Conditioning 3D Diffusion Models with 2D Images:

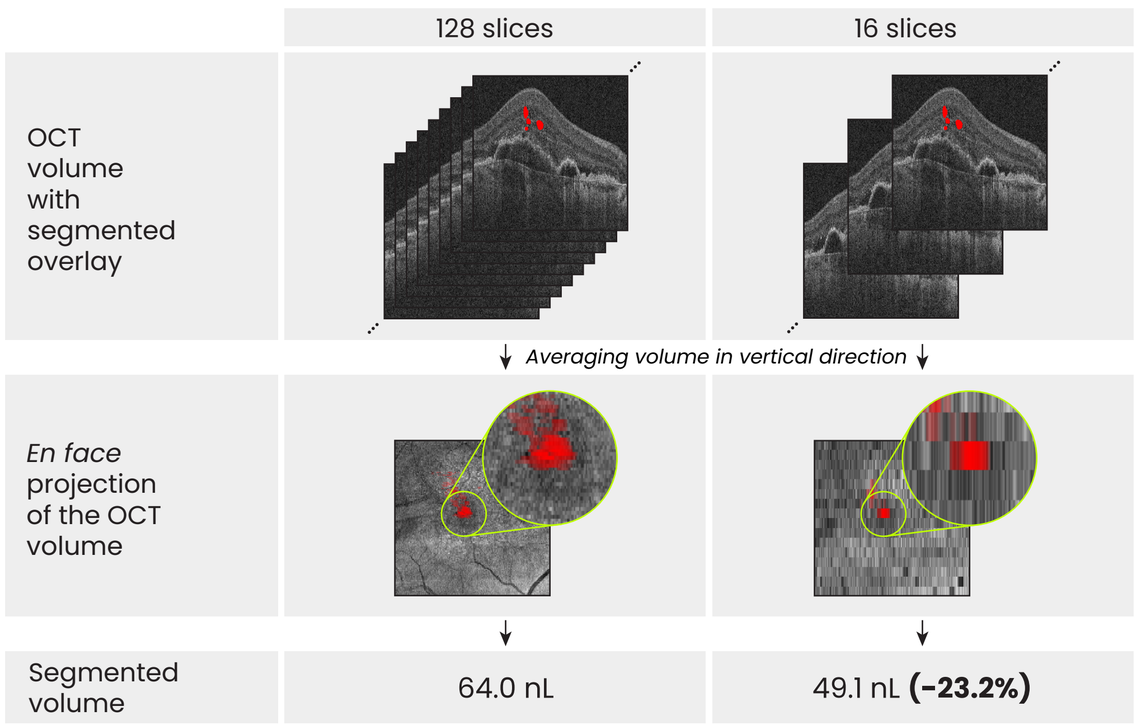
Towards Standardized OCT Volumes through En Face*-Informed Super-Resolution*

Coen de Vente[1,2,3], Mohammad Mohaiminul Islam[1,2], Philippe Valmaggia[4,5], Carel Hoyng[6], Adnan Tufail[7], Clara I. Sánchez[1,2] on behalf of the MACUSTAR consortium

[1]qurAI Group, Informatics Institute, University of Amsterdam, The Netherlands; [2]Amsterdam UMC location University of Amsterdam, Biomedical Engineering and Physics, The Netherlands; [3]DIAG, Department of Radiology and Nuclear Medicine, Radboudumc, The Netherlands; [4]Department of Biomedical Engineering, Universität Basel, Switzerland; [5]Department of Ophthalmology, University Hospital Basel, Switzerland; [6]Department of Ophthalmology, Radboudumc, Nijmegen, The Netherlands; [7]Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom.

Motivation

- Volumetric medical images commonly suffer from **high anisotropy** (*i.e.*, having different resolutions in different directions). For example, in optical coherence tomography (OCT), **slice spacing varies substantially**.
- This can result in inaccuracies in shape measurements or quantifications of biological objects of interest:



- **Aim:** Artifically upsample #slices in volumetric data.

Methods

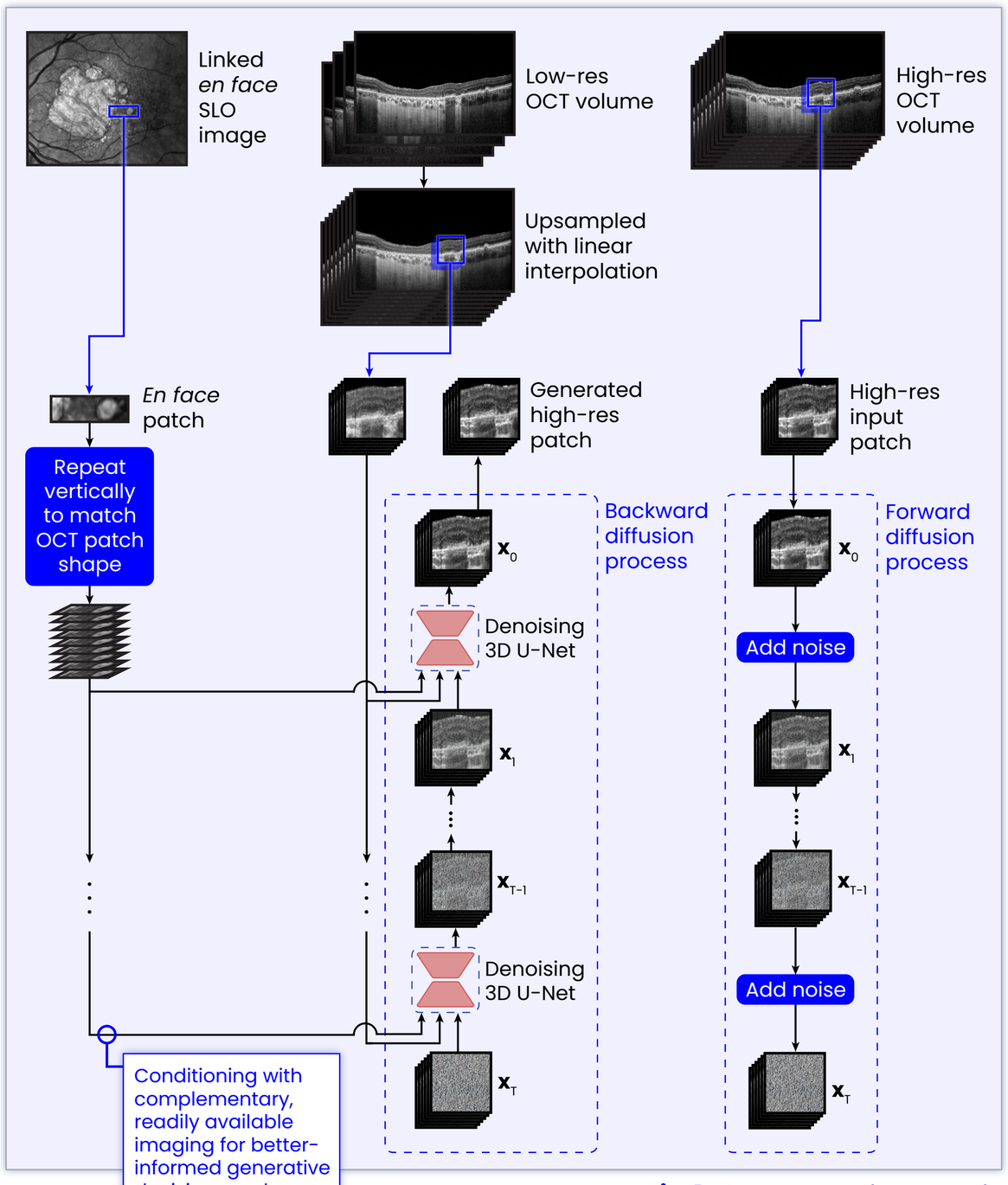


Fig. 1: Our proposed approach.

Experiments

- **Dataset:** MACUSTAR, a European multicenter study.
- **Training:** Patch size of 128 × 128 × 16.
- **Sampling:** Patch size of 496 × 496 × 16, using DDIM & RePaint[1], upsampling volumes from 30 to 240 slices.

	Tricubic	DDIM	$DDIM_{ef}$ (no CFG)	$DDIM_{ef}$
MSE ↓	**0.006** ± 0.002	**0.006** ± 0.003	**0.006** ± 0.003	**0.006** ± 0.003
SSIM ↑	**0.451** ± 0.116	0.444 ± 0.107	0.447 ± 0.107	0.447 ± 0.107
PSNR (dB) ↓	22.472 ± 1.418	22.401 ± 1.644	**22.495** ± 1.673	22.450 ± 1.683
$LPIPS_{axi}$ ↓	**0.120** ± 0.027	0.138 ± 0.030	0.138 ± 0.030	0.141 ± 0.031
$LPIPS_{cor}$ ↓	0.548 ± 0.103	**0.158** ± 0.047	**0.158** ± 0.048	0.162 ± 0.050
$LPIPS_{sag}$ ↓	0.540 ± 0.088	**0.144** ± 0.049	**0.144** ± 0.049	0.147 ± 0.050
$LPIPS_{2.5D}$ ↓	0.403 ± 0.072	**0.147** ± 0.041	**0.147** ± 0.042	0.150 ± 0.043
$LPIPS_{efproj}$ ↓	0.231 ± 0.055	0.063 ± 0.039	**0.060** ± 0.039	0.064 ± 0.039

Table I: Classical image similarity metrics (MSE, SSIM, and PSNR) and perceptual metrics (all LPIPS variants). Results show mean ± std. dev.

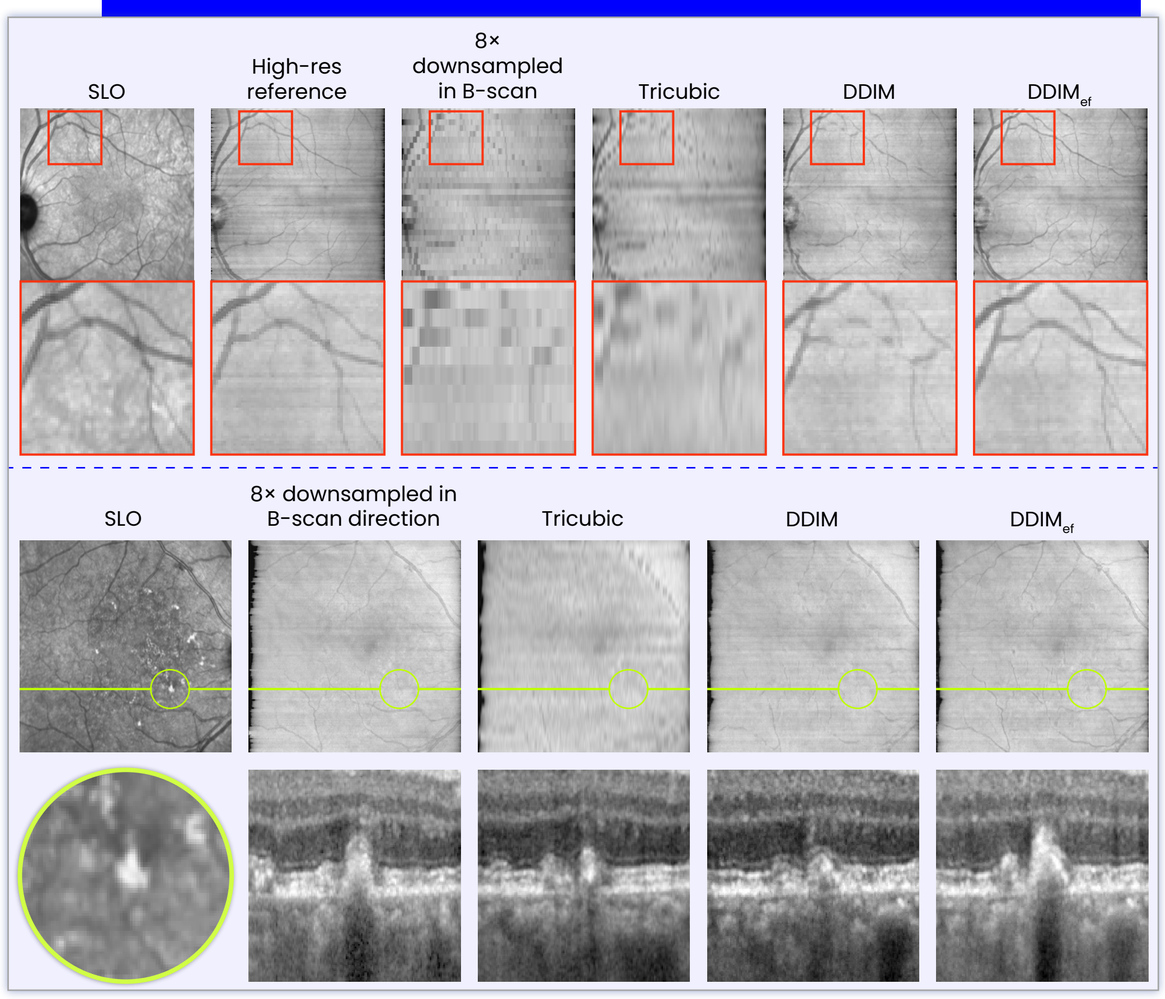


Fig. 2: Generated and baseline examples.

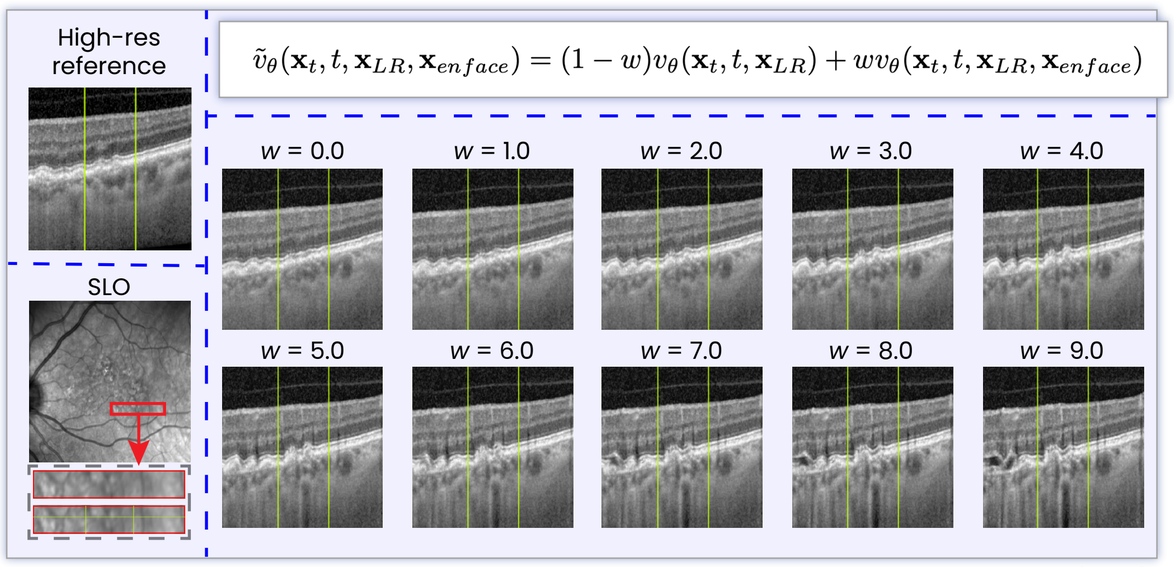


Fig. 3: Effect of classifier-free guidance (CFG).

Conclusion

- Conditioning 3D diffusion models with **complementary, readily available 2D imaging** data results in **improved super-resolution**, especially in terms of perceptual metrics and image sharpness.
- This could be an important step **towards standardized and high quality medical imaging**.

c.w.devente@uva.nl
coendevente.com

References
[1] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Van Gool, L. (2022). Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11461-11471).