Multimedia Tools and Applications (2020) 79:15397–15415
https://doi.org/10.1007/s11042-019-7397-7

Microscopic image super resolution using deep convolutional neural networks

Selen Ayas[1] · **Murat Ekinci**[1]

Received: 1 October 2018 / Revised: 17 January 2019 / Accepted: 22 February 2019 /
Published online: 9 March 2019

Abstract
Recently, deep convolutional neural networks (CNNs) have achieved excellent results in single image super resolution (SISR). Owing to the strength of deep CNNs, it gives promising results compared to state-of-the-art learning based models on natural images. Therefore, deep CNNs techniques have also been successfully applied to medical images to obtain better quality images. In this study, we present the first multi-scale deep CNNs capable of SISR for low resolution (LR) microscopic images. To achieve the difficulty of training deep CNNs, residual learning scheme is adopted where the residuals are explicitly supervised by the difference between the high resolution (HR) and the LR images and HR image is reconstructed by adding the lost details into the LR image. Furthermore, gradient clipping is used to avoid gradient explosions with high learning rates. Unlike the deep CNNs based SISR on natural images where the corresponding LR images are obtained by blurring and subsampling HR images, the proposed deep CNNs approach is tested using thin smear blood samples that are imaged at lower objective lenses and the performance is compared with the HR images taken at higher objective lenses. Extensive evaluations show that the superior performance on SISR for microscopic images is obtained using the proposed approach.

Keywords Microscopic imaging · Super resolution · Deep residual learning · Convolutional neural networks

1 Introduction

Microscopy of clinical specimens is a simple, rapid and inexpensive method for examining evidence of an infection and making a presumptive diagnosis of certain infectious disease. Moreover, early and accurate diagnosis generally increases the chance of appropriate treatment as early as possible. The World Health Organization (WHO) recommends the diagnosis of certain infectious diseases caused by parasites by microscopy [27, 28].

The gold standard method in clinical microbiology of certain infectious disease is to examine a stained smear sample under a microscope to determine the presence of parasites.

✉ Selen Ayas
selenguven@ktu.edu.tr

[1] Department of Computer Engineering, Karadeniz Technical University, 61080, Trabzon, Turkey

Therefore, visual microscopic examination is generally considered as the reference method in parasite analysis. However, the method has defects in terms of sensitivity and specificity for its use as a diagnostic tool because it is labor-intensive and time-consuming Moreover, it requires high workload.

Computer-aided microscopic image analysis has attracted more attention in microbiology and image processing. Recently, numerous efforts have been made to help clinicians to determine the presence of parasites [1, 15]. To analyze the microbiological structures like parasites, tissue and artifacts in microbiological images, the images acquired at various magnification level are first preprocessed to eliminate the effect of noise or enhance the image quality, and then segmented and classified as parasites in recent studies.

In this work, we propose an approach to enhance spatial resolution of microscopic systems without changing their design. The smear samples are imaged with low resolution (LR) and the image with better resolution is acquired as an output. By using LR images, we avoid spending extra effort and time to examine the slides at higher magnification factors. In the proposed study, we target thin smear blood samples for identification of *Plasmodium Falciparum* imaged with LR in order to significantly improve the resolution of the system using single image super resolution (SISR) which is widely used in many fields for enhancing the resolution of sensing systems.

1.1 Related work

The SISR aims to reconstruct a high resolution (HR) image from a given LR image by inferring all missing high frequency details. Due to super resolution (SR) is inherently ill-posed inverse problem with insufficient knowledge, the solution with the reconstruction constraint is not unique, i.e. there are multiple solutions for any given LR image. Several methods have been proposed to address the ill-posed problem.

Traditional methods include interpolation such as bilinear or bicubic [16], Lanczos resampling [7] and statistical based methods -edge-based [34] and oriented adaptive [21, 23, 42, 48] interpolations. Reconstruction based methods utilize various priors and constraints to regularize the SR problem including the iterative back projection (IBP) [13], maximum a posteriori probability (MAP) [33], projections onto convex sets (POCS) [2]. Although these methods are simple yet fast and have been studied comprehensively, the super-resolved image has usually blurring and aliasing artifacts in high frequency details, i.e. in edges and textures, and the quality of super-resolved image is unsatisfactory. Recently, learning based methods achieve great success by learning a mapping from LR to HR image patches when compared to early methods. The methods based on neighbor embedding [3, 4], sparse-coding [46, 47, 49], anchored neighborhood regression (ANR) [38] which is the combination of neighbor embedding and sparse coding, the improved version of ANR, namely A+ [39], and deep convolutional neural networks (CNNs) [5, 6, 17–19] are performed excellently in image SR.

1.2 Motivations and contributions

With the development of deep CNNs, deep CNNs based SISR has been shown to increase performance compared to previous learning based models, where deep CNNs directly learn end-to-end mapping between LR images and HR images. However, studies applying deep CNNs to medical imaging applications show poor performance, compared to natural images [24]. The main reason why these models have been performed poorly is because that commonly applied data augmentation techniques are not applied to medical images in order to

retain medical image statistics. Several studies have been carried out to investigate the best augmentation technique for capturing medical image statistics [8, 12]. Moreover, medical images require a higher accuracy since the slide images are used for diagnostic purposes. Taken over the whole, an efficient deep CNNs architecture for medical images needs to be well designed to get better accuracy.

To the best of our knowledge, microscopic image SR reconstruction has examined in only one study [32], yet. A deep neural network architecture was proposed in order to improve the performance of a microscope and the network model was tested using various tissue samples in their work. The study shows the applicability of the model to other imaging modalities. However, the main limitations of the study are that its effectiveness highly depends on the scaling factor and tissue types. The aforementioned study examined only one scaling factor in order to obtain the desired HR image. Furthermore, using various types of samples and stains in network training caused color distortion in the output images.

The motivation of the current work is three fold:

1. *Preprocessing*: For microscopic imaging where the image quality greatly effects the diagnosis of the disease, SR as a preprocessing step is quite desirable. Therefore, the proposed approach may be employed as a preprocessing stage to obtain a HR image that may later be passed to a microscopic segmentation and classification system. Hence, the classification accuracy may be improved significantly.
2. *Virtual microscope slide*: A large collection of digitally scanned images from a microscope sample at a low magnification, e.g. 20×, may be combined into one final output image using panoramic stitching techniques in order to construct a virtual microscope slide. Then, the proposed approach may be employed to acquire a HR panoramic image for viewing without a microscope at various high magnification, e.g. 40× or 100×. The final output image is close to the actual image as viewed with a microscope.
3. *Larger magnification*: A light optical microscope used by microbiologists magnifies the specimen up to 100×. However, some regions of the specimen within the viewing area may be still small to detect the parasites. Therefore, the viewed areas taken at the highest magnification, i.e. 100×, may be magnified to become larger than viewing window using the proposed approach.

In this paper, we apply deep CNNs to microscopic images in order to perform SR as in [32]. However, the network architecture is redesigned to perform SR task producing image with high quality even with a higher upsampling ratio and with various stain types. The contribution of the current work is two fold:

1. *Staining with different dyes*: The staining procedure effects the color of some parts of the structure to be observed and there are several types of microscopic stains can be used in microscopic imaging. The proposed approach can be effectively used for various types of samples and stains without any color distortion thanks to concept of residual learning.
2. *Multi-scale model*: The training of the network for different scaling factors results more parameters to be optimized and requires more time and memory. The proposed approach can be efficiently used for multi-scale model by combining datasets for each scaling factor to constitute a big dataset and then training the network using whole dataset.

Experimental results show that the proposed method yields higher SR performance.

The rest of the paper is organized as follows. The theory of the deep CNNs for SR task is briefly explained in Section 2. In Section 3, the proposed method is presented in detail. The proposed approach is validated with experimental clinical data in Section 4. Finally, the paper is concluded in the last section.

2 Deep convolutional neural network for super resolution

Most recently, due to its success in several computer vision tasks, such as image classification [22, 37], image identification [43, 51], object detection [9, 31], information retrieval [26, 35, 36], deep CNNs have been introduced to the field of image restoration including image SR [5, 6, 17–19], inpainting [45] and denoising [14, 45, 50] etc. Different from traditional learning based methods in image SR, CNNs can directly learn an end-to-end mapping between LR input and HR output. Therefore, CNN based SR methods have shown significant improvements.

CNNs are generally made up of three layer types: convolutional layers, pooling layers and fully connected layers. Since the function of pooling layer is to reduce the spatial size of the representation and the image details are very important for image SR problem, the CNNs applied for image SR contain no pooling or fully connected layers.

In a CNN architecture, the input feature map F_{l-1} is convolved with the convolutional kernels W_l which are the learnable filters, where l represents the l^{th} convolution layer. The obtained feature map is offset by the bias term b_l. A non-linear activation function is then applied to the end of each convolution layer. ReLU [10, 25] is the most popular activation function and it can be defined as $ReLU(a) = max(a, 0)$. Therefore, a basic convolutional layer can be represented as:

$$F_l = ReLU(W_l * F_{l-1} + b_l) \tag{1}$$

where $*$ denotes the 2D convolution operator along width and height directions. The output feature map F_l is given as an input feature map to another layer to compose a multi-layer CNN architecture.

3 Proposed method

In this section, we describe the proposed model of multi-scale deep CNNs based microscopic image SR. The model consists of two sub-processes: data preprocessing and network training. The SR reconstruction from real microscopic images require the accurate alignment of the LR input and HR output. The data preprocessing is used to register the two microscopic images taken of the same slide with different objective lenses. Therefore, we first analyze the data preprocessing. Then, we discuss the network architecture used in our experiments in more detail and describes the overall deep CNNs based SR algorithm for real microscopic images.

3.1 Data preprocessing

In microscopy, the field of view is widest on the lowest objective lens whereas when you switch to a higher objective, specimen to be examined becomes very close. Therefore, in

order to match the corresponding field of views of each images taken under the different objectives, an image alignment procedure whose structure is given in Fig.1 is used to generate a dataset including LR input and corresponding HR output.

Due to the specimens are scanned under a HR microscope with advanced technology, a simple alignment procedure is sufficient to match the corresponding images. It is well known that the 2-D normalized cross correlation (NCC) has been commonly used to measure the similarity between two images and is preferred for simple template matching applications [20, 40]. Therefore, 2-D NCC is applied to the corresponding images using the following steps:

1. The input images acquired with different objective lenses are converted to grayscale images.
2. Each HR image taken with a higher objective is down-sampled using bicubic interpolation method by a factor of s to generate template window.
3. The template window is slided in image function pixel-by-pixel and the 2-D NCC is computed between them using (2).

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\left\{\sum_{x,y}[f(x,y) - \bar{f}_{u,v}]^2 \sum_{x,y}[t(x-u,y-v) - \bar{t}]^2\right\}^{0.5}} \tag{2}$$

where $\bar{t}$ is the mean of the template and $\bar{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template.

4. The maximum value of the computed correlation values indicates the best match between the template function and subimage in image function.

3.2 Network architecture

In order to achieve image SR problem in microscopy images, a 20-layer deep convolutional neural network is utilized as illustrated in Fig. 2. Similar to the SR approaches, the network takes an interpolated LR image (to the desired size) as input to reconstruct HR image as output.

The input to network is a fixed-size 48 × 48 Y channel image. The input convolutional layer maps the input image into 64 channels. The input convolutional layer is followed by 18 blocks consist of a convolutional layer and ReLU activation function. In these blocks, 64 filters operate on 3 × 3 region across 64 feature maps. A single filter of size 3 × 3 with 64 channels for the last layer is used to produce residual image to be added back to the input image.

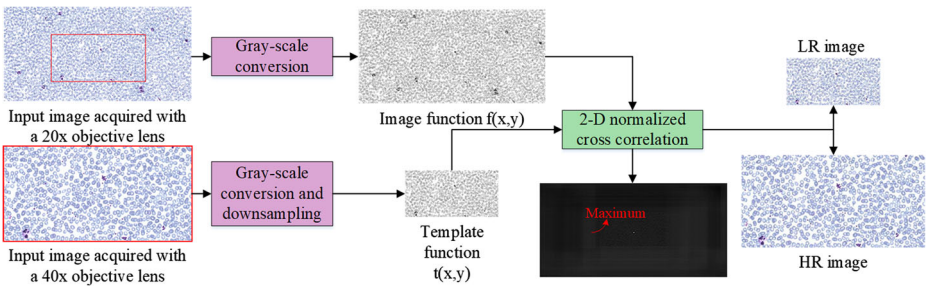

Fig. 1 Data preprocessing architecture for microscopic imaging

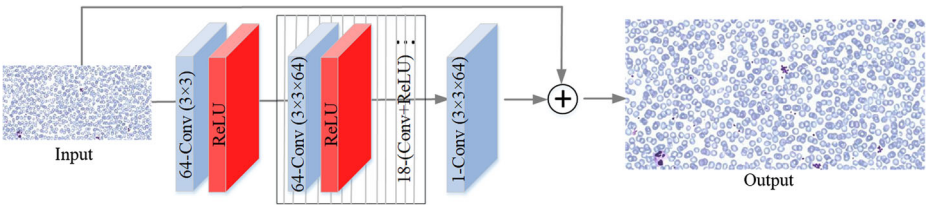

Fig. 2 Deep convolutional neural network architecture for microscopic imaging

Increasing the depth of the network leads to reduction in the size of the feature maps after each convolution layer. Thus, the reconstructed residual image is small when compared with the input interpolated LR image. In order to overcome this problem, zero-padding is applied around boundaries of the feature maps before applying convolutions to keep the feature map size fixed.

3.3 Network training

Let x, θ and y denote the interpolated LR image, the set of network parameters to be optimized and a HR image, respectively. Given a training dataset $\{x^{(i)}, y^{(i)}\}_{i=1}^{N}$, the goal is to find the best model f to generate a HR image $\hat{y} = f(x, \theta)$, where $\hat{y}$ is the estimation of the HR image y. The loss function to be minimized in order to find optimal parameters can be represented as:

$$\mathbb{E}(L) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left\|y^{(i)} - f(x^{(i)}, \theta))\right\|^2 \tag{3}$$

According to (3), the interpolated LR image is forwarded through the network being trained as in most of the SR works using deep CNNs. However, only the lost details in the interpolated LR image are predicted in each layer of the network. This end-to-end mapping requires more memory. Therefore, a residual image r, where most of the pixels are close to zero, defined as the difference of the ground truth and input images, i.e. $r = y - x$. The direct end-to-end mapping problem is transformed to a residual mapping problem. Thus, only the residual image is predicted so that the learning process leads to a more easier optimization process and better performance. The loss function in (3) can be rearranged as:

$$\mathbb{E}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left\|r^{(i)} - f(x^{(i)}, \theta))\right\|^2 \tag{4}$$

In the study, the loss function is minimized by Stochastic Gradient Descent with Momentum (SGDM) optimizer as defined in (5).

$$\theta_{l+1} = \theta_l - \gamma\nabla\mathbb{E}(\theta_l) + \alpha(\theta_l - \theta_{l-1}) \tag{5}$$

where l is the iteration number, γ is the learning rate and $\mathbb{E}(\theta)$ is the loss function. ∇ represents the gradient operator and α (namely, momentum) determines the contribution of the previous gradient step to the current iteration. The smaller learning rate ensures the

convergence but network training takes a long time. On the other hand, the usage of higher learning rate results in vanishing/exploding gradients. In order to boost the convergence by using higher learning rate, adjustable gradient clipping [29, 30] is used to suppress exploding gradients. The gradient clipping scheme with a given predefined range $[-\theta, \theta]$ avoids the gradient explosion by clipping the gradients to $[-\frac{\theta}{\gamma}, \frac{\theta}{\gamma}]$, where γ represents the current learning rate. Thus, the network converges quickly and produces effective results.

Most of the deep CNNs methods are needed to train from stratch for a different scaling factor. This network training results in more parameters to be optimized, and so it is memory and time consuming. A multi-scale model for different scaling factors can be efficiently trained by combining datasets for several scaling factors into a dataset. Therefore, we can obtain a well-trained model and the optimized parameters is transferred to the network containing different scaling factors.

4 Experiments and results

In this section, we first briefly introduce training and testing datasets, and quality metrics used in our study. Then, we present the experimental setup and network analysis. The proposed method is analyzed from two aspects, which are single model for multiple magnification objectives -multi-scale- and single model for different stained tissue section. Finally, we evaluate the performance of our model with the real microscopy images.

4.1 Training and testing datasets

The parasitology slides of thin smear blood samples are taken from the MaMic image database [44] which is scanned with a virtual microscopy platform. In order to train the network and evaluate the performance of our proposed SR approach, the microscopy images are acquired with 3 different magnification objectives ($10\times, 20\times, 40\times$). The images with 600×300 resolution are stored in jpeg file format, with 24 bits per pixel, in color.

The LR images are acquired with $10\times$ and $20\times$ objective lenses whereas the HR images are captured with $40\times$ objective lens. 26 LR images corresponding to different regions of the sample are generated for each of the two different parasitology slides and each of the objective lenses. Out of these images, 16 LR and their corresponding HR images are used as training dataset, while the other 10 images are used to evaluate the performance of upscaling factors $\times 2$ and $\times 4$. Since the big data generally get better results and the size of the training images are larger, we augment the training dataset by extracting 48×48 subimages with a stride of 12 from each image. By using the formula $(\frac{Width-FilterSize}{Stride} + 1) * (\frac{Height-FilterSize}{Stride} + 1) * NoOfImage$, where width and height are the image width and height, respectively, and filter size and stride are set to 48 and 12, respectively, more than **16000** image pair can be obtained for a specific magnification objective from the interpolated LR and HR images. Since the proposed network takes the interpolated LR image (to the desired size) as the input and reconstructs the HR image as the output, the one-to-one correspondence between them is established when the dataset is generated from them for training process. On the other hand, the whole test image is used for SR reconstruction. Similar to the SR literature, since the human visual system is more sensitive to details in intensity than in color, the SR reconstruction is only performed on the luminance component (Y channel in YCbCr color space) and, Cb and Cr channels are directly magnified to the desired size using bicubic interpolation in order to get a better display.

4.2 Quality metrics

The microscopic image SR performance is evaluated with widely used metrics: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [41]. PSNR is derived from the mean square error (MSE) which measures the average squared difference between reconstructed HR image and original HR image. MSE and PSNR are defined in (6) and (7), respectively.

$$MSE = \frac{1}{h * w} \sum_{i=1}^{h} \sum_{j=1}^{w} (\hat{H}(i, j) - H(i, j))^2 \tag{6}$$

$$PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) \tag{7}$$

where $\hat{H}$ and H represent the predicted and original HR images, respectively, h and w denote the image dimensions. A higher PSNR generally indicates that SR reconstruction is closer to the ground truth.

SSIM combines local image structure, luminance and contrast into a single local quality score and measures the image similarity [41]. The brightness $I(x, y)$, the contrast $c(x, y)$ and the structural $s(x, y)$ term are formulated as

$$I(\hat{H}, H) = \frac{2\mu_{\hat{H}}\mu_H + C_1}{\mu_{\hat{H}}^2 + \mu_H^2 + C_1} \tag{8}$$

$$c(\hat{H}, H) = \frac{2\sigma_{\hat{H}}\sigma_H + C_2}{\sigma_{\hat{H}}^2 + \sigma_H^2 + C_2} \tag{9}$$

$$s(\hat{H}, H) = \frac{\sigma_{\hat{H}H} + C_3}{\sigma_{\hat{H}}\sigma_H + C_3} \tag{10}$$

where $\mu_{\hat{H}}$ and μ_H are the local means for images $\hat{H}$ and H, respectively. $\sigma_{\hat{H}}$, σ_H and $\sigma_{\hat{H}H}$ represent the standard deviations and cross-covariance. C_1, C_2 and C_3 are constant values and C_3 defaultly set to $C_2/2$. SSIM index can be calculated as:

$$SSIM(\hat{H}, H) = \frac{(2\mu_{\hat{H}}\mu_H + C_1)(2\sigma_{\hat{H}H} + C_2)}{(\mu_{\hat{H}}^2 + \mu_H^2 + C_1)(\sigma_{\hat{H}}^2 + \sigma_H^2 + C_2)} \tag{11}$$

The SSIM index is a decimal value between 0 and 1. The higher SSIM value indicates that SR performance is better.

4.3 Implementation details

In our proposed model, we use a network of depth 20. Momentum and weight decay parameters are set to 0.9 and 0.0001, respectively. The learning rate is initially set to 0.1 for all layers and decreased by a factor of 10 every 10 training epochs. All the weights of filters in convolution layers are initialized using the method described in [11] and biases are initially set to zero. The activation function of the proposed model utilizes the ReLU function. The learning is stopped after 100 epochs (25800 iterations with batch size 64). For optimization, we use stochastic gradient descent with momentum. All models are trained using the Matlab on a NVIDIA GeForce GTX 1050 Ti and training takes roughly 4 hour.

4.4 Single model for multiple scales

Most of the deep CNNs studies use a trained model to construct HR images with the same scale factor. Similarly, a trained model with single scale factor, i.e. $\times 2.5$, is employed to reconstruct HR microscopic image with the same scale factor in [32] which is the first and only study proposed for deep learning based microscopic image SR. In our proposed study, we explore various scaling factors to construct different network models and conduct experiments to demonstrate the single model for multiple scale property of the proposed deep CNNs approach. To sum up, the single-scale and multi-scale model are explained as follows:

– *Single-scale model*: A network is created for each scale factor. It results more parameters to define and optimize a network. So, it is memory and time-consuming.
– *Multi-scale model*: Training datasets for each scale factor are combined to constitute a big dataset in order to train the network. A network is created for whole scale factors. Therefore, parameters are shared across all scale factors. It is an economical way to train, store and retrieve networks.

In the experiments, two scale factors ($\times 2$, $\times 4$) are considered for microscopic image SR. Firstly, the network is trained with a single scale factor and tested under the same scale factor. For single scale model trained with $\times 2$, LR and HR image pairs are composed of microscopic images taken under $20\times$ and $40\times$, respectively, whereas the LR and HR image pairs are captured under $10\times$ and $40\times$,respectively, for scale factor $\times 4$. Then, single-scale model is trained with a single scale factor and tested both in the same scale factor and the other scale factor for each dataset.

Experimental results are summarized in Table 1. The single scale models perform excellently when the HR images are restored with the corresponding training scale factors. However, the performance is decreased when the model is tested under another scale factor. For scale factor $\times 2$ for Dataset 1, the PSNR value achieved by the model trained with factor $\times 2$ is 27.4025 (in dB) whereas the model trained with factor $\times 4$ gives PSNR of 24.9003 (in dB). The network also performs well on the Dataset 2, with the same result as in Dataset 1. The same network is then trained with multiple scale factors, i.e. $\times 2$ and $\times 4$. The PSNR values for each scale are comparable to those achieved by the single scale models: 27.2336 vs. 27.4025 ($\times 2$), 22.9829 vs. 22.8977 ($\times 4$) for Dataset I and 29.6277 vs. 29.7153 ($\times 2$), 27.1281 vs. 27.0794 ($\times 4$) for Dataset 2. It indicates that the proposed approach can handle the multi-scale model when the HR reconstruction is performed at any scale used during training. In addition, it can be concluded that the proposed multi-scale

Table 1 Average PSNR (in dB) and SSIM results for the MaMic datasets, comparing bicubic interpolation and the deep CNNs output at single scale and multi-scale model

Dataset	Scale factor (Test/Train)	Bicubic	Single scale $\times 2$	Single scale $\times 4$	Multi scale $\times 2,4$
Dataset 1	$\times 2$	26.2849/0.8024	**27.4025/0.8376**	24.9003/0.6918	**27.2336/0.8294**
	$\times 4$	21.9487/0.5602	22.5964/0.5910	**22.8977/0.5622**	**22.9829/0.5707**
Dataset 2	$\times 2$	29.0063/0.8250	**29.7153/0.8529**	29.2688/0.8230	**29.6277/0.8491**
	$\times 4$	25.8648/0.7024	26.5085/0.7380	**27.0794/0.7496**	**27.1281/0.7526**

Bold text denotes that the test scale is used during training

network model exceeds the performance of single scale network for large scale, i.e. ($\times 2$, 4) gives PSNR 22.9829 and 27.1281 for Dataset 1 and Dataset 2, respectively, whereas ($\times 4$) gives PSNR 22.8977 and 27.0794. The LR images and the reconstructed images are given in Figs. 3, 4, 5 and 6. The proposed approach produces sharp edges although the LR input image is severely distorted.

4.5 Robustness to color variation

For microscopic imaging, it is well known that staining procedure changes the color of some parts of the structure being observed and allows to visualize better under the microscope. There are several types of microscopic stains and each can be used in microscopy imaging.

In Section 4.4, the experiments are conducted on two different datasets whose samples are stained with the same dye in its own dataset and higher quality microscopic images are reconstructed from LR ones using the proposed deep CNNs approach. In this section, we investigate the robustness of the proposed approach to color variation. To achieve this,

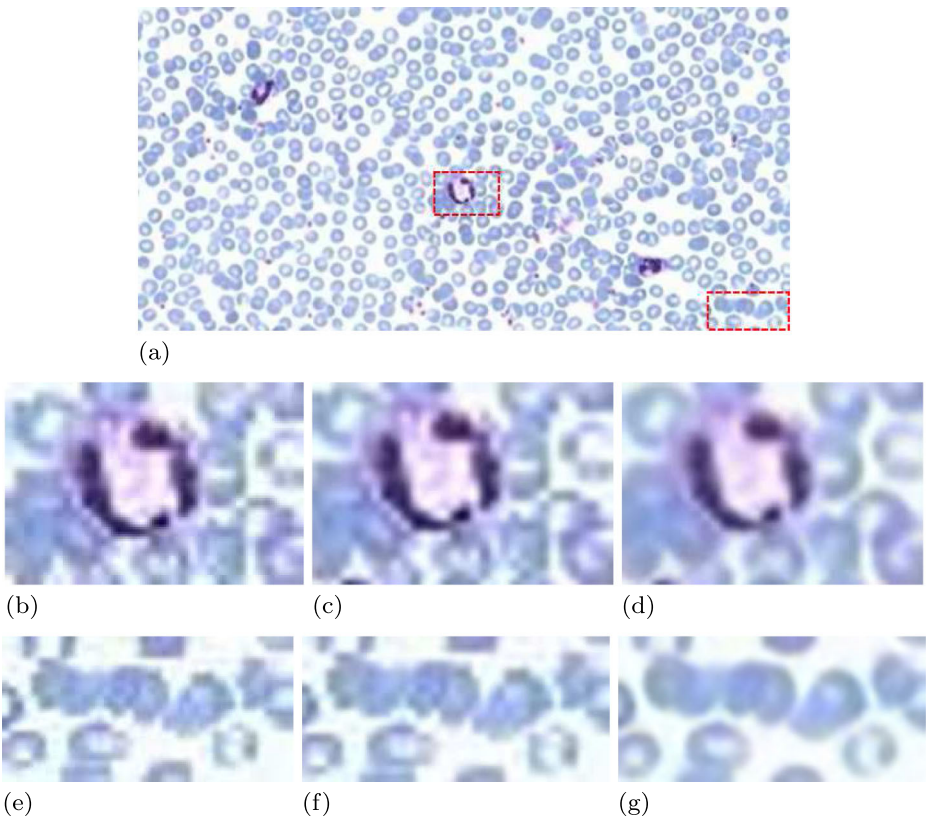

Fig. 3 Super resolution result of a thin smear blood samples using single scale model. The network was trained with $\times 2$ and tested under $\times 2$ scale factor. **a** Image of the deep CNNs result corresponding to $20\times$ input image. The red rectangle boxes denote the region of interesting patches (ROIs) to be magnified to see difference clearly. **b**, **e** Enlarged ROIs of the LR input image ($20\times$). **c**, **f** Enlarged ROIs of the image super resolved with bicubic interpolation (PSNR: 24.9961, SSIM: 0.7844). **d**, **g** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR: 26.7010, SSIM: 0.8380)

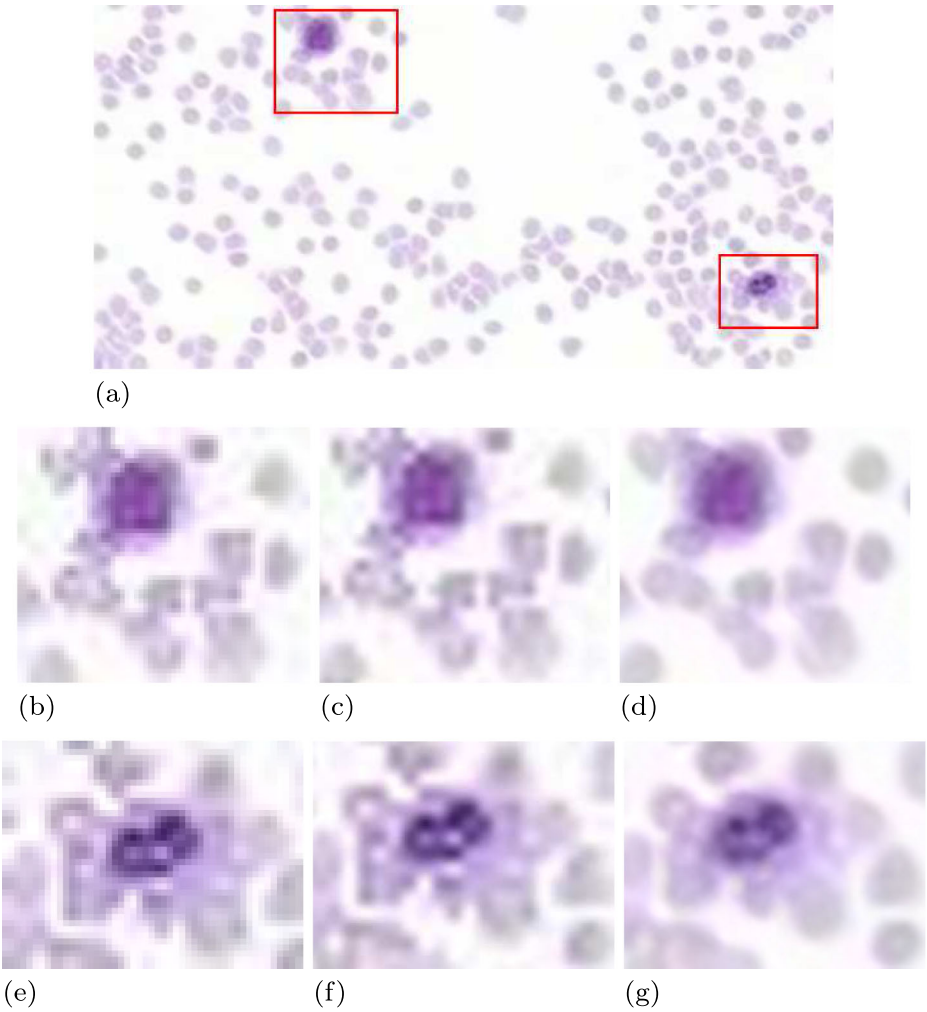

Fig. 4 Super resolution result of a thin smear blood samples using single scale model. The network was trained with $\times 4$ and tested under $\times 4$ scale factor. **b** Image of the deep CNNs result corresponding to $10\times$ input image. The red rectangle boxes denote the region of interesting patches (ROIs) to be magnified to see difference clearly. **b**, **e** Enlarged ROIs of the LR input image ($10\times$). **c**, **f** Enlarged ROIs of the image super resolved with bicubic interpolation (PSNR:28.4036, SSIM:0.7967). **d**, **g** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR:29.7667, SSIM:0.8406)

the proposed deep CNNs model is firstly trained using the microscopic images captured on slide of purple-stained smear sample, i.e. Dataset 2, and tested using the microscopic images of blue-stained smear sample slide, i.e. Dataset 1 or vice versa. As in the first experiment, different scaling factors to construct different network models are examined and average PSNR and SSIM results are listed in Table 2. Compared with Table 1, the experiments performed on these two scenarios yield similar results. Also, it is apparent that the results are slightly better than those obtained with the same dataset used for training and testing for Dataset 2 for scale factor $\times 2$ as it is unexpected: 29.7153 vs. 29.7987($\times 2$), 29.2688 vs. 29.3364($\times 4$), 29.6277 vs. 29.7259($\times 2, 4$). Figures 7 and 8 illustrate the validity and

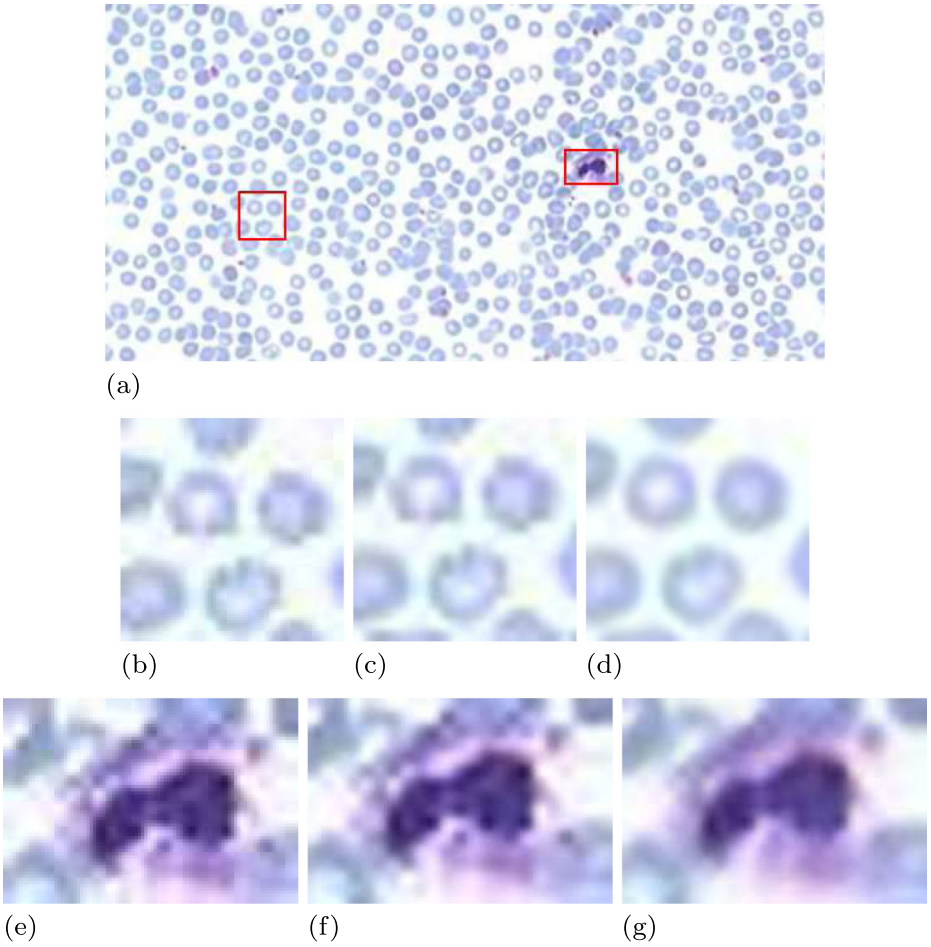

Fig. 5 Super resolution result of a thin smear blood samples using multi-scale model. The network was trained with $\times 2, 4$ and tested under $\times 2$ scale factor. **a** Image of the deep CNNs result corresponding to $20\times$ input image. The red rectangle boxes denote the region of interesting patches (ROIs) to be magnified to see difference clearly. **b**, **e** Enlarged ROIs of the LR input image ($20\times$). (c,f) Enlarged ROIs of the image super resolved with bicubic interpolation (PSNR:28.2829, SSIM:0.8494). **d**, **g** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR:29.7518, SSIM:0.8856)

applicability of the proposed approach in the color variation of the images. The figures in the first rows correspond to the SR result of a deep CNNs model which is trained and tested using the same dataset. The output images of a deep CNNs model which uses different dataset for training and testing are illustrated in the second rows. These results demonstrate that the proposed deep CNNs approach can be effectively used for various types of samples and stains without any color distortion in order to obtain better quality images.

4.6 Computational cost

A low-complexity and high-accuracy implementations are always desirable for many real-world applications. Moreover, the test time is more crucial for these applications compared

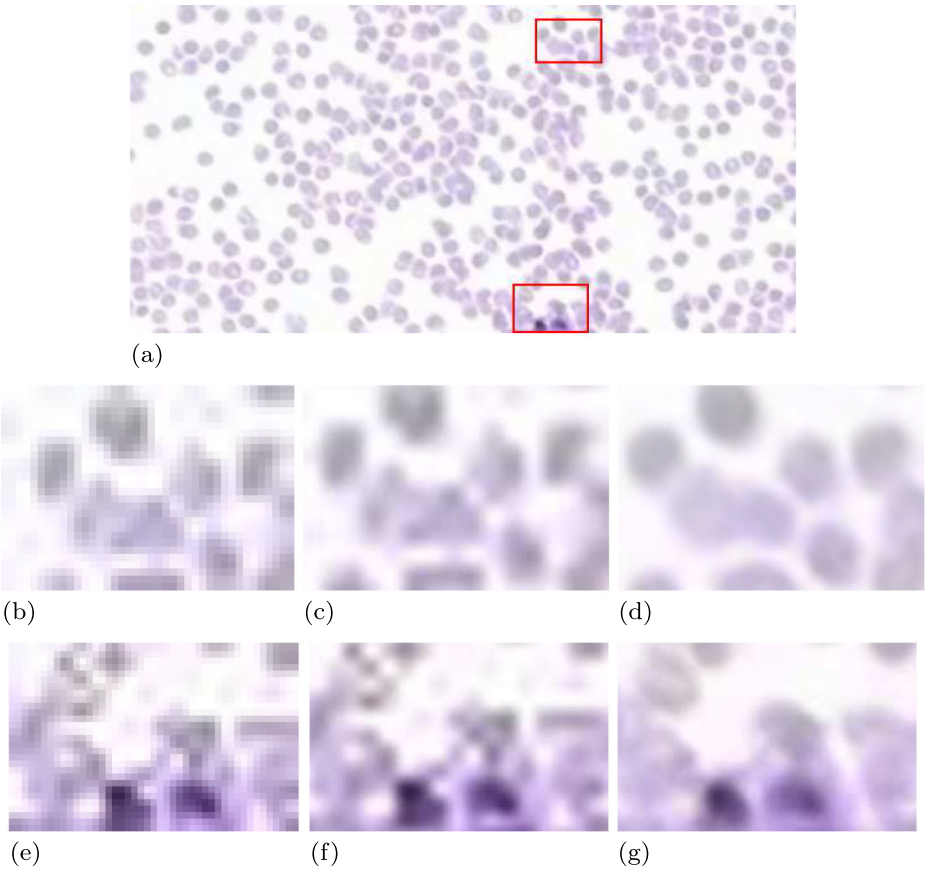

Fig. 6 Super resolution result of a thin smear blood samples using multi-scale model. The network was trained with ×2, 4 and tested under ×4 scale factor. **a** Image of the deep CNNs result corresponding to 10× input image. The red rectangle boxes denote the region of interesting patches (ROIs) to be magnified to see difference clearly. **b**, **e** Enlarged ROIs of the LR input image (10×). **c**, **f** Enlarged ROIs of the image super resolved with bicubic interpolation (PSNR:23.5932, SSIM:0.6053). **d**, **g** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR:25.3092, SSIM:0.6846).

Table 2 Average PSNR (in dB) and SSIM results of applying the Dataset 1 trained deep CNNs model on Dataset 2 input images or vice versa for the MaMic datasets, comparing bicubic interpolation and the deep CNNs output at single scale and multi-scale model

Dataset	Scale factor (Test/Train)	Bicubic	Single scale ×2	Single scale ×4	Multi scale ×2,4
Dataset 1	×2	26.2849/0.8024	**27.1187/0.8222**	24.5133/0.6727	**26.9979/0.8140**
	×4	21.9487/0.5602	22.5871/0.5793	**22.7730/0.5525**	**22.8824/0.5628**
Dataset 2	×2	29.0063/0.8250	**29.7987/0.8555**	29.3364/0.8218	**29.7259/0.8509**
	×4	25.8648/0.7024	26.4082/0.7358	**26.9736/0.7454**	**27.0459/0.7500**

Bold text denotes that the test scale is used during training

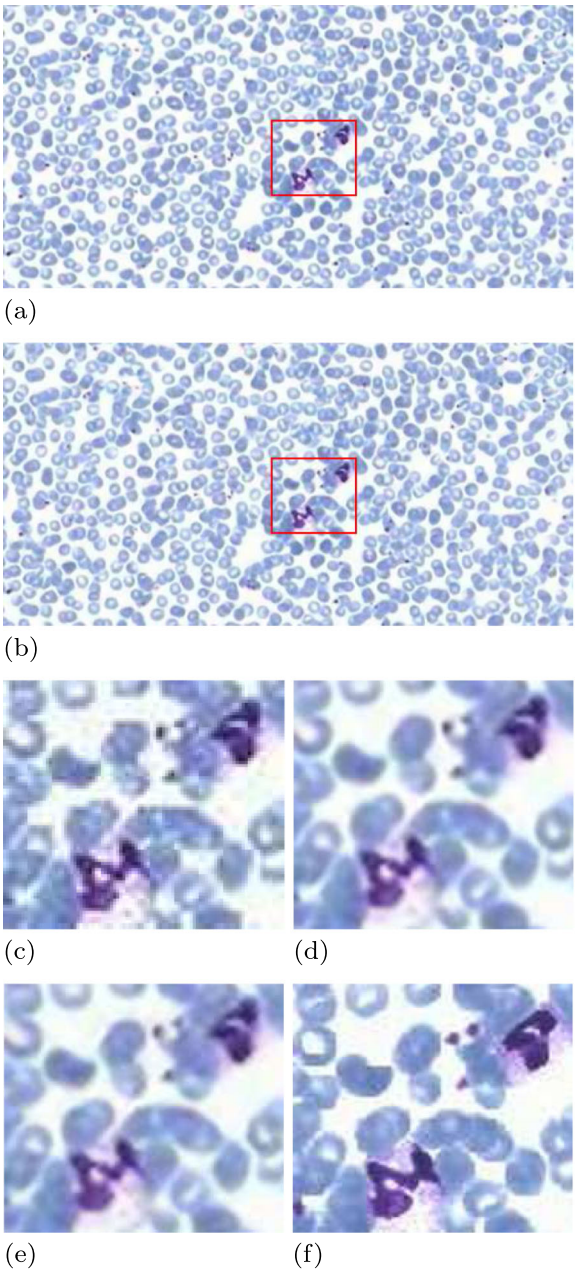

Fig. 7 **a** Super resolution result of applying the *Dataset 1* trained deep CNNs model with ×2 on a 20× *Dataset 1* input image. **b** Super resolution result of applying the *Dataset 2* trained deep CNNs model with ×2 on a 20× *Dataset 1* input image. The red rectangle boxes denote the region of interesting patches (ROIs) to be magnified to see difference clearly. **c** Enlarged ROIs of the LR input image (20×). **d** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR:26.0249, SSIM:0.8288); the network is trained with Dataset 1. **e** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR:25.9976, SSIM:0.8186); the network is trained with Dataset 2 consists of the images stained with a different dye. **f** Enlarged ROIs of the HR image (40×)

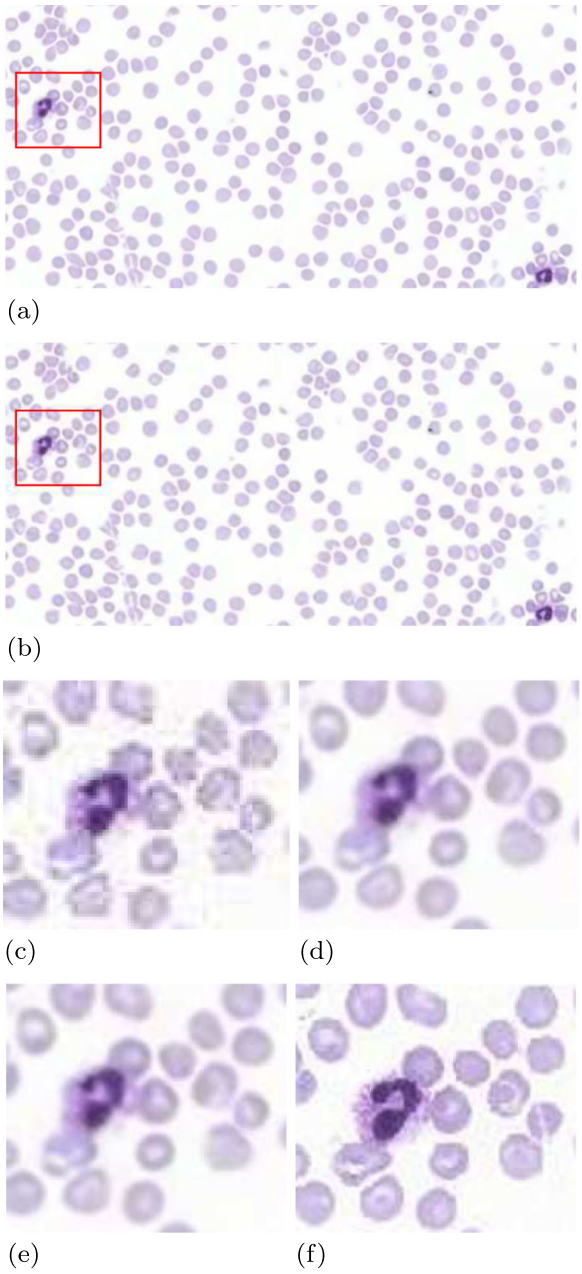

Fig. 8 **a** Super resolution result of applying the *Dataset 2* trained deep CNNs model with ×2 on a 20× *Dataset 2* input image. **b** Super resolution result of applying the *Dataset 1* trained deep CNNs model with ×2 on a 20× *Dataset 2* input image. The red rectangle boxes denote the region of interesting patches (ROIs) to be magnified to see difference clearly. **c** Enlarged ROIs of the LR input image (20×). **d** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR: 29.7103, SSIM:0.8516); the network is trained with Dataset 2. **e** Enlarged ROIs of the image super resolved with deep CNNs approach (PSNR:30.9416, SSIM:0.8795); the network is trained with Dataset 1 consists of the images stained with a different dye. **f** Enlarged ROIs of the HR image (40×)

Table 3 Time complexities of SR methods on different datasets (in seconds)

Dataset	Scale factor	Bicubic	Proposed method
Dataset 1	×2	0.05	0.24
	×4	0.05	0.23
Dataset 2	×2	0.04	0.24
	×4	0.05	0.25

with the training time. Therefore, an experiment is also conducted to demonstrate the efficiency of the proposed method in terms of time complexity. The average run time of the proposed method as well as the completing method are given in Table 3. The algorithms are implemented in MATLAB 2018b and executed on a 3.6 GHz processor with 8 GB RAM, Linux 64-bit PC. As shown in Table 3, the proposed method has reasonable computational complexity and can be efficiently used in real-world applications.

5 Conclusion

In this paper, we have proposed multi-scale deep CNNs based SR method capable of significantly enhancing spatial resolution of microscopic system even with a higher upsampling ratio and with various stain types. The proposed approach needs a single image taken with low objective lenses under a standard microscope and outputs the better resolution image without changing the hardware design. Different from the previous deep CNNs based microscopic image SR method [32], a multi-scale network model is put forward to eliminate the need to train the network from stratch for a different scaling factor. Furthermore, the proposed model reconstructs the HR image by means of residual learning in order to cope with the color distortion problem arising from the color variation between the various stain types. Evaluations performed on a public database scanned with a virtual microscopy platform with different upscaling factors show that the proposed deep CNNs model has achieved superior microscopic image SR performance in both visual and quantitative analysis.

Although superior performance has achieved on the public database, we think that better visual and quantitative results may be obtained by using high quality images since the quality of the images in the public database is very limited.

References

1. Ayas S, Ekinci M (2014) Random forest-based tuberculosis bacteria classification in images of zn-stained sputum smear samples. SIViP 8(1):49–61
2. Bauschke HH, Borwein JM (1996) On projection algorithms for solving convex feasibility problems. SIAM Rev 38(3):367–426
3. Bevilacqua M, Roumy A, Guillemot C, Morel MLA (2012) Low-complexity single image super resolution based on nonnegative neighbor embedding. In: Proceedings of the 23rd British machine vision conference, pp 1–10
4. Chang H, Yeung DY, Xiong Y (2004) Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004. CVPR 2004. IEEE, vol 1, pp I–I

5. Dong C, Loy CC, He K, Tang X (2016) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38(2):295–307
6. Dong C, Loy CC, Tang X (2016) Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. Springer, pp 391–407
7. Duchon CE (1979) Lanczos filtering in one and two dimensions. J Appl Meteorol 18(8):1016–1022
8. Eaton-Rosen Z, Bragman F, Ourselin S, Cardoso MJ (2018) Improving data augmentation for medical image segmentation. In: International conference on medical imaging with deep learning, pp 1–3
9. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
10. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp 315–323
11. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034
12. Hussain Z, Gimenez F, Yi D, Rubin D (2017) Differential data augmentation techniques for medical imaging classification tasks. In: AMIA annual symposium proceedings, American Medical Informatics Association, pp 979
13. Irani M, Peleg S (1993) Motion analysis for image enhancement: resolution, occlusion, and transparency. J Vis Commun Image Represent 4(4):324–335
14. Jain V, Seung S (2009) Natural image denoising with convolutional networks. In: Advances in neural information processing systems, pp 769–776
15. Jan Z, Khan A, Sajjad M, Muhammad K, Rho S, Mehmood I (2018) A review on automated diagnosis of malaria parasite in microscopic blood smears images. Multimed Tools Appl 77(8):9801–9826
16. Keys R (1981) Cubic convolution interpolation for digital image processing. IEEE Trans Acoust Speech Signal Process 29(6):1153–1160
17. Kim J, Kwon Lee J, Mu Lee K (2016) Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1646–1654
18. Kim J, Kwon Lee J, Mu Lee K (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1637–1645
19. Lai WS, Huang JB, Ahuja N, Yang MH (2017) Deep laplacian pyramid networks for fast and accurate superresolution. In: IEEE conference on computer vision and pattern recognition, vol 2, pp 5
20. Lewis JP (1995) Fast template matching. In: Vision interface, vol 95, pp 15–19
21. Li M, Nguyen TQ (2008) Markov random field model-based edge-directed image interpolation. IEEE Trans Image Process 17(7):1121–1128
22. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M (2014) Medical image classification with convolutional neural network. In: 2014 13th international conference on control automation robotics & vision (ICARCV). IEEE, pp 844–848
23. Li X, Orchard MT (2001) New edge-directed interpolation. IEEE Trans Image Process 10(10):1521–1527
24. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
25. Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814
26. Nie L, Wang X, Zhang J, He X, Zhang H, Hong R, Tian Q (2017) Enhancing micro-video understanding by harnessing external sounds. In: Proceedings of the 2017 ACM on multimedia conference. ACM, pp 1192–1200
27. Organization WH (2017) Global tuberculosis report 2017. World Health Organization
28. Organization WH (2017) World malaria report 2017. World Health Organization
29. Pascanu R, Mikolov T, Bengio Y (2012) Understanding the exploding gradient problem. arXiv:12115063
30. Pascanu R, Mikolov T, Bengio Y (2013) On the difficulty of training recurrent neural networks. In: International Conference on Machine Learning, pp 1310–1318
31. Ren S, He K, Girshick R, Sun J (2017) Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
32. Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A (2017) Deep learning microscopy. Optica 4(11):1437–1443
33. Schultz RR, Stevenson RL (1996) Extraction of high-resolution frames from video sequences. IEEE Trans Image Process 5(6):996–1011

34. Shi H, Ward R (2002) Canny edge based image expansion. In: IEEE international symposium on circuits and systems, 2002. ISCAS 2002. IEEE, vol 1, pp I–I
35. Song X, Feng F, Liu J, Li Z, Nie L, Ma J (2017) Neurostylist: neural compatibility modeling for clothing matching. In: Proceedings of the 2017 ACM on multimedia conference. ACM, pp 753–761
36. Song X, Feng F, Han X, Yang X, Liu W, Nie L (2018) Neural compatibility modeling with attentive knowledge distillation. arXiv:180500313
37. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) Breast cancer histopathological image classification using convolutional neural networks. In: 2016 international joint conference on neural networks (IJCNN). IEEE, pp 2560–2567
38. Timofte R, De Smet V, Van Gool L (2013) Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of the IEEE international conference on computer vision, pp 1920–1927
39. Timofte R, De Smet V, Van Gool L (2014) A+: adjusted anchored neighborhood regression for fast super-resolution. In: Asian conference on computer vision. Springer, pp 111–126
40. Tsai DM, Lin CT (2003) Fast normalized cross correlation for defect detection. Pattern Recogn Lett 24(15):2625–2631
41. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
42. Wang Q, Ward RK (2007) A new orientation adaptive interpolation method. IEEE Trans Image Process 16(4):889–900
43. Wang SH, Tang C, Sun J, Yang J, Huang C, Phillips P, Zhang YD (2018) Multiple sclerosis identication by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. Front Neurosci 12:1–11
44. WebMicroscope (2019) http://fimm.webmicroscope.net/Research/Momic/mamic
45. Xie J, Xu L, Chen E (2012) Image denoising and inpainting with deep neural networks. In: Advances in neural information processing systems, pp 341–349
46. Yang J, Wright J, Huang T, Ma Y (2008) Image super-resolution as sparse representation of raw image patches. In: IEEE conference on computer vision and pattern recognition, 2008. CVPR 2008
47. Yang J, Wright J, Huang TS, Ma Y (2010) Image super-resolution via sparse representation. IEEE Trans Image Process 19(11):2861–2873
48. Zhang L, Wu X (2006) An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans Image Process 15(8):2226–2238
49. Zeyde R, Elad M, Protter M (2010) On single image scale-up using sparse-representations. In: International conference on curves and surfaces, Springer
50. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process 26(7):3142–3155
51. Zhang YD, Pan C, Sun J, Tang C (2018) Multiple sclerosis identification by convolutional neural network with dropout and parametric relu. J Comput Sci 28:1–10

Selen Ayas received her BScE degree in Computer Engineering from Karadeniz Technical University (KTU), Trabzon, Turkey in 2011 and MScE degree in Computer Engineering from the same university in 2014. She is currently a full time Research Assistant and working toward the PhD degree in KTU. Her research interests are computer vision, machine learning and image processing including innovative algorithms such as super resolution and compressed sensing.

Murat Ekinci received his BScE in Electronical Engineering and MScE in Computer Engineering from Karadeniz Technical University (KTU), Turkey, in 1990 and 1993, respectively. He received his PhD degree from the University of Bristol, England, in 1997. He is currently a full time Professor in Computer Engineering Department at KTU. His research interest are computer vision, pattern recognition and image and video processing.