Perceptual colorization of the peripheral retinotopic visual field using adversarially-optimized neural networks

Hadar Cohen-Duwek (hadarco@openu.ac.il)
Yahia Showgan (yahiashowgan@gmail.com)
Elishai Ezra Tsur (elishai@nbel-lab.com)

Neuro-Biomorphic Engineering Lab (NBEL), Department of Mathematics and Computer Science,
The Open University of Israel, Ra'anana, Israel

Abstract

There is an apparent discrepancy between visual perception, which is colorful, complete, and in high resolution, and the saccadic, and spatially heterogeneous retinal input data. In this work, we computationally emulated foveated color maps and intensity channels as well as intra-saccadic motion data using a neuromorphic event camera. We used a convolutional neural network, U-Net, and adversarial optimization to demonstrate how retinal inputs can be used for the reconstruction of colorful images in high resolution. Our model may set the groundwork for the development of biologically plausible neural networks for computational vision perception.

Keywords: computational visual processing; computational cognition; neuromorphic vision; event cameras; color perception

Introduction

Retinal visual inputs reflect the intricate anatomical and physiological characteristics of the retina, where visual cues are captured by non-evenly distributed photoreceptors across the retinopatic field. While achromatic rod photoreceptors are found in M-shape-like distributaion across the retina (peak at the pheiphery and absent from the fovea), the chromatic cones photoreceptors are present in low density throughout the retina, peaking at the center of the fovea (Purves et al., 2001). Furthermore, retinal input is compressed and encoded by Retinal Ganglion Cells (RGC) to fit the limited capacity of the optic nerve. Most RGCs have a center-surround antagonist receptive field (RF), realizing spectral (color antagonism)-spatio-temporal filtering that transmits only color changes. The RGC's RF size varies, whereas its size becomes larger as the distance from the fovea increases. Therefore, in comparison to its periphery, the fovea features high visual resolution (Lee, 1996). Interestingly, electrophysiological studies (Field et al., 2010; Solomon et al., 2005) showed that while non-opponent red-green responsive P-type retinal ganglion cells (RGCs), are mostly found in the peripheral retina, opponent yellow-blue sensitive opponent P-type RGCs are mostly found in the retina's center. Consequently, RGCs transmit incomplete visual information from the retina to the brain. Retinal color processing is therefore considered to deteriorate in the peripheral retina, where the visual signals are of low resolution and color cues.

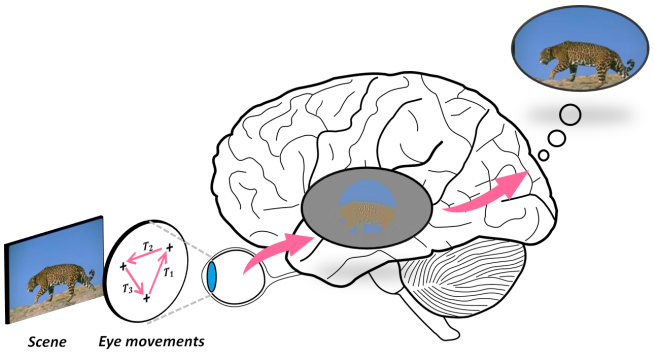

Figure 1: Illustration of perceptual image reconstruction. Reconstruction of a perceived image is based on retinal input during saccades.

There is an important apparent discrepancy between human visual perception, which is sharp and colorful across the full extent of the visual field, and the retinal input (Figure 1). Even though visual information is lacking in the peripheral retina as was described above, our subjective visual experience is that of sharp and colorful surfaces, both in the center and in the peripheral vision (Haun et al., 2017; Tyler, 2015). This apparent discrepancy was recently showcased by Cohen and colleagues (Cohen et al., 2020). Using virtual reality, they showed non-homogeneous color awareness as viewers routinely failed to notice color removals from the majority of their visual perceptive field. They concluded that our intuitive perception of a rich, colorful visual world under active, naturalistic viewing conditions is largely incorrect.

Another important characteristic of human vision it being based on saccade vision, where the eyes are constantly and successively fixating on changing locations several times per second (Gilchrist, 2011). While human visual perception is stable, in saccade vision, retinal objects' projections are continuously altered, supposedly resulting in motion blur and instability. The process of integrating successive fixations into a stable perceptual representation is known as trans-saccadic integration (Irwin, 1996; Melcher and Colby, 2008). During trans-saccadic integration, information from different visual scenes is combined across saccades, providing visual perceptive stability. Recent studies demonstrated that trans-saccadic integration does not simply correspond to

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

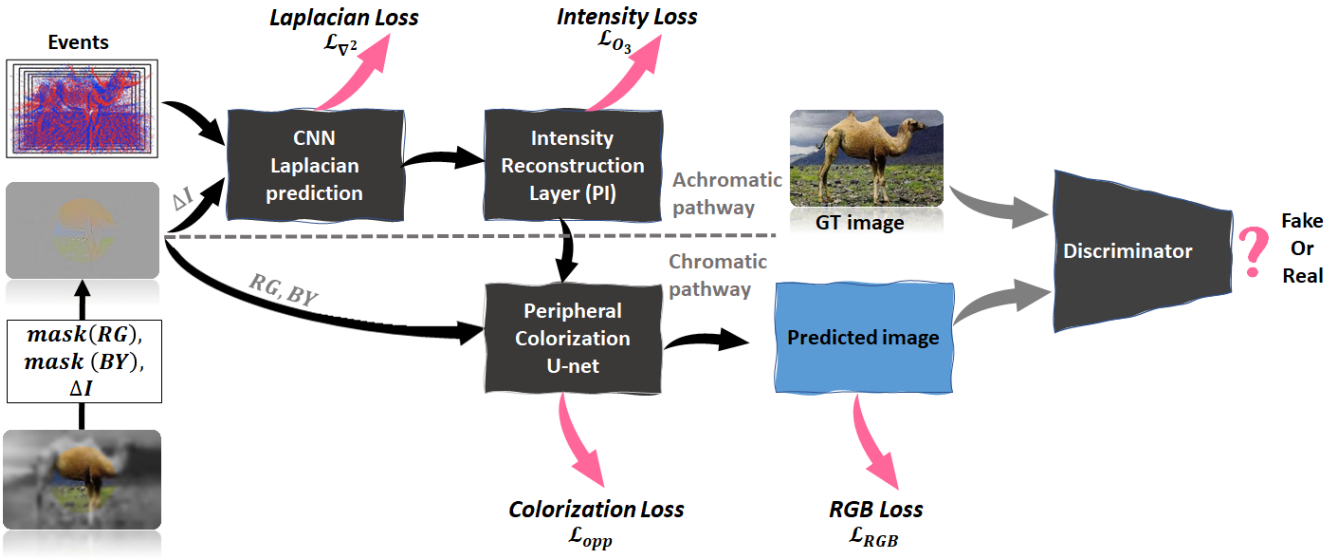

Figure 2: The architecture of the proposed reconstruction and colorization model. The retinal input contains three opponent channel and events from an event camera. Events represent intra-saccadic information. The predicted image and GT image can be used in adversarial training of the discriminator.

disconnected snapshots during each fixation (Stewart and Schütz, 2018; Wolf and Schütz, 2015). Therefore, the process could involve intrasaccadic motion information as well as cues from visual working memory. Notably, it was recently shown that intra-saccadic visual information is associated with visual stability via object correspondence and gaze correction (Schweitzer and Rolfs, 2021). Another feature of saccadic vision is saccadic suppression. Saccadic suppression occurs when visual information is not processed between fixation periods, and it is argued to be responsible for eliminating blurry stimuli (Krekelberg, 2010). Saccadic vision points out another dimension of the discrepancy between visual perception and retinal inputs: despite the eyes' constant movements, we experience the visual world as continuous and seamless through the mechanisms of trans-saccadic integration and saccadic suppression.

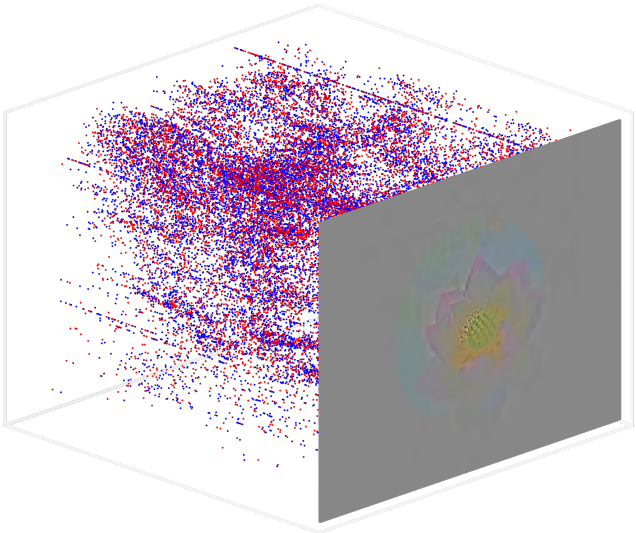

Figure 3: Emulation of a retinal input from the sunflower image (shown in Figure 4). Blue and red dots signify events, specify positive and negative changes in brightness, respectively. The events presented resulted in a total of three saccades.

To conclude, retinal input to the brain contains limited visual information, which is further challenged with instability and blurriness throughout the entire visual field. The visual system overcomes these impairments by employing intensive computations of stabilization, reconstruction, resolution enhancement, and colorization (Figure 1).

In this work, we used a Convolutional Neural Network (CNN) with chromatic and achromatic pathways. In the achromatic pathway, high-frequency temporal and spatial information was used to reconstruct high-resolution intensity images (representing intra-saccadic motion and spatial edges). The chromatic pathway enhances peripheral vision's perception of color, using U-Net for image colorization. Our proposed model allows the reconstruction of a full, colorful, and high-resolution image from incomplete retinal information. An adversarially trained discriminator (convolutional PatchGAN classifier (Isola et al., 2017)), was trained to reconstruct colorful outputs that cannot be distinguished from the real color images (Figure 2).

Methods

Generating retinal input

In this work, we acquired RGB and event data from the neuromorphic dataset N-Caltech101 (Orchard et al., 2015). To generate retinal inputs from a given image we computationally emulated foveated colors maps and intensity channels as well as intra-saccadic motion data (Figure 3). To emulate the RGC's On-Off color opponent receptive fields (Kuffler et al., 1984), input RGB images were converted to three opponent channels: the chromatic RG, BY channels, and the achromatic channel I using:

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} RG \\ BY \\ I \end{pmatrix} = M_{opp} \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ a & b & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

where M_{opp} is the color opponent transformation matrix in which $a = 0.2989, b = 0.587,$ and $c = 0.114$.

Achromatic derivative signal I_{on-off} were derived by convolving the achromatic intensity channel with the discrete Laplacian operator $L = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$ using: $I_{\text{on-off}} = I * L \approx \Delta I$.

We simulated the size of the receptive field, being small in the fovea and larger toward the peripheral retina (Perry and Geisler, 2002) by applying Gaussian filters with different scales on the opponent image. To simulate the deficient color perception in the peripheral vision we used a circular mask whose center was located at the chromatic channels (RG and BY) centers and whose radius R is 30 pixels. We zeroed out all of the pixels located outside of the mask's area.

To simulate the intra-saccadic motion, we used a recording taken by a neuromorphic event camera (silicon retina), available through the Caltech101 dataset. Event-based cameras are considered as following the principles of biological sensing, only capturing changes in scene reflectance, i.e., they report only changes in brightness. Event cameras transmit asynchronous events representing changes in relative intensity at the pixel level with a high milli-second range temporal resolution (Brandli et al., 2014; Gallego et al., 2022). Three saccade modalities were generated by virtually moving each image in front of an event camera (Fei-Fei et al., 2004)). Three saccades were fixed both in time and in target along the recording. Each recorded event-stream (events-data corresponding to three saccades) from each corresponding RGB image was assigned to an event frame, as was recently proposed by (Cohen-Duwek et al., 2021). Briefly, each event-camera file was converted to an event-frame tensor E, which the dimensions $H \times W \times T$ where H, and W are the spatial dimensions, and T=6 is the number of event-frames constituting each image in the dataset. Our retinal input signals were therefore tensors with the size $B \times H \times W \times 9$, where B is the batch size, and 9 stands for the following channels: 6 channels are the event-frame tensor ($90 \times 120 \times 6$) and 3 channels represent the retinal transformation described above, where the 7th channel is the foveated and masked RG channel, the 8th channel is the foveated and masked BY channel and the 9th channel is the Laplacian of the foveated intensity.

Reconstruction and colorization neural network

We utilized an artificial neural network for the reconstruction and colorization of the retinal inputs. Our neural network comprises the following phases (each described in length below): 1) A simple 5-layers Convolutional Neural Network (CNN) that predicts Laplacian from the event-frames (channels 0-6 in the input tensor) as well as the foveated intensity Laplacian (the last channel of the input tensor); 2) A Poisson solver layer that reconstructs the image from the predicted Laplacian input; and 3) A U-Net model for image colorization. We further used a Discriminator to train the network with adversarial examples. Our objective was to minimize the sum of the loss functions of each phase of the Generator while maximizing the ability of the Discriminator to detect "fake" and "real" examples.

Laplacian prediction. This phase is based on the 5-layers CNN, previously proposed by (Cohen-Duwek et al., 2021), which predicts the image Laplacian from event data. Here, however, as opposed to the originally proposed method, the CNN also incorporates the foveated intensity's Laplacian. We used the Mean Absolute Error (MAE) loss to compute the loss of this phase:

$$\mathcal{L}_{\nabla^2} = \lambda_{\nabla^2} MAE\left(Lap\,, \hat{L}ap\right) \quad (2)$$

where Lap is the Laplacian of the original intensity of the image and $\hat{L}ap$ is the predicted Laplacian (the output of the current phase of the Generator).

Poisson solver layer. The Poisson solver layer was implemented to solve the Poisson equation (Poisson Integration, PI) and to reconstruct surfaces (intensity reconstruction) from edge input. Although it does not contain learnable parameters, it was incorporated into the network to back-propagate errors for end-to-end training of the network. The realization of this layer was based on the PI algorithm, previously proposed by (Simchony et al., 1990), applied to the Laplacian and the Predicted Laplacian of the image intensity.

Here, we used the Structural Similarity Index Measure (SSIM) to compute the loss of this layer:

$$\mathcal{L}_{O_3} = \lambda_{PI}(1 - SSIM\left(PI(Lap), \hat{O}_3\right) \quad (3)$$

where $\hat{O}_3$ is the output of the Poisson solver layer.

Image colorization with U-Net. As an input to this phase, the predicted intensity channel ($\hat{O}_3$, the reconstructed intensity calculated at the Laplacian prediction stage) was concatenated with the opponent's color channel.
A U-Net neural network architecture, featuring encoder-decoder schemes with skip connections (Ronneberger et al., 2015, Isola et al., 2017) was utilized here for image colorization. We used the network to minimize two cost functions: (1) the MAE of the real (original) opponent colors of the image and the predicted opponent colors using:

$$\mathcal{L}_{opp} = \lambda_{opp}(MAE\left(O_1, \hat{O}_1\right) + MAE\left(O_2, \hat{O}_2\right)) \quad (4)$$

; and (2) the perceptual similarity of the original I_{RGB} and the predicted $\hat{I}_{RGB}$ RGB color images using two perceptual

similarity metrics: SSIM and LPIPS (Zhang et al., 2018) using:

$$\mathcal{L}_{RGB} = \lambda_{ssim}\left(1 - SSIM\left(I_{RGB}, \hat{I}_{RGB}\right)\right) + \lambda_{lpips} LPIPS\left(I_{RGB}, \hat{I}_{RGB}\right) \quad (5)$$

We defined $I_{RGB} = opp2rgb(O_1, O_2, O_3)$ and $\hat{I}_{RGB} = opp2rgb\left(\hat{O}_1, \hat{O}_2, \hat{O}_3\right)$, where $opp2rgb$ is the linear transformation from opponent channel to RGB color channel computed using the inverse of the opponent matrix M_{opp}.

Generative adversarial network (GAN)

An adversarially trained discriminator (convolutional PatchGAN classifier (Isola et al., 2017)), D, was trained to detect the generator's "fake" images, as the trained Generator, G, produces reconstructed and colorful outputs y that cannot be distinguished from "real" images x. An adversarial D attempts to maximize this objective against G 's attempt to minimize it. The GAN loss was computed using:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[logD(x, y)] + \lambda_D \mathbb{E}_{x,G(x)} \log\left[(1 - D(x, G(x))\right] \quad (6)$$

Where x is the retinal input, y is the GT image transformed to the opponent color space, $\mathbb{E}$ donates the expected value, and λ_D is a gain parameter. In the first term of Equation (6) GT examples are introduced to the discriminator and in the second term, the fake examples, created by the Generator, are presented to the discriminator.

To conclude, the final end-to-end minimization objective is:

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \mathcal{L}_{RGB} + \mathcal{L}_{opp} + \mathcal{L}_{O_3} + \mathcal{L}_{\nabla^2} \quad (7)$$

Implementation details

The model was implemented using TensorFlow and was trained on Google Colab. We divided the N-caltech101 dataset sequences into 6097 training sequences, 1306 validation sequences, and 1306 testing sequences. We use the Adam optimizer (Kingma and Ba, 2015), with a batch size of 16 and an initial learning rate of 0.001. At a plateau, we schedule a 20% learning rate with a minimum value of $2 \cdot 10^{-6}$. A plateau was defined as non-improving validation loss over 6 epochs:

$$Best_{new} < Best_{old}(1 - \alpha), \quad (8)$$

where α is the minimal required relative improvement (here we use α =0.005). We trained each model for 150 epochs. We set $\lambda_{\nabla^2} = 100$, $\lambda_{PI} = 25$, $\lambda_{opp} = 150$, $\lambda_{ssim} = 100$ and $\lambda_D = 10$ over the whole expriments.

Results

As part of our evaluation, we trained our reconstruction and colorization CNN in two training methods: (1) minimizing the generator loss $\mathcal{L}_G$ using:

$$\mathcal{L}_G = \mathcal{L}_{RGB} + \mathcal{L}_{opp} + \mathcal{L}_{O_3} + \mathcal{L}_{\nabla^2} \quad (9)$$

; and (2) adversarially minimizing the GAN's objective function (Eq. 7). Figure 4 shows the original ground truth (GT) images, and the reconstructed images using those two training methods. SSIM and LPIPS perceptual similarity measures are summarized in Table 1. Our results demonstrate that our proposed network is capable of reconstructing high-quality images from retinal input which is lacking peripheral information with both training methods (with and without a discriminator). While the results without the GAN optimization achieve a better perceptual score in both SSIM and LPIPS, the obtained images are more colorful in their peripheral areas. For example: (1) the sunflower leaves are reconstructed as green while colored brown in the GT image; and (2) the background flora in the black-and-white photograph of the elephant was reconstructed as green is colored gray in the GT image (Figure 4). Interestingly, green appears to be the GAN's color of choice for peripheral areas (e.g., the tail of the fish and the neck of the cougar (Figure 4)). The reason for this may be that many images in the dataset have large greenish areas in their background. Moreover, results without the GAN tend to appear achromatic or brownish at the periphery indicating that the without the GAN, the network has struggled to colorize achromatic areas.

To investigate the effect of intra-saccadic motion on the reconstructed images, we further trained our model with foveated images (RGB data without events). SSIM, LPIPS measures, and the reconstructed images are shown in Table 1 and Figure 4. Without events data (representing intra-saccadic motion), the reconstructed images are not as sharp in the periphery as they are with events data. The blurring effect is particularly noticeable in images with high-detail texture, such as the rooster image, or in the images with detailed peripheral information, such as the fishtail and the face images. LPIPS and SSIM measures demonstrate improved reconstruction when events are used. Interestingly, when GAN optimization is used without events, the colors appear to be corrupted as is evident in the rooster's background color, and the red spot on the cougar's face.

Table 1: Image quality evaluation metrics (D stands for the use of the GAN).

Method	SSIM ↑	LPIPS ↓
Events + D	0.7840	0.2120
Events - D	**0.8329**	**0.1643**
No Events + D	0.7651	0.2240
No Events - D	0.7682	0.2114

Discussion

In this work, we demonstrate how a deep neural network can be used to reconstruct high-resolution, colorful images of entire fields of view from limited and realistic retinal inputs. The reconstruction process handles the lack of peripheral visual cues using achromatic and chromatic pathways (Shapley, 2019). As we recently demonstrated (Cohen-Duwek et al., 2022; Cohen-Duwek and Ezra Tsur, 2022; Cohen Duwek and Ezra Tsur, 2021), the achromatic pathway is responsible for the reconstruction of whole surfaces from

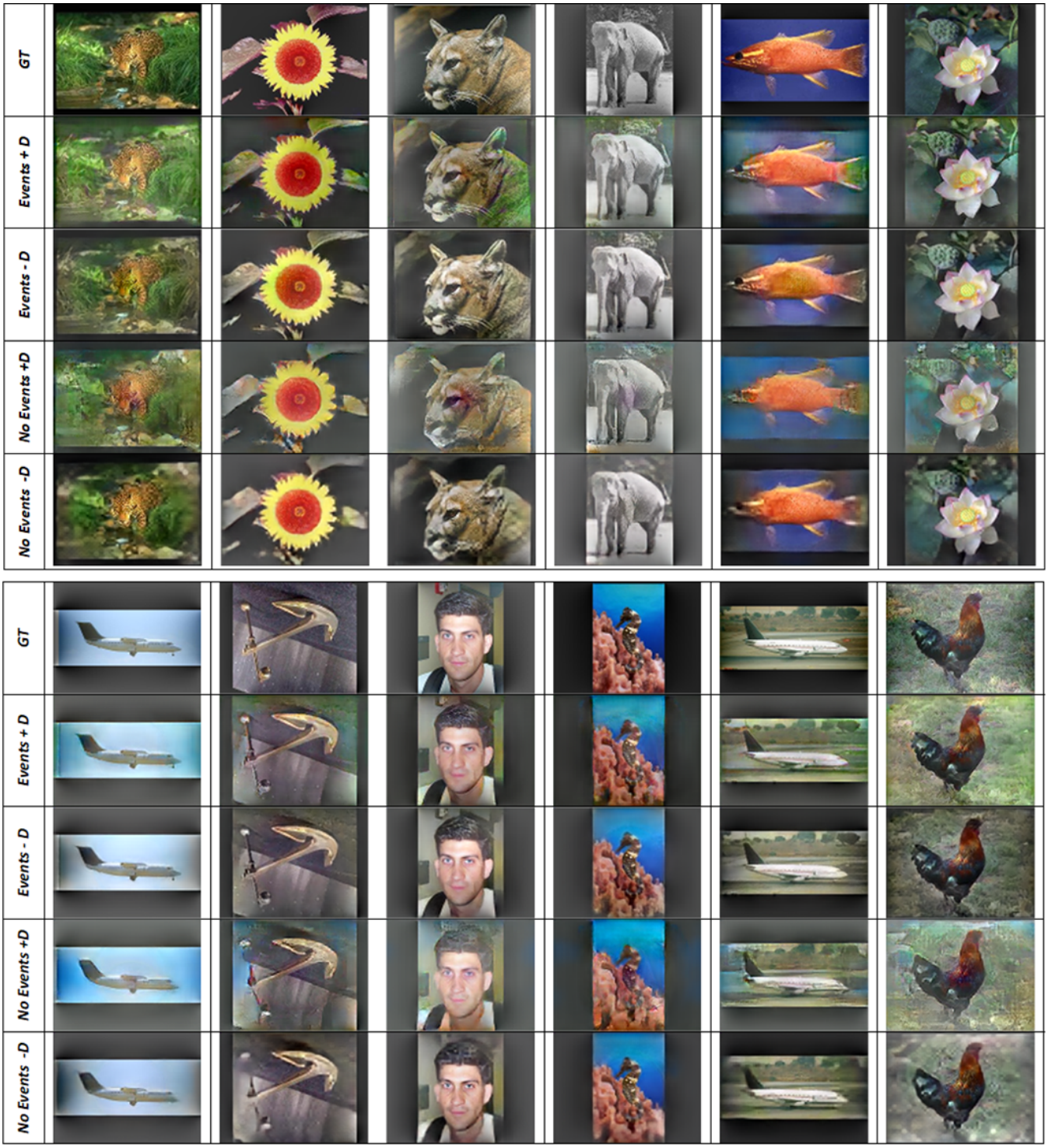

Figure 4: Selected image reconstructions. GT stands for ground truth (for reference); Event + D stands for using events in the input and adversarial training (Discriminator). Event – D stands for using event with regular optimization. No event stands for using only the opponent channels.

achromatic edge information (high spatial frequency, or events). In this work, we extended this model to handle foveated Laplacian of images. This data represents the center-surround response of achromatic RGCs, emulating the retinal input produced by both spatial intensity and intra-saccadic motion signals via fixational saccadic eye movements. The chromatic pathway predicts colors in peripheral vision based on natural image statistics using a U-Net architecture.

Our model was evaluated using two different training methods: a classical training procedure and an adversarial optimization. We found that, while non-adversarial training produces better perceptual errors and similarity measures (SSIM, LPIPS), adversarial training produces more colorful results in the images' peripheral areas. Furthermore, we found that the intra-saccadic information (here, the data acquired from an event camera) enhances the acquired visual cues, allowing the reconstruction of peripheral visual areas in a higher resolution.

The adversarial training of our model provides more colorful results in the peripheral areas. Despite the colorful results, Discriminator produces a lower LPIPS score. In this case, the LPIPS measures the degree of similarity between GT and its prediction and cannot account for visual or perceptual phenomena in which perception differs from GT, such as visual illusions (Cohen-Duwek et al., 2022). Accordingly, when colors are added to the prediction (for example, the elephant in Figure 4), the LPIPS decreases because colors are not present in GT.

Although most of the reconstructed images appear sharp and colorful when using adversarial training, some appear slightly corrupted. This problem appears to be caused by the presence of many unnatural images in the Caltech101 dataset (such as illustrations). Additionally, to fit the RGB Caltech101 images to the dimensions of the event camera's sensor, the GT RGB images had to be scaled and padded with zeros. Thus, a large number of images in our modified dataset had black borders, which may result in insufficient colorization of the peripheral areas. Interestingly, despite the limitations of the dataset we used, when we optimized the model with adversarial training, it was able to color both achromatic images (e.g., the elephant and the sunflower leaves (Figure 3)) and achromatic areas using learned statistics of natural images (Isola et al., 2017). A more realistic and natural dataset would likely result in better statistics of natural images as well as more realistic colorization. Furthermore, the N-Caltech101 dataset was acquired using a fixed saccadic motion. All three saccades in the dataset occurred at the same time and were targeted at the same target locations. Therefore, to reconstruct images with more realistic data, a new dataset, which contains random saccades, or attentional saccades (Itti et al., 1998), randomized both in timing and target location is desired.

To simplify the current work, we assumed that the peripheral vision is colorless. However, when presented with monochrome and large objects, observers are sensitive to color in the peripheral vision to some degree (Tyler, 2015). With a more realistic model of peripheral vision, such as was proposed by Haun, (2021), we would be able to incorporate more color information into the network. However, with the colorless periphery as input to the model, our model can be considered as a possible explanation for Cohen's experiment. Cohen and colleagues (2020) found that observers did not notice the absence of peripheral colors. Using our model, we demonstrate that peripheral color can be perceived (predicted colorization) based on achromatic input in the peripheral visual field.

As a further simplification, we used a feedforward architecture in this work. A more biologically plausible model, however, might be based on a recurrent architecture. A recurrent architecture incorporating memory, such as Long Short-Term Memory (LSTM), could potentially solve the problem of random saccades. In this context, convolutional LSTMs (Shi et al., 2015) may be considered to be working visual memory (Stewart and Schütz, 2018) for the task of trans-saccadic integration, which may be the function of the core mechanism for both high visual accuracy in the peripheral region as well as stabilization of vision across saccades. Additionally, by incorporating LSTMs into our model, we will be able to explain the VR experiments conducted by Cohen colleagues (Cohen et al., 2020) that found that color removal was not noticed by observers in this experiment. Working memory (LSTM) may be useful in "filling in" colors in areas that were colored before the color was removed in the experiment.

An important question that should be asked here is whether "Blackbox" models like deep artificial neural networks, which use non-biologically plausible optimization techniques (such as backpropagation) can model human perception. This question was previously discussed by leading researchers from both neuroscience and machine learning fields. It is generally agreed that current theories in systems neuroscience could be enhanced by a cohesive framework based on optimization (Richards et al., 2019). Following this idea and despite replacing one black box (the brain) with another black box (deep neural network), we were able to demonstrate that despite the limited amount of color observers perceive "in the blink of an eye" (Cohen and Rubenstein, 2020) and the limits of color awareness during real-world vision (Cohen et al., 2020), and despite the limitations of the retinal inputs, a neural network inspired by some principles of biological brains can perceive (or predict) a colorful and rich environment.

Acknowledgments

We would like to thank the members of the Neuro and Biomorphic Engineering Lab (NBEL) at the Open University of Israel for the fruitful discussions.

References

Brandli, C., Berner, R., Yang, M., Liu, S. C., and Delbruck, T. (2014). A 240 × 180 130 dB 3 µs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, *49*(10), 2333–2341.

Cohen-Duwek, H., and Ezra Tsur, E. (2022). Biologically Plausible Illusionary Contrast Perception with Spiking Neural Networks. *IEEE International Conference on Image Processing (ICIP)*.

Cohen-Duwek, H., Shalumov, A., and Tsur, E. E. (2021). Image Reconstruction From Neuromorphic Event Cameras Using Laplacian-Prediction and Poisson Integration With Spiking and Artificial Neural Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1333–1341.

Cohen-Duwek, H., Slovin, H., and Tsur, E. E. (2022). Computational modeling of color perception with biologically plausible spiking neural networks. *PLOS Computational Biology*, *18*(10), e1010648.

Cohen Duwek, H., and Ezra Tsur, E. (2021). Biologically Plausible Spiking Neural Networks for Perceptual Filling-In. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Cohen, M. A., Botch, T. L., and Robertson, C. E. (2020). The limits of color awareness during active, real-world vision. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(24), 13821–13827.

Cohen, M. A., and Rubenstein, J. (2020). How much color do we see in the blink of an eye? *Cognition*, *200*, 104268.

Fei-Fei, L., Fergus, R., and Perona, P. (2004). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2004-January*(January).

Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., MacHado, T. A., Jepson, L. H., Shlens, J., Gunning, D. E., Mathieson, K., Dabrowski, W., Paninski, L., Litke, A. M., and Chichilnisky, E. J. (2010). Functional connectivity in the retina at the resolution of photoreceptors. *Nature 2010 467:7316*, *467*(7316), 673–677.

Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A. J., Conradt, J., Daniilidis, K., and Scaramuzza, D. (2022). Event-Based Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(1), 154–180.

Gilchrist, I. (2011). Saccades. *The Oxford Handbook of Eye Movements*.

Haun, A. M. (2021). What is visible across the visual field? *Neuroscience of Consciousness*, *2021*(1).

Haun, A. M., Tononi, G., Koch, C., and Tsuchiya, N. (2017). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness*, *2017*(1).

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, *5*(3), 94–100.

Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*, 5967–5976.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

Kingma, D. P., and Ba, J. L. (2015, December 22). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Krekelberg, B. (2010). Saccadic suppression. *Current Biology*, *20*(5), 228–229.

Kuffler, S., Nicholls, J., and Martin, A. (1984). *From Neuron to Brain, 2nd Eds*.

Lee, B. B. (1996). Receptive field structure in the primate retina. *Vision Research*, *36*(5), 631–644.

Melcher, D., and Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences*, *12*(12), 466–473.

Orchard, G., Jayawant, A., Cohen, G. K., and Thakor, N. (2015). Converting Static Image Datasets to Spiking Neuromorphic Datasets Using Saccades. *Frontiers in Neuroscience*, *9*(NOV), 437.

Perry, J. S., and Geisler, W. S. (2002). Gaze-contingent real-time simulation of arbitrary visual fields. *Human Vision and Electronic Imaging VII*, *4662*, 57–69.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M. (2001). Anatomical Distribution of Rods and Cones. In *Neuroscience* (2nd ed.). Sinauer Associates.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., … Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience 2019 22:11*, *22*(11), 1761–1770.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241.

Schweitzer, R., and Rolfs, M. (2021). Intrasaccadic motion streaks jump-start gaze correction. *Science Advances*, *7*(30).

Shapley, R. (2019). Physiology of Color Vision in Primates. *Oxford Research Encyclopedia of Neuroscience*.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C., and Kong Observatory, H. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in Neural Information Processing Systems*, *28*.

Simchony, T., Chellappa, R., and Shao, M. (1990). Direct Analytical Methods for Solving Poisson Equations in Computer Vision Problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(5), 435–446.

Solomon, S. G., Lee, B. B., White, A. J. R., Rüttiger, L., and Martin, P. R. (2005). Chromatic Organization of Ganglion Cell Receptive Fields in the Peripheral Retina. *Journal of Neuroscience*, *25*(18), 4527–4539.

Stewart, E. E. M., and Schütz, A. C. (2018). Optimal trans-saccadic integration relies on visual working memory. *Vision Research*, *153*, 70–81.

Tyler, C. W. (2015). Peripheral color demo. *I-Perception*, *6*(6), 1–5.

Wolf, C., and Schütz, A. C. (2015). Trans-saccadic integration of peripheral and foveal feature information is close to optimal. *Journal of Vision*, *15*(16), 1–1.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595.