Simultaneous Detection of Regular Patterns in Ancient Manuscripts Using GAN-Based Deep Unsupervised Segmentation

Milad Omrani Tamrin(✉) and Mohamed Cheriet

École de technologie supérieure, 1100 Notre-Dame St W, Montreal, Qc, Canada
milad.omrani-tamrin.2@ens.etsmtl.ca, mohamed.cheriet@etsmtl.ca

Abstract. Document Information Retrieval has attracted researchers' attention when discovering secrets behind ancient manuscripts. To understand such documents, analyzing their layouts and segmenting their relevant features are fundamental tasks. Recent efforts represent unsupervised document segmentation, and its importance in ancient manuscripts has provided a unique opportunity to study the said problem. This paper proposes a novel collaborative deep learning architecture in an unsupervised mode that can generate synthetic data to avoid uncertainties regarding their degradations. Moreover, this approach utilizes the generated distribution to assign labels that are associated with superpixels. The unsupervised trained model is used to segment the page, ornaments, and characters simultaneously. Promising accuracies in the segmentation task were noted. Experiments with data from degraded documents show that the proposed method can synthesize noise-free documents and enhance associations better than the state-of-the-art methods. We also investigate the usage of overall generated samples, and their effectiveness in different unlabelled historical documents tasks.

Keywords: Ancient manuscripts · Degradations · Synthesize data · Unsupervised segmentation · Layout

1 Introduction

Ancient manuscripts play a vital source in cultural heritage as a fundamental role in history and social development. Since historical documents have deteriorated and aged, there are always difficulties in having a high level of acknowledgements rather than current analysis documents. Some ancient manuscripts have been available and copied on the online networks. However, it is quite challenging to find relevant and classified documents for users without having an electronic search-tool. Furthermore, to capture a categorized dataset manually, including different types of historical documents, is costly and would require high consumption of time to process. Such training recognition systems require

A. Del Bimbo et al. (Eds.): ICPR 2020 Workshops, LNCS 12667, pp. 279–291, 2021.
https://doi.org/10.1007/978-3-030-68787-8_20

a large amount of data [5]. Lately, the variety of manuscripts with different qualities has increased, which creates a more complex task. Besides, using heuristic image processing techniques to achieve artificial data is high maintenance [31]. The indexing technique is considered imperative to ensure having an automated search tool for document image classification. In contrast, the current approaches for document categorization depend on feature learning and hand-crafted features [4,6,7,9,16,21]. Recently, feature learning achieved state-of-the-art performance in Convolutional Neural Network (CNN). Despite CNN being recognized as state-of-the-art performance-based, it still found inaccuracy of recording document retrieval and text recognition. Besides, most of the feature learning methods are in supervised learning domains. Using such approaches requires large amount of label data to reach better results [1]. Preparing manually labelled data is time-consuming and needs expertise. Thus unsupervised approaches are good candidates to develop a robust textual information extraction such as page, ornaments, and texts without data annotation. Since having segmented data regarding ancient manuscripts is a binarization task, it is necessary to have sufficient data to train deep learning models without annotations. This task is counted as one of the critical steps seen as a limitation for the researchers in the deep learning domain [31]. Figure 1 demonstrates some samples of ancient manuscripts that have been damaged and eroded due to the passing of time. To handle the variance of such data, we propose to explore further the use of generative adversarial networks (GAN). Following recent work exploring generative

Fig. 1. Random samples of degraded manuscripts [29]

models in research settings with scarce data, [23] GAN shows promise in synthesizing new samples. This paper attempts to produce an artificial dataset using state of the art generative models for the ancient and degraded manuscripts to reconstruct the synthesized historical documents. The contribution of this paper is three-fold:

– We propose a joint augmentation, Deep convolutional GAN, which obtains a new state-of-the-art that increases the accuracy for synthesizing ancient manuscripts.
– We propose an appropriate unsupervised approach that leads to extracting the page, Ornament, and Characters of generated documents' to improve the segmentation performance.
– By a small amount of unlabeled data, we improved and validated the performance of generated data segmentation.

Section 2 review and study the related work. Section 3 pulls up the proposed approach in detail, including the generative framework and its advantages for synthesizing data in an unsupervised manner. The experimental result is presented in Sect. 4, and finally, Sect. 5 concludes the results obtained by the models and discuss their performance concerning our goal. To our knowledge, this work is the first use of Deep Convolutional Generative Adversarial networks (DCGANs) [26], which utilize an unsupervised segmentation model to synthesize the annotations of historical manuscripts.

2 Related Work

GAN is introduced by [15] to generate images that have been used in several problems, but we are briefly summarizing the relevance of our work. Document annotation introduced by [32] is a platform that helps users to annotate an image with labels. The process of annotating is time-consuming and requires some expertise. Later, the document image analysis community created a system that could augment data structure via XML. However, the system must know when asked about the dentition of document pages. Pix2Pix [18] and image to image transformation frameworks are also using source and target to function the mapping.

The idea of unsupervised segmentation was first introduced in W-Net [33] for semantic segmentation. In [12], an unsupervised medical image segmentation based on the local center of mass computation, was proposed. While the method if of interest, achieving those results using deep learning is a worthy challenge. As of late, the [22] generalized a fully convolutional network (FCN) to process the semantic segmentation task. The method was taken from pre-trained networks of a well-known structure called ImageNet [10] to do the downsampling and upsampling. In order to get a better prediction of segmentation, the last layer was combined. The U-Net [27] proposed another segmentation structure presented with a U-shaped architecture. The famous approach also had the same structure as AutoEncoder. There was also competition in document analysis [11,25,30]. However, most of such methods were required to have labelled data, including pre-training steps. In terms of unsupervised learning, Deep Belief Networks (DBNs) [17] worked to learn from training distribution to reach the common features. The model tries to maximize the latent likelihood of variables during the training process. The problem of such models measures the latent likelihood variables, which are limited. Additionally, [14] introduced a new approach that utilizes various rotations in the training data to distinguish the correct labels.

3 Overview of Proposed Method

Our proposed framework is based on unsupervised learning combined with a Deep Convolutional Generative Adversarial Networks (DCGAN) [26] model. Our approach follows one of the fundamental elements of feature extraction with different styles, including pages, ornaments, and characters, to efficiently have a high volume of data in an unsupervised manner. Albeit, it is shown in Fig. 2 the proposed approach reveals an adversarial process to identify the different features in reconstructing documents. The step of learning features is used in a later segmentation step to improve classification performance.

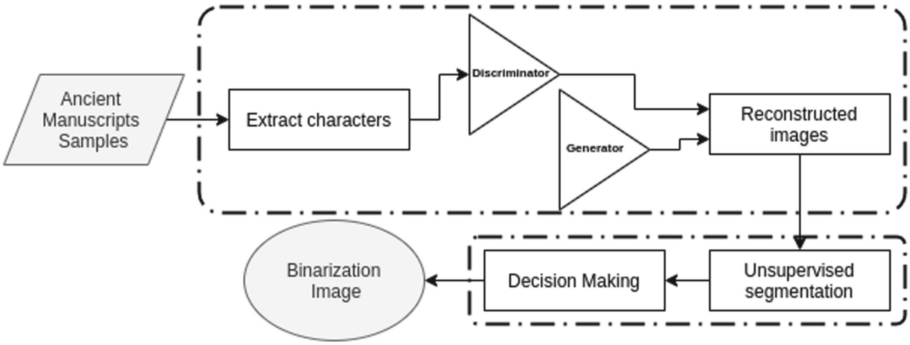

Fig. 2. Historical documents analysis diagram

3.1 Work Methodology

Our idea aims to develop an unsupervised segmentation model for reconstructing ancient manuscripts using DCGAN on different non labelled historical documents. Our effort is to achieve a rich performance of augmentation tasks. Additionally, we propose an unsupervised segmentation method using backpropagation to segment the document layouts. We offer an improved technique to assess the effects on documents' anomalies. The generator and discriminator play an adversarial game against each other to generate new samples from unlabeled data.

The Deep Learning Frameworks. The original architecture of GANs [15] has two components, including Generator G, which tries to fool the second note called Discriminator D. The generator aims to produce fake images using random noises vector z via original images' latent space. Moreover, the Discriminator by assessing the output $D(G(z))$ and generated images $G(z)$ from the latent space of $\mathbf{z} \sim P_{\text{noise}}(\mathbf{z})$ decides whether the output is real image x from $p_{data}(x)$ or is fake from $p(z)$ that represents the distribution of sample z. Follow the Eq. 1 in DCGAN, where the G utilize the transposed technique to apply up-sampling of image size, the convNet tries o find the correlated area of images. The objective of training consists of two processes. In the first step, the discriminator updates parameters by maximizing the expected log-likelihood, and in the second step, while the discriminator parameters are updated, the generator generates fake images.

$$\min_G \max_D V(G, D) = \mathbb{E}_{\mathbf{x} \sim P_{\text{data }}(\mathbf{x})}[\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim P_{\text{noise }}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \tag{1}$$

We consider the information from two figures, where Fig. 3 illustrates the reduction of cost function during the training of 5000 epochs and Fig. 4 depicts the results which shows that the model has improved the quality of generated images. Such an adversarial game continues between these two notes by targeting identifying the generated ones from the original samples. The learning part from the generator is also done during backpropagation steps from the statistical distribution of the training dataset through the discriminator. The $p_g(x)$ also refers to the distribution of vectors from a generator that tries to maximize its loss. Contrary, the discriminator is trying to max the rewards $V(D, G)$ through training operations [8]. In this work, we use a developed version of GANs called DC-GANs. in the next step, we use the output of GANs as the input of segmentation. Besides, the unsupervised segmentation by backpropagation is applied [20]. As shown in the Fig. 5, a reconstructing degraded ancient manuscript's proposed learning architecture is presented.

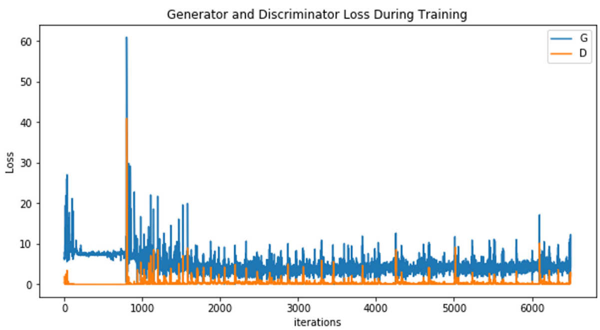

Fig. 3. Generator and Discriminator loss during 5k epochs

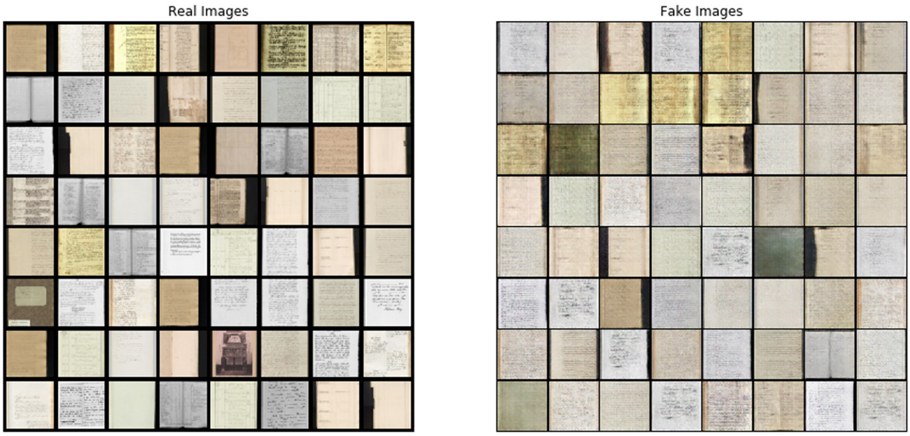

Fig. 4. Real vs fake images using DCGAN

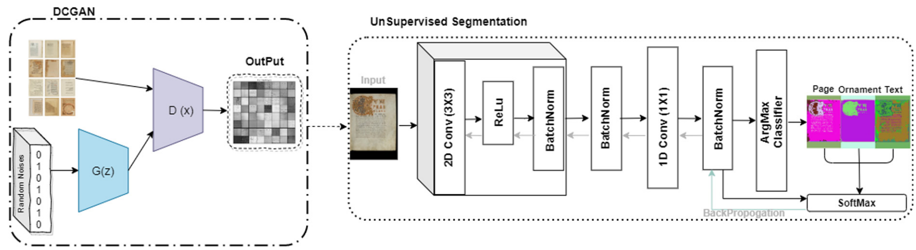

Fig. 5. Proposed Learning architecture

Restriction on Similar Features. The first challenge is to assign the pixels of common features to the same labels. We take the input images in colour, and then each pixel is normalized between [0,1]. The structure to get the feature map from the input image consists of a 2D Conv layers, ReLu activation, and batch normalization function. The method takes the map response by utilizing a linear classifier. $\{y_n = W_c x_n + b\}_{n+1}^{N} = 1$ where $W_c \in \mathbb{R}^{p \times q}$, $b_c \in \mathbb{R}^q$ and q represents the number of classes and p sets for the region's filters. Since the number of segments is unknown, we need to label the remaining clusters. The normalization procedure selects the maximum value in y'_n to achieve the most frequent cluster labels between the group of $\{1, ..., q\}$ clusters. Equation 2 demonstrates the *ith* cluster in y'_n.

$$C_i = y'_n \in \mathbb{R} | y'_{n,i} \geq y'_{n,j}, \forall_j \tag{2}$$

Restraint on Clustering Label. To reach decent quality, we need to extract superpixels C_{max} in each input image. Then we constrain other pixels to their superpixel as a cluster label. The simple linear iterative clustering proposed in [2] is used to get the superpixels. The algorithm benefits from the k-means clustering

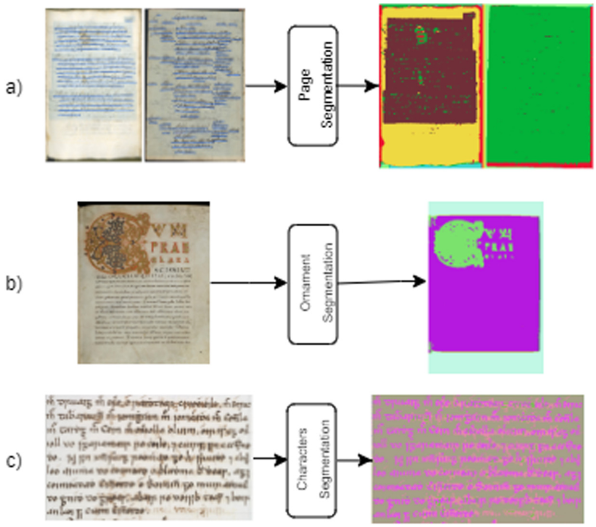

Fig. 6. Samples of generated data and segmented features, a) cBAD dataset b) DIBCO dataset, c) Irish dataset

approach to generate the superpixels. The motivation behind such a technique is that ancient manuscripts' datasets are usually highly imbalanced and contain no labels or masks for segmentation between foreground (page, ornaments, and layout) and background (stain and other natural noises). The aim of using a generating label is to design an automatic mask for documents. The objective is to qualify the capacity of good augmentation masks of the learned networks. In our scarce data setting, the Inception score may be of little use. In Fig. 6, we demonstrated some of the generated masks using our methods. It also reveals the network's dropout that sets out other values from random images of document masks to do the data transformation process. Any 3×3 pixel patch could then exhibit another historical document's characteristics and segment the natural noises of degraded documents.

3.2 Learning Networks

The case of document object detection in ancient manuscripts can be viewed as a binary classification task, where the author [19] would indicate the presence of anomaly within the entire document. Due to the two properties of DCGAN, the reconstructed images are fed into the unsupervised generating segments. While the training is done, the networks transfer the learned features to the unsupervised generation. Our full objective is described in Eq. 3:

$$L(G_x, D_{G(z)}, y'_n) = L_{DC-GAN}(D(x), D(G(z))) + L_{y'}(y'_n, c'_n) \tag{3}$$

This objective would minimize the G's loss and maximize the $logD(G(z))$. Training is split up into two main parts, where at first, the DC-GAN would update the network weights, and in the second part, the network parameters would predict the superpixels that are considered labels.

4 Experiments

This paper explores an unsupervised feature learning where the model input is the synthesized data from DCGAN. We adopt a standard testing protocol using generated data by DCGAN as our input for the unsupervised segmentation model. Given our approach to the restricted data, we can see a better improvement in various corrupted document images. Since the dataset has not provided the ground-truth, we use Otsu's algorithm to obtain the binary masks. Three different tasks are tested to extract the features, including Page, Ornament, and characters. Three datasets are used to illustrate the performance of the proposed method. Since the *Pytorch* framework is used in our models, we had to fix tensors' dimensions to train our generative model. The below Eq. 4) normalized the height and width of samples where the *mean* is a sequence of means for each channel, and *std* represents the standard deviations for each channel.

$$image = \frac{image - mean}{std} \tag{4}$$

Once the normalization is done, all the train samples are resized. Table 1 demonstrates the comparison of segmentation between original data and simulated data. The results show that True positive, which is the number of correctly segmented pixels as foreground, is higher than the false positive, falsely segmented.

Table 1. Performance for feature extraction in %

Method	TruePositive	FalsePositive	F1-Score
Page Segment	0.44	0.08	0.55
Orn Segment	0.35	0.16	0.47

4.1 Page Segmentation

Historical documents have different boundary regions, which can lead to displeasing results for document image processing approaches. In this experiment, we used the cBAD dataset [11] to apply our model as real samples for DCGAN to generate fake images; later we used the proposed unsupervised segmentation. To apply the page segmentation and predict the relevant page pixels, the binary masks are essential. 600 complete images are used to train for 200 epochs. Such an approach is suitable for uncovering the pages of historical documents where the ground-truths are not provided. Figure 7 demonstrate the result of our approach.

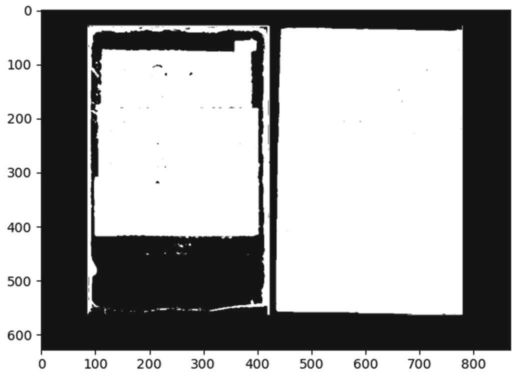

Fig. 7. Sample of generated data and segmented Page

4.2 Ornament Segmentation

Ornaments in historical documents play an essential role in discovering the symbols and signs of our ancestors. Therefore, to have a better understanding of such signs, large quantities of such images are considered to be an essential task. The data used to train our generator is gathered randomly from a different database, including 400 images. The model is trained with 5000 epochs to generate new ornaments. To apply the segmentation step, we trained with 10 epochs; later, we used Otsu to have a binary mask. Figure 8 demonstrates the result of our approach.

4.3 Character Segmentation

Characters are an essential part of manuscripts. They are the meaningful regions of documents. This experiment uses the Irish dataset to generate more images to overcome the lack of abnormal compilations within this domain. We used 149 images to train our DCGAN and generate fake samples. In the next chapter, to segment the characters and their regions, the proposed method proved great results by removing the degradation and separating the foreground from a noisy background. Figure 9 shows promising results with only 15 self-training epochs to achieve the segmented characters. To study the performance of binarization via learned representation and evaluate the proposed method for Character recognition, we use F-Measure (FM), pseudoFMeasure (Fps), PSNR and DRD standard metrics presented in [24]. Following the result in Table 2, we compare the quality of representation to other binarization techniques. Their architecture is based on different hyperparameters, which make the comparison difficult. However, our result is still significant.

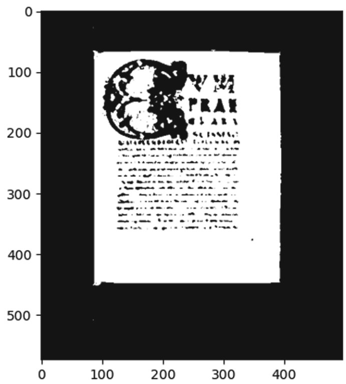

Fig. 8. Sample of generated data and segmented Ornament

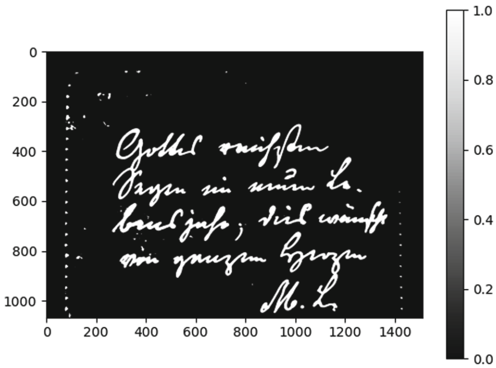

Fig. 9. Sample of generated data and segmented Character

Table 2. Comparison with other binarization technique

Method	FM	Fps	PSNR	DRD
Adak et al. [3]	73.45	75.94	14.62	26.24
Gattal et al. [13]	64.52	68.29	13.57	16.67
Saddami et al. [28]	46.35	51.39	11.79	24.56
Proposed method	**78.16**	**80.25**	**16.31**	**11.35**

5 Conclusion

We presented, in this paper, a novel deep generative segmentation approach to obtain the regular pattern in historical documents. We developed an augmentation method in order to generate synthetic data. The approach has combined two methods, including augmentation, to generate more data and unsupervised segmentation to detect patterns such as page, ornaments, and characters. The experimental results on three different benchmark datasets show that the model's performance is competitive regarding the state-of-the-art. Furthermore, the technique is able to overcome the lack of dataset with generating images and segment the layouts of generated images on three different ancient manuscript datasets. Future work will be devoted to a more detailed study of automatically extracting bounding boxes for our training set in a large quantity of different synthesized document types.

Acknowledgement. The authors thank the NSERC Discovery held by Prof. Cheriet for their financial support. We thank Ms. MG Jones, for assistance and comments that greatly improved the manuscript.

References

1. Abuelwafa, S., Pedersoli, M., Cheriet, M.: Unsupervised exemplar-based learning for improved document image classification. IEEE Access **7**, 133738–133748 (2019)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intelli. **34**(11), 2274–2282 (2012)
3. Adak, C., Chaudhuri, B.B., Blumenstein, M.: A study on idiosyncratic handwriting with impact on writer identification. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 193–198. IEEE (2018)
4. Afzal, M.Z., Kölsch, A., Ahmed, S., Liwicki, M.: Cutting the error by half: investigation of very deep CNN and advanced training strategies for document image classification. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 883–888. IEEE (2017)
5. Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D.: Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3722–3731 (2017)
6. Bukhari, S.S., Dengel, A.: Visual appearance based document classification methods: Performance evaluation and benchmarking. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 981–985. IEEE (2015)
7. Chen, S., He, Y., Sun, J., Naoi, S.: Structured document classification by matching local salient features. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pp. 653–656. IEEE (2012)
8. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2172–2180 (2016)
9. Das, A., Roy, S., Bhattacharya, U., Parui, S.K.: Document image classification with intra-domain transfer learning and stacked generalization of deep convolutional neural networks. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 3180–3185. IEEE (2018)
10. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)

11. Diem, M., Kleber, F., Fiel, S., Grüning, T., Gatos, B.: cbad: ICDAR 2017 competition on baseline detection. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 1355–1360. IEEE (2017)
12. Eaton-Rosen, Z., Bragman, F., Ourselin, S., Cardoso, M.J.: Improving data augmentation for medical image segmentation (2018)
13. Gattal, A., Abbas, F., Laouar, M.R.: Automatic parameter tuning of k-means algorithm for document binarization. In: Proceedings of the 7th International Conference on Software Engineering and New Technologies, pp. 1–4 (2018)
14. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728 (2018)
15. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
16. Harley, A.W., Ufkes, A., Derpanis, K.G.: Evaluation of deep convolutional nets for document image classification and retrieval. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 991–995. IEEE (2015)
17. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
18. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
19. Ji, B., Chen, T.: Generative adversarial network for handwritten text. arXiv preprint arXiv:1907.11845 (2019)
20. Kanezaki, A.: Unsupervised image segmentation by backpropagation. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1543–1547. IEEE (2018)
21. Kumar, J., Doermann, D.: Unsupervised classification of structurally similar document images. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 1225–1229. IEEE (2013)
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
23. Maroñas, J., Paredes, R., Ramos, D.: Generative models for deep learning with very scarce data. In: Vera-Rodriguez, R., Fierrez, J., Morales, A. (eds.) CIARP 2018. LNCS, vol. 11401, pp. 20–28. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-13469-3_3
24. Ntirogiannis, K., Gatos, B., Pratikakis, I.: ICFHR 2014 competition on handwritten document image binarization (h-dibco 2014). In: 2014 14th International Conference on Frontiers in Handwriting Recognition, pp. 809–813. IEEE (2014)
25. Pratikakis, I., Zagoris, K., Barlas, G., Gatos, B.: ICDAR 2017 competition on document image binarization (dibco 2017). In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 1395–1403. IEEE (2017)
26. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
27. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

28. Saddami, K., Afrah, P., Mutiawani, V., Arnia, F.: A new adaptive thresholding technique for binarizing ancient document. In: 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 57–61. IEEE (2018)
29. Schomaker, L.: Lifelong learning for text retrieval and recognition in historical handwritten document collections. arXiv preprint arXiv:1912.05156 (2019)
30. Simistira, F., et al.: ICDAR 2017 competition on layout analysis for challenging medieval manuscripts. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 1361–1370. IEEE (2017)
31. Tensmeyer, C.A.: Deep learning for document image analysis (2019)
32. Wei, H., Chen, K., Seuret, M., Würsch, M., Liwicki, M., Ingold, R.: Divadiawi-a web-based interface for semi-automatic labeling of historical document images. Digital Humanities (2015)
33. Xia, X., Kulis, B.: W-net: A deep model for fully unsupervised image segmentation. arXiv preprint arXiv:1711.08506 (2017)