VILNIUS UNIVERSITY

FACULTY OF MATHEMATICS AND INFORMATICS

DATA SCIENCE STUDY PROGRAMME

Master's Thesis

Impact of Downscaling Fundus Images and Vessel Masks on Thin Retinal Vessel Segmentation

Akies dugno vaizdų ir kraujagyslių anotacijų mastelio mažinimo įtaka smulkiųjų tinklainės kraujagyslių segmentavimui

Antanas Bukauskas

Supervisor : dr. Jolita Bernatavičienė

Vilnius
2026

Summary

The retinal microvasculature provides valuable biomarkers for systemic diseases, yet accurate segmentation of thin retinal vessels remains challenging, particularly under resolution constraints imposed by modern deep learning pipelines. While many retinal vessel segmentation studies report strong overall performance, they rarely distinguish between vessel calibers or quantify how image and mask downscaling disproportionately affects thin vessels that carry important clinical information. This work systematically investigates how fundus image and vessel mask resolution reduction impacts thin retinal vessel structure and subsequent segmentation accuracy.

We establish a baseline retinal vessel segmentation model using high-resolution images from the FIVES dataset and the HRFormer architecture and propose a morphological algorithm to separate thin and thick vessels in binary segmentation masks based on vessel thickness and image resolution. This enables class-specific evaluation without reframing segmentation as a multi-class problem. Using this framework, we quantify structural information loss (defined as F1-score error) introduced in ground truth masks by different mask downscaling strategies. We also assess segmentation performance across progressively lower fundus image resolutions while keeping model architecture and training conditions fixed.

Our results show that linear interpolation methods (bilinear, bicubic and distance transform) best preserve vessel structural information during mask downscaling, but require resolution-specific threshold calibration, as the conventional 50% intensity threshold is suboptimal across scales. Thin ground truth vessels are substantially more affected by resolution reduction than thick vessels: even under optimal conditions, thin vessel F1-score decreases from 0.9286 at 1024×1024 to 0.7014 at 256×256, that is 2.6 times more structural information loss than thick vessels for given resolutions. Overall, reducing mask resolution by a factor of 2 constitutes an equivalent drop in structural information for both thin and thick vessels which leaves use of downscaled resolutions in retinal research a questionable choice.

Fundus image resolution experiments reveal that 512×512 constitutes a practical lower bound for thin vessel analysis, as models trained at 1024×1024 and 512×512 achieve statistically indistinguishable performance, though a big disparity is left between thin and thick vessels (F1-score of 0.70 for thin and 0.91 for thick vessel segmentation). Further reducing the resolution significantly affects thin vessel predictions while thick vessels remain relatively intact. Recall is identified as the primary metric impacted by resolution reduction, indicating loss of vessel connectivity rather than boundary distortion. Despite comprising only about 15% of vessel pixels, thin vessels contribute approximately 78% of normalized segmentation errors at practical resolutions, highlighting their disproportionate influence on model output quality. These findings demonstrate that overall segmentation metrics can obscure clinically relevant segmentation failures.

Keywords: eye fundus segmentation, fundus image downscaling, high-resolution deep learning, retinal vessel morphology, thin vessels

Santrauka

Tinklainės mikrokraujagyslių tinklas turi vertingus biologinius žymenis sisteminių ligų diagnostikai, tačiau tiksli plonų tinklainės kraujagyslių segmentacija išlieka sudėtinga, ypač esant mastelio apribojimams, kuriuos lemia šiuolaikinės giliojo mokymosi sistemos. Nors daugelyje tinklainės kraujagyslių segmentavimo tyrimų pateikiami geri bendrieji rezultatai, juose retai atskiriami kraujagyslių storiai arba kiekybiškai įvertinama, kaip vaizdų ir kraujagyslių anotacijų mažinimas neproporcingai paveikia plonas kraujagysles, turinčias didelę klinikinę reikšmę. Šiame darbe sistemingai tiriama, kaip akių dugno vaizdų ir kraujagyslių anotacijų mastelio mažinimas veikia plonų tinklainės kraujagyslių struktūrą ir vėlesnį segmentavimo tikslumą.

Kaip ekspermentų atskaitos tašką mes naudojame tinklainės kraujagyslių segmentavimo modelį su aukštos raiškos FIVES duomenų rinkinio vaizdais ir HRFormer architektūra, bei pasiūlome morfologinį algoritmą, kuris atskiria plonas ir storas kraujagysles dvinarėse segmentavimo anotacijose pagal kraujagyslių storį ir vaizdo mastelį. Tai leidžia atlikti klasių atžvilgiu atskirą vertinimą nekeičiant uždavinio į daugiaklasę segmentaciją. Pasitelkus šį algoritmą, kiekybiškai įvertinamas struktūrinės informacijos praradimas (apibrėžtas kaip F1 rodiklio klaidos dydis), atsirandantis anotacijose dėl skirtingų mastelio mažinimo strategijų. Taip pat analizuojamas segmentavimo tikslumas palaipsniui mažinant akių dugno vaizdų mastelį, išlaikant nekintančią modelio architektūrą ir mokymo sąlygas.

Rezultatai rodo, kad tiesinės interpoliacijos metodai geriausiai išsaugo kraujagyslių struktūrinę informaciją anotacijų mažinimo metu, tačiau reikalauja slenkstinės kalibracijos skirtingiems masteliams, nes įprastas 50% intensyvumo slenkstis nėra optimalus visais masteliais. Plonos anotuotos kraujagyslės yra žymiai labiau paveikiamos mastelio mažinimo nei storos: net ir optimaliomis sąlygomis plonų kraujagyslių F1 rodiklis sumažėja nuo 0.9286 esant 1024×1024 iki 0.7014 esant 256×256, kas yra 2,6 karto didesnis struktūrinės informacijos praradimas nei storoms kraujagyslėms esant tiems patiems masteliams. Bendrai anotacijų mastelio sumažinimas per pusę lemia atitinkamai didelį struktūrinės informacijos praradimą plonoms ir storoms kraujagyslėms, kur sumažintų mastelių naudojimas yra abejotinas pasirinkimas tinklainės vaizdų srityje.

Akių dugno vaizdų mastelio mažinimo eksperimentai parodo, kad 512×512 yra žemiausias vartotinas mastelis plonų kraujagyslių analizei, nes modeliai, apmokyti su 1024×1024 ir 512×512 raiška, pasiekia statistiškai nesiskiriančius rezultatus, tačiau didelis skirtumas išlieka tarp plonų ir storų kraujagyslių (F1 rodiklis siekia apie 0.70 plonoms ir 0.91 storoms kraujagyslėms). Tolesnis mastelio mažinimas reikšmingai pablogina plonų kraujagyslių segmentavimą, tuo tarpu storos kraujagyslės išlieka palyginti nepakitusios. Nustatyta, kad atkūriamumas (recall) yra pagrindinis rodiklis, jautriausias mastelio mažinimui, rodantis kraujagyslių vientisumo praradimą. Nors plonos kraujagyslės sudaro tik apie 15% visų kraujagyslių pikselių, jos lemia maždaug 78% normalizuotų segmentavimo klaidų vartojamuose masteliuose, pabrėždamos neproporcingai didelę jų įtaką segmentavimo modelio išvesties kokybei. Šie rezultatai parodo, kad bendrieji segmentavimo rodikliai gali nuslėpti kliniškai reikšmingas segmentavimo klaidas.

Raktiniai žodžiai: akies dugno segmentavimas, akies dugno vaizdų spaudimas, aukštos raiškos gilusis mokymasis, tinlainės kraujagyslių morfologija, siauros kraujagyslės

Contents

Introduction

The retina provides a unique window into systemic health, with its microvascular network serving as a direct indicator of cardiovascular, metabolic and neurological conditions [15]. The retinal vasculature can be observed non-invasively through fundus imaging [1]. Changes in retinal vessel structural information including caliber variations, tortuosity alterations and microvascular dropout can act as indicators for developing diseases. Fundus imaging therefore enables early detection and longitudinal monitoring of conditions including diabetic retinopathy, hypertensive retinopathy, glaucoma and even cardiovascular disease risk [35].

Retinal vessel segmentation has attracted substantial research attention over the past two decades [8, 35, 42]. Accurate vessel segmentation enables downstream tasks such as vessel diameter measurement [24, 55], arteriovenous ratio calculation [21], tortuosity quantification [43, 58] and bifurcation analysis [17, 31]. These metrics provide objective biomarkers for disease progression and treatment response. However, while existing studies report overall segmentation performance using metrics like dice coefficient or F1-score, they rarely distinguish between vessel calibers despite their markedly different clinical significance and segmentation difficulty [3, 54]. The retinal vascular network exhibits a hierarchical structure ranging from large vessels through medium-caliber vessels (arterioles and venules) to the finest capillaries. Thin vessels, primarily comprising capillaries and small arterioles, provide critical information about microvascular health and early pathological changes. In diabetic retinopathy, for example, capillary dropout and microaneurysm formation occur before larger vessel abnormalities become apparent. Similarly, hypertensive microvascular changes affect small arterioles before major vessels show measurable alterations [15]. The accurate segmentation of these thin vessels therefore holds particular clinical value for early disease detection and monitoring. Yet thin vessel segmentation presents unique technical challenges. These structures occupy only a few pixels in width even at high resolutions, making them highly susceptible to image noise, contrast variations and artifacts [8]. They often exhibit lower contrast against the background compared to thick vessels due to reduced blood volume and thinner vessel walls. Their small size makes them vulnerable to partial volume effects where a single pixel contains contributions from both vessel and background tissue. Furthermore, thin vessels form dense, complex networks with frequent crossings and bifurcations, creating ambiguous regions that challenge both human annotators and automated algorithms [42].

The relationship between image resolution and segmentation accuracy remains poorly understood, particularly for thin vessels. Downscaling segmentation masks inevitably distorts or eliminates structural information in fine structures. For fundus vessel trees, this degradation manifests as broken connectivity in thin vessels where continuous capillaries fragment into disconnected segments [12, 28, 33], exaggerated vessel tortuosity due to discretization artifacts or altered vessel thickness as multi-pixel structures collapse to single-pixel representations or disappear entirely [43]. Analysis of low-resolution datasets such as DRIVE, which contains images at 565×584 resolution, reveals that human annotators prioritize preserving vessel connectivity even when thin vessels barely distinguish from background noise [14]. This annotation behavior suggests that connectivity represents a more

clinically meaningful feature than pixel-perfect boundary delineation for thin vessels. Despite the clear impact of resolution on thin vessel representation, papers proposing new vessel segmentation methods rarely specify their scaling procedures or acknowledge the resulting information loss. Many studies train models at resolutions different from the original image acquisition, applying downscaling to manage computational constraints without reporting the interpolation methods used or quantifying the structural changes introduced. This omission makes it difficult to compare results across studies, as models trained at different resolutions may be solving fundamentally different problems: one preserving fine capillary networks while another segments only major vessels that survive downscaling [25, 26].

Similarly, fundus image downscaling degrades visual features critical for vessel detection [12]. This degradation occurs even with sophisticated interpolation methods, as information genuinely lost through downscaling cannot be recovered. The choice of image resolution involves fundamental trade-offs. Higher resolutions preserve thin vessel details and enable more accurate structural measurements, but demand greater computational resources for storage, processing and model training. Modern deep learning architectures, particularly those based on convolutional neural networks with multiple downscaling layers, face severe memory constraints at high resolutions. Conversely, lower resolutions enable efficient training and inference but may sacrifice precisely the thin vessel information that holds great clinical value.

Our work gives insights to the following questions: How much vessel ground truth information is lost when downscaling annotated retinal vessel masks? At what point does fundus image resolution reduction degrade vessel information beyond the threshold where accurate segmentation remains possible? Do different downscaling methods preserve vessel structural information differently? And crucially, how do these choices affect not just overall segmentation metrics which may be dominated by easily detected thick vessels but specifically the detection of the clinically important thin vessel network?

Goal: This study aims to quantify the effects of fundus image and mask downscaling on thin retinal vessel structural information and to determine the subsequent impact on vessel segmentation accuracy.

Tasks:

1. Establish a robust baseline for retinal vessel segmentation by selecting an appropriate high-resolution model architecture, providing a consistent dataset split and adopting evaluation metrics that account for severe class imbalance between vessel and background pixels.

2. Develop and implement a morphological algorithm that separates thin and thick retinal vessels from binary segmentation masks using a resolution-dependent thickness criterion, enabling independent analysis of segmentation performance for each vessel caliber. Define what vessels are considered as thin.

3. Define and quantify structural information loss in retinal vessels caused by different segmentation mask downscaling strategies across multiple target resolutions and comparing how thin and thick vessels are affected individually.

4. Evaluate the impact of fundus image resolution on vessel segmentation accuracy by conducting controlled training experiments at progressively lower effective input resolutions, while keeping model architecture and training conditions fixed. Measure thin vessel contribution to the overall error.

The study proceeds as follows: the literature review section examines existing work in retinal vessel segmentation, giving context to disease-specific vascular changes, segmentation challenges, vessel morphology, model architectures and related works while identifying gaps in resolution-dependent analysis and thin vessel-specific evaluation; the data section describes datasets considered for this study, detailing their characteristics, annotation procedures and suitability for resolution experiments; the methodology section presents vessel separation algorithm, describes downscaling experiments and evaluation framework; the baseline establishment section justifies choices for model architecture, dataset selection, training procedures and vessel thickness thresholds through preliminary experiments and literature precedent; results and discussion section analyzes and provides key insights to experimental findings for mask downscaling and fundus image resolution effects; finally, limitations acknowledge constraints in experimental scope and generalizability, conclusions summarize contributions and future work outlines remaining challenges and promising research directions for advancing thin vessel segmentation.

Literature Review

Imaging methods. [1] Human eye comprises several critical structures visible during examination, including the optic disc (where the optic nerve enters the eye), retinal blood vessels (arteries and veins) and the macula (responsible for central vision). Abnormalities in each of these parts or surrounding area could be a sign of developing pathology. To observe changes in human eye, medical experts use several imaging modalities that enable visualization and analysis of retinal structures, each providing unique diagnostic capabilities.

Fundoscopy or fundus imaging represents the foundational examination technique, utilizing direct or indirect ophthalmoscopy to visualize the posterior segment through the pupil. This method provides real-time assessment of retinal vessels, optic disc and macula using white light illumination, though image quality depends on media clarity and pupil dilation.

Optical coherence tomography (OCT) employs low-coherence interferometry to generate high-resolution cross-sectional images of retinal layers. OCT enables precise measurement of retinal thickness, identification of intraretinal fluid and detailed visualization of photoreceptor layers and retinal pigment epithelium. This non-invasive technique proves essential for monitoring macular pathology and quantifying structural changes over time.

Optical coherence tomography angiography (OCTA) extends OCT capabilities by detecting motion contrast from flowing blood, creating depth-resolved vascular maps without contrast injection. OCTA reveals capillary networks and visualizes neovascularization with superior detail compared to traditional angiography methods.

Fluorescein angiography involves intravenous fluorescein injection followed by sequential fundus photography to document retinal circulation dynamics. This gold-standard technique demonstrates vascular leakage, perfusion defects and neovascularization through real-time fluorescence patterns, providing crucial information for diagnosing and managing retinal vascular diseases.

Of these eye examination methods, the most readily available data is for eye fundus images acquired through fundoscopy. In this study we will focus on fundus image processing.

Vessel Abnormalities Detected in Fundus Images. [15] Retinal vascular changes provide essential diagnostic markers across multiple pathological conditions. In nonproliferative diabetic retinopathy (NPDR), *intraretinal microvascular abnormalities* (IRMA) represent dilated capillary beds within the retina. *Venous beading*, characterized by focal dilatations alternating with constricted segments along retinal veins, indicates severe retinal hypoxia and impending progression to proliferative disease. *Neovascularization* manifests as abnormal vessel growth, with neovascularization at the disc (NVD) and neovascularization elsewhere (NVE) representing hallmarks of proliferative diabetic retinopathy (PDR) and retinal vein occlusion (RVO). These pathological vessels demonstrate irregular caliber, abnormal branching patterns. In glaucoma, *bayoneting of retinal vessels* occurs as vessels traverse the excavated optic cup margin, creating characteristic angular deflections. Hypertensive retinopathy (HR) produces distinctive vascular changes including *copper or silver wire arterioles*, where increased arterial wall thickness creates metallic light reflexes. *Congested retinal veins* reflect elevated retinal venous pressure associated with systemic hypertension. RVO generates col-

lateral vessels as alternative drainage pathways, accompanied by *dilated and tortuous veins* proximal to the occlusion site. Retinal artery occlusion (RAO) presents with *attenuated arteries* and may demonstrate *embolic manifestation* as visible occlusive material within arterial lumens.

Retinal Vessel Segmentation Challenges. Retinal vessel segmentation is commonly formulated as a pixel-wise classification task, where each pixel in a fundus image is assigned to either the vessel or background class. While many studies report strong overall segmentation performance, difficulties in segmenting thin vessels are often discussed primarily through qualitative visual inspection rather than quantitative, class-specific metrics [3, 54, 57]. As a result, claims regarding thin vessel degradation frequently lack direct numerical support, making it difficult to assess the magnitude and consistency of these errors across datasets and resolutions.

Thin retinal vessels play an important role in characterizing microvascular changes associated with diseases such as DR, HR and RVO. However, these vessels present inherent segmentation challenges due to their small caliber, low contrast and limited pixel representation, particularly in peripheral regions of the fundus image [8]. Suggesting that even a minor reductions in spatial resolution can render thin vessels indistinguishable from background tissue, leading to discontinuities or complete loss in segmentation outputs.

The hierarchical structure of the retinal vasculature, spanning large arteries and veins down to dense capillary networks, requires segmentation methods that operate effectively across multiple spatial scales [42]. While recent imaging systems are capable of acquiring high-resolution fundus images, many segmentation pipelines downscale these images early in processing or divide them into smaller patches, potentially discarding fine-grained vascular information and global context. The impact of these design choices on thin vessel preservation is rarely evaluated explicitly.

Additional challenges arise from complex vessel topology and pathological alterations. Vessel crossings and bifurcations introduce ambiguity in continuity estimation [32], while pathological changes such as increased tortuosity, arteriovenous nicking and neovascularization disrupt typical vessel appearance. Image artifacts and co-occurring lesions, including hemorrhages and exudates, further complicate segmentation by obscuring or mimicking vascular structures [42].

Segmentation performance is also influenced by class imbalance at both the pixel and dataset levels. Vessel pixels constitute a small fraction of the image, with thin vessels representing an even smaller subset. In addition, clinical datasets often contain disproportionate numbers of healthy or mildly affected cases compared to severe pathologies [35]. These imbalances can bias models toward dominant patterns and reduce sensitivity to subtle vascular features. Consequently, aggregate performance metrics may obscure systematic failures in thin vessel detection, underscoring the need for resolution-aware, class-specific quantitative evaluation.

Retinal Vessel Morphology. Morphological features of retinal vessels in fundus images can be quantified to support objective assessment of vascular structure and disease-related changes. Commonly analyzed features include vessel diameter, bifurcation geometry, tortuosity, and topological connectivity, all of which depend on accurate vessel delineation.

Several methods have been proposed for vessel thickness estimation. Huang et al. [24] fit cubic splines to vessel centerlines and estimate local vessel width by matching gray-level intensity

profiles perpendicular to the centerline at vessel edges. Zafer et al. [55] propose a vessel tracking approach that detects vessel edges using intensity gradients and computes vessel diameter as the perpendicular distance between fitted edge lines along the tracked centerline. Both approaches rely on precise boundary detection and are sensitive to segmentation quality and image resolution.

Bifurcation detection has also received considerable attention. Feng et al. [17] segment retinal vessels using a fully convolutional network and identify bifurcations by detecting Y- and T-shaped patterns in the binary vessel mask using morphological operations. Kalaie et al. [31] propose a hierarchical probabilistic graphical model based on Gaussian basis functions for intensity interpolation, classifying local cross-sections as vessel or junction points using class-specific gamma distributions. Their method jointly tracks vessels and detects bifurcations within a unified probabilistic framework.

Vessel tortuosity is another widely studied morphological characteristic. Moss et al. [36] introduce a tortuosity index based on detecting inflection points along vessel centerlines using tangent line angle changes and aggregating segment-level tortuosity ratios across vessel length. Zou et al. [58] propose a derivative-based tortuosity metric that incorporates both curved and near-linear segments, improving consistency with clinical visual assessment compared to purely curvature-based measures.

Connectivity of predicted vessel masks is often improved through post-processing under the assumption that the retinal vasculature forms a continuous tree. Morphological top-hat transforms are commonly applied to enhance thin vessels and reconnect fragmented structures [18, 19]. More recent approaches explicitly model connectivity. Su et al. [48] propose a topology-aware connectivity boosting method that estimates prediction uncertainty, identifies regions of intermediate confidence, and selectively reconnects broken vessel segments in these areas. Such methods highlight the importance of structural continuity, particularly for thin vessels that are prone to fragmentation during segmentation and downscaling.

Deep Neural Network Retinal Vessel Segmentation Methods. Most deep neural network (DNN) approaches to retinal vessel segmentation adopt a patch-based training strategy, in which fundus images are divided into a large number of small, often overlapping patches. Unlike traditional machine learning methods that rely on manually engineered features, DNN-based models learn hierarchical feature representations directly from data. As a result, extensive preprocessing is generally not required, but these models depend heavily on large and diverse training datasets to generalize effectively to unseen images. Since many early retinal vessel segmentation datasets contain only a limited number of expert-annotated fundus images, patch-based methods are commonly used to artificially increase the number of training samples.

Patch-based processing also addresses computational constraints. For conventional convolutional neural networks (CNNs), computational complexity increases quadratically with input resolution, making full-resolution training impractical on limited hardware. By restricting input size to small patches, models can be trained more efficiently while maintaining acceptable memory usage. BridgeNet [56], for example, incorporates residual connections within a CNN framework to generate vessel probability maps and introduces patch-based loss weighting to mitigate vessel-background class imbalance. Oliveira et al. [38] propose an FCN-based approach augmented with additional

input channels derived from stationary wavelet decomposition, demonstrating that domain-specific information can complement deep feature learning. MFA-UNet [6] further extends CNN-based architectures by incorporating self-attention mechanisms and multi-branch decoding modules to improve microvascular segmentation accuracy.

Despite their practicality, patch-based methods introduce several limitations. Training on large numbers of patches increases computational cost and training time as dataset size grows. More importantly, patch-based processing sacrifices global fundus image context, which is essential for modeling vessel connectivity, hierarchical structure and long-range dependencies. These limitations motivate the exploration of segmentation methods that operate directly on high-resolution fundus images.

Only a limited number of studies attempt retinal vessel segmentation using high-resolution inputs exceeding 1000×1000 pixels. These approaches typically modify standard network architectures to reduce computational complexity while preserving spatial detail. VesselView [46] applies a standard U-Net architecture to high-resolution inputs without substantial architectural changes and reports a notable decline in segmentation performance, highlighting the limitations of directly scaling conventional models. SuperVessel [23] addresses this issue using a dual-path network with a shared encoder, where one path performs fundus image super-resolution as an auxiliary task to refine vessel predictions. Pethmunee et al. [40] employ a DeepLabV3 backbone that leverages atrous and dilated convolutions alongside atrous spatial pyramid pooling to expand the receptive field while reducing computational cost. HRD-Net [34] introduces deformable convolutions to adapt the receptive field to irregular vessel geometry and adopts a multi-path design inspired by HRNet [51], preserving full-resolution representations in one branch while processing lower-resolution features in parallel.

Overall, existing deep learning approaches demonstrate strong performance on retinal vessel segmentation benchmarks but often rely on design choices such as patch-based processing or early downsampling that may disproportionately affect thin vessel representation. The impact of these choices on thin vessel preservation, particularly under varying resolution constraints, remains insufficiently quantified in current literature.

Related Work on Fundus Image Rescaling. In many retinal vessel segmentation studies, downscaling of fundus images is accompanied by a corresponding reduction in output mask resolution. While this approach simplifies model design and reduces computational cost, it introduces distortions in the ground truth representation that complicate result interpretation. Excessive resolution reduction can alter vessel structural information, limiting the applicability of segmentation outputs for downstream clinical and analytical tasks.

Domarkaitė [12] investigates how overall vessel segmentation performance is affected by different interpolation methods and target resolutions across multiple retinal datasets. The study employs a U-Net architecture trained on patches extracted from merged datasets. The results indicate that bilinear interpolation to a resolution of 750×750 yields optimal performance for datasets with sufficiently high original resolution. In contrast, datasets with lower native resolution experience degraded segmentation accuracy when being upscaled to higher resolutions, suggesting that the effectiveness of rescaling depends strongly on the available spatial detail in the source images.

Zihuang et al. [53] apply a pretrained Segment Anything Model 2 (SAM2) [44] with a custom decoder to ultra-widefield fundus images. Their work demonstrates that resizing images beyond their native resolution negatively affects segmentation quality. The model is successfully trained on full-resolution 2048×2048 images without patch-based processing and the authors show that maintaining high spatial resolution leads to improved performance compared to models trained on lower-resolution inputs.

Ramirez et al. [43] study the effect of fundus image resolution on vessel tortuosity measurements rather than segmentation accuracy. They report that some tortuosity indices remain largely invariant to resolution changes, while others scale proportionally with the downsampling factor. However, when image resolution is reduced below 400×400, the metrics begin to fluctuate unpredictably. The authors attribute this behavior to the loss of structural vessel information rather than simple geometric scaling.

Notably, existing studies do not explicitly quantify information loss in ground truth vessel masks as a function of resolution. Furthermore, when evaluating segmentation performance at different resolutions, ground truth masks are typically downscaled alongside the input images. As a result, reported metrics primarily reflect a model's ability to reproduce a deformed version of the ground truth rather than its capacity to infer vessel structures that are lost due to resolution reduction.

Related Work on Thin Vessel Segmentation Quantification. Publicly available retinal vessel segmentation datasets do not provide explicit annotations distinguishing thin and thick vessels. Consequently, researchers seeking class-specific evaluation must first derive such labels from binary ground truth masks. Only a limited number of studies attempt this type of quantitative separation.

Yan et al. [25] and Yang et al. [26] manually separate vessel ground truth masks into thin and thick vessel classes and train multi-module networks to segment each class separately. Both approaches employ three-stage architectures consisting of a thick vessel segmentation module, a thin vessel segmentation module and a fusion module that combines their outputs. Yan et al. use fully convolutional networks (FCNs), while Yang et al. adopt a U-Net-based design. Their results consistently show lower segmentation accuracy for thin vessels compared to thick vessels, although explicitly incorporating thin vessel supervision improves performance.

Christodoulidis et al. [10] propose an automated vessel separation approach based on vessel centerline extraction and width estimation along the normal direction of the centerline. The method combines multiple preprocessing steps to obtain an initial vessel mask, followed by multi-scale tensor voting to reconnect fragmented thin vessels. This strategy improves sensitivity for thin vessel segmentation, particularly in challenging regions.

Despite these contributions, existing evaluation strategies treat thin vessels as a distinct semantic class, penalizing predictions that confuse thin and thick vessels even though both represent correct vessel detections in a binary segmentation context. Moreover, these methods implicitly assume that model outputs can directly represent thin vessel predictions, despite being trained on binary vessel-background labels. This mismatch makes ground truth and prediction comparisons less consistent. Additionally, most works provide only high-level descriptions of how thin vessel masks are derived without going into details on methods used.

In contrast, this thesis focuses on resolution-aware thin vessel separation and evaluation without reframing the problem as multi-class segmentation. By separating vessel calibers at the evaluation stage rather than during training, the proposed approach enables class-specific performance analysis while remaining consistent with the binary nature of standard retinal vessel segmentation models.

1 Data

For initial testing we utilized five retinal vessel segmentation datasets but later settled on a singular dataset for training, validation and testing (provided our reasoning at section 3.2). Table Table 1 summarizes the properties of multiple vessel segmentation datasets.

Table 1 *List of popular open source retina datasets used for vessel segmentation.*

Dataset	**Size**	**Resolution**	**FOV**	**Devices**
FIVES [30]	800	2048×2048	50	Topcon TRC-NW8
DRIVE [14]	40	565×584	45	Canon CR-5 + 3CCD
STARE [47]	20	700×605	35	Topcon TRV-50
HRF [22]	45	3304×2336	45	Canon CR-1
CHASE_DB1 [20]	28	999×960	30	Nidek NM-200-D
Leuven-Haifa [49]	240	1444×1444	30	Visucam 500
LES_AV [39]	22	1620×1444	45	-
INSPIRE-AVR [37]	40	2392×2048	-	-
IOSTAR [45]	30	1024×1024	45	EasyScan camera1
ARIA [5]	50	768×576	50	-
UoA-DR [9]	200	2124×2056	45	Zeiss Visucam 500
DRIONS-DB [7]	110	600×400	-	HP-PhotoSmart-S20
IDRiD [41]	81	4288×2848	50	Kowa VX-10

1) FIVES (Fundus Image Vessel Segmentation) [30] is the largest publicly available vessel segmentation dataset, containing 800 high-resolution images captured at 2048×2048 pixels with a 50° field of view using the Topcon TRC-NW8 fundus camera. The images were collected from routine clinical examinations at multiple medical centers, encompassing both healthy subjects and patients with diabetic retinopathy, glaucoma, age-related macular degeneration (200 fundus images each), which contributes to the dataset's diversity in vessel appearance and image quality. This substantial training set size enables training deep learning models without requiring image patching. The elimination of patching artifacts and the preservation of global vascular structure context make FIVES our primary dataset for downscaling experiments.

2) DRIVE (Digital Retinal Images for Vessel Extraction) [14] comprises 40 images at 565×584 resolution with a 45° field of view, captured using a Canon CR-5 non-mydriatic 3CCD camera. Despite its small size and relatively low resolution, DRIVE remains one of the most widely cited benchmarks in vessel segmentation literature. Each image includes manual segmentations from two observers, providing insight into inter-annotator variability.

3) STARE (STructured Analysis of the Retina) [47] is the smallest dataset in our study, containing 20 images at 700×605 resolution with a narrow 35° field of view, acquired using a Topcon TRV-50 fundus camera. STARE is notable for including images with various pathologies, offering diversity in vascular appearance and image quality.

4) HRF (High-Resolution Fundus) [22] contains 45 ultra-high-resolution images at 3304×2336 pixels with a 45° field of view, captured using a Canon CR-1 fundus camera. The dataset is divided into three subsets: healthy subjects, diabetic retinopathy patients and glaucomatous patients. HRF's

exceptional resolution makes it particularly valuable for cross-validating findings from downscaling experiments, as it preserves fine vessel details that may be lost in lower-resolution datasets.

5) CHASE_DB1 (Child Heart and Health Study in England Database 1) [20] consists of 28 images at 999×960 resolution with a 30° field of view, acquired from children using a Nidek NM-200-D hand-held fundus camera. This dataset is unique in being captured from a pediatric population, resulting in images with different vascular characteristics and potentially higher image noise due to the handheld acquisition method.

2 Methodology

This section describes the experimental framework and algorithmic procedures used to investigate how spatial downscaling affects the representation and segmentation of retinal vessels of different calibers. Our methodology is structured to explicitly disentangle thin and thick vessel behavior, both at the level of ground truth annotations and model predictions, enabling a fine-grained analysis that is not possible with conventional binary vessel evaluation. We first introduce algorithms for separating vessel masks into thin and thick components and for consistently classifying predicted pixels under boundary uncertainty. Building on this foundation, we then quantify morphological ground truth mask distortions introduced by different downscaling strategies and resolutions and finally evaluate how reduced image resolution impacts segmentation model performance with particular emphasis on the vulnerability of thin vascular structures.

2.1 Algorithmic Thin Vessel Separation

To analyze the differential impact of downscaling on thin versus thick vessels, we require a method to separate these vessel classes in both ground truth masks and model predictions. Existing vessel segmentation datasets typically provide only binary masks without vessel diameter annotations. We therefore developed two complementary algorithms: one for separating ground truth masks into thin and thick vessel components and another for classifying predicted vessel pixels according to the ground truth separation. Both algorithms rely on morphological operations commonly used in literature [18, 19] which we adapt for our own purposes and a thickness threshold that distinguishes thin vessels (diameter at or below threshold) from thick vessels (diameter above threshold). The selection of this threshold is detailed in section 3.4.

2.1.1 Algorithm for Ground Truth Masks

The ground truth separation algorithm separates binary vessel masks into thin and thick vessel components through a multi-step morphological process. The input is a binary mask where white pixels represent vessels and black pixels represent background. The output consists of two masks: one for thin vessels and one for thick vessels. For visualization purposes throughout this paper, we display these as a single composite image where thin vessels appear red, thick vessels appear white and background remains black.

A thickness threshold value defines the boundary between thin and thick vessels. Vessel structures at or below this thickness are classified as thin, while all other vessel portions are classified as thick. Our morphological operations center on this threshold value, though we apply small offsets in certain steps to improve visual quality and connectivity.

Step 1: Upscaling and morphological opening. We begin by upscaling the reference vessel mask by a factor of two using nearest neighbor interpolation. This upscaling enables morphological operations with odd-valued kernel radii, eliminating sub-pixel thickness complications and improving precision when the result is subsequently downscaled. We then apply a modified morphological

opening (white top-hat transform) to the upscaled mask. First, we perform erosion using a circular structuring element with radius equal to the thickness threshold:

$$E = M \ominus B_r \tag{1}$$

where M is the upscaled vessel mask, B_r is a circular structuring element of radius r (the thickness threshold) and $\ominus$ denotes morphological erosion. This operation removes superimposed pixels of M from center of B_r effectively erasing thin vessel structures which do not fit into the B_r and making all vessels thinner by the erosion radius. The circular structuring element ensures smooth resulting edges.

Next, we dilate the eroded mask using a circular structuring element of radius $r + 1$:

$$D = E \oplus B_{r+1} \tag{2}$$

where $\oplus$ denotes morphological dilation. The slightly larger dilation radius compensates for imperfect edge reconstruction. Redilated vessels do not perfectly align with original thick vessel boundaries, so this offset prevents small disconnected artifact pixels at vessel edges. The resulting mask D approximates the thick vessel tree, having eliminated thin structures while restoring thick vessel dimensions.

We downscale this thick vessel approximation back to the original resolution using nearest neighbor interpolation, then obtain the thin vessel mask through logical conjunction:

$$M_{\text{thin}} = M_{\text{ground truth}} \wedge \neg D_{\text{downscaled}} \tag{3}$$

where $M_{\text{ground truth}}$ is the original ground truth mask, $D_{\text{downscaled}}$ is the downscaled thick vessel approximation and $\neg$ denotes logical negation. This operation extracts only vessel portions not classified as thick vessels.

Step 2: Thin vessel refinement. The morphological opening leaves small pixel clusters that arise from vessel tortuosity and regions where vessel thickness approximates the threshold value. These clusters do not represent genuine thin vessels and degrade mask quality.

We identify these artifacts using connected component analysis, which traces edges of individual thin vessel structures and labels all disconnected components. Each component is evaluated by attempting to fit it within a square structuring element of side length equal to the thickness threshold plus 2 pixels (the offset determined through visual inspection of multiple samples). Components that do not span at least this width are removed from the thin vessel mask which is essentially the same erosion operation 2.1.1 but done on each disconnected component individually and with a bigger structuring element only committing changes that completely erase the component.

This filtering improves visual quality by removing spurious disconnected fragments, though it also eliminates some genuine branching microvessels and brief thin sections of vessels, which are then reclassified as thick vessels. We accept this trade-off to prioritize connectivity and reduce noise in the thin vessel mask.

Step 3: Thick vessel refinement. To improve thick vessel mask connectivity, we apply the same

filtering process to the thick vessel mask. We obtain the thick vessel mask through:

$$M_{\text{thick}} = M_{\text{ground truth}} \wedge \neg M_{\text{thin}} \tag{4}$$

We again perform connected component analysis and remove components that do not fit within a square of side length equal to the thickness threshold plus 2 pixels. This produces the final thick vessel mask M_{thick}. The final thin vessel mask M_{thin} is defined as all vessel pixels not classified as thick vessels.

The resulting masks provide close approximations of thin and thick vessel tree components. Some pixel-level accuracy is sacrificed to improve connectivity and visual coherence which is a reasonable trade-off given that our experiments do not use image patching techniques. Since fundus images retain their global context as input to segmentation models, vessels should appear as continuous segments rather than fragmented patches. The algorithm therefore prioritizes preserving the topological structure of the vessel tree over perfect boundary delineation at ambiguous thickness values.

2.1.2 Algorithm for Model Prediction Masks

Segmentation models can introduce errors around thin vessel edges, causing predicted vessels to appear thicker than in the ground truth. Applying the ground truth separation algorithm directly to predictions would misclassify such thickened thin vessels as thick vessels. We therefore develop a second algorithm that leverages the ground truth separation to classify predicted pixels while accommodating boundary uncertainty. The algorithm was inspired by Wiedemann et. al. [52] approach on using buffer zones for quantifying road network segmentation predictions.

Step 1: Buffering ground truth pixels. We create a buffer zone around ground truth thin vessels to account for prediction boundary errors. We dilate the thin ground truth vessel mask M_{thin} by the thickness threshold r:

$$M_{\text{dilated}} = M_{\text{thin}} \oplus B_r \tag{5}$$

This dilation effectively expands thin vessel width up to five times, creating a generous search region for thin vessel predictions.

However, this dilated mask may encroach into nearby thick vessels, potentially misclassifying thick vessel predictions as thin. To prevent this, we create a protection mask by inverting the thick vessel mask M_{thick} and eroding it by one fourth the thickness threshold r:

$$M_{\text{protected}} = (\neg M_{\text{thick}}) \ominus B_{r/4} \tag{6}$$

This eroded inverted mask identifies regions where thin vessel buffer zones should not extend (areas in close proximity to thick vessels).

We combine these components with the ground truth thin vessel mask to produce the final buffer zone:

$$M_{\text{thin buffer}} = (M_{\text{dilated}} \wedge M_{\text{protected}}) \vee M_{\text{thin}} \tag{7}$$

The logical disjunction with M_{thin} ensures that all ground truth thin vessel pixels remain in the buffer zone even if they fall within the protected region, preventing genuine thin vessels from being excluded.

Step 2: Prediction classification. We overlay the buffer mask with the model prediction mask $M_{\text{prediction}}$. Predicted vessel pixels are classified as thin if they fall within the buffer zone:

$$M_{\text{thin prediction}} = M_{\text{prediction}} \wedge M_{\text{thin buffer}} \tag{8}$$

All other predicted vessel pixels are classified as thick:

$$M_{\text{thick prediction}} = M_{\text{prediction}} \wedge \neg M_{\text{thin prediction}} \tag{9}$$

This algorithm does not classify false positive predictions outside the buffer zone - such artifacts are automatically classified as thick vessels. While this limitation could be addressed by additionally applying the ground truth separation algorithm to the predicted mask, our study focuses primarily on structural inaccuracies in the vicinity of ground truth thin vessels rather than on isolated prediction artifacts. The buffer zone approach ensures that predictions near known thin vessels are evaluated appropriately while accepting that spurious predictions far from any ground truth vessels receive less refined classification.

2.1.3 Counting TP, FP and FN Pixels

The distinction between thin and thick vessels in this study does not constitute multi-class classification in the traditional sense, where each class is evaluated independently against all others. We do not penalize predictions that confuse thin and thick vessel pixels, as our focus is on binary vessel segmentation rather than vessel caliber classification. Our segmentation models receive no information about vessel thickness and are trained only to distinguish vessels from background.

This evaluation scheme treats thin and thick vessels asymmetrically but consistently. For thin vessel pixels, we classify predictions as true positive (TP) when correctly identified as any vessel class (thin or thick), false positive (FP) when background pixels are incorrectly labeled as vessel and false negative (FN) when thin vessel pixels are incorrectly labeled as background. The same scheme applies to thick vessel evaluation: thick vessel pixels are TP when predicted as any vessel, FP when background is mislabeled as vessel and FN when thick vessel pixels are mislabeled as background. A general formula is provided bellow where class can be interpreted as thickness class for thin or thick vessels. True negative (TN) pixels carry no value for our calculations and interpretations so we do not consider them.

$$TP_{\text{class}} = |M_{\text{class prediction}} \wedge M_{\text{ground truth}}| \tag{10}$$

$$FP_{\text{class}} = |M_{\text{class prediction}} \wedge \neg M_{\text{ground truth}}| \tag{11}$$

$$FN_{\text{class}} = |\neg M_{\text{class prediction}} \wedge M_{\text{class ground truth}}| \tag{12}$$

The overall TP, FP and FN metrics are considered as a sum of each thickness class for the respective metric. This approach allows us to measure vessel detection accuracy for each thickness class independently while acknowledging that the underlying segmentation task is binary. Since thin and thick vessel pixel classes are imbalanced with thick vessels occupying substantially more pixels, separate evaluation prevents thick vessel performance from overshadowing changes in thin vessel accuracy. This separation enables us to examine how downscaling differentially affects each vessel class.

2.2 Quantification of Structural Information Changes from Mask Downscaling

This experiment addresses two objectives: identifying downscaling methods that best preserve thin vessel ground truth mask structural information at different target resolutions and comparing thin vessel preservation relative to thick vessels across the downscaling process. Understanding these effects is useful because segmentation models that produce lower resolution outputs sacrifice accuracy in downstream tasks such as vessel tortuosity quantification, where pixel-level precision directly affects measurement reliability [43]. We hypothesize that thin vessels experience substantially greater information loss than thick vessels at lower resolutions due to their limited pixel footprint and greater susceptibility to discretization effects.

2.2.1 Experimental Protocol

We quantify structural information loss using precision, recall and F1-score. Precision tracks false positive pixels, which indicate thickened vessel walls relative to the original ground truth:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

Recall measures false negative pixels, revealing thinned vessel walls and potential loss of pixel-level connectivity:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

The F1-score balances these metrics, with higher values indicating minimal structural information distortion:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{15}$$

When we talk about structural information loss in this study we interpret it as error in F1 score:

$$Error = 1 - F1 \tag{16}$$

Measurements are computed separately for thin and thick vessels, as the larger pixel count in thick vessels would otherwise dominate the results and obscure changes in thin vessels.

We obtain thin and thick vessel segmentation masks from the test set (detailed in section 3.2) using the ground truth separation algorithm described in section 2.1.1. Then apply various resizing methods to downscale masks from the original 2048×2048 resolution to three target resolutions:

1024×1024, 512×512 and 256×256. All downscaling methods are performed independently for thin and thick vessel masks to preserve information about vessel classification at each resolution.

2.2.2 Interpolation Methods

We test five interpolation methods found in literature representing different approaches to handling spatial resampling:

Nearest neighbor interpolation assigns each output pixel the value of the nearest input pixel without averaging or smoothing. For a target pixel at position (x', y') in the downscaled image, the method finds the closest source pixel at position (x, y) in the original image:

$$I'(x', y') = I(\text{round}(x' \cdot s_x), \text{round}(y' \cdot s_y)) \tag{17}$$

where s_x and s_y are the scaling factors along each axis. This method preserves sharp edges and produces binary output for binary input, but introduces aliasing artifacts and blocky appearance, particularly along diagonal vessel segments.

Max pooling divides the input image into non-overlapping rectangular regions corresponding to each output pixel and selects the maximum value within each region:

$$I'(x', y') = \max_{(i,j) \in R(x', y')} I(i, j) \tag{18}$$

where $R(x',y')$ defines the pooling region for output pixel (x', y'). For binary vessel masks, this ensures that any vessel pixel within a region causes the output pixel to be classified as vessel, guaranteeing perfect connectivity preservation but artificially thickening vessels.

Bilinear interpolation computes each output pixel as a weighted average of the four nearest input pixels, with weights determined by distance. For an output pixel at fractional coordinates (x', y') between input pixels, the interpolated value is:

$$I'(x', y') = (1 - a)(1 - b)I(x_0, y_0) + a(1 - b)I(x_1, y_0) + (1 - a)bI(x_0, y_1) + abI(x_1, y_1) \tag{19}$$

where (x_0, y_0) and (x_1, y_1) are the integer coordinates of surrounding pixels and $a = x' - x_0$, $b = y' - y_0$ are the fractional offsets. This produces smooth intensity transitions and reduces aliasing compared to nearest neighbor, but generates non-binary output requiring thresholding.

Bicubic interpolation extends bilinear interpolation by considering a 4 $\times$4 neighborhood of 16 pixels around each output location, fitting a cubic polynomial surface to these values. The interpolated value is computed as:

$$I'(x', y') = \sum_{i=0}^{3} \sum_{j=0}^{3} I(x_i, y_j) \cdot W(x' - x_i) \cdot W(y' - y_j) \tag{20}$$

where $W(t)$ is Catmull-Rom spline cubic kernel function.

Distance transform interpolation operates differently by first computing the distance transform

of the binary vessel mask, which assigns each background pixel a value equal to its distance to the nearest vessel pixel. This distance map is then downscaled using bilinear interpolation and the result is thresholded to reconstruct a binary vessel mask. For a binary mask M, the distance transform is:

$$D(x, y) = \min_{(i,j)\in M} \sqrt{(x-i)^2 + (y-j)^2} \tag{21}$$

2.2.3 Thresholding for Non-Binary Methods

Since bilinear, bicubic and distance transform methods produce non-binary outputs with pixel intensities in the range $[0, 255]$, we threshold their results to obtain binary vessel masks. Pixels at or above the threshold are classified as vessel, while those below are classified as background. For each non-binary method at each resolution, we select threshold values that best preserve thin vessel pixels by brute-forcing all possible thresholds measuring resulting F1-scores for thin vessels. We apply this method to a representative mask with visually abundant thin vessels.

2.2.4 Evaluation Protocol

We assess information loss by rescaling the downscaled masks back to the original 2048×2048 resolution using nearest neighbor interpolation. This upscaling step allows direct pixel-level comparison with the original ground truth. To evaluate thin vessels, we overlay the rescaled thin vessel mask with the complete original ground truth mask and apply the pixel classification scheme described in section 2.1.3. The same procedure applies to thick vessel evaluation.

After obtaining pixel-level true positive, false positive and false negative counts for each sample, we calculate precision, recall and F1-score values. We average these metrics across all test set samples and compute standard deviations to assess consistency across different vessel morphologies and image characteristics. Results are analyzed separately for thin and thick vessels to reveal class-specific degradation patterns across resolutions and interpolation methods.

2.3 Segmentation of Downscaled Fundus Images

To assess the effect of reduced fundus image resolution on retinal vessel segmentation, we train segmentation models using input images downscaled to four resolutions: 1024×1024 (baseline), 512×512, 256×256 and 128×128. Model performance is evaluated separately for thin and thick vessels in order to analyze how different vessel calibers are affected by progressive loss of spatial detail.

2.3.1 Experimental Protocol

The experimental design isolates image resolution as the sole variable influencing segmentation performance. All models share the same baseline architecture (described in section 3.1), training procedure and optimization settings (described in section 3.3), ensuring that observed performance differences arise from changes in input resolution rather than differences in model capacity,

complexity or learning dynamics. To maintain a consistent model input size, each downscaled fundus image is upscaled back to the baseline resolution of 1024×1024 using bilinear interpolation (for both downscaling and upscaling) before being passed to the network. As a result, the effective information content of the input corresponds to the lower resolution, despite the fixed spatial dimensions at the model input layer.

Ground truth vessel masks remain unchanged across all experiments and are kept at the baseline resolution of 1024×1024 (downscaled using a method from previous experiment r2.2 with the best F1-score). No rescaling is applied to the segmentation masks. This design choice ensures consistent spatial alignment between predictions and ground truth while allowing the experiment to focus on the model's ability to recover vessel structures from degraded image information. In this setting, segmentation performance reflects the model's capacity to interpolate and infer fine vascular details when pixel-level information is reduced or lost due to prior downscaling. This means that if model performs similarly across several different resolutions, the downscaled fundus images retained all of the important structural information. The setup is similar to the SuperVessel model [23] in that the output has higher visual fidelity than the input with the difference being that the physical dimensions of input is the same as output in our experiment.

2.3.2 Evaluation Protocol

For each predicted segmentation and corresponding ground truth mask, thin and thick vessel regions are extracted using the procedures described in sections 2.1.1 and 2.1.2. Precision, recall and F1-score are then computed separately for thin and thick vessels. These metrics are compared across models trained with different effective input resolutions to quantify resolution-dependent performance changes for each vessel class.

In addition to vessel class specific metrics, we report the overall vessel segmentation F1-score, defined as:

$$F1_{\text{overall}} = \frac{2(TP_{\text{thin}} + TP_{\text{thick}})}{2(TP_{\text{thin}} + TP_{\text{thick}}) + (FP_{\text{thin}} + FP_{\text{thick}}) + (FN_{\text{thin}} + FN_{\text{thick}})} \tag{22}$$

Using this formulation, the contribution of thin vessels to the total segmentation error is computed as:

$$ErrorContribution_{\text{thin}} = \frac{FP_{\text{thin}} + FN_{\text{thin}}}{FP_{\text{thin}} + FP_{\text{thick}} + FN_{\text{thin}} + FN_{\text{thick}}} \tag{23}$$

An analogous expression is used to compute the error contribution of thick vessels. This decomposition allows the total F1-score error (structural information loss) to be attributed proportionally to thin and thick vessel segmentation failures, providing insight into which vessel class is more likely to affect the overall segmentation accuracy and segmentation quality.

For each resolution, we perform a permutation test under the null hypothesis that downscaled fundus images retain the same thin/thick/overall vessel structural information as the baseline resolution. The test compares the mean F1-scores obtained at the baseline resolution with those obtained after downscaling. Model outputs from the two conditions are repeatedly pooled and randomly reassigned into two groups of equal size and the difference in mean F1-score is computed for each

permutation. This process is repeated for 10000 iterations to approximate the empirical null distribution of the test statistic. The p-value is then estimated as the proportion of permutations for which the observed difference in mean F1-score is at least as large as the difference measured between the original, unpermuted groups.

Different segmentation architectures, including convolutional [27, 56], adversarial and transformer-based models [29], may exhibit varying sensitivity to input resolution due to differences in receptive fields and inductive biases [4, 11]. However, the objective of this experiment is not to compare architectural performance. Instead, the analysis focuses on a single architecture to establish a controlled baseline that characterizes how progressive downscaling of fundus images affects segmentation accuracy for thin and thick retinal vessels independently.

3 Baseline Establishment

This section defines the baseline configuration used throughout the study to ensure that subsequent analyses of downscaling effects are grounded in a stable and well-characterized reference setup. The baseline establishes consistent choices for model architecture, dataset selection, image and mask resolution and training procedure, allowing performance changes observed in later experiments to be attributed specifically to resolution-related factors rather than confounding design variations. In particular, the baseline is designed to preserve thin vessel structures as effectively as possible under practical computational constraints, providing a reliable point of comparison for evaluating how fundus image and vessel mask downscaling influence thin retinal vessel segmentation accuracy.

3.1 High-Resolution Segmentation Model

This study requires a segmentation model capable of processing high-resolution fundus images while preserving fine spatial details that are critical for thin vessel delineation. For this purpose, we adopt a transformer-based architecture specifically designed for dense prediction at high spatial resolution. The High-Resolution Transformer (HRFormer) [29] addresses a key limitation of standard Vision Transformers [13], which typically rely on aggressive patch-based downsampling that can suppress thin vascular structures. By maintaining high-resolution feature representations throughout the network, HRFormer is well suited for modeling narrow vessels and accurately capturing vessel boundaries.

Transformers were originally introduced in the context of natural language processing as an alternative to recurrent and convolutional sequence models [50]. Their core component, the self-attention mechanism, enables each element in a sequence to attend to all other elements and weigh their relative importance. Given an input sequence represented by feature vectors, self-attention computes three learned projections: queries (Q), keys (K) and values (V). The attention output is obtained as

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V, \tag{24}$$

where d denotes the dimensionality of the key vectors. This formulation allows the model to capture long-range dependencies and contextual relationships without relying on fixed spatial or temporal neighborhoods, which proved highly effective for language modeling tasks.

The success of transformers in language processing motivated their adaptation to vision tasks. In Vision Transformers, an image is typically decomposed into a sequence of non-overlapping patches, which are linearly projected into embedding vectors and treated analogously to word tokens. Self-attention is then applied to model global relationships between patches. While this formulation enables effective global context modeling, it also introduces limitations for dense prediction tasks. The use of relatively large patches (e.g., 16×16 pixels) and repeated spatial downsampling can lead to the loss of fine-grained information, which is particularly detrimental for thin retinal vessel segmentation where vessels may span only a few pixels in width.

HRFormer modifies this paradigm by preserving high-resolution representations throughout the network. Instead of collapsing spatial detail early in the pipeline, the architecture maintains parallel feature streams at multiple resolutions which is a detail that was inspired by HRNet model [51] and replicated in several other retinal vessel segmentation works [27, 34]. The first stage operates at the highest resolution and extracts local features using a convolutional stem, which provides an inductive bias well suited for capturing edges and vessel boundaries. In subsequent stages, additional lower-resolution streams are introduced at $4\times$, $8\times$, $16\times$and $32\times$downscaling. Each stream is processed independently using transformer blocks, while convolutional multi-scale fusion modules enable repeated information exchange between resolutions. This design allows the network to combine precise local detail from high-resolution streams with broader contextual information from lower-resolution streams.

A major challenge in applying self-attention to high-resolution images is its quadratic computational complexity with respect to the number of spatial tokens. HRFormer addresses this issue by employing local-window self-attention. Feature maps are partitioned into non-overlapping $K \times K$ windows and multi-head self-attention is computed independently within each window. This approach reduces computational complexity from quadratic to linear in spatial size, making it feasible to process high-resolution fundus images. However, local-window attention restricts direct interaction between neighboring windows, which can fragment elongated structures such as vessels.

To mitigate this limitation, HRFormer integrates a 3×3 depth-wise convolution within the feed-forward network (FFN) of each transformer block. The FFN follows an MLP $\rightarrow$Depth-Wise Convolution $\rightarrow$MLP structure, enabling information exchange across window boundaries and expanding the effective receptive field. This design allows vessel structures that span multiple windows to remain connected in the learned representation, which is needed for preserving vessel connectivity and topology.

The baseline segmentation model in this study is trained using the HRFormer-S (hrt_small) configuration (around 9.8M trainable parameters), as provided in the authors' official implementation `https://github.com/HRNet/HRFormer/tree/main/seg`. Preliminary experiments indicate that this smaller variant trains faster while achieving performance comparable to larger HRFormer configurations. Since the objective of this work is not architectural comparison but a controlled analysis of resolution effects, HRFormer-S provides a suitable balance between computational efficiency and segmentation accuracy. Furthermore, the baseline model achieves competitive performance on the FIVES dataset test set, making it an appropriate reference for studying the impact of fundus image and vessel mask downscaling on thin retinal vessel segmentation.

3.2 Dataset and Resolution

Higher-resolution fundus images theoretically provide greater spatial detail, which can improve the representation of thin retinal vessels. In practice, the effective information content depends not only on image resolution but also on sensor characteristics, optical quality and acquisition conditions [8]. Very high spatial resolution does not necessarily guarantee improved vessel visibility if image noise or contrast limitations dominate. Nevertheless, preserving higher resolution remains advanta-

geous for modeling fine vascular structures, particularly capillaries that occupy only a few pixels in width.

In this study, the highest resolution at which the segmentation model can be trained reliably is 1024×1024. Training at higher resolutions requires smaller batch sizes due to GPU memory constraints, which leads to unstable optimization and increased overfitting. Empirically, these effects outweigh the potential benefits of higher spatial detail, making 1024×1024 a practical upper bound for model training in this experimental setup.

Several publicly available retinal vessel segmentation datasets are widely used in the literature, including FIVES, DRIVE, STARE, CHASE_DB1 and HRF. Among these, FIVES and HRF provide image resolutions exceeding 1024×1024. FIVES contains a substantially larger number of samples than other datasets, enabling model training on full images without resorting to patch-based strategies that may disrupt global vessel connectivity. Although HRF offers ultra-high-resolution images, its limited number of samples makes it more suitable for cross-dataset evaluation rather than primary model training.

Based on these considerations, FIVES is selected as the primary dataset for fundus image and vessel mask downscaling experiments. To assess whether relying solely on FIVES compromises model generalization, an additional exploratory data analysis experiment is conducted in which FIVES, DRIVE, STARE and CHASE_DB1 are merged into a single training set and the resulting model is evaluated on HRF. This combined-dataset training does not yield improved segmentation performance compared to a model trained exclusively on FIVES. This observation is consistent with prior work showing that models trained on FIVES generalize well to unseen retinal datasets [16]. For all baseline experiments, fundus images are downscaled to 1024×1024 using bilinear interpolation. Vessel segmentation masks are downscaled using bilinear interpolation followed by intensity thresholding at a value of 64 to recover binary masks. No additional image augmentation is applied, allowing the analysis to focus specifically on the effects of resolution changes.

3.3 Experimental Setup

All vessel segmentation models are trained using a composite loss function that combines focal loss and weighted cross-entropy loss, with a weighting factor of 0.4 applied to the cross-entropy loss component. This formulation is selected to address class imbalance between vessel and background pixels while maintaining stable optimization behavior. Model optimization is performed using stochastic gradient descent (SGD) with a momentum of 0.9, weight decay of 1×10^{-4} and an initial learning rate of 0.02. Training proceeds for a maximum of 250 epochs and the model checkpoint corresponding to the minimum validation loss is selected for evaluation.

The dataset split follows the original partitioning provided by the FIVES dataset authors, consisting of 600 fundus images for training and 200 images for testing. To enable validation-based model selection, the training set is further subdivided into training and validation subsets using a 5:1 ratio. This split is performed randomly while maintaining class balance across diseases, resulting in 25 validation images per disease and a total of 100 validation images.

All segmentation models and experimental pipelines are implemented in Python using the Py-

Torch framework (version 1.7.0). Supporting scripts for data preprocessing, training, evaluation and metric computation are also developed in Python. All experiments are conducted on a workstation equipped with an NVIDIA RTX 5090 GPU with 32 GB of VRAM, an AMD Ryzen 9950X3D CPU (16 cores, 5.7 GHz) and 64 GB of DDR5 system memory.

3.4 Thickness of Thin Vessels

The separation of retinal vessels into thin and thick classes requires the definition of a vessel thickness threshold, as publicly available retinal vessel datasets provide only binary vessel annotations without explicit caliber information. In this study, vessel thickness is defined in pixel units and evaluated at the native resolution of 2048×2048, which corresponds to the highest spatial fidelity available in the dataset. A single global threshold is used for all images to ensure consistency across samples and experiments.

Several candidate threshold values are examined through qualitative inspection of the resulting thin and thick vessel masks. Lower thresholds tend to classify only the finest capillary segments as thin, while reassigning substantial portions of secondary branches to the thick vessel class. Conversely, higher thresholds increasingly include medium-caliber vessels in the thin class, reducing the distinction between vessel classes and weakening the interpretability of class-specific performance metrics. The evaluation focuses on preserving visually continuous capillary networks while avoiding misclassification of the main arterial and venous trunks.

Based on this analysis, a thickness threshold of 8 pixels at 2048×2048 resolution is selected. This value consistently classifies branching capillary vessels and narrow peripheral structures as thin, while assigning the majority of primary arteries and veins to the thick vessel class. The threshold also yields stable behavior across different fundus images and disease categories, reducing sensitivity to local variations in vessel width.

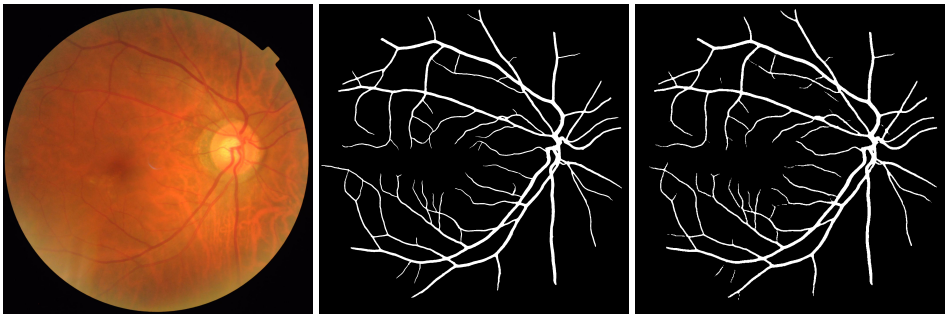
It is important to note that this threshold does not represent a physiological definition of vessel caliber but rather a practical and resolution-dependent criterion for algorithmic separation. Vessel thickness varies continuously and transitional regions near the threshold may be assigned differently depending on local structural information and imaging conditions. Nevertheless, the chosen value provides a reproducible and interpretable division that supports the study's objective of analyzing how downscaling differentially affects thin and thick retinal vessel segmentation.

4 Results and Discussion

This section reports the experimental results in direct correspondence with the tasks defined in this study and provides an interpretation of the observed outcomes.

4.1 Baseline model performance

The baseline segmentation model is trained on the FIVES dataset using fundus images downscaled to 1024×1024 resolution with bilinear interpolation. Corresponding retinal vessel segmentation masks are resized to the same resolution using bilinear interpolation, followed by pixel intensity thresholding at a value of 64 to recover binary masks. Under this configuration, the model achieves state-of-the-art results with an overall average accuracy of 0.9882, precision of 0.9147, recall of 0.8627 and an F1-score of 0.8842 on the test set.



(a) *Fundus image* ***(b)*** *Ground truth* ***(c)*** *Baseline prediction*

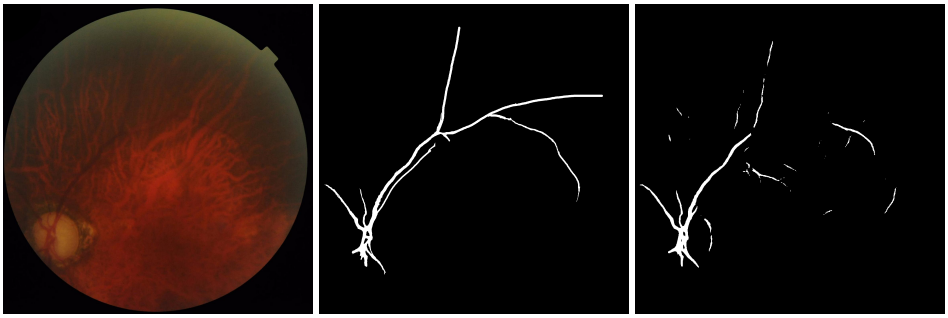
Figure 1 *An example of baseline retinal vessel segmentation model output; both thin and thick vessels have a good contrast with the background resulting in a well segmented fundus images*

The high overall accuracy relative to the F1-score reflects the pronounced class imbalance between background and vessel pixels in retinal images. Since background pixels constitute the majority of the image, accuracy alone provides a limited description of segmentation quality and tends to overestimate model performance. Precision, recall and F1-score therefore offer more informative measures for evaluating vessel segmentation effectiveness.

The precision value of 0.9084 indicates that most pixels predicted as vessels correspond to true vessel regions, suggesting that the model introduces relatively few false positive vessel detections as shown in Figure 1. In contrast, the lower recall value of 0.8358 reveals that a notable proportion of ground truth vessel pixels are not detected by the model demonstrated by Figure 2. This imbalance between precision and recall indicates that the model more frequently fails to identify existing vessels than it incorrectly labels background pixels as vessels. Such behavior is consistent with conservative prediction tendencies commonly observed in vessel segmentation models trained under strong class imbalance [33].

These baseline results establish a reference level of segmentation performance at high resolution, against which the effects of fundus image and vessel mask downscaling are evaluated in subsequent experiments. In particular, the observed recall limitation motivates further analysis of

how resolution reduction disproportionately affects the detectability of thin retinal vessels, which are more susceptible to being missed due to their limited pixel footprint.



(a) *Fundus image* ***(b)*** *Ground truth* ***(c)*** *Baseline prediction*

Figure 2 *An example of baseline retinal vessel segmentation model output; glaucomatous fundus image obscures vessels where even the thick vessels start to blend in with the background making expert annotation as well as model prediction difficult*

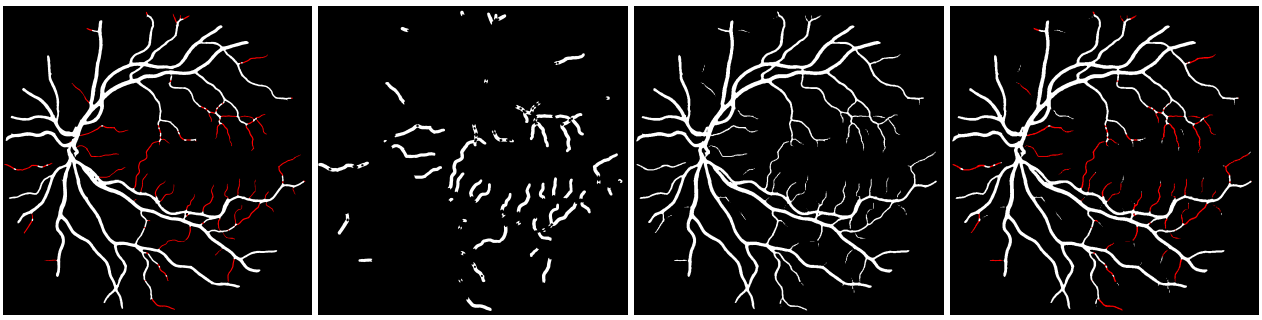
4.2 Algorithm for Thin Vessel Separation

The proposed morphological separation algorithm is applied to all 200 ground truth segmentation masks in the FIVES test set. Overall, the algorithm successfully separates thin and thick vessel components, producing visually coherent vessel class masks across the majority of samples. The quality of separation varies depending on vessel tree complexity, local vessel topology and proximity of vessel thickness to the predefined threshold.

Visual inspection indicates that the algorithm performs consistently in regions where vessels are well separated and exhibit clear differences in caliber. Thin capillary networks in peripheral regions are generally identified correctly, while major arteries and veins are preserved as continuous thick vessel structures. In these cases, the resulting thin and thick vessel masks maintain both connectivity and anatomical plausibility. More challenging scenarios arise near bifurcation points, where thick vessels branch into multiple thinner segments and in transitional regions where vessel diameter gradually changes rather than exhibiting a sharp boundary (Figure 7). In such areas, parts of the vessel may be inconsistently classified depending on local morphology and orientation.

Using a thickness threshold of 8 pixels at the native 2048×2048 resolution yields an approximate 85:15 ratio between thick and thin vessel pixels across the test set (Figure 6). This distribution aligns with visual expectations, as capillary networks occupy a relatively small fraction of total vessel area despite their dense spatial coverage. Figure 3 illustrates an example of a ground truth mask after application of the separation algorithm, alongside the corresponding prediction classification using the dilated thin vessel buffer. The example demonstrates how the algorithm assigns predicted vessel pixels to the thin vessel class when they fall within the buffered region surrounding ground truth thin vessels. Having the buffer radius equal to the vessel threshold thickness in most cases effectively captured thin vessel segmentations in prediction masks that deviate from the ground truth though not in all cases (Figure 8).

Several systematic limitations of the algorithm are observed. First, the method assumes that



(a) *Separated GT* ***(b)*** *Buffer thin vessels* ***(c)*** *Prediction* ***(d)*** *Separated prediction*

Figure 3 *An example of thin vessel (red pixel) separation algorithm output both for ground truth and for model output predictions with thin vessels inferred ground truth*

vessel boundaries in the binary masks are well defined. This assumption may not hold in cases of low-contrast vessels, annotation noise or partial labeling errors, leading to inaccurate thickness estimation. Second, the use of circular structuring elements implicitly assumes locally circular vessel cross-sections. This approximation becomes less accurate for highly tortuous vessels or for vessels viewed at oblique angles, where the apparent cross-section deviates from circularity [18]. Third, vessel segments with diameters close to the selected thickness threshold are inherently ambiguous. Minor variations in local morphology, pixel discretization or annotation quality can cause these regions to oscillate between thin and thick classifications. As a result, the algorithm may locally overestimate or underestimate vessel thickness in transitional regions. While these effects introduce pixel-level inaccuracies, they are spatially limited and do not significantly disrupt overall vessel connectivity.

Despite these limitations, the algorithm provides a consistent and reproducible approximation of thin and thick vessel structures across the dataset. Given that the primary objective of this work is comparative analysis across resolutions rather than absolute vessel caliber estimation, the observed inaccuracies are acceptable. The separation results are sufficiently stable to support class-specific evaluation of structural information degradation and segmentation performance in subsequent experiments.

4.3 Mask Downscaling Structural Information Loss

Tables Table 2 and Table 3 present the quantitative assessment of structural information preservation across different interpolation methods and resolutions. As anticipated, thin vessels experience substantially greater information loss compared to thick vessels across all downscaling methods and resolutions. For instance, at 256×256 resolution, the best-performing methods (bilinear, bicubic and distance transform interpolation) achieve an F1-score of 0.7014 for thin vessels compared to 0.8838 for thick vessels. The difference in structural information loss was 2.6 times greater for thin vessels across all resolutions when compared with thick ones. This disparity reflects the fundamental challenge that thin vessels, often only a few pixels wide at the original resolution, are disproportionately affected by discretization errors inherent in downscaling.

The data demonstrates consistency across vessel classes. When one method outperforms another at a given resolution for either precision, recall or F1-score, this superiority holds across all

Table 2 *Segmentation mask interpolation information loss across different resolutions for thin vessels on FIVES test set*

Interpolation Method	Mask Resolution at Threshold	Precision	Recall	F1
Nearest Neighbor	1024x1024	**0.9102 ± 0.0077**	0.9049 ± 0.0063	0.9075 ± 0.0060
	512x512	**0.8369 ± 0.0093**	0.8256 ± 0.0119	0.8312 ± 0.0088
	256x256	**0.6926 ± 0.0162**	0.6768 ± 0.0227	0.6845 ± 0.0176
Max Pooling	1024x1024	0.8305 ± 0.0079	**1.0000 ± 0.0000**	0.9074 ± 0.0047
	512x512	0.6268 ± 0.0133	**1.0000 ± 0.0000**	0.7705 ± 0.0101
	256x256	0.4310 ± 0.0147	**1.0000 ± 0.0000**	0.6022 ± 0.0144
Bilinear	1024x1024 & 96	0.8926 ± 0.0081	0.9676 ± 0.0029	**0.9286 ± 0.0045**
	512x512 & 96	0.8175 ± 0.0095	0.8808 ± 0.0095	**0.8479 ± 0.0076**
	256x256 & 32	0.6808 ± 0.0171	0.7236 ± 0.0176	**0.7014 ± 0.0154**
Bicubic	1024x1024 & 98	0.8924 ± 0.0082	0.9677 ± 0.0029	**0.9285 ± 0.0045**
	512x512 & 100	0.8175 ± 0.0095	0.8806 ± 0.0096	**0.8478 ± 0.0076**
	256x256 & 34	0.6808 ± 0.0171	0.7236 ± 0.0176	**0.7014 ± 0.0154**
Distance Transform	1024x1024 & 96	0.8926 ± 0.0081	0.9676 ± 0.0029	**0.9286 ± 0.0045**
	512x512 & 96	0.8175 ± 0.0095	0.8808 ± 0.0095	**0.8479 ± 0.0076**
	256x256 & 32	0.6808 ± 0.0171	0.7236 ± 0.0176	**0.7014 ± 0.0154**

Table 3 *Segmentation mask interpolation information loss across different resolutions for thick vessels on FIVES test set*

Interpolation Method	Mask Resolution at Threshold	Precision	Recall	F1
Nearest Neighbor	1024x1024	**0.9677 ± 0.0028**	0.9664 ± 0.0031	0.9671 ± 0.0029
	512x512	**0.9397 ± 0.0056**	0.9373 ± 0.0055	0.9385 ± 0.0055
	256x256	**0.8829 ± 0.0100**	0.8788 ± 0.0111	0.8809 ± 0.0104
Max Pooling	1024x1024	0.9295 ± 0.0055	**1.0000 ± 0.0000**	0.9635 ± 0.0030
	512x512	0.8152 ± 0.0124	**1.0000 ± 0.0000**	0.8981 ± 0.0076
	256x256	0.6558 ± 0.0181	**1.0000 ± 0.0000**	0.7919 ± 0.0135
Bilinear	1024x1024 & 96	0.9586 ± 0.0036	0.9873 ± 0.0014	**0.9727 ± 0.0024**
	512x512 & 96	0.9284 ± 0.0063	0.9558 ± 0.0040	**0.9419 ± 0.0051**
	256x256 & 32	0.8725 ± 0.0105	0.8955 ± 0.0101	**0.8838 ± 0.0100**
Bicubic	1024x1024 & 98	0.9586 ± 0.0036	0.9872 ± 0.0014	**0.9727 ± 0.0024**
	512x512 & 100	0.9284 ± 0.0063	0.9557 ± 0.0040	**0.9419 ± 0.0051**
	256x256 & 34	0.8725 ± 0.0105	0.8955 ± 0.0101	**0.8838 ± 0.0100**
Distance Transform	1024x1024 & 96	0.9586 ± 0.0036	0.9873 ± 0.0014	**0.9727 ± 0.0024**
	512x512 & 96	0.9284 ± 0.0063	0.9558 ± 0.0040	**0.9419 ± 0.0051**
	256x256 & 32	0.8725 ± 0.0105	0.8955 ± 0.0101	**0.8838 ± 0.0100**

tested resolutions for both thin and thick vessels. The small standard deviations (typically under 0.02) indicate that this behavior is consistent across all test samples rather than driven by outliers.

Nearest neighbor interpolation, despite its widespread use in literature and computational simplicity, is outperformed by linear interpolation methods on F1-score for both vessel classes. However, it achieves marginally higher precision than other methods, suggesting it introduces fewer false positives through artificial vessel thickening. This precision advantage diminishes as resolution decreases, making it less relevant for highly downscaled masks.

Max pooling presents an interesting trade-off. It achieves perfect recall (1.0000) across all resolutions and vessel classes by design, as the operation selects the maximum value within each pooling window, ensuring no vessel pixels are lost where vessels exist. This guarantees preservation of vessel connectivity, which proves beneficial for thin vessels at 1024×1024 resolution where it performs comparably to nearest neighbor interpolation (F1-score of 0.9074 vs. 0.9075). However, max pooling completely disregards vessel wall thickness, resulting in progressively degraded precision as resolution decreases. At 256×256, precision drops to 0.4310 for thin vessels and 0.6558 for thick vessels, indicating severe artificial vessel thickening. Despite these limitations, max pooling's connectivity preservation makes it worth considering for applications where maintaining vessel topology is paramount.

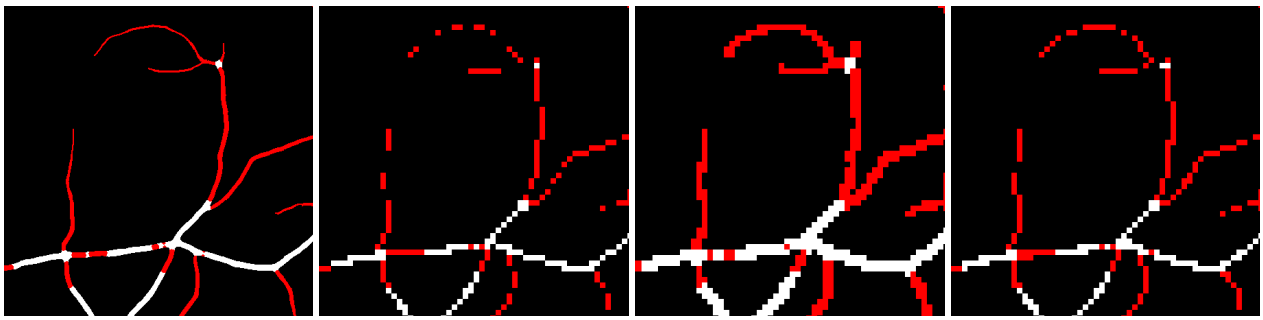
Linear interpolation methods, namely bilinear, bicubic and distance transform, achieve nearly identical results and represent the closest approximation to ground truth with minimal information loss. At 1024×1024, all three methods produce statistically indistinguishable results for thin vessels (F1-score of approximately 0.9286), substantially outperforming both nearest neighbor (0.9075) and max pooling (0.9074). This superiority persists across all resolutions. Notably, none of these methods perform optimally at the conventional 127 (50%) pixel intensity threshold. Instead, optimal thresholds vary by resolution, with values around 96-100 for 1024×1024 and 512×512 and substantially lower values (32-34) for 256×256. This suggests that linear interpolation methods require resolution-specific threshold calibration to balance precision and recall effectively.

The progression of information loss across resolutions reveals the severity of downscaling impact. For thin vessels using the best-performing methods (linear interpolations), F1-scores decline from 0.9286 at 1024×1024 to 0.8479 at 512×512 (2.13$\times$more structural information lost) and further to 0.7014 at 256×256 (1.96$\times$additional loss). The same structural information degradation rate is observed for thick vessels. This suggests that structural information degradation rate increases linearly with downscaling factor though for thick vessels the substantially smaller absolute F1-score decline reflects thick vessels' greater robustness to resolution reduction due to their larger pixel footprint.

A good recall score is slightly more important than precision [33]. Lower precision indicates artificially thickened vessels, a distortion that can potentially be corrected through morphological thinning operations. In contrast, lower recall signifies that vessel segments have been completely erased, which is a loss that cannot be recovered through post-processing. This consideration is particularly relevant for thin vessels, where recall drops from 0.9676 at 1024×1024 to 0.7236 at 256×256

using linear methods, indicating that approximately 25% of thin vessel pixels are lost at the lowest resolution.

An example is given in Figure 4 on how thin vessel connectivity may be lost when aggressively downscaling ground truth masks which would not be the case if human annotator were to label this fundus image [14]. Meaning that this is no longer a representation of original ground truth and the underlying segmentation task would be different from that which uses the original ground truth labels. Figure 9 and Figure 10 display how choosing an appropriate resolution and downscaling method can still preserve mask connectivity and wall thickness.



(a) *Ground truth* ***(b)*** *Nearest neighbor* ***(c)*** *Maxpool* ***(d)*** *Bilinear*

Figure 4 *Visual thin vessel structure (red pixels) degradation when downscaling to* 256×256 *resolution using different interpolation methods; only maxpooling (c) preserves the vessel connectivity but at the cost of precision; nearest neighbor interpolation (b) appears to make the vessel severely fractured; even linear interpolation methods (d) alter vessel's tortuosity*

Based on these findings, we select bilinear interpolation with its resolution-specific threshold as the standard method for mask downscaling in subsequent experiments, as it provides the optimal balance of information preservation, computational efficiency and implementation simplicity.

4.4 Impact of Fundus Image Resolution on Vessel Segmentation

Table Table 4 presents segmentation performance across different fundus image resolutions while maintaining a constant mask resolution of 1024×1024. This experimental design isolates the impact of fundus image information loss from mask resolution effects, revealing how degraded visual features affect the model's ability to distinguish vessels from background.

Thick vessels achieve substantially higher segmentation accuracy than thin vessels across all resolutions, as expected given their greater visual prominence and pixel footprint. However, the relationship between resolution and performance differs markedly between vessel classes.

Models trained at 1024×1024 and 512×512 resolutions produce nearly identical results. For thin vessels, F1-scores are 0.7060 and 0.6946 respectively, representing a negligible difference of approximately 1.6%, which is not statistically significant ($p = 0.3902$). Thick vessel performance is even more stable, with F1-scores of 0.9116 and 0.9077 ($p = 0.7133$). This suggests that downscaling from 1024×1024 to 512×512 preserves sufficient fundus image information for the model to maintain segmentation accuracy, making 512×512 a viable resolution for computational efficiency without sacrificing performance.

Table 4 *Performance metrics by vessel class across different rescaled fundus image resolutions and* 1024×1024 *resolution mask on FIVES test set*

Vessel Class	Resolution	Precision	Recall	F1	p-value
Overall	1024x1024	**0.9147 ± 0.0523**	**0.8627 ± 0.1155**	**0.8842 ± 0.0984**	baseline
	512x512 → 1024x1024	0.9131 ± 0.0422	0.8585 ± 0.1193	0.8794 ± 0.1026	0.6350
	256x256 → 1024x1024	0.9029 ± 0.0488	0.8367 ± 0.1120	0.8648 ± 0.0939	0.0485
	128x128 → 1024x1024	0.8590 ± 0.0556	0.7458 ± 0.1021	0.7953 ± 0.0854	<0.0001
Thin	1024x1024	**0.8080 ± 0.0740**	**0.6420 ± 0.1558**	**0.7060 ± 0.1308**	baseline
	512x512 → 1024x1024	0.8011 ± 0.0685	0.6300 ± 0.1587	0.6946 ± 0.1359	0.3902
	256x256 → 1024x1024	0.7761 ± 0.0761	0.5677 ± 0.1479	0.6463 ± 0.1256	<0.0001
	128x128 → 1024x1024	0.6276 ± 0.0862	0.2923 ± 0.0991	0.3903 ± 0.1046	<0.0001
Thick	1024x1024	**0.9291 ± 0.0535**	**0.9022 ± 0.1179**	**0.9116 ± 0.1013**	baseline
	512x512 → 1024x1024	0.9281 ± 0.0414	0.8991 ± 0.1217	0.9077 ± 0.1054	0.7133
	256x256 → 1024x1024	0.9191 ± 0.0478	0.8848 ± 0.1156	0.8977 ± 0.0971	0.1710
	128x128 → 1024x1024	0.8778 ± 0.0565	0.8286 ± 0.1116	0.8493 ± 0.0903	<0.0001

Performance diverges substantially at 256×256 resolution. Thin vessel F1-score drops to 0.6463, representing an 7.0% decline from the 512×512, with the difference being statistically significant ($p < 0.0001$). This degradation accelerates dramatically at 128×128, where F1-score falls to 0.3903, corresponding to a 43.8% drop from the baseline and effectively indicating segmentation failure for thin vessels. Thick vessels exhibit more resilient but still notable degradation, declining to F1-scores of 0.8977 at 256×256 and 0.8493 at 128×128.

Precision remains remarkably stable for thick vessels across resolutions down to 256×256, remaining above 0.9191 and only drops noticeably at 128×128 (0.8778). This stability indicates minimal false positive predictions, meaning the model does not generate spurious vessel artifacts or excessively thicken vessel walls even when working with moderately degraded input images. Thin vessel precision also remains relatively stable between 1024×1024 and 256×256 (0.8080-0.7761), before dropping sharply at 128×128 to 0.6276.

Recall emerges as the primary metric affected by resolution reduction. For thin vessels, recall declines progressively from 0.6420 at 1024×1024 to 0.6300 at 512×512, 0.5677 at 256×256 and finally 0.2923 at 128×128. This pattern indicates that as fundus image quality degrades, thin vessels become increasingly difficult to distinguish from background tissue, resulting in vessel segments disappearing from predictions. This loss is particularly detrimental as it damages vessel tree connectivity, which is a critical morphological feature that cannot always be recovered through post-processing.

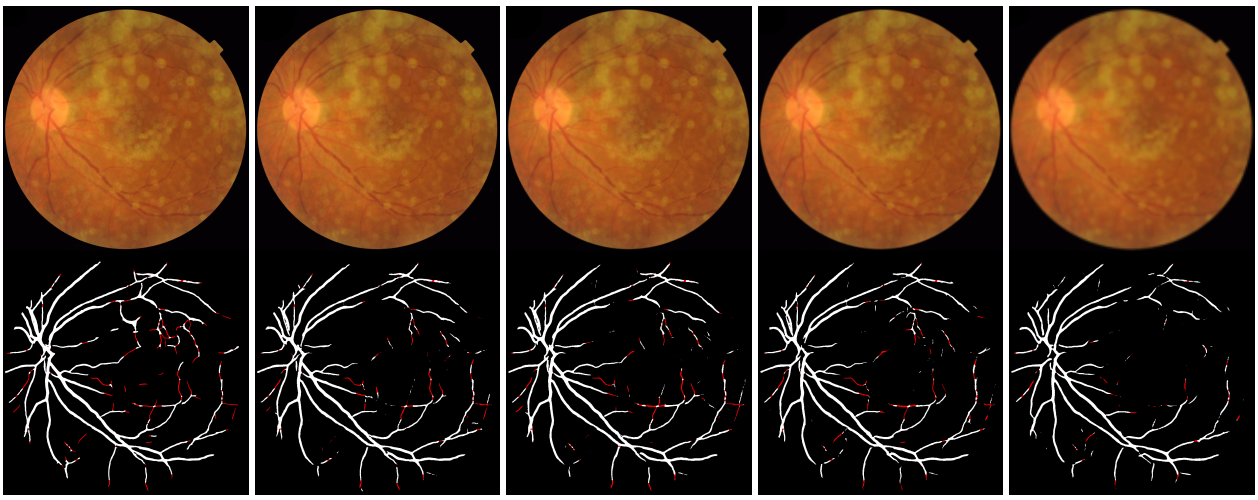
Thick vessels follow a similar but less severe recall degradation pattern, declining from 0.9022 at 1024×1024 to 0.8286 at 128×128. The preservation of relatively high precision alongside declining recall suggests that when the model does predict vessels at lower resolutions, those predictions are generally accurate, but the model fails to detect substantial portions of the vessel tree.

Despite thin vessel pixels constituting only approximately 14.4% of total vessel mask pixels, they contribute disproportionately to overall segmentation error. At 1024×1024 resolution, thin vessels account for approximately 38.49% of total F1-score error (loss of structural information), with similar contributions at 512×512 (38.18%) and 256×256 (37.73%). This contribution decreases slightly

at 128×128 (36.71%), not because thin vessel segmentation improves, but because thick vessel errors begin to accumulate more rapidly at this extreme downscaling level. When normalizing these values to account for class imbalance (measuring error rate relative to class prevalence) the disparity becomes stark. Thin vessels are responsible for 78.81% of normalized errors at 1024×1024, 78.59% at 512×512, 78.27% at 256×256 and 77.52% at 128×128.

The standard deviation values reveal substantial performance variability across test samples, particularly for thin vessels. Thin vessel recall exhibits high variability across all resolutions, indicating that thin vessel detection quality varies considerably between images. Some fundus images permit reasonable thin vessel segmentation, while others yield poor results. This variability likely reflects differences in image quality, vessel contrast and background tissue characteristics across the dataset [8].

The high variability in thin vessel metrics propagates to overall performance, where recall standard deviations remain notably higher than precision standard deviations. In contrast, precision shows lower variability across all resolutions and vessel classes, suggesting that when the model does predict vessels, it does so with relatively consistent accuracy across samples. This disparity indicates that the primary challenge lies not in false positive control but in the inconsistent ability to detect vessels, particularly thin ones, under varying image conditions.



(a) *Ground truth* ***(b)*** 1024×1024 ***(c)*** 512×512 ***(d)*** 256×256 ***(e)*** 128×128

Figure 5 *An example of how reducing effective image resolution affects thin vessel (red pixel) segmentation; while there is no major difference between* 1024×1024 *and* 512×512 *resolutions, downscaling fundus images further noticeably reduces amount of thin vessel segments found*

The results paint a challenging picture for thin vessel segmentation at reduced resolutions (Figure 5). Even at the baseline 1024×1024 resolution, a thin vessel F1-score of 0.7060 indicates substantial room for improvement. While the model reliably identifies thick vessels (F1-score of 0.9116), it struggles with the capillary network that contains much of the clinically relevant vascular detail. This limitation becomes pronounced at 256×256 and severe at 128×128, where thin vessel detection performance degrades to a level that limits practical applicability.

5 Work Limitations

Several constraints in our experimental design and methodology warrant consideration when interpreting these findings.

Our baseline model operates at 1024×1024 resolution rather than the native 2048×2048 resolution of FIVES images due to computational limitations. Training at full resolution proved infeasible since only batch size 1 could fit in memory, causing rapid overfitting and training collapse. This necessitates downscaling ground truth masks to 1024×1024 for our baseline, introducing a small amount of information loss into all subsequent comparisons. Our reported performance metrics therefore represent an upper bound slightly below what might be achievable with native resolution training.

The interpretability of fundus image segmentation results across resolutions depends substantially on baseline model accuracy. Our findings characterize how resolution reduction affects a specific model architecture at a particular performance level. As more sophisticated and accurate segmentation models emerge, the relationship between resolution and performance may shift. Higher resolutions might become essential for future architectures that can exploit finer details more effectively or conversely, advanced models might achieve acceptable thin vessel segmentation at lower resolutions through superior feature extraction. Our conclusions establish baseline expectations but may require revision as the field advances.

Substantial computational resources are required for training models on high-resolution images even with architectures optimized for efficiency. The FIVES dataset's 500 training images at 1024×1024 resolution already demand significant GPU memory and computation complexity. This resource barrier may prevent many researchers from exploring resolutions above 512×512, creating a practical ceiling independent of theoretical benefits. The computational cost-benefit trade-off will vary by application and our findings should inform rather than prescribe resolution choices.

Most critically, our conclusions depend heavily on the operational definition of "thin vessel". We define thin vessels using a thickness threshold derived from the algorithmic separation process described in section 2.1.1. This definition is dataset-specific and threshold-dependent where different thickness values would yield different vessel classifications and consequently different accuracy metrics. Few works measure thin vessel segmentation accuracy separately so the academic community currently lacks consensus on standardized thin vessel measurements, making future cross-study comparisons difficult. Future work would benefit from establishing common thin vessel definitions based on clinical relevance, vessel diameter distributions or anatomical classifications. Without such standardization, "thin vessel segmentation accuracy" remains an inherently relative metric that varies across datasets, methodologies and research groups.

6 Conclusions and Future Work

This study establishes fundamental constraints on fundus image and retinal vessel segmentation mask resolution reduction for retinal vessel segmentation, with particular focus on thin vessel preservation.

1. We established a baseline retinal vessel segmentation model based on the HRFormer architecture, which is designed to efficiently process high-resolution fundus images. The model achieved state-of-the-art performance on the high-resolution FIVES dataset. The dataset enabled training and inference on full-resolution images without resorting to patch-based segmentation. This design choice allowed the model to preserve global vascular context, which is useful for maintaining vessel continuity and accurately representing complex vessel topology.

   Precision and recall proved to be informative and complementary evaluation metrics for analyzing structural changes in the retinal vasculature with error in F1 representing loss in structural information. While precision primarily reflected artificial vessel thickening and false positive predictions, recall captured the model's ability to detect and preserve vessel connectivity. Across all experiments, recall was the most sensitive to resolution reduction, indicating that connectivity loss and missed vessel segments are the dominant failure points. These errors represent irreversible information loss, as missing vessel segments are harder to recover through post-processing, in contrast to boundary distortions associated with reduced precision.

2. We proposed an algorithmic method for separating retinal vessel mask pixels into two groups based on a predefined vessel thickness threshold and applied this method to analyze the effects of fundus image and vessel mask downscaling on thin retinal vessel segmentation. This approach enabled class-specific evaluation of segmentation performance without reformulating the task as a multi-class problem.

   For the FIVES dataset, a vessel thickness threshold of 8 pixels at the native resolution yielded a consistent and visually plausible separation between thin and thick vessels. This threshold effectively included capillary networks within the thin vessel class while retaining the majority of major arteries and veins as thick vessels. As a result, the method supported stable, reproducible and interpretable class-specific analysis across all experiments.

3. Our findings demonstrate that linear interpolation methods provide the most effective preservation of vessel structural information during segmentation mask downscaling across all tested resolutions. However, these methods require resolution-specific threshold calibration, as the conventional 50% intensity threshold does not consistently produce optimal results at different scales.

   Thin vessels are affected most by reductions in ground truth mask resolution. Even when using the best-performing interpolation methods, the thin vessel F1-score decreases from 0.9286 at 1024×1024 to 0.7014 at 256×256, indicating substantial structural deviation from the original ground truth. This decline corresponds to approximately 2.6 times greater relative

structural information loss compared to thick vessels over the same resolution range. Overall, reducing the mask resolution by a factor of two results in an approximately twofold reduction in preserved structural information.

4. For applications requiring thin vessel analysis, 512×512 represents the lower practical bound for fundus image resolution. This resolution offers computational savings without significant loss in segmentation accuracy. Models trained at 1024×1024 and 512×512 achieve nearly identical performance, with thin vessel F1-scores of 0.7060 and 0.6946 respectively and no statistically significant difference ($p = 0.3902$). Resolution reduction below 512×512 results in significant performance degradation and should be avoided unless thick vessel segmentation alone suffices for the clinical task. This recommendation aligns with the observation that publicly available vessel segmentation datasets maintain resolutions at or above 512×512, reflecting an implicit community consensus on a minimum viable resolution.

   The high standard deviation observed in thin vessel metrics presents a significant challenge for clinical deployment. Models trained at 1024×1024 and 512×512 resolutions exhibit thin vessel recall standard deviations of approximately 0.16, indicating substantial performance fluctuations across patients and imaging conditions. Such variability suggests that reliable clinical use would require additional quality control mechanisms or preprocessing strategies to identify cases where thin vessel segmentation is likely to fail.

   Our evaluation scheme reveals that segmentation models are more prone to damaging thin vessel structural connectivity or missing thin vessels entirely than to misclassifying detected vessels. This behavior manifests as progressively declining recall with relatively stable precision across resolutions, indicating that resolution reduction primarily leads to vessel disappearance rather than boundary distortion.

   When examining overall model performance, improvements in thick vessel segmentation yield larger absolute gains in F1-score due to class imbalance, as thick vessels constitute approximately 85% of vessel pixels. However, normalized error analysis reveals a different trend. Thin vessels account for approximately 78% of normalized segmentation errors at practical resolutions, demonstrating that thin vessel segmentation quality disproportionately influences model utility for comprehensive retinal vascular analysis despite their limited pixel prevalence.

Several potential research directions stem from our findings.

1. Specialized loss functions that weight thin vessel pixels more heavily or incorporate topological constraints to preserve connectivity could substantially improve thin vessel recall. Techniques such as focal loss, boundary-aware loss or centerline-focused objectives warrant systematic evaluation.

2. Systematically removing thin vessels from ground truth and measuring segmentation impact would isolate annotation quality effects from algorithmic limitations. Tracking false positive predictions in ambiguous regions could reveal whether these ”errors” represent genuine mistakes or valid vessel detections missed by annotators.

3. Deep learning-based super-resolution algorithms may better preserve or reconstruct fine vascular structures compared to traditional interpolation. Systematic comparison would determine whether these techniques genuinely recover vessel details or hallucinate anatomically incorrect structures.

4. Comprehensive replication across DRIVE, STARE, CHASE_DB1 and other datasets would establish generalizability of our conclusions. Different datasets exhibit varying image quality, vessel appearance and annotation philosophies that may interact with resolution effects.

5. Developing methods to separate capillaries from major vessels based on morphological, topological or appearance features would enable clinically meaningful performance evaluation beyond arbitrary thickness thresholds.

6. Investigating preprocessing techniques such as adaptive contrast enhancement or illumination correction could reduce the high performance variability (recall standard deviation exceeding 0.16) observed across test samples and improve model robustness for clinical deployment.

Acknowledgments

The authors acknowledge use of large language models in refining written contents of this study: Claude Sonnet 4.5 [2] was given the following prompt ”Enhance and make academic text for given subject more readable for master thesis titled: ”Impact of Downscaling Fundus Images and Vessel Masks on Thin Retinal Vessel Segmentation”. Avoid using pompous speech that stresses importance, avoid using same length paragraphs, keep terminology consistent, use present tense, keep latex notations.” which is appended by comprehensive text or detailed bullet points written by the author. Each section/subsection was passed to model individually with the same instructions. Large language model’s output was additionally reviewed and corrected (if needed) by the author.

References

[1] M. Abràmoff, C. N. Kay. “Image Processing.” In: *Retina*. Elsevier, 2013, pages 151–176. ISBN: 9781455707379. https://doi.org/10.1016/b978-1-4557-0737-9.00006-0. URL: http://dx.doi.org/10.1016/B978-1-4557-0737-9.00006-0.

[2] Anthropic. URL: https://claude.ai/.

[3] İ. Atli, O. S. Gedik. “Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation.” In: *Engineering Science and Technology, an International Journal* 24.2 (2021), pages 271–283. ISSN: 2215-0986. https://doi.org/10.1016/j.jestch.2020.07.008. URL: http://dx.doi.org/10.1016/j.jestch.2020.07.008.

[4] A. Bakhtiarnia, Q. Zhang, A. Iosifidis. *Efficient High-Resolution Deep Learning: A Survey*. 2022. https://doi.org/10.48550/ARXIV.2207.13050. URL: https://arxiv.org/abs/2207.13050.

[5] P. Bankhead, C. N. Scholfield, J. G. McGeown, T. M. Curtis. “Fast Retinal Vessel Detection and Measurement Using Wavelets and Edge Location Refinement.” In: *PLoS ONE* 7.3 (2012). Edited by T. Serrano-Gotarredona, e32435. ISSN: 1932-6203. https://doi.org/10.1371/journal.pone.0032435. URL: http://dx.doi.org/10.1371/journal.pone.0032435.

[6] J. Cao, J. Chen, Y. Gu, J. Liu. “MFA-UNet: a vessel segmentation method based on multi-scale feature fusion and attention module.” In: *Frontiers in Neuroscience* 17 (2023). ISSN: 1662-453X. https://doi.org/10.3389/fnins.2023.1249331. URL: http://dx.doi.org/10.3389/fnins.2023.1249331.

[7] E. J. Carmona, M. Rincón, J. García-Feijoó, J. M. Martínez-de-la-Casa. “Identification of the optic nerve head with genetic algorithms.” In: *Artificial Intelligence in Medicine* 43.3 (2008), pages 243–259. ISSN: 0933-3657. https://doi.org/10.1016/j.artmed.2008.04.005. URL: http://dx.doi.org/10.1016/j.artmed.2008.04.005.

[8] J. Cervantes, J. Cervantes, F. García-Lamont, A. Yee-Rendon, J. E. Cabrera, L. D. Jalili. “A comprehensive survey on segmentation techniques for retinal vessel segmentation.” In: *Neurocomputing* 556 (2023), page 126626. ISSN: 0925-2312. https://doi.org/10.1016/j.neucom.2023.126626. URL: http://dx.doi.org/10.1016/j.neucom.2023.126626.

[9] R. J. Chalakkal, W. H. Abdulla, S. Sinumol. “Comparative Analysis of University of Auckland Diabetic Retinopathy Database.” In: *Proceedings of the 9th International Conference on Signal Processing Systems*. ICSPS 2017. ACM, 2017, pages 235–239. https://doi.org/10.1145/3163080.3163087. URL: http://dx.doi.org/10.1145/3163080.3163087.

[10] A. Christodoulidis, T. Hurtut, H. B. Tahar, F. Cheriet. “A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images.” In: *Computerized Medical Imaging and Graphics* 52 (2016), pages 28–43. ISSN: 0895-6111. https://doi.org/10.1016/j.compmedimag.2016.06.001. URL: http://dx.doi.org/10.1016/j.compmedimag.2016.06.001.

[11] A. Dede, H. Nunoo-Mensah, E. T. Tchao, A. S. Agbemenu, P. E. Adjei, F. A. Acheampong, J. J. Kponyo. "Deep learning for efficient high-resolution image processing: A systematic review." In: *Intelligent Systems with Applications* 26 (2025), page 200505. ISSN: 2667-3053. https://doi.org/10.1016/j.iswa.2025.200505. URL: http://dx.doi.org/10.1016/j.iswa.2025.200505.

[12] J. Domarkaitė. "Investigation of eye fundus image quality on vascular segmentation using deep neural networks." In: *Vilnius University Library* (2022).

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2020. https://doi.org/10.48550/ARXIV.2010.11929. URL: https://arxiv.org/abs/2010.11929.

[14] *DRIVE*. URL: https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction/data.

[15] *EyeWiki*. URL: https://eyewiki.org.

[16] J. Fadugba, P. Köhler, L. Koch, P. Manescu, P. Berens. *Benchmarking Retinal Blood Vessel Segmentation Models for Cross-Dataset and Cross-Disease Generalization*. 2024. https://doi.org/10.48550/ARXIV.2406.14994. URL: https://arxiv.org/abs/2406.14994.

[17] X. Feng, G. Cai, X. Gou, Z. Yun, W. Wang, W. Yang. "Retinal Mosaicking with Vascular Bifurcations Detected on Vessel Mask by a Convolutional Network." In: *Journal of Healthcare Engineering* 2020 (2020), pages 1–13. ISSN: 2040-2309. https://doi.org/10.1155/2020/7156408. URL: http://dx.doi.org/10.1155/2020/7156408.

[18] M. M. Fraz, A. Basit, S. A. Barman. "Application of Morphological Bit Planes in Retinal Blood Vessel Extraction." In: *Journal of Digital Imaging* 26.2 (2012), pages 274–286. ISSN: 1618-727X. https://doi.org/10.1007/s10278-012-9513-3. URL: http://dx.doi.org/10.1007/s10278-012-9513-3.

[19] M. M. Fraz, P. Remagnino, A. Hoppe, S. Velastin, B. Uyyanonvara, S. A. Barman. "A supervised method for retinal blood vessel segmentation using line strength, multiscale Gabor and morphological features." In: *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. 2011, pages 410–415. https://doi.org/10.1109/ICSIPA.2011.6144129.

[20] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, S. A. Barman. "An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation." In: *IEEE Transactions on Biomedical Engineering* 59.9 (2012), pages 2538–2548. https://doi.org/10.1109/TBME.2012.2205687.

[21] Y. Hatanaka, H. Tachiki, K. Ogohara, C. Muramatsu, S. Okumura, H. Fujita. "Artery and vein diameter ratio measurement based on improvement of arteries and veins segmentation on retinal images." In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2016, pages 1336–1339. https://doi.org/10.1109/EMBC.2016.7590954.

[22] *HRF*. URL: https://www5.cs.fau.de/research/data/fundus-images/.

[23] Y. Hu, Z. Qiu, D. Zeng, L. Jiang, C. Lin, J. Liu. "SuperVessel: Segmenting High-Resolution Vessel from Low-Resolution Retinal Image." In: *Pattern Recognition and Computer Vision*. Springer Nature Switzerland, 2022, pages 178–190. ISBN: 9783031189104. https://doi.org/10.1007/978-3-031-18910-4_15. URL: http://dx.doi.org/10.1007/978-3-031-18910-4_15.

[24] L. Huang, F. Liu. "Width measurement of retinal vessels using cubic spline fitting and extended edge-searching mode." In: *International Journal of Imaging Systems and Technology* 34.2 (2023). ISSN: 1098-1098. https://doi.org/10.1002/ima.22996. URL: http://dx.doi.org/10.1002/ima.22996.

[25] Z. Yan, X. Yang, K.-T. Cheng. "A Three-Stage Deep Learning Model for Accurate Retinal Vessel Segmentation." In: *IEEE Journal of Biomedical and Health Informatics* 23.4 (2019), pages 1427–1436. https://doi.org/10.1109/JBHI.2018.2872813.

[26] L. Yang, H. Wang, Q. Zeng, Y. Liu, G. Bian. "A hybrid deep segmentation network for fundus vessels via deep-learning framework." en. In: *Neurocomputing* 448 (2021), pages 168–178.

[27] Z. Ye, Y. Liu, T. Jing, Z. He, L. Zhou. "A High-Resolution Network with Strip Attention for Retinal Vessel Segmentation." In: *Sensors* 23.21 (2023), page 8899. ISSN: 1424-8220. https://doi.org/10.3390/s23218899. URL: http://dx.doi.org/10.3390/s23218899.

[28] Y. Yuan, L. Wang, L. Zhang. "Improving vessel connectivity in retinal vessel segmentation via adversarial learning." In: *Knowledge-Based Systems* 262 (2023), page 110243. ISSN: 0950-7051. https://doi.org/10.1016/j.knosys.2022.110243. URL: http://dx.doi.org/10.1016/j.knosys.2022.110243.

[29] Y. Yuan, R. Fu, L. Huang, W. Lin, C. Zhang, X. Chen, J. Wang. "HRFormer: High-Resolution Transformer for Dense Prediction." In: *CoRR* abs/2110.09408 (2021). URL: https://arxiv.org/abs/2110.09408.

[30] K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, J. Ye. "FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation." In: *Scientific Data* 9.1 (2022). ISSN: 2052-4463. https://doi.org/10.1038/s41597-022-01564-3. URL: http://dx.doi.org/10.1038/s41597-022-01564-3.

[31] S. Kalaie, A. Gooya. "Vascular tree tracking and bifurcation points detection in retinal images using a hierarchical probabilistic model." In: *Computer Methods and Programs in Biomedicine* 151 (2017), pages 139–149. ISSN: 0169-2607. https://doi.org/10.1016/j.cmpb.2017.08.018. URL: http://dx.doi.org/10.1016/j.cmpb.2017.08.018.

[32] H. Li, J. Zhang, Q. Nie, L. Cheng. "A retinal vessel tracking method based on bayesian theory." In: *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE. 2013, pages 232–235.

[33] G. Lin, H. Bai, J. Zhao, Z. Yun, Y. Chen, S. Pang, Q. Feng. “Improving sensitivity and connectivity of retinal vessel segmentation via error discrimination network.” In: *Medical Physics* 49.7 (2022), pages 4494–4507. ISSN: 2473-4209. https://doi.org/10.1002/mp.15627. URL: http://dx.doi.org/10.1002/mp.15627.

[34] J. Liu, D. Zhao, J. Shen, P. Geng, Y. Zhang, J. Yang, Z. Zhang. “HRD-Net: High resolution segmentation network with adaptive learning ability of retinal vessel features.” In: *Computers in Biology and Medicine* 173 (2024), page 108295. ISSN: 0010-4825. https://doi.org/10.1016/j.compbiomed.2024.108295. URL: http://dx.doi.org/10.1016/j.compbiomed.2024.108295.

[35] V. Mayya, S. Kamath S., U. Kulkarni. “Automated microaneurysms detection for early diagnosis of diabetic retinopathy: A Comprehensive review.” In: *Computer Methods and Programs in Biomedicine Update* 1 (2021), page 100013. ISSN: 2666-9900. https://doi.org/10.1016/j.cmpbup.2021.100013. URL: http://dx.doi.org/10.1016/j.cmpbup.2021.100013.

[36] H. E. Moss, J. Cao, M. Wasi, S. E. Feldon, M. Shahidi. “Variability of Retinal Vessel Tortuosity Measurements Using a Semiautomated Method Applied to Fundus Images in Subjects With Papilledema.” In: *Translational Vision Science amp; Technology* 10.14 (2021), page 32. ISSN: 2164-2591. https://doi.org/10.1167/tvst.10.14.32. URL: http://dx.doi.org/10.1167/tvst.10.14.32.

[37] M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. van Ginneken, J. C. Folk, M. D. Abramoff. “Automated Measurement of the Arteriolar-to-Venular Width Ratio in Digital Color Fundus Photographs.” In: *IEEE Transactions on Medical Imaging* 30.11 (2011), pages 1941–1950. ISSN: 1558-254X. https://doi.org/10.1109/tmi.2011.2159619. URL: http://dx.doi.org/10.1109/tmi.2011.2159619.

[38] A. Oliveira, S. Pereira, C. A. Silva. “Retinal vessel segmentation based on Fully Convolutional Neural Networks.” In: *Expert Systems with Applications* 112 (2018), pages 229–242. ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2018.06.034. URL: http://dx.doi.org/10.1016/j.eswa.2018.06.034.

[39] J. I. Orlando, J. Barbosa Breda, K. van Keer, M. B. Blaschko, P. J. Blanco, C. A. Bulant. “Correction to: Towards a Glaucoma Risk Index Based on Simulated Hemodynamics from Fundus Images.” In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Springer International Publishing, 2018, E1–E1. ISBN: 9783030009342. https://doi.org/10.1007/978-3-030-00934-2_106. URL: http://dx.doi.org/10.1007/978-3-030-00934-2_106.

[40] T. Pethmunee, S. Kansomkeat, P. Bhurayanontachai, S. Intajag. “Confident Learning-Based Label Correction for Retinal Image Segmentation.” In: *Diagnostics* 15.14 (2025), page 1735. ISSN: 2075-4418. https://doi.org/10.3390/diagnostics15141735. URL: http://dx.doi.org/10.3390/diagnostics15141735.

[41] S. P. Prasanna Porwal. *Indian Diabetic Retinopathy Image Dataset (IDRiD)*. 2018. https://doi.org/10.21227/H25W98. URL: https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid.

[42] K. Radha, Y. Karuna. “Retinal vessel segmentation to diagnose diabetic retinopathy using fundus images: A survey.” In: *International Journal of Imaging Systems and Technology* 34.1 (2023). ISSN: 1098-1098. https://doi.org/10.1002/ima.22945. URL: http://dx.doi.org/10.1002/ima.22945.

[43] N. Ramírez. “Impact of image resolution on retinal vessel tortuosity assessment.” In: *Optica Pura y Aplicada* (2024). ISSN: 2171-8814. https://doi.org/10.7149/opa.57.1.51171. URL: http://dx.doi.org/10.7149/OPA.57.1.51171.

[44] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, et al. *SAM 2: Segment Anything in Images and Videos*. 2024. https://doi.org/10.48550/ARXIV.2408.00714. URL: https://arxiv.org/abs/2408.00714.

[45] *RetinaCheck*. URL: https://www.retinacheck.org/.

[46] R. Santos-Mateos, A. Velev-Santos, X. M. Pardo. “VesselView: A CNN for Segmentation of Vessels in High-Resolution Retinal Fundus Images.” In: *Pattern Recognition and Image Analysis*. Springer Nature Switzerland, 2025, pages 258–270. ISBN: 9783031995682. https://doi.org/10.1007/978-3-031-99568-2_21. URL: http://dx.doi.org/10.1007/978-3-031-99568-2_21.

[47] *STARE*. URL: https://cecas.clemson.edu/~ahoover/stare/.

[48] H. Su, L. Gao, Z. Wang, Y. Yu, J. Hong, Y. Gao. “A Hierarchical Full-Resolution Fusion Network and Topology-Aware Connectivity Booster for Retinal Vessel Segmentation.” In: *IEEE Transactions on Instrumentation and Measurement* 73 (2024), pages 1–16. https://doi.org/10.1109/TIM.2024.3411133.

[49] J. Van Eijgen, J. Fhima, M.-I. Billen Moulin-Romsée, J. A. Behar, E. Christinaki, I. Stalmans. “Leuven-Haifa High-Resolution Fundus Image Dataset for Retinal Blood Vessel Segmentation and Glaucoma Diagnosis.” In: *Scientific Data* 11.1 (2024). ISSN: 2052-4463. https://doi.org/10.1038/s41597-024-03086-6. URL: http://dx.doi.org/10.1038/s41597-024-03086-6.

[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. “Attention Is All You Need.” In: *CoRR* abs/1706.03762 (2017). URL: http://arxiv.org/abs/1706.03762.

[51] J. Wang, K. Sun, T. Cheng, B. Jiang, et al. “Deep High-Resolution Representation Learning for Visual Recognition.” In: *CoRR* abs/1908.07919 (2019). URL: http://arxiv.org/abs/1908.07919.

[52] C. Wiedemann, C. Heipke, H. Mayer, O. Jamet. “Empirical Evaluation Of Automatically Extracted Road Axes.” In: (1998).

[53] Z. Wu, X. Xiong. “Vessel-SAM2: Adapting Segment Anything 2 for Patch-Free Retinal Vessel Segmentation in Ultra-High Resolution Fundus Images.” In: *IEEE Sensors Letters* 9.9 (2025), pages 1–4. ISSN: 2475-1472. https://doi.org/10.1109/lsens.2025.3595139. URL: http://dx.doi.org/10.1109/LSENS.2025.3595139.

[54] Y. Xu, Y. Fan. “Dual-channel asymmetric convolutional neural network for an efficient retinal blood vessel segmentation in eye fundus images.” In: *Biocybernetics and Biomedical Engineering* 42.2 (2022), pages 695–706. ISSN: 0208-5216. https://doi.org/10.1016/j.bbe.2022.05.003. URL: http://dx.doi.org/10.1016/j.bbe.2022.05.003.

[55] Y. Zafer, C. Ikibas, K. Cemal. “A simple analytic approach for tracking retinal vessels and measuring their diameter.” In: *Proceedings of the First International Conference on Bioinformatics*. BIOINFORMATICS. SciTePress - Science, 2010, pages 13–18. https://doi.org/10.5220/0002694200130018. URL: http://dx.doi.org/10.5220/0002694200130018.

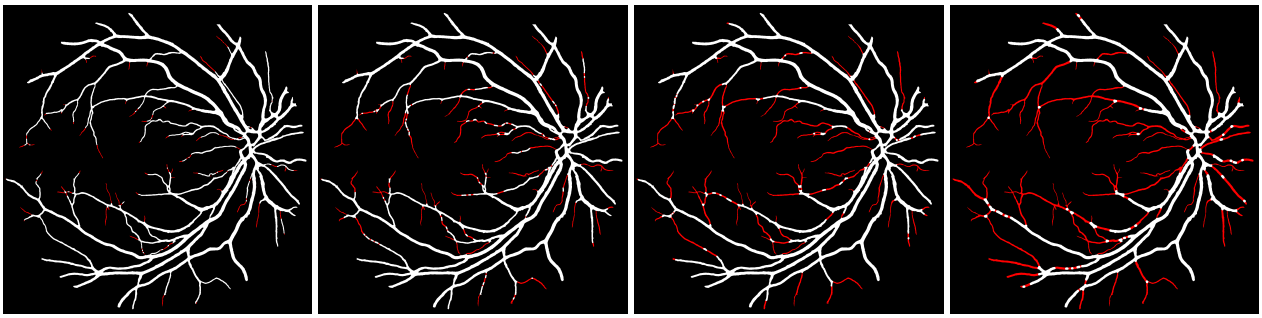
[56] Y. Zhang, M. He, Z. Chen, K. Hu, X. Li, X. Gao. “Bridge-Net: Context-involved U-net with patch-based loss weight mapping for retinal blood vessel segmentation.” In: *Expert Systems with Applications* 195 (2022), page 116526. ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2022.116526. URL: http://dx.doi.org/10.1016/j.eswa.2022.116526.

[57] L. Zhou, Q. Yu, X. Xu, Y. Gu, J. Yang. “Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement.” In: *Computer Methods and Programs in Biomedicine* 148 (2017), pages 13–25. ISSN: 0169-2607. https://doi.org/10.1016/j.cmpb.2017.06.016. URL: http://dx.doi.org/10.1016/j.cmpb.2017.06.016.

[58] Y. Zou, F. Abdolahi, J. A. Gluckstein, K. K. Gokoffski, R. Zinjal, A. E. Felder, M. Shahidi. “A Method for Peripapillary Retinal Vessel Tortuosity Measurement and Its Clinical Correspondence.” In: *Translational Vision Science amp; Technology* 14.10 (2025), page 34. ISSN: 2164-2591. https://doi.org/10.1167/tvst.14.10.34. URL: http://dx.doi.org/10.1167/tvst.14.10.34.

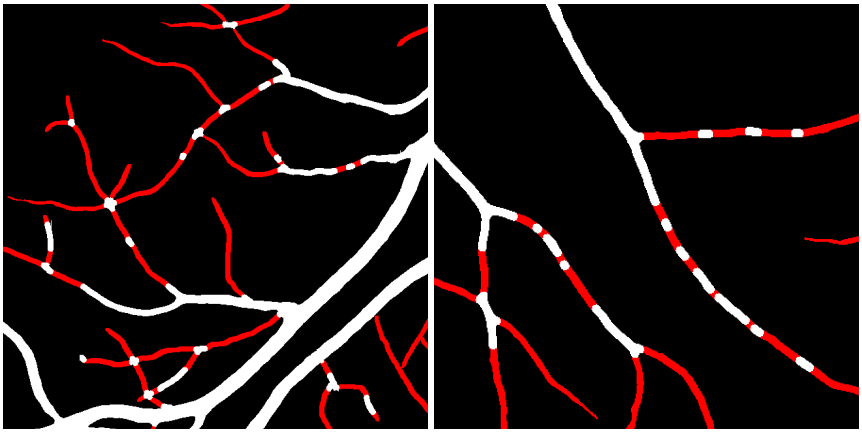
Appendix

Link to github: `https://github.com/AntanasBuk/ThinVesselSeg`



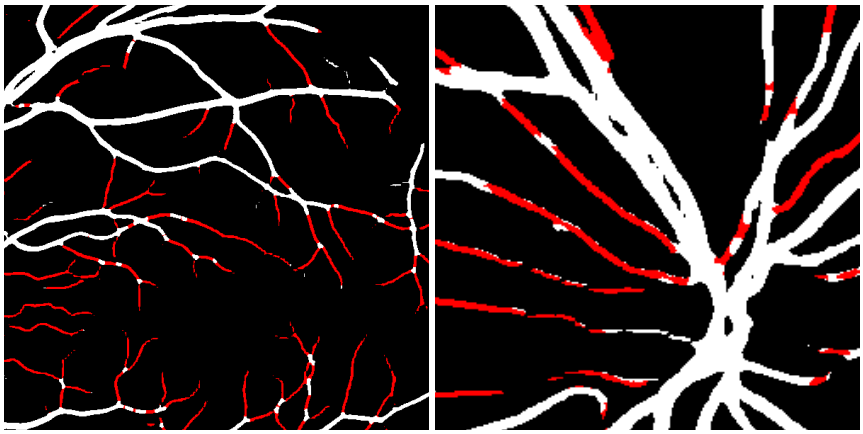
(a) *ø 5* ***(b)*** *ø 8* ***(c)*** *ø 10* ***(d)*** *ø 15*

Figure 6 *An example of thin vessel (red pixel) separation algorithm output for different thickness thresholds. Here a threshold of 5 only captures the ends of thin vessels; a threshold of 8 provides a good approximation on which pixels should be considered as thin; a threshold of 10 already starts to capture medium sized vessels; a threshold of 15 captures parts of major thick vessels*



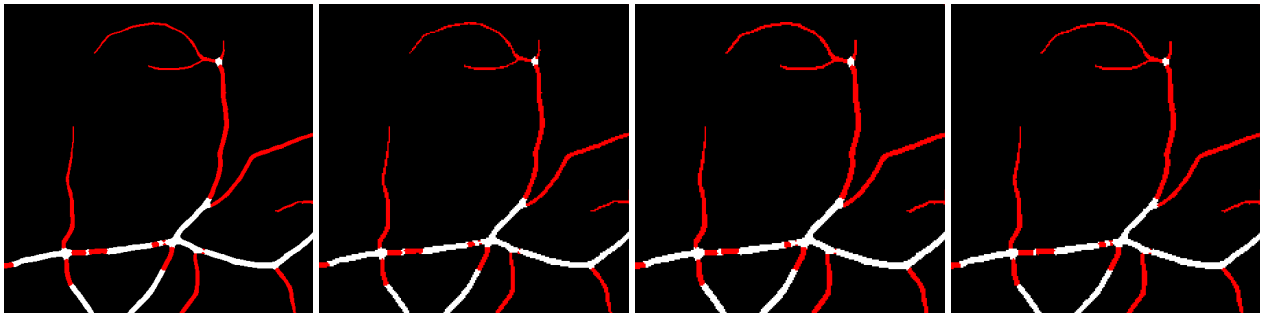
(a) *Bifurcation problems* ***(b)*** *Threshold problems*

Figure 7 *Bifurcation points (a) are interpreted as thick vessels (white pixel); vessels with width varying around thickness threshold (b) impacts separation connectivity*



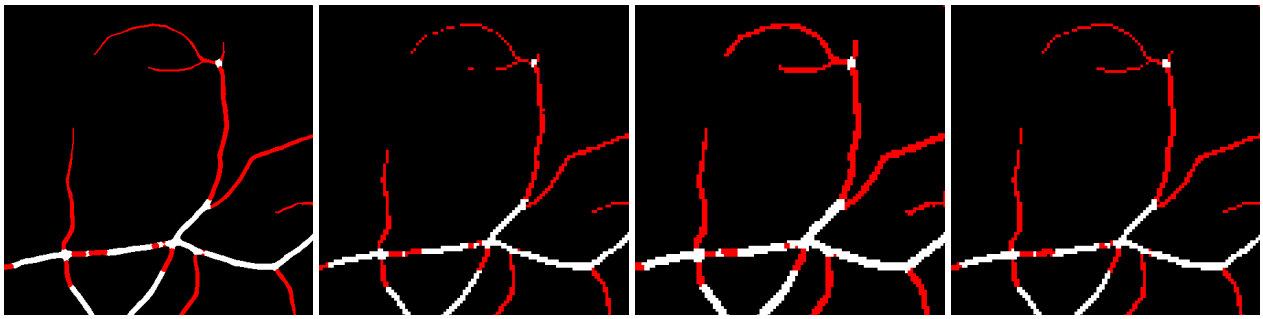
***(a)** Good buffer overlap* ***(b)** Bad buffer overlap*

***Figure 8** An example of good (a) vessel prediction interpretation for thin vessels (red pixel) where most of thin vessel segments are within thin vessel buffer zone; and an example of bad (b) vessel prediction interpretation for thin vessels where gaps are left on thin vessel walls and false overlaps with thick vessels (white pixel)*



***(a)** Ground truth* ***(b)** Nearest neighbor* ***(c)** Maxpool* ***(d)** Bilinear*

***Figure 9** Visual thin vessel (red pixels) structure degradation when downscaling to 1024×1024 resolution using different interpolation methods*



***(a)** Ground truth* ***(b)** Nearest neighbor* ***(c)** Maxpool* ***(d)** Bilinear*

***Figure 10** Visual thin vessel structure (red pixels) degradation when downscaling to 512×512 resolution using different interpolation methods*