Self-Supervised Bulk Motion Artifact Removal in Optical Coherence Tomography Angiography

Jiaxiang Ren[1] Kicheon Park[2] Yingtian Pan[2] Haibin Ling[1*]
[1]Department of Computer Science [2]Department of Biomedical Engineering
Stony Brook University
{jiaxren,hling}@cs.stonybrook.edu {ki.park,yingtian.pan}@stonybrook.edu

Abstract

Optical coherence tomography angiography (OCTA) is an important imaging modality in many bioengineering tasks. The image quality of OCTA, however, is often degraded by Bulk Motion Artifacts (BMA), which are due to micromotion of subjects and typically appear as bright stripes surrounded by blurred areas. State-of-the-art methods usually treat BMA removal as a learning-based image inpainting problem, but require numerous training samples with nontrivial annotation. In addition, these methods discard the rich structural and appearance information carried in the BMA stripe region. To address these issues, in this paper we propose a self-supervised content-aware BMA removal model. First, the gradient-based structural information and appearance feature are extracted from the BMA area and injected into the model to capture more connectivity. Second, with easily collected defective masks, the model is trained in a self-supervised manner, in which only the clear areas are used for training while the BMA areas for inference. With the structural information and appearance feature from noisy image as references, our model can remove larger BMA and produce better visualizing result. In addition, only 2D images with defective masks are involved, hence improving the efficiency of our method. Experiments on OCTA of mouse cortex demonstrate that our model can remove most BMA with extremely large sizes and inconsistent intensities while previous methods fail.

1. Introduction

As a fast and non-invasive optical imaging technology with high spatiotemporal resolution, optical coherence tomography (OCT) is an emerging medical imaging modality not only for laboratory research but also for clinical applications [8]. Optical coherence tomography angiography (OCTA) has been reported to effectively diagnose and assess many retinal conditions, such as diabetic retinopathy [21], macular degeneration [23] and choroidal neovascular membrane [19]. Not only in ophthalmology, OCT has also been employed to visualize the prognosis of tumor vasculature for a better understanding of microenvironment [22]. Benefitting from high spatiotemporal resolution and non-invasive features, OCT has been utilized to study neurovascular changes and brain function *in vivo*.

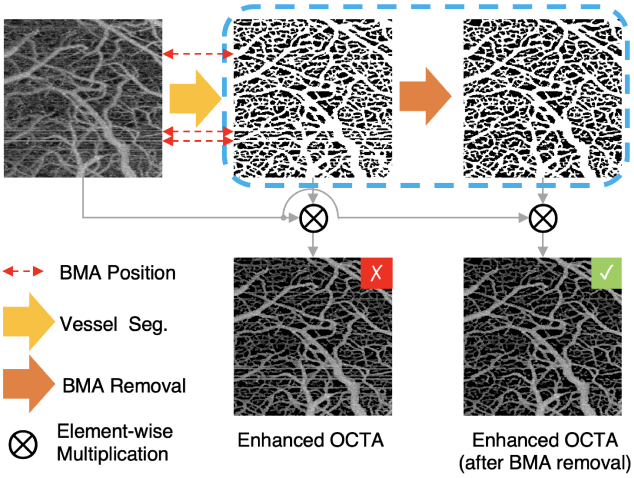


Figure 1. Illustration of OCTA affected by Bulk Motion Artifacts (BMA). Our work focuses on the BMA removal task in vessel mask (blue rectangle in dashed line). The enhanced OCTA after BMA removal has better visualization quality.

A typical OCTA image of mouse cortex is shown in Fig. 1. Previous approach [15] employed image enhancement to remove background noises for better qualitative and quantitative analysis. Specifically, a vessel mask was generated to highlight vasculature while suppressing the background noise. Unfortunately, this approach failed to remove Bulk Motion Artifact (BMA), which was caused by micromotion of the object, as highlighted with double arrows in Fig. 1.

BMA is one of the most frequent artifacts affecting many

*Corresponding author.

medical imaging modalities (CT, MRI and OCT), which seriously degrades image quality and affects quantitative analysis. Over decades of research, numerous models have been developed to remove BMA, but effective solutions for all circumstances remain missing. In OCTA, BMA causes a severely blurred image and usually appears as horizontal stripes of different widths with high-intensity in the center and low contrast in the surrounding area. As illustrated in Fig. 1, the horizontal stripes across OCTA are the areas affected by BMA. In the following sections, we use BMA and stripe (artifacts) interchangeably.

Previous approaches remove BMA with image registration or supervised denoising models, requiring duplicate scans or numerous training data with annotations. The current state-of-the-art model [14] inpaints the BMA-affected areas to generate a clear vasculature mask for enhancement. However, the inpainting model abandons the information in noisy areas so has limited capability to fill large gaps. Besides, lots of images with annotations are needed for training. We observed that, even if the intensity value of BMA is far stronger than normal area, there were still textures within BMA. Furthermore, BMA only affects vertical gradient rather than horizontal ones due to the OCTA imaging process. These features could serve as structural references of vasculature in noisy BMA area. However, there are three main challenges for BMA removal: (1) information within the motion artifacts is too noisy; (2) how to inject structure information from the BMA area into the recovery framework; and (3) OCTA data are limited and the pixel-level annotations are difficult to acquire.

In this paper, we propose a self-supervised BMA removal method to solve the issues above. First, the gradient-based structural information and appearance feature are extracted from BMA and injected into the model to capture more connectivity. Specifically, we keep the BMA-affected area, even the noisy one, as an input channel instead of discarding it directly. Furthermore, based on the observation that BMA mainly affects vertical gradients rather than horizontal ones, we introduce horizontal gradients to make our model aware of potential structures under BMA. Adding the gradient map makes the training stabler and convergence faster. With structural and appearance information and easily collected defective masks, we train our model in a self-supervised manner that the clear areas are used for training while the BMA areas are only for inference.

With the rich structural and appearance information carried in the BMA stripe region as reference, our model can remove large BMA and handle the background noise at the same time. Our BMA removal model is end-to-end and only requires 2D image with defective mask, which is more practical. Compared with other context-guided denoising models, our content-aware model can learn guidance directly from noisy images and is thus more robust. Experiments on a mouse cortex OCTA dataset demonstrate that our model can remove most BMA under different intensities and extreme area sizes while existing methods fail. The specific contributions of our paper are:

- We propose a new structural denoising model that can leverage the rich structural and appearance information carried in the BMA stripe region.
- Our model is trained in a self-supervised manner, and training does not involve manual annotation.
- For image enhancement, our method can alleviate BMA and background noises simultaneously, and thus significantly improve image quality.

2. Related Work

In this section, we will first discuss OCTA and BMA. Then, we will briefly introduce some general denoising methods followed by related OCT denoising methods.

2.1. OCTA and BMA in Awake Animal Study

OCT is a high-speed 3D imaging technique with micron-level resolution. Previous methods employ gradient-based filters [7, 11, 12] to generate vessel masks to highlight vasculature for better analysis. However, the image reconstruction is sensitive to micromotion caused by heartbeat, respiration, and mechanical vibration. Such motions lead to misalignment and phase noise which cause BMA. Severe BMA usually results in extremely high-intensity noise. Previous gradient-based approaches [7, 12] can not handle such artifacts because BMA has similar gradient information with vessel branches. Moreover, long-lasting BMA could result in a broad and blurred area, making it hard to correct. Last but not least, BMA is often accompanied by background noise, which makes it even harder to find vessels within the affected area. The interleaved background noise and BMA could lead to poor visualization and erroneous quantification.

2.2. General Denoising Method

BM3D [4] is a classic denoising model based on sparse representation. Inspired by the observation that natural images consist of repeated textures, BM3D groups related image patches and filters them together to shrink noise. It handles Gaussian noise well but fails to remove structural noise. After convolutional neural network (CNN) is firstly applied for the denoising task in [9], lots of works have been proposed to keep updating the state-of-the-art performance [3, 25, 29, 30]. Recently, Wu *et al.* [26] propose a single image dehazing method based on contrastive learning and achieves state-of-the-art performance on both synthetic and real-world datasets. Still, this method is trained

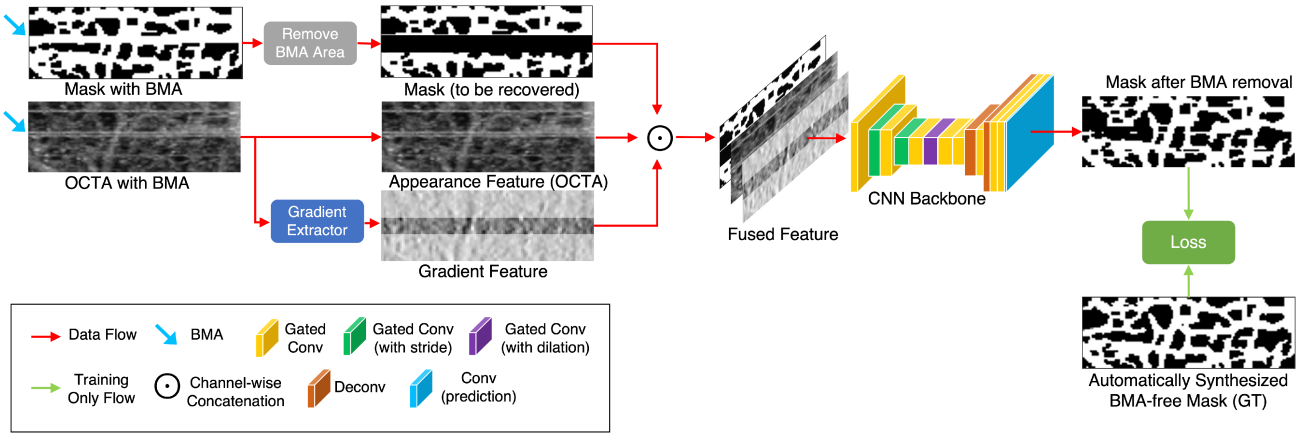


Figure 2. Framework of the proposed Content-Aware BMA Removal model (CABR).

in a supervised manner. It is not the case for some medical image modalities without ground truth.

More and more unsupervised learning based denoising methods [6, 10, 13] have been proposed and achieved promising performance. However, these approaches are not intended to remove structural noise. Broaddus *et al*. [1] employ structural kernels to remove spatially correlated noise in fluorescence microscope images. However, the performance of this method depends on the selection of blind mask, which is not fully automatic and needs prior knowledge for noise distribution.

2.3. OCT Denoising Method

Most traditional OCTA denoising methods [2, 24] are based on duplicate sampling, in which BMA-affected area is replaced by BMA-free area at the same location but different sampling period. The downside of such approaches is the considerably long sampling time, which reduces the practicability for researches and clinical diagnoses. Devalla *et al*. [5] propose a deep learning based OCT denoising model for retinal images. It is relatively easier to denoise OCT images of the retina than other tissues, such as mouse bladder and cortex, because of shallower imaging depth, less background noise, and BMA. Li *et al*. [16] remove BMA in an enhancement approach. Firstly, this method applies optimally oriented flux (OOF) [11, 12, 15] to generate vessel mask with defects in BMA areas and then employs tensor voting [18] to restore the defective mask. Finally, an OCTA image is enhanced with the corrected mask for BMA removal. However, this method may fail to remove moderate and severe BMA.

The current state-of-the-art model [14] removes BMA with the combination of a BMA detection module, a vessel segmentation module and a mask inpainting module [28]. The BMA detection module predicts BMA affected rows in OCTA. The segmentation module takes OCTA with BMA areas removed as input and produces a corrupted mask. After that, the mask inpainting module takes the probability map of the segmentation module as input and fills the missing mask. Manually annotated masks are used to train the segmentation module and inpainting module. The whole pipeline [14] is relatively ad-hoc and the performance depends on every step. Furthermore, numerous 2D and 3D training data are involved in training. Last but not least, this model learns only from surrounding BMA-free areas while ignoring the structural information within noisy data. So this context-based model often fails when dealing with severe BMA.

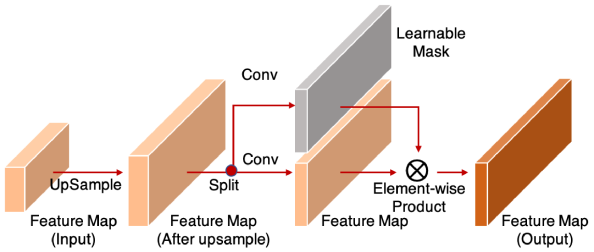


Figure 3. Deconvolution layer with GatedConv.

3. Method

In this section, we will start with problem formulation and challenges. Then, to deal with the context learning issue, we will discussing two sets of information injection and the proposed Content-Aware BMA Removal Model (CABR). Finally, we will illustrate a self-supervised training strategy to solve data insufficiency.

3.1. Problem Formulation

We first define the BMA-affected OCTA image as $\boldsymbol{I} \in \mathbb{R}^{H \times W}$, the defective vasculature mask $\boldsymbol{M} \in \{0, 1\}^{H \times W}$,

the Gradient Statistics $\boldsymbol{G} \in \mathbb{R}^{H \times W}$ (See Sec. 3.2), and the corresponding row label $\boldsymbol{l} \in \{0,1\}^H$, where W and H are image width and height, respectively. $\boldsymbol{l}_i = 0$ means the i-th row of $\boldsymbol{I}$ is clear while $\boldsymbol{l}_i = 1$ for BMA. For each $\boldsymbol{I}$, the corresponding mask $\boldsymbol{M}$ is automatically generated with OOF [11, 12]. The BMA affected region can be discerned easily and the row label $\boldsymbol{l}$ indicates which rows in image are affected by BMA. The proposed BMA removal framework takes $(\boldsymbol{I}, \boldsymbol{M}, \boldsymbol{G}, \boldsymbol{l})$ as input and outputs the vasculature prediction $\boldsymbol{M}'$ with BMA removed:

$$\boldsymbol{M}' = f(\boldsymbol{M}, \boldsymbol{I}, \boldsymbol{G}, \boldsymbol{l}). \tag{1}$$

The main issue of previous context-guided inpainting models is that little information from noisy area is automatically learned. Therefore these surrounding-based generative models produce plausible but unconvincing results, especially for large missing areas. But the certainty is just what medical image researchers demand. So the information within noisy area should be utilized to improve the prediction confidence.

Data insufficiency is another issue to apply supervised methods in medical images sometimes. It is difficult or even impossible to get noise-free or ground truth for some imaging modalities. Supervised methods may not work well under this condition. So we should leverage any available information to solve the data insufficiency. These two issues will be solved in the following parts.

3.2. Content-Aware BMA Removal

The uncertainty of context-guided inpainting models is the main drawback for the tasks demanding accurate results rather than plausible ones. We have observed that, even with heavy noise, some structure and texture information exist in BMA affected areas. To leverage such vessel clues, we introduce two information injection approaches, *e.g.* Gradient Statistics (GS) and Appearance Feature (AF). The proposed CABR can automatically learn reliable references from the injected information for BMA removal. The framework of CABR is illustrated in Fig. 2.

Gradient Statistics (GS). BMA usually causes high-intensity horizontal stripes in OCTA images. We observed that the horizontal gradients are hardly affected by BMA so we introduce the horizontal gradients as the structural information for CABR. Specifically, we apply the vertical Sobel operator in BMA areas to extract GS. The absolute value of response is injected into the model to capture more reference for BMA removal. Sobel operator is linear and efficient so only marginal computational cost is involved.

Appearance Feature (AF). GS provides a good reference for branch vessels but may not discern capillaries at relatively low intensities. Inspired by the deep denoising models, we define the BMA-affected areas as AF and keep AF, even noisy ones, as the additional input. With sufficient training samples, the BMA removal model can automatically learn the texture information to capture better vessel connectivities. The generation of training samples is discussed in Sec. 3.3.

Model Design. To overcome the limitations of context-guided inpainting model, we propose the content-based CABR for BMA removal in OCTA. The framework is illustrated in Fig. 2. With the two sets of information injection, our model can learn from both the content within BMA and the context outside BMA. It is thus more comprehensive and robust.

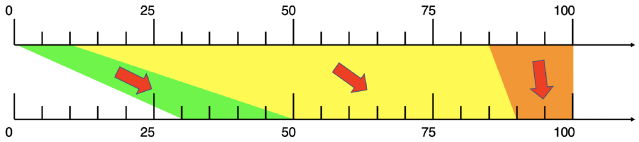
The imbalance of BMA sizes is another challenge for BMA removal. Most BMAs are thin stripes and are easy to remove while severe BMAs, on the contrary, are only small part of dataset but harder to remove. For previous context-guided inpainting models, the imbalanced distribution leads to performance gap between gentle and severe BMAs. These models often fail to fill in extra wide stripes due to lacking both references and hard training samples.

Different from previously free-form inpainting frameworks [17,20,28], our CABR focuses on correcting the vasculature in the centerline areas of input image. Reflection padding is used for the stripe located near boundary. This schema reduces the learning cost and model complexity. Furthermore, we can select more hard cases to train CABR so that it works better when dealing with these broadly missing areas.

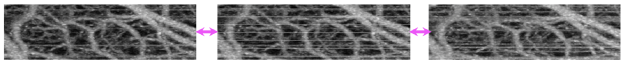
Model Architecture. CABR has an encoder-decoder CNN backbone, which stacks gated convolution layers (GatedConv) [28] to handle the stripe with various width in a noisy OCTA. As illustrated in Fig. 2, the backbone consists of 10 GatedConv layers [28], 1 GatedConv layer with dilation [27], 2 DeConv layers and 1 regular Conv layer with *sigmoid* activation function for prediction. GatedConv is a kind of attention convolution that learns a dynamic feature gate for each channel and each spatial location. We choose GatedConv due to its SOTA performance in inpainting. In addition, it saves around 50% trainable parameters than regular convolution with the same channel number. Considering the relatively small dataset, such light-weight module is preferred to reduce overfitting. Deconvolution used in CABR is an upsampling layer followed by a GatedConv, as illustrated in Fig. 3.

3.3. Self-supervised Training

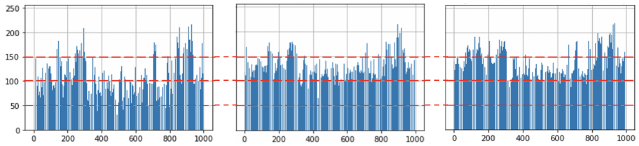
As we mentioned before, data insufficiency is one of the main challenges in OCTA. Here, we train our model in a self-supervised way to resolve problem. We make use of the majority BMA-free areas to acquire acceptable training masks at low cost and generate as many as noisy images for training. With these easily collected masks and synthetic noisy images, we can train the BMA removal model in a self-supervised manner.



(a) The schema of intensity mapping used in AdBMA.



(b) OCTA images.



(c) Intensity values of the row marked by double arrows in (b).

Figure 4. In (b) (c), from left to right: clear images, synthetic images and images affected by BMA.

Ground Truth Mask. The ground truth of vasculature mask for *in vivo* OCTA is either impossible or very difficult to acquire. Since OOF [12] can produce relatively accurate vessel mask for the most area (96.4%) without BMA, we resort to this defective mask with minimally additional manual annotation as the ground truth. To be specific, we only manually annotate the remaining 3.6% BMA areas and combine defective mask in BMA-free areas to get the vasculature mask. This approach can produce high-recall masks at low-cost. The manually annotated part is only involved in testing and is not used in training. All these masks are verified by experts.

BMA Synthesis. We propose the Adaptive BMA Synthesis (AdBMA) approach to generate noisy images from BMA-free areas. After inspecting the data distributions, we find that BMA has different effects on high and low intensity pixels. In other words, BMA mainly uplifts the low-intensity pixels while keeping the maximums unchanged. So we employ a remapping schema to mimic the BMA influence on clear images. Given a clear image $\boldsymbol{I}$ and two sets of hyperparameters $(\mathcal{P}_{\text{low}}, \mathcal{P}_{\text{high}})$ and $(\mathcal{P}_{\text{base}}, \mathcal{P}_{\text{low}'}, \mathcal{P}_{\text{high}'})$, we synthesize BMA in four steps: (1) randomly map the intensities below $\mathcal{P}_{\text{low}}$ into $[\mathcal{P}_{\text{base}}, \mathcal{P}_{\text{low}'})$ (half-open interval); (2) map the intensities between $\mathcal{P}_{\text{low}}$ and $\mathcal{P}_{\text{high}}$ into $[\mathcal{P}_{\text{low}'}, \mathcal{P}_{\text{high}'})$; (3) map the intensities above $\mathcal{P}_{\text{high}}$ into $[\mathcal{P}_{\text{high}'}, \max(\boldsymbol{I}))$; and (4) clip intensities within image maximum (usually 255) to avoid overflow. Gaussian noise is added in steps (2) and (3) to ensure uncertainty. The remapping schema is illustrated in Fig. 4a.

The synthesized noisy images have similar distribution with the real BMA, but more importantly they have corresponding ground truth masks. The benefits of synthesizing BMA are twofold: (1) we can generate as many plausible noisy images as needed; and (2) it fits our framework and enables learning from noisy data.

Fig. 4b shows a BMA-free OCTA image, an image with synthetic BMA and an image with real BMA. Moreover, Fig. 4c is the bar plot of intensity values for the row marked by double arrows in Fig. 4b. We can see that the synthetic OCTA image has not only plausible appearance but also similar intensity distribution. Thus the distribution gap between training and testing data is minimized.

Training Scheme. The train/val splitting is area-based, which is different from the typical supervised approach. Specifically, every image is split into the BMA-free (clear) area for training and the BMA-affected (noisy) area for evaluation. During training, only the loss from the clear area with synthetic BMA is used. During testing, only the Dice score in the BMA-affected area is measured. Thus the model works in a *self-supervised* way and can utilize as many training samples as possible.

4. Experiments

We evaluate the proposed BMA removal framework on a mouse cortex OCTA dataset. The results are compared with the state-of-the-art motion correction model [14] and a GatedConv-based inpainting model [28]. We re-implement the GatedConv-based model for inpainting the stripe region.

4.1. Dataset

We collect 39 OCTA images of mouse cortex with vessel masks. The noise level, defined as (BMA-affected area/Total area), varies throughout the dataset. Note that, not all images have BMA noise. We validate on the BMA-affected areas from 14 OCTAs. According to the noise level, OCTAs in the validation set are further classified as the easy set (noise$<2\%$, 4 OCTAs), medium set ($2\leq$noise$<4\%$, 7 OCTAs) and hard set (noise$\geq 4\%$, 3 OCTAs). We further collect 6 samples to extend the validation set. The BMA removal performance is measured by the Sørensen–Dice coefficient (Dice) in the rest of the paper, unless otherwise specified.

Training Details. The hyperparameters for BMA synthesis are $(\mathcal{P}_{\text{low}} = \mathcal{P}_{10\%}, \mathcal{P}_{\text{high}} = \mathcal{P}_{85\%})$ and $(\mathcal{P}_{\text{base}} = \mathcal{P}_{30\%}, \mathcal{P}_{\text{low}'} = \mathcal{P}_{50\%}, \mathcal{P}_{\text{high}'} = \mathcal{P}_{90\%})$ where $\mathcal{P}_{k\%}$ is the k-th percentile of input image $\boldsymbol{I}$. As for the re-implemented GatedConv-based inpainting model, we randomly select $2\sim20\%$ rows within each patch for training. As for CABR, we select center rows in each patch with width randomly from 1 to 11 for training. Furthermore, we use data augmentation, such as random cropping, horizontal and vertical flipping, on the training set.

Implementation. The CNN backbone used for all experiments is illustrated in Fig. 2. All convolutional layers use 3×3 kernels. The downsampling rate, stride rate and dilation rate are set to 2. The backbone has 16 feature maps in

Table 1. Performance on the mouse cortex OCTA dataset. GatedConv* is re-implemented and trained on the OCTA dataset. "All*" is the extended validation set with "All" and 6 extra samples.

Method	Easy	Medium	Hard	All	All*
OOF [12]	45.76	42.70	43.72	43.79	42.93
Li *et al.* [14]	68.83	68.51	60.70	66.93	70.21
GatedConv* [28]	77.41	73.48	74.37	74.80	77.03
CABR$_{(SynBMA)}$	84.99	81.98	79.22	82.24	83.92
CABR$_{(SynBMA,GE)}$	**85.71**	**82.67**	**79.69**	**82.90**	**84.54**

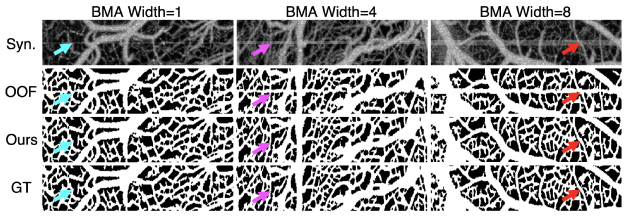


Figure 5. Denoising result for synthesized BMA of width {1,4,8}.

the first level, which increases up to 64 when the level gets deeper. All models are trained with Dice loss and Adam optimizer for 1,200 epochs. The learning rate starts from 10^{-4} and halves if no improvement is observed for 20 epochs. We used 256×496 input size and batch size 12 for GatedConv training. We used 64×496 input size and batch size 48 for CABR training. All models are trained on a single NVIDIA GeForce GTX Titan Xp GPU.

4.2. BMA Removal Results

To evaluate the performance of the proposed CABR framework, we compare it with the state-of-the-art OCTA motion correction model by Li *et al.* [14]. The result of OOF [12] is the baseline. To the best of our knowledge, there are no other publications that pertain directly to BMA removal in OCTA. So we also re-implement a GatedConv-based inpainting model [28], for inpainting the BMA-affected region. The re-implemented inpainting model has a similar amount of parameters as CABR and is trained with the same setting.

Quantitative Result. The Dice scores on the mouse cortex OCTA dataset of different methods are in Tab. 1. The result shows a clear improvement of our method over the OOF baseline. Besides, our approach significantly outperforms the state-of-the-art model [14] by 15.97% and 14.33% in the validation set and the extended validation set, respectively. The proposed CABR also achieves the best performance in all subsets, which evidences the effectiveness and advantage of content-based learning. Even in the hard subset with larger and more dense BMA across OCTA, our CABR surpasses the re-implemented GatedConv by 5.32%, suggesting that CABR can learn more connectivities from noise input for BMA removal. The experiment results also show that adding GS further improves the performance of CABR.

Qualitative Result. The BMA removal results for the synthesized BMA of width 1, 4 and 8 are illustrated in Fig. 5. It shows that our method can handle severe BMA and recover vasculature while the baseline [12] fails.

The experiment results for two validation samples are illustrated in the first and third rows of Fig. 6. We compare our method with OOF [12] (only for reference, no BMA removal effect), Li *et al.*'s method [14], our re-implemented GatedConv-based inpainting model [28]. We can see from the first and third rows of Fig. 6 that our method removes most of the false positive masks caused by BMA while achieving the best vasculature connectivity at the same time. Li *et al.*'s method [14] can also remove some BMA but the low recall is the main drawback. Specifically, thin BMA stripes can be restored while thicker ones can not. On the contrary, the re-implemented GatedConv shows better connectivity than Li *et al.*'s method [14] but some BMAs are remaining. The cyan and pink arrows highlight more details in the zoom-in panels of Fig. 6.

The second OCTA is harder to handle because the stripes are relatively broader and more densely distributed. From the third and fourth rows of Fig. 6, we can see that Li *et al.*'s method [14] can not fill the gap in the second zoom-in area and the GatedConv-based inpainting model has more artifacts. Our method produces a complete mask with better connectivity and fewer artifacts.

We also compare with the self-supervised StructN2V [1] and the results in Fig. 7 suggest that StructN2V is not suitable for our task: incorrect structure removal (cyan/red arrow), suffering from wide BMA (pink arrow), *etc*.

OCTA Enhancement. OCTA enhancement based on vesselness mask is a regular procedure for biomedical analysis. Usually, this procedure highlights vasculature while suppressing background noise, and thus improves contrast. The enhanced OCTA is generated by multiplying the raw OCTA with corresponding vessel mask. The second and fourth rows of Fig. 6 illustrate the enhancement results of different approaches on two sets of BMA-affected OCTAs. Our method removes most of the BMA artifacts and produces an OCTA with stronger contrast, as illustrated in the zoom-in panels as the dashed rectangle areas. From the second OCTA, where BMA in the bottom area degrades image quality severely, we can see that our method still removes most of the BMAs while preserving vessel structures, thus showcasing the effectiveness of our proposed method.

4.3. Ablation Study

To study the effects of each module, we conduct the ablation experiments on the mouse cortex OCTA dataset. All experiments are based on the same training setting in

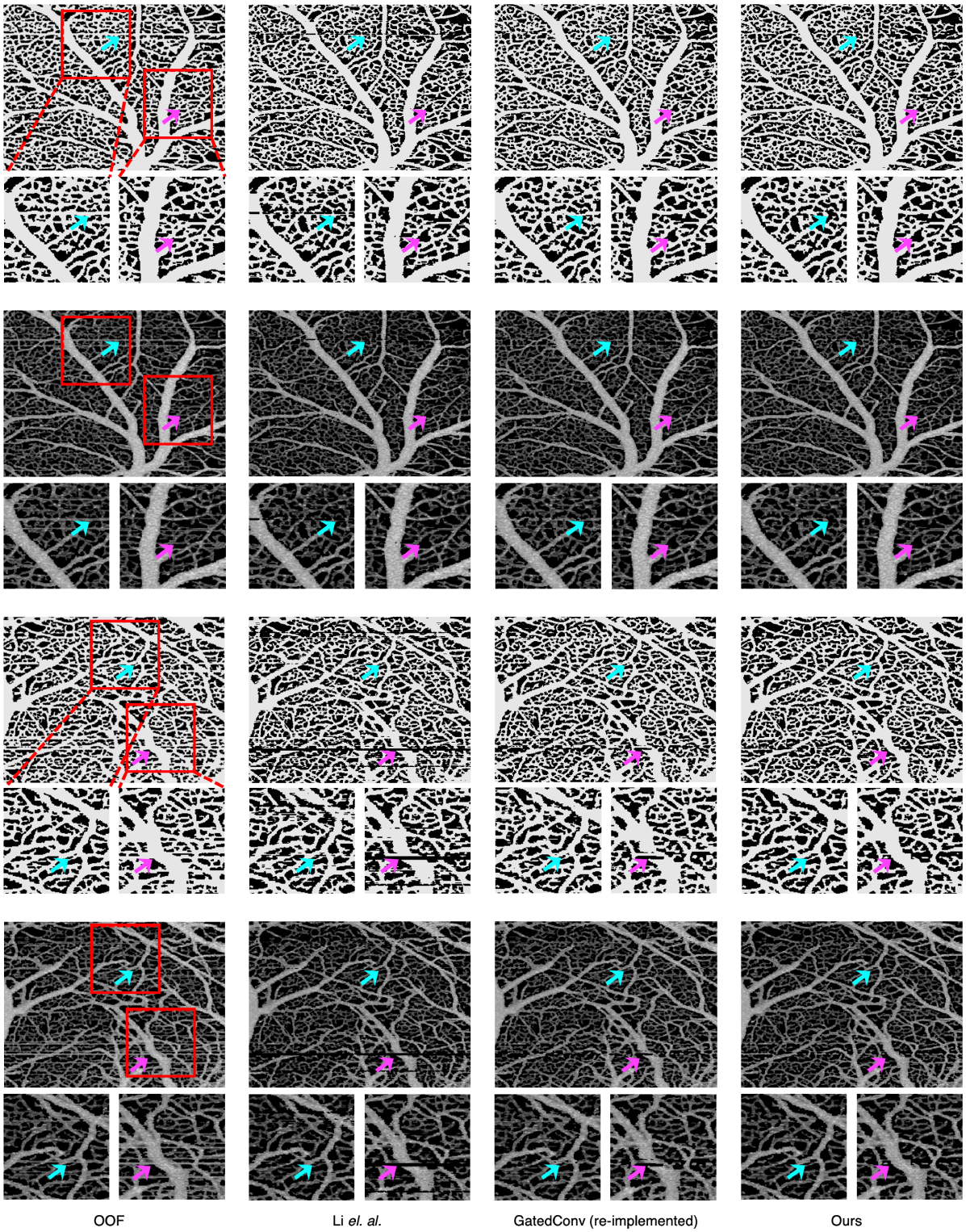


Figure 6. Two sets of OCTA masks and enhanced images of different methods. The first and third rows are the masks of OOF [12], Li's method [14], our re-implemented GatedConv-based inpainting model [28] and the proposed CABR, respectively. The corresponding enhanced images are in the second and fourth rows. The zoom-in results within the rectangles in dashed line are also illustrated for better comparison. The cyan and pink arrows highlight more details in the zoom-in panels.

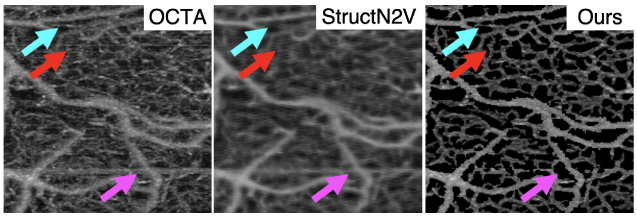


Figure 7. The denoising result of StructN2V [1] and our method.

Table 2. Ablation study of CABR on the mouse cortex OCTA dataset. "Abs" means the absolute value. "Abs-BMA" means absolute GS only in BMA. "Conv" means the regular convolution.

Appearance		Gradient Statistics			Module		Dice↑(%)
Gauss	AdBMA	Naive	Abs	Abs-BMA	Conv	Gated	
✓						✓	81.29
	✓					✓	82.24
	✓	✓				✓	80.08
	✓		✓			✓	82.38
	✓		✓		✓		82.57
	✓			✓		✓	**82.90**

Sec. 4.1. For AF injection, we compare the proposed AdBMA with a naive implementation that adds Gaussian noise. Both methods generate high-intensity stripes like BMA artifacts. However, the Gaussian method does not have a blurring effect so the result is less plausible. For GS injection, the *naive* stands for the gradient of Sobel operator. *Abs* and *Abs-BMA* stand for the absolute value of Sobel results on the whole image and only the BMA area, respectively. To report the contribution of GatedConv module, we also conduct an ablative study and find that GatedConv does outperform regular convolution by 0.33% with only about 50% trainable parameters.

Ablation of Information Injection. As shown in Tab. 2, we first investigate the effectiveness of AF injection with synthetic BMA. "Gauss" means adding Gaussian noise directly to generate training samples. Our proposed AdBMA outperforms the Gaussian-based synthetic approach by 1.66%, which verifies that (1) texture information within BMA, even noisy, still benefits the recovery of vasculature; and (2) the blurring effect from AdBMA improves the performance. We also evaluate the effectiveness of GS as shown in Tab. 2. The performance of CABR is improved by 2.30% for Abs and 2.82% for Abs-BMA, respectively, which states the validity of GS in BMA removal. During experiments, we find that GS leads to faster convergence. CABR benefits less from GS than the synthetic BMA feature. One possible reason is that convolution layers may have learned similar or even more powerful kernels than the Sobel operator.

Table 3. Ablation study of CABR for different model sizes on the mouse cortex OCTA dataset. # Channels is the channel number for the first layer of CNN backbone. # Parameters is the number of trainable parameters for the whole model.

# Channels	# Parameters	Dice↑(%)
8	33.7K	78.70
16	133.9K	**82.90**
32	534.0K	82.44

Ablation of Model Size. With the same modules in the CNN backbone, we investigate how the model size influences the performance of BMA removal. Tab. 3 lists the result of CABR with different sizes. A larger channel number means more parameters and a larger model size. The 8-channel model only achieves 78.70% Dice coefficient, which is less than 16-channel and 32-channel models. The results demonstrate that 8-channel is insufficient to capture all vessel information. Also, a larger model size, *i.e.*, 32-channel in the first layer, does not boost the performance. The 16-channel model outperforms the 32-channel by 0.46% with about 75% less trainable parameters, suggesting that the OCTA dataset may not support training the model with over 140K parameters.

5. Conclusion

In this paper, we propose a self-supervised content-aware model to remove BMA and improve image quality in OCTA. First, the GS-based structural information and AF are extracted from the BMA area and injected into the model to capture more vasculature connectivity. Second, we train the model in a self-supervised manner with easily collected defective masks. With the rich structural and appearance information carried in the BMA stripe region as references, our model can remove BMAs and produce a better vasculature mask. Compared with other context-guided inpainting models, our content-aware model can learn guidance directly from noisy images and thus is more robust. Experiments on a mouse cortex OCTA dataset demonstrate that our model can remove most BMAs, even the huge and inconsistent ones, while existing methods fail. The current BMA synthesis approach is manually designed and still has room for improvement. Incorporating domain knowledge of BMA formulation or introducing a generative model for synthesis are potential research directions.

Acknowledgment. This work was supported in part by NSF Grants 1814745 and 2006665, and NIH grants R01DA029718 and RF1DA048808.

References

[1] Coleman Broaddus, Alexander Krull, Martin Weigert, Uwe Schmidt, and Gene Myers. Removing structured noise with self-supervised blind-spot networks. In *ISBI*, 2020. 3, 6, 8

[2] Acner Camino, Miao Zhang, Simon S Gao, Thomas S Hwang, Utkarsh Sharma, David J Wilson, David Huang, and Yali Jia. Evaluation of artifact reduction in optical coherence tomography angiography with real-time tracking and motion correction technology. *Biomed. Opt. Express*, 7(10):3905–3915, 2016. 3

[3] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *TPAMI*, 39(6):1256–1272, 2016. 2

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *TIP*, 16(8):2080–2095, 2007. 2

[5] Sripad Krishna Devalla, Giridhar Subramanian, Tan Hung Pham, Xiaofei Wang, Shamira Perera, Tin A Tun, Tin Aung, Leopold Schmetterer, Alexandre H Thiery, and Michaël JA Girard. A deep learning approach to denoise optical coherence tomography images of the optic nerve head. *Sci. Rep.*, 9(1):1–13, 2019. 3

[6] Shreyas Fadnavis, Joshua Batson, and Eleftherios Garyfallidis. Patch2self: denoising diffusion mri with self-supervised learning. In *NeurIPS*, 2020. 3

[7] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *MICCAI*, 1998. 2

[8] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991. 1

[9] Viren Jain and Sebastian Seung. Natural image denoising with convolutional networks. In *NeurIPS*, 2008. 2

[10] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *CVPR*, 2019. 3

[11] Max WK Law and Albert CS Chung. Three dimensional curvilinear structure detection using optimally oriented flux. In *ECCV*, 2008. 2, 3, 4

[12] Max WK Law and Albert CS Chung. An oriented flux symmetry based active contour model for three dimensional vessel segmentation. In *ECCV*, 2010. 2, 3, 4, 5, 6, 7

[13] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *ICML*, 2018. 3

[14] Ang Li, Congwu Du, and Yingtian Pan. Deep-learning-based motion correction in optical coherence tomography angiography. *J. Biophotonics*, 2021. 2, 3, 5, 6, 7

[15] Ang Li, Jiang You, Congwu Du, and Yingtian Pan. Automated segmentation and quantification of oct angiography for tracking angiogenesis progression. *Biomed. Opt. Express*, 8(12):5604–5616, 2017. 1, 3

[16] Ang Li, Guang Zeng, Congwu Du, Huiping Zhang, and Yingtian Pan. Automated motion-artifact correction in an octa image using tensor voting approach. *Appl. Phys. Lett.*, 113(10), 2018. 3

[17] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *ECCV*, 2018. 4

[18] Gérard Medioni, Chi-Keung Tang, and Mi-Suen Lee. Tensor voting: Theory and applications. In *RFIA*, 2000. 3

[19] Lea Querques, Chiara Giuffrè, Federico Corvi, Ilaria Zucchiatti, Adriano Carnevali, Luigi A De Vitis, Giuseppe Querques, and Francesco Bandello. Optical coherence tomography angiography of myopic choroidal neovascularisation. *Br. J. Ophthalmol.*, 101(5):609–615, 2017. 1

[20] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li. Structureflow: Image inpainting via structure-aware appearance flow. In *ICCV*, 2019. 4

[21] Wasim A Samara, Abtin Shahlaee, Murtaza K Adam, M Ali Khan, Allen Chiang, Joseph I Maguire, Jason Hsu, and Allen C Ho. Quantification of diabetic macular ischemia using optical coherence tomography angiography and its relationship with visual acuity. *Ophthalmology*, 124(2):235–244, 2017. 1

[22] Benjamin J Vakoc, Ryan M Lanning, James A Tyrrell, Timothy P Padera, Lisa A Bartlett, Triantafyllos Stylianopoulos, Lance L Munn, Guillermo J Tearney, Dai Fukumura, Rakesh K Jain, et al. Three-dimensional microscopy of the tumor microenvironment in vivo using optical frequency domain imaging. *Nat. Med.*, 15(10):1219–1223, 2009. 1

[23] Nadia K Waheed, Eric M Moult, James G Fujimoto, and Philip J Rosenfeld. Optical coherence tomography angiography of dry age-related macular degeneration. *Dev. Ophthalmol.*, 56:91–100, 2016. 1

[24] David Wei Wei, Anthony J Deegan, and Ruikang K Wang. Automatic motion correction for in vivo human skin optical coherence tomography angiography through combined rigid and nonrigid registration. *J. Biomed. Opt.*, 22(6), 2017. 3

[25] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods*, 15(12):1090–1097, 2018. 2

[26] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, 2021. 2

[27] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *ICLR*, 2016. 4

[28] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *ICCV*, 2019. 3, 4, 5, 6, 7

[29] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *TIP*, 27(9):4608–4622, 2018. 2

[30] Yuqian Zhou, Jianbo Jiao, Haibin Huang, Yang Wang, Jue Wang, Honghui Shi, and Thomas Huang. When awgn-based denoiser meets real noises. In *AAAI*, 2020. 2