Textual Deblurring using Convolutional Neural Network

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

07-10-2021 / 16-10-2021

CITATION

Nadeem, Muhammad Shahroz; Hussain, Sibte; Kurugollu, Fatih (2021): Textual Deblurring using Convolutional Neural Network. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16760632.v1>

DOI

[10.36227/techrxiv.16760632.v1](https://doi.org/10.36227/techrxiv.16760632.v1)

Textual Deblurring using Convolutional Neural Network

Muhammad Shahroz Nadeem
University of Derby
Derby, United Kingdom.
M.Nadeem@derby.ac.uk

Sibt ul Hussain
NUCES-Fast
Islamabad, Pakistan
sibtul.hussain@nu.edu.pk

Fatih Kurugollu
University of Derby
Derby, United Kingdom.
F.Kurugollu@derby.ac.uk

Abstract

The problem of blind textual deblurring of images is a classical problem in image restoration. Here we propose a new formulation of Convolutional Neural Network (CNN) for solving this problem. Our architecture has been carefully designed after thorough empirical evaluation with detailed experimentation of the recent advances in the domain of DL. Our method outperforms the previous state of the art CNN based blind deblurring method, while reducing the training time, memory and depth. These traits make our method a potential candidate to be integrated as a defacto method for textual deblurring in mobile devices. Precisely, our contributions are: (i) a CNN model that outperforms the state of art method for blind textual deblurring; our this method is also resource friendly and takes less time to train. (ii) analysis on the learning behavior of CNN for textual deblurring by encoding structural information of images through image gradients. (iii) recommendations based on detailed empirical validations of different design choices for building an image deblurring network.

1. Introduction

Rapid digitisation has resulted in corporations' adoption of paperless management systems. However, paper based textual documents still exist in large amount of data that are stored in textual form; like documents, notes, financial ledgers, books and religious writings. Digitising them requires special equipment like high resolution cameras and scanners that captures such data with higher pixel quality and lower noise. Google has made certain efforts in making high resolution cameras (Art Camera) which they are using to preserve artwork [20]. However, such equipment are expensive and not readily available.

In contrast, now almost everybody has a low end mobile connected with camera. Especially, during the covid-era online sharing of written documents scanned through hand held devices has become a norm. However, capturing textual images using these cameras require paying special

attention to image capturing process. For instance a poor lighting condition or a simple shake of hand during image capturing process can lead to a poor quality blurred image [8].

Traditional blind image deblurring methods utilise manually perceived hand crafted features, employing edges or gradients with local image statistics to reconstruct or recover the blur kernel [18, 4, 3, 12, 19]. Recent, CNN based methods have removed this limitation [8, 1, 16, 15]. To this end we develop, by careful analysis and evaluation of different CNN architectures and model designing choices, to come up with an architecture that gives state of the art performance on the benchmark datasets for text deblurring. Additionally, we propose a new loss function for textual deblurring of images, called the “**Structured Loss**”. In this work we show that structured loss is better at deblurring text in images in contrast to simply using L_2 loss, due to its ability to focus more on edges. In this loss function, first order derivative is added to the L_2 loss, which forces the CNN model to generate clear and sharp human readable textual images.

We would like to emphasize that we have achieved all the performance gains, while reducing the memory footprint of the model. This makes our model a suitable candidate for execution of textual deblurring in small hand held device like mobile phones or tablets. It should be also noted that our models' spatial support is not defined by the kernel size of the model. Rather, it takes the whole ($300 \times 300 \times 3$) images for processing. Furthermore, in contrast to other methods, during comparison we have computed all the evaluation matrices Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) using whole images, rather than the (160×160) central patch of the image. Since, we believe that in real life setting text can be captured on any part of the image during image capturing process.

The main contributions of the paper are as follows:

1. Proposed a new loss function for textual deblurring called the “**Structured Loss**”. This loss function utilises the first order derivative of the image, which is added to the L_2 components in the loss function. This

new loss function facilitates the CNN network to focus more on edges thus generates sharp textual images.

2. Perform empirical evaluation of the design choices for the task of textual deblurring. Exhaustive investigation provides a study about the effects of sampling techniques, types of normalisation, loss functions, and data dimensionality has on the performance of textual deblurring.
3. a new resource efficient CNN model that outperforms the CNN model of Hardis et al. [8] while having a 8 times smaller memory footprint. Providing a PSNR value of 18.6 in comparison to the state-of-the-art value of 16.06.

The remaining of the paper is organised as follows. Section 2 provides a brief overview of the problem domain. Section 3 provides extensive literature review of both traditional and deep learning based methods. Section 4 describes our approach, dataset utilised in the experimentation process, evaluation metrics, the proposed loss function and models’ architecture. Section 5 elaborate the effects of different design choices, which aided in the designing and finalising of our CNN model. In section 6 visual and empirical analysis of the achieved performance is provided. Section 7 provides some promising research directions. Finally section 7.1 concludes the paper.

2. Background

To evaluate the deblurring quality of a method the closeness of the perceived image to the captured image is measured. Specifically, image quality can be defined as presence of some desirable characteristics such as sharpness, tone reproduction, contrast and colour; while absence of some undesirable characteristics such as blur, noise and incorrect exposure. We can address these problems either by being careful during image capturing and using specialised hardware such as image stabilisers, hybrid cameras, coded shutters or by post processing the captured image by specialised algorithms.

In image processing to recover high quality image from a poor quality one, we apply different image restoration algorithms. Image restoration is an interesting and classical image processing problem [14]. Image deblurring is a special case of restoration problem in which a blurred image is processed to produce a sharpened image. Like majority of other noises, blur is introduced due to physical and manual errors such as motion and defocus blur. Other cases include noise, incorrect exposure and distortion effects, *etc.*

Image processing algorithms formalize the image restoration problem as recovery of blurring kernel K , i.e. if we can recover the blur kernel K , we can quasi recover

the original image by convolving the image with this blurring kernel k , specifically:

$$I_{out} = I_{in} * K + n \quad (1)$$

where I_{out} is the deblurred image, I_{in} is the blur image given as input, K is the kernel and n is the random noise added during image capturing system, here $*$ is used to imply convolution as shown in equation 1. Generally speaking these algorithms can be categorised into two broad categories.

Non blind deblurring: The non blind deblurring methods explicitly take the kernel K or Point spread function (PSF) and blur image as input to generate the deblur image – *c.f.* Figure 1. Mostly used and available tools such as (Photoshop, GIMP, *etc.*) use non blind deblurring techniques to recover images. However, these are limited in their use as user has to specify hand crafted kernels to generate final deblur images. Furthermore, these hand crafted kernels remove blur to the extent for which they are build for.

Blind deblurring: For blind deblurring methods neither the PSF nor any other prior information to generate the deblur image is known. This is why blind deblurring is difficult and considered an ill posed problem [8, 18, 3]. A problem which has multiple solutions is called an ill posed problem.

Figure 1: Blur image generation via convolution with a kernel (or PSF).

3. Literature review

All the previous deblurring methods typically try to approximate the functional form of PSF used to generate the blur image. Once we have the approximation of PSF, any

non blind deblurring method can be used to deblur the image. The performance of all such methods depend on how good is the approximation function. Many such methods have been proposed for blind deblurring [1, 16, 15, 3, 4, 18, 12].

3.1. Hand-crafted feature methods.

Prior to Deep Learning (DL), traditional methods designed for vision tasks utilised manually perceived and hand crafted features to extract relevant information for the PSF approximation. Furthermore, they also put constraints on specific structure of underlying documents such as either the document needs to be bi-modal (black and white without any complex background or graphics) or to be in specific orientation, or in specific lighting conditions, *etc.* Thus these models are not generalizable to real life settings of mobile connected cameras, where user will be taking pictures without any predefined constraints. For instance, Li et al. [12] method works very well on strictly two tone images but failed on images with complex background or noise. It was not a very robust or scalable method and works only for a subset of blur kernels. They used inverse mapping function to deblur images, and evaluated their performance on total misclassified pixel count.

Cho et al. [4] use text image specific properties with optimization techniques to produce their final output. These properties include contrast difference between text with near background and uniformity in color of printed text characters, *etc.* Clearly, these assumptions do not hold for real-world images. This method failed to deblur textual images if the text present in the images was in smaller fonts or connected. Moreover, it also worked only on a subset of blur kernels and failed to remove motion blur as well. The method works very well for text images as well as natural images.

Chen et al. [3] designed a very good method for blind deblurring as it was successful in suppressing ring effect and had sufficient PSNR improvement. However, the performance of their method depended on which text segmentation method was being used. Due of this their method was not generalisable as text segmentation method has to be selected based on the type of images.

Pan et al. [18] generated perceptually good deblurred images and, their methods was also less complex compared to previous methods. But for their method to work, textual images had to be horizontally aligned, bi-modal and uniform otherwise the method produced poor deblurred images.

All these manually hand crafted feature selection methods failed to be generalizable, robust and scalable. Furthermore, they were limited to certain deblurring kernels.

3.2. Deep Learning methods

Typical DL methods have seen many advancements in recent time. Primarily due to their ability to automatically learn feature representations from data. Due to this DL methods are state-of-the-art in object detection [23, 5], image segmentation [13, 6, 11], depth perception [2, 25, 21], *etc.*

Naturally, many DL methods have been proposed for blind deblurring of images. Nah et al. [15] proposed a multi-scale Convolutional Neural Network (CNN) and a new dataset captured using high speed camera. Their method works it way up from coarse-to-fine, and finally gives a high resolution deblurred image. Further, according to them removing batch normalisation layer helps network to cater for range flexibility thus improving deblurring performance. Noroozi et al. [17] also proposed a multi-scale DeblurNet^{Wild} for deblurring trained on blurry and sharp image pairs with skip connections to produce the residual image.

Zhang et al. [29] on the other hand combined two Generative Adversarial Networks (GAN), namely learning-to-blur (BGAN) and learning-to-deblur (BDGAN). Similarly, Kupyn et al. [9, 10] also utilised GANs to produce deblurred images. They employ depth-wise separable convolutions.

Recurrent networks have also been used for deblurring task. Zhang et al. [28] reconstruct deblurred images after passing it through a series of convolutional and Recurrent Neural Network (RNN) based deconvolution network. They show that deblurring can be modelled as a infinite impulse response (IIR) model.

instead of deblurring the whole image, Zhang et al. [27] perform patch wise deblurring. Motivated by Spatial Pyramid Matching, their Deep Multi-Patch Hierarchical Network (DMPHN) perform deblurring in a fine-to-coarse manner in a novel stacking approach.

Ren et al. [22] propose an autoencoder with skip connection with sigmoid non linearity. In addition they employ a Fully connected network, combined together with softmax. Their method called the selfdeblur method capture the deep prior to produce the blur kernels.

Majority of the DL methods are mostly based for scenic images. Hardis et al. [8] proposed a deblurring method for deblurring textual images. They investigate how CNNs perform textual deblurring. They have made two major contributions in their work. Firstly, they have introduced a new dataset for textual image deblurring. Secondly, they have proposed a CNN model that gives state of the art performance on dataset produced by them. They have compared the performance of their CNN with classical text deblurring techniques.

4. Our Approach

We have developed a CNN based algorithm that gives better performance than the previous state of the art [8], while being resource friendly and faster to train. We have explored different design choices. Empirically, proven what their affects are on the final performance. Which combinations provides better learning opportunity and which cause the loss in performance. With this we have developed a road map on what practices to follow on problems that are similar to this in nature.

We also have throughly and empirically evaluated different loss functions and judge which simplify learning process for the CNN model. By directly minimizing loss or maximizing PSNR and implemented custom loss functions which aid the quick convergence of CNN.

4.1. Dataset

For our deblurring experiments we used the publicly available dataset of [8]. This dataset consists of 66,741 images. Each image has a dimensionality of $(300 \times 300 \times 3)$. Thus totaling up-to approximately 16.5 GB in size.

The dataset was uniformly divided into several chunks to ease training. Each chunk had 3000 images. 17 chunks were dedicated for training. 3 chunks for validation and 2 chunks for testing. Testing set was used only once after doing complete configuration and tuning on validation set once the training was complete.

4.2. Evaluation Metrics

For judging the quality of image restoration algorithms two types of evaluation metric are generally used; subjective and objective [7]. Subjective metrics are based on human judgment and we compare how close the images are perceptually whereas objective metrics rely on numeric criteria [7]. In these metrics we compare images based on their pixel values. To compare performance for image deblurring we will use two evaluation metrics, one objective and one subjective.

Peak Signal to Noise Ratio (PSNR): PSNR is the logarithmic ratio between power of maximum possible pixel value to Mean Square Error (MSE) (*c.f.* Equation 2) of two images, *i.e.* it measures how good an image is reconstructed by pixel level comparison. PSNR (*c.f.* Equation 3) is computed in decibels where higher value represents better results.

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (y_{i,j} - x_{i,j})^2 \quad (2)$$

$$PSNR = 10 * \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (3)$$

Structural Similarity Index Measure (SSIM): SSIM is a function of image windows x and y and is generally defined as (*c.f.* Equation 4). Specifically, SSIM measures luminance (*c.f.* Equation 5), contrast (*c.f.* Equation 6) and structural similarity (*c.f.* Equation 7) between two image windows x and y . SSIM value ranges from zero to one. SSIM value can only be one if and only if $x = y$, Zero value indicates no correlation between images thus no structural similarity. Constants c_1 , c_2 and c_3 are added to avoid null denominator.

$$SSIM(x, y) = l(x, y) c(x, y) s(x, y) \quad (4)$$

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (5)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (6)$$

$$s(x, y) = \frac{\sigma_{x,y} + c_3}{\sigma_x\sigma_y + c_3} \quad (7)$$

PSNR is poor at discriminating structural difference between images. There are many transformations/degradations that can be applied to an image and same PSNR can be achieved even though generated images are perceptually and structurally different [24]. So we have used both PSNR and SSIM because we wanted our model to generate both numerically and perceptually good image deblurring results.

4.3. Structured Loss

We Implemented the structured loss function, to analyse what effect it has on CNNs learning and whether it guides the CNN to produce sharp images or not. The magnitude of first order derivative of the image was added with L_2 loss. The participation of both components was controlled by alpha and beta hyper parameter. Here $mag_{\nabla} I$ represent the magnitude of gradient image produced by applying first order derivative filter.

Structured loss was basically the result of a (3×1) and a (1×3) kernel $([1, 0, -1])$ convolved with the output images. Basically, we were adding the first order derivative component to the loss in order for the CNN to produce sharp images as shown in Equation 8. The goal of this loss function was to enable CNN to quickly distinguish the edges and generate a sharp image by encoding font shapes structures.

$$StructuredLoss = \alpha MSE + \beta ||mag_{\nabla} I_o - mag_{\nabla} I_p||^2 \quad (8)$$

Figure 2: Proposed architecture of CNN

4.4. Our Architecture

Our proposed architecture is a 7 layer deep CNN which has (11×11) convolutional kernels and 128 channels per layer excluding the last layer. Last layer is a (1×1) convolution layer with 3 output channels shown in Figure 8. Our this model outperforms the current state-of-the-art PSNR achieved by Hardis et al. [8] for textual deblurring.

All of this was done with a reduced memory foot print due to the fact that we used pooling and unpooling layers in our model. For pooling operation we used max pooling layer which picks the max pixel value while reducing dimensionality. And, for unpooling deconvolution layer was used. This turns out to be better unpooling layer [26] in our experiments as it learns which features to pick while up sampling the image. Due, to this we have greatly reduced the memory foot print of our model.

Our this model requires 3 days to train on a single K40 GPU which is also less then [8], as their model takes one week to train. Our this initial model training was however slow because it required very low learning rate of $1 \times e^{-6}$ to learn. To speed up the convergence we also experimented with batch and instance normalization. Both did fine on training and we were able to set the learning rate as high as $1 \times e^{-1}$

Batch normalization however failed during validation as this layer learns the mean and variance of the whole dataset during training and applies the learned parameters during testing and validation. This caused the output image to be distorted as combined mean and variance of the whole

dataset was being used to normalize a neuron output.

Instance normalization however worked better as it normalizes each image individually rather with respect to a batch, it is applied evenly during both training and validation/test. We then decided to go with instance normalization instead of batch normalization.

We also used structural loss function which simplifies the learning criteria for the CNN model. Due to which the model quickly focuses on textual information present in the images. This loss function focuses the CNN learning capacity on text prior due to which we have achieved good textual deblurring. Many variants were then derived from this model to further fine tune and squeeze performance.

5. Design choices

Here we discuss in detail all the experiments and results done by fine tuning and putting to test different configurations for our architecture. We discuss what affect each configuration has on the proposed architecture performance. Further, we also shed light on why such a configuration works well or poorly for our this problem of blind deblurring. We also empirically prove why the configuration we selected is the best trade off between performance, resource consumption and time. We also perform comparison of our results with existing state of the art method and show superiority of our method (Shown in Table 2).

5.1. Effects of Down and Up-Sampling

We trained two models where we tested different upsampling and downsampling techniques. For downsampling

we experimented with max, average and L_2 -norm pooling. Max pooling performed far better than other downsampling techniques. Max pooling helped in suppressing unnecessary information. Max pooling in conjunction with convolution produced better deblurring.

For upsampling, we experimented with fixed unpooling and transposed convolution. Transpose convolution was a better upsampling technique. Because, it could learn what features to give importance to while upsampling. The images produced this way were better deblurred. Fixed unpooling layer produced images which were poor in perceptual quality as compared to transposed convolution layer. With the addition of down sampling images dimensionality was reduced, which had a great affect on reducing the computational complexity. Model became faster at convergence but still took one week to train. We named this model as variant 2.

5.2. Effects of Normalization

Although results that were being produced by variant-2 were promising. However, the convergence speed was still very slow and took approximately 1 week to train. This could be pointed to the low learning rate that was set in our base model. This slow convergence speed was a very serious bottleneck in our experimentation process. In order to perform more experimentation and explore different design choices quickly with the available resources and time we needed to boost the learning rate higher. For this we had to add a normalization layer in our CNN. So, we tried batch and instance normalisation. This model was named Variant-4.

With both the normalisation techniques we were able to set high learning rates of up to $1 \times e^{-1}$. Both worked fine during training phase. The problem aroused during validation phase. Batch normalisation caused image distortion because it tries to normalise the whole batch using the learned mean and variance during training. This proves the claim made by Nah et al. [15] that removing batch normalisation improves deblurring performance.

Instance normalisation however did the trick as it only normalises each image with respect to its own mean and variance; not across batch. Therefore, fixes the range flexibility problem in deblurring [15]. To conclude instance normalisation is a better normalisation technique for textual deblurring in images.

5.3. Effects of Loss Function

We trained another variant which was exactly like Variant-4 with the exception of the loss function. We called it Variant-3 For Variant-3 we removed our custom loss function and replaced it with the defacto loss function used in image restoration problems: L_2 loss. With this change we exactly wanted to see how better our custom loss function

compares to L_2 loss at deblurring of textual images. And, what effect can a loss function have on the learning process.

After, evaluation we found out that our custom loss function was better at deblurring the textual images (*c.f.* Table 2). Even though the PSNR is barely better in the graphs we argue that this can be further improved by simply changing the ratio of participation of structural part and L_2 part. This avenue has not yet being fully explored and needs further experimentation.

5.4. Effects of Increasing Data Dimensionality

we increased the depth of Variant-3 from 7 to 9. We added convolution layers after each deconvolution layer in hopes that these layers will consolidate the up sampled images by restoring image features. We call it Variant-5.

Although, we had previously argued that adding these layers reduce computational complexity but sampling also has a drawback. When we down sample an image we reduce the dimensionality which causes reduction in computational cost but, we also reduce the amount of information that we are passing to the next convolution layer.

Keeping this in mind we ran another Variant-6 to check either sampling techniques were a hindrance to the performance of the model. This model achieved a PSNR of 18.57 db which is better then Variant-3 and Variant-4. However, the improvement is better that also at a higher computational cost. This proves that downsampling techniques does reduce a little performance but at the cost of saving computational complexity. After this experiment, we decided to keep the downsampling in our final variant. Because, the performance gain was not that significant to the gain in computational complexity.

5.5. Experimentation Discussion

For all the above variants same initialization technique was used and were similar in all the other parameters expect those which were mentioned (*c.f.* Table 2). Kernel size and number of channels was same. except variant 4.1 in which number of channels was reduced to 80. So far, Variant-4.1 is the best model configuration which offers best trade-off between computational complexity, training time, forward pass time and, image deblurring PSNR and SSIM shown in Table 1.

We would like to emphasize that we have achieved all the performance gains, while reducing the memory footprint of the model. This makes our model a suitable candidate for execution of textual deblurring in small hand held device like mobile phones or tablets. It should be also noted that our models' spatial support is not defined by the kernel size of the model. Rather, it takes the whole ($300 \times 300 \times 3$) images for processing. Furthermore, in contrast to other methods, during comparison we have computed all the evaluation matrices Peak Signal to Noise Ratio (PSNR) and Struc-

	Depth	Learning rate	Normalization	Pooling/unpooling	Loss function
Variant-1	7	$1e^{-6}$	✗	✗	structure loss
Variant-2	7	$1e^{-6}$	✗	✓	structure loss
Variant-3	7	$1e^{-2}$	✓	✓	l_2 loss
Variant-4	7	$1e^{-2}$	✓	✓	structure loss
Variant-5	9	$1e^{-2}$	✓	✓	l_2 loss
Variant-6	7	$1e^{-2}$	✓	✗	l_2 loss

Table 1: All trained variant configurations

	Validation PSNR (Decibells)	Validation SSIM	Memory (*4 MBs)	Parameters (millions)	Forward Pass Time (seconds)	Depth
Variant-1	5.9	0.552	44.40	9.96	0.735	7
Variant-2	18.6	0.83	30.06	9.96	0.533	7
Variant-3	18.2	0.81	30.06	9.96	0.652	7
Variant-4	18.25	0.81	30.06	9.96	0.654	7
Variant-5	17.9	0.79	30.06	13.9	0.871	9
Variant-6	18.47	0.82	44.40	9.96	0.937	7
Hardis L15	16.06	–	253.34	4.49	0.45	15
L 10 L	14.01	–	263.34	2.47	–	10
L 10 M	14.01	–	216.90	1.86	–	10
L 10 S	13.80	–	180.54	1.36	–	10

Table 2: Experimentation Results Summary: Top 6 rows show our variants derived from the Base Model. Bottom 4 rows show current state of the art methods variants. Note that our method uses the complete image for performance evaluation instead of the central (160×160) patch used by the current state of the art.

tural Similarity Index Measure (SSIM) using whole images, rather than the (160×160) central patch of the image [8]. Since, we believe that in real life setting text can be captured on any part of the image during image capturing process.

After experimentation with different design choices we recommend that pooling operation are a must to reduce computational complexity, instance normalization boosts quick convergence thus making training faster, increasing depth makes CNNs learning long and difficult; better results can be achieved by relatively smaller networks if smartly designed and adding structural component helps CNN to produce sharp deblurred images.

6. Results

Figure 3 shows sample images for Variant-2 which has the highest PSNR of all models. Here we can see that images are clearly deblurred with distinguishable text. However, still the PSNR achieved is low. That is because, if we compare variants’ output images to the original images and, closely observe we can see that the intensity of pixel values is low in variants’ output image as compared to the label images. Therefore, the PSNR achieved can be further improved by some kind of thresholding technique. However,

we argue that the main objective for textual deblurring is to be able to deblur image in such a way that the text is human readable. In the figure 3, it is clearly visible that the the output image image are near identical in terms of textual information. However, it must be noted that the individual letters in a word of a sentence are convoluted to some absurd representation. Through manual inspection of the text, the vowels and some particular letters with similar shapes are often misclassified for example, letter “a”, “e”, “o”, “i”, “t” and “l”. However, still with these inconsistencies the human brain is able to correctly guess the word.

7. Future work

We have evaluated many new architectural choices. However, we believe that this performance can still be further stretched. Addition of skip connections can be one possible way forward to improve results. We are still in the process of evaluating our custom design loss function and trying to find out the right mixture of structural and L_2 loss component, with which we can focus on the deblurring text in the images. We believe that this loss function is the key to solving this problem. Furthermore, it has the potential to becoming a standard loss for deblurring.

Blur Image	Original Image	Output Image
<p>Software", "Interacting with y, VOL 11, No 5.</p> <p>6) "Technology and the Early assroom", Position Statement: d young children- Ages Three Young Children, 15-11,16.</p> <p>7) "Information Technology & Reading, Writing, Science, & Center for Educational Research Paris.</p> <p>Smith (1983) "Learning with Micros, Ellis Harwood fiers, England.</p> <p>2) "Microcomputers in Primary</p>	<p>Software", "Interacting with y, VOL 11, No 5.</p> <p>6) "Technology and the Early assroom", Position Statement: d young children- Ages Three Young Children, 15-11,16.</p> <p>7) "Information Technology & Reading, Writing, Science, & Center for Educational Research Paris.</p> <p>Smith (1983) "Learning with Micros, Ellis Harwood fiers, England.</p> <p>2) "Microcomputers in Primary</p>	<p>Software", "Interacting with y, VOL 11, No 5.</p> <p>6) "Technology and the Early assroom", Position Statement: d young children- Ages Three Young Children, 15-11,16.</p> <p>7) "Information Technology & Reading, Writing, Science, & Center for Educational Research Paris.</p> <p>Smith (1983) "Learning with Micros, Ellis Harwood fiers, England.</p> <p>2) "Microcomputers in Primary</p>
<p>[2] K. Gatermann and M. W systems in chemistry. Ad</p> <p>[3] B. Huber and J. Versche 18(1):91-108, 1998.</p> <p>[4] A.P. Morgan, A.J. Somme to nonlinear analytic sys</p> <p>[5] A.J. Somnese, J. Versc A. Dickenstein and I.Z. Applications, volume 14 2005.</p>	<p>[2] K. Gatermann and M. W systems in chemistry. Ad</p> <p>[3] B. Huber and J. Versche 18(1):91-108, 1998.</p> <p>[4] A.P. Morgan, A.J. Somme to nonlinear analytic sys</p> <p>[5] A.J. Somnese, J. Versc A. Dickenstein and I.Z. Applications, volume 14 2005.</p>	<p>[2] K. Gatermann and M. W systems in chemistry. Ad</p> <p>[3] B. Huber and J. Versche 18(1):91-108, 1998.</p> <p>[4] A.P. Morgan, A.J. Somme to nonlinear analytic sys</p> <p>[5] A.J. Somnese, J. Versc A. Dickenstein and I.Z. Applications, volume 14 2005.</p>
<p>maximum disruption of the overall attack would be inherently some- however, because in order to in- into the network the attacker has substantial resources into the sys- net benefit to the system up to those nodes become malicious. In attacker has effectively amortized his attack by all the good service p to that point. For this reason, attack is likely to be effective only sh it is the timing of the disruption some reason (e.g., if the attack is with some other important event.)</p>	<p>maximum disruption of the overall attack would be inherently some- however, because in order to in- into the network the attacker has substantial resources into the sys- net benefit to the system up to those nodes become malicious. In attacker has effectively amortized his attack by all the good service p to that point. For this reason, attack is likely to be effective only sh it is the timing of the disruption some reason (e.g., if the attack is with some other important event.)</p>	<p>maximum disruption of the overall attack would be inherently some- however, because in order to in- into the network the attacker has substantial resources into the sys- net benefit to the system up to those nodes become malicious. In attacker has effectively amortized his attack by all the good service p to that point. For this reason, attack is likely to be effective only sh it is the timing of the disruption some reason (e.g., if the attack is with some other important event.)</p>
<p>t, which is intuitive pleasing since behavior observed in the three-state fast fading (see the last plot) the opp from error states. This can be explain crease the transmitted power (relat ver is observed at 2.5 to 3.5 dB) -the dynamics of the channel tend to ns, so that right after an error the p her than average. We refer to this cas case of "anticorrelated" errors. Th ite Markov model. be also checked by simple calculation as in this section. In the following F</p>	<p>t, which is intuitive pleasing since behavior observed in the three-state fast fading (see the last plot) the opp from error states. This can be explain crease the transmitted power (relat ver is observed at 2.5 to 3.5 dB) -the dynamics of the channel tend to ns, so that right after an error the p her than average. We refer to this cas case of "anticorrelated" errors. Th ite Markov model. be also checked by simple calculation as in this section. In the following F</p>	<p>t, which is intuitive pleasing since behavior observed in the three-state fast fading (see the last plot) the opp from error states. This can be explain crease the transmitted power (relat ver is observed at 2.5 to 3.5 dB) -the dynamics of the channel tend to ns, so that right after an error the p her than average. We refer to this cas case of "anticorrelated" errors. Th ite Markov model. be also checked by simple calculation as in this section. In the following F</p>
<p>ate pain specialists respectively.</p> <p>Overall, the ti deemed 77.6% and outcomes as more a >5 on a 10-point a combined appropri fuded. Of the ment of pain, d abstraction. dividing the numb prescribed by both therapies prescribe specialists concurr assessment that th inappropriate than five cases in which</p>	<p>ate pain specialists respectively.</p> <p>Overall, the ti deemed 77.6% and outcomes as more a >5 on a 10-point a combined appropri fuded. Of the ment of pain, d abstraction. dividing the numb prescribed by both therapies prescribe specialists concurr assessment that th inappropriate than five cases in which</p>	<p>ate pain specialists respectively.</p> <p>Overall, the ti deemed 77.6% and outcomes as more a >5 on a 10-point a combined appropri fuded. Of the ment of pain, d abstraction. dividing the numb prescribed by both therapies prescribe specialists concurr assessment that th inappropriate than five cases in which</p>

Figure 3: Textual deblurring results for the highest performing variant-2.

As the problem of blind deblurring is very hard as high amount of blur and noise can cause image degradation to such a extent that model fails. Even after achieving good image restoration we still believe that this problem is far from solved and can be helped by guided approach to blind deconvolution. As the blur space we are trying to model is huge so a possible solution can be a guided approach to learning anti-PSF functions. A CNN that can successfully predict anti-PSF function can be given to a non blind method for deblurring. Further, we can also train such a network in conjunction to this network; that minimizes the difference between the original image and output image. By doing so we build a network that is able to fill in the missing information left behind the generative network. By these two ways we believe that further performance gain in image restoration can be achieved.

7.1. Conclusion

We have proposed a CNN model that tries to solve the problem of blind text deblurring. It outperforms the previous state of the art method while remaining resource friendly and computationally fast. Precisely, our this approach takes less time to converge/train, has reduced forward pass time thus enable real time response and reduced memory footprint making it possible to be deployed in small hand held devices.

We have proposed a new custom designed loss function which simplifies learning for textual deblurring and has the potential to become a standard choice loss function in computer vision problems relating to image transfer tasks.

Furthermore, we have demonstrated what effects are exhibited by the new advancements in CNN on this problem of textual deblurring. We have empirically evaluated different design choices to find which are better and how they affect the performance of the architecture in terms of memory consumption, convergence speed and image restoration quality.

References

- [1] A. Chakrabarti. A neural approach to blind motion deblurring. In *European conference on computer vision*, pages 221–235. Springer, 2016.
- [2] W. Chen, Z. Fu, D. Yang, and J. Deng. Single-image depth perception in the wild. *Advances in neural information processing systems*, 29:730–738, 2016.
- [3] 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*, pages 369–376. IEEE, 2011.
- [4] H. Cho, J. Wang, and S. Lee. Text image deblurring using text-specific properties. In *European Conference on Computer Vision*, pages 524–537. Springer, 2012.
- [5] R. Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1440–1448, 2015.
- [6] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):142–158, 2016.
- [7] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *Pattern recognition (icpr), 2010 20th international conference on*, pages 2366–2369. IEEE, 2010.
- [8] M. Hradiš, J. Kotera, P. Zemčík, and F. Šroubek. Convolutional neural networks for direct text deblurring. In *Proceedings of BMVC*, pages 2015–10, 2015.
- [9] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.
- [10] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019.
- [11] Y. Lee and J. Park. Centermask: Real-time anchor-free instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13906–13915, 2020.
- [12] T.-H. Li and K.-S. Lii. A joint estimation approach for two-tone image deblurring by blind deconvolution. *IEEE Transactions on Image Processing*, 11(8):847–858, 2002.
- [13] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [14] X.-J. Mao, C. Shen, and Y.-B. Yang. Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*, 2016.
- [15] S. Nah, T. Hyun Kim, and K. Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017.
- [16] T. M. Nimisha, A. Kumar Singh, and A. N. Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4752–4760, 2017.
- [17] M. Noroozi, P. Chandramouli, and P. Favaro. Motion deblurring in the wild. In *German conference on pattern recognition*, pages 65–77. Springer, 2017.
- [18] 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*, pages 2901–2908, 2014.
- [19] 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*, pages 2901–2908, 2014.
- [20] V. Papakonstantinou and P. De Hert. Legal challenges posed by online aggregation of museum content: The cases of europeana and the google art project. *SCRIPTed*, 9:314, 2012.
- [21] M. Poggi, F. Aleotti, F. Tosi, and S. Mattoccia. Towards real-time unsupervised monocular depth estimation on cpu.

- In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5848–5854. IEEE, 2018.
- [22] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3341–3350, 2020.
 - [23] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.
 - [24] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.
 - [25] D. Wofk, F. Ma, T.-J. Yang, S. Karaman, and V. Sze. Fast-depth: Fast monocular depth estimation on embedded systems. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6101–6108. IEEE, 2019.
 - [26] M. D. Zeiler, G. W. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2018–2025. IEEE, 2011.
 - [27] H. Zhang, Y. Dai, H. Li, and P. Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019.
 - [28] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. Lau, and M.-H. Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2529, 2018.
 - [29] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2737–2746, 2020.