2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Skip-Connected Deep Convolutional Autoencoder for Restoration of Document Images

Guoping Zhao, Jiajun Liu*, Jiacheng Jiang, Hua Guan and Ji-Rong Wen
Beijing Key Laboratory of Big Data Management and Analysis Methods
School of Information, Renmin University of China
Beijing 100872, China
Email: {guopingzhao, jiajunliu, jiachengjiang, guanhua, jrwen}@ruc.edu.cn

***Abstract*—The denoising and deblurring of images are the two essential restoration tasks in the document image processing task. As the preprocessing stages of the processing pipeline, the quality of denoising and deblurring heavily influences the result of subsequent tasks, such as character detection and recognition. In this paper, we propose a novel neural method for restoring document images. We named our network Skip-Connected Deep Convolutional Autoencoder (SCDCA), which is composed of multiple layers of convolution followed by a batch normalization layer and the leaky rectified linear unit (Leaky ReLU) activation function. Inspired by the idea of residual learning, we use two types of skip connections in the network. One is identity mapping between convolution layers and the other is used to connect the input and output. Through these connections, the network learns the residual between the noisy and clean images instead of learning an ordinary transformation function. We empirically evaluate our algorithm on an open and challenging document images dataset. We also assess our restoring results using the optical character recognition (OCR) test. Experimental results have demonstrated the effectiveness and efficiency of our proposed algorithm by comparing with several state-of-the-art methods.**

I. Introduction

Although most of today's documents are in electronic form, many older documents do not have the original electronic version. These old documents are usually saved as scanned images or photographs. The text on the document images cannot be directly identified, nor can they be retrieved by modern search and analytical tools. OCR is the process of transforming printed or handwritten documents into a structured and digitized format. However, due to the noise, blur, ink blot, fading, paper aging and other reasons, the document image is often not clear enough and the recognition quality of the image is poor. This is precisely why many important documents are not digitalized and are kept offline. In many cases, preprocessing of document images is an essential process before OCR, and sometime can decide if a document image meets the requirements for OCR. The main purpose of the preprocessing stage is to enhance the image quality for further processing steps by reducing or correcting the noise and blur in the document images. Typical examples of degraded document image are shown in Fig. 1.

* Corresponding Author

Image restoration techniques have been a subject of growing interest in the past decades. It attempts to produce a new image that has less noise, less blur and is closer to the original noise-free image. As in previous work, the corrupted low-quality image can be formulate as: $y = D(x) + n$, where y presents the degraded image, x presents the original high-quality image, D is the degradation function and n is noise. The image restoration is also known as an inverse problem, because its goal is to estimate an unknown original image from a (possibly noisy, blurry or low-resolution) observation. That is to say, the task of restoration is estimating the x from the observation y.

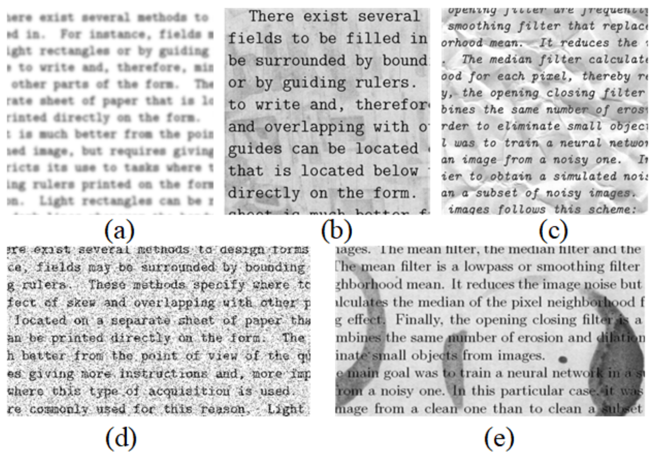

Fig. 1. Typical degradation of document images. (a)blurred. (b)poor paper. (c)wrinkles. (d)noising image (e)ink marks.

Conventional restoration algorithms mainly focus on natural scene images, and as the demand for OCR increases drastically, there have been studies to solve document restoration in recent years [1]–[6]. The algorithm in [2] is based on the document image foreground segmentation, and it calculates a desired intensity probability density function as the prior for deblurring. [3] analyzes the different properties of document image from natural-scene images, and takes these domain-specific properties into the optimization process. [4] uses the L_0-regularized intensity and gradient prior. And the methods in [6], [7] are based on deep neural network. [8] is based on the two-tone prior.

In this paper, we propose a deep convolutional auto-encoder with skip-connections, which is named SCDCA, to restore the document image from a low-quality observation. The SCDCA consists of two components: one is the encoder, and the other is the decoder. Many previous studies have proved that a deep structure is beneficial to obtain good results in computer vision problems, because the deeper network has more parameters and it naturally has powerful feature representation learning capability. In addition, increasing the depth of the network results in larger network receptive field, enabling the kernels of convolutional layers to receive more contextual information. But on the other hand, as the depth of the network increases, the network is often exposed to the vanishing gradient problem. So, our approach adopts the idea of residual learning proposed in [9], which resolves the vanishing gradient problem in deep networks by adding identity mappings as the skip connections between convolution layers. Meanwhile, we add a special skip-connection from the input to the output to pass information more directly between layers. Our network no longer learns a reconstruction function directly from the degraded image to the original image, but instead it learns the residuals between them. Experiments show that the residual learning technique accelerates the training process and improves the restoration performance. In addition, unlike some discriminative models that handle a certain noise level or a fix blur kernel, the SCDCA has the capability to handle more complex situation.

In summary, our main contributions are in the following aspects: first, we proposed a novel network architecture, which we name "the skip-connected deep convolutional autoencoder", that learns the residual between input and output instead of learning a transformation function from input to output. Second, we formulate the residual learning and explain its advantage. Finally, through empirical experiments, we proved that the SCDCA not only has excellent performance in image restoration, but also boosts the accuracy of the subsequent OCR.

II. Related Work

Analytical techniques for solving the inverse problems have been studied for a long time. There exists a huge amount of literature addressing the topic of image restoration problems, state-of-the-art reviews can be seen in [10] and [11]. In general, the image restoration includes many aspects, such as denoising [12]–[16], deblurring [2]–[4], [6], [17], demosaicking [16], [18],super-resolution [19], [20], etc. The two most relevant aspects of this paper are denoising and deblurring.

The dictionary learning methods in [21] use a sparse and redundant representations over trained dictionaries to remove zero-mean white and homogeneous Gaussian additive noise in the given image. The Block-matching and 3D filtering (BM3D) reported in [12] achieves very good performance for denoising images corrupted by Additive White Gaussian noise (AWGN). The BM3D is based on the fact that an image has a locally sparse representation in transform domain. It first grouping the similar 2D image fragments into 3D data arrays, and then carrying the collaborative filting on these 3D groups, finally using a inverse 3D transform to get the denoising image.

It is worthwhile taking note that the image restoration approaches have been more often learning-based methods in recent years. In particular, deep neural network based approaches have been dominant in this area over time. Stacked Denoising Autoencoders (SDA) is reported in [22], just like original strategy for building deep networks, it stacking layers of denoising autoencoders which are trained locally to denoise corrupted versions of their input images. The [13] use a plain multilayer perceptron (MLP) on image patches, and get better results than BM3D. The methods in [15] by propose shrinkage fields, a random field-based architecture that combines the image model and the optimization algorithm in a single unit, and achieving the goals both computational efficiency and high restoration quality. The [23] proposed a new deep convolutional network structure to capture the characteristics, and trained two submodules in supervised manner for deconvolution and artifact removal. The generative adversarial network (GAN) is used to restore the clear high-resolution image from the blurry low-resolution input in [7]. In order to handle face and text image at same time, it create a special GAN that contains a generator, but has two discriminators. The [24] combined the BM3D with CNN.

The most similar approaches to our work are proposed in [25] and [6]. The [25] use a deep convolutional encoder-decoder network with symmetric skip connections to for image denoising and super-resolution. Our work is also a encoder-decoder architecture, but the difference with [25] is that we not use deconvolution layer in the network, and the setting mode of skip-connections is also not the same. The methods in [6] is based on a 15 layers convolutional neural network. Compared to [6], we add the batch normalization layers [26] and skip-connections [9] in our network, which accelerate the model convergence of the model and boost its performance.

III. Proposed Method

We propose skip-connected deep convolutional autoencoder (SCDCA) for restoration of document images. The proposed method, as show in Fig. 2, aims to learn an end-to-end noise or blur removal function F from low-quality image to his restoration version. The SCDCA contains two stages: encoding stage and decoding stage. The encoder brings the data from the input to a bottleneck layer. Then, the decoder takes this encoded input and converts it back to an image. The latent space is the space in which the data lies in the bottleneck layer.

A. Residual Learning Formulation

In generator, we can formulate the the observed image by: $y = D(x) + n$, where y presents the degraded observation, x presents the original high-quality image, D is the degradation function and n is noise. In most case in the field of document image restoration, because the image damage is not only from scanner and camera, but also from external contaminationso

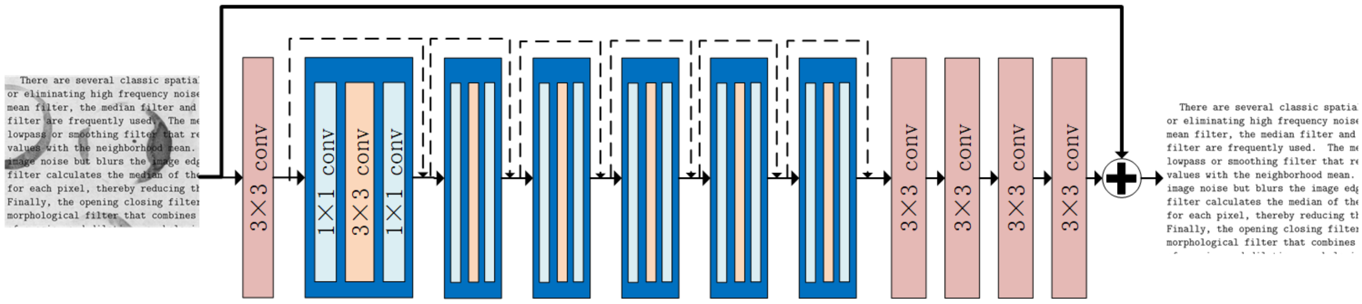

Fig. 2. The architecture of the proposed restoration network. Note that the two types of skip connection. One is used between residual blocks, and the other is used from input to output. The effect of first type skip connection is restoring vanishing degraded. The second type skip connection force the network to learn the residual between low-quality and high-quality document image.

the D is often not a single blur kernel or has a certain pattern. The autoencoder aims to learn the latent representation of the input, and can reconstruction the original data from noisy effects. We draw on the idea from Denoising Autoencoder [22], mapping the added noise image into a latent space, where is more robust to noise, and then reconstruct it with a constraint that the output is closer to the clean image. What we are different from [22] is that we not directly mapping the low-quality document image to the latent space, but mapping the residual parts, such as noiser or blur, to the latent space. And then through the decoder, we reconstruct this residual part. Finally, we subtracts the residual from input image, and obtaining the restoration.

We formulate our approach:

$$z = EN_{\theta_e}(y) \tag{1}$$

$$x = y - DE_{\theta_d}(z) \tag{2}$$

y is the observed document image, x is the restoration of y, z is the hidden representation of noise or blur. The $EN()$ express the encoder, $DE()$ express the decoder, and the θ_e, θ_d is the parameters of the encoder and decoder. Combined the Eq.(1) and Eq.(2), the SCDCA maps the input to output via:

$$x = y - DE_{\theta_d}(EN_{\theta_e}(y)) \tag{3}$$

The parameters θ_e, θ_d is learned by minimizing the L_2 loss between the restoration and ground truth.

We use the symbol $\hat{x}$ to indicate the ground truth of input image. The final loss function of our SCDCA is expressed as:

$$\begin{aligned} \mathbb{L} &= \frac{1}{2}\left\|x - \hat{x}\right\|_2^2 \\ &= \frac{1}{2}\left\|y - \hat{x} - DE_{\theta_d}(EN_{\theta_e}(y)\right\|_2^2 \end{aligned} \tag{4}$$

B. Encoder Network

The role of the encoder network is to extract features from the input and map it to a latent space.

In general, the autoencoder plays the role of dimension reduction, making the size of feature map in bottleneck layer smaller. But, in our restoration task, in order to avoid losing the original image contextual information, we did not reduce the feature size.

TABLE I
ARCHITECTURE OF THE ENCODING NETWORK.

Tpye	Kernel	Stride	Padding	Outputs	Activation
conv	3×3	1×1	1	32	ReLU
conv	1×1	1×1	0	64	Leaky ReLU
conv	3×3	1×1	1	64	Leaky ReLU
conv	1×1	1×1	0	32	ReLU
conv	1×1	1×1	0	64	Leaky ReLU
conv	3×3	1×1	1	64	Leaky ReLU
conv	1×1	1×1	0	32	ReLU
conv	1×1	1×1	0	128	Leaky ReLU
conv	3×3	1×1	1	128	Leaky ReLU
conv	1×1	1×1	0	32	ReLU
conv	1×1	1×1	0	128	Leaky ReLU
conv	3×3	1×1	1	128	Leaky ReLU
conv	1×1	1×1	0	32	ReLU
conv	1×1	1×1	0	256	Leaky ReLU
conv	3×3	1×1	1	256	Leaky ReLU
conv	1×1	1×1	0	32	ReLU
conv	1×1	1×1	0	256	Leaky ReLU
conv	3×3	1×1	1	256	Leaky ReLU
conv	1×1	1×1	0	32	ReLU

As shown in TABLE I, the encoder is consist of one 3×3 convolution layer followed six residual blocks. After every convolution layer, there is a batch normalization layer and a Leaky ReLU [27] activation. The basic unit of the encoder is called residual block, which is fisrt proposed in [9] using to resolving the vanishing degraded problem in image classification. Since then, many studies have proved that this kind of structure also had good effect on many low-level vision problems [28]–[31], and many variants of residual block were invented. We modified the normal structure of residual block in order to make it suitable for document image restoration.

The most obvious change is that we keep the output size of all convolution layers consistent with the original image by padding and setting the stride of convolution kernels to 1. The second, we take the Leaky ReLU activation function at the first two convolution layers in a residual block is , and use ReLU in the third convolution layer. In addition, the number of convolution kernels is doubled every two residual blocks until it reaches 256. To satify the additivity of skip connection between residual blocks, the output depth of last

1×1 convolution layer in each residual block is 32. It is depends on the number of convolution kernel in the first $3{\times}3$ convolution layer. The overview of the encoder can be seen in TABLE I.

C. Decoder Network

In original autoencoder [22], the role of the decoder network is to restore the encoded input back to the original image. But in our residual learning model, the output of decoder is the noise image or blur image extracted from input. Therefore, we get the result we expected through subtracting the model's output from the input image.

TABLE II
ARCHITECTURE OF THE DECODING NETWORK.

Tpye	Kernel	Stride	Padding	Outputs	Activation
conv	3×3	1×1	1	32	Leaky ReLU
conv	3×3	1×1	1	32	Leaky ReLU
conv	3×3	1×1	1	16	Leaky ReLU
conv	3×3	1×1	1	3/1	Leaky ReLU

Compared to the encoder, the architecture of decoder is very simple. As shown in TABLE II, the decoder consists of four convolution layers, and the kernel size of these convolution layers is 3×3. The batch normalization layers are stacked after every convolution, and all activation functions are Leaky ReLU. The only difference is the number of convolution kernels, the first two convolution layers have 32 kernels, the third layer have 16 kernels, and the last layer have 3 kernels. We set number of kernels in the last convolution layer to 3 is to obtain a three-channel output, which is the same as the original input shape. When dealing with grayscale image, setting the number of kernels in last convolution layer to 1.

TABLE III
POST-PROCESSING TRANSFORMATION

Operation	Tpye	Value
Activation	ReLU	
Arithmetic	Add	-1
Arithmetic	Multiply	-1
Activation	ReLU	
Arithmetic	Multiply	-1
Arithmetic	Add	1

As we all know, the value range of each pixel in the image is [0, 255]. For convenience, in image processing, it is usually divided by 255 to normalize the range to [0, 1]. Therefore, we need to ensure that every pixel of final output is between 0 and 1. Direct truncation may affect the transfer of the gradients, so before calculating the loss between input and output, we add a post-processing transformation. The post-processing transversion utilize the property of ReLU function ($max(0, x)$), maintaining the gradients transitivity while transforing the scale. We formulate the post-processing transformation as:$1 - [1 - [x]_+]_+$, and $[\cdot]_+$ is the symbol of $max(0, x)$. The detail can be seen in TABLE III.

IV. EXPERIMENT

A. Experiment Setting

1) DataSet: We used the DataSet in [32], whose training set contains 144 document images, coming from 18 different fonts. There are 8 images for each font on average. In addition, there are eight different types of sludge, two image sizes. The images in training have corresponding clean images. There are also 72 document images in [32], but have no corresponding clean images provided. Therefore, the test set cannot be used to quantitatively assess the effectiveness of the model, only can prove the effectiveness from visual effects. So in the quantitative evaluation phase, we need to redivide the original training set to a new traiing set and a test set. We have designed two experiments to measure the restoration power of our model. One is that we evaluate the model's ability of removing Additive White Gaussian Noise(AWGN), in this case, we use the clean label image with add varying degrees of gauss noise as input, and origin clean image as label. The second is to evaluate its capacity of removing real sludge, in this case, we trained our network directly use the dirty image as input and corresponding clean image as label. We adopted the peak-signal-to-noise-ratio (PSNR) as the quality metric. A higher PSNR generally indicates that the reconstruction is of higher quality. In addition, we have set up an extra experiment, we use a OCR algorithm to test the model's output, and evaluate the effectiveness by the accuracy of OCR test. We named the experiment as OCR test. In this test, we no need to redivide the original dataset.

2) Parameters Setting: Althrough our SCDCA model can accept an arbitrary size of input image, but in order to make the training set more diverse we divide the training image into small patches. We set the patch size to 40×40, and stride to 10. That is to say, we crope a 40×40 image patch every 10 pixels. We initializing the parameters in SCDCA by the He initialization reported in [33] and use Adam with betas of (0.9,0.999), eps of 1e-8, weight decay of 5e-4, batch size of 128, and the learning rate is initialize of 1e-2 decayed by 5 every 30 epochs. We trained the SCDCA model on TensorFlow framework with a Nvidia Tesla K40c GPU.

3) Data Augmentation: Because the size of the data set [32] is very small, as long as 144 labeled images, so it is necessary to take some data amplification methods. We rescaledthe size to [0.8,0.9,1.1,1.2] of origin image, and expanded 4 tiems number of images. We had also taken taken the horizontal flip to doubled its number. Because of the properties of the document image, such as the text is always horizontal,so we do not adopt the methods (e.g. rotation) what will change this data distribution.

B. Denoising

To evaluate our method, we compared our results to a range of state-of-the-art methods:BM3D [12], SSDA [14], RED30 [25], DCNP [30]. The implementations of these methods have been release online by the authors or other researchers. For this experiment, We used the clean version image in training set. We randomly selected 24 images for testing, the remaining 120 images for training. We test additive Gaussian with zero mean and standard deviation $\sigma = 10, 20, 30, 40, 50$.

TABLE IV
THE AVERAGE PSNR(dB) RESULT OF $\sigma = 10, 20, 30, 40, 50$

Methods	BM3D	SSDA	RED30	DCNP	Ours
$\sigma = 10$	34.28	34.23	34.41	34.91	35.14
$\sigma = 20$	28.79	29.12	29.86	31.28	31.77
$\sigma = 30$	25.40	26.61	27.87	28.04	28.93
$\sigma = 40$	22.90	22.99	26.01	26.88	27.49
$\sigma = 50$	20.20	20.68	25.44	26.25	26.57

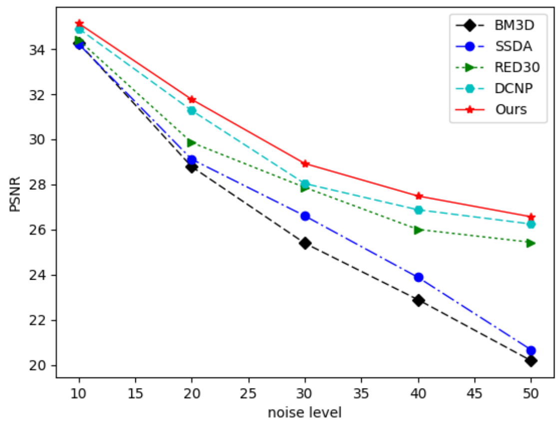

Fig. 3.

TABLE IV shows the quantitative comparisons between our method and other previous methods. For convenience, we also give the legend in Fig. 3. It can be seen that SSDA slightly better than the BM3D, and when the noise is slight($\sigma < 10$) or serious($\sigma > 50$) the SSDA is very close to BM3D. When the $\sigma <= 20$ the RED30 is closer to BM3D and SSDA, and with the σ increasing, the RED30 became close to DCNP. Out SCDCA and DCNP lead RED30 with obvious advantages. Our SCDCA can outperform the SSDA by a little margin, making a gain about 0.25 dB at the points $\sigma = 10, 50$ and a gain about 0.9 dB at $\sigma = 30$. Through Fig. 3, we can clearly see that our method has overall advantages over others. More importantly, with the increase of noise, the effect of SCDCA decreased more slowly. But the effects of BM3D and SSDA are almost linearly decrease

C. Real Scene

In this task, we trained our network directly use the dirty image as input and corresponding clean image as label. Although we claim that this is an experiment based on real scenes, but the sludges are also added artificially Just as previously described, the sludges(ink blot, fading, wrinkles, etc.) are closer to real scenes in life.

The original average psnr between 144 training dirty images and its corresponding clean images is 16.33 dB. We random choose 14 images as test, and remaining 130 images for training. Afer the trained 300 epochs, the average psnr of this 14 images is 38.74. We trained the RED30 [25] using the same setting, and the psnr between output and input is 36.62, 1.1 dB lower than our SCDCA model. Visual comparisons is given in Fig. 4. We can see that compared with middle image, our result is more clear and with less black ink points.

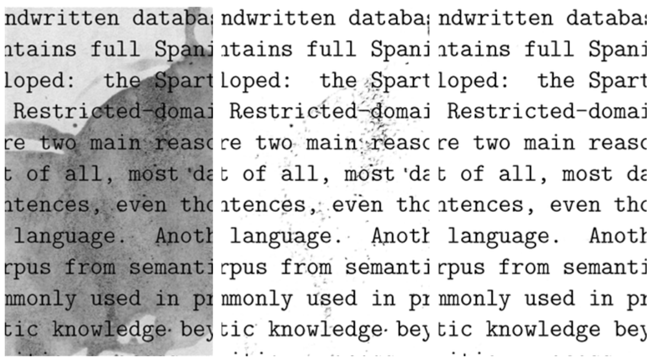

Fig. 4. Example patches of the result.The left is origin input, the middle is the output of RED30, and the right is our result.

D. OCR Test

In the OCR test, we compare the OCR result of the origin blurry images and the output images of our model. We use the open source OCR engine Tesseract. We did not any fine-tuning on Tesseract, only use its origin model. Through this experiment, we have proved that SCDCA not only makes the document image clear and recognizable, but also improves the accuracy of OCR tasks. We show an example in Fig. 5.

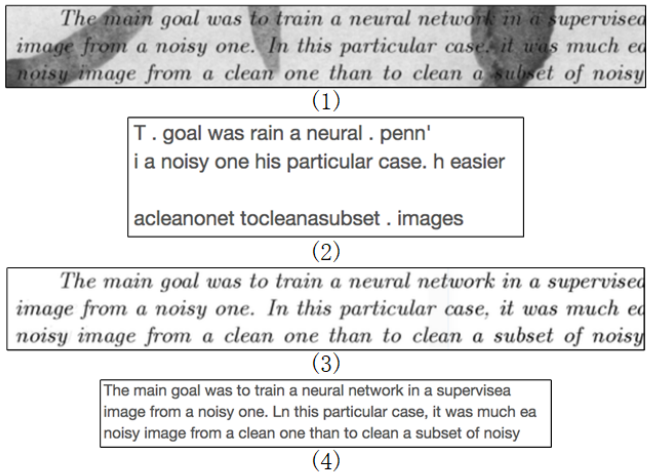

Fig. 5. Example patches of the OCR result. (1) is the blur image patch; (2) is the OCR result of blur image in (1); (3) is the output patch of (1) after SCDCA processing; (4) is the OCR result of (3). It can be seen from this example that document restoration significantly improves the accuracy of OCR.

We can see that, the OCR result of blur image is terrible, the recognition rate is less than 20%. While the OCR result of restoration image is very impressive, only two words are identified wrong. This experiment demonstrated that the restoration of document image is very helpful to improvement of the effect of OCR task.

V. Conclusion

In this paper, we propose a skip-connected deep convolutional autoencoder (SCDCA) for restoring document images. We use two types of skip-connections in SCDCA, one is used to directly pass the features from front convolution layers to later layers. The other is adopted to make the model learning the residuals between low-quality images and high-quality images, instead of learning a restoration function prior between them. We did three groups of experiments to evaluate our methods. The first one is on denoising task on different noise level, and blind deblurring. Then we test the model on a data set that simulates the really smudges. Finally, we elucidate the effectiveness of our restoration model by comparing the result of a OCR program. In the further, we intend to combine the restoration task with OCR task through an end-to-end network, and directly generate the accurate recognition results from low-quality document images.

Acknowledgment

This work was partially supported by National Natural Science Foundation of China (No.61602487), the Research Funds of Renmin University of China (No.2015030275, No.2018030202), and the Fundamental Research Funds for the Central Universities.

References

[1] J. Banerjee, A. M. Namboodiri, and C. V. Jawahar, "Contextual restoration of severely degraded document images," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 517–524.

[2] X. Chen, X. He, J. Yang, and Q. Wu, "An effective document image deblurring algorithm," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 369–376.

[3] H. Cho, J. Wang, and S. Lee, "Text image deblurring using text-specific properties," in *European Conference on Computer Vision*. Springer, 2012, pp. 524–537.

[4] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring text images via l0-regularized intensity and gradient prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2901–2908.

[5] L. Xiao, J. Wang, W. Heidrich, and M. Hirsch, "Learning high-order filters for efficient blind deconvolution of document photographs," in *European Conference on Computer Vision*. Springer, 2016, pp. 734–749.

[6] M. Hradiš, J. Kotera, P. Zemcík, and F. Šroubek, "Convolutional neural networks for direct text deblurring," in *Proceedings of BMVC*, vol. 10, 2015.

[7] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang, "Learning to super-resolve blurry face and text images," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[8] X. Jiang, H. Yao, and S. Zhao, "Text image deblurring via two-tone prior," *Neurocomputing*, vol. 242, pp. 1–14, 2017.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013.

[11] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[13] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?" in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012, pp. 2392–2399.

[14] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 341–349.

[15] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[16] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 191, 2016.

[17] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "l_0-regularized intensity and gradient prior for deblurring text images and beyond," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 2, pp. 342–355, 2017.

[18] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: residual interpolation for color image demosaicking," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1288–1300, 2016.

[19] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*. Springer, 2014, pp. 184–199.

[20] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[21] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[22] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[23] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in Neural Information Processing Systems*, 2014, pp. 1790–1798.

[24] D. Yang and J. Sun, "Bm3d-net: A convolutional neural network for transform-domain collaborative filtering," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 55–59, Jan 2018.

[25] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 2802–2810.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[27] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, no. 1, 2013.

[28] L. Tran, X. Liu, J. Zhou, and R. Jin, "Missing modalities imputation via cascaded residual autoencoder," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[29] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[30] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[31] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[32] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.