JDHBI - Journal of Digital Health and Biomedical Informatics, 2026, 1:1
doi: 10.5753/jdhbi.2026.8362
This work is licensed under a Creative Commons Attribution 4.0 International License.

Research Paper

Retinal Vessel Segmentation in OCTA Images with U-Net: A Comparative Analysis of Training Strategies

Henrique Christoph Bohn [Universidade do Vale do Rio dos Sinos | *hbohn@edu.unisinos.br*]
Cristiano André da Costa [Universidade do Vale do Rio dos Sinos | *cac@unisinos.br*]
Fabio Lavinsky [Wills Eye Hospital, Sidney Kimmel Medical College, GEPO Lavinsky Eye Institute | *ylavinsky@hotmail.com*]
Ismael Garcia Santos [Universidade do Vale do Rio dos Sinos | *ismael@edu.unisinos.br*]
Gregory Stein [Universidade do Vale do Rio dos Sinos | *SteinGregory@edu.unisinos.br*]
Felipe André Zeiser [Universidade do Vale do Rio dos Sinos | *felipezeiser@unisinos.br*]

Graduate Program in Applied Computing (PPGCA) — Universidade do Vale do Rio dos Sinos (UNISINOS) — Av. Unisinos, 950 - Cristo Rei, São Leopoldo, RS, Brazil — ZIP Code 91330-002

Abstract. Optical coherence tomography angiography (OCTA) is a non-invasive, depth-resolved imaging modality that enables visualization of retinal vascular slabs, including superficial and deeper plexuses. Accurate vessel segmentation is essential for deriving biomarkers such as vessel density and foveal avascular zone area, which support the assessment of diabetic retinopathy and age-related macular degeneration. This study investigates whether task-specific data augmentation or ImageNet-based transfer learning provides greater benefit for large-vessel segmentation under limited labeled data. We evaluated four U-Net-based configurations on the OCTA_6mm subset of OCTA-500 using a 70/15/15 subject-level split: a U-Net without pretraining or augmentation, the same model with augmentation, and two variants with ImageNet-pretrained ResNet34 and EfficientNet-B0 encoders. The augmented U-Net achieved Dice = 0.887 and IoU = 0.799, within the range of previously reported OCTA-500 baselines from 2D projection maps and close to the reported IPN volumetric baseline (DSC = 0.882). The pretrained encoders did not outperform the U-Net without pretraining. To isolate the effect of augmentation from concurrent training-recipe changes, we conducted a controlled multi-seed ablation varying only the augmentation policy. Augmentation yielded a modest but consistent Dice gain of $+0.007$ over a matched no-augmentation baseline, driven almost entirely by geometric transformations; intensity and elastic components added no measurable benefit. These findings suggest that, at this data scale, task-preserving geometric invariances are more effective than natural-image pretraining for 2D large-vessel OCTA segmentation.

Keywords: Optical Coherence Tomography Angiography; Retinal Vessel Segmentation; U-Net; Data Augmentation; Transfer Learning

Received: 29 May 2026 • **Accepted:** 07 Jul 2026 • **Published:** 13 Nov 2026

1 Introduction

Optical coherence tomography angiography (OCTA) is a non-invasive imaging modality that captures the retinal microvasculature through motion contrast. Static tissue yields a stable signal across repeated scans, whereas moving red blood cells induce temporal variations that reveal blood flow without exogenous contrast agents Kashani *et al*. (2017); Gao *et al*. (2016). The resulting 3D angiographic volume can be projected into 2D en face maps of distinct vascular plexuses, from which biomarkers such as vessel density, retinal layer measurements, and foveal avascular zone (FAZ) area are derived. These measurements are clinically relevant in diabetic retinopathy (DR), age-related macular degeneration (AMD), retinal vein occlusion (RVO), and choroidal neovascularization. Accurate vessel segmentation is therefore a prerequisite for reliable OCTA quantification, but it remains challenging. Motion artifacts can interrupt or duplicate vessels, while projection artifacts may transfer superficial flow into deeper layers, reducing anatomical specificity and creating false-positive patterns Gao *et al*. (2016); Kashani *et al*. (2017). Disease-related morphology further increases inter-subject variability, and, compounding these difficulties, annotated OCTA data remain scarce. When labeled data are limited, two practical strategies are commonly adopted to compensate: increasing the effective diversity of the training set through data augmentation, or importing learned representations through transfer learning Tajbakhsh *et al*. (2020).

We therefore compare these two strategies directly for large-vessel segmentation on the OCTA_6mm subset of the OCTA-500 dataset. OCTA-500 is a public retinal OCTA dataset comprising OCT/OCTA volumes, projection maps, clinical labels, and pixel-level vascular annotations. We evaluate a U-Net trained without pretraining, the same model with augmentation, and two ImageNet-pretrained encoder variants, ResNet34 and EfficientNet-B0. Because a single-run comparison can confound the effect of augmentation with incidental variation in the training process (e.g., weight initialization and stochastic optimization dynamics), we further isolate its contribution through a controlled multi-seed ablation. The goal is to determine which training strategy is most useful for 2D large-vessel segmentation in OCTA when labels are limited and the target modality differs substantially from natural images. Specifically, we ask whether task-preserving data augmentation or ImageNet-pretrained encoders provide the greater practical benefit under a fixed U-Net-based segmentation setting.

2 Background and Related Works

OCTA is a functional extension of structural OCT, a depth-resolved imaging technique introduced in 1991 that enabled micrometer-scale visualization of retinal anatomy Kashani *et al*. (2017); Gao *et al*. (2016). Unlike structural OCT, which relies on back-reflected light, OCTA exploits temporal signal variation from moving blood cells, using Doppler, speckle-variance, or decorrelation contrast to generate dye-free angiograms Gao *et al*. (2016). It preserves depth information through volumetric acquisition and slab-wise en face visualization, but remains sensitive to eye motion and shadowgraphic projection artifacts, which can make superficial vessels appear in deeper slabs Gao *et al*. (2016); Kashani *et al*. (2017). These physical constraints directly affect segmentation and quantification.

Vessel segmentation in OCTA has been addressed with thresholding, filtering, active contours, and deep learning. Classical pipelines are simple but brittle under noise, low contrast, artifacts, and layer-segmentation errors Kashani *et al*. (2017); Gao *et al*. (2016). U-Net Ronneberger *et al*. (2015) is a strong learning-based baseline because its encoder–decoder structure with skip connections preserves context and fine vessel detail, reporting DSC = 89.1% on OCTA_6mm large-vessel segmentation Li *et al*. (2023a). In contrast, Image Projection Network (IPN) segments directly from the OCTA volume using a 3D-to-2D architecture, reaching DSC = 0.882 by exploiting volumetric cues unavailable from a single projection map; IPN-V2 later extended this direction in OCTA-500 Li *et al*. (2020, 2023b). This contrast between efficient 2D projections and richer but heavier volumetric models motivates testing the limits of a carefully trained 2D U-Net.

Transfer learning is another common response to scarce medical annotations, but its effectiveness depends on the source–target domain relationship. Reviewing 121 medical image classification studies from 425 screened papers, Kim et al. Kim *et al*. (2022) found that transfer learning is widely used to reduce data requirements and computational cost. However, choices regarding the backbone and fine-tuning remain largely empirical. Raghu et al. Raghu *et al*. (2019) showed that on retinal and chest X-ray tasks, ImageNet pretraining yielded only small gains, whereas smaller CNNs achieved comparable medical performance despite lower ImageNet accuracy. Their analysis suggests that useful feature reuse is concentrated mainly in lower layers, with part of the benefit attributable to optimization rather than to high-level semantic transfer.

This caveat is central for OCTA, whose signal arises from interferometric motion contrast and vascular decorrelation rather than natural-image color, texture, or object semantics Kashani *et al*. (2017); Gao *et al*. (2016). The same modality gap appears in OCTA-specific multimodal work, where OCT–OCTA fusion yields only marginal gains over OCTA alone Abtahi *et al*. (2022). We therefore compare two ImageNet-pretrained encoders with different capacities: ResNet34 (≈21M parameters) and EfficientNet-B0 (≈4M parameters). Capacity matters because the target training set is small; a larger encoder may overfit during fine-tuning, whereas a lighter one may retain useful low-level filters with fewer trainable parameters.

3 Materials and Methods

This retrospective experimental study compared deep learning training strategies for retinal vessel segmentation in OCTA images. Experiments used the publicly available OCTA-500 dataset Li *et al*. (2023b), which contains de-identified OCT/OCTA data and corresponding annotations. No new patient data were collected, and no intervention or participant contact was performed. Thus, this work represents a secondary analysis of previously published, anonymized data and did not require additional institutional ethics approval.

We frame the study as a controlled comparison rather than an architecture search. A common preprocessing stage normalizes the B5 maps and binarizes the ground-truth masks. Augmentation then optionally expands training-set diversity. The same U-Net backbone is used for four configurations: model without pretraining or augmentation, model with augmentation, and two ImageNet-pretrained encoders (ResNet34, EfficientNet-B0) under two-phase fine-tuning. Therefore, each comparison isolates a single factor: data diversity versus a representational head start under limited labels.

The pipeline was implemented in Python 3.10 with PyTorch 2.7.1 on an NVIDIA GeForce RTX 4050 GPU (6.4 GB VRAM, 32 GB RAM).

3.1 Data and Preprocessing

All experiments use OCTA-500 Li *et al*. (2023b), which comprises 500 subjects imaged on a single 70 kHz SD-OCT system (RTVue-XR, Optovue; 840 nm) at Jiangsu Province Hospital between 2018 and 2020, split by field of view into OCTA_6mm (300 subjects, 6×6 mm), and OCTA_3mm (200 subjects, 3×3 mm). The 3 mm field resolves the fovea in finer detail, whereas the 6 mm field trades resolution for coverage and, importantly for this study, for diagnostic diversity: it includes AMD, DR, CNV, CSC, RVO, and other conditions, with only ≈30% normal eyes (Figure 1). Each subject provides OCT and OCTA volumes, six pre-computed projection maps, text labels, and pixel-level annotations for large vessels, capillaries, arteries, veins, 2D/3D FAZ, and retinal layers. We work with OCTA_6mm because it provides broader visualization of the larger retinal vasculature and greater clinical heterogeneity, thereby stressing segmentation models under realistic variability.

For large vessels, the relevant input is the B5 map (OCTA_ILM_OPL), the maximum en face OCTA projection between the internal limiting membrane (ILM) and the outer plexiform layer (OPL), which best captures the inner retinal vasculature Li *et al*. (2023b). This choice carries an important dependency: B5 is generated from retinal layer segmentation, which can fail under disease Zhang *et al*. (2020). Therefore, any 2D pipeline inherits this upstream risk, a limitation revisited in the Discussion. We use the 300 B5 maps (400×400 px) with their *GT_LargeVessel* masks. Subjects were divided into training, validation, and test sets using a 70/15/15 split, corresponding to 210, 45, and 45 subjects, respectively, with a fixed seed for reproducibility. Images were normalized to $[0, 1]$. Mean vessel density is ≈9.5% of pixels, a moderate class imbalance that motivates the loss design in Section 3.3.

The split was random and not stratified by image quality,

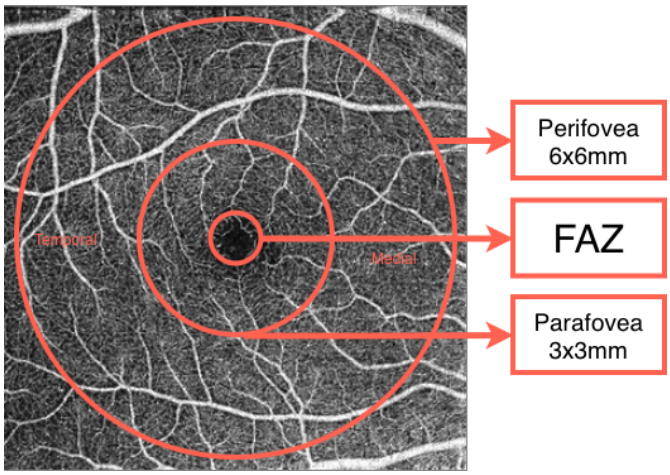


Figure 1. The two fields of view in OCTA-500. The $3 \times 3\,\mathrm{mm}^2$ field concentrates sampling near the fovea, resolving the FAZ in finer detail; the $6 \times 6\,\mathrm{mm}^2$ field used here trades resolution for broader vascular coverage and greater diagnostic diversity.

vessel density, pathology, or artifact level. Although images with severe artifacts should ideally be excluded or analyzed separately through quality control, this choice preserves a simple and reproducible protocol while allowing possible distributional differences between partitions.

3.2 U-Net Architecture

The U-Net trained without pretraining follows the standard encoder–decoder formulation with skip connections and serves as a non-pretrained baseline, motivated by evidence that ImageNet transfer is not universally beneficial in medical imaging and that simpler convolutional models can remain competitive under domain shift Raghu *et al*. (2019); Kim *et al*. (2022). The encoder has four blocks, each with two 3×3 convolutions, Batch Normalization, ReLU activation, and 2×2 max-pooling. Channels double across levels from 64 to 512, followed by a 1024-channel bottleneck. The symmetric decoder uses transposed convolutions for upsampling and concatenates the corresponding skip connections before each convolutional pair. A final 1×1 convolution with sigmoid activation produces the segmentation map. The network contains 31,036,481 trainable parameters. Figure 2 summarizes the architecture used in this study.

3.3 Loss Function

The training objective combines Binary Cross-Entropy (BCE) and Dice Loss. BCE penalizes pixel-wise classification errors, while Dice Loss optimizes the overlap between the predicted vessel map and the ground-truth mask. This is suitable for OCTA vessel segmentation because vessels occupy a small portion of the image: Dice Loss mitigates class imbalance, and BCE preserves stable local supervision. Let p denote the predicted probability map and y the ground-truth mask. The final loss is:

$$\mathcal{L} = \mathcal{L}_{\mathrm{BCE}}(p,\, y) + \big(1 - \mathrm{Dice}(p,\, y)\big) \tag{1}$$

3.4 Data Augmentation

For Rounds 2, 3, and 4, data augmentation was applied to the training set using *albumentations*, an open-source Python library for applying image transformations in machine learning pipelines. The transformations included rotations of up to $\pm 30°$ ($p = 0.5$), horizontal flips ($p = 0.5$), vertical flips ($p = 0.3$), brightness and contrast variations ($p = 0.4$), Gaussian noise ($p = 0.2$), and elastic transformations ($p = 0.2$). Each transformation was applied simultaneously to the image and its corresponding mask. These choices were intended to reflect plausible sources of variation in OCTA imaging: differences in acquisition orientation motivate rotations, anatomical asymmetry between left and right eyes supports horizontal flips, and variability in device settings and examination conditions motivates intensity perturbations and noise-based augmentation Kashani *et al*. (2017). No augmentation was applied to the validation or test sets.

3.5 Augmentation Ablation Protocol

The single-run comparison between Round 1 and Round 2 conflates the effect of data augmentation with concurrent changes in the training recipe (L2 regularization, gradient clipping, longer training, and a more patient early-stopping criterion adopted only in Round 2). To isolate the contribution of augmentation and its individual components, we conducted a controlled, multi-seed ablation study. The Round 2 training recipe was held fixed (identical optimizer, weight decay, gradient clipping, learning-rate scheduler, and early-stopping configuration), and only the augmentation policy was varied across five settings: *none* (no augmentation); *geometric* (rotation and horizontal/vertical flips); *intensity* (brightness/contrast and Gaussian noise); *geometric+intensity*; and *full* (geometric and intensity plus elastic deformation), the last being identical to the augmentation used in Round 2 (Figure 3).

Each policy was trained with three random seeds ($\{42, 123, 2024\}$). The seed determines the train/validation/test partition, so for a given seed, all five policies are evaluated on the same test subjects, yielding a paired comparison. We report the mean test Dice and IoU across seeds, along with the inter-seed standard deviation. This inter-seed dispersion quantifies run-to-run reproducibility and is conceptually distinct from the inter-subject variability reported in Table 1: the former measures how much the result changes when the experiment is repeated, the latter how much performance varies across patients within a single run.

3.6 Training Protocol

The Adam optimizer was used in all rounds with an initial learning rate of 10^{-4} and a batch size of 4, limited by the available VRAM. A *ReduceLROnPlateau* scheduler reduced the learning rate by a factor of 0.5 when the validation metric reached a plateau, with a patience of 8 epochs in Round 2 and 5 epochs in Rounds 3 and 4. Early stopping was configured with a patience of 20 epochs in Round 2 and 15 epochs in Rounds 3 and 4, allowing Round 2 to explore the parameter space for longer before training was interrupted. To improve stability during the additional epochs, Round 2 also used L2 regularization (weight_decay $= 10^{-5}$) and gradient clipping with a maximum norm of 1.0 at each update step. These measures helped prevent unstable gradients without noticeably slowing convergence under the combined BCE + Dice loss.

In Rounds 3 and 4, fine-tuning was conducted in two phases to reduce the risk that the randomly initialized decoder would destabilize the representations learned by the

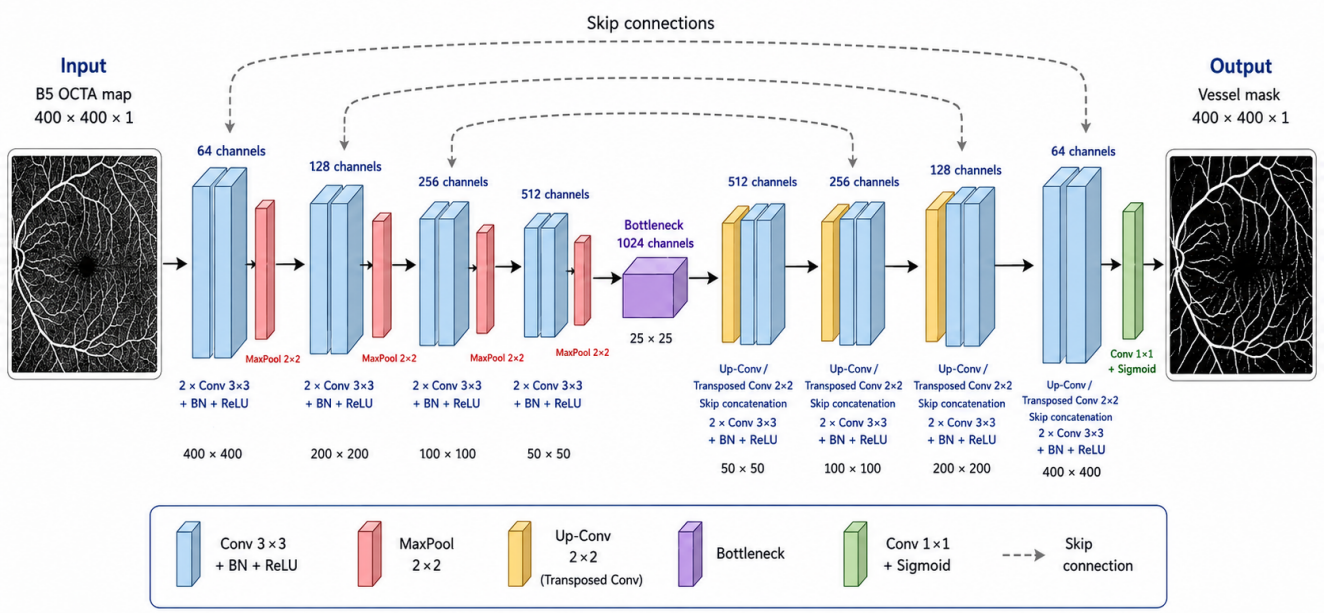


Figure 2. U-Net architecture used for large-vessel segmentation in OCTA. The network receives a B5 OCTA map with size $400 \times 400 \times 1$ and follows a symmetric encoder–decoder design. The encoder progressively reduces spatial resolution while increasing the number of feature channels from 64 to 512, followed by a 1024-channel bottleneck. The decoder restores the original resolution through transposed convolutions and concatenates skip connections from the encoder to recover fine vessel details. A final 1×1 convolution, followed by a sigmoid activation, produces the vessel probability map.

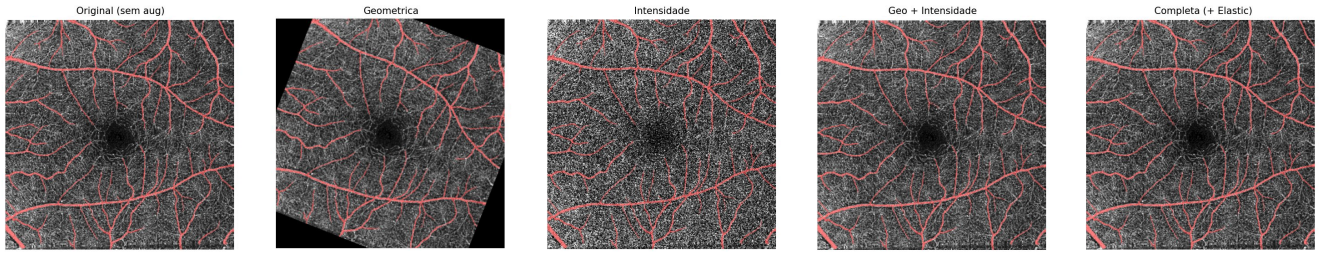


Figure 3. Effect of each augmentation policy on a representative test subject. From left to right: the original image (no augmentation); *geometric* (rotation and flips); *intensity* (brightness/contrast and Gaussian noise); *geometric+intensity*; and *full* (adding elastic deformation). The vessel mask (red overlay) is transformed jointly with the image. The intensity panel illustrates the comparatively strong Gaussian-noise setting used in Round 2, whose contribution is quantified in the ablation (Section 4.2).

pre-trained encoder during the first iterations. In Phase 1, which lasted 15 epochs, the encoder was frozen, and only the decoder was trained, with a learning rate of 3×10^{-4}. In Phase 2, lasting up to 85 epochs, the entire model was unfrozen and trained with a learning rate of 10^{-4} for ResNet34 and 5×10^{-5} for EfficientNet-B0. The lower learning rate for EfficientNet-B0 was adopted to make fine-tuning more conservative, given the limited size of the target dataset and the sensitivity of compact architectures to domain-specific adjustments.

Figure 4 presents an overview of the implemented pipeline, from data input to final evaluation.

3.7 Ground Truth Annotations

The ground-truth masks used in this study correspond to the GT_LargeVessel annotations provided by OCTA-500 for the OCTA_6mm subset. These binary masks delineate large retinal vessels and are paired with the B5 projection map of each subject. In the present pipeline, masks were converted to binary format using a threshold of 127, with vessel pixels assigned to 1 and background pixels assigned to 0. These annotations were used as the reference standard for model training, validation, and test-set evaluation.

4 Evaluation and Results

Performance was quantified using four metrics on the test set: Dice, Intersection over Union (IoU), Precision, and Recall. Dice measures the relative overlap between the predicted mask P and the ground truth mask G:

$$\mathrm{Dice} = \frac{2\,|P \cap G|}{|P| + |G|} \tag{2}$$

IoU is related to Dice by $\mathrm{IoU} = \mathrm{Dice} \,/\, (2 - \mathrm{Dice})$. Precision and Recall characterize the error with respect to false positives and false negatives, respectively. For all test-set evaluations, probability maps were binarized using a fixed threshold of 0.5, and the resulting binary masks were compared with the corresponding ground-truth annotations.

4.1 Test Set Results

Table 1 summarizes the metrics on the 45 held-out test subjects, together with values reported in the literature for the same OCTA_6mm subset. Figure 5 presents the results visually.

It is worth noting that the comparison with results from the literature is indirect, as differences in experimental protocols, including data partitioning, preprocessing, and training configurations, may influence the reported values.

4.2 Augmentation Ablation

Table 2 reports the multi-seed augmentation ablation. Each value is the mean over three seeds, and the $\pm$ term denotes the inter-seed standard deviation. The ΔDice column compares each policy with the *none* baseline, which uses the same training recipe and therefore isolates the effect of augmentation from the recipe changes described in Section 3.6.

Importantly, the none condition in Table 2 is not the same experiment as Round 1 in Table 1. Round 1 used the original no-augmentation recipe, whereas the none condition in the ablation uses the Round 2 recipe without augmentation. Therefore, Table 2 isolates the effect of augmentation under a matched training recipe, while Table 1 reflects the performance of the four main experimental configurations.

Two patterns stand out. First, the full augmentation improves Dice by $+0.0070$ over the matched baseline. A small effect, but a consistent one, since it exceeds the inter-seed standard deviation ($\approx$0.005) and holds across all three seeds. Second, the gain is essentially geometric: rotations and flips alone recover $+0.0068$, while adding intensity perturbations ($+0.0016$, below the inter-seed dispersion and thus not demonstrable) or elastic deformation contributes nothing measurable beyond geometry. The strong Gaussian-noise setting visible in Figure 3 is, in this light, inert at best. Figure 6 summarizes the per-policy means.

4.3 Training Curves

Figure 7 shows the loss and validation curves for the representative augmented run (the *full* policy at the seed whose test Dice, 0.893, matches the multi-seed mean). The training loss sits above the validation loss throughout. This behavior is expected when augmentation is applied only to the training set, which makes those samples harder, and therefore not a sign of overfitting. Validation Dice climbs steeply, surpasses the IPN reference near epoch 20, and settles around 0.89, with early stopping at epoch 103. Precision and recall, erratic for the first $\sim$15 epochs, converge to a balanced 0.89–0.90, indicating no systematic bias toward false positives or negatives.

For context, the pretrained encoders behaved as their capacities suggest. Two-phase fine-tuning was essential: with the decoder frozen first, ResNet34 rose from Dice 0.69 to 0.87 within ten epochs, and EfficientNet-B0 from 0.35 to 0.86 over fifteen epochs. Unfreezing added little. ResNet34 plateaued at $\approx$0.884 and EfficientNet-B0 at $\approx$0.891, leaving both below the augmented without pretraining model.

4.4 Qualitative Analysis

Figure 8 illustrates the three performance regimes with representative test subjects. In the strong case (10071, Dice 0.930), the main arcades are recovered confidently, and errors are confined to the tips of thin branches. In the median case (10146, Dice 0.891), misses appear in low-contrast peripheral regions. In the difficult case (10169, Dice 0.843), degradation tracks reduced image quality and motion artifacts, intrinsic to OCTA acquisition Kashani *et al*. (2017), and were compounded by atypical morphology common in advanced DR or AMD Li *et al*. (2023b). The error maps make the failure mode legible: the residual is dominated by false negatives (blue) rather than false positives (red).

That asymmetry is informative. Across the harder cases, the drop is driven more by recall than by precision: the model omits faint peripheral vessels more readily than it hallucinates spurious ones. This is consistent with the limits of image-level supervision. The symmetric BCE + Dice objective cannot restore vascular evidence that motion artifacts or weak local contrast have already erased from the input Kashani *et al*. (2017). Figure 9 shows the converse, high-quality regime, where contours align tightly with the major vasculature and preserve the FAZ, with deviations limited to small-caliber peripheral branches.

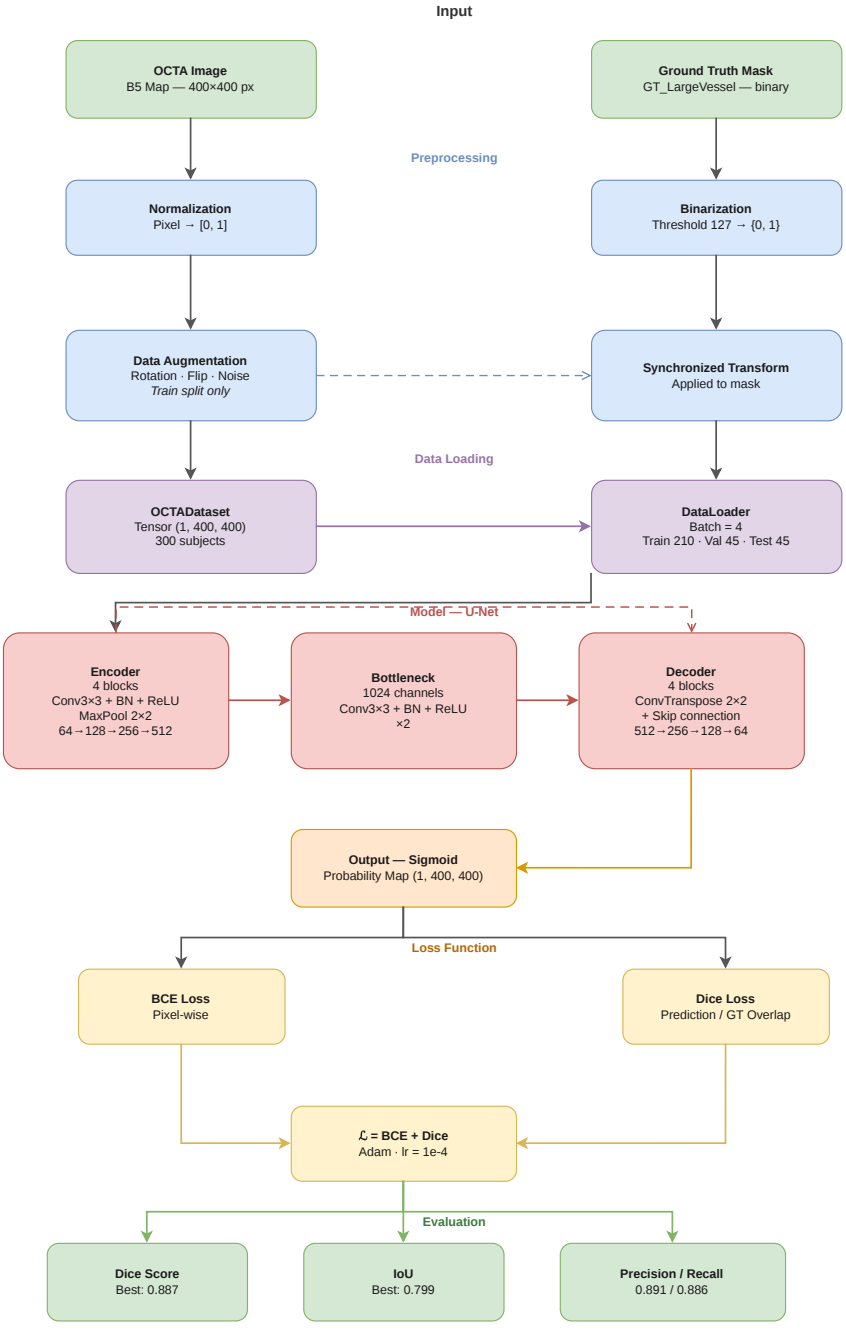


Figure 4. Pipeline of the proposed solution. The workflow comprises six stages: data input (OCTA image and ground-truth mask), preprocessing with synchronized augmentation, loading via *OCTADataset* and *DataLoader*, segmentation by the U-Net with encoder, bottleneck, and decoder connected by skip connections, computation of the combined BCE + Dice loss, and evaluation using Dice, IoU, Precision, and Recall metrics on the test set.

Table 1. Performance on the test set (45 subjects, OCTA_6mm). Standard deviations refer to inter-subject variability. Literature values were extracted from the respective papers; Precision and Recall were not reported in those works for the OCTA_6mm subset.

Model	Dice	IoU	Precision	Recall
R1: U-Net without augmentation	0.869 ± 0.034	0.770 ± 0.049	0.883	0.857
R2: U-Net with augmentation	$\mathbf{0.887 \pm 0.034}$	$\mathbf{0.799 \pm 0.050}$	**0.891**	**0.886**
R3: U-Net + ResNet34	0.871 ± 0.035	0.774 ± 0.051	0.867	0.878
R4: U-Net + EfficientNet-B0	0.878 ± 0.036	0.784 ± 0.052	0.869	0.889
U-Net Li *et al.* (2023b)	0.891	0.805	–	–
IPN Li *et al.* (2020)	0.882	0.788	–	–
IPN-V2 Li *et al.* (2023b)	0.892	0.806	–	–

Table 2. Multi-seed augmentation ablation on the OCTA_6mm test set (3 seeds). All policies share the Round 2 training recipe; only the augmentation varies. The $\pm$ term denotes inter-seed standard deviation (run-to-run reproducibility), distinct from the inter-subject variability in Table 1.

Augmentation policy	Dice	IoU	ΔDice vs. none
None	0.885 ± 0.005	0.795 ± 0.008	—
Geometric	0.892 ± 0.005	0.806 ± 0.006	$+0.0068$
Intensity	0.887 ± 0.007	0.798 ± 0.010	$+0.0016$
Geometric + Intensity	0.892 ± 0.005	0.806 ± 0.006	$+0.0067$
Full (+ Elastic)	$\mathbf{0.892 \pm 0.006}$	$\mathbf{0.806 \pm 0.008}$	$+0.0070$

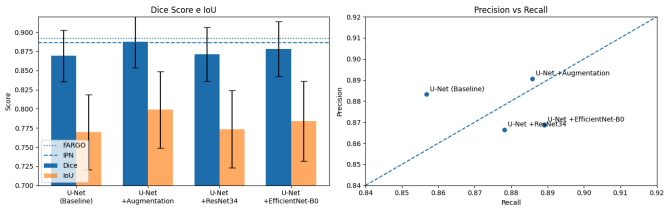


Figure 5. Performance comparison across the four configurations on the test set. Left: Dice and IoU with error bars (inter-subject standard deviation) and reference lines for the IPN Li *et al.* (2020) and IPN-V2 Li *et al.* (2023b) methods. Right: Precision vs. Recall; the dashed diagonal indicates balance between the two metrics.

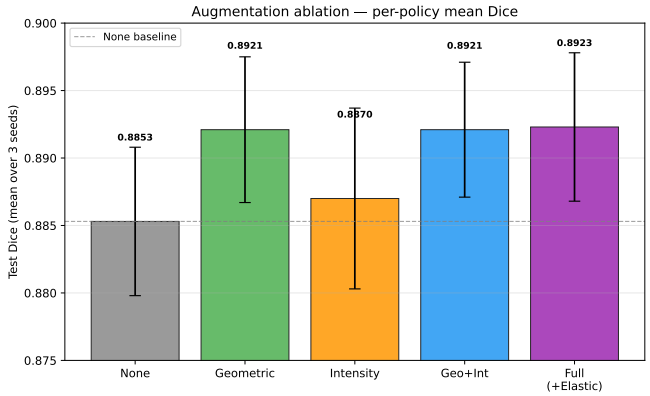


Figure 6. Mean test Dice per augmentation policy (three seeds; error bars are inter-seed standard deviation). The geometric component drives the improvement over the none baseline; intensity and elastic transforms add no measurable benefit.

5 Discussion

The main finding of this study is that for large-vessel segmentation on the OCTA_6 mm subset of OCTA-500, a non-pretrained U-Net trained with data augmentation achieved the best performance among the tested configurations, whereas ImageNet-pretrained encoders did not improve over the model trained without pretraining. However, the controlled multi-seed ablation showed that the isolated effect of augmentation was modest. When the training recipe was kept fixed, augmentation increased Dice by only $+0.007$ over the matched no-augmentation baseline, and this gain was driven almost entirely by geometric transformations. These results suggest that, in this setting, task-preserving geometric invariances are more useful than representations transferred from natural images.

The distinction between the single-run comparison and the controlled ablation is central to interpreting the results. The apparent improvement from Round 1 to Round 2 cannot be attributed solely to augmentation, because Round 2 also introduced changes in the training recipe, including L2 regularization, gradient clipping, longer training, and a more patient early-stopping criterion. By contrast, the ablation fixed these factors and varied only the augmentation policy. Under this controlled setting, rotations and flips accounted for nearly all of the observed improvement, whereas intensity perturbations and elastic deformation added no measurable benefit. This pattern is consistent with the anatomy of retinal vessels. Geometric transformations preserve plausible spatial configurations, while aggressive intensity changes or elastic distortions may move samples away from the distribution of real B5 OCTA projection maps.

The limited benefit of ImageNet pretraining is also consistent with the substantial domain gap between natural images and OCTA. OCTA contrast is derived from motion-related decorrelation and interferometric signal variation, rather than



Figure 7. Training curves for the representative augmented run (early stopping at epoch 103). Left: training and validation loss; the higher training loss reflects augmentation applied only to the training set, not overfitting. Center: validation Dice, which exceeds the IPN reference (0.882) early and approaches the test Dice (0.892). Right: validation precision and recall, converging to 0.89–0.90 with no false-positive or false-negative bias.

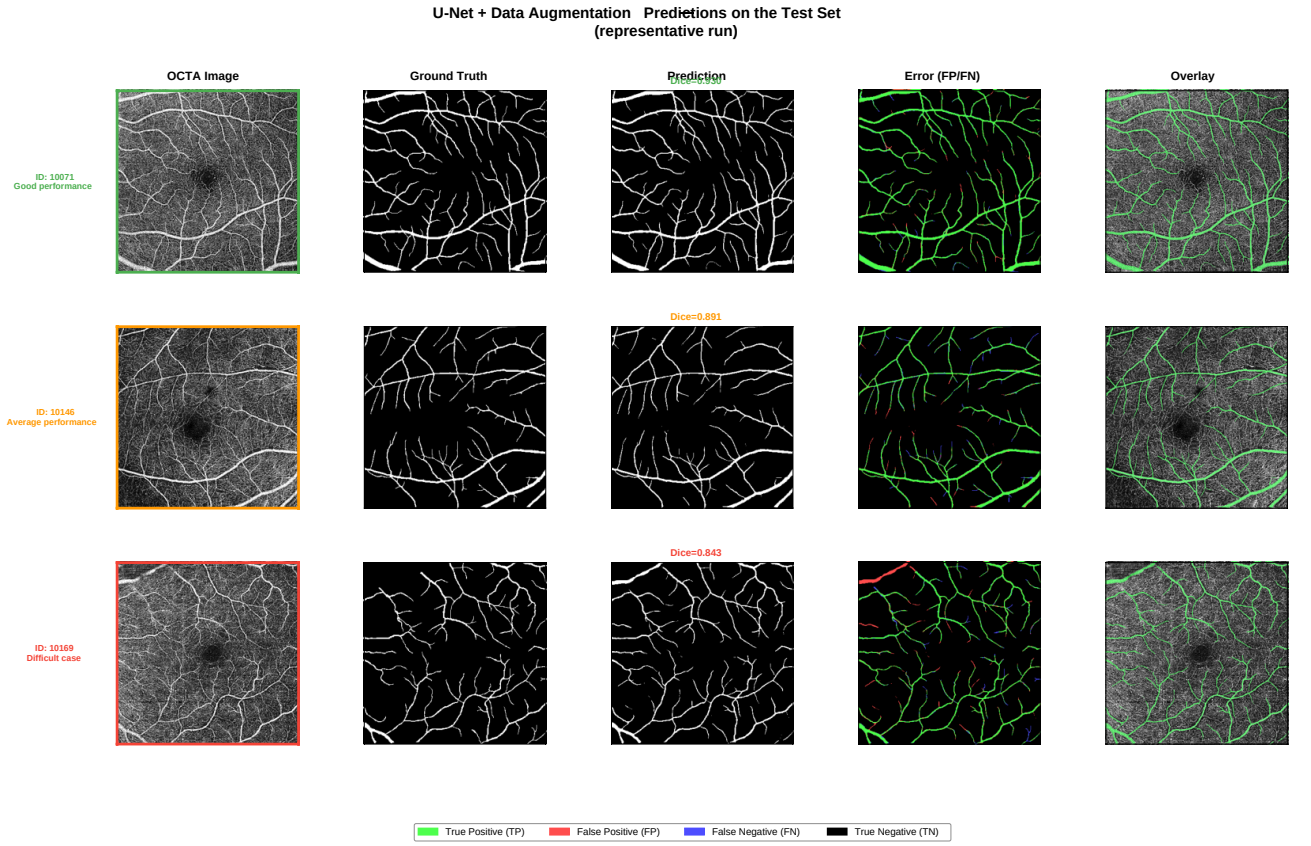


Figure 8. Predictions of the representative augmented model on three test subjects spanning the observed performance range. Columns: OCTA B5 map, ground truth, predicted mask (threshold 0.5), error map (green: true positive; red: false positive; blue: false negative), and green overlay. Residual error is dominated by false negatives in low-contrast regions.

from the color, texture, and object-level semantics present in natural-image datasets Kashani *et al*. (2017); Gao *et al*. (2016). Although low-level filters learned from ImageNet may still be useful, deeper semantic representations are less likely to transfer effectively to this modality. This interpretation agrees with previous observations in medical imaging, where ImageNet pretraining can provide limited gains when the target task and image domain differ substantially from the source data Raghu *et al*. (2019); Kim *et al*. (2022). In our experiments, the larger ResNet34 encoder underperformed the augmented U-Net without pretraining, while the lighter EfficientNet-B0 model performed more closely to it, suggesting that both model capacity and domain mismatch may influence fine-tuning with limited labeled data.

These findings do not imply that transfer learning is generally ineffective for OCTA. Rather, its utility likely depends on the task, data scale, and source-target similarity. For example, in artery-vein classification, Alam *et al*. (2020) reported a clear benefit from ImageNet pretraining, particularly in a more data-limited setting. That task also differs from large-vessel localization, because artery-vein classification requires discriminative vascular features that may be harder to learn from a few labeled examples. In the present study, where the target is binary large-vessel segmentation and 210 subjects are available for training, the U-Net without pretraining appears sufficient to learn the relevant low-level and structural features.

The comparison with volumetric approaches further highlights the practical value of a well-trained 2D model. The augmented 2D U-Net surpasses the IPN volumetric baseline (DSC=0.882) Li *et al*. (2020) using only a projection map, and remains close to the OCTA-500 reference U-Net (DSC=0.891) Li *et al*. (2023b), differing by only 0.004, which is within the inter-subject variability reported in Table 1. Therefore, the result should not be interpreted as definitive evidence of superiority over volumetric methods, but rather as evidence that 2D projection-based segmentation remains competitive for large-vessel localization. This is relevant in practical settings, where 2D models are simpler to train, require less memory, and can be more easily integrated into image-analysis pipelines.

The qualitative results help explain where the model succeeds and where it remains limited. In high-quality images, the augmented U-Net accurately delineated the main vascular arcades and preserved the foveal avascular zone, with errors mostly restricted to small-caliber peripheral branches. In more difficult cases, performance decreased primarily due to false negatives in low-contrast regions or in areas affected by motion artifacts. This failure pattern suggests that the model is less prone to hallucinating vessels than to missing faint vascular structures when the input signal is weak. Such errors are expected in OCTA, where artifacts, projection effects, and reduced local contrast can obscure vascular continuity Kashani *et al*. (2017); Gao *et al*. (2016).

This study has several limitations. First, the analysis used only the B5 projection map, discarding other OCTA projections and the original 3D volume. Consequently, the model inherits the limitations of projection-based analysis, including dependence on retinal layer segmentation, which can be affected by disease-related abnormalities Zhang *et al*. (2020). Second, the train, validation, and test split was random and not stratified by image quality, pathology, vessel density, or artifact level. Although this preserves a simple and reproducible protocol, future studies should include explicit quality control and stratified evaluation, especially to assess the impact of severe artifacts. Third, the transfer-learning comparison does not fully isolate pretraining from the encoder architecture; a stricter design would compare each pretrained encoder with the same architecture, initialized without pretraining, under an identical fine-tuning protocol. Finally, this work focused only on large-vessel segmentation. Capillaries, arteries, veins, and 3D FAZ segmentation pose different challenges and may require volumetric, multimodal, or multi-scale approaches Xu *et al*. (2022).

Overall, the results indicate that, for 2D large-vessel segmentation in OCTA at this data scale, careful training and anatomically plausible geometric augmentation are more beneficial than ImageNet-based encoder pretraining. Future improvements are likely to depend less on stronger augmentation policies or larger pretrained encoders and more on higher-quality annotations, explicit artifact handling, stratified evaluation, and the incorporation of additional projections or volumetric information when the target structure requires it.

6 Conclusion

For 2D large-vessel segmentation on the OCTA_6mm subset of OCTA-500, task-specific data augmentation was more effective than ImageNet-based transfer learning under the tested U-Net-based setting. The augmented U-Net achieved Dice = 0.887 and IoU = 0.799, remaining competitive with previously reported OCTA-500 baselines and close to volumetric approaches despite using only 2D projection maps. However, the controlled multi-seed ablation showed that the isolated effect of augmentation was modest, with a Dice gain of only $+0.007$ over a matched no-augmentation baseline.

The results further indicate that this gain was driven mainly by geometric transformations, whereas intensity perturbations and elastic deformation added no measurable benefit. ImageNet-pretrained encoders did not improve performance, suggesting that natural-image representations are less advantageous than task-preserving geometric invariances for this OCTA segmentation task. Future work should prioritize stratified evaluation by image quality and pathology, explicit artifact analysis, and the use of additional projections or volumetric information for more complex targets such as capillaries, arteries, veins, and FAZ segmentation.

Declarations

Funding

The authors received funding from the Coordination for the Improvement of Higher Education Personnel (CAPES) – Brazil – Finance Code 001.

Authors' Contributions

Henrique Christoph Bohn designed the study, implemented the experiments, analyzed the data, interpreted the results, and wrote the manuscript. Cristiano André da Costa and Felipe André Zeiser supervised the study, contributed to the methodological design, interpreted the results, and revised the manuscript. Fábio Lavinsky contributed

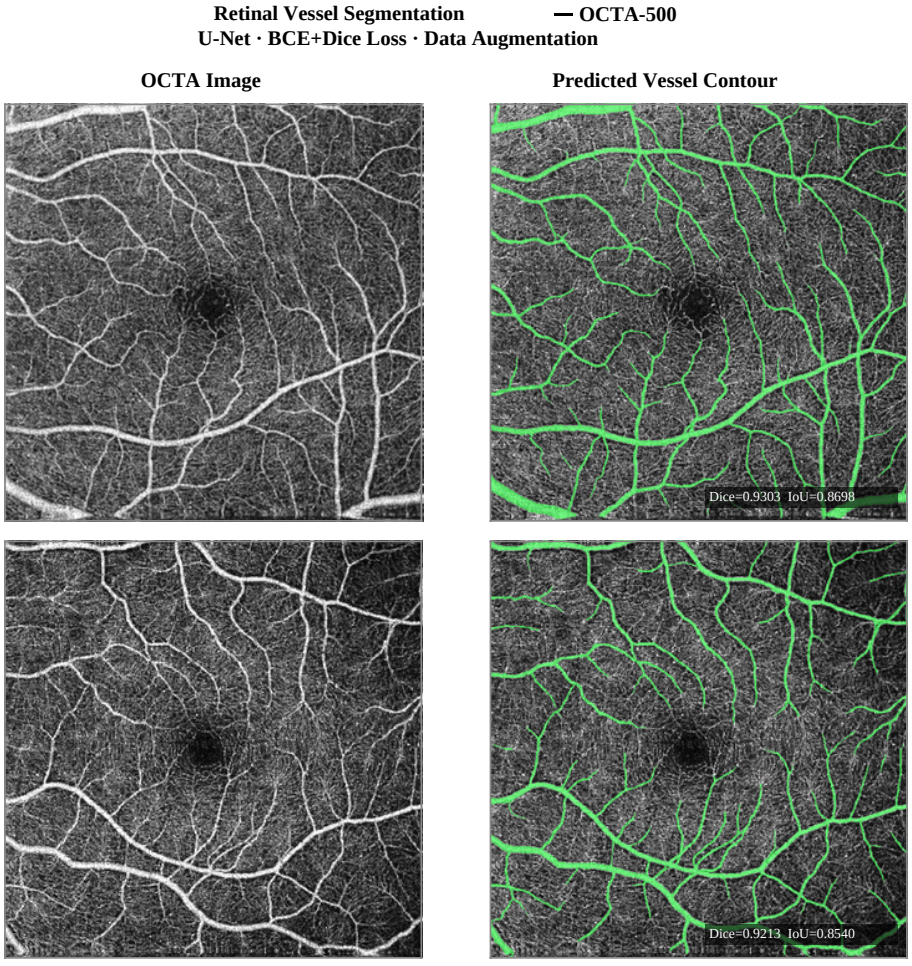


Figure 9. Segmentation examples from the representative model on two high-quality test subjects. Left: OCTA B5 map; right: predicted contour overlaid in green. Both exceed Dice 0.91 and IoU 0.84, with accurate delineation of the main vasculature and preservation of the FAZ; deviations concentrate in small-caliber peripheral branches where local contrast is reduced.

clinical expertise and supported the interpretation of retinal OCTA findings. Ismael Garcia Santos and Gregory Stein contributed to the methodological and computational aspects of the study and critically revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and materials

The data analyzed in this study are publicly available in the OCTA-500 dataset Li *et al*. (2023b). The processed data, trained models, and source code are available from the corresponding author upon reasonable request.

Further relevant information

This study used only the publicly available, de-identified OCTA-500 dataset Li *et al*. (2023b); therefore, no additional ethics approval was required.

References

Abtahi, M., Le, D., Lim, J. I., and Yao, X. (2022). MF-AV-Net: An open-source deep learning network with multimodal fusion options for artery-vein segmentation in OCT angiography. *Biomedical Optics Express*, 13(9):4870–4888. DOI: 10.1364/BOE.468483.

Alam, M., Le, D., Son, T., Lim, J. I., and Yao, X. (2020). AV-Net: Deep learning for fully automated artery-vein classification in optical coherence tomography angiography. *Biomedical Optics Express*, 11(9):5249–5263. DOI: 10.1364/BOE.399514.

Gao, S. S., Jia, Y., Zhang, M., Su, J. P., Liu, G., Hwang, T. S., Bailey, S. T., and Huang, D. (2016). Optical coherence tomography angiography. *Investigative Ophthalmology & Visual Science*, 57(9):OCT27–OCT36. DOI: 10.1167/iovs.15-19043.

Kashani, A. H., Chen, C.-L., Gahm, J. K., Zheng, F., Richter, G. M., Rosenfeld, P. J., Shi, Y., and Wang, R. K. (2017). Optical coherence tomography angiography: A comprehensive review of current methods and clinical applications. *Progress in Retinal and Eye Research*, 60:66–100. DOI: 10.1016/j.preteyeres.2017.07.002.

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., and Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, 22(1):69. DOI: 10.1186/s12880-022-00793-7.

Li, M., Chen, Y., Ji, Z., Xie, K., Yuan, S., Chen, Q., and Li, S. (2020). Image projection network: 3D to 2D image segmentation in OCTA images. *IEEE Transactions on Medical Imaging*, 39(11):3343–3354. DOI: 10.1109/TMI.2020.2992244.

Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., and Chen, Q. (2023a). OCTA-500: A retinal dataset for optical coherence tomography angiography study. *IEEE Transactions on Medical Imaging*, 42(6):1539–1556. DOI: 10.1109/TMI.2022.3228490.

Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., and Chen, Q. (2023b). OCTA-500: A retinal dataset for optical coherence tomography angiography study. *IEEE Transactions on Medical Imaging*, 42(6):1539–1556. DOI: 10.1109/TMI.2023.3235614.

Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In *Advances in Neural Information Processing Systems*, volume 32.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer. DOI: 10.1007/978-3-319-24574-4_28.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., and Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693. DOI: 10.1016/j.media.2020.101693.

Xu, Q., Li, M., Ji, Z., Yuan, S., Liu, Q., and Chen, Q. (2022). Foveal avascular zone volume: A new index based on 3D U-Net segmentation for OCTA images. *Frontiers in Neuroscience*, 16:887420. DOI: 10.3389/fnins.2022.887420.

Zhang, H., Ji, Z., Fan, W., Yuan, S., Liu, Q., and Chen, Q. (2020). Robust layer segmentation against complex retinal abnormalities for en face OCTA generation. In *Ophthalmic Medical Image Analysis - MICCAI 2020 Workshop*, volume 12437 of *Lecture Notes in Computer Science*, pages 112–121. Springer. DOI: 10.1007/978-3-030-59722-1_62.