PHOTO: DARCY PETERKA

OCTDiff: Bridged Diffusion Model for Portable OCT Super-Resolution and Enhancement

Ye Tian[1]; Angela McCarthy[3]; Gabriel Gomide[3]; Nancy Liddle[2]; Jedrzej Golebka[3]; Royce W.S. Chen[3]; Jeffrey M. Liebmann[3]; Kaveri A. Thakoor[1,2,3]

Columbia University Department of [1]Biomedical Engineering, [2]Computer Science and Columbia University Irving Medical Center Department of [3]Ophthalmology; New York, NY

12/2025

0. Outline of Presentation

- Background: Portable OCT Images
- Algorithm: OCTDiff:
  - – Adaptive Noise Aggregation (ANA)
  - – Multiscale Cross Attention (MSCA)
  - – Loss Function with Clinical Quality Score
- Results: Quantitative and Qualitative
- Conclusions, Implications and Future Directions

1. Introduction - Portable Optical Coherence Tomography (pOCT)

Imaging Depth	2.9mm
Center Wavelength	840nm
Scan Range	8mm × 8 mm

Zeiss Cirrus HD-OCT 5000 (high-res OCT) @CUMC

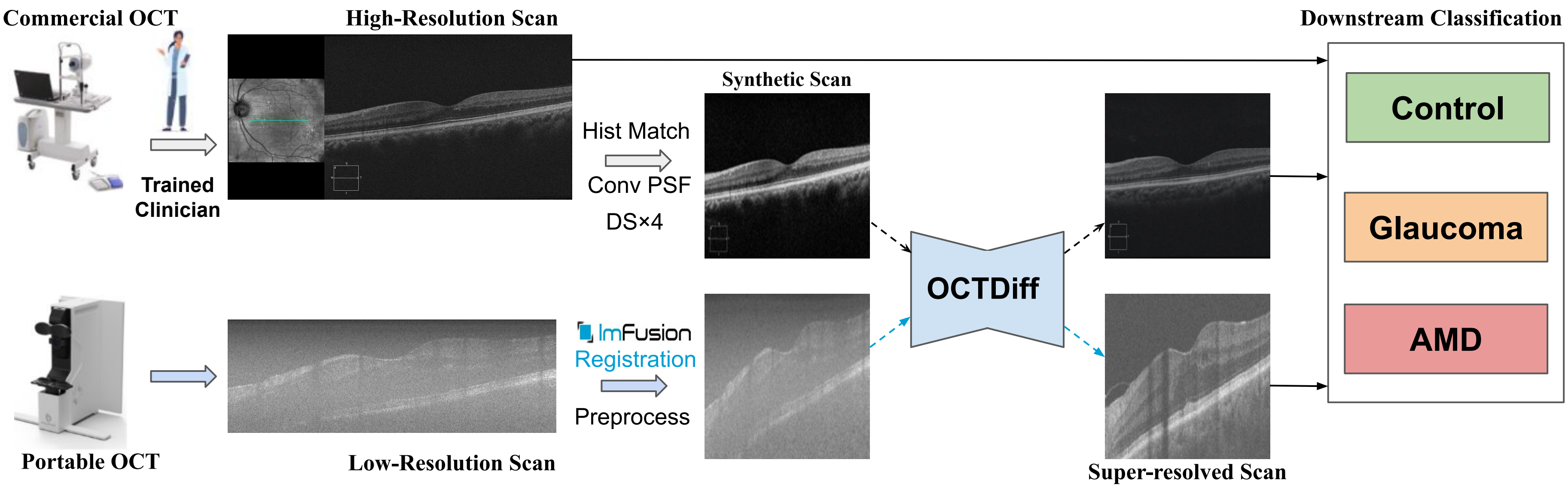
- ~$40K
- Fixed in Exam Room
- Expert only

Imaging Depth	2mm
Center Wavelength	840nm
Scan Range	6.5 × 6.5 mm

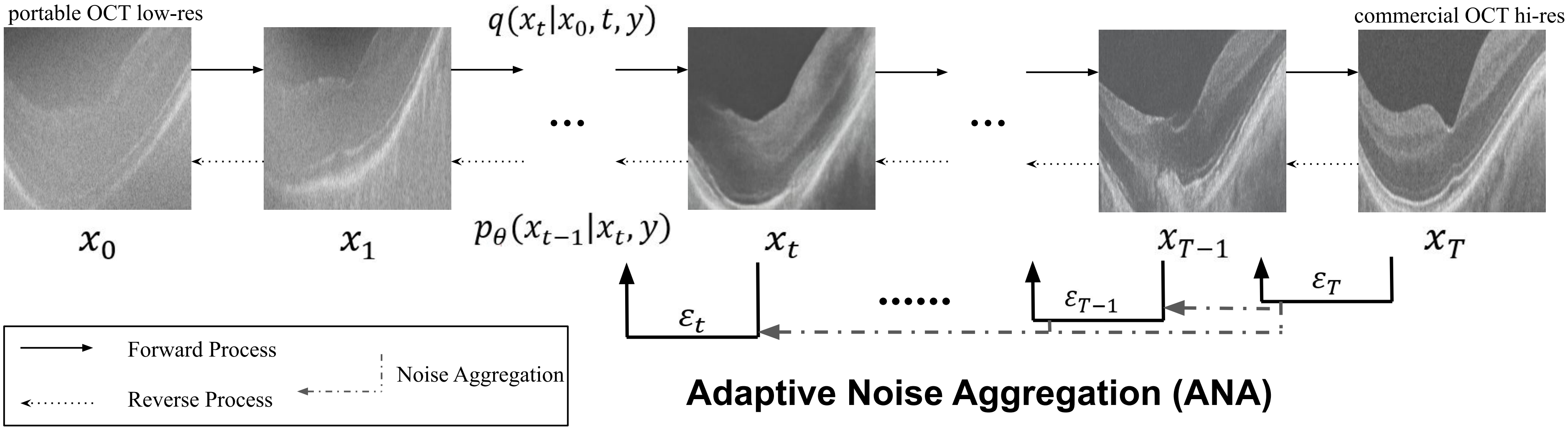
Philophos KUOS 100, portable OCT

- ~$10K
- Carried on Flight
- Novice with User Manual

2. Method


Commercial OCT
Trained
Clinician
High-Resolution Scan
Hist Match
Conv PSF
DS×4
Synthetic Scan
OCTDiff
Downstream Classification
Control
Glaucoma
AMD
Portable OCT
Low-Resolution Scan
ImFusion
Registration
Preprocess
Super-resolved Scan

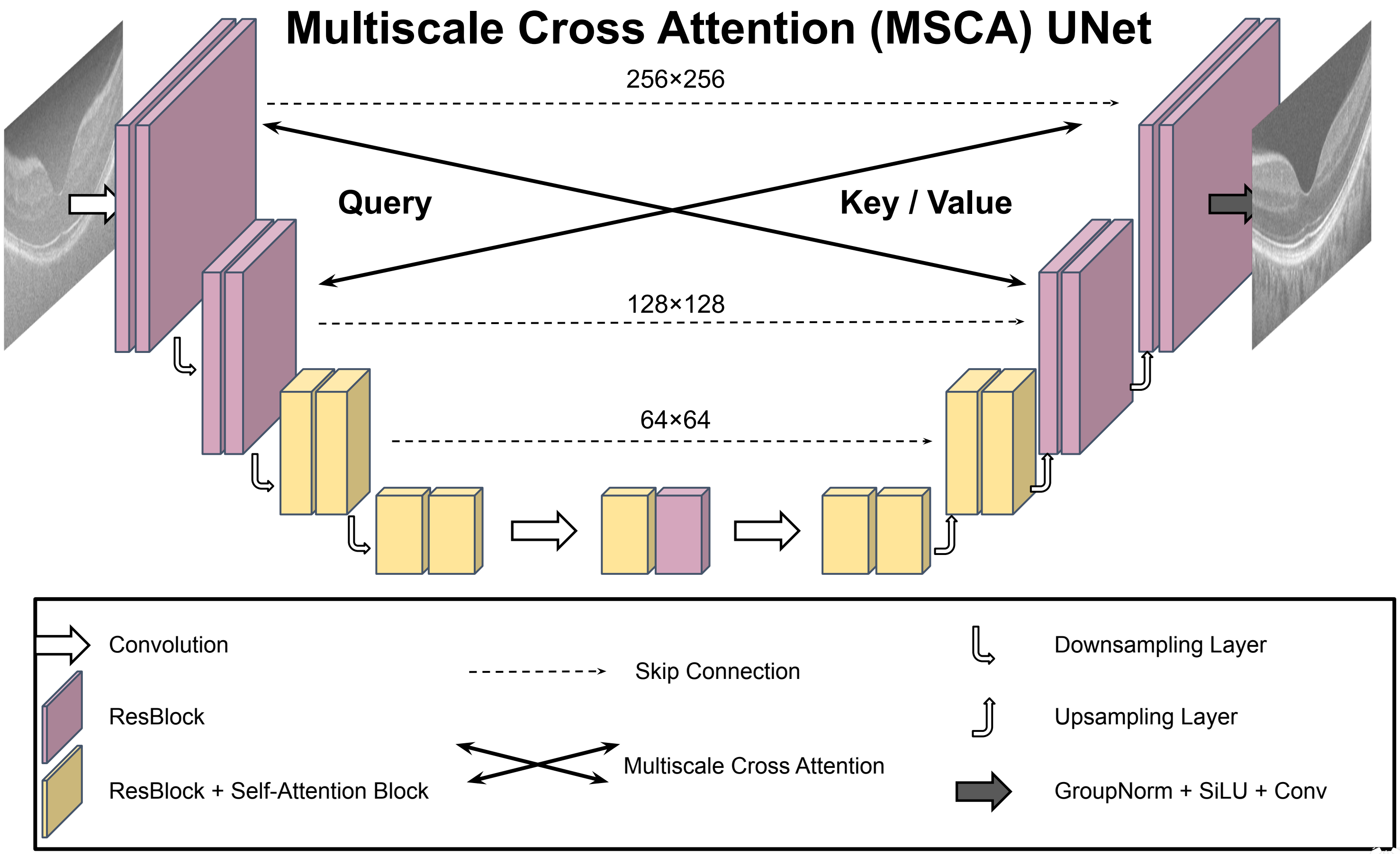
2. Method - Adaptive Noise Aggregation (ANA)



Adaptive Noise Aggregation (ANA)

$$\bar{\varepsilon}_t = \frac{1}{Z_t} \sum_{\tau=t}^{T-1} \exp(-\alpha(\tau - t)) \cdot \hat{\varepsilon}_T$$

2. Method

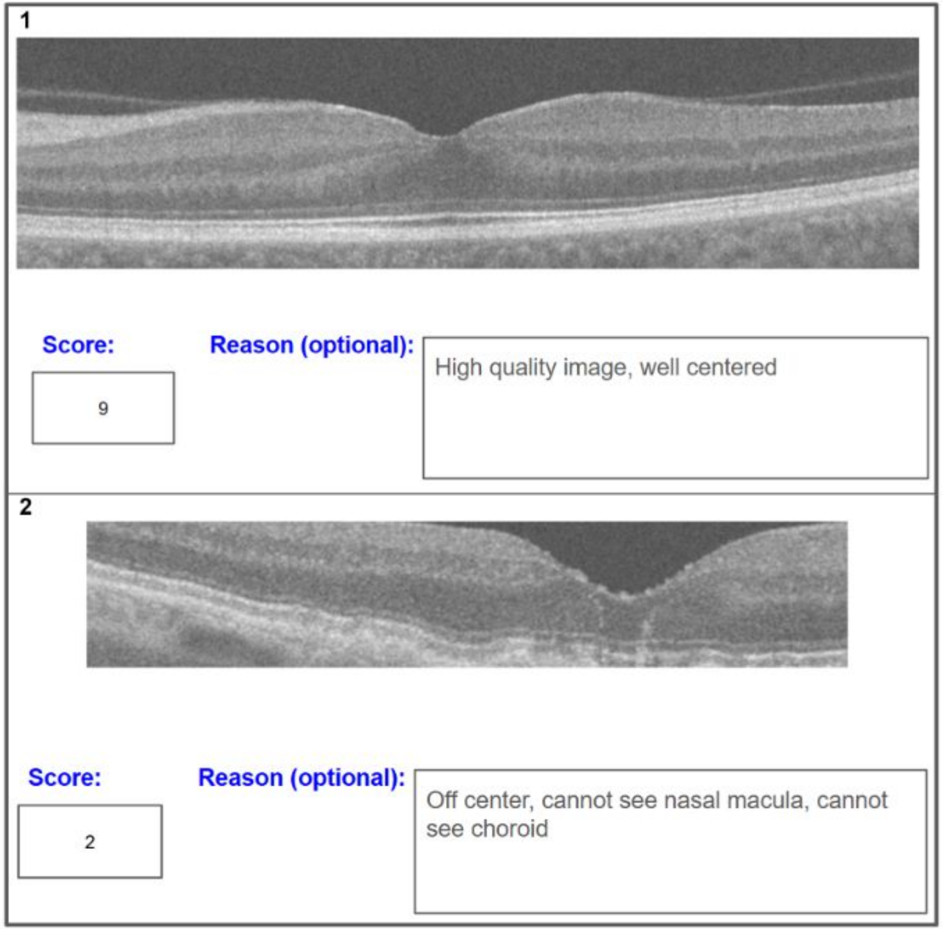


Multiscale Cross Attention (MSCA) UNet
256×256
Query
Key / Value
128×128
64×64
Convolution
Skip Connection
Downsampling Layer
ResBlock
Upsampling Layer
Multiscale Cross Attention
ResBlock + Self-Attention Block
GroupNorm + SiLU + Conv

2. Method

Loss Function with Clinical Quality Score

$$\mathrm{MSE}_{\mathrm{focal}} = \frac{1}{N}\sum_{i=1}^{N}\left(1 - S_{\mathrm{quality}}^{(i)}\right)^{\gamma} \cdot (x_i - \hat{x}_i)^2$$



3. Results

Table 1: Quantitative comparison of models on Philophos 84 and Synthetic 500 datasets. Arrows indicate the desirable direction for each metric.

Model	Philophos 84 Dataset			Synthetic 500 Dataset[1]		
	SSIM% ↑	PSNR ↑	LPIPS% ↓	SSIM% ↑	PSNR ↑	LPIPS% ↓
SRCNN	39.7	18.4	49.3	91.7	28.2	9.21
VDSR	26.1	17.9	47.7	85.2	33.0	12.7
CycleGAN	58.4	29.2	28.3	95.3	30.3	17.9
Swin2SR	78.8	35.9	18.3	96.9	38.2	4.7
CDM	71.9	33.2	31.5	98.6	34.2	14.3
BBDM	87.2	35.3	27.9	98.1	**42.7**	6.2
OCTDiff(ours)	**93.6**	**38.8**	**16.1**	**98.9**	41.0	**1.7**

3. Results

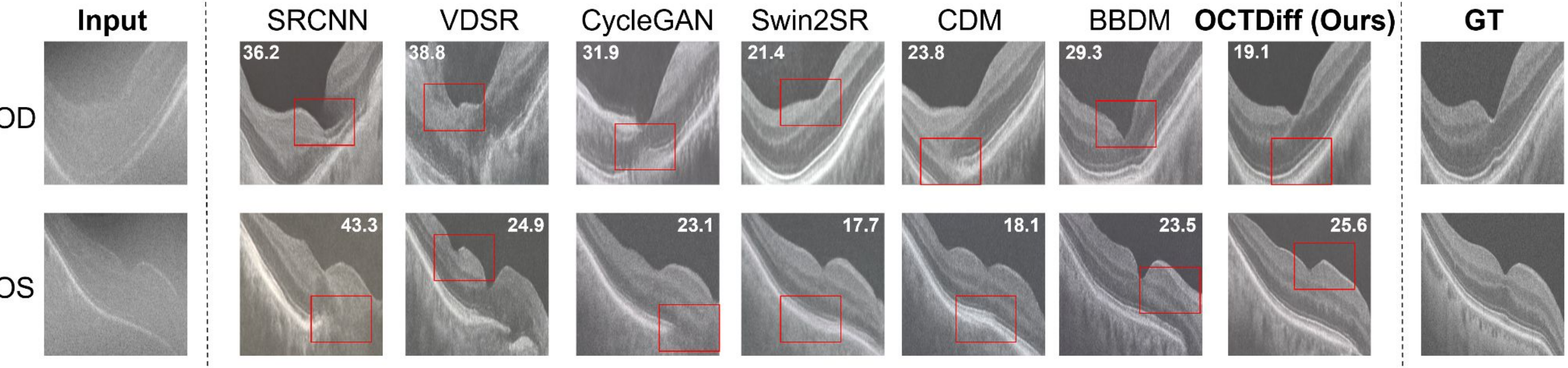


Figure 4: Two examples of reconstructed images from baseline models and our proposed OCTDiff. The first row shows a right-eye (OD) scan, and the second row shows a left-eye (OS) scan. The first and last columns correspond to the input low-resolution image and the ground truth (GT) high-resolution image, respectively. Baseline methods include SRCNN [38], VDSR [39], CycleGAN [40], Swin2SR [41], CDM[28] and BBDM[11]. Red boxes highlight regions where notable degradation or artifacts compared to GT. Each image is annotated with its BRISQUE score [42] to quantify perceptual quality.

3. Results - Ablation Study

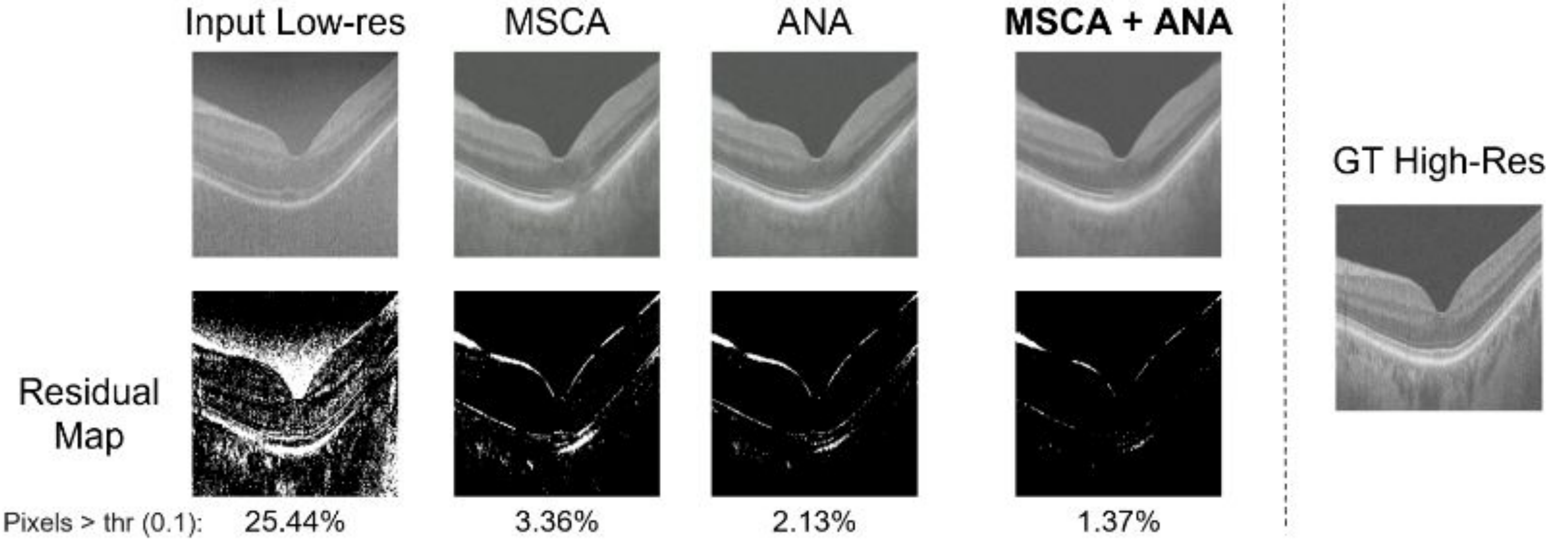


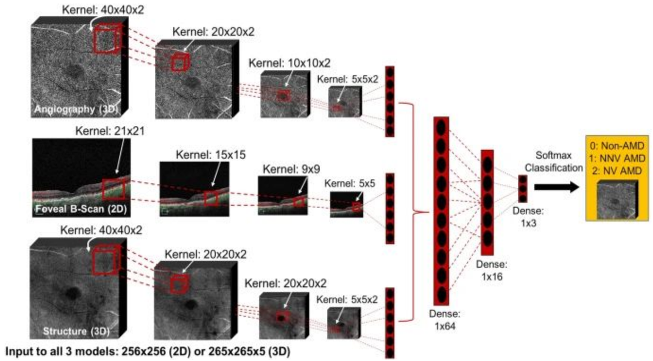
Figure 5: An example of output image when toggling MSCA and ANA. The second row are the corresponding residual maps of the images at top compare with GT on the right. At bottom, the ratio of pixels that have a difference over 10% are shown.

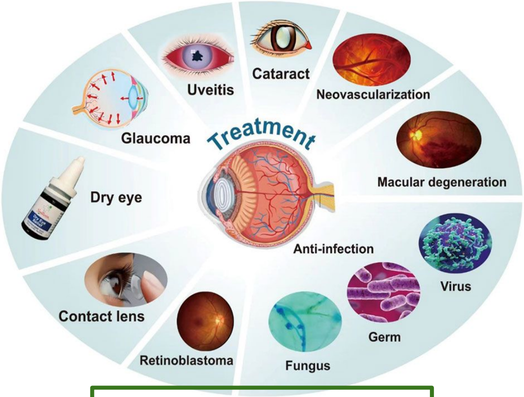
3. Result - Downstream Classification

- We trained three classification architectures: **ViT**, vanilla **CNN**, and **SwinT**.
- This evaluation was conducted independently for two representative ophthalmic tasks: **glaucoma** diagnosis and age-related macular degeneration (**AMD**) classification.

Disease Class	Model	High Res.	Low Res.	OCTDiff	BBDM
Glaucoma	ViT	74.4 ±1.9	50.1 ±1.8*	74.5 ±3.1	62.0 ±2.5*
	CNN2D	93.4 ±5.0	83.3 ±2.2*	93.5 ±0.5	81.0 ±1.0*
	SwinT	75.5 ±2.6*	49.5 ±3.1*	55.6 ±4.3	57.1 ±3.2
AMD	ViT	86.5 ±1.9	48.9 ±2.7*	85.8 ±1.6	79.5 ±2.1
	CNN2D	94.6 ±0.9*	83.0 ±1.4*	96.4 ±0.4	91.8 ±0.7*
	SwinT	82.3 ±1.4*	49.4 ±2.4*	66.3 ±1.8	64.9 ±2.6

4. Discussion and Future Directions





Data
(Fundus/**OCT**/VF/
Report/EyeTrack)

AI
(ML/**DL**)

Application
(Diseases/Tasks)

AI-Advanced Healthcare
Early Detection / Remote Diagnosis

4. Discussion and Future Directions

(1) **latent space** method and make image features as input

(2) test our model on **natural SR datasets** with similar image orientation

(3) actively extending training data size especially using **different low-cost OCTs** other than Philophos such that our model can meet the industrial standard

(4) physically embedding OCTDiff into low-cost OCT via **NVIDIA Jetson Orin Nano** GPU for point-of-care image-capture and real-time diagnosis.

(5) post-hoc evaluation by ophthalmologists are also essential to ensure there are no **hallucinations** to enable safe clinical use.

Thank you!

Ye Tian
Michael Lau
Angela McCarthy
Gabriel Gomide
Nancy Liddle
Jedrzej Golebka
Vish Prathikanti
Dr. Royce W.S. Chen
Dr. Jeffrey M. Liebmann
Dr. Kaveri A. Thakoor

Artificial Intelligence for Vision Science (AI4VS) Lab

Ye Tian (yt2793@columbia.edu)

Research to Prevent Blindness

Department of Biomedical Engineering
COLUMBIA | ENGINEERING
20 Years of Excellence
2000–2020

COLUMBIA ENGINEERING
Biomedical Engineering

CS@CU COMPUTER SCIENCE

COLUMBIA UNIVERSITY
DATA SCIENCE INSTITUTE

COLUMBIA | OPHTHALMOLOGY