2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782262

OCTA-based AMD Stage Grading Enhancement via Class-Conditioned Style Transfer

Haochen Zhang[1], Anna Heinke[2], Krzysztof Broniarek[3], Carlo Miguel B. Galang[2], Daniel N. Deussen[2], Ines D. Nagel[2]
Katarzyna Michalska-Małecka[3], Dirk-Uwe G. Bartsch[2], William R. Freeman[2], Truong Q. Nguyen[1], Cheolhong An[1]
[1]Electrical and Computer Engineering Department, UC San Diego, USA
[2]Jacobs Retina Center, Shiley Eye Institute, UC San Diego, USA
[3]Ophthalmology Department, Medical University of Gdańsk, Poland
{haz035, tqn001, chan}@ucsd.edu {aheinke, dbartsch, wrfreeman}@health.ucsd.edu

***Abstract*—Optical Coherence Tomography Angiography (OCTA) is a promising diagnostic tool for age-related macular degeneration (AMD), providing non-invasive visualization of sub-retinal vascular networks. This research explores the effectiveness of deep neural network (DNN) classifiers trained exclusively on OCTA images for AMD diagnosis. To address the challenge of limited data, we combine OCTA data from two instruments—Heidelberg and Optovue—and leverage style transfer technique, CycleGAN, to convert samples between these domains. This strategy introduces additional content into each domain, enriching the training dataset and improving classification accuracy. To enhance the CycleGAN for downstream classification tasks, we propose integrating class-related constraints during training, which can be implemented in either supervised or unsupervised manner with a pretrained classifier. The experimental results demonstrate that the proposed class-conditioned CycleGAN is effective and elevates DNN classification accuracy in both OCTA domains.**

I. Introduction

The retina, a vital part of the human eye, transforms light into neural signals through a photosensitive layer of optic nerve tissue. At the core of retina lies the macula, essential for detecting color and light intensity, without which detailed and vivid color perception is compromised [1]. Unfortunately, various eye diseases, such as age-related macular degeneration (AMD), choroidal neovascularization (CNV), and diabetic macular edema (DME), can impact the health of the macula. AMD, in particular, is a prevalent cause of vision loss, contributing to a significant health concern [2]. Since the progression of AMD often associates with abnormal vascular changes, diagnostic tools detecting pathological vessels is crucial for optimal treatment and vision preservation.

Optical coherence tomography angiography (OCTA), an emerging imaging technique, facilitates visualization of retinal and inner choroidal circulation without dye injection, becoming crucial for detecting CNV lesions in neovascular AMD [3]. OCTA offers quantitative metrics like foveal avascular zone area and capillary network density, aiding in the assessment of retinal diseases and morphological changes in vascular networks responding to treatments [4]. Although not yet a gold standard in AMD treatment decision-making, the non-invasive nature of OCTA and its ability to detect subclinical changes position it as an ideal diagnostic and monitoring tool.

Recent research [5] highlighted the limited capability of en-face OCTA images in identifying CNV, and OCTA alone struggled with accurate CNV activity determination. To unlock the full potential of OCTA in clinical applications, more effective analysis methods are needed. Given the superior performance of deep neural networks (DNNs) in surpassing human abilities in natural image classification [6], our study focuses on developing deep classifiers for AMD stage grading *based solely on OCTA images*. These classifiers directly learn informative features from the data, proving more efficient than humans in discerning correlations within multiple projections, en-face OCTA images at different depths [7].

To enhance the performance of a deep classifier in AMD stage grading, a recent investigation [8] scrutinized en-face OCTA images produced by the Heidelberg OCTA instrument, identifying a substantial impact from segmentation errors of retina layers. In [8], two strategies were proposed to handle these errors: one involves direct analysis of the 3D raw OCTA volume, thereby mitigating the influence of segmentation errors, while the other entails expanding the training dataset to encompass the data distribution affected by segmentation errors. Given the impracticality of exporting 3D raw data for OCTA instruments like Optovue and Topcon, hindering the broader application of a 3D classifier, this paper explores the latter approach. To facilitate this exploration, we amassed an additional OCTA dataset generated by the Optovue OCTA instrument and trained deep classifiers using cross-instrument data, leveraging style transfer techniques.

In detail, as the two datasets comprise distinct patient groups, we employ the CycleGAN framework [9] for unpaired style transfer. Utilizing a well-trained CycleGAN translator facilitates the conversion of Heidelberg samples to Optovue and vice versa, introducing additional contents into each domain. This process also enhances sample diversity within each domain through cycle consistency and identity outputs. Unlike many existing style transfer methods that focused on realistic visual quality, our primary objective is enhancing clas-

* This work is supported by grant NIH R01 EY033847-02.

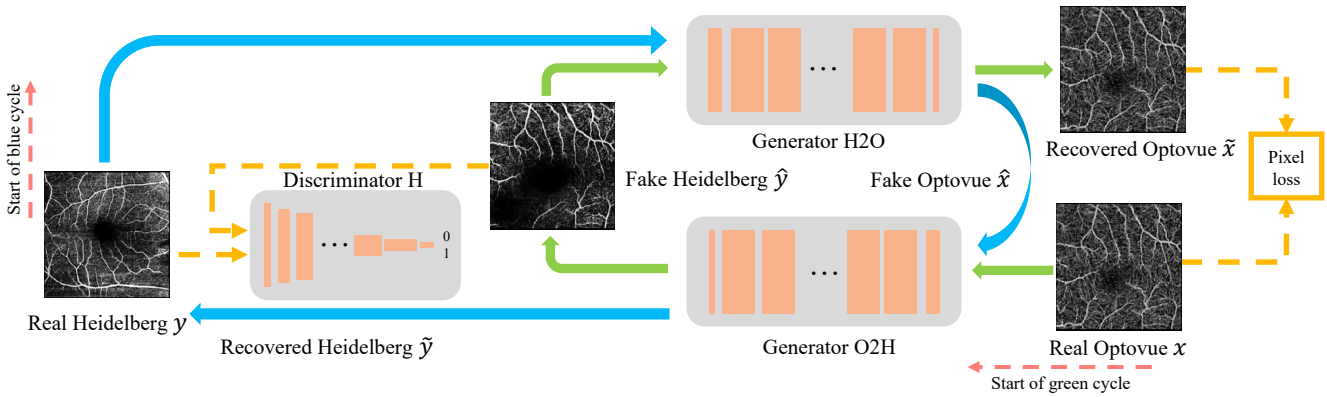


Fig. 1. Illustration of our framework utilizing CycleGAN to transfer samples between Heidelberg and Optovue domains. For simplicity, only one channel of OCTA projection is displayed, and the identity loss is omitted. For more details please refer to Fig. 2

sification accuracy. To optimize machine recognition quality of transferred images, we introduce additional constraints using a pretrained deep classifier. This constraint could be implemented in either fully-supervised or unsupervised manners: in domains with available labels, we ensure the transferred image belongs to the same class; in the absence of labels, we aim to make the representatives of the image before and after transferring as similar as possible. Experimental results demonstrate that the CycleGAN translator effectively enhances classification performance in each domain, and our proposed class-conditioned approach further boosts the accuracy.

II. Related work

1) OCTA analysis: As OCTA enables the visualization of the retinal vascular network, the majority of OCTA-based studies have concentrated on segmentation tasks [10], [11]. For example, Alam et al. [12] utilized U-Net for artery-vein classification, employing transfer learning to address limited dataset challenge. Besides segmentation, some OCTA studies delve into deep learning-based classification, such as Le et al. [13] which used the VGG16 network for diabetic retinopathy stage grading. Lin et al. [14] incorporated boundary shape and distance maps for simultaneous classification and segmentation. To the best of our knowledge, only one previous study focused on AMD diagnosis using OCTA exclusively. In Zhang et al.'s work [8], they investigated the impact of layer segmentation errors and proposed a 3D OCTA volume network with a 2D EfficientNet backbone, trained with additional en-face OCTA supervision. Diverging from this approach, our paper takes a unique route by collecting a new dataset using a different instrument to broaden training data diversity and, consequently, enhance accuracy.

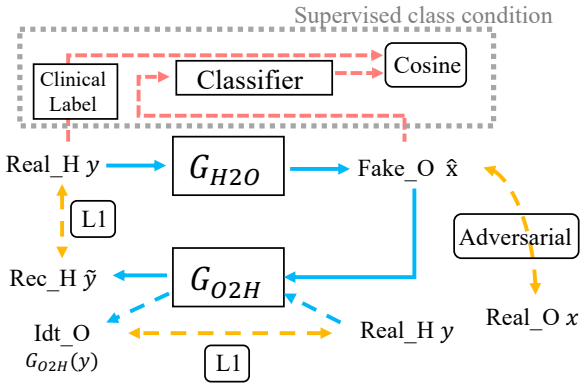
2) Neural style transfer: Style transfer aims to manipulate the visual appearance of an image by altering its thematic style while preserving its content [15]. As a pioneer work, Gatys et al. [16] introduced a seminal neural style transfer algorithm, which utilized DNN to separate and recombine content and style representations. Then, numerous popular works have been published to improve in both speed and scalability. For example, Johnson et al. [17] proposed an image transformation network with perception loss that achieved real-time style transfer. Ulyanov et al. [18] found instance normalisation significantly improved stylisation quality than batch normalization. One notable progress is CycleGAN [9], which extended style transfer to unpaired image domains, allowing the adaptation of artistic styles between diverse datasets, for example in area of medical imaging. Although in recent years, diffusion model [19], [20] has shown powerful ability in image generation. It can hardly replace CycleGAN currently because of its specially ability to trained on unpaired dataset. This is also why we use CycleGAN instead of popular diffusion model as translator framework.

III. Method

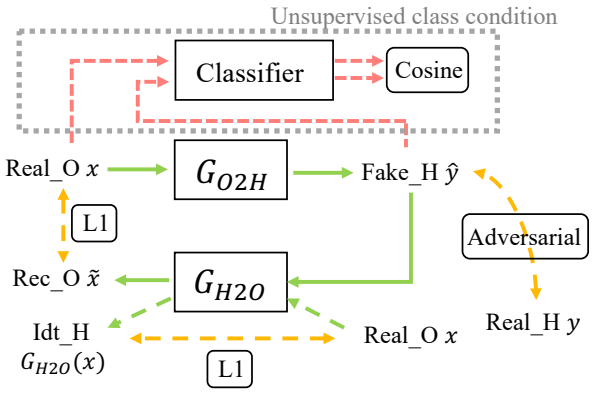
A. Overview of CycleGAN framework

Existing image translation methods often depend on paired datasets, where corresponding images in two domains share the same content. However, in our case, the Heidelberg and Optovue datasets contain retinas from different patients, making paired data unavailable. To overcome this limitation, we employ the CycleGAN framework [9] which addresses the absence of paired data by introducing a cycle-consistency constraint. For instance, in the Green cycle of Fig. 1, a real Optovue sample x undergoes translation to the Heidelberg domain using the generator $G_{O2H}(\cdot)$, resulting in $\hat{y}$. Subsequently, $\hat{y}$ is translated back to the Optovue domain through another generator $G_{H2O}(\cdot)$, producing the recovered Optovue image $\tilde{x}$. By imposing an L1 loss between x and $\tilde{x}$, we ensures faithful reconstruction, preserving both content and style in Optovue domain. Adversarial loss between y and $\hat{y}$ enforces the transferred image share same style as samples in Heidelberg domain. An analogous procedure is employed for the blue cycle, and both cycles share the generators $G_{O2H}(\cdot)$ and $G_{H2O}(\cdot)$. Thus in this manner, a robust and accurate mapping can be learned even without paired training samples.

Additionally, following the recommendation in [9], we incorporate an identity loss, which as depicted in Fig. 2, measures the L1 difference between x and $G_{H2O}(x)$, along



(a) Blue cycle with supervised class condition



(b) Green cycle with unsupervised class condition

Fig. 2. The disassembly diagram of Fig. 1 showing the blue (a) and green (b) cycles, respectively, with the proposed supervised (a) and unsupervised (b) class constraints.

with the L1 difference between y and $G_{O2H}(y)$. It plays a pivotal role in preserving channel relationships, which is critical in our en-face OCTA data. In summary, the vanilla training loss of CycleGAN without class constraint is:

$$\begin{aligned} l_{cyclegan}(x,y) = & \, l_{cyc}(x,\tilde{x}) + l_{cyc}(y,\tilde{y}) \\ & + \alpha(l_{GAN}(x,\hat{x}) + l_{GAN}(y,\hat{y})) \\ & + \beta(l_{idt}(x, G_{H2O}(x)) + l_{idt}(y, G_{O2H}(y))) \end{aligned} \tag{1}$$

where α, β are loss weights.

B. Explicit class condition

Diverging from most existing style transfer works, our primary objective is facilitating the classifier training with transferred images, rather than emphasizing realistic visual quality. Prior research [21] has established that, in image restoration area, optimal classification performance is not achievable when images undergo restoration with distortions or perceptual losses alone. We extend this principle to style transferring. Therefore in this section, we investigate explicit constraint to elevate the machine recognition quality of transferred images. Leveraging a pretrained classifier, we propose implementing this constraint through either a fully-supervised approach, as demonstrated in Fig 2 (a), or an unsupervised approach, as illustrated in Fig 2 (b), details of which will be presented in the following paragraphs.

1) Supervised class constraint: In the ideal scenario where class annotations are available, a straightforward approach to incorporate class constraints involves applying a pretrained classifier $cls(\cdot)$ to the transferred image and enforcing it to remain within the same class. For example, illustrated in Fig 2 (a), we feed the fake Optovue image $\hat{x}$ into the classifier and then compute the classification loss with respect to the label of real Heidelberg image y. We use cosine loss rather than cross entropy, as it has demonstrated effective with limited data [22]. It is worth noticing that the classifier is pretrained and frozen during CycleGAN training process. Thus, to minimize the additional classification loss, the generator $G_{H2O}(\cdot)$ must learn to preserve category information while performing style transfer. Combined with the vanilla loss, the final CycleGAN training loss with supervised class constraints is:

$$\begin{aligned} l_{total} = & \, l_{cyclegan}(x,y) \\ & + \gamma(l_{cos}(label(x), cls(\hat{y})) + l_{cos}(label(y), cls(\hat{x}))) \end{aligned} \tag{2}$$

where γ is loss weight, $cls(\cdot)$ is a pretrained classifier.

2) Unsupervised class constraint: Alternatively, this class constraint can be implemented in an unsupervised manner when labels are unavailable in certain cases. As illustrated in Fig 2 (b), we apply the same classifier to the real Optovue image x to generate a pseudo label for the classification loss. This pseudo label, akin to the soft labels utilized in teacher-student distillation learning approaches [23], encapsulates essential category information. Moreover, this approach allows for computing classification loss at the feature level, wherein the classifier is replaced with a feature extractor. In the scenario presented in Fig 2 (b), the classifier extracts the image representations for both real Optovue image x and the transferred fake Heidelberg image $\hat{y}$. Despite differences in style, these images share identical semantic content. Consequently, we utilize a cosine distance as loss function to make their image representatives as close as possible. When integrated with the vanilla loss, the final CycleGAN training loss incorporating unsupervised class constraints is:

$$\begin{aligned} l_{total} = & \, l_{cyclegan}(x,y) \\ & + \gamma(l_{cos}(cls(x), cls(\hat{y})) + l_{cos}(cls(y), cls(\hat{x}))) \end{aligned} \tag{3}$$

where γ is loss weight, $cls(\cdot)$ is a pretrained feature extractor.

C. Classifier training with CycleGAN

In this section, we outline the process of training a classifier using the augmented training set generated by CycleGAN. Leveraging a well-trained CycleGAN translator, trained with the proposed class condition, we acquire the capability to transfer Heidelberg samples into Optovue style and vice versa. This augmentation introduces samples with previously unseen content in each domain, illustrated by 'Fake_O' and 'Fake_H' in Fig. 3. Despite sharing the same style as real samples, these synthetic data exhibit entirely different vascular patterns, enriching the diversity of the training dataset.

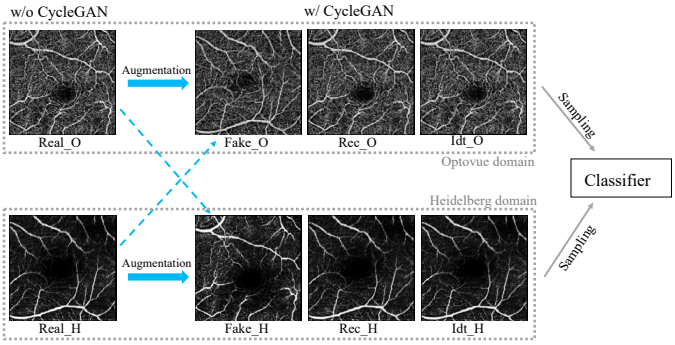


Fig. 3. Illustration of the CycleGAN-based data augmentation for DNN classifier training. Synthetic samples are generated via CycleGAN as shown in Fig. 2. The classifier can be trained with real or synthetic data from each domain, employing predetermined sampling strategies.

Furthermore, the CycleGAN translator allows for intra-domain sample augmentation by utilizing the outputs of cycle consistency and identity, exemplified by 'Rec_O' and 'Idt_O' in the Optovue domain. While in an ideal scenario 'Real_O', 'Rec_O', and 'Idt_O' would be identical (if the loss $l_{cyclegan}$ reaches zero, which is practically unattainable), they remain slightly different but adhere to the distribution of the Optovue domain. Analogous to conventional data augmentation techniques like random distortion, these variations hold the potential to enhance classifier training, particularly when combined with our proposed class constraint. Utilizing different predetermined sampling strategy, the classifier can be trained with real or synthetic data from the Heidelberg domain, the Optovue domain, or a combination of both.

IV. Experiments

A. Implement details

1) Dataset: In addition to the Heidelberg OCTA dataset utilized in [8], we collected an extra Optovue Solix OCTA dataset from Medical University Gdańsk, Poland. This dataset comprises of 580 en-face OCTA projections, sharing the same target categories—active, remission, dry, and normal—as the Heidelberg dataset. Please refer to [8] for clinical definitions of each category. For dataset division, we initially select 15 samples per category for Optovue and 25 samples per category for Heidelberg to form the testing set. Then, for classifier training, we randomly choose a validation set from the remaining samples to conduct a 5-fold validation experiment. In CycleGAN training, since it is an unpaired translator, we use both the training and validation sets. However, no testing samples are involved in the training process for either network.

2) Implementation: We implement all DNNs using the PyTorch platform. For CycleGAN, we compute adversarial loss using LSGAN [24] in a PatchGAN manner, with the loss weights α and β consistent with the original paper [9]. The class constraint weight γ is set to 0.1. The networks are trained using the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. All generators and discriminators undergo 200 epochs with a learning rate of 0.0002, followed by a linear decrease over the next 600 epochs. As for the classifier, we maintain the training settings exactly the same as in [8]. This includes the use of Adam optimizer with 10^{-5} weight decay, an initial learning rate of 10^{-3} decreasing via a cosine scheduler with a minimum value of 10^{-5}. The loss function is the cosine loss [22].

B. Results

Given the limited instances of OCTA based AMD stage graders, especially on the Optovue OCTA dataset, we implement the recently published 3D Heidelberg classifier [8] and apply it to both Heidelberg and Optovue en-face images. We also report the 3D performance on Heidelberg as a reference. Additionally, we train a multimodal AMD grader [25] as well, but restrict it to use the OCTA modality only due to the absence of OCT B-scans in our Optovue dataset.

Based on the different sampling strategies illustrated in Fig. 3, we have various experiment settings to consider:

1) Train the classifier with either 'Real_H' or 'Real_O' exclusively or a mixture of both without CycleGAN augmentation;
2) Train the classifier with a combination of both domains, augmented by CycleGAN, encompassing all eight versions depicted in Fig. 3;
3) Train the classifier within one instrument domain, augmented by CycleGAN, for example using 'Real_O', 'Fake_O', 'Rec_O' and 'Idt_O' for Optovue domain.

By combining all these cases with the type of class condition employed during CycleGAN training, we present the results in Table I and Table II, where the accuracy of each classifier is evaluated on Heidelberg and Optovue test sets, respectively.

Let's firstly examine the classifier without CycleGAN augmentation in Table I. Trained within a single domain, the classifier demonstrates optimal performance on the corresponding test set but significantly underperforms on the other. Simply combining the two domains helps mitigate imbalanced performance but achieves suboptimal outcomes on both test sets. In Table II, introducing CycleGAN augmentation to the two-domain mixture training improves performance on both test sets, particularly with our proposed class constraint. As observed using CycleGAN trained with the supervised class constraint, the classifier achieves optimal performance on both test sets, even slightly surpassing domain-specific training without CycleGAN. When applied to domain-specific training, the CycleGAN translator further enhances accuracy. In Heidelberg domain with CycleGAN augmented data, the performance significantly improves on the Heidelberg test set, approaching the performance with 3D raw data. It also improves performance on the Optovue test set, although not optimally. A similar case can be observed in Optovue domain.

These experiments demonstrate the effectiveness of CycleGAN translator as an advanced data augmentation technique in enhancing classification performance. It is noteworthy that in all experiments involving CycleGAN, the performance improvement is marginal, or even worse, without our proposed class constraint. The results empirically affirm that pixel-wise loss is insufficient, and class-related conditions are essential

TABLE I
CLASSIFICATION ACCURACY WITHOUT CYCLEGAN AUGMENTATION. REAL_{H,O} MEANS A MIXTURE OF REAL_H AND REAL_O. REAL_H 3D, WHICH USES 3D DATA AS INPUT, IS PROVIDED FOR REFERENCE.

Backbone	Training sample	Target domain	Test Acc (Heid/Opto)
CustomCNN [25]	Real_O	Opto only	0.31 / 0.45
	Real_H	Heid only	0.56 / 0.3
EfficientNet [8]	Real_O	Opto only	0.33 / **0.7333**
	Real_H	Heid only	**0.72** / 0.5167
	Real_{H,O}	Heid+Opto	0.67 / 0.7
	Real_H 3D	Heid only	*0.8* / -

TABLE II
CLASSIFICATION ACCURACY WITH CYCLEGAN AUGMENTATION. PLEASE REFER TO FIG. 3 FOR EXAMPLE OF DIFFERENT TRAINING SAMPLE TYPE.

Training sample	Target domain	Class condition	Test Acc (Heid/Opto)
Real_O, Rec_O Fake_O, Idt_O	Opto only	No	0.51 / 0.7167
	Opto only	Supervised	0.55 / **0.8**
	Opto only	Unspv	0.64 / 0.7833
Real_H, Rec_H Fake_H, Idt_H	Heid only	No	0.73 / 0.65
	Heid only	Supervised	**0.78** / 0.6
	Heid only	Unspv	0.77 / 0.6833
Real_{H,O}, Rec_{H,O} Fake_{H,O}, Idt_{H,O}	Heid+Opto	No	0.68 / 0.7333
	Heid+Opto	Supervised	**0.76** / **0.7667**
	Heid+Opto	Unspv	0.74 / 0.75

when employing image processing techniques, such as style transferring, to enhance classification performance. Regarding the two types of proposed class constraints, the supervised one consistently outperforms its unsupervised counterpart, potentially due to the additional label information provided.

V. CONCLUSION

In this study, we employed an unpaired style transfer algorithm, CycleGAN, for cross-instrument data augmentation to enhance the accuracy of AMD stage grading using en-face OCTA images exclusively. The proposed approach not only enhances diversity within each domain but also generates unseen data based on samples from the other domain. Our experiments prove the CycleGAN translator is particularly beneficial if trained with the proposed class-related constraints. We hope this study draws more attention to the significance of semantic-related loss design in image processing algorithms aimed at improving classification as downstream task.

REFERENCES

[1] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[2] R. R. Bourne, J. B. Jonas, S. R. Flaxman, J. Keeffe, J. Leasher, K. Naidoo, M. B. Parodi, K. Pesudovs, H. Price, R. A. White *et al.*, "Prevalence and causes of vision loss in high-income countries and in eastern and central europe: 1990–2010," *British Journal of Ophthalmology*, vol. 98, no. 5, pp. 629–638, 2014.

[3] R. F. Spaide, J. G. Fujimoto, N. K. Waheed, S. R. Sadda, and G. Staurenghi, "Optical coherence tomography angiography," *Progress in retinal and eye research*, vol. 64, pp. 1–55, 2018.

[4] A. Heinke, W. R. Freeman, D.-U. G. Bartsch, L. Cheng, C. M. B. Galang, A. Warter, F. Kalaw, H. Zhang, T. Q. Nguyen, and C. An, "Quantitative evaluation of morphological changes in anti-vegf treated choroidal neovascularization due to age related macular degeneration using optical coherence tomography angiography." *Investigative Ophthalmology & Visual Science*, vol. 63, no. 7, pp. 1348–F0182, 2022.

[5] M. Cavichini, K. C. Dans, M. Jhingan, M. J. Amador-Patarroyo, S. Borooah, D.-U. Bartsch, E. Nudleman, and W. R. Freeman, "Evaluation of the clinical utility of optical coherence tomography angiography in age-related macular degeneration," *British Journal of Ophthalmology*, vol. 105, no. 7, pp. 983–988, 2021.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *ICCV*, 2015, pp. 1026–1034.

[7] A. Heinke, H. Zhang, D. Deussen, C. M. B. Galang, A. Warter, F. G. P. Kalaw, D.-U. G. Bartsch, L. Cheng, C. An, T. Nguyen *et al.*, "Artificial intelligence for OCTA-based disease activity prediction in age-related macular degeneration." *RETINA*, pp. 10–1097, 2022.

[8] H. Zhang, A. Heinke, C. M. B. Galang, D. N. Deussen, B. Wen, D.-U. G. Bartsch, W. R. Freeman, T. Q. Nguyen, and C. An, "Robust amd stage grading with exclusively octa modality leveraging 3D volume," in *ICCV Workshop*, 2023, pp. 2411–2420.

[9] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232.

[10] Y. Guo, A. Camino, J. Wang, D. Huang, T. S. Hwang, and Y. Jia, "MEDnet, a neural network for automated detection of avascular area in OCT angiography," *Biomedical optics express*, vol. 9, no. 11, pp. 5147–5158, 2018.

[11] M. Li, Y. Chen, Z. Ji, K. Xie, S. Yuan, Q. Chen, and S. Li, "Image projection network: 3D to 2D image segmentation in OCTA images," *Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3343–3354, 2020.

[12] M. Alam, D. Le, T. Son, J. I. Lim, and X. Yao, "AV-Net: deep learning for fully automated artery-vein classification in optical coherence tomography angiography," *Biomedical optics express*, vol. 11, no. 9, pp. 5249–5257, 2020.

[13] D. Le, M. Alam, C. K. Yao, J. I. Lim, Y.-T. Hsieh, R. V. Chan, D. Toslak, and X. Yao, "Transfer learning for automated octa detection of diabetic retinopathy," *Translational Vision Science & Technology*, vol. 9, no. 2, pp. 35–35, 2020.

[14] L. Lin, Z. Wang, J. Wu, Y. Huang, J. Lyu, P. Cheng, J. Wu, and X. Tang, "BSDA-net: A boundary shape and distance aware joint learning framework for segmenting and classifying OCTA images," in *MICCAI*, 2021, pp. 65–75.

[15] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," *transactions on visualization and computer graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.

[16] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *CVPR*, 2016, pp. 2414–2423.

[17] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016, pp. 694–711.

[18] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *CVPR*, 2017, pp. 6924–6932.

[19] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *NeurIPS*, vol. 33, pp. 6840–6851, 2020.

[20] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, vol. 1, no. 2, p. 3, 2022.

[21] D. Liu, H. Zhang, and Z. Xiong, "On the classification-distortion-perception tradeoff," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, pp. 1–10, 2019.

[22] B. Barz and J. Denzler, "Deep learning on small datasets without pre-training using cosine loss," in *WACV*, 2020, pp. 1371–1380.

[23] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[24] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *ICCV*, 2017, pp. 2794–2802.

[25] K. A. Thakoor, J. Yao, D. Bordbar, O. Moussa, W. Lin, P. Sajda, and R. W. Chen, "A multimodal deep learning system to distinguish late stages of AMD and to compare expert vs. AI ocular biomarkers," *Scientific reports*, vol. 12, no. 1, pp. 1–11, 2022.