

Print Defect Mapping with Semantic Segmentation

Augusto C. Valente Cristina Wada Deangela Neves Deangeli Neves
Fábio V. M. Perez Guilherme A. S. Megeto Marcos H. Cascone Otavio Gomes
Instituto de Pesquisas Eldorado
firstname.lastname@eldorado.org.br

Qian Lin
HP Labs, HP Inc.
qian.lin@hp.com

Abstract

Efficient automated print defect mapping is valuable to the printing industry since such defects directly influence customer-perceived printer quality and manually mapping them is cost-ineffective. Conventional methods consist of complicated and hand-crafted feature engineering techniques, usually targeting only one type of defect. In this paper, we propose the first end-to-end framework to map print defects at pixel level, adopting an approach based on semantic segmentation. Our framework uses Convolutional Neural Networks, specifically DeepLab-v3+, and achieves promising results in the identification of defects in printed images. We use synthetic training data by simulating two types of print defects and a print-scan effect with image processing and computer graphic techniques. Compared with conventional methods, our framework is versatile, allowing two inference strategies, one being near real-time and providing coarser results, and the other focusing on offline processing with more fine-grained detection. Our model is evaluated on a dataset of real printed images.

1. Introduction

The absence of print defects is one of the most significant factors that influence customer satisfaction when experiencing a printer product. For this reason, mapping such defects before the final revision in preparation for market launch is extremely important to the manufacturing printer industry.

In large-scale quality control processes, print defect detection still requires human experts in an approach that is repetitive, time-consuming, and prone to human error and subjective evaluation [18]. In order to improve the efficacy and efficiency of that process, automated methods to map print defects are highly desired.

There is a wide variety of print defects that may arise

in the printing process, such as streaks, banding, spots, and patches. The causes of each of them vary according to the printer technology type. In this paper, we mainly focus on banding and streaks, since they are easier to be synthesized, and are both common to inkjet and laser electrophotographic printers.

In inkjet printers, banding is generally caused by small electromechanical or chemical problems in the printer components, such as damaged nozzles in the print heads, or ink density variation [3]. Streak defects, on the other hand, can be a result of misaligned or contaminated print heads.

In laser printers, the non-uniform line spacing observed in banding defects is mainly caused by fluctuations in the velocity of the optical photoconductor (OPC) drum [6]. Moreover, imperfections in the cleaning mechanism on the intermediate transfer belt (ITB), damages in the belt surface itself, or the color cartridges may generate streak artifacts on printings.

Some works have proposed automated intelligent systems for assessing print quality [17]. Conventionally adopted approaches [16, 18, 19, 20] focus on print quality defect detection using standard machine vision methods, coarsely indicating the location of print defects. However, none of these solutions employ defect detection at the pixel level and map different types of defects in a single model. In this work, we specifically attack this problem, which we refer to as *print defect mapping*.

Since 2012, deep convolutional neural networks (DCNNs) have achieved remarkable results in many computer vision tasks, including image classification, object detection, and semantic segmentation [4, 11, 14]. In this paper, we employ semantic segmentation to detect and map print artifacts at the pixel level. We use DeepLab-v3+ [5] as the backbone of our framework. We chose this particular architecture because of its potential to simultaneously retrieve rich contextual information and sharp object bound-

aries, which are essential for characterizing print defects.

While deep learning techniques show impressive results in multiple computer vision tasks, they usually require large annotated datasets. Manually collecting and annotating data can be both expensive and time-consuming. When it comes to acquiring well-annotated data for training semantic segmentation models, which requires fine-grained label maps, that task becomes even more difficult and costly [23].

To overcome that problem, we create artificial print artifacts samples with image processing techniques to generate our training dataset. We initially collect a set of digital images with diverse content and add the dark streaks and the color bandings. In addition, we simulate the effect of printing and scanning documents by estimating color transformations from real devices. With that, we aim to make our training data close to real use cases.

Inspired by works in image quality assessment [2, 7], we also implement two different paradigms for print defect mapping: No-Reference (NR) and Full-Reference (FR). In the first one, the framework receives only the printed images, not requiring the original printable image. In the second paradigm, both the printed and the original digital images form the input, which may give additional cues to where the defects may occur.

Our main contributions are as follows:

- We present an end-to-end deep learning framework for pixel-wise print defect mapping. To the best of our knowledge, this is the first work to map print defects at the pixel level.
- We propose a pipeline for simulating a print-and-scan effect and to artificially create print artifacts in diverse background images for two types of defects—dark streaks and banding. The proposed pipeline also leverages automatic pixel annotation, resulting in cost-free data in terms of human annotation.
- Our framework supports two input modes, Full-Reference (FR) and No-Reference (NR), that can be interchangeably used with small changes to its architecture. Moreover, it also adopts two inference strategies: *Resized image* and *Patch-based*, taking a trade-off between speed and quality into account.

2. Literature Review

2.1. Print Defect Mapping

Previous approaches to Print Defect Mapping focus only on the coarse detection of defects. Generally, no more than one defect type is targeted, eliminating the need for mapping each input to a category. Furthermore, the detection task usually relies on the application of a set of standard image operations and similarity measures.

Wang *et al.* [18] investigated the detection of local print defects in the form of spots and patches by analyzing the standard deviation of pre-extracted regions of interest. For each region, they use the valley-emphasis threshold to find potential local defects. Then, a set of hand-crafted features, such as edge strength and average RGB value, are extracted and fed into an SVM model to determine whether the overall quality of the printed image is good enough. They evaluate their model in terms of accuracy for global quality prediction, instead of localized defect detection as we propose.

Zhang *et al.* [21] adopted a similar approach to detect light streaks in printer test images. Firstly, they pre-process standard print test pages in order to select only the smooth areas of the image. They divide these areas into blocks and compute the projection of the distance between each pixel within a block and the average pixel value in the block to detect streaks. Finally, they use logistic regression to remove false-positives from the previous step.

Spivakovsky *et al.* [16] proposed a SSIM-based comparison between printable images and their scanned versions to detect defects. Similarly, Xiao *et al.* [19] aligned reference and scanned images with feature-based registration, and compared them based on color difference distribution to detect text fading. Yangping *et al.* [20] proposed a system based on grayscale and gradient differences for template matching.

While standard image processing techniques may work when detecting a specific type of artifact, it is challenging to cover multiple occurrences of different print defect types in the same image. Unlike conventional methods, our framework takes full advantage of deep neural networks, which have proved to generalize remarkably well in several computer vision tasks, to map both dark streaks and color banding defects pixel-wise.

2.2. Semantic Segmentation

Semantic segmentation is a computer vision task that consists of assigning a label to every pixel in an image. Results on public benchmarks have considerably improved with the use of fully-convolutional DCNNs, such as FCN [11], U-Net [14], and DeepLab [4]. Applications of semantic segmentation are diverse, including medical image analysis [14], remote sensing [22] and autonomous driving [15]. In this work, we also employ semantic segmentation techniques, specifically DeepLab-v3+, to the problem of print defect mapping.

The motivation behind the application of semantic segmentation to print defect mapping is the precision gains in locating and identifying the defective patterns using a more detailed inspection. Besides, this approach would provide better generalization since the model will focus on deviations of the ideal printer operation rather than on specific characteristics of the defect.

2.3. Print Defect Simulation

Obtaining real data to train a print defect mapping algorithm is hard for several reasons. One would first need to use multiple defective printers and print several documents in order to obtain a good range of defects and backgrounds. Additionally, carefully annotating defect locations in the images would take a long time. As deep learning methods require significant amounts of labeled data, we generate synthetic training data using image processing techniques. That consisted in simulating the actual print defects and also the overall look of a printed-and-scanned page. We detail the implementation in the next section.

We based our artifact model in [6], where the authors created a framework to reproduce defects from laser printers. They estimate parameters from actual print defects and use both geometrical modeling and scanned templates. However, their approach still depends on printing the pages to achieve realism.

3. Methodology

3.1. Synthetic data

The ability to create realistic synthetic artifacts has many advantages. First, it can generate large amounts of training data, which is a demand for deep neural networks. Second, it can produce different types of artifacts, such as fine pitch banding and streaks, that generally occurs uniformly on the printed page; or repetitive artifacts, like spots, which may occur randomly across the page.

We address two frequent types of print defects: dark streaks and color bandings. They essentially result in horizontal or vertical lines with an excess or lack of a certain color component. Dark streaks are usually thinner textured lines, while color bandings are rectangular regions with near homogeneous drop or excess in a color channel. Figure 1 shows examples of such defects.

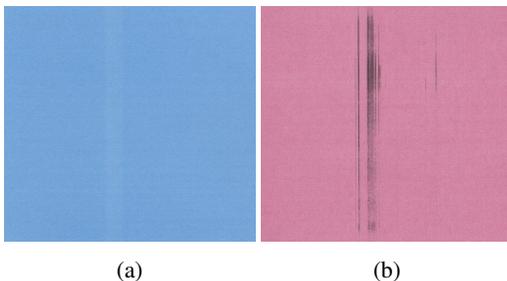


Figure 1: Print defect examples: (a) color banding with an increase in the magenta channel, and (b) dark streak.

To simulate dark streaks, we blend a random thin rectangular image region with a dark-colored texture pattern. This texture pattern was procedurally generated by applying Per-

lin noise to create a gray level perturbation. The pattern varies the color intensity, randomly along the streak length, and the edges are set to be lighter than the center. Figure 2 shows an example of a simulated dark streak.

When simulating color banding, on the other hand, we first convert the input image to CMYK, which is the color space adopted in the printing process. After that, we overlay rectangular regions that affect each color channel separately and have slight variations in color. Pixel intensities can be either decremented/incremented to simulate lack/excess. The amount of pixel intensity modification varies across the band's width, being either brighter or darker in the middle. This variation is subject to a bimodal Gaussian distribution, as shown in Figure 2.

The last step in our synthetic data pipeline is to imitate the appearance of printing and scanning an image. This process consists of learning the RGB tone curves of printer and scanner devices. We printed and scanned test pages in different devices and modeled the color transformation as a polynomial, whose coefficients are determined by applying a linear least square (LLS) regression between the original digital image and its respective printed-and-scanned version on a real device.

$$F(x, y, R, G, B) = a_0 + a_1y + a_2x + a_3R + a_4G + a_5B + a_6yR + a_7yG + a_8yB + a_9xR + a_{10}xG + a_{11}xB + a_{12}RG + a_{13}RB + a_{14}GB + a_{15}RGB \quad (1)$$

The input variables of the polynomial function shown in Equation 1 are, respectively, the normalized image x and y coordinates, and its respective red, green, and blue color channels. Figure 2 shows an example result of the print-scan simulation process.

3.2. Network architecture

As mentioned in Section 2, considering the potential of print defect mapping using a semantic segmentation approach, we investigated some DCNN architectures, particularly U-Net [14] and DeepLab-v3+ [5], since both attain state-of-art results in other tasks. In our preliminary experiments, DeepLab-v3+ outperformed U-Net when segmenting thinner lines. For this reason, we decided to use DeepLab-v3+ architecture as a base for our model.

DeepLab-v3+ [5] combines the benefits of two methods: an atrous spatial pyramid pooling (ASPP) module [4] and an encode-decoder structure. In encoding-decoding configurations, the network constructs a feature representation of the image in lower resolution and then applies trained fractionally-strided convolutions to recover label maps in higher resolution. The low-resolution intermediary representation provides good context understanding but hinders the detection of small or thin objects. In order to compensate for this issue, the ASPP module, based on dilated or

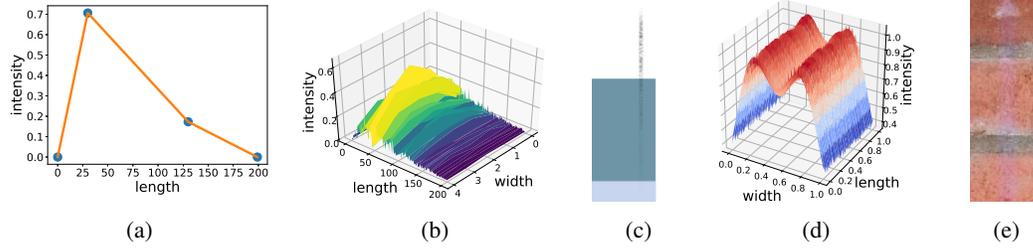


Figure 2: Examples of pixel intensity profile of dark streak and color banding simulations. Dark streak: (a) intensity profile across the defect length, (b) 3D intensity profile and (c) the corresponding simulated defect. Color banding: (d) 3D intensity profile and (e) color banding defect applied to the magenta channel. Best viewed on screen.



Figure 3: Print-scan simulation example: (a) original image and (b) result after the effect is applied.

atrous convolutions, leverages its increased receptive field to avoid sacrificing spatial resolution, and recovers labeled pixels by applying bilinear upsampling. Less use of downsampling helps fine-grained detection, but the context understanding is not as broad as in encoder-decoder networks. Another advantage of using DeepLab-v3+ architecture is that it does not require huge datasets, as it was trained on the Pascal VOC 2012 dataset [5].

To configure DeepLab-v3+ for print defect mapping, we performed variations in the architecture hyperparameters and evaluated their impact on the defect identification.

The ASPP module performs multiple pooling scales through four parallel atrous convolutions using a set of 256 filters of size 1×1 or 3×3 in dilated configuration and using depth-wise separable convolutions. The encoder-decoder structure, in its turn, has a DCNN in the encoder backbone, performing atrous convolution followed by the ASPP structure. In the decoder module, the DCNN output is concatenated with the bilinearly upsampled ASPP output. The concatenated result is convolved with a 3×3 filter and bilinearly upsampled to the desired output.

We tested different output strides (8 and 16), a parameter that controls the ratio of input image resolution to the output image resolution. In general, an output stride of 8 gave better results, but at the cost of slower training. In the DCNN, we verified the use of ImageNet-pretrained ResNet-50 and ResNet-101 models. The former achieved similar performance while being faster to train.

ResNet-50 is composed of building blocks, each one formed by stacked sets of 3 convolutional layers. We tested different dilation values— $[1, 1, 1]$, $[1, 2, 1]$, and $[1, 2, 4]$ —in these 3 layers, achieving the best performances with $[1, 2, 1]$, and $[1, 1, 1]$ in some cases.

3.3. No-Reference and Full-Reference metrics

We designed the framework of this work to support two metrics used in Image Quality Assessment (IQA).

IQA is a broad research field that focuses on quantitative representing the human perception of image quality. IQA is generally divided into two areas: Full-Reference (FR) and No-Reference (NR). FR explores metrics that quantify the quality of a distorted image, given its reference image. On the other hand, NR metrics measure the quality of distorted images without requiring their reference image.

FR methods are usually based on image operations. The existence of reference images provides useful information about their distorted counterparts, resulting in a wide range of evaluation possibilities. In contrast to that, having no ground truth image to establish quality criteria makes NR a more challenging problem, which has seen more recent advances with the rise of deep learning [2, 8, 10, 12].

Both FR and NR extend their application to other fields in which measuring output image quality is essential to improve the end-user viewing experience, such as Image Capturing and Image Processing. Since print defects are directly linked to the overall quality of printed images, we added both sets of metrics.

3.4. Generating training data

Our training set consists of 2949 digital images—including documents, photos, and slides—with minimum dimensions of 1920×1080 . The samples are preprocessed online during training by a set of techniques, described in Section 3.1, that generate their potentially defective printed versions. Thus, the virtual number of training samples is much higher than the raw number of digital images in the dataset, since each image results in potentially infinite defective versions of itself.

We first simulate the appearance of printing and scanning effects on digital images. Then, we randomly decide if the image will be defective based on an assigned probability of 0.9. If so, we randomly add the artifacts described in Section 3.1 to it. The probability of streaks occurrence varied from 0.3 to 0.5, while the sum of channel banding probabilities is its complementary. Finally, we generate the ground truth mask indicating which pixels have artifacts.

We resized all processed images and their respective masks to a common size of 1920×1080 pixels using bicubic interpolation, and keeping their aspect ratio. Their respective masks are interpolated with nearest neighbors to preserve label information.

3.5. Training pipeline

To train the NR network, we extract random patches of 513×513 pixels of the processed image and its respective mask. Then, we feed this set into the network, which considers the corresponding set of masks as the ground truth. Figure 4a shows a diagram of the full training pipeline for the NR network.

For FR, we extract random patches of 513×513 pixels of the reference/defective images and the mask. We concatenate the extracted patches from reference and defective images forming a 6-channel image, which is fed to the network, as shown in Figure 4b. The change to the network input introduces more weights in the first layer. To initialize them, we simply copied the pretrained weights from the first three channels, since the new channels also correspond to an RGB image. We also tried to use the difference between the

reference and target images as the network input. However, we found that concatenation provided better results.

We used the multi-class cross-entropy loss and weighted pixels labeled as background with 0.05 to account for the imbalance between background and defects. We explored the intersection over union loss [1] but did not get better results. We also tried to train a one-vs-all class setting, implemented as the sum of the binary cross-entropy loss for each class against all other classes but got worse results.

We used the Adam optimizer [9] with an initial learning rate of 10^{-4} , and weight decay of 10^{-5} . We adopted a strategy of accumulating gradients between batches. This allows us to achieve a larger effective batch size. We used a batch size of 5, and accumulate gradients for 10 iterations, resulting in an effective batch size of 50.

We employed online data augmentation during training (90-degree clockwise and counter-clockwise rotation and horizontal flipping). We normalized each image by subtracting its mean and dividing by its standard deviation. We randomly split the dataset into 90% for training and 10% for validation. We also verified if we increased the number of training samples to 11 102 would give better results, but no significant improvement was observed.

3.6. Inference strategy

We evaluate our framework considering two approaches: *resized image* and *patch-based*. In the first, we resize images down to a resolution of 1280×720 with bicubic interpolation before feeding it to the model. In the second, we extract patches of 513×513 pixels from the original image

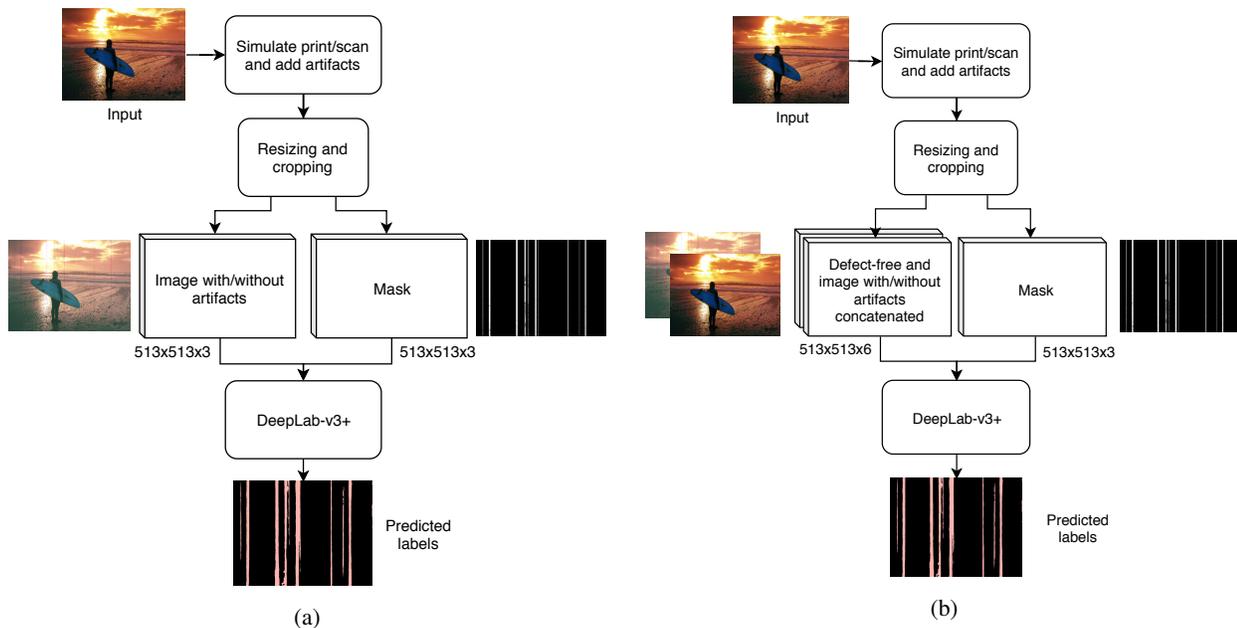


Figure 4: Training pipeline for (a) no-reference and (b) full-reference methods.

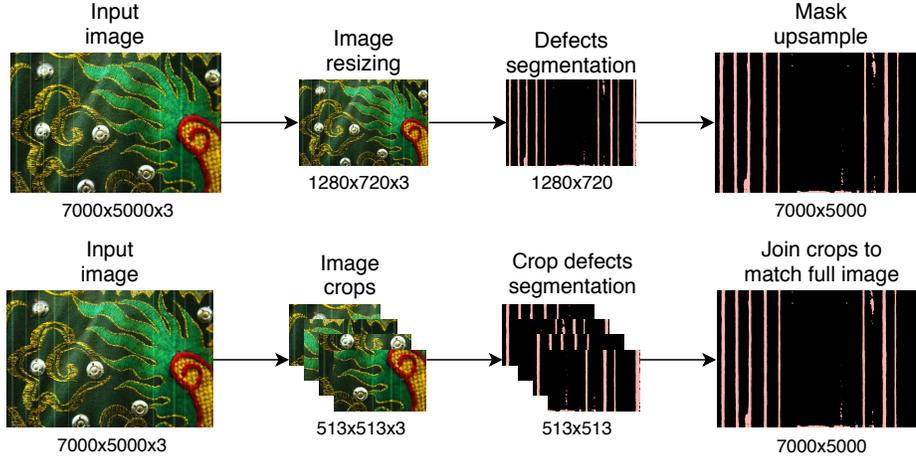


Figure 5: Inference pipeline with resized image approach (top) and patch-based approach (bottom).

and run inference separately on each of them. In the end, we combine the predictions from each image patch.

These inference strategies are supported by both FR and NR pipelines and are illustrated in Figure 5.

4. Experiments

We evaluated our method on two datasets, one with semi-synthetic defects and another with real samples from defective printers, both on an Intel Xeon E5-2620 CPU and NVIDIA Titan X Maxwell GPU. Due to differences between the two datasets, we trained two types of models for evaluation, one considering banding defects with multiple labels and the other associating banding with a single label.

Table 1: Evaluation on semi-synthetic and real data for both inference strategies and full-reference (FR) and no-reference (NR) methods. Values indicate the average and standard deviation for four runs, to account for randomness.

Method		mIoU	Time (s /image)
Semi-synthetic data			
Resized Image	FR	0.25 ± 0.02	2.5
	NR	0.32 ± 0.01	1.8
Patch-based	FR	0.33 ± 0.05	20.5
	NR	0.45 ± 0.04	20.5
Real data			
Resized image	FR	0.35 ± 0.03	1.7
	NR	0.43 ± 0.02	1.7
Patch-based	FR	0.37 ± 0.03	17.8
	NR	0.49 ± 0.04	20.0

For the semi-synthetic dataset, we added synthetic artifacts to 103 background images using the process described in Section 3.1, printed and scanned them at 600 dpi (around 7000×5000 pixels), and aligned the digitized and reference images with feature matching. After alignment, the images had around 2000×3000 pixels. The printers and scanners used in this dataset were the same used to design our print-scan simulation described in Section 3.1. Manual annotation was not needed as we recorded defect locations during artifacts generation. Multiple labels were used for banding, depending on the affected color channel (cyan, magenta, yellow, or black), allowing us to examine more detailed predictions for this type of defect. Results for the semi-synthetic dataset are shown in Table 1 and Figure 6.

The real dataset served as a benchmark of how well the synthetic training data was translating to the real-world. It has 30 scanned images with real banding and streaks defects, collected from defective printers, and scanned at 600 dpi. We manually annotated the defects on each scanned image and manually aligned the digitized and reference images. After alignment, the images had average dimensions of 2500×3000 . A single label was assigned to banding defects, due to the difficulty of manually analyzing in retrospect which color channel had been affected. Table 1 and Figure 7 show the results for the real dataset.

Following the semantic segmentation literature, we evaluated our models with the mean intersection over union (mIoU) over classes. Previous works on print defect detection [18, 19, 21] do not evaluate their approaches in terms of mIoU, and each of them is evaluated differently, making direct comparisons among works hard.

5. Discussion

Table 1 shows that the NR method outperforms FR on all datasets and inference approaches. At first glance, this re-

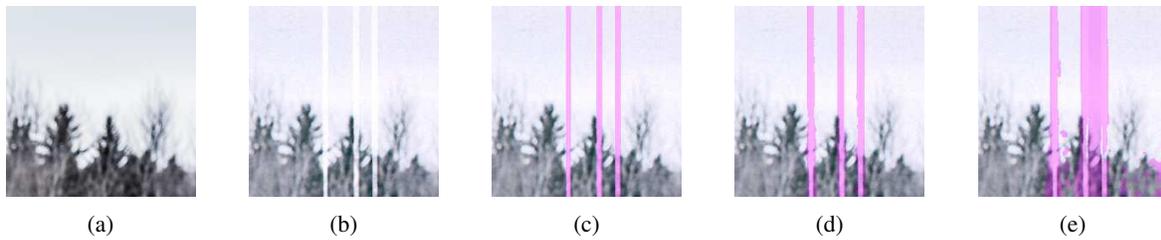


Figure 6: Sample results on the semi-synthetic dataset with patch-based inference mode: (a) reference image input, (b) defective image input, (c) ground truth overlay, predictions from (d) no-reference and (e) full-reference model. Banding defects are marked as magenta. Best viewed on screen.

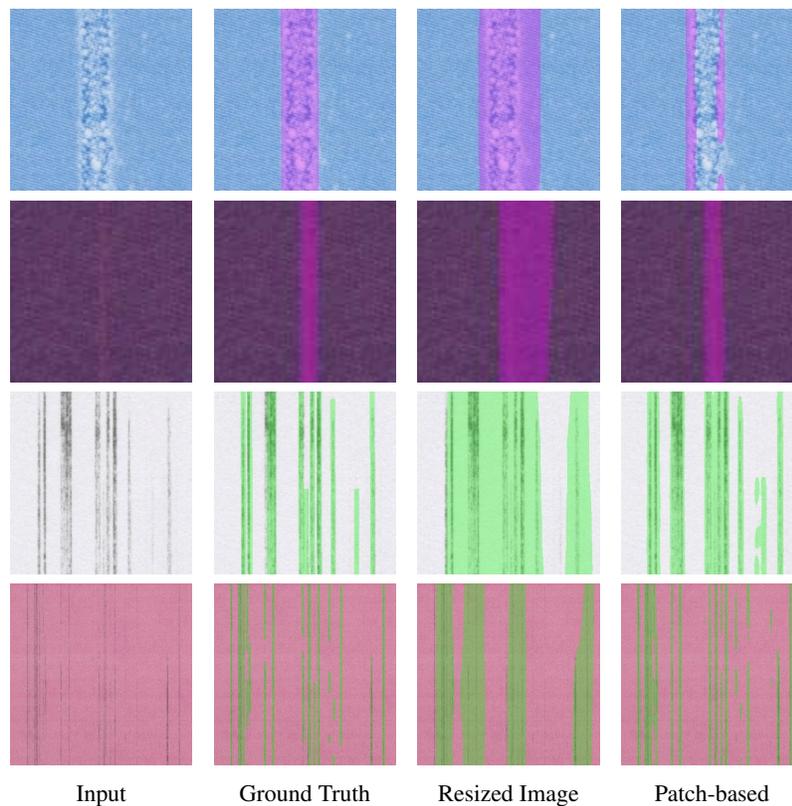


Figure 7: Sample results on the real dataset for the no-reference method with resized image and patch-based inference strategies. Streaks are marked as green and banding defects are marked as magenta. Note that patch-based inference produces finer results than resized image. Best viewed on screen.

sult seems surprising as the reference image could provide cues for the model. Nevertheless, we hypothesize that it may be difficult for the FR network to realize that it should take into account the reference and scanned images separately when predicting the defective pixels since concatenating them does not explicitly convey that.

The human eye is capable of detecting defects without a reference image. Hopefully, a machine learning model may be able to do the same. State of the art in computer vision shows that deep learning methods usually work well

with minimal supervision than other methods, which may strongly rely on handcraft features. It may be the case that the NR method works better with DCNNs because the input space (3 channels) is simpler than in the FR method (6 channels), facilitating convergence.

Figures 6 and 7 show that the patch-based inference provided more fine-grained detection compared to the resized image mode. However, this advantage comes with the cost of it being considerably slower (Table 1). Choosing between one of the inference methods depends on the goal

of the problem. If the system must be fast, resized image is the best choice, while patch-based inference is favored when quality is more important.

Previous works [18, 21] require input images to have specific aspects such as the ones of standard print test pages, which are mostly solid-color filled. Conversely, our framework is more general, working for documents, photos, and combinations of both. Furthermore, our framework does not require calibrating thresholds for different occurrences of streak defects, as in [21].

Considering the lack of similar works, we found it suitable to use an intermediate step in the framework proposed by Wang *et al.* [18] as a base for comparison. Their method provides a fine defect detection map from which we could compute the IoU and compare it with ours. Due to input restrictions in their framework, we evaluated it only with solid-color filled defective images, as depicted in Figure 8.

We chose *patch-based* inference as it provides better detection of streaks, and the NR method to agree with Wang *et al.*'s detection method, which does not require the reference image. Our model was able to detect more streaks, even the fading ones, when compared to their approach. This can be verified by a higher IoU value of 0.299, outperforming their mapping, which achieved an IoU of 0.183. Despite that, we would like to emphasize that Wang *et al.*'s aim was different from ours. Their defect detection was an intermediate step to locate and get the defect features to subsequently feed a predictor and get an overall quality assessment of the printed image.

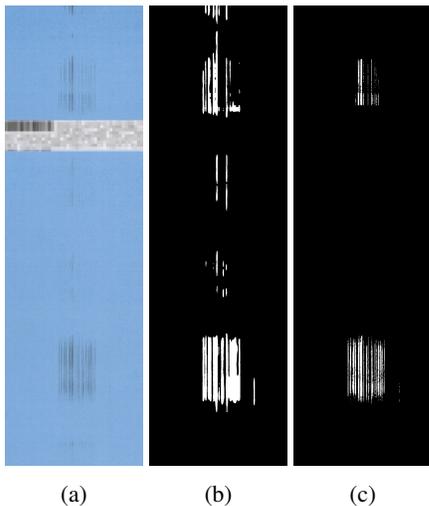


Figure 8: Comparison between methods: (a) real image crop with streaks (b) map of our method (c) method by [18]. Best viewed on screen.

Although we use the mIoU to evaluate the performance of our models, we realized from our experience annotating the evaluation dataset that two human annotators might

produce different segmentation masks, introducing biases towards annotation. This problem is recurrent in other applications of semantic segmentation [13]. In print defect mapping, since defects may be very thin, annotation agreement may even be smaller. Thus, for more realistic evaluation, we may use some metric relaxation in future works.

6. Conclusion and Future Works

We presented an end-to-end technique that uses DCNNs to perform semantic segmentation for print defect mapping. To the best of our knowledge, this is the first work to describe a fully-automated pipeline for detecting print defects at the pixel level.

We simulated printing and scanning of documents, thus eliminating the need for this laborious task. Moreover, our method continuously simulates data, making our pipeline cost-free in terms of annotation. Our results show that our models can transfer the knowledge learned from synthetic data to real print defects, demonstrating its applicability for print quality assessment in industry.

We can improve our framework by updating specific parts of the pipeline. The DCNN can be swapped to a more modern architecture. Models may also take advantage of attention mechanisms to improve the results by making the network focus more on specific areas of the image. The pre-processing pipeline can be further improved with better defects simulations and with stronger data augmentation. Additionally, performance can increase with larger datasets.

The framework can be improved by including more types of defects so that a single approach can be used to detect a wide range of print defects. Nevertheless, our results show that it is possible to have two defects in the same model, indicating that we may also get good results when adding more defect types. Since we use synthetic defects, the hard part would be to simulate the defects.

Alongside with the defect mapping, the framework could also predict a global quality score in a multi-task learning approach. A global quality score can be useful to quantify the overall condition of a printer and can be used to discard defective pages automatically. However, we need to define some heuristics in order to produce synthetic scores for the training set.

Acknowledgements

We thank Eric Maggard from HP Inc. at Boise, ID for providing the image dataset with real print defects, and Jianyu Wang, Terry Nelson, Renee Jessome, Steve Astling, Eric Maggard, Mark Shaw, and Jan P. Allebach for providing us the source code from [18].

References

- [1] M. Berman, A. Rannen Triki, and M. B. Blaschko. The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4413–4421, 2018.
- [2] S. Bosse, D. Maniry, K. Mller, T. Wiegand, and W. Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, Jan 2018.
- [3] J. C. Briggs. Banding characterization for inkjet printing. In *2000 Image Processing, Image Quality, Image Capture Systems Conference*, March 2000.
- [4] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, April 2018.
- [5] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision – ECCV 2018*, pages 833–851, Cham, 2018. Springer International Publishing.
- [6] W. Jang and J. P. Allebach. Simulation of print quality defects. *Journal of Imaging Science and Technology*, 49(1):1–18, 2005.
- [7] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, June 2014.
- [8] J. Kim and S. Lee. Fully deep blind image quality predictor. *IEEE Journal of Selected Topics in Signal Processing*, 11(1):206–220, Feb 2017.
- [9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
- [10] Y. Li, L. Po, L. Feng, and F. Yuan. No-reference image quality assessment with deep convolutional neural networks. In *2016 IEEE International Conference on Digital Signal Processing (DSP)*, pages 685–689, Oct 2016.
- [11] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, June 2015.
- [12] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu, and Y. Zhang. Blind predicting similar quality map for image quality assessment. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6373–6382, June 2018.
- [13] V. Ribeiro, S. Avila, and E. Valle. Handling inter-annotator agreement for automated skin lesion segmentation. *arXiv preprint arXiv:1906.02415*, 2019.
- [14] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [15] M. Siam, M. Gamal, M. Abdel-Razek, S. Yogamani, and M. Jagersand. Rtseg: Real-time semantic segmentation comparative study. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1603–1607, Oct 2018.
- [16] A. Spivakovsky, O. Perry, O. Haik, and A. Malki. Image defect detection, May 10 2018. US Patent US20180131815A1.
- [17] A. Verikas, J. Lundström, M. Bacauskiene, and A. Gelzinis. Review: Advances in computational intelligence-based print quality assessment and control in offset colour printing. *Expert Syst. Appl.*, 38(10):13441–13447, Sept. 2011.
- [18] J. Wang, T. Nelson, R. Jessome, S. Astling, E. Maggard, M. Shaw, and J. P. Allebach. Local defect detection and print quality assessment. *Electronic Imaging*, 2016(13):1–7, 2016.
- [19] Z. Xiao, M. Nguyen, E. Maggard, M. Shaw, J. Allebach, and A. Reibman. Real-time print quality diagnostics. *Electronic Imaging*, 2017(12):174–179, 2017.
- [20] W. Yangping, X. Shaowei, Z. Zhengping, S. Yue, and Z. Zhenghai. Real-time defect detection method for printed images based on grayscale and gradient differences. *Journal of Engineering Science & Technology Review*, 11(1), 2018.
- [21] R. Zhang, E. Maggard, R. Jessome, M. Cho, and J. P. Allebach. Block window method with logistic regression algorithm for streak detection. *Electronic Imaging*, 2019:300–1–300–7, 2019.
- [22] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, Dec 2017.
- [23] A. Zlateski, R. Jaroensri, P. Sharma, and F. Durand. On the importance of label quality for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1479–1487, 2018.