines only, rather than horizontal rules alongside graph paper rules, tears, coffee stains, and so on. Limitation stem from the use of underpowered linear algorithms.

To address a wider variety of noise types and foreground the impact of writing implement choice, Applied Media Analytics collected the noisy STIL Handwriting Dataset that we test on in this work. The denoising algorithm that we propose is based on stacked convolutional autoencoders [22] and is most similar to many networks used for deconvolution [23, 24, 25], which has shown promise in applications like image deblurring [25] and segmentation [26].

3. APPROACH

Our exploration of the writer identification included a large variety of feature extraction methods, divided broadly between manual and deep learning-based techniques, both of which we assessed with and without denoising. Results in Sec. 4 show only most performant combinations of algorithms.

Traditional methods in dealing with structured and unstructured noise typically call for some type of data augmentation [27], which has worked well for CNNs. However, given that Jain's [2] work outperforms Fiel's [4] on noisy data by a large margin (Fig.1), there needs to be an equivalent mechanism for manual features in order to provide a usable solution.

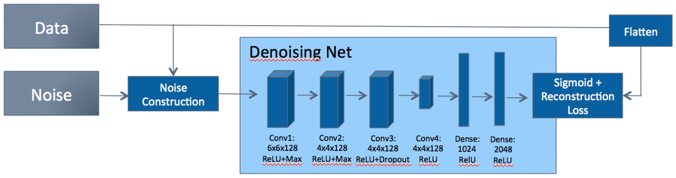


Fig. 2: Binary Denoising Neural Networks

While our initial cues were taken from fully convolutional networks without dense layers like [25, 26], there was a quick realization that while smaller in parameter space, training times for fully convolutional networks have the potential to be 3–4× slower using our CUDA kernels and Maxwell class GPUs. Given this, we abandoned deconvolutional layers in favor of fully-connected layers followed by a reshaping operation, which optimized in an acceptable timeframe in light of our time constraints. Secondly, because the application space differs from those of [25, 26], the proposed denoising methodology necessitates different cost functions and strategies.

Our network is an encoding/decoding network, but it a slight abuse of the terminology in that the "decoding" portion of our network does not reflect the "encoding" portion. Secondly, we differ from cited literature because we use bi-

[1] For a particular query document, a given distance/similarity metric is used to rank all candidate documents from the corpus, listed from most similar to least similar. A correct hit for the soft top-k is defined as *at least one* document image from the same writer as the query being included in the top k. A correct hit for the hard top-k criterion is defined as the case when *all* of the k documents retrieved are correct. Over an evaluation set, the overall hard or soft top-k metric is the hit percentage for each query document in the set. An additional metric used in STIL-HD for evaluation is the top-1 metric, which is a hit when the top-ranked author is correct.

nary rather than continuous targets. We found that doing so will raise the signal to noise ratio and decrease convergence time. Fig. 2 depicts the pipeline, which follows the following steps: (1) Binarize image patch (threshold at 0.45 of document maximum) to get x_i, (2) Add noise n_i to image patch x_i with function $g(x_i, n_i)$ to yield $\hat{x}_i$ (3) Train denoising network $f(\cdot)$ with clean binary image target x_i

In image processing, evaluation of the relative difference between "clean" and "noisy" is typically conducted with mean squared error (MSE) or peak signal to noise ratio, and several works [25, 28] use normalized MSE as the initial cost function. We found that binary target labels obtained by low thresholds of the original image are effective, and so the proposed approach, instead, defines the cost between predicted output and target using a scaled sigmoid function without normalization. The neural network $f(\cdot)$ is the light-blue highlighted architecture in Fig. 2, and we define the last layer output as a function of $h^{\ell-1}$, the state of the previous layer:

$$f(\hat{x}) = \sigma(\alpha(W_\ell * h^{\ell-1} + b_\ell)) \qquad (1)$$

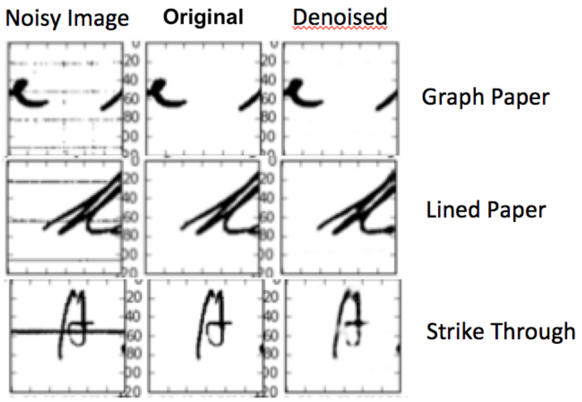
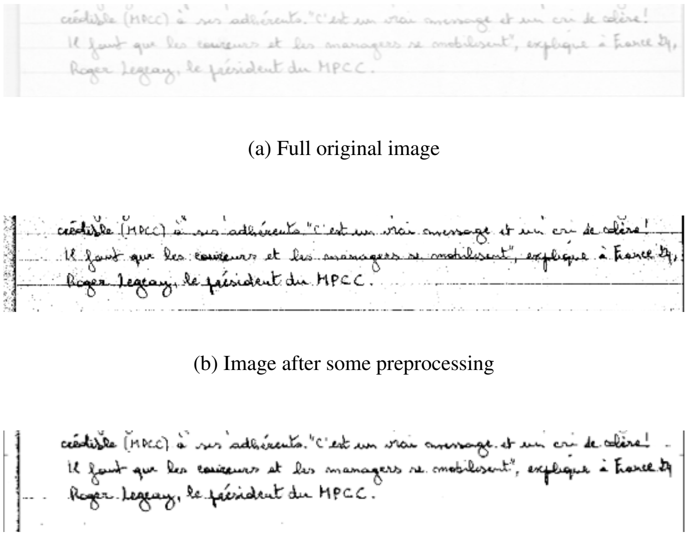


Fig. 3: Example denoised image patches with 56×56

Initially, we found that this approach was effective in removing noise in patches where there are large amounts of foreground content. For example, any of the patches in Fig. 3 would work well in cross-validation samples almost immediately during training since both foreground content and background noise are present. In contrast, for image patches with no foreground content and just structured noise, the denoising algorithm was extremely slow to converge. Such areas typically make up a larger portion of the entire document than the salient portions, and are arguably more important to denoise. In most recognition pipelines, there is a segmentation step followed by feature extraction, where improper regularization would yield otherwise meaningless components.


cérébrale (MPCC) à ses adhérents. "C'est un vrai carnage et un cri de colère!
Il faut que les coureurs et les managers se mobilisent", explique à France 3,
Roger Legeay, le président du MPCC.

(a) Full original image

(b) Image after some preprocessing

(c) Full denoised image

Fig. 4: Example complete denoised image from the proposed neural network.

As such, we slightly adjusted cost function $\mathcal{L}$ for batches of size B by optimizing the neural network $f(\cdot)$ with a slight modification of the final cross-entropy penalty to consider situations where noise patterns had been added to blank patches of the input:

$$\mathcal{L} = -\sum_{i}^{B} (1 + \lambda \mathbb{I}(x_i)) x_i^T \log\left(f(g(x_i^{(2d)}, n_i))\right), \qquad (2)$$

where n_i is the noise added to image patch x_i by function $g(\cdot, \cdot)$, λ is a parameter that we set, and the indicator function $\mathbb{I}$ denotes when x_i is a patch with no handwriting. We use the superscript $^{(2d)}$ to distinguish $x_i \in \mathbb{R}^{3136}$ from $x_i^{(2d)} \in \mathbb{R}^{56 \times 56}$ for simplicity. Though not shown in (2), we used momentum and Nesterov momentum. For completeness, the neural network input is generated by:

$$g(x_i^{(2d)}, n_i) = \frac{1}{255} \min\left(2 \cdot 255 - x_i^{(2d)} - n_i, 255\right) \qquad (3)$$

and x_i, is a reshaped/flattened version of $x_i^{(2d)}$. Noise n_i is randomly sampled from a set of images downloaded from the web, and added to document patches with some randomly determined adjustments to position, intensity, and so on.

To obtain images like Fig. 4, the deep net itself was convolved over the entire image in raster scan fashion. This required an overlap between each window proportional to the convolutional filter size, in order to consider edge effects. Since our CNNs use input patches of 56×56 or 120×120, we set overlap between successive windows to 10 or 20, respectively.

Table 1: STIL Handwriting: Top-1 by Guise (%)

Approach	lined pen	blank marker	colored pencil	off-white pencil	graph pen	lined pencil	yellow lined pen	disguised blank pen	All
FLASH ID	33	0	8	42	25	8	33	33	32
CNN	44.4	10.0	45.6	50.0	46.7	15.6	14.4	14.4	30.1
Seam [2]	57.8	40.0	75.6	73.3	60.0	53.3	43.3	23.6	53.4
Vert [2]	51.1	37.8	72.2	71.1	62.2	54.4	38.9	22.4	51.3
Hough-based	62.2	36.2	72.4	71.8	22.8	54.7	44.2	22.7	48.4
Pixwise-SVM [29],Seam	78.2	11.5	73.3	67.2	50.6	55.2	45.2	28.7	51.2
DeNN,Seam	90.0	21.1	77.8	82.2	88.9	62.2	51.1	34.4	63.5
DeNN,Vert	87.8	20.0	78.9	81.1	87.8	57.8	55.6	30.0	62.4

Table 2: Average Error

	Lined	Graph	Stains	All
Hough-based	0.077	0.090	0.0892	0.082
Adaptive [30]	1.5e-3	7.2e-3	3.2e-2	6.2e-3
Pixwise-SVM [29]	4.14e-4	6.31e-4	4.21e-4	5.52e-4
DeNNs [2]	3.24e-4	8.20e-4	1.32e-4	4.23e-4

4. RESULTS

The evaluation of our denoising neural network is done through the top 1 criterion used by the commercial and academic communities[2] in Table 1 and the average error in Table 2.[3] In considering the effects of denoising, we have tried other denoising algorithms, including [29, 31, 18] paired with seam features [2]. We also compared to conventional signal processing (e.g., wavelets and discrete cosine transform approaches) and machine learning techniques.

Overall accuracy rates for off-the-shelf adaptive denoising [30] actually *decreased* performance using seam features [2] to 55.0% from 72.0% on soft top-10 and 7.2% from 9.7% for hard top-3, and 31.1% from 53.4% for top-1 accuracy. We also tested against an implementation of [18] for rule-line removal, but this also degraded manual feature performance (soft top-10 68.3%, hard top-3 3.8%), even after experimenting with threshold values. The most successful denoising algorithms are shown in the Tables.

Denoising the image with the proposed denoiser helps in every guise category except when writers are asked to use markers and write on a blank sheet of paper. Here, denoising actually *reduced* accuracy. Qualitative visual inspection did not identify any discernible difference between the images that would cause such a drop in performance. In fact, using markers fooled the ranking algorithm *more* than when writers were asked to deliberately disguise their hand. Our conversations with subject matter experts on this interesting point revealed that markers could change the mindset of the writer, mask writing styles due to implement thickness, and lower the information content of the writing itself. Further software development requires incorporation of this additional reasoning and is a subject for future work.

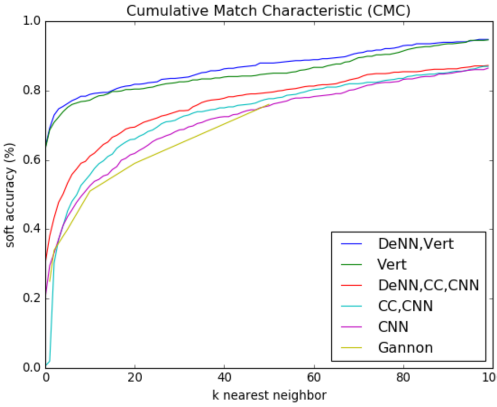


Fig. 5: Comparing all feature extraction techniques by varying k from 0 to 100.

5. CONCLUSIONS

Image denoising with encoder/decoder networks has improved accuracy, except when markers are used as the writing implement. Immediate avenues for improvement involve coupling denoising with segmentation algorithms [26] and discriminating RNNs like [6, 5]. Other directions aiding performance include the global (document-level) analysis and denoising, perhaps where features [32, 33] can be applied in tandem with current local features. While deep learning features did not perform well in our experiments, we view this primarily as a data issue, and there are paths forward using semi-supervised learning techniques (e.g. [34]).

[2] The fraction of times the best guess author correct for a given document.

[3] Average Error = $\frac{1}{MN} \sum_{i,j}^{M,N} I_{i,j} \oplus \hat{I}_{i,j}$

6. ACKNOWLEDGEMENTS

We would like to thank Rajiv Jain and David Doermann for their discussions, code, and general responsiveness. Without their input and work, despite being external collaborators, the results achieved would not have been possible. Additionally, we would like to thank Stefan Fiel for his initial conversations.

7. REFERENCES

[1] Sciometrics: Flashid: Language-independent handwriting biometric. `http://sciometrics.com/products/sciometrics-flash-id.html` (2015)

[2] Jain, R., Doermann, D.: Writer identification using an alphabet of contour gradient descriptors. In: Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. (Aug 2013) 550–554

[3] Fiel, S., Sablatnig, R.: Writer identification and writer retrieval using the fisher vector on visual vocabularies. In: Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. (Aug 2013) 545–549

[4] Fiel, S., Sablatnig, R.: Writer identification and retrieval using a convolutional neural network. In: Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II. (2015) 26–37

[5] Yang, W., Jin, L., Liu, M.: Deepwriterid: An end-to-end online text-independent writer identification system. CoRR **abs/1508.04945** (2015)

[6] Graves, A., Fernández, S., Schmidhuber, J.: Multi-dimensional recurrent neural networks. CoRR **abs/0705.2011** (2007)

[7] Louloudis, G., Stamatopoulos, N., Gatos, B.: ICDAR 2011, writer identification competition. 11th International Conference on Document Analysis and Recognition 1475–1479

[8] Louloudis, G., Gatos, B., Stamatopoulos, N., Papendreou, A.: ICDAR 2013 writer identification competition. 12th International Conference on Document Analysis and Recognition (2013) 1397–1401

[9] Malik, M.I., Ahmed, S., Marcelli, A., Pal, U., Blumenstein, M., Alewijns, L., Liwicki, M.: Icdar2015 competition on signature verification and writer identification for on- and off-line skilled forgeries (sigwicomp2015). In: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on. (Aug 2015) 1186–1190

[10] Laboratory, S.T.I.: Science technology integration laboratory (2013) Data File, Unpublished Dataset, Cited with Permission.

[11] Ball, G.R., Srihari, S.N., Srinivasan, H.: Segmentation-Based And Segmentation-Free Methods for Spotting Handwritten Arabic Words. In Lorette, G., ed.: Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule (France), Université de Rennes 1, Suvisoft (October 2006) http://www.suvisoft.com.

[12] LeCun, Y., Cortes, C., Burges, C.J.: The MNIST database of handwritten digits. `http://yann.lecun.com/exdb/mnist/` (2015)

[13] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09. (2009)

[14] Jain, R., Doermann, D.: Offline writer identification using k-adjacent segments. In: Document Analysis and Recognition (ICDAR), 2011 International Conference on. (Sept 2011) 769–773

[15] Kaggle: Frobnication tutorial. `https://www.kaggle.com/c/denoising-dirty-documents` (2015)

[16] Bach, K., Lichman, M.: Uci machine learning repository (2013) University of California, School of Information and Computer Science.

[17] Snchez, J., Romero, V., Toselli, A., Vidal, E.: ICFHR2014 competition on handwritten text recognition on transcriptorium datasets (HTRtS). In: International Conference on Frontiers in Handwriting Recognition (ICFHR). (2014) 181–186

[18] Abd-Almageed, W., Kumar, J., Doermann, D.: Page rule-line removal using linear subspaces in monochromatic handwritten arabic documents. In: Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on. (July 2009) 768–772

[19] Imtiaz, S., Nagabhushan, P., Gowda, S.D.: Rule line detection and removal in handwritten text images. In: Proceedings of the 2014 Fifth International Conference on Signal and Image Processing. ICSIP '14, Washington, DC, USA, IEEE Computer Society (2014) 310–315

[20] Cao, R., Tan, C.L.: Separation of overlapping text from graphics. In: Document Analysis and Recognition (ICDAR), 2001 6th International Conference on. (2001) 44–48

[21] Kumar, J., Doermann, D., Abd-Almageed, W.: Rule line dataset. `http://lampsrv02.umiacs.umd.edu/projdb/project.php` (2009) University of Maryland: Language and Media Processing Laboratory.

[22] Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: Proceedings of the 21th International Conference on Artificial Neural Networks - Volume Part I. ICANN'11, Berlin, Heidelberg, Springer-Verlag (2011) 52–59

[23] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I. (2014) 818–833

[24] Zeiler, M.D., Taylor, G.W., Fergus, R.: Adaptive deconvolutional networks for mid and high level feature learning. In: IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011. (2011) 2018–2025

[25] Xu, L., Ren, J.S., Liu, C., Jia, J.: Deep convolutional neural network for image deconvolution. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., eds.: Advances in Neural Information Processing Systems 27. Curran Associates, Inc. (2014) 1790–1798

[26] Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. CoRR **abs/1511.00561** (2015)

[27] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., eds.: Advances in Neural Information Processing Systems 25. Curran Associates, Inc. (2012) 1097–1105

[28] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. CoRR **abs/1501.00092** (2015)

[29] Kumar, J., Doermann, D.S.: Fast rule-line removal using integral images and support vector machines. In: 2011 International Conference on Document Analysis and Recognition, ICDAR 2011, Beijing, China, September 18-21, 2011. (2011) 584–588

[30] Bradley, D., Roth, G.: Adaptive thresholding using the integral image. J. Graphics Tools **12**(2) (2007) 13–21

[31] Priest, C.: Image processing + machine learning in r: Denoising dirty documents tutorial series. `http://blog.kaggle.com/2015/12/04/image-processing-machine-learning-in-r-denoising-dirty-doc` Accessed: 2013-07-08.

[32] Siddiqi, I., Vincent, N.: Combining global and local features for writer identification. Proceedings of the 11. Int. Conference on Frontiers in Handwriting Recognition, Montreal (2008)

[33] m. Cheung, Y., Deng, J.: Off-line text-independent writer identification using a mixture of global and local features. In: Computational Intelligence and Security (CIS), 2011 Seventh International Conference on. (Dec 2011) 1524–1527

[34] Rasmus, A., Valpola, H., Honkala, M., Berglund, M., Raiko, T.: Semi-supervised learning with ladder network. CoRR **abs/1507.02672** (2015)