2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-7358-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISBI53787.2023.10230473

HIGH-THROUGHPUT MICROSCOPY IMAGE DEBLURRING WITH GRAPH REASONING ATTENTION NETWORK

Yulun Zhang[1], *Donglai Wei*[2], *Richard Schalek*[3], *Yuelong Wu*[3], *Stephen Turney*[3], *Jeff Lichtman*[3], *Hanspeter Pfister*[3], *Yun Fu*[4]

[1]ETH Zürich, Switzerland, [2]Boston College, USA, [3]Harvard University, USA , [4]Northeastern University, USA

ABSTRACT

High-quality (HQ) microscopy images afford more detailed information for modern life science research and quantitative image analyses. However, in practice, HQ microscopy images are not commonly available or suffer from blurring artifacts. Compared with natural images, such low-quality (LQ) microscopy ones often share some visual characteristics: more complex structures, less informative background, and repeating patterns. For natural image deblurring, deep convolutional neural networks (CNNs) achieve promising performance. But they usually suffer from large model sizes, heavy computation costs, or small throughput, which are critical for high-throughput microscopy image deblurring. To address those problems, we collect HQ electron microscopy and histology datasets and propose a graph reasoning attention network (GRAN). Specifically, we treat deep feature points as embedded visual components, build a graph describing the relationship between all pairs of visual components, and perform reasoning in the graph with a graph convolutional network. The reasoning results are then transferred as attention and residual learning is introduced to achieve graph reasoning attention block (GRAB). We conduct extensive experiments to demonstrate the effectiveness of our GRAN.

Index Terms— Microscopy image, image deblurring, graph reasoning attention network, adversarial training

1. INTRODUCTION

Combined with a multibeam scanning electron microscope (mSEM), automated tape-collecting ultramicrotome (ATUM) based volume serial sectioning can routinely generate images at 225 MB/s continuously for months at a time. These projects create PB-sized data sets of brain tissue for connectomic analysis. As the connectomics field moves toward studying larger and larger volumes, projects become more negatively affected by slowdowns resulting from the acquisition of low-quality (LQ) images. Imperfections in the flatness of a section, or improper number or placement of the focus support points, result in poorly focused images. Since 27% of the total section acquisition time comes during the focusing stage, re-imaging out-of-focus images significantly decreases imaging throughput. Out-of-focus rates typically range from 8-15% of the total number of images. However, extreme cases reach 30%. Though pre-image acquisition solutions aimed at preventing out-of-focus imaging are being pursued, practical solutions are far from being realized. A fast post-acquisition approach to computationally restore the out-of-focus image to an in-focus one is heavily desired.

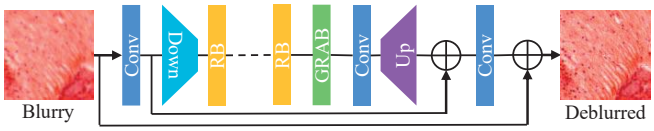


Fig. 1. Illustration of our graph reasoning attention network (GRAN). 'Conv' means convolutional layer. 'RB' means residual block [9]. 'Down' and 'Up' modules denote feature downscaling and upscaling. 'GRAB' denotes graph reasoning attention block. '$\oplus$' means element-wise sum.

Recently, deep networks (e.g., convolutional neural network and Transformer) have been showing their advantages for image restoration (e.g., image super-resolution and deburring) [1, 2, 3, 4, 5, 6, 7, 8]. Deep convolutional neural network (CNN) was firstly introduced for image super-resolution (SR) in SRCNN [1, 2]. In image restoration applications, Zhang *et al.* [3] proposed to learn denoising prior. By utilizing hierarchical features, RDN [6, 7] further improved image restoration performance. Data-driven enhancement was proposed for blurry retinal images in [8]. But, those methods have some limitations that hinder their performance. They either neglect the image characteristics in the microscopy domain or suffer from some intrinsic drawbacks. First, in the microscopy domain, images usually have repeating structural patterns and complex structures. Very large deep networks (e.g., EDSR [9]) could suffer from heavy parameter numbers and computation cost problems. Second, microscopy images usually have a large background region, being far less informative. While, most previous CNN-based techniques neglect to consider the spatial image pixels distinctively. They can hardly distinguish semantic regions (e.g., target and background). Third, most previous image deblurring methods mainly depend on convolutional operations, failing to capture the non-local relationship among input feature points.

To alleviate those issues and limitations, we collect HQ electron microscopy (EM) and histology (HIST) datasets. Such datasets provide extensive and informative materials for life science research and image analyses. We also propose

a lightweight graph reasoning attention network (GRAN) for high-throughput microscopy image deblurring. As illustrated in Fig. 1, we would conduct feature downscaling and upscaling to save computation cost and favor larger receptive field size. In the coarse scale, we treat feature points as visual components and construct a fully-connected relationship graph among them. We further perform reasoning with a graph convolutional network (GCN). We then achieve a graph reasoning block (GRAB), which can be easily inserted into other networks. For microscopy image deblurring, we also investigate the behavior of adversarial training.

2. GRAPH REASONING ATTENTION NETWORK

2.1. Motivations

Different from natural images, there are some specific characteristics in microscopy ones, such as repeating visual patterns, relatively complex and detailed structures, and less informative background. To handle these cases, a more distinguishable mechanism is desired. In high-level visual applications, lots of attention mechanisms have been proposed to focus on more informative channels [10] or spatial positions [11]. For the microscopy image deblurring, we further investigate the reasoning and attention mechanisms to make better use of informative visual regions and patterns.

2.2. Framework

We denote $\mathbf{x}$ and $\mathbf{y}$ as the blurry and clean images respectively. The deblurred output is denoted as $\widehat{\mathbf{y}}$. We insert our proposed graph reasoning attention block (GRAB) in the residual network to form GRAN. As shown in Fig. 1, we conduct feature extraction in the coarse scale (i.e., features between downscaling and upscaling), which could efficiently reduce computation costs. Our GRAN reconstructs deblurred output $\widehat{\mathbf{y}}$ with the blurry input $\mathbf{x}$ by

$$\widehat{\mathbf{y}} = G_{\theta_G}(\mathbf{x}), \tag{1}$$

where the generative network G_{θ_G} is parameterized by θ_G. For the blurry input $\mathbf{x}$, G_{θ_G} reconstructs its corresponding deblurred output $\widehat{\mathbf{y}}$. Furthermore, in the training phase, we introduce a discriminator D_{θ_D} and obtain GRAN_GAN.

2.3. Graph Reasoning Attention Block (GRAB)

Here, we mainly focus on features in the coarse scale between downscaling and upscaling, where most operations are conducted. Let's denote the deep feature set from the coarse scale as V. Considering the relationship among each feature point $\mathbf{v}_i$, we target to further enhance it with GRAB (Fig. 2). Each feature point can be viewed as a visual component.

Recently, visual reasoning [12, 13, 14, 15] has been investigated in deep networks, where the relationship among visual components is modeled and mined. In this work, for the visual components, we are inspired to build the relationship reasoning model. Technically, using weight parameters W_φ and W_ϕ, we first embed the visual components into two

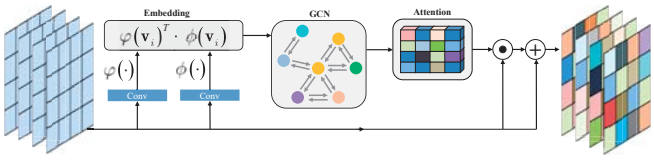


Fig. 2. Illustration of our graph reasoning attention block (GRAB), which helps further enhance the visual components.

embedding spaces. Following the annotations in [15], we calculate the pairwise affinity to build the relationship via

$$R(\mathbf{v}_i, \mathbf{v}_j) = \varphi(\mathbf{v}_i; W_\varphi)^T \phi(\mathbf{v}_j; W_\phi), \tag{2}$$

where we use $\varphi(\mathbf{v}_i; W_\varphi)$ and $\phi(\mathbf{v}_j; W_\phi)$ to denote two embeddings. We get the relationship between $\mathbf{v}_i$ and $\mathbf{v}_j$ via Eq. (2) to form a fully-connected graph.

Given a set of graph nodes (i.e., visual components) V and edges (i.e., feature component relationships) R, we can form a graph $G(V, R)$. For each visual component pair, we measure its affinity edge and obtain affinity matrix R via Eq. (2). If a graph edge has a large affinity value, its corresponding visual component pair is highly correlated in terms of the semantic relationship. Such a relationship can be obtained with visual reasoning [12, 13, 15] in a relationship graph.

To conduct reasoning in this fully-connected graph, we then turn to utilize GCN [16]. For each node, we first define its neighbors with their graph relationships and further compute its response. Some previous related methods [17] incorporate reasoning to enhance the input. In this work, the sigmoid activation function is utilized to obtain reasoning attention. We then further introduce residual learning to connect the output and input features via

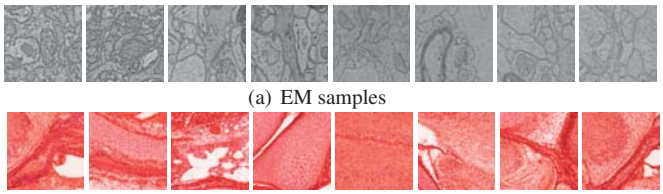
$$V = \sigma\left(\left(\left(RV^T W_g\right) W_r\right)^T\right) \odot V + V, \tag{3}$$

where $\odot$ means element-wise multiplication. W_r is the weight matrix of the residual structure. σ is the sigmoid activation function. The affinity matrix R has a size of $P \times P$. The GCN weight matrix W_g has a size of $C \times C$. In Fig. 2, we illustrate this reasoning process. Using graph reasoning attention, we achieve enhanced visual components.

2.4. Adversarial Training

Most previous microscopy image enhancement methods mainly utilize generative models, which might lead to blurring and over-smoothing artifacts. We innovate a discriminative model to alleviate the blurring artifacts and generate sharper outputs. In natural image restoration, adversarial learning has been widely investigated. However, they ignore capturing the relationship among the entire feature components, hindering their deep representation abilities.

Currently, there are lots of adversarial training strategies used in natural image restoration tasks [4, 18, 5]. Here, we mainly focus on the effect of adversarial training for microscopy image deblurring. Consequently, we introduce the classic discriminator (denoted as D_{θ_D}) in [4] in our deblurring framework. Following the work in the generative adversarial network (GAN) [19], we aim to solve an adversarial



(a) EM samples

(b) HIST samples

Fig. 3. Samples of our collected EM and HIST datasets.

min-max problem. This can be described via

$$\min_{\theta_G} \max_{\theta_D} D_{\theta_D}(\mathbf{y}) - D_{\theta_D}(G_{\theta_G}(\mathbf{x})), \quad (4)$$

where the generator G_{θ_G} tries to fool the discriminator D_{θ_D}. Namely, for the fake inputs, G_{θ_G} misleads D_{θ_D} to give higher scores. While D_{θ_D} aims to give higher scores for target clean images and smaller scores for restored ones. We train generator G_{θ_G} to reconstruct outputs that are highly similar to target clean images in an alternative way.

3. DATASET

3.1. Electron Microscopy (EM) Dataset

To comprehend the diverse appearances of different biological organisms, our electron microscopy dataset (see Fig. 3(a)) contains brain tissues from two different species: one from the temporal gyrus cortex of a human, and the other from the hindbrain of a zebrafish. Both samples were fixed and stained with osmium (rOTO), cut into 30nm-thick sections, and imaged using Zeiss MultiSEM 505 Electron Microscope at 4nm resolution with 570pA beam current and 200ns/pixel scanning speed. For both datasets, we first took in-focus images, then adjusted the objective lens and the stigmators to introduce different types of out-of-focus effects. Due to the imperfections in the calibrations of the microscope, images with different focus conditions may experience both translational and rotational drifts. Therefore, we post-processed the out-of-focus images to align them with their corresponding in-focus images using affine transforms.

3.2. Histology (HIST) Dataset

The histological sample (see Fig. 3(b)) of a free-tailed bat embryo was kindly provided by the Harvard Museum of Comparative Zoology. The bat embryo was sectioned into 10μm thick slices, stained with alum-cochineal and orange G, and mounted on multiple glass slides with tens of sections on each slide. To obtain the ground truth for the defocusing effect of an optical microscope system, we took one slide from the bat embryo slide collection and acquired a z-stack covering 22μm depth range around the focal plane with 2μm step size using a ZEISS Axio Scan.Z1 slide scanner in brightfield mode. The image pixel resolution was 1μm.

4. EXPERIMENTAL RESULTS

4.1. Settings

We split the high-resolution (e.g., 4K resolution) EM and histology images into sub-images, where we further remove images with large backgrounds. As a result, we obtain 768 training, 12 validation, and 160 test EM images (2K resolution). We also collect 704 training, 6 validation, and 22 test histology images (1K resolution). To mimic the blurring process, we synthesize low-quality (LQ) images by using blurring kernels. Specifically, following common settings in IRCNN [3] and RDN [7], we utilize 25×25 Gaussian blur kernel of standard deviation 1.6. The additive Gaussian noise ($\sigma = 2$) is further added to the blurry images. We adopt PSNR and SSIM [20] as evaluation metrics.

Table 1. Investigations of GRAB and its positions. The best PSNR (dB) values on EM and HIST validation data.

Case	Baseline	Low-level	Middle-level	High-level
EM	27.12	27.23	27.39	27.45
HIST	33.87	33.99	34.14	34.21

4.2. Implementation Details

The image super-resolution (SR) network EDSR [9] (40 residual blocks) is utilized as the backbone. We insert GRAB (followed by a Conv layer) before element-wise adding in the coarse scale. The size of all Conv layers is 3×3, except for that in GRAB, where the kernel size is 1×1. For input and output layers, we set the channel number as 3. The remaining Conv layer has 64 channels. To keep the feature map size fixed, we apply the zero-padding strategy. In the training stage, we set the batch size as 16 and input patch size as 96 × 96. We train the model by ADAM optimizer [21] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We set the initial learning rate as 10^{-4} and decrease it to half every 200 epochs.

4.3. Ablation Study

We study the effects of GRAB and its positions by using EDSR baseline [9] as the baseline. We use EM and HIST validation data to report the results.

4.3.1. Effect of GRAB

In Tab. 1, simple EDSR_baseline with GRAB contributes to the performance obviously, no matter where the GRAB is inserted. Our proposed GRAB is demonstrated to be effective with these comparisons. This is mainly because GRAB could take advantage of the non-local relationships among the deep feature points, resulting in better deep representations.

4.3.2. Effect of GRAB Position

We further investigate the insert positions of GRAB. Here, we only focus on the cases, where only one GRAB is inserted in low-level, middle-level, and high-level positions in the coarse scale of the baseline. As we can learn from Tab. 1, GRAB inserted at a higher level would achieve more performance improvements. This is mainly because that GRAB at a higher level would have a larger receptive field size and achieve better graph reasoning attention from the higher-level feature. Consequently, in our following comparisons, we insert the GRAB in the high-level position of the network.

Table 2. PSNR (dB) and SSIM comparison. We mark the best and second-best values in red and blue.

Method	EM		HIST	
	PSNR	SSIM	PSNR	SSIM
Blurry	27.59	0.5514	29.97	0.7915
EDSR [9]	28.46	0.6655	35.12	0.9183
RDN [7]	28.48	0.6664	35.13	0.9186
GRAN (ours)	28.59	0.6693	35.40	0.9215
GRAN_GAN (ours)	27.40	0.6107	33.23	0.8837

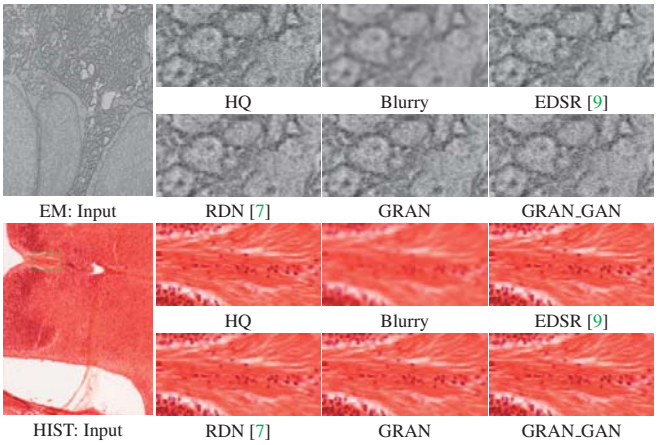


Fig. 4. Visual results about microscopy image deblurring.

4.4. Comparison with Other Methods

We compare our GRAN with several advanced techniques, including EDSR [9] and RDN [7].

4.4.1. Quantitative Results

We provide quantitative comparison in Tab. 2. GRAN represents the results obtained by our generator only. GRAN_GAN indicates that adversarial learning is applied. It can be seen that our GRAN outperforms EDSR [9] and RDN [7] by a notable margin, obtaining the best deblurring performance on both EM and HIST datasets. Furthermore, our GRAN utilizes moderate-scale parameters and thus supports fast model inference. This observation indicates that it is a high-precision microscopy image deblurring method and also a highly practical one. On the other hand, our GRAN_GAN obtains lower PSNR/SSIM values. It is reasonable because GRAN_GAN uses perceptual and GAN losses.

4.4.2. Visual Results

We provide visual results in Fig. 4 and compare with methods listed in Tab. 2, on the EM (top two rows) and HIST (bottom two rows) datasets. As can be seen, our GRAN obtains comparable or even better visual results than those of EDSR [9] and RDN [7]. Such an achievement could be notable because EDSR and RDN are much larger networks than ours. Moreover, with adversarial learning, our GRAN_GAN displays visible superiority over others. In the EM images, for example, there are some noise-like details in the ground truth (i.e., HQ) images. Such details are blurred and can hardly be reconstructed by most methods. While the outputs of GRAN_GAN appear to be more faithful to the HQ ones. In the HIST images, there are many tiny tissues (e.g., blood vessels), which contain more structured details. Our GRAN_GAN achieves better visual results by recovering more such details. Such comparisons and observations illustrate the superiority of our proposed model in microscopy image deblurring tasks.

Table 3. Parameters and FLOPs comparison. To calculate FLOPs, we set the input size as $400\times400\times3$.

Method	EDSR [9]	RDN [7]	GRAN
# param. (M)	38.37	21.98	3.20
FLOPs (G)	6136.38	3514.59	145.72

Table 4. Running time (s) comparison on EM and HIST data.

Method	EDSR [9]	RDN [7]	GRAN
EM	12.07	9.26	0.95
HIST	6.09	4.59	0.54

4.4.3. Model Size and Computation Cost

In Tab. 3, we give the model size (i.e., parameter number), computation cost (i.e., FLOPs), and performance on the HIST dataset for each CNN-based deblurring method. We analyze the comparison of parameter numbers and FLOPs first. As shown in Tab. 3, the model sizes of EDSR [9] and RDN [7] are about 11 and 6 times larger than ours respectively. The FLOPs of EDSR and RDN are 42 and 24 times larger than ours. The FLOPs gap between our method and the compared ones becomes much larger than that in terms of model size. This is mainly because we introduce feature rescaling operation and extract deep features in the coarse scale.

4.4.4. Running Time

In Tab. 4, we compare the running time among each method. The average image resolution in EM and HIST test sets are 1344×1560 and 1024×1024 respectively. The average time of EDSR and RDN is about 12 and 9 times longer than ours in EM and HIST data. We further consider Tabs. 2 and 4 together. We can conclude that our GRAN achieves a much better trade-off between performance and running time.

5. CONCLUSION

We collect high-quality EM and HIST datasets and propose a GRAN model for high-throughput microscopy image deblurring. To investigate the non-local relationship in the features, we build the relationship graph among the deep feature points. We perform reasoning on the relationship graph with GCN and obtain a graph reasoning attention block (GRAB). It can be inserted in basic networks and could further contribute to feature representation ability and performance improvement. We further introduce adversarial training for perceptual enhancement. We demonstrate the lightweight, fast, and effective properties of our GRAN with extensive experiments.

Acknowledgments. This research is supported in part by NSF IIS-2124179 and NIH Project #5U24NS109102.

6. REFERENCES

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, 2014. 1

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *TPAMI*, 2016. 1

[3] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang, "Learning deep cnn denoiser prior for image restoration," in *CVPR*, 2017. 1, 3

[4] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017. 1, 2

[5] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *CVPR*, 2018. 1, 2

[6] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, "Residual dense network for image super-resolution," in *CVPR*, 2018. 1

[7] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, "Residual dense network for image restoration," *TPAMI*, 2020. 1, 3, 4

[8] He Zhao, Bingyu Yang, Lvchen Cao, and Huiqi Li, "Data-driven enhancement of blurry retinal images via generative adversarial networks," in *MICCAI*, 2019. 1

[9] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPRW*, 2017. 1, 3, 4

[10] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *CVPR*, 2018. 2

[11] Yunpeng Chen, Yannis Kalantidis, Jiashu Li, Shuicheng Yan, and Jiashi Feng, "a^2-nets: Double attention networks," in *NIPS*, 2018. 2

[12] Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap, "A simple neural network module for relational reasoning," in *NeurIPS*, 2017. 2

[13] Xinlei Chen, Li-Jia Li, Li Fei-Fei, and Abhinav Gupta, "Iterative visual reasoning beyond convolutions," in *CVPR*, 2018. 2

[14] Bolei Zhou, Alex Andonian, Aude Oliva, and Antonio Torralba, "Temporal relational reasoning in videos," in *ECCV*, 2018. 2

[15] Kunpeng Li, Yulun Zhang, Kai Li, Yuanyuan Li, and Yun Fu, "Visual semantic reasoning for image-text matching," in *ICCV*, 2019. 2

[16] Thomas N. Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017. 2

[17] Yunpeng Chen, Marcus Rohrbach, Zhicheng Yan, Yan Shuicheng, Jiashi Feng, and Yannis Kalantidis, "Graph-based global reasoning networks," in *CVPR*, 2019. 2

[18] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *ECCVW*, 2018. 2

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *NIPS*, 2014. 2

[20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *TIP*, 2004. 3

[21] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2014. 3