Image reconstruction for the artificial compound eye based on deep learning

Jioh Lee, Cheolsun Kim, Youngin Choi, Heung-no Lee

Jioh Lee, Cheolsun Kim, Youngin Choi, Heung-no Lee, "Image reconstruction for the artificial compound eye based on deep learning," Proc. SPIE 12438, AI and Optical Data Sciences IV, 1243815 (15 March 2023); doi: 10.1117/12.2648154

SPIE.

Event: SPIE OPTO, 2023, San Francisco, California, United States

Image reconstruction for the artificial compound eye based on deep learning

Jioh Lee[a], Cheolsun Kim[b,c], Youngin Choi[a], and Heung-no Lee[*a,b]

[a]Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea, 61005.
[b]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea, 61005.
[c]Smart Electronics Research Center, Korea Electronics Technology Institute, Jeonju-si, Jeollabuk-do, Republic of Korea, 54853.

ABSTRACT

The visual system of arthropods, called the compound eye, has distinctive features such as a wide field of view, high-speed motion detection, and infinite depth of field. These features have attracted researchers to build artificial compound eyes. However, the compound eye is limited in spatial resolution by its structural constraints such as the number and size of ommatidia that compose the compound eye. These constraints also can be found in the existing artificial compound eye. In previous work, a design method overcame these limitations and achieved resolution improvements by increasing the acceptance angle of ommatidia and using numerical optimization based on compressive sensing (CS). However, the limitation is that prior information such as a sparsifying basis is needed to solve the numerical optimization problem, and obtaining the solution to this problem is computationally time-consuming. In this paper, we propose a deep learning-based artificial compound eye. The deep learning architecture takes a measurement from the compound eye as input and learns how to reconstruct the original image. The experimental result demonstrates that the proposed deep learning approach provides improved performance in image reconstruction for the artificial compound eye.

Keywords: compound eye, computational imaging, numerical optimization, deep learning

1. INTRODUCTION

The apposition compound eye is the type of eye that can be found in arthropods that exhibits several distinctive features, including a wide field of view, high speed of motion detection, and infinite depth of field.[1–3] It is composed of optical units called ommatidia.[4–6] Many researchers have attempted to replicate this structure in various ways, including simulating compound eyes with computers[7] and implementing real optical devices.[8–10] These artificial compound eyes have a range of potential, including use as digital cameras on microrobots, motion detection, endoscopy, and image-guided surgery.[8–10] However, the apposition compound eye has a lower spatial resolution compared to vertebrate eyes due to its optical structure,[11–13] and achieving a comparable level of resolution would require an impractical design.[12] Despite the development of artificial compound eyes that exploit the structural benefits of the compound eye, the limitation of the low spatial resolution remains.

In previous work, a design method[7] for improving the spatial resolution of artificial compound eyes has been proposed, including increasing the acceptance angle of ommatidia and adopting reconstruction techniques based on compressive sensing (CS). CS is a method of reconstructing the original signal from measurements by solving a convex optimization problem or linear inverse problem.[14] There are manifold algorithms that solve these problems by exploiting sparsity constraints to obtain a unique solution.[15–17] However, these algorithms are limited by the fact that the sparsity of the input signal affects the performance of reconstruction. While any images in nature can be represented as a sparse linear combination of basis such as discrete Fourier transform, wavelet transform, and discrete cosine transform,[18,19] we are not able to represent all images as a spare signal with a fixed sparsifying basis. As a result, the performance of image reconstruction algorithms based on CS

*heungno@gist.ac.kr

AI and Optical Data Sciences IV, edited by Bahram Jalali, Ken-ichi Kitayama,
Proc. of SPIE Vol. 12438, 1243815 · © 2023 SPIE
0277-786X · doi: 10.1117/12.2648154

can vary depending on the choice of sparsifying basis. In addition, solving these problems is computationally time-consuming.

Recently, several researchers have employed deep learning frameworks to solve the linear inverse problem.[20–24] In this study, we propose a method for reconstructing images from measurements taken by an artificial compound eye using a deep learning approach. The proposed deep learning architecture takes measurements from the artificial compound eye as inputs and learns to map these measurements to the original image. We compare the quality of the reconstructed images and the computation time of the proposed deep learning method to the reconstruction algorithm based on CS[17] in a noisy environment. The results of our experiments show that the deep learning-based reconstruction method for the artificial compound eye outperforms the CS-based method in terms of the quality of the reconstructed image and computation time.

2. ARTIFICIAL COMPOUND EYE

2.1 System description

An ommatidium of the apposition type of compound eye consists of the cornea, crystalline cone, rhabdom, and photoreceptor cell.[4–6] In biologically inspired artificial compound eyes, each ommatidium can be modeled as a set of microlenses and photodetector placed on a flattened,[25] curved,[26] spherical,[10] hemispherical[8,9] surface. In this study, we consider that each ommatidium is placed on a hemispherical surface. Our artificial compound eye has ommatidia that collect light intensities within their receptive field, defined by acceptance angle $\Delta\phi$, and are located on the spherical coordinate (r, ϕ, θ). We note that the acceptance angle of ommatidium in an artificial compound eye can be set by adjusting the microlens diameter D, photodetector diameter d, light wavelength λ, focal length f of ommatidial optics, and the material of the microlens.[12] Figure 1(a) shows an illustration of the parameters of the ommatidia.

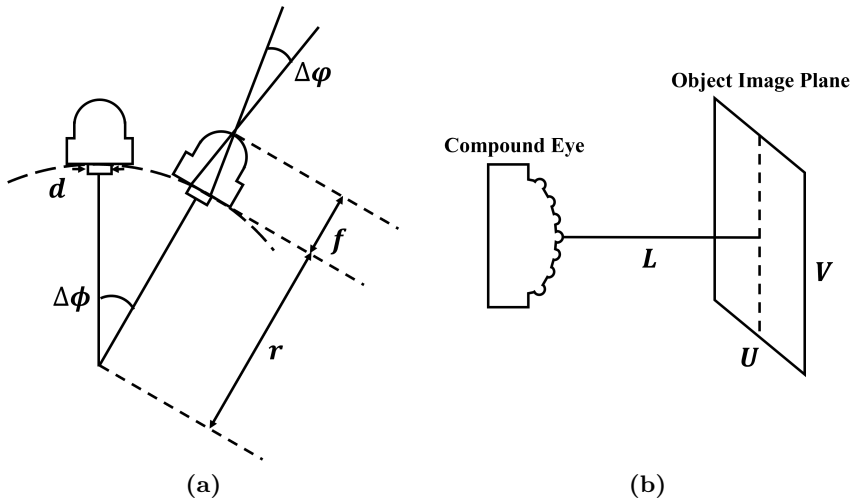
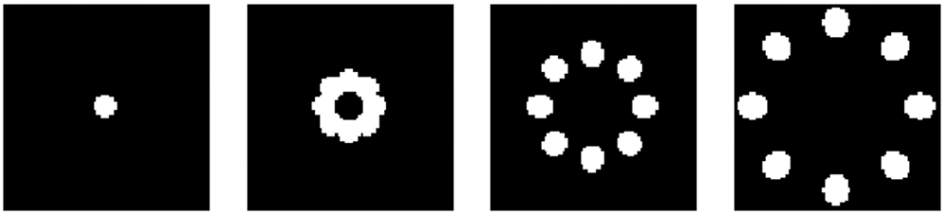


Figure 1. Illustrations of parameters. (a) parameters of the ommatidia. $\Delta\phi$: acceptance angle, $\Delta\varphi$: angle between ommatidia in spherical coordinate, d: diameter of photodetector, f: focal length of microlens, r: radius of compound eye. (b) The imaging system of the artificial compound eye.

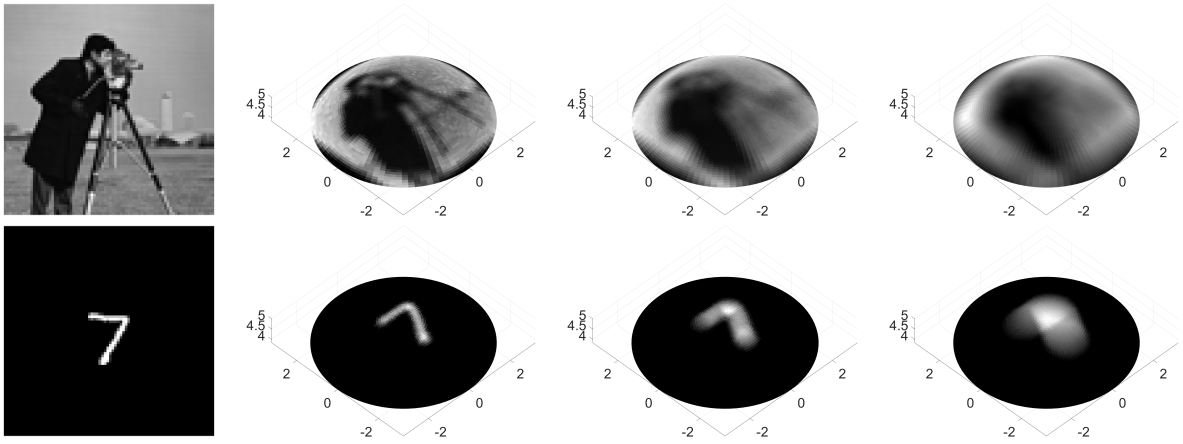
Figure 1(b) shows the schematic of the imaging system of the artificial compound eye. The object image plane is a size of $U \times V$mm and the compound eye observes the object image plane from a distance of L mm. We assume that each pixel is a size of 1×1mm. In this paper, the optical transfer function of the ommatidia is simply designed to sum up the intensities of all light within their receptive field, defined by the acceptance angle, while ignoring light intensities outside of this field. Each measurement of an ommatidium corresponds to a single pixel in the image captured by the compound eye. Figure 2 shows the receptive field of ommatidia located at a set of specific coordinates.

The linear system of equations for the artificial compound eye is described in the following details, as referenced in.[7] Let N denote the number of pixels of an image that we aim to reconstruct and consider the artificial



(a) **(b)** **(c)** **(d)**

Figure 2. the receptive field of the parameter set. Each ommatidium is placed on spherical coordinates. 80×80mm of the object image plane and 1 pixel corresponds to 1×1mm. $r = 5\text{mm}, \Delta\varphi = 10°, L = 50\text{mm}$ (a) $\phi = 0°, \theta = 0°$. (b) $\phi = 10°$. (c) $\phi = 10°$. (d) $\phi = 30°$. $\theta = 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°$ for (b), (c), (d) in a counterclockwise direction.



(a) **(b)** **(c)** **(d)**

Figure 3. Measurements of the artificial compound eye according to acceptance angle. $N = 80 \times 80, M = 3601, \Delta\phi = 1°, \Delta\theta = 4°, r = 5\text{mm}, L = 50\text{mm}$. (a) input images (b) $\Delta\varphi = 5°$. (c) $\Delta\varphi = 10°$. (d) $\Delta\varphi = 20°$

compound eye with M ommatidia. Let y_i denote the measurement of i^{th} ommatidium and x_j denotes value of j^{th} pixel. The object image can be represented as a vector $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T \in \mathbb{R}^{N \times 1}$. It is important to note that the optical transfer function of ommatidia is designed to sum up the intensities of all light within a corresponding receptive field, which is determined by the acceptance angle while ignoring light intensities outside of this field. Thus, transfer function of i^{th} ommatidium can be expressed as a vector $\mathbf{a_i} = [a_{i1}, a_{i2}, \ldots, a_{iN}] \in \mathbb{R}^{1 \times N}$ and $a_i j = 1$ means that the j^{th} pixel is within the receptive field of i^{th} ommatidium and $a_i j = 0$ otherwise. As a result, measurement of i^{th} ommatidium y_i can be calculated as follows:

$$y_i = \mathbf{a_i} \cdot \mathbf{x} \tag{1}$$

Thus, the linear system of equation with M ommatidia can be represented as follows:

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^{M \times 1}$ is the measurement vector and $\mathbf{n}$ is a measurement noise vector modeled as white Gaussian noise. $\mathbf{A} = [\mathbf{a_1}^T, \mathbf{a_2}^T, \ldots, \mathbf{a_M}^T]^T \in \mathbb{R}^{M \times N}$ denotes sensing matrix of the artificial compound eye, where each row of $\mathbf{A}$ represents the transfer function of the corresponding ommatidium. Figure 3 shows the measurements of an input image captured by the artificial compound eye.

In prior research, numerous algorithms[15–17] for solving the linear inverse problem have been proposed to reconstruct the original input image x. In this paper, we propose the reconstruction framework based on deep learning for the artificial compound eye, as an alternative to CS-based reconstruction algorithms.

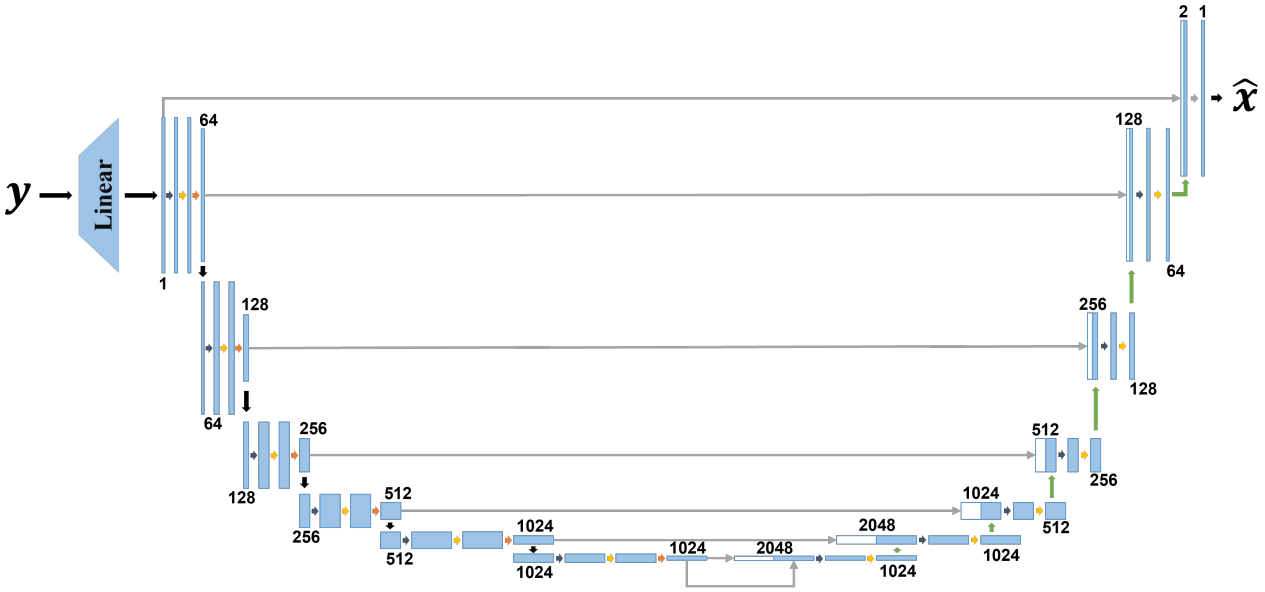


Figure 4. The network architecture of the proposed framework.

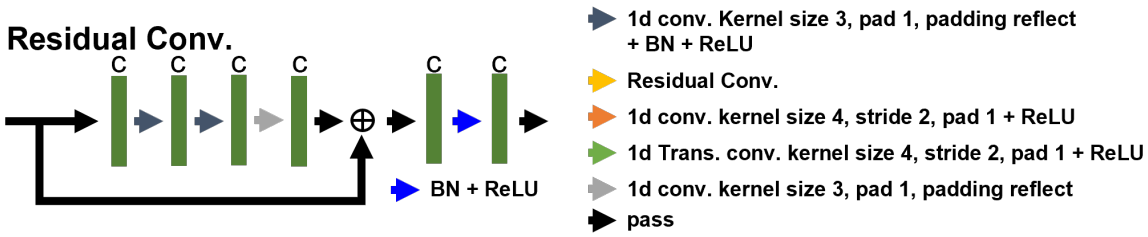


Figure 5. The details of operations in the network architecture of the proposed framework.

2.2 Deep learning approach

Many researchers have applied deep learning approaches to solving inverse problems.[20–24] Convolutional neural network (CNN) can learn a mapping from input to output through training on structured datasets and efficiently produce output results based on inputs at a lower cost than CS recovery algorithms.[20,21,23] In spectroscopy, it has been demonstrated that a CNN architecture with residual connections is more effective at solving the inverse problem and recovering the original spectra than a conventional CNN architecture under noisy environments.[24] Also, U-Net-based CNN architecture is suitable for image reconstruction.[22,27] In this paper, we propose U-Net-based deep learning architecture with residual convolution block to reconstruct the original image from measurements of the artificial compound eye.

Figure 4 shows the overall network architecture of the proposed framework. As an input to our network, we compute measurement vector $\mathbf{y}$ by multiplying the input vector $\mathbf{x}$ and sensing matrix $\mathbf{A}$ and feed this forward to the network. In an input layer of the network, we apply a linear layer and initialize its parameters by the pseudo inverse of sensing matrix $\mathbf{A}^+ \in \mathbb{R}^{N\times M}$ in order to increase the dimension of the input vector from M to N. We note that we use 1-D convolution operation throughout the network due to the 1-D input vector. The remainder of the network consists of two parts: a contracting path on the left side of the network and an expansive path on the right side of the network. The contracting path is composed of repeated blocks of 1-D convolution with a kernel size of 3, followed by batch normalization (BN) and a rectified linear unit (ReLU), a residual convolution (as shown in Figure 5), and a 1-D convolution with a kernel size of 4 and a stride of 2, followed by a ReLU. In place of using a pooling operation for downsampling at the end of each block, we utilize a convolution with a kernel size of 4 and a stride of 2. The expansive path also consists of repeated blocks but utilizes a transposed

convolution with kernel size 4 and stride of 2, followed by ReLU at the end of each block for upsampling. Each block of the expansive path takes the concatenation of the output features of the previous block and output features from the corresponding block of the contracting path, which pass a single 1-D convolution layer. In the output layer, we concatenate the output from the input layer with the output from the final block of the expansive path, apply 1-D convolution, and obtain reconstructed image $\hat{\mathbf{x}}$. The specific number of features of convolution operations is shown in Figure 4.

3. EXPERIMENT AND RESULTS

In this paper, the artificial compound eye that has 3601 ommatidia is considered. We note that each ommatidium is located on spherical coordinates (r, ϕ, θ). We set r to 5mm, ϕ to the range of [0, 40], $\Delta\varphi$ to 1°, θ to the range of [0, 360), $\Delta\theta$ to 4°, L to 50mm, U to 80mm, and V to 80mm. We also consider that each pixel is the size of 1×1mm. To simulate noisy environments, we add white Gaussian noise of 40 decibels (dB) of Signal-to-Noise-Ratio (SNR), which is defined as follows:

$$SNR_{dB} = 10 \times \log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{3}$$

where P denotes the power of the signal. In this paper, we use the alternating direction method[17] to compare the performance of the CS-based reconstruction algorithm with our proposed framework, and we consider the discrete cosine transform (DCT) as the sparsifying basis. The CS-based reconstruction simulation is conducted using MATLAB on a CPU.

MNIST is a dataset containing 60,000 training images and 10,000 test images of handwritten digits with a size of 24×24 pixels. We consider the image to be placed at the center of the object image plane and captured by the artificial compound eye. Using the training dataset, we trained our network to minimize the following loss function:

$$MSE = \frac{1}{K}\sum_{k=1}^{K} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \tag{4}$$

where $\mathbf{x}$ denotes the training image, $\hat{\mathbf{x}}$ denotes the reconstructed image, and K denotes the total number of images. By the backpropagation, the network is trained to reduce the mean squared error (MSE) over the training images. we trained our network with 200 epochs and a batch size of 128 and evaluated it via test images. We used Adam[28] as an optimizer for updating our network parameters, with parameters $\beta_1 = 0.5, \beta_2 = 0.9$, and a learning rate of 0.00005. The training process is conducted in a distributed environment using PyTorch on two NVIDIA TITAN RTX GPUs.

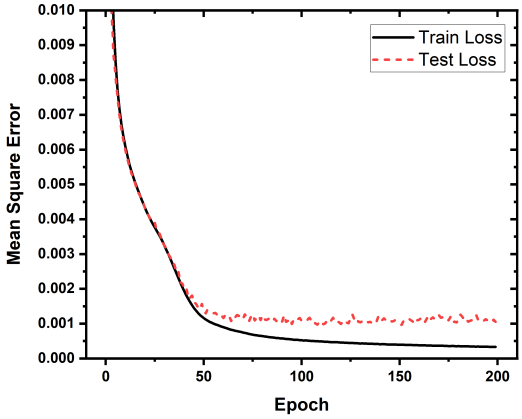


Figure 6. The training loss and test loss evaluated on every epoch.

Figure 6 shows the training loss and test loss evaluated at each epoch. The training loss and test loss converge to an MSE of 0.00033 and 0.00094, respectively. We calculate the MSE of the CS-based reconstruction method on the test images. The MSE of the CS-based reconstruction method using a DCT sparsifying basis is 0.0361. We also measure the average time duration of single image reconstruction. For CS-based reconstruction, we evaluate the average time using MATLAB. The average time duration is 62.8613 seconds. For the proposed deep learning-based method, we use a profiler that is implemented in PyTorch to measure the average time of forwarding of a single measurement on CPU. According to the PyTorch profiler, the single forward duration time is 436.368ms. The training, test loss, and profiling result shows that the proposed deep learning-based reconstruction method outperforms in terms of MSE and speed.

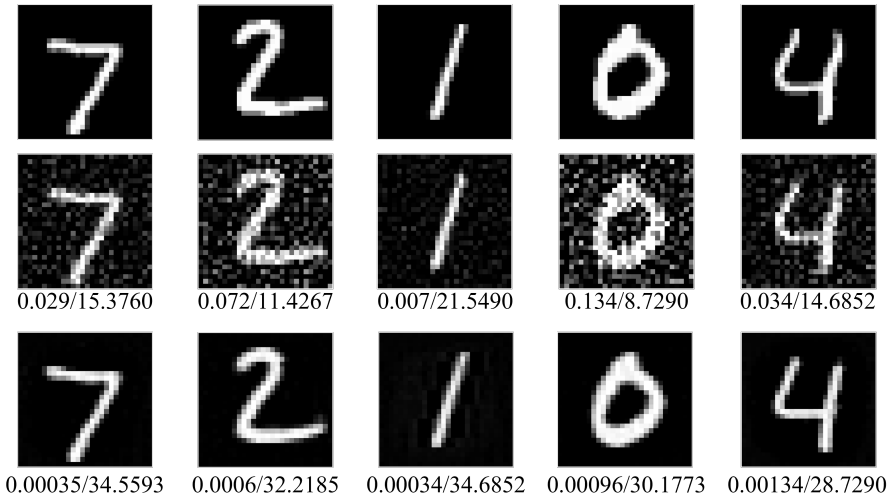


Figure 7. Results of each reconstruction method on the first 5 images of the test dataset in a noisy environment. The first row displays the original input images. The second row illustrates the reconstruction results of the CS-based method. The third row shows the results of the proposed method. The number below each image indicates MSE/PSNR.

Figure 7 shows the first five images of the test dataset and reconstruction results of the CS-based method and proposed method. As previously stated, the MSE of the CS-based reconstruction method is higher than the deep learning-based method. Peek-Signal-to-Noise-Ratio (PSNR) of the CS-based reconstruction result on the first 5 test images is 15.3760 dB, 11.4267 dB, 21.5490 dB, 8.7290 dB, 14.6852 dB, respectively. On the other hand, the Peek-Signal-to-Noise-Ratio (PSNR) of the deep learning-based reconstruction result is 34.5593 dB, 32.2185 dB, 34.6852 dB, 30.1173 dB, 28.7290 dB, respectively. The proposed method shows far superior performance of the quality of the reconstructed image.

4. CONCLUSION

This paper presents the deep learning-based framework for image reconstruction from measurements captured by an artificial compound eye. We formulated a mathematical model of the compound eye and ommatidia and trained our network on specific datasets. In addition, we compared the reconstruction performance of our proposed method to that of a CS-based reconstruction method. The MSE of our proposed method on test images was 0.00094, while the MSE of the CS-based reconstruction method was 0.0361. The average time duration for reconstruction with the proposed method was 436.368ms, while the average time duration for the CS-based reconstruction method was 62.8613s. The experimental results demonstrate that our network successfully learned a structured representation from the dataset and a mapping from measurement to the original image, leading to superior performance in terms of MSE and speed when compared to the CS-based reconstruction method. These techniques can be used in small, high-resolution computational endoscopic cameras.

ACKNOWLEDGMENTS

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIP) (NRF-2021R1A2B5B03002118) and This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program(IITP-2021-0-01835) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)

REFERENCES

[1] Dudley, R., [*The Biomechanics of Insect Flight Form, Function, Evolution*], Princeton University Press (2000).

[2] Warrant, E. and Nilsson, D.-E., [*Invertebrate Vision*], Cambridge University Press (2006).

[3] Zufferey, J.-C., Beyeler, A., and Floreano, D., [*Optic Flow to Steer and Avoid Collisions in 3D*], 73–86, Springer Berlin Heidelberg, Berlin, Heidelberg (2010).

[4] Land, M. F., "The optics of animal eyes," *Contemporary Physics* **29**(5), 435–455 (1988). doi: 10.1080/00107518808222601.

[5] Dan, E. N., "Vision optics and evolution," *BioScience* **39**(5), 298–307 (1989).

[6] Land, M. F. and Nilsson, D.-E., [*Animal Eyes*], Oxford University Press (2012).

[7] Lee, W.-B., Jang, H., Park, S., Song, Y. M., and Lee, H.-N., "Compu-eye: a high resolution computational compound eye," *Optics Express* **24**(3), 2013–2026 (2016).

[8] Song, Y. M., Xie, Y., Malyarchuk, V., Xiao, J., Jung, I., Choi, K.-J., Liu, Z., Park, H., Lu, C., and Kim, R.-H., "Digital cameras with designs inspired by the arthropod eye," *Nature* **497**(7447), 95–99 (2013).

[9] Hu, Z.-Y., Zhang, Y.-L., Pan, C., Dou, J.-Y., Li, Z.-Z., Tian, Z.-N., Mao, J.-W., Chen, Q.-D., and Sun, H.-B., "Miniature optoelectronic compound eye camera," *Nature Communications* **13**(1), 1–10 (2022).

[10] Lee, M., Lee, G. J., Jang, H. J., Joh, E., Cho, H., Kim, M. S., Kim, H. M., Kang, K. M., Lee, J. H., Kim, M., et al., "An amphibious artificial vision system with a panoramic visual field," *Nature Electronics* **5**(7), 452–459 (2022).

[11] Barlow, H. B., "The size of ommatidia in apposition eyes," *Journal of Experimental Biology* **29**(4), 667–674 (1952).

[12] Land, M. F., "Visual acuity in insects," *Annual Review of Entomology* **42**(1), 147–177 (1997). doi: 10.1146/annurev.ento.42.1.147.

[13] Alkon, D. L., [*Neural principles in vision*], Springer Science Business Media (2012).

[14] Donoho, D. L., "Compressed sensing," *IEEE Transactions on information theory* **52**(4), 1289–1306 (2006).

[15] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences* **2**(1), 183–202 (2009).

[16] Chambolle, A. and Pock, T., "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision* **40**(1), 120–145 (2011).

[17] Yang, J. and Zhang, Y., "Alternating direction algorithms for l1-problems in compressive sensing," *SIAM journal on scientific computing* **33**(1), 250–278 (2011).

[18] Aharon, M., Elad, M., and Bruckstein, A., "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing* **54**(11), 4311–4322 (2006).

[19] Candes, E. J. and Tao, T., "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE transactions on information theory* **52**(12), 5406–5425 (2006).

[20] Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., and Ashok, A., "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 449–458 (2016).

[21] Palangi, H., Ward, R., and Deng, L., "Distributed compressive sensing: A deep learning approach," *IEEE Transactions on Signal Processing* **64**(17), 4504–4518 (2016).

[22] Jin, K. H., McCann, M. T., Froustey, E., and Unser, M., "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing* **26**(9), 4509–4522 (2017).

[23] Mousavi, A. and Baraniuk, R. G., "Learning to invert: Signal recovery via deep convolutional networks," in [*2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*], 2272–2276, IEEE (2017).

[24] Kim, C., Park, D., and Lee, H.-N., "Compressive sensing spectroscopy using a residual convolutional neural network," *Sensors* **20**(3), 594 (2020).

[25] Kitamura, Y., Shogenji, R., Yamada, K., Miyatake, S., Miyamoto, M., Morimoto, T., Masaki, Y., Kondou, N., Miyazaki, D., and Tanida, J., "Reconstruction of a high-resolution image on a compound-eye image-capturing system," *Applied Optics* **43**(8), 1719–1727 (2004).

[26] Floreano, D., Pericet-Camara, R., Viollet, S., Ruffier, F., Brückner, A., Leitel, R., Buss, W., Menouni, M., Expert, F., and Juston, R., "Miniature curved artificial compound eyes," *Proceedings of the National Academy of Sciences* **110**(23), 9267–9272 (2013).

[27] Jethi, A. K., Murugesan, B., Ram, K., and Sivaprakasam, M., "Dual-encoder-unet for fast mri reconstruction," in [*2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*], 1–4, IEEE (2020).

[28] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).