LETTER

Deep Convolutional Neural Networks for Manga Show-Through Cancellation

Taku NAKAHARA[†], ***Nonmember***, **Kazunori URUMA**[††a)], **Tomohiro TAKAHASHI**[†††], ***and*** **Toshihiro FURUKAWA**[†], ***Members***

SUMMARY Recently, the demand for the digitization of manga is increased. Then, in the case of an old manga where the original pictures have been lost, we have to digitize it from comics. However, the show-through phenomenon would be caused by scanning of the comics since it is represented as the double sided images. This letter proposes the manga show-through cancellation method based on the deep convolutional neural network (CNN). Numerical results show that the effectiveness of the proposed method.

key words: *manga, image show-through cancellation, deep convolutional neural network, residual learning, batch normalization*

1. Introduction

Recently, manga (Japanese comics) is very popular in a lot of countries, and it leads to increasing of the demand for the digitization of manga. Then, in the case of an old manga where the original pictures have been lost, we have to digitize the manga from the double sided printed image. However, the digital image obtained by scanning of double sided images is often degraded by the interference from the back side. This phenomenon is referred to as show-through [1].

Several show-through cancellation methods have been proposed [1]–[5] in order to cancel the show-through of the general scanned image. Ophir et al. have proposed the image separation method for show-through cancellation based on the total variation (TV) regularization and mean squared error (MSE) fidelity [2]. In [3], it is introduced a symmetric linear image mixture model, and an independent component analysis (ICA) based show-through cancellation method have been proposed. In [4], non-negative matrix factorization (NMF) approach for show-through cancellation method have been proposed, and Nagayasu et al. have presented show-through cancellation method using Kalman filter with colored driving source [5]. Although these methods achieve to reduce the effect of the show-through, the transmittance parameter to represent the rate of the show-through from the back side image is required. Moreover, some methods also require the back side image satisfying the accurate registration between the front side and the back side. However, the transmittance parameter is usually unknown, and it is difficult to perform the accurate registration between the front side and back side since there is a warp due to a thickness of paper when the paper is scanned. Furthermore, these methods do not work well for manga data because they have not considered the characteristic representation of manga such as the screen tone, the line drawing etc.

Active research for convolutional neural networks (CNNs) has led to the development of the image processing. Deep learning techniques for image restoration have achieved a good performance [6], [7]. Dong et al. [6] have presented a fully CNN for single image super-resolution, and this network directly learns an end-to-end original mapping between low- and high-resolution images. Zhang et al. [7] have proposed an end-to-end trainable single deep CNN for blind Gaussian denoising, single image super-resolution, and JPEG image deblocking. This network adopts the residual learning to remove the latent clean image from noisy observation, and this strategy successes in denoising with various noise level.

This letter proposes the manga show-through cancellation method based on the deep CNN. Utilizing the generative model of the show-through [1], the abundant manga data degraded by show-through are generated, and CNN modeled by [7] learns the degradation of the show-through using the set of the original manga data and the degraded data. Then, in order to be robustness for the various transmittance parameters, the show-through manga data to learn CNN are generated by using various transmittance parameters. The advantage of the proposed method is that it is required neither the transmittance parameter nor the back side image, and the novelty of the proposed method is to introduce the deep CNN approach for show-through cancellation and to consider the show-through cancellation for manga. Numerical results show the effectiveness of the proposed method compared with the conventional method [5].

2. Main Work

2.1 Image Show-Through Cancellation Problem

This subsection presents the generative model of show-through and the show-through cancellation problem [1]. Let

Manuscript received March 12, 2018.
Manuscript revised July 7, 2018.
Manuscript publicized August 2, 2018.

[†]The authors are with Department of Management Science, Tokyo University of Science, Tokyo, 125–8585 Japan.
[††]The author is with Department of Computer Science, Kogakuin University, Tokyo, 163–8677 Japan.
[†††]The author is with Department of Human and Information Science, Tokai University, Hiratsuka-shi, 259–1292 Japan..
a) E-mail: uruma@cc.kogakuin.ac.jp

DOI: 10.1587/transinf.2018EDL8051

$x^f_{i,j}$ and $x^b_{i,j}$ denote the (i, j)th pixel value of the front side original image and the back side original image, respectively. The transmittance parameter to represent the rate of the show-through from the back side image and the point spread function (PSF) are denoted as $s \in [0, 1]$ and $h_{i,j}$, respectively. Here, the (i, j)th pixel value of the front side observed image $y^f_{i,j}$ and the (i, j)th pixel value of the back side observed image $y^b_{i,j}$ are given as follows,

$$\begin{aligned} y^f_{i,j} &= x^f_{i,j} - s \textstyle\sum_p \sum_q h_{p,q}(l_{max} - x^b_{i-p,j-q}) \\ y^b_{i,j} &= x^b_{i,j} - s \textstyle\sum_p \sum_q h_{p,q}(l_{max} - x^f_{i-p,j-q}) \end{aligned}, \tag{1}$$

where l_{max} is the maximum luminance value of image. Then, considering show-through model for all the pixels in an image, Eq. (1) is described as the following vector form,

$$\begin{aligned} \boldsymbol{y}^f &= \boldsymbol{x}^f - sH(\boldsymbol{l}_{max} - \boldsymbol{x}^b) \\ \boldsymbol{y}^b &= \boldsymbol{x}^b - sH(\boldsymbol{l}_{max} - \boldsymbol{x}^f) \end{aligned}, \tag{2}$$

where $\boldsymbol{y}^f$, $\boldsymbol{y}^b$, $\boldsymbol{x}^f$, $\boldsymbol{x}^b$ and $\boldsymbol{l}_{max} \in R^M$ are the vectors composed of $y^f_{i,j}$, $y^b_{i,j}$, $x^f_{i,j}$, $x^b_{i,j}$ and l_{max}, respectively. $H \in R^{M\times M}$ denotes convolution operator matrix composed of $h_{p,q}$, and M is the number of pixels in image. Therefore, the show-through cancellation problem is to restore the $\boldsymbol{x}^f$ from $\boldsymbol{y}^f$.

2.2 Denoising Convolutional Neural Networks (DnCNNs) [7]

This subsection presents the denoising method based on the denoising CNN (DnCNN) model proposed by Zhang et al. in [7]. The goal of the paper [7] is to restore a clean image $\boldsymbol{x}$ from a noisy observed image $\boldsymbol{y}$ which follows an image degradation model $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{v}$ is a noise image. In order to restore $\boldsymbol{x}$ from $\boldsymbol{y}$, Zhang et al. adopt the residual learning formulation, that is, DnCNN aims to learn the residual mapping function $R(\boldsymbol{y}) \approx \boldsymbol{v}$.

In order to obtain $R(\boldsymbol{y})$, Zhang et al. use the MSE as the loss function between the residual image and the estimated image from noisy input as follows,

$$L(\Theta) = \frac{1}{2N}\sum_{n=1}^{N} \|R(\boldsymbol{y}_n; \Theta) - (\boldsymbol{y}_n - \boldsymbol{x}_n)\|^2_2, \tag{3}$$

and the trainable parameters Θ in DnCNN is estimated using N training image pairs. Then, the DnCNNs are constructed from three layer types. The first layer is constructed from 64 convolutional filters of size $3 \times 3 \times c$, and the rectified linear units (ReLU), i.e., $max(0, \cdot)$, where c represents the number of image channels. Layers from 2 to $(D-1)$ are constructed by using 64 numbers convolutional filters of size $3 \times 3 \times 64$, batch normalization (BN) [8], and ReLU. D is depth of the DnCNNs. For the last layer, c convolutional filters of size $3\times3\times64$ are used to reconstruct the output. By incorporating convolution with ReLU, DnCNN can gradually separate image structure from the noisy observation through the hidden layers. Then we obtain the restored clean image $\boldsymbol{x} = \boldsymbol{y} - R(\boldsymbol{y})$ using $R(\boldsymbol{y})$.

General CNN mapping to the clean image $\boldsymbol{x}$ from $\boldsymbol{y}$ would be close to identity mapping when the noise level is low. As this result, it leads to the low denoising effect. However, the residual learning can prevent falling out this, and therefore, the method proposed in [7] gives well denoising results for various noise level without using the parameter such as noise level.

2.3 CNN for Manga Show-Through Cancellation

This subsection presents the manga show-through cancellation method based on CNN. Generally, we cannot know the accurate transmittance parameter s, and it is even possible to be zero. Therefore, in order to prevent being the identity mapping, this letter adopts to learn the residual mapping function. That is, this letter proposes to learn the residual mapping $R(\boldsymbol{y}^f) \approx -s\boldsymbol{h}(\boldsymbol{l}_{max} - \boldsymbol{x}^b)$. Then this letter also use the MSE based loss function as follows,

$$L(\Theta) = \frac{1}{2N}\sum_{n=1}^{N} \|R(\boldsymbol{y}^f_n; \Theta) - (\boldsymbol{y}^f_n - \boldsymbol{x}^f_n)\|^2_2, \tag{4}$$

and the trainable parameters Θ in DnCNN is estimated by using abundant show-through manga data generated by Eq. (2). Here, in order to be robustness for the various transmittance parameter, the show-through manga data to learn the network are generated using various transmittance parameters. The loss function is minimized using Adam algorithm proposed in [9].

Same as the method in [7], the proposed manga show-through cancellation CNN consists of the three layer types. The first layer is convolution and non-linear mapping of ReLU layer, and this layer extracts an 8-dimensional feature for $3 \times 3 \times c$ size patches. In this letter, image channel is $c = 1$ because we deal with grayscale manga. From the second to $(D - 1)$th layers, each layer is constructed from convolution, BN [8] and ReLU. The input 8-dimensional vectors are convoluted by $3 \times 3 \times 8$ size patches. The last layer is convolution layer which outputs the residual manga image. The output image is given by $3\times3\times8$ size filter. Note that the batch normalization is used from the second layer to the $(D - 1)$th layer, and this strategy is based on the previous studies [7], [12]. Then, Fig. 1 shows the architecture of the proposed show-through cancellation CNN. Finally, the restored manga image is obtained as $\boldsymbol{x}^f = C(\boldsymbol{y}^f - R(\boldsymbol{y}^f))$, where $C(\bullet) : R^M \to R^M$ is the function to clamp the pixel values as follows,

$$(C(a))_i = \begin{cases} 0 & (a)_i < 0 \\ (a)_i & 0 \le (a)_i < l_{max} \\ l_{max} & l_{max} \le (a)_i \end{cases}. \tag{5}$$

Here, $(a)_i$ is the ith element value of a vector $\boldsymbol{a}$.

3. Experiments

In this section, using Manga109 dataset [10], we present the

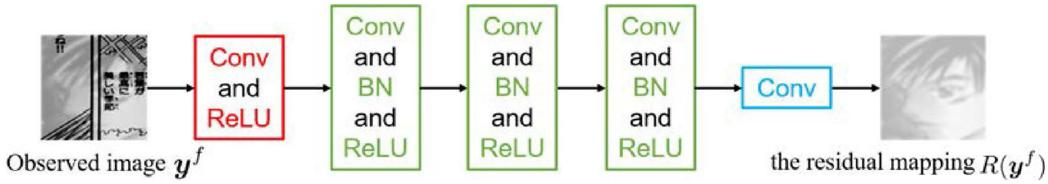


Fig. 1 The architecture of the proposed show-through cancellation CNN.

explanation of the constructing of the proposed CNN and the numerical examples of the show-through cancellation.

In training of the proposed CNN, we use 190 and 30 manga images for the training images and the validation images, respectively. The validation images are used to measure the generalization error. For reasons of implementation simplicity, as the pre-proccesing, the training images and validation images are resized to 500 × 500 size by the bicubic method without keeping the aspect ratio. Note that it is pointless to keep the aspect ratio because there is many expressions such as the deformation of the characters in manga. The 50 × 50 size image is extracted from the 500 × 500 show-through manga image, and about 200000 images are used for training. Then the manga data with the show-through are constructed based on Eq. (2), where the transmittance parameter is set to $s \in [0.1, 0.4]$, and a 5×5 Gaussian filter with $\sigma = 2$ is used as PSF matrix H. The maximum luminance value is set to $l_{max} = 1$. The number of layers of CNN is set to $D = 5$, and a mini-batch size is 128. The dimension of the input layer is set to 50 × 50. For parameters of Adam algorithm in [9], we use $\alpha = 0.01$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The proposed CNN is trained to 50 epochs. The learning rate 10^{-3} and 10^{-4} is used for the 50 epochs.

This letter compares the proposed method with the method of Nagayasu et al. [5] because the Nagayasu's method achieves higher performance than the Sharma's method [1]. Nagayasu's method requires both the accurate value of transmittance parameter s and the back side image satisfying the accurate registration between the front side and the back side. In order to compare the performance of the show-through cancellation methods, this experiment uses 500 × 500 size 10 test images to demonstrate. These test images are called as Image1–Image10, where each image represent each page of manga such as Fig. 2. Since the size of input layer is 50 × 50, the image is processed by each block. Note that these images are not included in training/validation data to construct the proposed CNN.

Figure 2 shows the results of show-through cancellation of Image1, and Fig. 3 shows the zoomed images of show-through cancellation results of Image1, Image2 and Image3. As is shown, the proposed method achieves a better show-through cancellation than the conventional method. Although the Nagayasu's method reduces the effect of the show-through phenomenon, the front side information is also turned off. Tables 1 and 2 show the peak signal to noise ratio (PSNR) and structural similarity (SSIM) [11] for $s = 0.2, 0.4$. We can see that the proposed method achieves

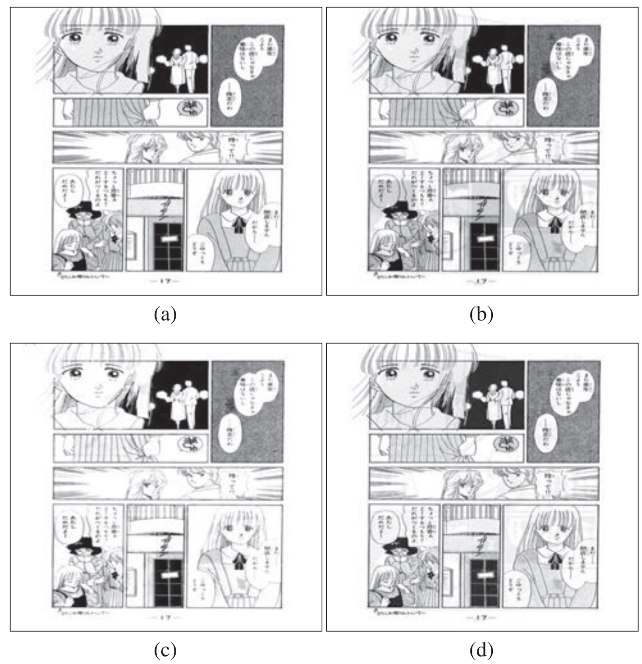


Fig. 2 Visual comparison of Image1: (a) original image, (b) show-through image ($s = 0.2$), (c) conventional method [5] and (d) proposed method. ©Riku Kurita

higher performance than the Nagayasu's method. The limitations of the proposed method is that there is not enough effects for the show-through of straight drawing. It is because there is a lot of straight drawing at frame regions in manga and because they are drawn strongly. Therefore, it is difficult to cancel the straight show-through appropriately.

This letter explains the experimental results using real manga data which are commercially available comics. Authors bought several comics and scanned the some pages of the comics by a multi function printer. Then authors obtained the show-through manga data, and the proposed method was applied for the show-through cancellation. As the results, the proposed method cancels the show-through appropriately similar to the results of the proposed method shown in Figs. 2 and 3. Note that this letter can not show these results due to copyright restrictions.

Finally, using the Manga 109 image data, this letter shows the experimental results similar to the experiment using a real data. That is, the images of Manga 109 are printed to the front side and back side of the medium quality paper by inkjet printer. Then, the show-through manga data is scanned as a 7015 × 4960 size digital image by a multi function printer, and it is resized to 7050 × 5000 by the padding

(a) (b) (c) (d)

Fig. 3 Visual comparison of the zoomed Image1 (1st row, $s = 0.2$), Image2 (2nd row, $s = 0.2$) and Image3 (3rd row, $s = 0.4$): (a) original image, (b) show-through image, (c) conventional method [5] and (d) proposed method. ©Riku Kurita

Table 1 Comparison of PSNR [dB] for $s = 0.2$ and $s = 0.4$.

Image	$s = 0.2$		$s = 0.4$	
	Conv. [5]	Proposed	Conv. [5]	Proposed
Image1	18.685	**29.350**	18.311	**27.014**
Image2	19.224	**32.062**	18.403	**28.500**
Image3	20.183	**30.703**	18.565	**27.883**
Image4	20.105	**31.399**	19.819	**28.420**
Image5	20.128	**32.676**	19.977	**28.235**
Image6	20.566	**32.721**	20.143	**28.225**
Image7	19.177	**32.997**	19.294	**28.114**
Image8	19.809	**31.050**	19.621	**28.316**
Image9	19.767	**32.298**	19.055	**27.511**
Image10	19.561	**31.337**	18.688	**27.129**
Average	19.720	**31.659**	19.187	**27.935**

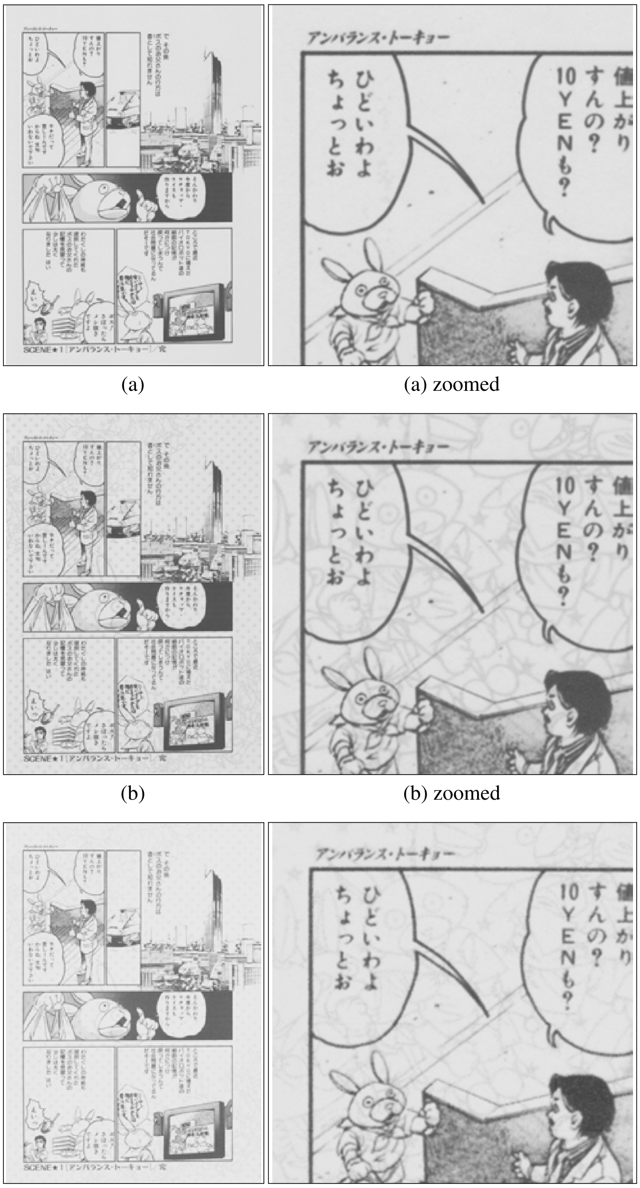
Table 2 Comparison of SSIM for $s = 0.2$ and $s = 0.4$.

Image	$s = 0.2$		$s = 0.4$	
	Conv. [5]	Proposed	Conv. [5]	Proposed
Image1	0.811	**0.960**	0.799	**0.941**
Image2	0.858	**0.968**	0.840	**0.951**
Image3	0.842	**0.958**	0.795	**0.926**
Image4	0.854	**0.926**	0.834	**0.897**
Image5	0.844	**0.971**	0.838	**0.935**
Image6	0.857	**0.972**	0.846	**0.935**
Image7	0.828	**0.974**	0.842	**0.949**
Image8	0.858	**0.945**	0.855	**0.924**
Image9	0.842	**0.972**	0.832	**0.945**
Image10	0.835	**0.975**	0.815	**0.952**
Average	0.843	**0.962**	0.830	**0.935**

processing. The image is processed every each the 50×50 size block. Note that the printed manga data is not included in the training image set, validation image set and test image set. Figure 4 shows the results of the show-thorough cancellation of the proposed method. Although it is difficult to quantitatively evaluate, the proposed method seems to cancel the show-through well.

4. Conclusion

This letter proposes manga show-through cancellation method based on the CNN which adopts the residual learn-



(a) (a) zoomed

(b) (b) zoomed

(c) (c) zoomed

Fig. 4 Experiment results for both side printed comic data: (a) original image, (b) scaned show-through image and (c) the result of proposed method. ©Minako Uchida

ing formulation and the batch normalization. The proposed CNN model achieves the show-through cancellation without using transmittance parameter. The experimental results show that the proposed method achieves to remove the show-through appropriately compared with the previous method based on the Kalman filter [5].

References

[1] G. Sharma, "Show-through cancellation in scans of duplex printed documents," IEEE Trans. Image Process., vol.10, no.5, pp.736–754, 2001.

[2] B. Ophir and D. Malah, "Show-through cancellation in scanned images using blind source separation techniques," Proc. IEEE Int. Conf. Image Process., vol.3, pp.233–236, 2007.

[3] A. Tonazzini, E. Salerno, and L. Bedini, "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique," Int. J. Document Anal. Recognit., vol.10, no.1, pp.17–25, 2007.

[4] F. Merrikh-Bayat, M. Babaie-Zadeh, and C. Jutten, "Using non-negative matrix factorization for removing show-through," Proc. LVA/ICA, Lecture Notes in Computer Science, vol.6365, pp.482–489, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[5] R. Nagayasu, N. Tanabe, and T. Furukawa, "Show-through cancellation method using Kalman filter algorithm with colored driving source in duplex scanning," Journal of Image Society of Japan, vol.55, no.4, pp.415–418, 2016.

[6] C. Dong, C.C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell., vol.38, no.2, pp.295–307, 2016.

[7] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process., vol.26, no.7, pp.3142–3155, 2017.

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Proc. Int. Conf. Machine Learn., pp.448–456, 2015.

[9] D.P. Kingma and J.L. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980, 2014.

[10] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using Manga109 dataset," Multimedia Tools and Applications, vol.76, no.20, pp.21811–21838, Springer, 2016.

[11] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.

[12] P. Xiang, L. Wang, J. Cheng, B. Zhang, and J. Wu, "A deep network architecture for image inpainting," Proc. IEEE Int. Conf. Computer and Communications, pp.1851–1856, 2017.