https://doi.org/10.2352/issn.2169-2629.2019.27.65
Reprinted from
Journal of Imaging Science and Technology® 63(4): 040404-1–040404-9, 2019.

Deep Learning Approaches for Whiteboard Image Quality Enhancement

Mekides Assefa Abebe▲ and Jon Yngve Hardeberg▲
The Norwegian Colour and Visual Computing Laboratory; NTNU, Norway
E-mail: mekides123@gmail.com

Abstract. *Different whiteboard image degradations highly reduce the legibility of pen-stroke content as well as the overall quality of the images. Consequently, different researchers addressed the problem through different image enhancement techniques. Most of the state-of-the-art approaches applied common image processing techniques such as background foreground segmentation, text extraction, contrast and color enhancements and white balancing. However, such types of conventional enhancement methods are incapable of recovering severely degraded pen-stroke contents and produce artifacts in the presence of complex pen-stroke illustrations. In order to surmount such problems, the authors have proposed a deep learning based solution. They have contributed a new whiteboard image data set and adopted two deep convolutional neural network architectures for whiteboard image quality enhancement applications. Their different evaluations of the trained models demonstrated their superior performances over the conventional methods. © 2019 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2019.63.4.040404]

1. INTRODUCTION

Whiteboards are known to be very important illustration tools for different types of communications. With the recent and rapid technological advancements, people increasingly communicate over the internet. Individuals currently also prefer to take a picture of whiteboard contents or presentation slides over the conventional hard-copy paper notes. Most people, however, utilize their personal mobile phones, computers and very affordable streaming solutions. Therefore, the quality of their experiences will mostly be limited by the capabilities of such technologies.

Many of the camera technologies, which are affordable and being utilized by many, are known to have limited qualities. The majority of the webcam and videoconferencing cameras, in particular, have inadequate dynamic range, smaller color gamut and lower resolution [1]. The bandwidth and other architectural limitations of the current streaming networks additionally add on the degradations of the streaming content [2, 3]. Usually, it is the whiteboard regions of such content whose qualities are highly degraded. Several factors, such as specular reflections (due to the reflective nature of the whiteboard materials), non-uniform illuminations (because of the unprofessional room lightings), shadows, occlusions, and related others, are known to contribute on the matter [4].

Consequently, over the past years, several whiteboard image enhancement approaches have been proposed by several researchers. The greater number of the proposed solutions were only intended to detect and extract handwritten texts, by filtering and subtracting the respective whiteboard backgrounds [5–8]. Few other researchers, on the other hand, were rather interested in the enhancement of the different quality attributes of all whiteboard image contents [9, 10]. Almost all these methods, with the exception of one recent method considering the naturalness of the whiteboard background appearances [4], set the background of the whiteboard images to white and enhance the color and contrasts of the pen-stroke content.

Many of the state-of-the-art methods mostly create appealing results for their intended applications. However, in the presence of more severe whiteboard image degradations (such as specular reflections, overexposed regions and color tints or white-balancing problems) and pen-stroke contents (like filled polygons or other shapes of illustrations), most methods fail to generate enhanced results and additionally introduce visible artifacts. The methods, also by design, do not have a capacity to restore the degraded or lost pen-stroke content details. However, the majority of the existing whiteboard image and video archives are highly distorted and, hence, more efficient and comprehensive enhancement solutions are required.

In this regard, we have explored enhancement approaches which can inclusively restore almost all types of whiteboard image distortions. From our inspection, we were able to determine that adaptation of some of the deep learning (DL) approaches to whiteboard image enhancement purposes could be the most effective solution. In the past, different DL approaches have been used for many other applications [11–13]. However, up to the time of writing, we did not come across any DL based whiteboard image quality enhancement solutions.

Therefore, in this work, we have proposed two deep convolutional neural network architectures which are adapted from two common image denoising (Fully convolutional and deconvolutional [14]) and image segmentation (UNet [15]) architectures. We have adopted and trained the two architectures and evaluated their performances for whiteboard image quality enhancement applications. The results of the

▲ IS&T Members.

Received Mar. 23, 2019; accepted for publication May 15, 2019; published online Aug. 22, 2019. Associate Editor: Michael Murdoch.
1062-3701/2019/63(4)/040404/9/$25.00

two architectures are also evaluated with respect to the best performing conventional state-of-the-art whiteboard image enhancement algorithms.

Moreover, to train the proposed models, finding suitable whiteboard image data sets was very challenging. Therefore, we have created our own data set by collecting quality whiteboard images through various sources and simulating the different whiteboard image quality degradations. More detailed information on our data set generation processes is given in the coming sections.

In general, the novel contributions of this work contains:

- The generation of a new whiteboard image data set for image quality enhancement purposes,
- The introduction of deep learning approaches for whiteboard image quality enhancement applications, and
- The evaluation of the conventional whiteboard image quality enhancement approaches related to deep learning approaches.

2. RELATED WORK

In this work, we are introducing DL approaches for the application of whiteboard image quality enhancements. Over the years, several other enhancement approaches have been also introduced. Therefore, a brief review of the state-of-the-art on related topics is presented as follows.

2.1 *Whiteboard Image Quality Enhancement Methods*

Researchers from various research communities have been addressing the whiteboard image quality degradation problems, caused by different limitations of the acquisition and streaming technologies [4]. However, instead of enhancements, most of the proposed solutions were designed for text recognition and extraction purposes. Most solutions consisted of different ways of hand written text detection [5], segmentation [6, 7], and whiteboard content classification [8] algorithms. Some are multiple frame based enhancement approaches, which are proposed with the intention of recovering occluded information and removal of redundant data by propagating and discarding information from consecutive frames of whiteboard videos [16, 17].

In case of solutions for whiteboard image quality enhancements, the number of prior methods are very limited [4, 9, 10, 18, 19]. Increasing the visibility of pen-stroke content through different color and contrast enhancement approaches was the main goal for most of these enhancement methods. Some of the methods enhance the color saturation of pen strokes, while processing the background of the whiteboard to be completely white [18, 19]. Others, on the other hand, preserved the natural appearance of the whiteboard images by applying different white-balancing, saturation correction and contrast enhancement techniques [4, 9, 10].

Nevertheless, currently available whiteboard image/video archives consist of very low quality whiteboard contents [20, 21]. The pen-stroke content of such whiteboards are mostly visually undiscernible due to the low image resolution, non-uniform illumination, specular highlights, overexposure, and other related problems [4]. Therefore, in the presence of such severe quality degradations, most of the described enhancement solutions fail to recover important details. The methods mostly use median filter [9, 10, 18, 19] or polynomial surface fitting based techniques [4] to estimate the background color of whiteboard images. We have noticed such techniques producing different processing artifacts, mainly in the presence of more complex pen-stroke content (such as filled polygons or other types of shaped illustrations). Therefore, more holistic and powerful whiteboard image enhancement solutions, with a capability of restoring severely degraded and more complex pen-stroke contents, are highly required.

2.2 *Whiteboard Image Quality Attributes*

As described in the previous section, most state-of-the-art whiteboard image enhancement approaches emphasize on producing more legible whiteboard contents [9, 10, 18, 19]. As a result, most of their results tend to have unnatural whiteboard appearances. Few other approaches, nonetheless, aim to preserve the natural appearance of real world whiteboards [4]. However, up to the time of writing, there is no standard specification of whiteboard image quality, which are expected from the various enhancement approaches. Observers' expectations as well as the relationship among the legibility of pen-stroke content and the naturalness of the resulted whiteboards, with respect to the overall image quality, are not very well investigated.

Related to natural image researches, overall image quality is usually assessed by the two most defining attributes, the usefulness (legibility) and the naturalness [22–25] of images. According to prior natural image studies, increasing the saturation, brightness as well as contrast of image contents, to a certain amount, showed an improved legality and naturalness scores. Increasing the values beyond that point, contrarily, reduced the naturalness of the results. However, in most scenarios, observers preferred more colorful and contrasted images, despite their unnatural appearances [22–24, 26]. On the other hand, manipulations of images' color temperature and hue deteriorate the naturalness as well as the overall image quality [22, 26].

In our application, the most vital contents of whiteboard images are the pen strokes. Hence, in addition to color and contrast attributes, the whiteboard image legibility is also expected to be affected by other text related factors such as, the character spacings, the thickness of the pen strokes, as well as the typefaces and sizes of characters. Related studies for hand/computer written characters on other emissive and reflective medias (very different from whiteboards) showed the not excessively bold or light characters with bigger open counters and easily recognizable shapes to be more legible [27, 28]. Other studies on background–text relations, additionally, showed how positive contrast (emanated from white characters which are written on dark backgrounds) hinders the discernibility of contents than that of the negative

contrast (resulted from dark characters written on white backgrounds) [29, 30].

Concerning whiteboard image quality, we have also recently conducted a series of perceptual studies assessing many whiteboard related quality factors [31]. The effects of several whiteboard image background processing together with color, contrast, and brightness enhancements on the legibility, naturalness, as well as the general quality of whiteboard images have been investigated. The conclusions of our study highly resemble those of the prior natural image studies that we have described in the previous paragraphs. Our findings indicate the enhancement of legibility with increased saturation and contrast of pen-stroke contents. Degradation of naturalness is also observed with most of our evaluated background enhancements. But, when it comes to overall whiteboard image quality, observers tend to prefer more legible pen-stroke content even with the most unnatural looking whiteboard backgrounds. Generally, all these image quality studies showed that there is always a compromise to be made depending on the intended application.

2.3 *Deep Learning for Image Quality Enhancements*

Unlike the other conventional image enhancement techniques, Deep learning (DL) techniques are gaining a lot of attention in the last couple of decades [32, 33]. DL approaches are outperforming state-of-the-art machine learning approaches in many computer vision related applications [34]. The most common DL based computer vision applications include image classification, object recognition and scene understanding, motion tracking, human pose estimation and action recognition, and related others. Similarly, the color imaging community recently have also been using DL techniques for applications like blind image quality evaluation, perceptual modeling and different image enhancement purposes [35–37]. This rapid integration of DL techniques in many research fields is mainly a result of different factors such as, the availability of large data sets, the advancements of parallel GPU computing, the introduction of powerful programming frameworks, and many other algorithmic revolutions [38–40].

Regarding image quality enhancement, the amount of proposed DL based solutions, up to the time of writing, are very limited. The most common enhancement applications of DL techniques that we have come across so far include: dehazing [12, 13], image denoising [14, 41, 42], enhancement of underexposed [43] and low resolution images [37, 44], and white balancing [45–47]. Most of these methods use autoencoders like end-to-end trainable DL architectures, which mainly comprises two connected encoding and decoding networks. The encoding as well as decoding parts of the architectures are, in turn, formed by a combination of two or more Convolutional, Deconvolutional, Pooling, Upsampling, or other regularization layers [42]. Some image denoising methods additionally include skip connections between intermediate encoding and decoding layers, for better preservation of important image details [14].

In addition to the DL architectures, the most common concern among many image quality enhancement methods is the availability of training data. Finding high quality ground truth images for many image degradation problems is very challenging. In consequence, most DL based enhancement techniques generate and use synthetic data sets for the training and evaluation of their DL networks [41, 43]. Even so, in some color imaging problems the modeling and simulation of common degradations could be even more challenging. In such cases, real acquisitions of different degradations and ideal scenarios or more complex unsupervised DL techniques are required. Moreover, due to processor speed and memory limitations, most DL solutions reduce the spatial resolutions of their data set images to be less than or equal to 256×256. This restriction, however, will have a negative effect on the end results of image enhancement applications of fine contents (such as whiteboards).

So far, we were not able to find any DL based approaches designed particularly for whiteboard image enhancement purposes. Our search for any open synthetic or real whiteboard image data sets for quality enhancement purposes was not successful either. The only whiteboard image data set that we could find was the whiteboard image class of the open image data set [20, 21]. However, the open image data set is prepared for object recognition applications with only 1000 degraded whiteboard images (with no ground truth undegraded images) and their labeled bounding boxes.

Therefore, for a complete and more accurate whiteboard image enhancement applications, a new data set which can completely represent the real whiteboard image quality degradation problems is very essential. Along with the data set, for a better-quality enhancement as well as restoration and preservation of pen-stroke content, an efficient DL model needs to be introduced.

3. METHOD

As it is explained in the previous sections, different conventional approaches of the state-of-the-art whiteboard image quality enhancement are unable to recover highly degraded whiteboard image contents. The different DL techniques, which have been effectively applied for various other types of image enhancement applications, have also been briefly stated. However, up to this time, the problem of enhancing severely degraded whiteboard images is not properly addressed. Consequently, we have proposed DL approaches for efficient enhancement of such whiteboard images. For the implementation, we have created our own data set and adopted two proper deep convolutional neural network architectures.

3.1 *Data Set Generation*

The accuracy of any DL model is known to be determined by the underlying data set. For reasonable approximation of the unknown input to output mapping function, the amount and types of the data set content need to be well distributed. Related to whiteboard image quality enhancement applications, this type of well-organized data set is not currently

available. Therefore, for representing most whiteboard image degradation scenarios, we have generated a novel whiteboard image data set.

To create the data set, we have first collected high quality whiteboard images from different sources and synthetically simulated different quality degradations. Our attempt for the acquisition of high quality whiteboard images through professional camera and studio setup was unsuccessful due to our limited access to powerful and diffuse light sources (which can uniformly illuminate different sizes of whiteboard surfaces). Also, the content distributions of images, created only in a single studio room and with a small set of hand writing styles, would not be an accurate representative of most real whiteboard image archives. Instead, we have collected our ground truth images through the internet, our colleagues, and friends.

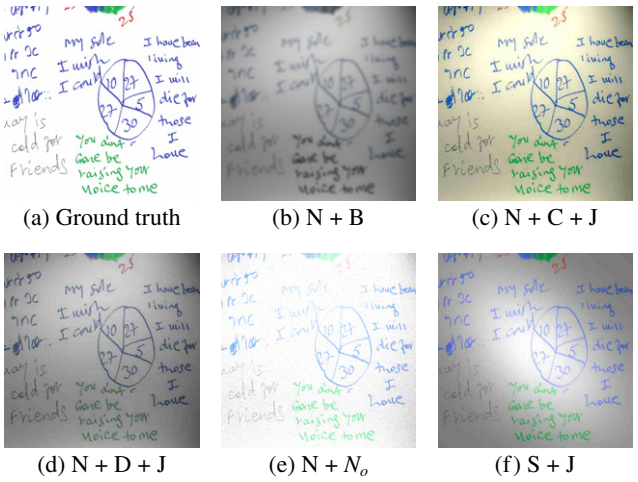
However, we needed to set whiteboard image quality criteria for the selection of the ground truth images. To that end, we have followed our previous study of perceptual whiteboard image quality attributes [31]. According to our study, observers mostly choose whiteboard images which have very white, uniform, and bright backgrounds together with more saturated and highly contrasted pen-stroke content. Also, when evaluating overall whiteboard image quality, observers showed leaning preferences toward content legibility rather than the naturalness of the whiteboard backgrounds. Meanwhile, we have noticed strong resemblance of the most preferred whiteboard images, from our previous study [31], to the whiteboard images created from interactive electronic whiteboards, tablets, and laptop computers. We have found the electronically created whiteboard images to be the most perceptually legible. For this reason, we have decided to collect and use such types of whiteboard images as our ground truth data set and, up to the time of writing, we managed to collect 340 images of resolution 512×512.

The collected ground truth images are then synthetically degraded to simulate the most common whiteboard image degradations that we had encountered in our past experiences [4, 31]. In total, various combinations of the following seven forms of quality degradations are applied on each ground truth image (Figure 1a). Sample images of the different degradations, from our data set, are given in Fig. 1.

1. *Noise*: For representing the different noise artifacts of low quality whiteboard images, we have added a combination of Gaussian and Salt and pepper noises. We have used the mean value of 0.5 and variance of 0.01 for the Gaussian white noise. Whereas, the noise density value of 0.02 is used for the Salt and pepper noises, which will approximately affect 20% of the image pixels (Fig. 1e).
2. *Non-uniform illumination:* Also, to create more realistic non-uniform illuminations, we have chosen to extract and apply the illuminations of real whiteboard images (Fig. 1b). The illumination approximation of the non-uniformly illuminated real whiteboard images L_{ap} is computed as $L_{ap} = B_y \times S_p \times B_x$. The discrete orthogonal basis functions of B_y,S_p, and B_x are in turn computed from the luminance values L_{real} of the source image I_{real} according to Matthew et al. [48]. Finally, the luminance channel of our simulated image L_{sim} is created from that of the ground truth image L_g as $L_{sim} = L_g - \max(L_g) + L_{ap}$.
3. *Gaussian Blur:* We have also simulated the common blurring artifact of different imaging devices by applying Gaussian blur [49]. We have filtered the images with a 3D Gaussian smoothing kernel of size $2 \times (2 \times \sigma) + 1$ and standard deviation of $\sigma = 3$ (Fig. 1b).
4. *Image Compression:* Often most streaming contents are highly compressed due to low network bandwidth and faster transmission purposes [3]. As a result, compression artifacts are very common among video-conferencing whiteboards. To represent this issue in our data set, we applied JPEG compressions with three different quality levels (20%, 50%, and 100%) [50].
5. *Specular highlights:* Moreover, the luminance values of the data set images are further modified, using Eq. (1), for generating different specular highlights. The mean μ and variance ν parameters of the multivariate normal probability density function, Eq. (1), were varied to generate highlights at different spatial positions and with different extents (Fig. 1f).

$$L_{sim} = \frac{1}{\sqrt{|\mu|(2\pi)^2}} e^{-0.5(L_g-\nu)\mu^{-1}(L_g-\nu)} \qquad (1)$$

6. *Color desaturation:* Considering many limitations of conventional digital cameras, we have additionally included one representation of their color clipping and color desaturation issues. The representative degraded images (Fig. 1d) are created by reducing the saturation values of the corresponding ground truth images with various scaling factors. For this work, we have only included images with 40% saturation reduction.



(a) Ground truth (b) N + B (c) N + C + J

(d) N + D + J (e) N + N_o (f) S + J

Figure 1. Example data set images for demonstrating the different whiteboard image degradations, applied on each ground truth image. The abbreviations N, B, C, J, D, S, and N_o stand for the types of applied degradations, which are Non-uniform illumination, Blurring, Color tint, JPEG compression, Desaturation, Specularity, and Noise, respectively.

7. *Color tint:* The other common whiteboard image quality degradation source is the white-balancing steps of the acquisition processes. Whiteboard images captured with the wrong or without white balancing, usually, appear to have color tints similar to the illumination colors (Fig. 1c). Therefore, to incorporate such effects, we have first estimated the illuminants A_g of our ground truth images I_g following the principal component analysis method [51]. The simulated images with different color tints I_{sim} are then created using different colored illuminants A_{new}, $I_{sim} = (I_g/A_g) \times A_{new}$. Totally, four color tints (Reddish, Greenish, Bluish, and Yellowish) are included in our data set.

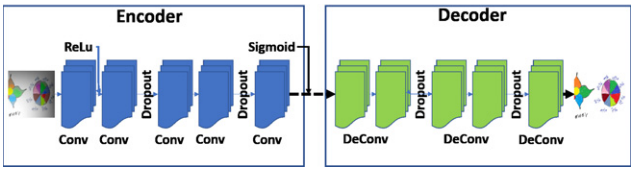
In general, around 43 combinations of these quality degradation are used on each quality image of our data set (a combination of noise and 2 non-uniform illuminations, 2 blurred non-uniform illuminations, 3 types of specular highlights each compressed with 3 JPEG compression levels, color desaturated 2 non-uniform illuminations with three levels JPEG compression, and 2 non-uniform illuminations combined with the 4 types of color tints and 3 JPEG compression levels). In total, our current generated data set contains 14620 pairs of low quality (degraded) and high quality (ground truth) whiteboard images of 512 × 512 resolutions. The data set is further divided into training (11870 pairs), validation (1820 pairs), and test (699 pairs) sets of images for training and evaluation purposes.

3.2 *Architectures of the Proposed Convolutional Neural Networks*

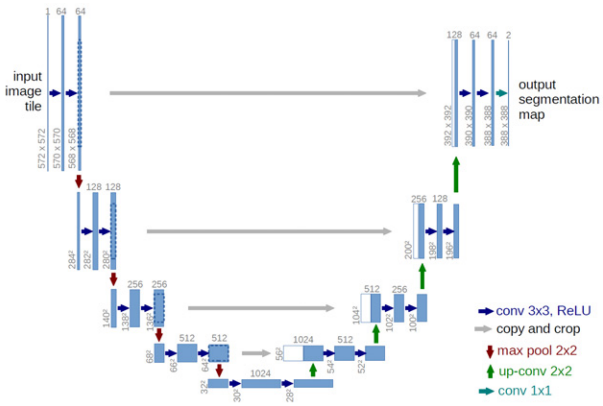
Most of the DL based image enhancement methods, described in the prior sections, use different Autoencoder like DL architectures for more effective end-to-end and supervised learning. We also believe, with our new data set, that the same type of architectures could be effectively adopted to the applications of whiteboard image quality enhancements. Therefore, in this work, we have created and thoroughly evaluated two DL architectures by following methods like Mao et al. [14] and Ronneberger et al. [15].

The frameworks of the two evaluated architectures, shown in Figure 2, contain two connected encoding and decoding networks. The encoder network is mainly used to map the input into its latent representation, by applying different sets of convolutional, downsampling, and other regularization operations. The decoder networks, on the other hand, is used to reconstruct the latent representation back to the enhanced form of the input data [42]. Similar to the state-of-the-art DL networks, the kernel sizes of all our convolution and deconvolution layers are set to be 3 × 3 for their proven excellent performances.

Most image denoising DL algorithms follow some form of fully convolutional and deconvolutional architectures (with no downsampling and upsampling operations), Fig. 2(a). Most methods prefer such architectures for their capabilities of important image detail preservation. It is said that application of different downsampling and upsampling operations usually result in important information loss [14].



(a) Full convolution deconvolution architecture



(b) UNet architecture (taken from [15])

Figure 2. The proposed deep convolutional neural network architectures.

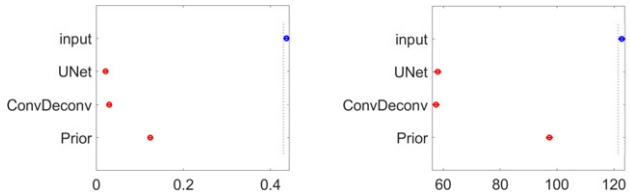
However, such types of DL networks are also known for their high computational costs. Therefore, we have added another efficient and alternative architecture, shown in Fig. 2(b). The architecture is created by adopting the UNet segmentation model [15] and replacing their segmentation maps by our ground truth images. Unlike the fully convolutional and deconvolutional architecture, the UNet encoding and decoding networks include several downsampling and upsampling layers, together with many other convolutional operations. The down and upsampling operations mostly help for avoiding overfitting and reducing the overall computational cost. However, to compensate for the information losses due to the downsampling operations, some skip connections (among the corresponding encoder and decoder layers) are required. Skip connections, once in every couple of hidden layers, are valuable in recovering important image information (by forward information passing) and finding local minimum (by backward gradient passing).

3.3 *Training Procedures*

The above-mentioned networks are implemented, trained, and evaluated using tensorflow and Keras frameworks. For learning the end-to-end mapping from the degraded whiteboard images to the ground truth original images, the weights of the individual architectures are trained on our proposed data set. The encoder and decoder networks of both architectures are trained by minimizing the cross-entropy (log) loss of the network predictions from the ground truth images. Among different tested learning parameters, the moving window Adadelta optimizer (of all its parameters set to default) produces better results for our two proposed architectures.

Table I. Model evaluations of the two proposed architectures.

Models	Unet		Fully Conv. And Deconv.	
Data set	Loss	RMSE	Loss	RMSE
Test	0.0400	0.00046	0.0423	0.00084
Training	0.0407	0.00044	0.0433	0.00088
Validation	0.0429	0.00046	0.0428	0.00088
Training time (1 epoch)	286 sec.		1266 sec.	
Converges after	303 epochs		145 epochs	



(a) iCID Evaluation (b) S-CIELAB Evaluation

Figure 3. ANOVA multiple comparison results of the mean strengths of the iCID and S-CIELAB metrics results.

We have trained our networks with early stopping cross-validation mechanism, on Linux computer with GeForce RTX 2080 GPU. As it can be seen from the summary information (provided in Table I) of our training processes of the fully convolutional–deconvolutional and the UNet models, the processes terminated after 145 and 303 epochs, respectively. We have stopped the training of the two models by setting the baseline validation loss criteria of the early stopping algorithm to a value of 0.043, which leads to high quality enhanced whiteboard images.

4. EVALUATION AND DISCUSSION

After selecting the best performing training mechanisms and their corresponding parameters, described in the previous paragraphs, our models were finally trained with the early stopping criteria of baseline validation loss set to 0.043. The final training loss and accuracy values of the two models along with the summary of their execution time are given in Table. I. Both models were able to achieve the termination criteria at the very different number of epochs and with high variability of execution speeds. This type of learning difference is understandable due to different hidden layer operations that constructed the two models.

The presented fully convolutional and deconvolutional model have a total of 376, 326 trainable parameters and takes around 1266 seconds for training a single epoch of our training data set. The UNet model, on the other hand, only takes around 286 seconds with its 1,962,659 total trainable parameters, for the same amount of epoch training. The higher computational time of the fully convolutional and deconvolutional model is due to smaller batch sizes we used for the training. The model's higher memory requirement forced us to reduce the batch size to only 5 images (than the 16 batch size used during the training process of the UNet model).

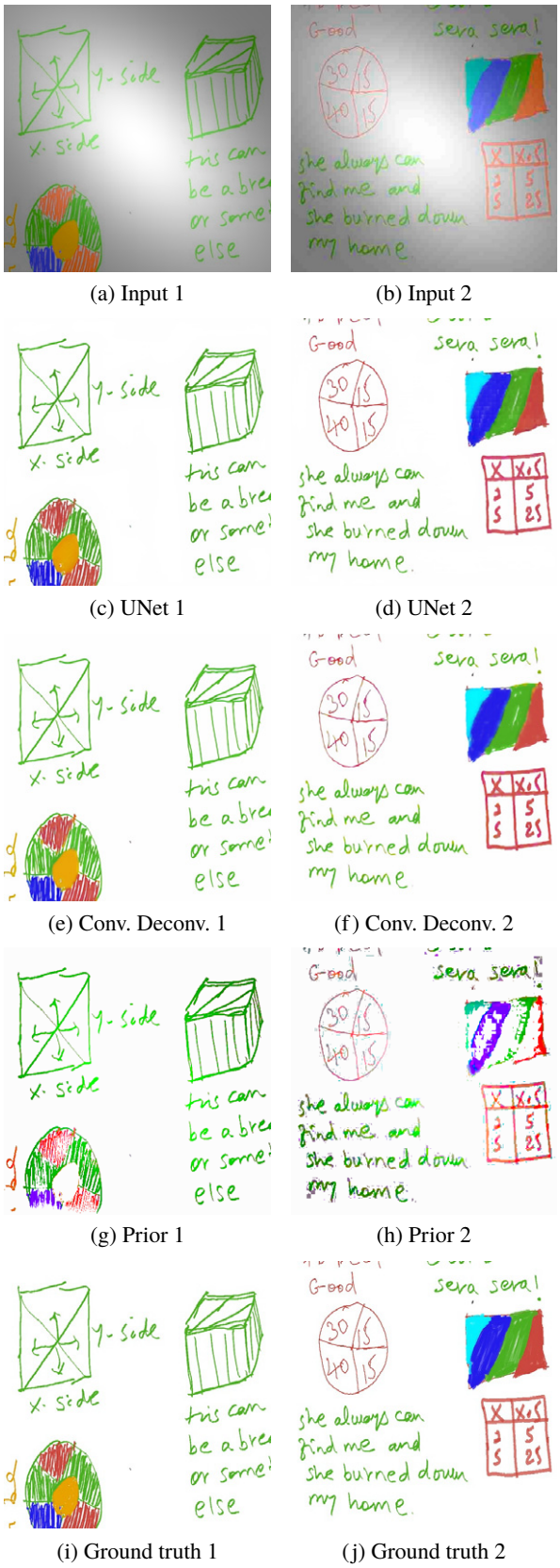
To assess the generalization and predictive capabilities of both models, we have additionally evaluated the computed loss and root mean square error values of the final best models (Table I) on the three sets of our data set. In all cases, our trained UNet model shows higher accuracies.

The evaluation results demonstrate that using networks like the evaluated fully convolutional and deconvolutional model will help increase the accuracy of the models' enhancement results. However, deeper versions of such DL networks may lead to reduced performances on unseen whiteboard images due to overfitting. The very high memory space complexity of such architectures is also a critical issue. To elevate these and other related problems, UNet like architectures with several various sized downsampling and upsampling operations are recommended. As shown in our evaluation results, the UNet model greatly improved the computational complexity of the model with higher accuracy values. The possible information loss, because of the downsampling process, can be effectively reduced through various skip connections, as shown in Fig. 2(b). Higher number of skip connections mostly helps to retain most of the original information.

In addition to our model evaluations, we have furthermore verified the proposed models' whiteboard image quality enhancement capabilities, in comparison to the conventional models. To represent most of the conventional methods, we have selected the best performing whiteboard enhancement method of our prior whiteboard image quality

Table II. Full reference image quality evaluations of the enhancement methods with our test set images. The ground truth test set images of the generated data set were used as our quality references.

Methods	SSIM		CID		iCID		S-CIELAB	
	mean	std.	mean	std.	mean	std.	mean	std.
Input	0.6919	0.1564	0.5608	0.1026	0.4368	0.1517	122.7556	9.4663
Prior method	0.9012	0.1341	0.1268	0.1322	0.1237	0.1228	97.3365	13.6304
Conv. Deconv.	0.9881	0.0116	0.0216	0.0311	0.0292	0.0278	**57.4867**	27.4462
UNet	**0.9914**	0.0086	**0.0125**	0.0194	**0.0208**	0.0206	58.0957	17.0345



(a) Input 1 (b) Input 2

(c) UNet 1 (d) UNet 2

(e) Conv. Deconv. 1 (f) Conv. Deconv. 2

(g) Prior 1 (h) Prior 2

(i) Ground truth 1 (j) Ground truth 2

Figure 4. Sample enhancement results of the evaluated methods for the test set images of our data set.

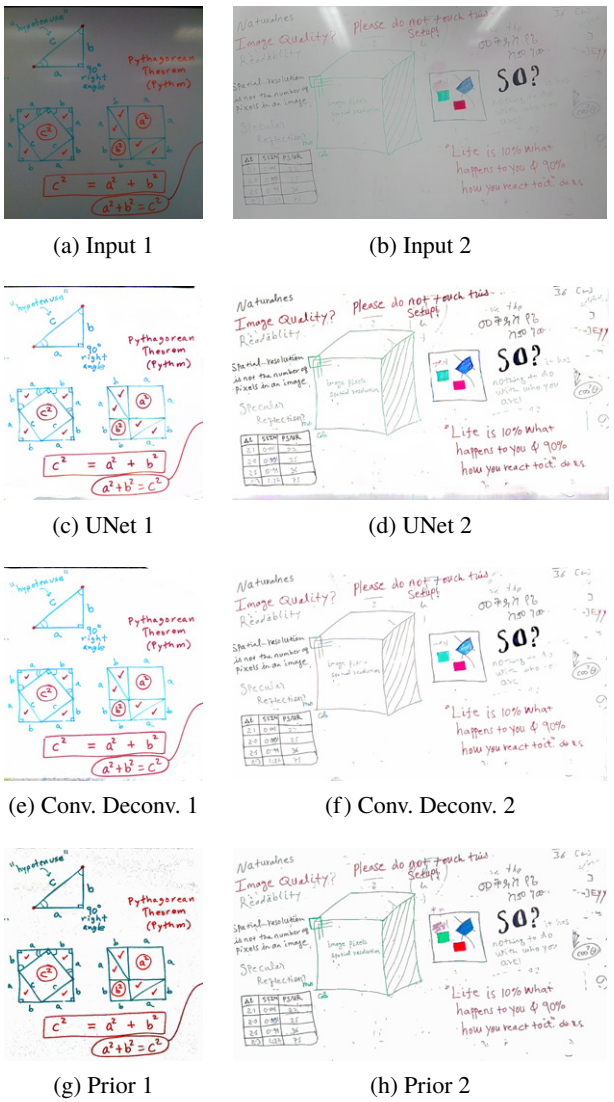
evaluation [31]. In total, we have assessed the results of three whiteboard image quality enhancement methods, the Prior method [31], Fully convolutional and deconvolutional (which we abbreviate as Conv. Deconv.), and the UNet models. The comparison is performed on the 699 test images of our data set (which are never used to train the two models) as well as other real whiteboard images.

The enhancement results of the three methods, for all 699 degraded test images, are first generated and their quality differences from the ground truth images are calculated. We have used the structural similarity index metric (SSIM), S-CIELab color difference, color-image-difference (CID), and improved color-image-difference (iCID) image quality metrics for the evaluation of the structural as well as spatial color similarities of the enhanced and reference images [52–54]. The averages as well as the standard deviations of the image quality differences of our evaluation results are provided in Table II. We have also presented, in Figure 4, two sample image results of the evaluated methods (from the 699 test images), for additional visual assessment purposes.

The image quality results (given in Table II), show the higher performances of the proposed UNet and Conv. Deconv. based models than the prior enhancement method. The two DL based models, in all the computed image quality metrics, gives better average quality results with much lower variances among the 699 test images. Our additional analysis of variances among the evaluated methods, through one-way ANOVA computations, resulted in a p-value $= 0$ for all considered image quality metrics. The statistical significance of the two DL based models' improvements over the degraded images as well as the prior conventional enhancement method results are also clearly visible in the multiple comparison plots of the mean image quality values, given in Figure 3.

For further visual assessment purposes, we have additionally provided example test set (Fig. 4) and captured real world whiteboard (Figure 5) image results. One can see, from results like Fig. 5, and Fig. 4, that the DL based models were able to restore more complex and filled pen-stroke content which were severely degraded. The conventional prior enhancement methods, on the other hand, failed to accurately recover such types of whiteboard contents. Also, most pen strokes which are located under specular highlight regions or very noisy backgrounds were more accurately and legibly recovered by our DL based methods than the conventional prior method. The two models, similarly, surpasses the conventional method in terms of color restoration accuracy.

In general, the two proposed DL approaches showed very promising results in terms of recovering severely degraded whiteboard image contents. Pen-stroke contents under specular regions, noise, compression, or other difficult degradation conditions are made to be effectively recovered. The processing artifact problems of prior enhancement methods, discussed in the related work sections, are also eliminated with the proposed DL techniques.



(a) Input 1 (b) Input 2

(c) UNet 1 (d) UNet 2

(e) Conv. Deconv. 1 (f) Conv. Deconv. 2

(g) Prior 1 (h) Prior 2

Figure 5. Sample enhancement results of the evaluated methods on real whiteboard images.

5. FUTURE WORK

For most DL based approaches, it is always preferable to have as many data set images as possible. We also do not believe the previously presented combinations of the seven quality degradations to completely represent all scenarios of existing whiteboard image quality problems. For example, current trained models do not effectively handle whiteboard images with motion degradations and different translational and rotational motion displacements of pen-stroke contents need to be considered in the future. Therefore, we will be regularly adding more ground truth images as well as new forms of whiteboard image degradation simulations.

It should also be noted that setting the early stopping criteria of the training process to lower values and increasing the resolution of the data set images could lead to more accurate models. With future access to a more powerful GPU computers, it will be highly recommended to train the presented models with much lower minimum loss criteria as well as higher resolution image data set.

6. CONCLUSION

In this work, we have considered the application of deep learning approaches for whiteboard image quality enhancement applications. The different problems of the conventional whiteboard image enhancement techniques were targeted to be resolved. Particularly, we have proposed two deep convolutional neural network architectures with our new generated whiteboard image data set for degraded whiteboard image enhancement purposes. Our overall training and evaluation results of the proposed deep learning models showed us that shallower DL architectures, with some downsampling and upsampling operations and additionally compensated by skip connections, will immensely improve the models' efficiency as well as quality enhancement performances. The proposed deep learning models, trained on our currently available data set, showed superior performances, specially for whiteboard regions with more complex pen-stroke contents and severe degradations. We also believe that, with increased resolution and number of data set images and more longer training time, the current performances of the proposed model can be improved much further.

REFERENCES

[1] M. A. Abebe, "*Perceptual Content and Tone Adaptation for HDR Display Technologies*," Ph.D. dissertation (Université de Poitiers, 2016).

[2] M. Shahid, M. A. Abebe, and J. Y. Hardeberg, "Assessing the quality of videoconferencing: from quality of service to quality of communication," *IS&T Electronic Imaging: Image Quality and System Performance XV* (IS&T, Springfield, VA, 2018), pp. 1–7.

[3] B. Belmudez, *Audiovisual Quality Assessment and Prediction for Videotelephony*, 1st ed. (Springer International Publishing, Switzerland, 2015).

[4] C. A. A. Duque, M. A. Abebe, M. Shahid, and J. Y. Hardeberg, "Color and quality enhancement of videoconferencing whiteboards," *IS&T Electronic Imaging: Color Imaging XXIII* (IS&T, Springfield, VA, 2018), pp. 1–13.

[5] H. Lu and M. Kowalkiewicz, "Text segmentation in unconstrained hand-drawings in whiteboard photos," *Int'l. Conf. on Digital Image Computing Techniques and Applications (DICTA)* (IEEE, Piscataway, NJ, 2012), pp. 1–6.

[6] T. Plotz, C. Thurau, and G. A. Fink, "Camera-based whiteboard reading: New approaches to a challenging task," *Int'l. Conf. on Frontiers in Handwriting Recognition* (Newcastle Univesrsity, UK, 2008), pp. 385–390.

[7] M. Wienecke, G. A. Fink, and G. Sagerer, "Towards automatic video-based whiteboard reading," *Proc. 7th Int'l. Conf. on Document Analysis and Recognition, 2003* (IEEE, Piscataway, NJ, 2003), pp. 87–91.

[8] S. Vajda, L. Rothacker, and G. A. Fink, *A Method for Camera-Based Interactive Whiteboard Reading* (Springer, Berlin, Heidelberg, 2012), pp. 112–125.

[9] Z. Zhang and L.-W. He, "Whiteboard scanning and image enhancement," Digit. Signal Process. **17**, 414–432 (2007).

[10] M. Gormish, B. Erol, D. G. Van Olst, T. Li, and A. Mariotti, "Whiteboard sharing: capture, process, and print or email," Proc. SPIE **7879**, 9 (2011).

[11] R. Shanmugamani, *Deep Learning for Computer Vision* (Packt Publishing, Birmingham, UK, 2018).

[12] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: an end-to-end system for single image haze removal," IEEE Trans. Image Process. **25**, 5187–5198 (2016).

[13] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2017).

[14] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *Adv. Neural Information Process. Syst.* (Curran Associates Inc., Red Hook, NY, 2016), pp. 2802–2810.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597.

[16] P. E. Dickson, W. R. Adrion, and A. R. Hanson, "Whiteboard content extraction and analysis for the classroom environment," *10th IEEE Int'l. Symposium on Multimedia* (IEEE, Piscataway, NJ, 2008), pp. 702–707.

[17] P. E. Dickson, C. Kondrat, W. R. Adrion, T. D. Richards, and R. B. Szeto, "Improved whiteboard processing for lecture capture," *IEEE Int'l. Symposium on Multimedia (ISM)* (IEEE, Piscataway, NJ, 2016), pp. 649–654.

[18] L. W. He, Z. Liu, and Z. Zhang, "Why take notes? Use the whiteboard capture system," Technical Report MSR-TR-2002-89, Microsoft Research (September, 2002), [Online; accessed 12-April-2017].

[19] L. W. He and Z. Zhang, "Real-time whiteboard capture and processing using a video camera for remote collaboration," IEEE Trans. Multimedia **9**, 198–206 (2007).

[20] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, T. Duerig, and V. Ferrari, "The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale," arXiv:181100982, 2018.

[21] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, S. Kamali, M. Malloci, J. Pont-Tuset, A. Veit, S. Belongie, V. Gomes, A. Gupta, C. Sun, G. Chechik, D. Cai, Z. Feng, D. Narayanan, and K. Murphy, "Openimages: A public dataset for large-scale multi-label and multi-class image classification," *Dataset available from* https://storage.googleapis.com/openimages/web/index.html, 2017.

[22] T. Janssen and F. Blommaert, "Predicting the usefulness and naturalness of color reproductions," J. Imaging Sci. Technol. **44**, 93–104 (2000).

[23] A. Kuijsters, W. A. IJsselsteijn, M. T. M. Lambooij, and I. Heynderickx, "Influence of chroma variations on naturalness and image quality of stereoscopic images," *IS&T's Electronic Imaging: Human Vision and Electronic Imaging XIV,* (IS&T, Springfield, VA, 2009), p. 72401.

[24] H. de Ridder, "Naturalness and image quality: saturation and lightness variation in color images of natural scenes," J. Imaging Sci. Technol. **40**, 487–493 (1996).

[25] R. Halonen, S. Westman, and P. Oittinen, "Naturalness and interestingness of test images for visual quality evaluation," Proc. SPIE **7867**, 12 (2011).

[26] H. de Ridder, F. J. Blommaert, and E. A. Fedorovskaya, "Naturalness and image quality: chroma and hue variation in color images of natural scenes," Proc. SPIE **2411**, 11 (1995).

[27] A. Haley, "Its about legibility," (Monotype Imaging Inc, Woburn, MA, 2017).

[28] T. Zlokazova and I. Burmistrov, "Perceived legibility and aesthetic pleasingness of light and ultralight fonts," *Proc. European Conf. on Cognitive Ergonomics 2017* (ACM, New York, NY, 2017), pp. 191–194.

[29] L. Rello and J. P. Bigham, "Good background colors for readers: A study of people with and without dyslexia," *Proc. 19th Int'l. ACM SIGACCESS Conf. on Computers and Accessibility* (ACM, New York, NY, 2017), pp. 72–80. Available: http://doi.acm.org/10.1145/3132525.3132546.

[30] M. Grozdanovic, D. Marjanovic, G. L. Janackovic, and M. Djordjevic, "The impact of character/background colour combinations and exposition on character legibility and readability on video display units," Trans. Inst. Meas. Control **39**, 1454–1465 (2017).

[31] M. Abebe and J. Y. Hardeberg, "Evaluation of naturalness and readability of whiteboard image enhancements," *IS&T Electronic Imaging: Color Imaging XXIV: Displaying, Processing, Hardcopy, and Applications* (IS&T, Springfield, VA, 2019).

[32] L. Deng and D. Yu, "Deep learning: methods and applications," Found. Trends Signal Process. **7**, 197–387 (2014).

[33] G. Anthes, "Deep learning comes of age," Commun. ACM **56**, 13–15 (2013).

[34] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: a brief review," Comput. Intell. Neurosci. **2018**, 13 (2018).

[35] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," IEEE Trans. Neural Netw. Learning Systems **26**, 1275–1286 (2015).

[36] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," IEEE Signal Process. Mag. **34**, 130–141 (2017).

[37] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," ACM Trans. Graph. **36** (2017).

[38] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: a system for large-scale machine learning," *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (USENIX Association, Savannah, GA, 2016), pp. 265–283.

[39] X. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," IEEE Access **2**, 514–525 (2014).

[40] D. Kirk, "NVIDIA cuda software and gpu parallel computing architecture," *Proc. 6th Int'l. Symposium on Memory Management* (ACM, New York, NY, 2007), pp. 103–104.

[41] 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. **26**, 3142–3155 (2017).

[42] 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," J. Machine Learning Res. **11**, 3371–3408 (2010).

[43] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," Pattern Recognit. **61**, 650–662 (2017).

[44] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295–307 (2016).

[45] Z. Lou, T. Gevers, N. Hu, and M. P. Lucassen, "Color constancy by deep learning," *BMVC* (BMVA Press, Durham, UK, 2015), pp. 76.1–76.12.

[46] S. W. Oh and S. J. Kim, "Approaching the computational color constancy as a classification problem through deep learning," Pattern Recognit. **61**, 405–416 (2017).

[47] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," *European Conf. on Computer Vision* (Springer, Cham, Switzerland, 2016), pp. 371–387.

[48] M. Harker and P. OLeary, "Discrete orthogonal polynomial toolbox: DOPBox version 1.8," Mar. 2014. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/41250-discrete-orthogonal-polynomial-toolbox-dopbox-version-1-8?focused=962ed076-fe6b-7f3c-0d5e-d37913c6545c&tab=example.

[49] S. Kapur, *Computer Vision with Python* (Packt Publishing, Birmingham, 2017), Vol. 3, Chapter 2: Filters and Features, p. 206.

[50] D. Taubman and M. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice* (Springer Publishing Company, Incorporated, 2013).

[51] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," J. Opt. Soc. Am. A **31**, 1049–1058 (2014).

[52] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker, "Image-difference prediction: From grayscale to color," IEEE Trans. Image Process. **22**, 435–446 (2013).

[53] J. Preiss, F. Fernandes, and P. Urban, "Color-image quality assessment: from prediction to optimization," IEEE Trans. Image Process. **23**, 1366–1378 (2014).

[54] M. Pedersen and J. Y. Hardeberg, *Full-Reference Image Quality Metrics (Foundations and Trends(r) in Computer Graphics and Vision)* (Now Publishers Inc, Boston, 2012).