Disentangled Retinal Fundus Synthesis with Non-Circular Vessel-Topology Evaluation

Syed Abdullah Basit(✉) and Tanvir Alam

College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
syba62904@hbku.edu.qa

Abstract. Self-supervised disentanglement of vessel anatomy from appearance is commonly validated by an encoder-projection cosine between source and swap. **This validation is circular**: it measures invariance of the very encoder used to generate the swap, not preservation of anatomy. A totally vessel-destroyed image (Gaussian blur, σ=8) still reads as 0.92 "anatomy retention" while a non-circular vessel-topology metric — clDice via an independently-trained FIVES segmenter — collapses to 0.003, below an unrelated-pair floor of 0.11; ablations move the cosine by +0.08 and −0.15 while clDice stays flat. We propose segmenter-based clDice, read against an unrelated-pair floor, as the evaluation protocol for vessel-bearing medical image synthesis. Applying it to a latent-diffusion testbed, a vessel-mask ControlNet, and a MUNIT-style GAN shows that conclusions drawn under it depend on controls that are easy to omit: a clDice margin inverts once resolutions are matched, since resolution dominates the metric, and the mask baseline exhibits a *second* circularity mode — it is conditioned on the evaluating segmenter's own output. Non-circularity is therefore necessary but not sufficient: a topology metric must be paired with matched resolution and compute, an audit of conditioning–evaluator independence, and complementary evidence beyond a single scalar.

Keywords: Evaluation methodology · Disentangled representation learning · Retinal vessel topology · Latent diffusion · Metric circularity.

1 Introduction

Retinal fundus images underpin non-invasive screening for diabetic retinopathy, glaucoma, and hypertensive disease, where automated assessment hinges on faithful preservation of the *vessel tree* [16]. Synthesising fundus images that preserve a patient's vessel topology while varying illumination, colour, and capture artefacts serves data augmentation, privacy-preserving sharing, and controlled studies of acquisition shift [18]. This requires *disentanglement*: separately exposing vessel anatomy and appearance so that one transfers without disturbing the other. Throughout, "anatomy" denotes vessel topology specifically, not retinal anatomy in general; §5 quantifies where that scope falls short (optic disc).

A natural self-supervised approach learns an anatomy encoder E_a and an appearance encoder E_s jointly, with an invariance objective on E_a (e.g., NT-Xent contrastive [3]) and a variance objective on E_s (e.g., VICReg [2]); the

obvious sanity check for anatomy preservation under the swap is then the cosine $\cos(E_a(A), E_a(\text{swap}))$ between the same encoder's own embeddings, which we term "anatomy retention". **We show that this check is circular.** The swap is generated conditional on $E_a(A)$, and E_a is trained to be invariant to view augmentation, so the cosine measures the encoder's invariance — a property it was optimised to satisfy — not whether the swap's vessel tree matches A's. Gaussian-blurring at σ=8 destroys the vessel tree (clDice 0.003, below the 0.11 floor) yet the cosine still reports 0.924, monotonically blind to vessel destruction.

We therefore replace it with a **non-circular** protocol: clDice [15], the centerline-Dice topology metric, computed on vessel masks produced by an *independently-trained* U-Net segmenter (FIVES [10]), and always reported against an unrelated-pair floor so preservation reads as a margin over chance. The segmenter is never exposed to the generative model, so the protocol measures vessel preservation without invoking the encoder it was designed to evaluate. To exercise it we build a self-supervised latent-diffusion testbed (§3.1) and compare it against a vessel-mask ControlNet and a MUNIT-style GAN.

Contributions. **(i)** We diagnose encoder-cosine circularity along two orthogonal axes: the blur sweep above, and component ablations in which the cosine *rises* (+0.08) when the vessel-fidelity loss is removed and *falls* (−0.15) when contrastive invariance is removed, while clDice stays flat. **(ii)** We propose the non-circular protocol above and validate its stability under a second, architecturally-independent segmenter. **(iii)** We show non-circularity is necessary but not sufficient: a clDice margin inverts once resolutions are matched, and the mask baseline exhibits *conditioning–evaluator coupling*, a second circularity mode a non-circular metric does not remove. We give a checklist — match resolution and compute, audit conditioning–evaluator independence, report complementary non-topology evidence.

2 Related Work

Disentangled image-to-image translation. A line of work decomposes an image into content and style codes with separate encoders, recombining them for translation; MUNIT [9], DRIT++ [11], and StarGAN-v2 [4] are prototypical. These papers evaluate primarily with distributional and perceptual scores (FID, LPIPS diversity) rather than a source-to-swap encoder cosine. Our target is the cosine check itself — the content-preservation diagnostic any self-supervised content/style encoder pair invites, since source and swap already live in the embedding space used to condition the swap — which to our knowledge has not been stress-tested against ground-truth structure. Biases of disentanglement metrics more broadly have been studied [12]; our critique targets this self-referential cosine wherever it validates an encoder-conditioned swap, medical or otherwise.

Medical synthesis, spatial conditioning, and topology metrics. Retinal generation preserving clinical content has used vessel-conditioned GANs [5], cycle-consistent

translation [20], and diffusion models [13], mostly evaluated with pixel-wise or perceptual losses plus mask-based comparisons. None systematically addresses the circularity that arises when the disentanglement encoder is itself the evaluator — nor, as we show in §4.5, the related failure when a model's conditioning signal comes from the evaluator. ControlNet [19] injects spatial conditioning into a frozen UNet via zero-initialised residuals; our multi-scale anatomy injector follows this principle, and a vessel-mask ControlNet variant serves as a comparison system. clDice [15] is a centerline-Dice variant for tubular structures; earlier work pairs it with topology-aware losses [8], whereas we use it strictly for evaluation, via a segmenter independent of the generative model.

3 Method

We first summarise the synthesis testbed used to exercise the protocol (§3.1), then state the protocol itself (§3.2).

3.1 Synthesis Testbed: Anatomy-Style Disentangled Latent Diffusion

The testbed exercises the protocol rather than being a contribution in itself; its full configuration is in the supplementary material.

Setup and encoders. Inputs are RGB fundus images x normalised to $[-1, 1]$, $H{=}W{=}512$. A frozen Stable Diffusion VAE [14] maps $x \mapsto z = \mathrm{VAE}_{\mathrm{enc}}(x) \cdot s$, $s{=}0.18215$, giving the diffusion target $z \in \mathbb{R}^{4\times H/8\times W/8}$. Two trainable encoders consume x: an anatomy encoder E_a producing a spatial map $\mathbf{m}_a \in \mathbb{R}^{32\times h\times w}$ and a projection $\mathbf{p}_a$ (used only for contrastive training), and an appearance encoder E_s producing $\mathbf{v}_s \in \mathbb{R}^{768}$ plus its projection. A $1{\times}1$ convolution bottleneck with InstanceNorm, $\mathrm{Bot} : \mathbb{R}^{32} \to \mathbb{R}^{8}$, gives a vessel-rich low-channel code $\mathbf{c} = \mathrm{Bot}(\mathbf{m}_a)$ at latent resolution; its narrowness matters for the comparison in §4.6.

UNet and conditioning. A UNet ϵ_θ predicts the noise added to z_t. Conditioning enters through four paths (Figure 1): anatomy by *input-concat* of $\mathbf{c}$ with z_t and by zero-initialised *multi-scale ControlNet residuals* at every UNet scale; style by *cross-attention* on $\mathbf{v}_s$ expanded to $N{=}4$ tokens and by *AdaGN* in every ResNet block. Anatomy and style conditioning are dropped independently during training, so three UNet passes combine at inference into compositional classifier-free guidance on anatomy (α_a) and style (α_s) separately (supplementary).

Training objectives. The loss combines (a) diffusion MSE; (b) anatomy contrastive NT-Xent [3] on $\mathbf{p}_a$ across two augmented views; (c) appearance VICReg [2] on $\mathbf{v}_s$; (d) style reconstruction — a low-timestep $\hat{z}_0$ prediction, VAE-decoded and re-encoded by E_s, with cosine loss against a swapped reference style, forcing ϵ_θ to use $\mathbf{v}_s$ rather than the anatomy path; (e) vesselness fidelity

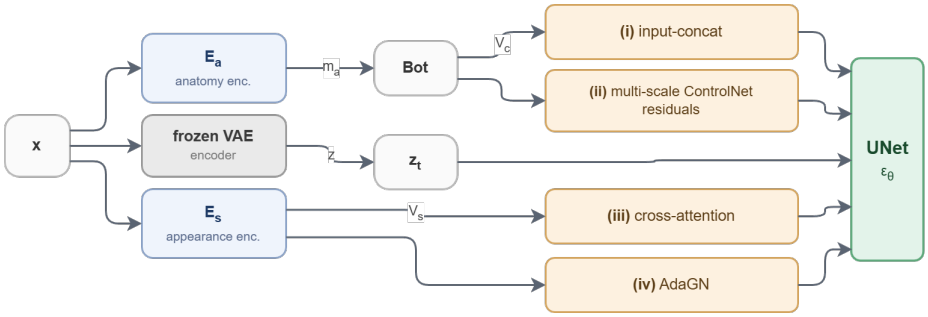

Fig. 1. Testbed conditioning pipeline. Encoders E_a, E_s and a frozen VAE consume x; anatomy ($\mathbf{c}$, paths (i)–(ii)) and appearance ($\mathbf{v}_s$, paths (iii)–(iv)) reach the UNet through four routes (§3.1).

— a small decoder predicts a vesselness mask from $\mathbf{c}$, trained against the thresholded multiscale-Hessian vesselness [7] of x's green channel. Term (e) is the only loss explicitly requiring the anatomy stream to encode vessels; ablating it (§4.3) drives the encoder cosine *up* while clDice stays flat. Loss weights and optimisation are in the supplementary material.

3.2 Non-Circular Vessel-Topology Protocol

The circular metric. Let $\mathrm{swap}(A, B)$ denote ϵ_θ's sampled output conditioned on $\mathbf{c}_A = \mathrm{Bot}(E_a(A))$ and $\mathbf{v}_{s,B} = E_s(B)$. The encoder-cosine "anatomy retention" reads $R_a(A, B) = \cos(E_a(A), E_a(\mathrm{swap}(A, B)))$. Here E_a is the same encoder that produced the conditioning, and it is trained for invariance to view augmentation; the cosine therefore measures how well that invariance generalises across the generative model's image manifold, not whether the swap's vessel tree matches A's.

The non-circular replacement. Let $\mathrm{Seg} : x \mapsto m \in \{0,1\}^{1\times H\times W}$ be a binary vessel segmenter — a U-Net with a ResNet-34 encoder — trained on the FIVES dataset [10] *prior to and independently of* the generative model, then frozen. We report vessel topology preservation as

$$\mathrm{clDice}\big(\mathrm{Seg}(\mathrm{swap}(A, B)),\, \mathrm{Seg}(A)\big), \tag{1}$$

where clDice is the harmonic mean of skeleton-versus-mask precision and sensitivity [15]. The segmenter has the two properties the protocol needs: *independence* — it was never optimised jointly with ϵ_θ, E_a, or E_s; and *topology sensitivity* — clDice penalises connectivity errors more strongly than dense Dice. We additionally report the unrelated-pair floor $\mathrm{clDice}(\mathrm{Seg}(B), \mathrm{Seg}(A))$, so preservation reads as a margin over chance.

Two circularity modes. Write C for a model's conditioning signal, M for the evaluator. *(i) Evaluator–generator coupling*: $M = E_a$, the encoder that produced C — the encoder cosine, removed by choosing an independent segmenter. *(ii) Conditioning–evaluator coupling*: $C = M(A)$ — the model is conditioned on the evaluator's own output, so M grades a model built from its own predictions, right or not. A non-circular M does *nothing* about (ii): it relocates the circularity to the conditioning path. Any mask-conditioned model scored by its own segmenter exhibits it, including the ControlNet of §4.5, so auditing $C \perp M$ is part of the protocol.

4 Experiments

4.1 Setup

Data and segmenters. We use 41,390 fundus images from two public DR-screening corpora, Kaggle EyePACS [6] and APTOS 2019 [1], quality-filtered by a Laplacian-variance gate and resized to 512^2. The evaluating segmenter is a U-Net with a ResNet-34 encoder (segmentation-models-pytorch [17]) trained on FIVES's 800 ground-truth pairs [10] to validation Dice > 0.7, then frozen; *Seg2* is a ResNet-18 counterpart (test Dice 0.889; supplementary).

Training and sampling. AdamW at 10^{-4} peak LR, bf16, on one RTX 3090 (24 GB); a 512^2 run takes $\sim$30 h and each 256^2 ablation $\sim$6 h. The 512^2 mask baseline (§4.5) is warm-started from its 256^2 checkpoint and fine-tuned at 3×10^{-5}, matching our model's resolution and sampler. Loss weights, parameter groups, and schedule are in the supplementary material. Sampling uses DDIM 20 steps with α_s=1.0, α_a=3.0 unless stated otherwise.

Reported quantities. Each cell in Table 1 is a mean over N=300 paired swaps (seed 42, identical pair indices across rows). PSNR/SSIM are A$\rightarrow$A self-reconstruction; "Style" is $\cos(E_s(B), E_s(\text{swap}))$ and "EncRet" the circular $\cos(E_a(A), E_a(\text{swap}))$ we critique, both on the encoders' contrastive projection heads — the spaces trained for invariance, which is what makes EncRet circular.

4.2 The Encoder Cosine Is Blind to Vessel Destruction

To show the encoder cosine does not measure what its name implies, we construct a controlled degradation *outside the generative model*: Gaussian-blur the input with $\sigma \in [0, 16]$ and measure both segmenter clDice and encoder cosine between original and blurred. The blur eliminates fine vessels while preserving the coarse layout the encoder is invariant to, so any valid "anatomy retention" metric must decline with clDice.

At σ=8 the segmenter-measured topology has collapsed to clDice 0.003, below the 0.11 unrelated-pair floor, yet the encoder cosine still reports 0.924; even at σ=16 (a smooth gradient) it reports 0.78. Across the sweep clDice falls from

Table 1. Evaluation over N=300 paired swaps (seed 42, identical pairs across rows). *CI*: bootstrap 95% interval on mean clDice. *Seg2*: a second, architecturally-independent segmenter (ResNet-18), tracking clDice on every row. *fl*: unrelated-pair floor. *CDist*: non-encoder colour-histogram distance to B (lower better). *OD*: classical-CV optic-disc overlap (floor .094/.105). *Style/EncRet*: the circular encoder cosines we critique (undefined for mask baselines). Bold = best per column; the testbed leads only on PSNR and OD.

Variant	clDice ↑	95% CI	Seg2 ↑	Dice ↑	fl	SSIM ↑	PSNR ↑	CDist ↓	OD ↑	Style/EncRet
Ours, full (512^2)	0.559	[.548,.569]	0.554	0.591	.109	0.615	**20.43**	0.067	**0.476**	0.877 / 0.830
Ours, full (256^2)	0.314	[.307,.321]	0.308	0.396	.112	0.616	22.00	**0.060**	0.664	0.783 / 0.818
− multi-scale residuals	0.293	[.285,.300]	0.292	0.368	.112	0.603	21.43	0.061	0.661	0.821 / 0.824
− vesselness fidelity loss	0.342	[.335,.349]	0.338	0.417	.112	**0.622**	22.43	0.065	0.706	0.649 / **0.899**
− contrastive	0.338	[.331,.345]	0.323	0.418	.112	0.594	21.02	0.065	0.633	0.739 / **0.664**
Mask ControlNet baseline (256^2)	0.239	[.235,.243]	0.241	0.289	.112	0.536	16.04	0.092	0.236	0.663 / —
Mask ControlNet baseline (512^2)	**0.582**	[.577,.588]	**0.579**	**0.617**	.109	0.608	16.99	0.066	0.193	**0.905** / —

0.999 to 0.0003 while the cosine never drops below 0.78 (N=40; plotted in the supplementary material). The two decouple monotonically: the encoder cosine cannot serve as a vessel-preservation metric because, by design, it does not see vessels at the scale where vessels live.

4.3 Disentanglement Quality and Ablations

Table 1 shows our main result (512^2, top row) with a from-scratch 256^2 reference and three component ablations. The vessel-mask ControlNet baseline appears at both resolutions and is analysed in §4.5, where the resolution-matched comparison reverses the ranking we reported at 256^2.

Resolution dominates vessel topology: clDice rises from 0.314 at 256^2 to 0.559 at 512^2 (§4.4 isolates resolution itself from rendering quality). The remaining components shift clDice by at most ± 0.03 — yet the encoder cosine moves substantially and *in opposite directions*. Removing the vesselness fidelity loss (λ_{struct}=0) leaves clDice flat ($0.314 \rightarrow 0.342$) while the cosine *rises* to 0.899: with no structural pull on the bottleneck, E_a becomes more invariance-dominated, exactly what the cosine rewards. Removing contrastive invariance ($\lambda_{\text{contrastive}}$=0) again leaves clDice flat ($0.314 \rightarrow 0.338$) while the cosine *falls* to 0.664: E_a is no longer optimised for invariance, so its self-cosine drifts toward the inter-image baseline.

The cosine thus swings $+0.08$ and -0.15 while clDice barely moves, with the ordering unchanged under *Seg2* — a controlled demonstration of circularity independent of §4.2's degraded-input one. Secondarily, multi-scale residual injection adds +0.02 clDice, and style transfer drops when either the vesselness ($0.78 \rightarrow 0.65$) or contrastive ($0.78 \rightarrow 0.74$) loss is removed: both aid disentanglement *quality* even where neither is needed for topology.

4.4 Resolution Confound and Distributional Quality

Since resolution dominates the clDice gain (§4.3), it is unclear whether $256^2 \rightarrow 512^2$ improves vessel *rendering* or merely the segmenter's sensitivity to sharper input.

We separate the two on real images alone, with no generative model involved: for N=200 real images, clDice$(\mathrm{Seg}(x), \mathrm{Seg}(\downarrow_{256}\uparrow_{512}(x)))$ is the score a *perfectly preserved* vessel tree would receive after a 256px bottleneck. This ceiling is 0.909, far above both generative numbers (0.314, 0.559): the segmenter tolerates 256px inputs, so the gap reflects rendering quality, not segmenter unreliability, and matching resolution before comparing systems is mandatory rather than cosmetic. FID between 200 real and 200 swaps at 512^2 is 148.5, a sanity check only at this N.

4.5 Vessel-Mask ControlNet Baseline

The most directly competitive baseline replaces the learned anatomy encoder with a hard-coded vessel mask from the same independent FIVES segmenter, projected to the same bottleneck channel slot, with everything else identical; with no learned anatomy projection to be invariant on, the contrastive loss is disabled for this row and EncRet is undefined. Compared at 256^2, the baseline underperforms the testbed on every metric (+0.07 clDice, +6 dB PSNR, better ColorDist and OD). **This comparison is confounded by resolution**: the testbed's headline configuration runs at 512^2, and §4.4 shows resolution dominates clDice. We therefore warm-start the baseline to 512^2 (the nets are fully convolutional), matching the testbed's resolution, optimiser, and schedule.

At matched resolution the ranking inverts and the baseline wins on vessel topology: clDice 0.582 $[0.577, 0.588]$ against 0.559 $[0.548, 0.569]$, non-overlapping intervals, with the gap holding under the second segmenter (0.579 vs. 0.554), on dense Dice, and on style cosine. The testbed leads on reconstruction fidelity (+3.4 dB PSNR) and, decisively, on optic-disc preservation (OD 0.476 vs. 0.193, floor 0.094). A clDice margin measured across a resolution mismatch is therefore uninterpretable, however non-circular the metric itself.

This row also exhibits circularity mode (ii) (§3.2): the baseline is *conditioned on* $\mathrm{Seg}(A)$ and *scored by* clDice$(\mathrm{Seg}(\mathrm{swap}), \mathrm{Seg}(A))$, i.e. $C = M(A)$, so its clDice rewards reproducing the evaluator's own predictions — what a mask-conditioned model is trained to do. The score is not invalid, but neither is it clean anatomical fidelity. *Seg2* does not settle it: trained on the same FIVES ground truth, its masks correlate with the first's, so it tests segmenter architecture rather than $C \perp M$; a conclusive test needs a conditioning segmenter trained on disjoint data, beyond our budget. The resolution confound and the OD/PSNR gaps hold independently of the coupling.

4.6 A Classical GAN, and Downstream Utility

A classical GAN. To situate both models against the literature our critique targets, we add a single-domain MUNIT-style [9] content/style autoencoder-GAN at 256^2, trained for 15k iterations. It reaches clDice 0.706 ($[0.698, 0.714]$, floor 0.112), above both the testbed (0.559) and the resolution-matched mask baseline (0.582) despite lower resolution and budget; a ColorDist diagnostic excludes the degenerate explanation that its decoder copies A, and we attribute the margin

to its far wider content path. Mode (ii) does *not* apply — it never sees the evaluating segmenter — so under clDice alone the testbed ranks last of the three, a ranking the encoder cosine would not reveal.

Downstream utility. Topology preservation is not utility. Augmenting a class-balanced DR-grading train set (500/class) with one testbed swap per real image gives a *mixed* result on an identical held-out real set: accuracy $0.520 \rightarrow 0.512$, quadratic-weighted κ $0.732 \rightarrow 0.759$. DR grade depends on non-vessel pathology a vessel/appearance decomposition cannot carry across a swap — the gap §5 finds at the optic disc — so a model can preserve topology well and still be unhelpful downstream. Both protocols are in the supplementary material.

5 Discussion and Limitations

Optic-disc fidelity. The optic disc renders as an OD-sized bright blob at roughly the right location but without internal structure (vessel-emergence boundary, cup-to-disc cues). A classical-CV OD detector, independent of both the segmenter and the generative model, gives OD-Dice 0.476 at 512^2 against a floor of 0.094 — a real margin, but far below our clDice on the vessel tree. The resolution-matched mask baseline scores 0.193, barely above the same floor: whatever the learned anatomy stream costs in clDice, it is what carries non-vessel macro-anatomy through the swap at all. The bottleneck, trained by NT-Xent invariance and a vesselness loss that suppresses blob-like structures, under-represents this anatomy, and §4.6's mixed result follows.

Seed variance and segmenter dependence. Over n=4 independent 256^2 trainings, seed std is small (clDice $\pm$0.016, EncRet $\pm$0.024): the EncRet ablation swings ($+0.08$, -0.15) are 3–7σ while clDice barely moves, placing the circularity well outside seed noise. The 512^2 runs are single ($\sim$30 h each), bounded by the 256^2 seed std. The protocol also depends on its segmenter; a different one shifts absolute numbers, which we mitigate by always reporting the floor. Recomputing every row under *Seg2* (Table 1) leaves the ablation ordering and the baseline comparison unchanged — establishing robustness to segmenter *architecture*, not conditioning–evaluator independence (§4.5); an ensemble over segmenters trained on disjoint corpora would settle both, and remains future work.

Style transfer is also an encoder cosine. Our "Style" column, $\cos(E_s(B), E_s(\text{swap}))$, is circular in the same way, which is why we pair it with the non-encoder ColorDist. The two diverge: the 512^2 mask baseline posts the highest Style cosine of any row (0.905) on an unremarkable ColorDist (0.066). Style's ordering among close variants should not be over-interpreted; we report it only for continuity with prior work.

6 Conclusion

The encoder-cosine "anatomy retention" metric is *circular*: it measures invariance of the encoder that generated the swap, not preservation of anatomy. Our non-circular replacement — clDice via an independently-trained segmenter, read against an unrelated-pair floor — is stable under a second segmenter, but applying it across three architectures shows non-circularity is necessary and not sufficient. We recommend that work evaluating conditional medical-image synthesis (i) prefer segmenter-based topology metrics over encoder-projection cosines, (ii) match resolution and compute before reading any margin, (iii) audit whether a model's conditioning signal and its evaluator share a source, and (iv) report complementary evidence so no single scalar carries the claim. Conditions (ii) and (iii) are not hypothetical: each changes a conclusion above.

7 Acknowledgement

This study was supported by Qatar Research Development and Innovation (QRDI) Council fund under Path Towards Precision Medicine PPM-07-0418-240048.

References

1. Asia Pacific Tele-Ophthalmology Society: APTOS 2019 blindness detection. Kaggle competition, https://www.kaggle.com/c/aptos2019-blindness-detection (2019)
2. Bardes, A., Ponce, J., LeCun, Y.: VICReg: Variance-invariance-covariance regularization for self-supervised learning. In: The Tenth International Conference on Learning Representations (ICLR). OpenReview.net (2022), https://openreview.net/forum?id=xm6YD62D1Ub
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research, vol. 119, pp. 1597–1607. PMLR (2020), http://proceedings.mlr.press/v119/chen20j.html
4. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: StarGAN v2: Diverse image synthesis for multiple domains. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8185–8194 (2020). https://doi.org/10.1109/CVPR42600.2020.00821
5. Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendonça, A.M., Campilho, A.: End-to-end adversarial retinal image synthesis. IEEE Transactions on Medical Imaging (2018). https://doi.org/10.1109/TMI.2017.2759102
6. EyePACS, California Healthcare Foundation: Diabetic retinopathy detection. Kaggle competition, https://www.kaggle.com/c/diabetic-retinopathy-detection (2015)
7. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 1998, Lecture Notes in Computer Science, vol. 1496, pp. 130–137. Springer Berlin Heidelberg (1998). https://doi.org/10.1007/BFb0056195

8. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
9. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Computer Vision – ECCV 2018, Lecture Notes in Computer Science, vol. 11207, pp. 179–196. Springer International Publishing (2018). https://doi.org/10.1007/978-3-030-01219-9_11
10. Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., Zhang, Q., Wang, Y., Ye, J.: FIVES: A fundus image dataset for artificial intelligence based vessel segmentation. Scientific Data (2022). https://doi.org/10.1038/s41597-022-01564-3
11. Lee, H.Y., Tseng, H.Y., Mao, Q., Huang, J.B., Lu, Y.D., Singh, M., Yang, M.H.: DRIT++: Diverse image-to-image translation via disentangled representations. International Journal of Computer Vision pp. 2402–2417 (2020). https://doi.org/10.1007/s11263-019-01284-z
12. Liu, X., Thermos, S., Valvano, G., Chartsias, A., O'Neil, A., Tsaftaris, S.A.: Measuring the biases and effectiveness of content-style disentanglement. arXiv preprint arXiv:2008.12378 (2020)
13. Pinaya, W.H.L., Tudosiu, P.D., Dafflon, J., Da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J.: Brain imaging generation with latent diffusion models. In: Deep Generative Models – MICCAI Workshop 2022. Lecture Notes in Computer Science, Springer (2022). https://doi.org/10.1007/978-3-031-18576-2_12
14. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10674–10685 (2022). https://doi.org/10.1109/CVPR52688.2022.01042
15. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P.W., Bauer, U., Menze, B.H.: clDice – a novel topology-preserving loss function for tubular structure segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16555–16564 (2021). https://doi.org/10.1109/CVPR46437.2021.01629
16. Wong, T.Y., Kamineni, A., Klein, R., Sharrett, A.R., Klein, B.E.K., Siscovick, D.S., Cushman, M., Duncan, B.B.: Quantitative retinal venular caliber and risk of cardiovascular disease in older persons: The cardiovascular health study. Archives of Internal Medicine (2006). https://doi.org/10.1001/archinte.166.21.2388
17. Yakubovskiy, P.: Segmentation models PyTorch. GitHub, https://github.com/qubvel/segmentation_models.pytorch (2020)
18. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical Image Analysis (2019). https://doi.org/10.1016/j.media.2019.101552
19. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3813–3824 (2023). https://doi.org/10.1109/ICCV51070.2023.00355
20. Zhang, Z., Yang, L., Zheng, Y.: Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018). https://doi.org/10.1109/CVPR.2018.00963