

IMAGE QUALITY ENHANCEMENT OF SCANNED PHOTOS: COMPARISON OF DEEP LEARNING TECHNIQUES

Victor Rocha Santos^b, Tiago Pagano^b, Rafael Borges^e, Lucas Kirsten^e, Lucas Ortega^c, Máira Matos^c, José Vinícius Paranhos^d, Ingrid Winkler^a, Erick Giovani Sperandio Nascimento^a.

^a Dep. de Modelagem Computacional, Centro Universitário SENAI CIMATEC, Brasil

^b PPG em Modelagem Computacional e Tecnologia Industrial, Centro Universitário SENAI CIMATEC, Brasil

^c Grad. de Engenharia da Computação, Centro Universitário SENAI CIMATEC, Brasil

^d Téc. Desenvolvimento de Sistemas, Centro Universitário SENAI CIMATEC, Brasil

^e HP Inc., Brazil R&D

Abstract: Currently, millions of photos are captured daily, and several factors can influence the quality of an image, causing distortions. Research has shown that there are several ways to remove defects from images. This study aims to comparatively analyze the potential of Deep Learning techniques to improve scanned images with shadow, glare, crumpled paper, external lightning, change of perspective and wave distortion defects. Based on a review of the literature on recent deep learning architectures, we have selected three, which were trained and refined to improve the quality of the images. The results indicate that the nets were able to attenuate and remove some defects. On this basis, these initial experiments demonstrate that deep learning models are promising for the studied defects.

Keywords: Deep Learning; Image enhancement; Image denoising

MELHORIA DA QUALIDADE DE IMAGEM DE FOTOS DIGITALIZADAS: COMPARAÇÃO DAS TÉCNICAS DE APRENDIZAGEM PROFUNDA

Resumo: O objetivo deste trabalho é analisar comparativamente o potencial de técnicas de aprendizagem profunda para melhorar imagens digitalizadas com defeitos de sombra, reflexo de luz, papel amassado, luminosidade, mudança de perspectiva e ondulação. Com base em uma revisão da literatura sobre arquiteturas recentes de aprendizagem profunda, selecionamos três técnicas para remoção destes defeitos em imagens. Os resultados obtidos apontaram que as redes conseguiram atenuar alguns defeitos com intensidades variadas e, em alguns casos, removê-los. Conclui-se que estes experimentos iniciais demonstram que modelos de aprendizagem profunda são bastante promissores para a resolução de alguns dos defeitos estudados e que avanços significativos foram alcançados na melhoria da qualidade das imagens.

Palavras-chave: Aprendizagem profunda; Melhoria da qualidade de imagens; Remoção de ruídos em imagens

1. INTRODUCTION

Currently, millions of photos are captured and shared daily around the world, and several factors can influence the quality of an image. Distortions can occur at the time of capture, storage, transmission and display. In addition, other drawbacks may occur such as the camera failure, inadequate lighting, photographer's inexperience and several others, which can result in low quality images. When trying to scanning printed images and photos, the chances of capture defects are increased, especially when a digital camera is used instead of a more traditional scanning device.

To improve image quality, in addition to conventional computer vision methods, several machine learning (ML) techniques, more specifically deep learning, advanced in super resolution and image enhancement [1]. Li [2] developed a framework based on convolutional neural networks (CNN), training a synthetic dataset that automatically corrects different types of geometric distortions in a single image. This approach, however, deals only with defects in geometric perspective, that focused on the process is the minor part of the cases.

Vu [3] has proposed a way to improve the quality of images with a lower computational cost, that can be trained for super resolution and image enhancement problems. Chiang [4] has developed a method to improve images of handwritten documents, that not only removes interference and watermarks but also significantly increases the readability of documents. Pursuing similar goals, Bannigidad [5] and Rani [6] devised ways to restore images of degraded documents.

However, although these studies have achieved advances, solving many of common defects in image, it is still an open research problem. Thus, this study aims to comparatively analyze the potential of deep learning (DL) techniques to improve scanned images with shadow, glare, crumpled paper, external lightning, change of perspective and wave distortion defects.

This paper is organized as follows. Section II describes the experimental framework used. Sections III introduce terminology and relevant concepts used throughout this paper. Section IV describes the results observed. Finally, Section V presents our conclusions and suggestions for further research needed.

2. METHODOLOGY

The research methods were divided into four steps.

First, we created algorithms to simulate the six defects studied: shadow, glare, crumpled paper, external lightning, change of perspective and wave distortion defects. The algorithms were implemented in the Python programming language using two libraries for image processing: OpenCV and Pillow. During execution, each input image receives zero to two defects and generates three defective images.

Second step was to choose a dataset to add the defects. COCODataset [7] was selected due to its large number of images and for being public. An initial dataset of 10,500 images, served as input to the defect generation script, resulting in a total of 31,500 defective images.

After the generation of the dataset with artificial defects, we implemented models of neural networks to improve the images. For this, training, testing and

validation were planned, as well as the meta-parameters of the networks. The operation of the DL techniques implemented are shown below.

In the BRDNET training, the dataset was divided into 70: 15: 15 (training, validation and testing). The settings chosen for loss and metric were, respectively, MSE and MAE, while the optimizer used was Adam. The training lasted 40 epochs and used 20 images per batch (batch size).

For UNET training, sigmoid was used as an activation function in the output layer to capture the intensities of each pixel and LeakyReLU in the others. All layers have same padding to guarantee the extraction of features from the edges of the images and, after 1st and 3rd layer, BatchNormalization was used to improve the model's performance. To represent the bottleneck in the network, forcing the condensed learning of the defective features, MaxPooling and UpSampling of dimensions (2x2) were applied in the 2nd and 4th convolutional layers, respectively. The dataset was subdivided into 70: 15: 15 (training, testing and validation). The settings chosen for loss and metric were, respectively, MSE and MAE, while the optimizer used was Adam. The training lasted 50 epochs and used 130 images per batch (batch size).

VAE technique consists of a model with an encoder with the following characteristics: two two-dimensional convolutional layers with kernel 3x3 and with 20 and 10 filters, respectively, followed by a 2x2 size Maxpooling and kernel 3x3, with a last 2x2 size Maxpooling. All convolution layers used LeakyRelu as activation function and a BatchNormalization. The layers corresponding to the average and variance, of dimension 1000, pass through the VAE standard sampling function. The decoder has a 1000-dimension sample as input, followed by a resizing to 25x25x1. Subsequently having 3 sets with 1 skip connection adding tensors of the same dimension, convolution with 10 filters and UpSampling of 2x2 dimension. The settings used for the training were the standard loss of the VAE architecture, with the mean square error and the optimizer the RMSProp as the metric. For training, the dataset was divided into 70: 15: 15 (training, validation and testing). The training lasted 100 epochs and used 20 images per batch (batch size).

In the fourth step, after the network's results, 200 images of prediction of each technique were evaluated with two evaluation metrics that are commonly used. In WADIQAM metric, the value ranges from 0 to 100, and lower values imply better image quality, meanwhile, in SSIM metric the value ranges from 0 to 1 and the closer to one, it implies better image quality.

3. MACHINE LEARNING TECHNIQUES

In this section, we describe the DL techniques that were implemented and tested throughout our work.

3.1. BRDNet (Batch-Renormalization Denoising Network)

Convolutional Neural Networks (CNNs) were introduced and developed in the 1990s with works related to the recognition of handwritten digits [8]. Convolution is the multiplication of matrices, producing feature maps used as input for other filter's layers [9]. In CNNs, the filters can be compared to the neurons of the Multilayer Perceptron (MLPs) updating their values to learn more about the data.

CNN's development has brought a significant improvement to the field of ML applied to 2D images, such as the use of local receptive fields that, unlike MLPs, use kernels to learn features without having to create dense layer pixels, improving generalization and reducing the number of parameters.

Proposed by Tian [10], Residual Learning (RL) consists of adding skip connections of previous layers to other layers in the convolutional network, causing less performance degradation and better loss results in deep networks.

According to He [11], the Batch Renormalization (BRN) consists of applying Batch Normalization with samples of all batches for data normalization, as opposed to the common batch normalization by preventing network training problems such as vanishing gradient, in addition to enhancing network convergence.

3.2. UNET

AE are neural networks that learn to generate compressed representations of a given entry and, from there, make an approximate reconstruction of that entry [12]. In other words, the architecture in question has internal layers that are narrower than those of input and output, forcing it to learn the correlations and to form an internal representation of these patterns, through which the network is able to map input and output. Every autoencoder can be subdivided into an encoder, which is the section of the network that compresses the input and a decoder, which decodes the reduced representation.

Denoising Autoencoders are a specific type of autoencoder that learns to reconstruct a given entry from its defective version. The Convolutional Denoising Autoencoders variant applies the mathematical convolution operation to extract features from the input, most used in image processing [13].

3.3. Variational Autoencoders (VAE)

Variational Bayes Autoencoder, simply called VAE, is an efficient alternative to learn something with a probabilistic model [14]. They are structures like the AEs, having in their structure an encoder capable of translating information to a smaller sample space than it represents, and decoder can reconstruct that information to the original data. However, the latent space distribution of an AE, even if not linear, makes it impossible to generate new information from any point in the space.

To overcome these limitations and create a model capable of learning and generating new data, Tjandra [14] inserted a spatial sample capable of generating new information from a dataset X with N samples where the network's output is generated from a continuous random variable z . Considering that for each sample x there is a corresponding variable z and the value to $z(i)$ is generated from a conditional distribution [14, 15].

Still on VAE, it was not explored for tasks of image denoising and image enhancement, however its application can generate a latent space with a good approximation, based not on X directly, but on random signals X -based.

4. RESULTS AND DISCUSSION

An analysis of the results was carried out, involving machine learning techniques for solving defects in images.

4.1. BRDNet

Figure 1 shows the results of the BRDNet network. Results generated by the BRDNet technique, it is possible to observe that defects with low intensity can be corrected, although the images remain with minimal distortions, since the defect filling in the generated images is done based on the pixels that surround it.

Figure 1. Defective image / predicted image / original image



Source: Own author

4.2. UNET

Figure 2 shows the results of the UNET network. The model performed an attenuation of defects in general. For ripple defects and change in perspective, the model filled in the offset of the pixels of the edges with colors found in the vicinity of the image. For glare and shadow defects, the model was able to attenuate them to varying intensities of the defect. For the crumpled paper defect, no significant improvement was noticed.

Figure 2. Defective image / predicted image / original image



Source: Own author

4.3. VAE

Figure 3 shows VAE's results. The proposed model partially met its objective, in some cases it smoothes the defects and in the worst cases it fails to perform the removal or smoothing of the defects. The model also has difficulties with reconstructing the image structure and coloring.

Figure 3. Defective image / predicted image / original image



Source: Own author

Values in the table 1 refer to the averages of the evaluations and differently from what was observed in the qualitative analysis of the images, the metrics indicate that there was no improvement in the quality of the images, with BRDNet being the architecture that obtained the best results.

Table 1. Comparison between the mean evaluation

	BRDNet		UNET		VAE	
Mean	SSIM	WADIQAM	SSIM	WADIQAM	SSIM	WADIQAM
Predict	0.770	13.987	0.736	70.932	0.760	13.698
Defective	0.773	07.567	0.773	07.523	0.778	06.476

5. CONCLUSION

This study aimed to comparatively analyze the potential of deep learning techniques to improve scanned images with shadow, glare, crumpled paper, external lightning, change of perspective and wave distortion defects. These initial experiments demonstrate that deep learning models are very promising for the resolution of some of the studied defects and that significant advances have been achieved in improving the quality of the images.

Although the nets have learned to attenuate with varying intensities and, in some cases, remove the glare defects, shadow and crumpled paper, there has been a substantial loss of resolution within the reconstructed region. Thus, it might be interesting to continue training and explore hyper-parameters with the techniques in question to generate better results.

Further research is needed to investigate other deep learning techniques, such as the generative adversarial networks (GANs). In addition, we consider improving the overall training methodology, especially regarding the used dataset. We suggest bringing improvements to the generation of synthetic defects in images, as well as collecting real images by scanning printed versions of images from a new dataset.

Acknowledgments

This work was supported by Hewlett Packard Company/HP Brazil, and Brazilian Ministry of Science and Technology, through the Informatics Law.

6. REFERENCES

- ¹ O'MAHONY, N. et al. Deep learning vs. traditional computer vision. In: **SPRINGER. Science and Information Conference**. [S.l.], 2019. p. 128–144.
- ² LI, X. et al. Blind geometric distortion correction on images through deep learning. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2019. p. 4855–4864.
- ³ VU, T. et al. Fast and efficient image quality enhancement via desubpixel convolutional neural networks. In: **Proceedings of the European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2018. p. 0–0.

- ⁴ CHIANG, J.-S. et al. Adaptive image enhancement method for document. In: **IEEE.2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)**. [S.I.], 2017. p. 417–420.
- ⁵ BANNIGIDAD, P.; GUDADA, C. Restoration of degraded kannada handwritten paper inscriptions (hastapratī) using image enhancement techniques. In: **IEEE.2017 International Conference on Computer Communication and Informatics (ICCCI)**. [S.I.], 2017. p. 1–6.
- ⁶ RANI, N. S.; JAIN, A. S.; KIRAN, H. A unified preprocessing technique for enhancement of degraded document images. In: SPRINGER. **International Conference on ISMAC in Computational Vision and Bio-Engineering**. [S.I.], 2018. p. 221–233.
- ⁷ LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: SPRINGER. **European conference on computer vision**. [S.I.], 2014. p. 740–755.
- ⁸ LECUN, Y. et al. Handwritten digit recognition with a back-propagation network. In: **Advances in neural information processing systems**. [S.I.: s.n.], 1990. p. 396–404.
- ⁹ LECUN, Y. et al. Object recognition with gradient-based learning. In: **Shape, contour and grouping in computer vision**. [S.I.]: Springer, 1999. p. 319–345.
- ¹⁰ TIAN, C.; XU, Y.; ZUO, W. Image denoising using deep cnn with batch renormalization. **Neural Networks**, Elsevier, v. 121, p. 461–473, 2020.
- ¹¹ HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.I.: s.n.], 2016. p. 770–778.
- ¹² WANG, J.; HE, H.; PROKHOROV, D. V. A folded neural network autoencoder for dimensionality reduction. **Procedia Computer Science**, Elsevier, v. 13, p. 120–127, 2012.
- ¹³ GONDARA, L. Medical image denoising using convolutional denoising autoencoders. In: **IEEE.2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)**. [S.I.], 2016. p. 241–246.
- ¹⁴ TJANDRA, A. et al. Stochastic gradient variational bayes for deep learning-based asr. In: **IEEE.2015 IEEE Workshop on Automatic Speech Recognition and Understanding(ASRU)**. [S.I.], 2015. p. 175–180.
- ¹⁵ XU, Q. et al. The difference learning of hidden layer between autoencoder and variational autoencoder. In: **IEEE.2017 29th Chinese Control And Decision Conference(CCDC)**. [S.I.], 2017. p. 4801–4804.