

<https://doi.org/10.1038/s44387-025-00038-2>

# Artificial intelligence-assisted retinal imaging enables dense pixel sampling from sparse measurements

**Vineeta Das, Andrew J. Bower, Nancy Aguilera, Joanne Li & Johnny Tam**

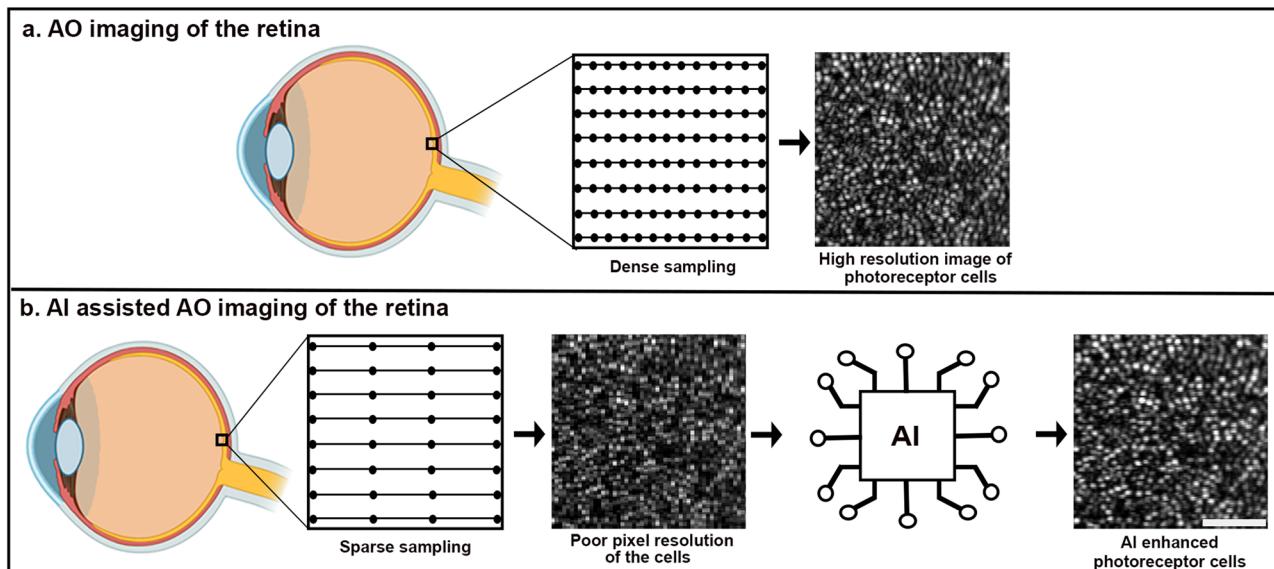
High resolution *in vivo* adaptive optics (AO) imaging has facilitated cellular level assessment of microscopic cone photoreceptors. However, the necessity for dense pixel sampling for good pixel resolution imposes a tradeoff with acquisition speed, leading to motion artifacts and extensive data generation. We introduce an artificial intelligence (AI) assisted imaging framework utilizing residual in residual transformer generative adversarial network (RRTGAN), an AI method that works alongside AO imaging to restore the pixel resolution of sparsely sampled images, circumventing the need for dense sampling. Our results show that RRTGAN can enable data-efficient imaging, restoring high-quality images from just one-fourth of the data and closely matching ground truth images. Cone spacing estimates across four participants aligned well with histology at various retinal locations. These results demonstrate AI assisted imaging's potential to overcome pixel sampling and imaging speed tradeoff, an important step toward improving the efficiency of routine AO imaging in the clinic.

Advances in biomedical optics have enabled cellular-level visualization of the living human retina<sup>1–4</sup>. For point scanning imaging systems such as confocal microscopes, there is a tradeoff between increasing the density of pixel sampling for improved resolution vs. decreasing pixel sampling for improved acquisition speed. In the case of ophthalmic imaging instruments such as optical coherence tomography (OCT), dense sampling is essential to ensure that the retina is captured at a pixel resolution sufficient for visualizing the microscopic cellular structures. However, dense sampling lengthens the time needed for each frame/volume to be acquired, which in turn exacerbates motion artifacts and image distortions due to continuous eye movements that occur during imaging.

Adaptive optics (AO) is a technology that can be used to sense and correct optical aberrations in the eye, enabling high resolution *in vivo* visualization of retinal cells<sup>5</sup>. The integration of AO with OCT has facilitated cellular-level, three-dimensional visualization of microscopic retinal cells<sup>6–8</sup>. Higher imaging speed is desirable from adaptive optics optical coherence tomography (AOOCT) systems to more rapidly acquire images and minimize motion artifacts, but this is often at the expense of insufficient pixel sampling depending on the cell being imaged. Although the introduction of high speed swept source lasers such as Fourier domain mode-locked (FDML) lasers<sup>9</sup> have improved scan rates<sup>10,11</sup>, acquisition speed still remains a challenge. This limitation is exacerbated by the small field of view (FOV) inherent to AO systems ( $\sim 0.5$  mm  $\times$   $\sim 0.5$  mm), necessitating multiple overlapping acquisitions followed by image processing and montaging to visualize larger areas of the retina. Due to the increased pixel sampling

required to image smaller cells such as cone photoreceptors, large amounts of data (on the order of terabytes) are typically generated during the acquisition, which can be resource intensive and inefficient. An alternate approach to increasing data efficiency and optimizing the imaging time is to decrease the pixel sampling (i.e. sparsely sample the retina). However, this may lead to sub-optimal image quality resulting from diminished pixel resolution.

In recent years, artificial intelligence (AI) based image enhancement methods have been applied to advanced optical microscopy techniques, including the application of AI for image super resolution<sup>12–15</sup>, reconstruction<sup>16</sup>, denoising<sup>17,18</sup>, and image translation<sup>19–21</sup>. Among the various AI methods, generative adversarial networks (GANs)<sup>22</sup>, which employ two neural networks (generator and discriminator) in an adversarial fashion, have been particularly effective for image enhancement in medical images such as improving image quality<sup>23–25</sup>, denoising OCT images<sup>26,27</sup>, color normalization and virtual staining in histopathology images<sup>28</sup>, and image synthesis in radiology<sup>29</sup>. GANs are typically comprised of a convolutional neural network (CNN)<sup>30</sup> to represent local characteristics of images. However, this local processing does not take into consideration the structural relationships across larger regions in the images. Increasingly, transformer networks<sup>31</sup> have received widespread attention in computer vision due to their ability to capture global features in images<sup>15,32–34</sup>. We hypothesize that the structural correlations present in the layered arrangement of the retinal cells can be better represented using the non-local attention mechanism offered by the transformer. Further, incorporating



**Fig. 1 | Artificial intelligence (AI) assisted imaging uses sparse sampling to recover cone visualization in the human retina.** **a** AAOCT imaging based on densely sampling provides high quality images of the cone photoreceptor cells at the expense of large data sizes and long imaging time. **b** AI-assisted AAOCT imaging sparsely samples

the retina and uses AI to restore the pixel resolution of the photoreceptor cells. Incorporating AI into the imaging framework overcomes the tradeoff between pixel sampling and acquisition time, enabling data and time efficient imaging of the retina. The bright dots in the retinal images are individual cone photoreceptor cells. Scale bar: 50  $\mu$ m.

transformer into the GAN framework helps to ensure faithful restoration of the retinal images that maintain structural consistency of the cells. In this paper, we combine the merits of transformer into GAN to develop an AI-assisted imaging framework where AI works alongside imaging to restore pixel sampling from sub-optimally or sparsely sampled AAOCT images.

For AO retinal imaging, sparse sampling is not typically preferred as it could result in the inability to resolve cells. In this paper, we demonstrate that with the help of AI, the pixel sampling of the sparsely sampled images can be restored to the level of the densely sampled images, making sparse sampling a viable option for more efficient imaging. Our AI method effectively restores pixel sampling from only 1/4<sup>th</sup> of the acquired data while ensuring reliable visualization of the retinal cells. Overall, the AI-assisted AAOCT imaging strategy circumvents the tradeoff between image degradation due to reduced pixel sampling and acquisition time, overcomes speed limitations of imaging systems due to hardware, and opens up the possibility for an alternate strategy to enhance and accelerate imaging for point-scanning optical systems.

## Results

### AI-assisted imaging achieves dense pixel resolution with only a fraction of the data

The overall goal is to integrate AI into the AAOCT imaging framework, where instead of performing dense sampling of the retina to visualize the cells (Fig. 1a), we explore the possibility of performing sparse sampling with only 1/4<sup>th</sup> of the data required for dense sampling and then use AI to improve the visualization of the poorly sampled cells (Fig. 1b). Our AI algorithm, residual in residual transformer GAN (RRTGAN), incorporates a modified generator based on transformers that are connected in a residual fashion to effectively model the correlations of the different retinal layers captured using the AAOCT to restore the visualization of the individual photoreceptor cells (Supplementary Fig. 1, Supplementary Methods).

The sparsely sampled images showed a pixelated appearance of the cone photoreceptor cells (Fig. 2a, f). RRTGAN was successful in improving the pixel resolution and enhancing the visualization of cells from sparsely sampled AAOCT volumes (Fig. 2) acquired from study participants (Supplementary Table 1). RRTGAN enhanced and ground truth (densely sampled) cone photoreceptor images showed similarity in cellular structure of the cones (Supplementary Video 1). RRTGAN also performed better than other competitive AI frameworks (ESRGAN and SwinIR) (Fig. 2). Qualitative visual comparison of the images showed clearer and sharper visualization of

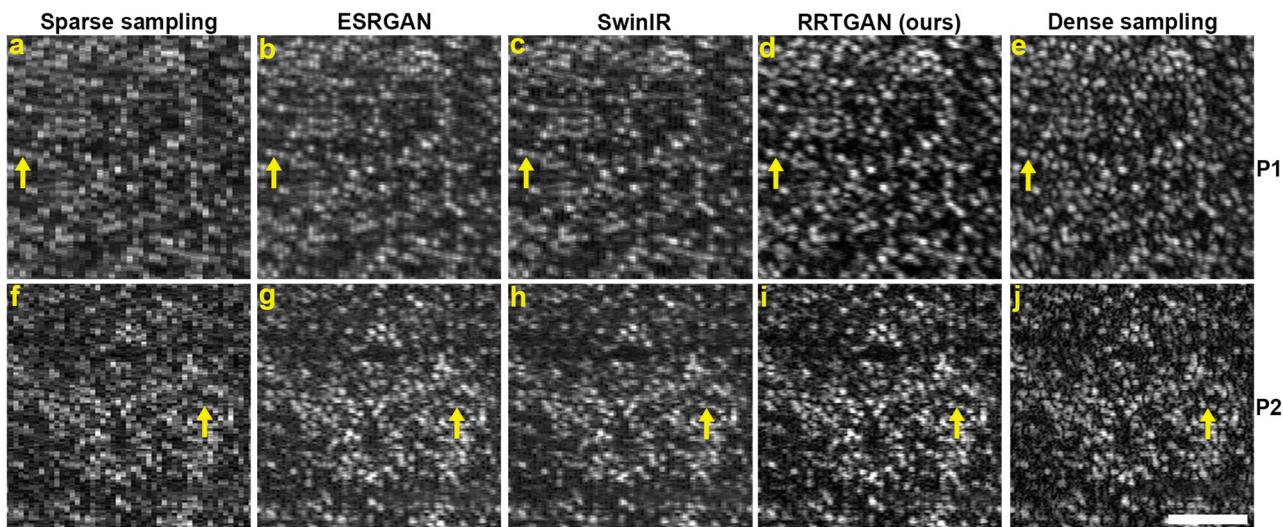
cones using RRTGAN compared to ESRGAN<sup>14</sup> and SwinIR<sup>15</sup>. This is likely due to the self-attention mechanism of the dense transformer blocks enabling the learning of non-local features and the skip connections bypassing abundant low frequency information, allowing the generator to focus on learning high frequency information and aiding in accurate restoration of the structure and shape of the cones in the enhanced images.

Quantitative comparison among the methods using objective image quality assessment metrics (peak signal to noise ratio (PSNR), deep image structure and texture similarity (DISTS)<sup>35</sup>, and learned perceptual image patch similarity (LPIPS)<sup>36</sup>) further corroborated our findings on the performance of RRTGAN (Supplementary Fig. 2) across study participants (Supplementary Table 1). For all three metrics, the improvement in RRTGAN performance was statistically significant ( $p < 0.05$ ). RRTGAN images exhibited the highest PSNR indicating better pixel level similarity with the optimally sampled ground truth images. There was an average reduction of 29% in DISTS and 31% in LPIPS for RRTGAN compared to other networks, indicating better perceptual similarity of RRTGAN enhanced and ground truth images. Likewise, RRTGAN also demonstrated the lowest Fréchet Inception Distance (FID)<sup>37</sup> score of 85.8 compared to ESRGAN (104.4) and SwinIR (127.8) indicating that the enhanced images were more similar to the ground truth images compared to other methods. Overall, our results indicate that RRTGAN outperformed existing AI methods and could successfully enhance the visualization of the cone photoreceptors from sparsely sampled AAOCT.

### AI restores the shape of the cones from sparsely sampled images

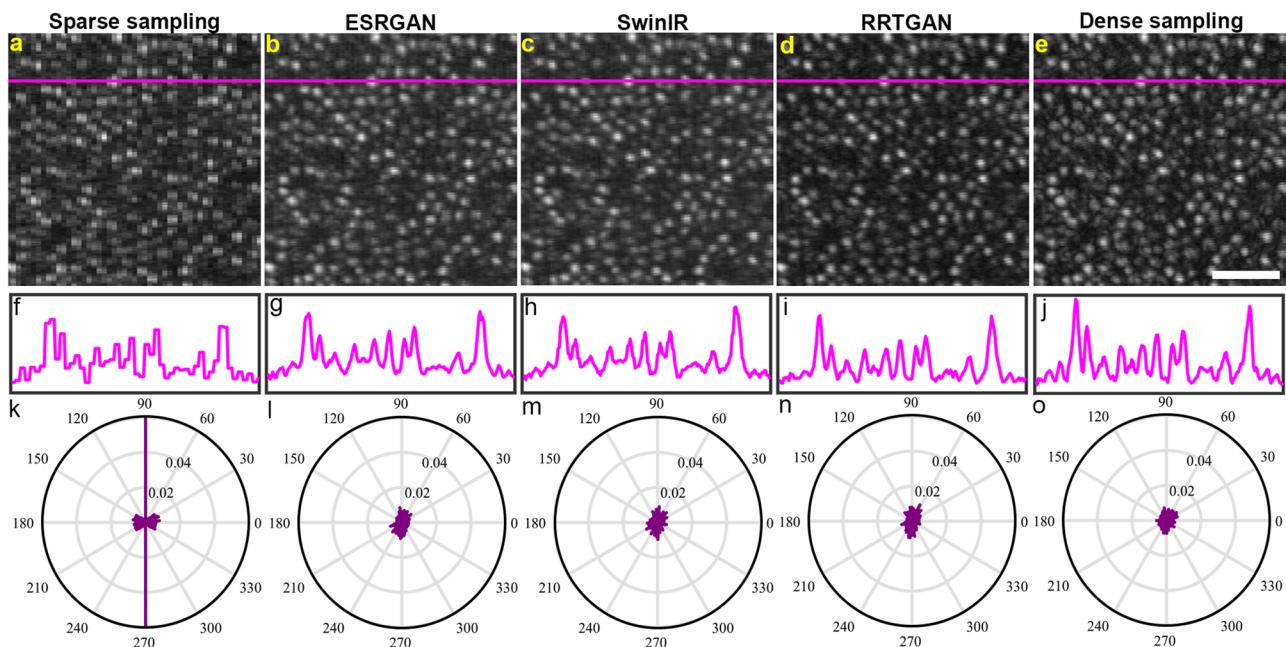
Having demonstrated the efficacy of RRTGAN, we also wanted to evaluate the structural integrity of cones in the enhanced images. Visual inspection of the sparsely sampled images displayed a pixelated appearance (Fig. 3a) of the cones. It was interesting to observe that the individual cone cells did not have pixelation in all AI-enhanced images (Fig. 3b-d). Line intensity profiles through the images showed square peaks at the locations of the cones in the sparsely sampled images (Fig. 3f), confirming that the individual cone cells appear pixelated due to reduced pixel sampling. Sharper peaks corresponding to the cones were observed for the cones in the AI-enhanced and ground truth images (Fig. 3g-j).

Using normalized histograms to measure the directionality of the edges (Methods, Edge Directionality) in the sparsely sampled images also revealed



**Fig. 2 | Residual in residual transformer generative adversarial network (RRTGAN) restores the pixel resolution of sparsely sampled images to match dense sampling.** **a, f** Sparsely sampled images of the cone photoreceptors from two participants (P1 and P2) having a pixelated appearance due to reduced pixel resolution. Pixel resolution of the cones enhanced by AI using **(b, g)** enhanced super-

resolution generative adversarial network (ESRGAN), **c, h** SwinIR, and **d, i** RRTGAN (ours). **e, j** Ground truth densely sampled images for visual comparison. The bright dots in the images represent the individual cone cells. The yellow arrows show cells that are more distinguishable in RRTGAN compared to the sparsely sampled, ESRGAN, and SwinIR images. Scale bar: 50  $\mu$ m.



**Fig. 3 | Artificial intelligence (AI) restores shape of the cones from sparsely sampled images.** Pixelated appearance of the cones visualized in **a** sparsely sampled images. The cells have smooth edges in **b** enhanced super resolution generative adversarial network (ESRGAN), **c** SwinIR, **d** residual in residual transformer generative adversarial network (RRTGAN), and **e** ground truth images. **f-j** Line intensity profile of the images across the magenta line in **(a-e)** shows square peaks for the sparsely sampled image and sharp pointy peaks for the AI-enhanced and

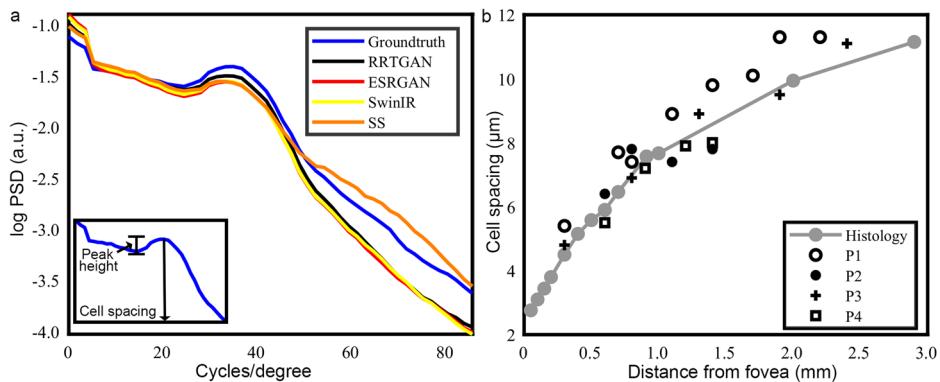
ground truth images. **k** Normalized histogram for the sparsely sampled image exhibits edges concentrated in fixed directions corresponding to the sharp edges, reflecting presence of square pixels and hence the pixelated appearance. **l-o** The normalized histogram of edges in the AI-enhanced and the ground truth images have edges distributed across all angles indicating the cones in these images no longer have sharp corners from the sparse sampling. Scale bar: 50  $\mu$ m.

a higher concentration at the vertical directions, further confirming the presence of sharp edges (Fig. 3k). On the contrary, the AI enhanced and the densely sampled ground truth images had edges uniformly distributed across different angles, corroborating the smoother appearance of the cones seen in the AI enhanced images (Fig. 3l-o). Computing the variance in the edge direction from the normalized histogram showed less variance for the AI methods compared to the sparsely sampled images (Supplementary Table 2). These results reinforce our observation demonstrating the

potential of AI for removing sharp edges and restoring the smooth shape of the microscopic photoreceptor cells.

#### RRTGAN preserves cell spacing and contrast comparable to densely sampled images

In addition to demonstrating the efficacy of RRTGAN in improving both the visualization and circular shape of cones, we also objectively evaluated how well RRTGAN preserves the spatial distribution of the cones in the



**Fig. 4 | Cone cell spacing of the artificial intelligence (AI) enhanced images are comparable to the ground truth densely sampled images. a** Circumferentially averaged power spectral density (PSD) of the sparsely sampled (SS), AI-enhanced and ground truth images of participant P3 imaged at 1.3 mm temporal to the fovea. The visible peak corresponding to cone cell spacing is observed in all the images. The inset shows a zoomed version of the peak with the vertical black line indicating the location of the fundamental spatial frequency associated with the cells. The height of

the peak indicates the cellular contrast. **b** Comparison of cone spacing estimated from residual in residual transformer generative adversarial network (RRTGAN) enhanced images with previously published histology data<sup>39</sup>. Symbols in black indicate cell spacing estimated from RRTGAN-enhanced images for four participants (P1, P2, P3, and P4) at different retinal locations. Cone spacing replotted from histology<sup>39</sup> shown in gray.

images. Examination of the radially averaged power spectral density showed a peak indicating the fundamental spatial frequency (representative of cell spacing) associated with the spatial arrangement of the cones in the images<sup>38</sup>. The peaks for the ground truth densely sampled and AI images were consistent (Fig. 4a and Supplementary Table 3), indicating high fidelity of the RRTGAN enhancement.

Furthermore, the height of the local peak of the radially averaged power spectral density provided an objective way to quantify the cellular contrast. Compared to the sparsely sampled cones, ESRGAN and SwinIR enhanced images, the RRTGAN enhanced images have the highest peak height, indicating the best cellular contrast and visualization of the cells (Table 1). The peak height of RRTGAN and the ground truth images are remarkably close, which agrees with our earlier results, indicating the improved visualization offered by RRTGAN. These results demonstrate RRTGAN's effectiveness in boosting cellular contrast in addition to providing structural and perceptual similarity.

The cell spacing estimated using the circumferentially averaged power spectral density for RRTGAN-enhanced images across all participants were within the expected ranges when compared with previously published normative data (Fig. 4b), confirming that RRTGAN did not create or delete additional cells. These results demonstrate the possibility of using AI to restore pixel sampling from insufficiently sampled data, which could enable AOOCT to more efficiently visualize cells from different retinal locations in the living human eye.

#### AI-assisted imaging generalizes across the retina

Cone distribution across the retina varies significantly, with the highest density concentrated near the fovea, gradually decreasing as the distance from the fovea increases<sup>39</sup>. Additionally, cone cells are also much smaller near the fovea and progressively increase in size with increasing distance from fovea. Having trained and validated the performance of RRTGAN on images collected approximately 1 mm from the fovea, we now evaluate its effectiveness on images from retinal locations that were never seen by RRTGAN during training. We tested the performance of RRTGAN at retinal locations 0.8 mm, 1.4 mm, 1.8 mm, and 2.5 mm temporal to the fovea, and found that RRTGAN successfully recovered the cone cellular structures at all the four locations (Fig. 5). Objective quantification further confirmed the similarity of RRTGAN-enhanced and the ground-truth images (Supplementary Table 4). This result demonstrates that RRTGAN can generalize across different imaging locations and can faithfully recover the cellular structures for never seen cells of different sizes.

Having demonstrated the generalizability of RRTGAN across different retinal locations featuring a range of cell sizes, we also verified the

consistency of the cells by imaging two contiguous retinal locations with overlapping regions. Consistency of cells across regions is essential for stitching together individual images with smaller FOV in order to investigate changes to the retina occurring over wider retinal regions. Two RRTGAN-enhanced images obtained from contiguous locations with ~50% overlap showed that the cells in the overlapping areas are consistent in both images, verifying that the AI-enhanced images preserve the spatial locations of the cones (Supplementary Video 2). By extending this approach across 23 overlapping retinal locations, this feat allowed us to apply RRTGAN to a much larger retinal region spanning across 3 mm temporal to the fovea (Fig. 6). This shows that the AI-enhanced images can be successfully used to build large montages to visualize larger areas of the retina.

Finally, thus far, we showed the RRTGAN effectively enhanced the cone photoreceptor layer (also referred to as the ellipsoid zone layer or the inner segment/outer segment (IS/OS) junction), which is only a portion of the overall 3D AOOCT volume. Since RRTGAN operates on the cross-sectional 2D AOOCT images (B-scans), it naturally operates across all of the neighboring layers that are in focus on the outer retinal AOOCT volumes. We found that RRTGAN was also successful in restoring the S cone layer<sup>40</sup> (cones sensitive to short wavelength light) as well as the cone outer segment tips (COST) layer of the retina (Fig. 7). In addition to enhancing the cone photoreceptors, since RRTGAN enhances data across the entire AOOCT B-scan, it was also effective in restoring the visualization of neighboring layers (Supplementary Fig. 3). This result shows an advantage of the RRTGAN training using cross-sectional 2D AOOCT images (i.e. OCT B-scans), extending its benefits across multiple 3D layers that can be reconstructed to obtain multiple rotated 2D views of the tissue (i.e. OCT en face view).

#### Discussion

We demonstrated that RRTGAN successfully restored the pixel sampling from sparsely sampled AOOCT images. The resulting RRTGAN-enhanced images enabled both the visualization of cone photoreceptor cells across the retina as well as the quantification of structural metrics such as cell-to-cell spacing despite using only one fourth of the imaging data. The enhanced images also successfully restored the 3D cone structure and enabled efficient visualization of the IS/OS junction and COST layers. Because RRTGAN is trained on cross-sectional B-scans rather than en face photoreceptor images, it also enables improved visualization of neighboring layers, including the retinal pigment epithelium (RPE) and choroidal vessels. This approach is also more data efficient as it can be applied to 2D slices and generalized across a 3D volume, enabling more rapid cellular-scale 3D imaging of the eye in a clinical setting.

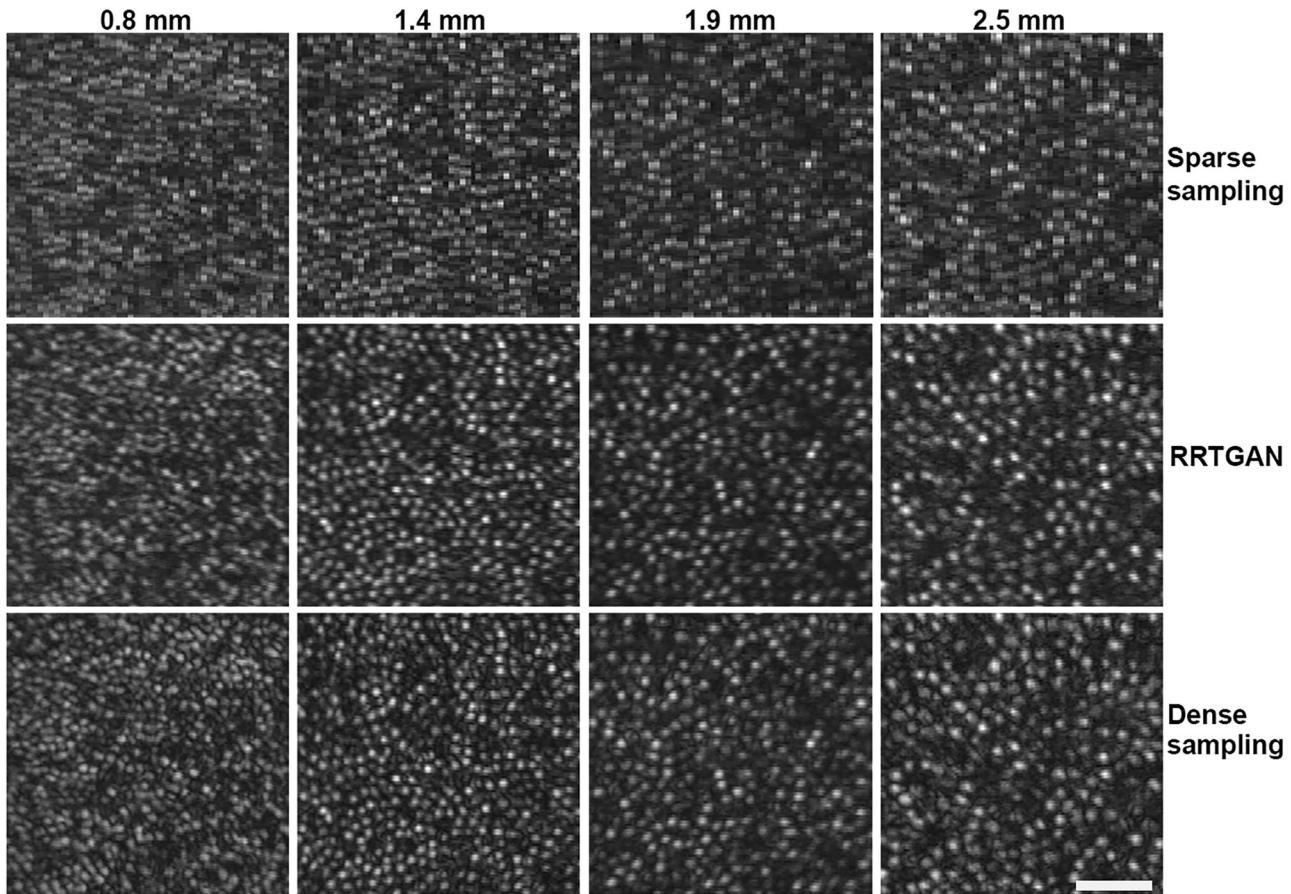
The success of RRTGAN can be attributed to the self-attention mechanism of transformers in effectively learning global correlations from the retinal layers of the AOOCT images together with the ability of the GAN to restore cellular structures. The improvement of RRTGAN over the CNN based ESRGAN<sup>14</sup> indicated that the restoration benefitted from non-local self-similarity modeling offered by the transformers. RRTGAN also generated superior results compared to the state-of-the-art SwinIR framework. This could be attributed to the residual in residual connections in RRTGAN that aided in improving the flow of gradients through the network. In terms of computational complexity, RRTGAN also has fewer parameters than ESRGAN (Supplementary Table 5). It was interesting that although trained on 2D cross-sectional image slices instead of 3D AOOCT volumes, RRTGAN was still able to preserve the 3D structure of the cells for visualization. This further illustrates the consistency of the network in independently enhancing neighboring B scans (one at a time) in a manner that still results in a seamless transition when viewing cells that are split across neighboring B scans.

**Table 1 | Comparison of cellular contrast of the different methods**

Image	Peak height (a. u.)
Sparse sampling	0.06 ± 0.03
ESRGAN	0.09 ± 0.03
SwinIR	0.09 ± 0.03
RRTGAN (ours)	0.12 ± 0.02
Dense sampling	0.11 ± 0.06

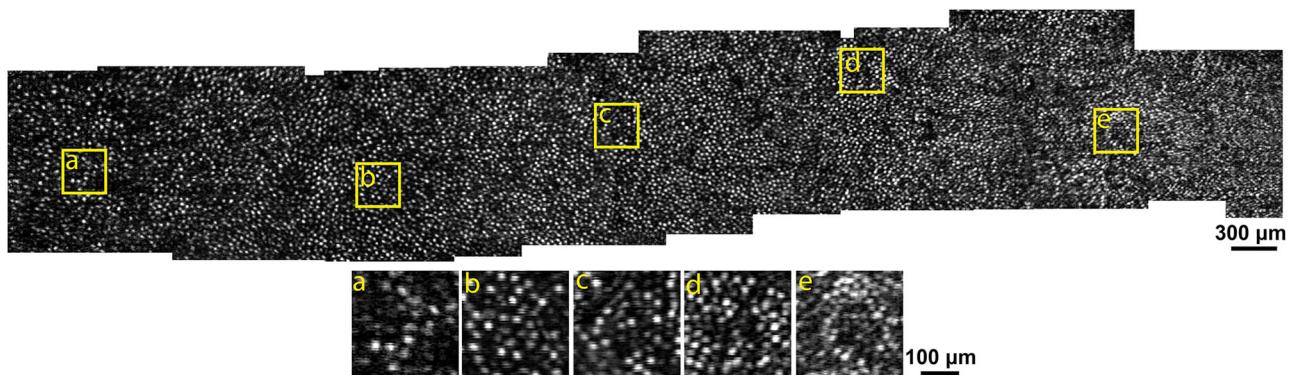
Our validation confirmed that the RRTGAN-enhanced images accurately preserved the cellular structure, which is crucial for assessing retinal health, particularly for performing morphometric measurements from the cells. This integration of AI into the AOOCT image acquisition process represents an approach, diverging from the traditional post-acquisition applications of AI. RRTGAN working alongside AOOCT imaging guarantees data efficiency as the images with 75% fewer samples can be restored to match the ones acquired with 100% sampling. Our AOOCT system can quickly generate terabytes of data across a typical imaging session (3D volumes from 20 to 25 locations per eye, with ~50 GB for 85 volumes acquired at each location). With RRTGAN, the requirement of dense sampling can be relaxed and thereby substantially reducing the data size. Further, the time saving achieved by acquiring less data also allows imaging at more retinal locations in the same amount of imaging time, particularly advantageous in a clinical setting where reducing the time spent per patient is important. This also results in substantial savings in the computational time needed to post-process AOOCT volumes to correct for eye motion<sup>27</sup>. Given the success of RRTGAN in enhancing the visualization of cone cells, we anticipate that such an approach can potentially be translated to other speed-constrained imaging applications such as AOOCT angiography<sup>41</sup>, which requires multiple scans of the same location in quick succession.

In conclusion, we introduced an AI-based RRTGAN framework to improve the visualization of cells from sparsely sampled, rapidly acquired, low pixel resolution images. By adopting an AI-assisted imaging strategy<sup>24,27</sup>, we demonstrated the capability of AI not only in improving the visualization of cells and reducing the time needed to acquire images, but also providing an opportunity to image at higher speeds as hardware solutions improve.



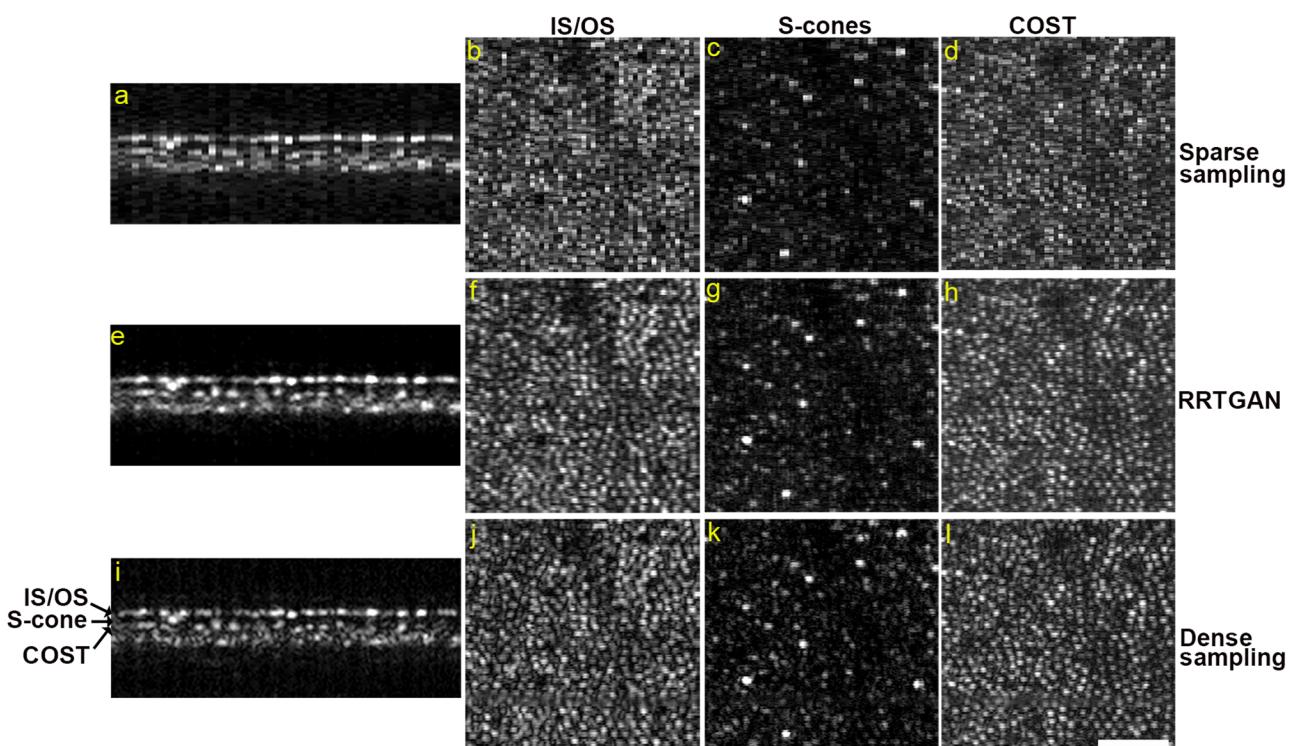
**Fig. 5 | Residual in residual transformer generative adversarial network (RRTGAN) and the ground truth densely sampled images have similar appearance of the cones for experimental data at different retinal locations.** Sparsely sampled, RRTGAN enhanced, and ground truth images are similar across all retinal

locations (0.8, 1.4, 1.9, and 2.5 mm temporal to the fovea). Images are shown from participant P1 (not used for model training). RRTGAN is able to effectively enhance images with varying cell size and spacing across different retinal locations and restores images to resemble cells seen in ground truth images. Scale bar: 50  $\mu$ m.



**Fig. 6 | Residual in residual transformer generative adversarial network (RRTGAN) enables efficient large-scale visualization of cone photoreceptors across the retina.** The image shows the visualization of the cone photoreceptors using the RRTGAN-enhanced images. The montage is constructed by stitching

together more than 20 overlapping images from the right eye of participant P1. The yellow squares (a–e) indicate regions that have been further zoomed for better visualization at retinal locations a 2.6, b 2.1, c 1.7, d 1.2, and e 0.8 mm temporal to the fovea.



**Fig. 7 | Residual in residual transformer generative adversarial network (RRTGAN) restores the 3D structure of cones.** a, e, i Cross-sectional adaptive optics optical coherence tomography (AOOCT) B-scans of the sparsely sampled, RRTGAN-enhanced, and ground truth densely sampled retina of participant P4 at 1.2 mm temporal to the fovea. b Top-down view of the inner segment/outer segment

(IS/OS) junction that appear as bright dots are hard to distinguish visually in the sparsely sampled images and have better visualization in (f) RRTGAN enhanced and (j) ground truth densely sampled images. Similar observations are seen for (c, g, k) S-cones and (d, h, l) cone outer segment tips. The arrows in (i) highlight the IS/OS junction, S-cones, and COST layers of the retina in the B-scan. Scale bar: 50 μm.

Addressing both the technical limitations of AOOCT systems and the practical burden on patients and clinicians is essential for broader clinical adoption. Our findings highlight the transformative potential of AI in retinal imaging and contribute meaningfully to the first step towards translation of AOOCT into routine ophthalmic practice.

Ireland). Light was raster scanned in the horizontal direction using a 3.3 kHz resonant scanner (SC-30, Electro-Optical Products Corp, Fresh Meadows, NY, USA) and in the vertical direction using a tip-tilt scanner (S-334, PI USA, Auburn, MA, USA). Wavefront correction was performed using a deformable mirror (DM97, Alpao, Montbonnot, France).

## Methods

### Adaptive optics optical coherence tomography instrument

A custom built swept-source AOOCT instrument (Supplementary Fig. 4) was used, based on an FDML light source<sup>42</sup> with central wavelength of 1060 nm and bandwidth of 77 nm (NG-FDML-1060-750-8B-FA, Optores, Munich, Germany) combined with a wavefront sensor (WFS) beacon at 940 nm (SLD-mCS-481-HP2-PM, Superlum, Carrigtwohill, Co. Cork,

### Adaptive optics optical coherence tomography imaging of human participants

Participants with no history or signs of ocular disease were recruited for this study. All participants underwent a comprehensive ophthalmic assessment. In total, four eyes from four healthy participants (age:  $42.8 \pm 13.0$  years) from the National Eye Institute Eye Clinic (National Institutes of Health, Bethesda, Maryland, USA) were imaged using the custom-built AOOCT

retinal imaging system. Eyes were dilated with 2.5% phenylephrine hydrochloride (Akorn Inc.) and 1% tropicamide (Sandoz, A Novartis Division). This study was approved by the Institutional Review Board of the National Institutes of Health (NCT02317328). Research procedures adhered to the tenets of the Declaration of Helsinki. Written, informed consent was obtained from all participants after the nature of the research and possible consequences of the study were explained.

### Data for training and validating the artificial intelligence methods

AOOCT volumes from four eyes of four participants were acquired at 12.7 volumes per second with  $512 \times 512$  A-scan sampling across 1.5 degrees field of view. Volumes were acquired from 5 to 23 retinal locations ranging from 0 to 3 mm temporal to the fovea. At each location, 45–85 volumes were acquired. Following image acquisition, the volumes were digitally flattened to correct for axial eye motion based on the outer retinal layers. The volumes were cropped at the edges to remove image portions with scanner distortion artifacts near the turnaround points of the scanner, resulting in final size of  $512 \times 400 \times 378$  pixels.

To train the AI methods, AOOCT B-scans were extracted from volumes of three participants at locations 0.8–1.2 mm. The region of interest ( $40 \times 400$  pixels) was cropped from the B-scans to form densely sampled ground truth images. These images were then down-sampled by a factor of 4 in the lateral direction to obtain sparsely sampled images ( $40 \times 100$  pixels). More than 10,000 densely/sparsely sampled image pairs were used for training.

To test the model, a leave one subject out based validation scheme was adopted. During testing, each B-scan of the volume is enhanced using AI ( $80 \times 100$  pixels; note the cropping is different than during training). The AI enhanced volumes were then registered to correct for eye motion artifacts and averaged to visualize the cones. The proposed RRTGAN model was also tested on a participant that had never been seen during the training process to validate the generalizability of the method.

### Validation metrics

**Objective image quality assessment.** Four image quality assessment metrics were used in this study: peak signal to noise ratio (PSNR), deep image structure and texture similarity (DISTs)<sup>35</sup>, learned perceptual image patch similarity (LPIPS)<sup>36</sup>, and Fréchet Inception Distance (FID)<sup>37</sup>. PSNR is a simple and widely used measure of the quality of the enhanced and the corresponding ground truth images. It computes the distortion by measuring pixel differences between the enhanced and the ground truth images, with higher values indicating better quality. DISTs and LPIPS extract deep features to assess the image quality of the AI-enhanced images. DISTs measures the textural and structural similarity between enhanced and ground truth images using deep features extracted from five layers of the VGG16 network. LPIPS quantifies human perceptual similarity between enhanced and ground truth images by computing the L2 distance between unit-normalized and scaled features extracted from multiple layers of the VGG network. FID evaluates the AI frameworks by assessing the quality of the generated images compared to the ground truths. A lower FID score guarantees better quality of the generated images.

**Cell spacing and contrast measurement.** Cell spacing and contrast was quantified to assess the efficacy of RRTGAN for cone photoreceptor enhancement. Cell spacing was estimated using the circumferentially averaged power spectrum<sup>38</sup> of each image region of interest. The peak spatial frequency of the spectrum (i.e., the cone photoreceptor fundamental frequency) is an estimate of cell spacing. To convert from pixels to  $\mu\text{m}$ , a paraxial ray trace on a three-surfaced simplified model eye<sup>43</sup> is used after replacing the axial length, corneal curvature, and anterior chamber depth with measurements of these values obtained from each participant (IOL Master, Carl Zeiss Meditec)<sup>44</sup>. Cellular contrast was approximated by measuring the height of the local peak in the circumferentially averaged power spectrum.

**Edge directionality.** Edge directionality of the images was measured to assess the degree of pixelation in the images. It was calculated by first

obtaining the edges of the individual cone cells. At the location of the edges, their direction was computed as  $\arctan \frac{G_y}{G_x}$ , where  $G_y$  and  $G_x$  were the vertical and horizontal edges, respectively. Histograms of the directions were used to visualize the distribution of edge directionalities. Higher magnitudes at 90 and 270 degrees in the histogram were indicative of presence of sharp edges and are useful to detect when cones exhibited a pixelated appearance. Uniformly distributed edges across all angles indicated smooth edges of the cones.

### Statistical analysis

Statistical analysis was performed using a one-tailed paired Student's *t* test to determine if there was statistically significant improvement in cell visualization between the sparsely sampled and the AI enhanced images. The test was applied to the objective image quality metrics (peak signal to noise ratio (PSNR), deep image structure and texture similarity (DISTs), and learned perceptual image patch similarity (LPIPS)) for both image sets for all four participants ( $N = 4$ ).

### Data availability

All data is included in the manuscript and/or supporting information.

### Code availability

The PyTorch implementation of RRTGAN is publicly available at <https://github.com/NIH-NEI/RRTGAN>.

Received: 13 May 2025; Accepted: 25 August 2025;

Published online: 09 December 2025

### References

1. Burns, S. A. et al. Adaptive optics imaging of the human retina. *Prog. Retinal Eye Res.* **68**, 1–30 (2019).
2. Miller, D. T. & Kurokawa, K. Cellular-scale imaging of transparent retinal structures and processes using adaptive optics optical coherence tomography. *Annu. Rev. Vis. Sci.* **6**, 115–148 (2020).
3. Morgan, J. I., Chui, T. Y. & Grieve, K. Twenty-five years of clinical applications using adaptive optics ophthalmoscopy. *Biomed. Opt. express* **14**, 387–428 (2022).
4. Roorda, A. & Duncan, J. L. Adaptive optics ophthalmoscopy. *Ann. Rev. Vis. Sci.* **1**, 19–50 (2015).
5. Liang, J., Williams, D. R. & Miller, D. T. Supernormal vision and high-resolution retinal imaging through adaptive optics. *JOSA A* **14**, 2884–2892 (1997).
6. Zawadzki, R. J. et al. Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal *in vivo* imaging. *Opt. express* **13**, 8532–8546 (2005).
7. Kocaoglu, O. P. et al. Adaptive optics optical coherence tomography at 1 MHz. *Biomed. Opt. Express* **5**, 4186–4200 (2014).
8. Bower, A. J. et al. Integrating adaptive optics-SLO and OCT for multimodal visualization of the human retinal pigment epithelial mosaic. *Biomed. Opt. Express* **12**, 1449–1466 (2021).
9. Klein, T. et al. Megahertz OCT for ultrawide-field retinal imaging with a 1050nm Fourier domain mode-locked laser. *Opt. Express* **19**, 3044–3062 (2011).
10. Liu, Z. et al. Ultrahigh-speed multimodal adaptive optics system for microscopic structural and functional imaging of the human retina. *Biomed. Opt. Express* **13**, 5860–5878 (2022).
11. Migacz, J. V. et al. Megahertz-rate optical coherence tomography angiography improves the contrast of the choriocapillaris and choroid in human retinal imaging. *Biomed. Opt. Express* **10**, 50–65 (2018).
12. Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2017).
13. Shocher, A., Cohen, N. & Irani, M. “zero-shot” super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2018).

14. Wang, X. et al. Esrgan: enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (ECCV, 2018).
15. Liang, J. et al. Swinir: image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (IEEE, 2021).
16. Singh, R. et al. Artificial intelligence in image reconstruction: the change is here. *Phys. Med.* **79**, 113–125 (2020).
17. Krull, A., Buchholz, T.-O. & Jug, F. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (IEEE, 2019).
18. Wu, W. et al. Dual residual attention network for image denoising. *Pattern Recognit.* **149**, 110291 (2024).
19. Isola, P. et al. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2017).
20. Zhu, J.-Y. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (IEEE, 2017).
21. Wang, X. et al. Deep reinforcement learning for image-to-image translation. Preprint at <https://arxiv.org/abs/2309.13672> (2023).
22. Goodfellow, I. et al. Generative adversarial nets. In *Advances in neural information processing systems*, 27 (NeurIPS, 2014).
23. Guerreiro, J. et al. Super-resolution of magnetic resonance images using Generative Adversarial Networks. *Comput. Med. Imaging Graph.* **108**, 102280 (2023).
24. Liu, J. et al. Graded image generation using stratified CycleGAN. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference* (Springer, 2020).
25. Li, J. et al. Artificial intelligence assisted clinical fluorescence imaging achieves in vivo cellular resolution comparable to adaptive optics ophthalmoscopy. *Commun. Med.* **5**, 105 (2025).
26. Mehdizadeh, M. et al. Employing texture loss to denoise OCT images using generative adversarial networks. *Biomed. Opt. express* **15**, 2262–2280 (2024).
27. Das, V. et al. Revealing speckle obscured living human retinal cells with artificial intelligence assisted adaptive optics optical coherence tomography. *Commun. Med.* **4**, 68 (2024).
28. Alajaji, S. A. et al. Generative adversarial networks in digital histopathology: current applications, limitations, ethical considerations, and future directions. *Mod. Pathol.* **37**, 100369 (2024).
29. Zhang, J. et al. BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis. *Comput. Methods Prog. Biomed.* **217**, 106676 (2022).
30. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).
31. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of International Conference on Learning Representations*, (ICLR 2021).
32. Deng, Z. et al. Rformer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark. *IEEE J. Biomed. Health Inform.* **26**, 4645–4655 (2022).
33. Dalmaz, O. & Yurt, M. and T. Çukur, ResViT: residual vision transformers for multimodal medical image synthesis. *IEEE Trans. Med. Imaging* **41**, 2598–2614 (2022).
34. Liu, J. et al. Hybrid transformer for lesion segmentation on adaptive optics retinal images. In *Medical Imaging 2022: Computer-Aided Diagnosis* (SPIE, 2022).
35. Ding, K. et al. Image quality assessment: unifying structure and texture similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 2567–2581 (2020).
36. Zhang, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2018).
37. Heusel, M. et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems*, 30 (NeurIPS, 2017).
38. Cooper, R. F., Aguirre, G. K. & Morgan, J. I. Fully automated estimation of spacing and density for retinal mosaics. *Transl. Vis. Sci. Technol.* **8**, 26 (2019).
39. Curcio, C. A. et al. Human photoreceptor topography. *J. Comp. Neurol.* **292**, 497–523 (1990).
40. Zhang, F. et al. Cone photoreceptor classification in the living human eye from photostimulation-induced phase dynamics. *Proc. Natl. Acad. Sci.* **116**, 7951–7956 (2019).
41. Kurokawa, K., Liu, Z. & Miller, D. T. Adaptive optics optical coherence tomography angiography for morphometric analysis of choriocapillaris. *Biomed. Opt. express* **8**, 1803–1822 (2017).
42. Huber, R., Wojtkowski, M. & Fujimoto, J. Fourier Domain Mode Locking (FDML): A new laser operating regime and applications for optical coherence tomography. *Opt. express* **14**, 3225–3237 (2006).
43. Bennett, A. & Rabbets, R. Proposals for new reduced and schematic eyes. *Ophthalmic Physiol. Opt.* **9**, 228–230 (1989).
44. Liu, J. et al. Automated photoreceptor cell identification on nonconfocal adaptive optics images using multiscale circular voting. *Invest. Ophthalmol. Vis. Sci.* **58**, 4477–4489 (2017).

## Acknowledgements

The authors would like to thank W. M. Zein, L.A. Huryn, B. P. Brooks, D. Claus, I. Espinal-Supp, N. Chen, C. Appleman, S. Yin, J. Suy, A. Papadogiannis, G. Babilonia-Ayukawa, B. Gildersleeve, A. Khasar, J. Oparah, E. Sarbah-Yalley, R. Woodard, S. Bhatta, and M. Arango for assistance with clinical procedures. The Office of Data Science Strategy, National Institutes of Health provided a seed grant enabling us to explore cloud-based computational resources. We would also like to thank J. Dennis, M. Moore, M. Dittrich, A. Oumantsev, S. Corder, J. Yee, S. Ponomarev, D. Nicholls, and W. Akpan for providing assistance with IT infrastructure and computational resources. We thank G. Dold and D. Ide for assistance with custom machining and A. Dubra, Z. Liu and D. X. Hammer for assistance with adaptive optics instrumentation. This work was supported by the Intramural Research Program of the National Institutes of Health, National Eye Institute. The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

## Author contributions

V.D., A.J.B., and J.T. conceived and designed the experiments. V.D., A.J.B., J.L., and N.A. collected the data. V.D., and J.T. analyzed the data. V.D., A.J.B., and J.T. contributed to materials and analysis tools. V.D. and J.T. prepared the manuscript with input from all authors.

## Funding

Open access funding provided by the National Institutes of Health.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44387-025-00038-2>.

**Correspondence** and requests for materials should be addressed to Johnny Tam.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025