

Tensor-based Feature Extraction for Pupil Recognition in Cataract Surgery

Binh Duong Giap, *Member, IEEE*, Karthik Srinivasan, Ossama Mahmoud, Shahzad Ihsan Mian, Bradford Laurence Tannen, and Nambi Nallasamy, *Member, IEEE*

Abstract—Cataract surgery remains the definitive treatment for cataracts, which are a major cause of preventable blindness worldwide. Adequate and stable dilation of the pupil are necessary for the successful performance of cataract surgery. Pupillary instability is a known risk factor for cataract surgery complications, and the accurate segmentation of the pupil from surgical video streams can enable the analysis of intraoperative pupil changes in cataract surgery. However, pupil segmentation performance can suffer due to variations in surgical illumination, obscuration of the pupil with surgical instruments, and hydration of the lens material intraoperatively. To overcome these challenges, we present a novel method called tensor-based pupil feature extraction (TPFE) to improve the accuracy of pupil recognition systems. We analyzed the efficacy of this approach with experiments performed on a dataset of 4,560 intraoperative annotated images from 190 cataract surgeries in human patients. Our results indicate that TPFE can identify features relevant to pupil segmentation and that pupil segmentation with state-of-the-art deep learning models can be significantly improved with the TPFE method.

I. INTRODUCTION

Cataract surgery addresses the pathologic clouding of the natural lens of the eye through removal of the clouded lens material and replacement with an artificial intraocular lens (IOL) implant. It is one of the most commonly performed surgeries worldwide and essential to addressing preventable blindness. There were more than 20 million cataract surgeries performed in the world in 2015, of which 3.6 million cases were in the United States of America and more than 4.2 million cases were in the European Union [1]. Cataracts most commonly occur as a natural part of the aging process, but can also occur due to trauma, medication side effects, and metabolic disorders. Despite improvements in cataract surgery technology and the phacoemulsification procedure, cataracts are one of the leading causes of blindness worldwide.

Cataracts are accessed surgically by passing instruments through the pupil, which is the natural aperture in the iris tissue. The pupil is pharmacologically dilated at the time of surgery, but pupillary instability during surgery is a major risk factor for cataract surgery complications. As such, pupil recognition in cataract surgery is an essential task that allows

surgeons to understand changes in pupil morphology during surgery. However, recognition performance is affected by many factors occurring during cataract surgery, such as variations in illumination, obscuration by surgical instruments, and hydration of cataractous lens material altering the appearance of the pupil itself. In the recent past, with the rapid development of deep learning, studies [2] – [5] proposed methods for pupil recognition in various scenarios, predominantly outside of the context of cataract surgery.

Lee *et al.* [2] applied UNet model to recognize and separate pupil from nystagmography video achieve accurate trajectory of pupil for diagnosis of various vestibular disorders. The group showed the potential for applying deep learning methods to diagnosing vestibular disorders since the UNet model achieved 94.85% dice coefficient. In other scenarios, Han *et al.* [3] proposed an indirect use of CNN for pupil center detection. Accordingly, the pupil region is first segmented by CNN models and the center of the segmented region is then determined. The authors demonstrated that the proposed method outperforms other conventional methods that directly find the pupil center as a regression result. Gowroju *et al.* [4] presented a modified UNet model to improve the accuracy of pupil segmentation in non-surgical images. The proposed model achieved a notable accuracy when compared to the original UNet model while consuming less time.

In the field of cataract surgery, Sokolova *et al.* [5] trained and tested a Mask R-CNN on a small (82-image) dataset of cataract surgery frames and showed moderate performance in the pupil segmentation task. This study was limited by the size of its dataset and the need for analysis of performance across varying phases of surgery and in the context of pupil obscuration by a variety of surgical instruments.

In this paper, we propose a novel method called tensor-based pupil feature extraction (TPFE) for pupil recognition in cataract surgery using deep learning models. The proposed method is designed to extract features of the pupil region effectively with the goal of utilizing TPFE's feature-rich output as input to pupil recognition systems. In order to study its effectiveness, we undertook experiments to evaluate the segmentation performance of a set of deep learning models with and without TPFE pre-processing. In the subsequent

This work was supported in part by the GME Innovations Fund (N.N., B.T.), The Doctors Company Foundation (N.N., B.T.), NIH K12EY022299 (N.N.), and Fogarty/NIH Grant D43TW012027 (K.S.).

B. D. Giap, S. I. Mian, and B. L. Tannen are with the Department of Ophthalmology & Visual Sciences, University of Michigan, 1000 Wall St., Ann Arbor, MI 48105, USA (e-mail: gbinhduo@med.umich.edu; smian@med.umich.edu; btannen@med.umich.edu).

K. Srinivasan is with the Department of Vitreo Retinal, Aravind Eye Hospital, Chennai, Tamil Nadu 600077, India (e-mail: karthiksrinivasan@aravind.org).

O. Mahmoud is with Department of Ophthalmology and Visual Sciences, University of Michigan, 1000 Wall St., Ann Arbor, MI 48105, USA, and Wayne State University School of Medicine, 540 E Canfield St., Detroit, MI 48201, USA (e-mail: ossamam@med.umich.edu).

N. Nallasamy (Corresponding author) is with the Department of Computational Medicine & Bioinformatics, 100 Washtenaw Ave., Ann Arbor, MI 48109, USA and the Department of Ophthalmology and Visual Sciences, University of Michigan, 1000 Wall St., Ann Arbor, MI 48105, USA (e-mail: nnallas@med.umich.edu).

sections, we describe the TPFE method, report its impact on pupil segmentation performance in our experiments, and consider potential implications and future directions for this work.

II. METHOD

A. Tensor-based Pupil Image Representation

A tensor is defined as a multilinear mapping over a set of vector spaces [6]. Tensors represent a higher-order generalization of vectors and are commonly utilized for multidimensional data representation. Tensors are widely used to represent videos, color images, and hyperspectral images [7]-[9] through various data arrangements. In order to ensure the clarity of the notation in this study, we first present the notation and terminology related to tensors used in this paper [10], [11].

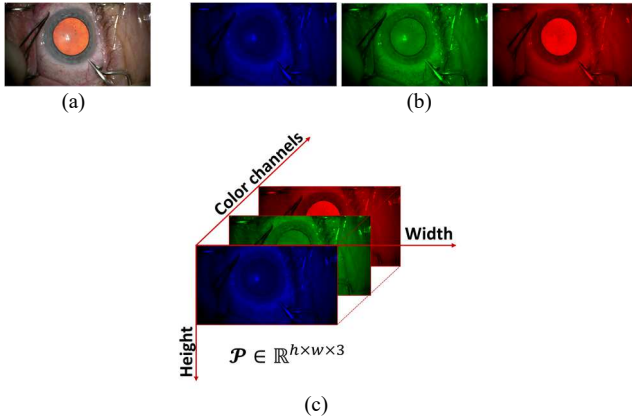


Fig. 1. Tensor-based pupil image construction. (a) A given color pupil image $I \in \mathbb{R}^{h \times w}$; (b) Blue, green, and red color channels I^c , $c \in \{b, g, r\}$ of (a); and (c) The third-order tensor $\mathcal{P} \in \mathbb{R}^{h \times w \times 3}$ of (a).

i) An N^{th} -order tensor denoted as \mathcal{P} is formally defined as an element of the tensor product of N vector spaces, where the order N of a tensor is the number of dimensions and $\mathcal{P} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$.

ii) *Fibers* are the higher-order analog of matrix rows and columns. A fiber of tensor \mathcal{P} can be determined by keeping an index, while fixing others. A third-order tensor \mathcal{P} involves column, row, and tube fibers denoted as $p_{:,m,l}$, $p_{n,:}$, and $p_{nm,:}$, respectively.

iii) *Slices* are two-dimensional portions of a tensor. A slice is defined by fixing all indices but two. A third-order tensor \mathcal{P} contains slices denoted as $P_{n,:}$, $P_{:,m}$, and $P_{:,l}$, which are called horizontal, lateral, and frontal slices, respectively.

iv) A tensor $\mathcal{P} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ has N ways to transform its data into a matrix called mode- k unfolding of tensor \mathcal{P} or tensor matricization. It is denoted by $P_{(k)}$, where $k \in \{1, 2, \dots, N\}$.

v) The mode- k product is the result of the multiplication of a tensor by a matrix in mode- k . The mode- k product $\mathcal{P} \times_k U$ of a tensor $\mathcal{P} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with a matrix $U \in \mathbb{R}^{J \times I_k}$ is a tensor of size $(I_1 \times \dots \times I_{k-1} \times J \times I_{k+1} \times \dots \times I_N)$. Therefore, we have

$$(\mathcal{P} \times_k U)_{i_1 \dots i_{k-1} j i_{k+1} \dots i_N} = \sum_{i_k} x_{i_1 i_2 \dots i_N} u_{j i_k}. \quad (1)$$

In this study, to effectively exploit the relationship among color channels and spatial information, we construct a third-order tensor with three frontal slices denoting three color channels of a given color pupil image. In particular, given a color pupil image denoted $I \in \mathbb{R}^{h \times w}$ of size $h \times w$, it is split into I^b, I^g , and I^r , which correspond to the three-color channel images in RGB color space. A third-order tensor denoted as $\mathcal{P} \in \mathbb{R}^{h \times w \times 3}$ sized $h \times w \times 3$ is then constructed by adopting each color channel I^c , $c \in \{b, g, r\}$ to be a frontal slice $P_{:,l}$, as depicted in Fig. 1.

B. High-Order Singular Value Decomposition of a Color Pupil Image

A given N -order tensor $\mathcal{P} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ can be expressed as a linear combination of the outer products in different modes. Specifically, the high-order singular value decomposition (HOSVD) expresses tensor $\mathcal{P} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ as follows:

$$\mathcal{P} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 \dots \times_N U^{(N)}. \quad (2)$$

It can also be expressed elementwise as:

$$p_{i_1 i_2 \dots i_N} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \dots \sum_{j_N=1}^{I_N} s_{j_1 j_2 \dots j_N} u_{i_1 j_1}^{(1)} u_{i_2 j_2}^{(2)} \dots u_{i_N j_N}^{(N)}, \quad (3)$$

where $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ denotes the core tensor, which has the same size with the given tensor \mathcal{P} . $U^{(k)} \in \mathbb{R}^{I_k \times I_k}$, $k = \{1, 2, \dots, N\}$ are matrices containing the left singular vectors of the mode- k unfolding matrices $P_{(k)}$ of tensor \mathcal{P} . These matrices $U^{(k)}$ are also called inverse factors and determined through singular value decomposition of mode- k unfolding matrices $P_{(k)}$ as follows:

$$P_{(k)} = U^{(k)} \Sigma^{(k)} V^{(k)T}, \quad (4)$$

where $\Sigma^{(k)}$ denotes the singular value matrices and $V^{(k)}$ contains the right singular vectors of $P_{(k)}$. Furthermore, the core tensor denoted \mathcal{S} is determined by:

$$\mathcal{S} = \mathcal{P} \times_1 U^{(1)T} \times_2 U^{(2)T} \times_3 \dots \times_N U^{(N)T}. \quad (5)$$

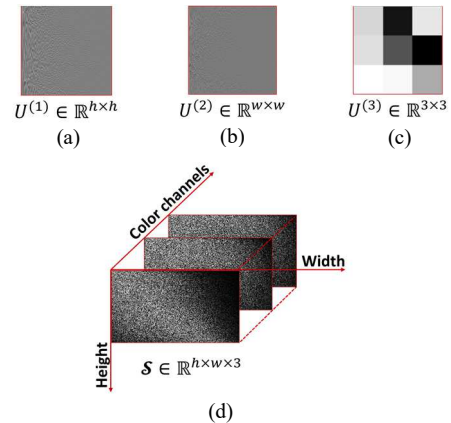


Fig. 2. The results of HOSVD decomposition of a third-order tensor $\mathcal{P} \in \mathbb{R}^{h \times w \times 3}$ in Fig. 1(c). (a) Inverse factor $U^{(1)}$; (b) Inverse factor $U^{(2)}$; (c) Inverse factor $U^{(3)}$; and (d) Core tensor \mathcal{S} .

As with the given tensor, \mathcal{P} , mode- k unfolding matrices $S_{(k)}$ of the core tensor \mathcal{S} can be obtained by:

$$S_{(k)} = \sum^{(k)} V^{(k)} (U^{(N)} \otimes \dots \otimes U^{(k+1)} \otimes U^{(k-1)} \otimes \dots \otimes U^{(1)}), \quad (6)$$

where \otimes denotes the Kronecker product of two matrices. At the end, a mode- k unfolding matrix $P_{(k)}$ of the given tensor \mathcal{P} can be reconstructed by:

$$P_{(k)} = U^{(k)} S_{(k)} (U^{(N)} \otimes \dots \otimes U^{(k+1)} \otimes U^{(k-1)} \otimes \dots \otimes U^{(1)})^T. \quad (7)$$

In this study, we first construct a third-order tensor $\mathcal{P} \in \mathbb{R}^{h \times w \times 3}$ of a given pupil image $I \in \mathbb{R}^{h \times w}$ as discussed in subsection II-A. By using HOSVD, \mathcal{P} is then decomposed into multiple components: three inverse factors denoted $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$ and a core tensor \mathcal{S} , as depicted in Fig. 2. HOSVD decomposition is briefly presented in Algorithm 1.

Algorithm 1: HOSVD

Input: A Color Pupil Image $I \in \mathbb{R}^{h \times w}$

Output: $U^{(1)}, U^{(2)}, U^{(3)}$, and \mathcal{S}

1. $I^c, c \in \{b, g, r\} \leftarrow \text{Split}(I)$;
 2. $\mathcal{P} \in \mathbb{R}^{h \times w \times 3} \leftarrow \text{Reshape}(I^c)$;
 3. $\text{get}(U^{(k)}, \Sigma^{(k)}, V^{(k)})$ via Eq. (4);
 4. $\text{get}(S_{(k)})$ via Eq. (6);
 5. $\mathcal{S} \leftarrow \text{Reshape}(S_{(k)})$
 6. Return $U^{(1)}, U^{(2)}, U^{(3)}, \mathcal{S}$.
-

C. Tensor-based Pupil Feature Extraction

In this study, we investigated the information carried by elements of inverse factor $U^{(3)}$ and sought to emphasize elements of $U^{(3)}$ seen to carry pupil-related information and de-emphasize elements seen to carry noise and background data. The feature-extracted image was then obtained by inverting HOSVD with the original \mathcal{S} , $U^{(1)}$, $U^{(2)}$ and a modified $U_m^{(3)}$.

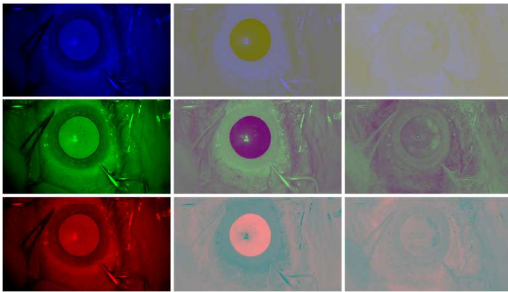


Fig. 3. The information of Fig. 1(a) carried by nine elements of inverse factor $U^{(3)} \in \mathbb{R}^{3 \times 3}$. The position of the image corresponds to the location of the investigated element in $U^{(3)}$.

The results of the mentioned analysis are depicted in Fig. 3. As seen, three elements in the first column of $U^{(3)}$ can reconstruct much of the three color channels, which are blue, green, and red, of the given input image. Accordingly, both pupil and interference information are mostly present and would be recovered from these elements. On the other hand, information specific to the pupil region appears to be present within the elements of the second column of the inverse factor $U^{(3)}$. As seen in Fig. 3, the third column of $U^{(3)}$ contains primarily interference information.

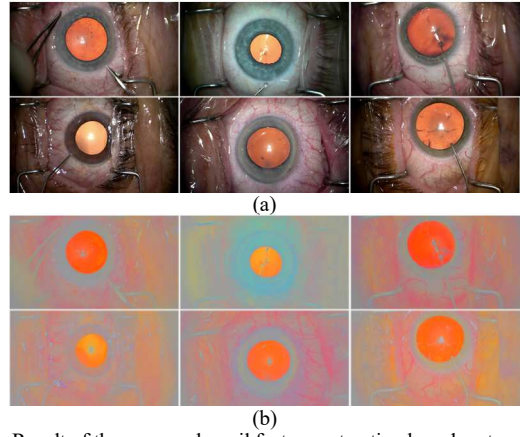


Fig. 4. Result of the proposed pupil feature extraction based on tensor. (a) Original images from different cataract surgeries. (b) Feature images of (a) by the proposed method.

In order to highlight the pupil region and eliminate the interference information in the given image, the elements of the first column were de-emphasized by subtracting off the mean of three elements. The proposed feature extraction algorithm is briefly presented in Algorithm 2 and the result of the proposed method is depicted in Fig. 4.

Algorithm 2: TPF

Input: Pupil Image $I \in \mathbb{R}^{h \times w}$

Output: Feature Image $F \in \mathbb{R}^{h \times w}$

1. $\{U^{(1)}, U^{(2)}, U^{(3)}, \mathcal{S}\} \leftarrow \text{Algorithm 1}(\mathcal{P})$;
 2. $\mu \leftarrow \text{mean}\{U^{(3)}(1,1), U^{(3)}(2,1), U^{(3)}(3,1)\}$;
 3. $U^{(3)}(i,1) = (U^{(3)}(i,1) - \mu), i \in \{1,2,3\}$;
 4. $\text{get}(P_{(k)}^{(u)}), k \in \{1,2,3\}$ via Eq. (7).
 5. $\mathcal{P}^{(u)} = \{P_{::l,l=1,2,\dots,12}^{(u)}\} \leftarrow \text{Reshape}(P_{(k)}^{(u)})$;
 6. $F^c, c \in \{b, g, r\} \leftarrow \text{Reshape}(\mathcal{P}^{(u)})$;
 7. $F \leftarrow \text{Merge}(F^b, F^g, F^r)$;
 8. Return F .
-

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Dataset

To evaluate the performance of the proposed method, we adopted a dataset consisting of 4,560 intraoperative images of size 480×270 pixels from 152 cataract surgeries, which were performed by surgeons at Kellogg Eye Center, University of Michigan from 2020 to 2021. The pupil region on the images was manually annotated using MATLAB version R2022a. The study was approved (HUM00160950) by the Michigan Medicine IRB in May of 2019 and was carried out in accordance with the tenets of the Declaration of Helsinki.

B. Pupil Recognition by Deep Learning Models

In order to assess performance of the proposed framework in improving the accuracy of pupil recognition, we adopted six state-of-the-art deep learning models: Xception [12], HR-Net [13], DeepLabV3+ [14], FPN [15], UNet [16], and LinkNet [17], each with VGG16 [18] as the backbone network for our experiments. The selected models were all pre-trained on ImageNet dataset [19] to reduce training time.

TABLE I
SEMANTIC SEGMENTATION RATE (%) OF DEEP LEARNING MODELS ON VALIDATION SET

NETWORK ARCHITECTURE	Precision (%)	Recall (%)	IoU (%)	Dice (%)
Xception	99.28	81.92	81.75	89.29
Xception + TPFE	98.97	85.71	85.37	91.50
HR-Net	99.24	86.88	86.72	92.36
HR-Net + TPFE	99.31	89.36	89.16	93.84
DeepLabV3+	99.37	86.71	86.58	92.39
DeepLabV3+ + TPFE	99.13	89.49	89.21	93.95
FPN	99.38	88.78	88.62	93.53
FPN + TPFE	99.45	88.97	89.20	93.88
UNet	99.33	87.79	87.62	92.96
UNet + TPFE	99.20	89.01	88.79	93.58
LinkNet	99.35	88.35	88.18	92.30
LinkNet + TPFE	99.10	89.50	89.21	93.83

The dataset was randomly split into training and validation subsets consisting of 75% and 25% of the dataset, respectively. Thus, 3,420 images from 114 videos were included for training, 1,140 images from 38 videos were included for validation. Separate instances of each deep learning model were trained using either the original images or the feature images. A batch size of 8 and a learning rate of 10^{-4} were fixed for training. We trained each for a maximum of 100 epochs corresponding to 42,700 training iterations. To this end, precision, recall, Intersection over Union, and Dice coefficient of all six models on the validation set were compared with and without use of the feature-extracted images. The results of the experiments are shown in Table I. As seen, the accuracy of the deep learning models is considerably improved when combined with the proposed feature extraction TPFE method. Each of the 6 models tested demonstrated a statistically significant improvement in performance ($p \ll 0.05$ on Wilcoxon [20] signed rank testing) when utilizing the method.

IV. CONCLUSION

In this paper, we have proposed a novel feature extraction method named TPFE for pupil recognition systems in cataract surgery using deep learning models. The effectiveness of TPFE was comprehensively demonstrated through a set of experiments using state-of-the-art deep learning models. The results reveal that the accuracy of pupil recognition systems in cataract surgery can be significantly improved through utilization of the proposed TPFE method.

ACKNOWLEDGMENT

This work was supported in part by the GME Innovations Fund (N.N., B.T.), The Doctors Company Foundation (N.N., B.T.), NIH K12EY022299 (N.N.), and Fogarty/NIH Grant D43TW012027 (K.S.).

REFERENCES

- [1] A. Grzybowski, "Recent developments in cataract surgery," *Ann. Transl. Med.*, vol. 8, no. 22, pp. 1540-1545, Nov. 2020.
- [2] Y. Lee, S. Lee, S. Jang, H. J. Wang, Y. J. Seo, and S. Yang, "Pupil detection and segmentation for diagnosis of Nystagmus with U-Net," in *2022 International Conference on Electronics, Information, and Communication (ICEIC)*, Jeju, Korea, Republic of, 2022, pp. 1-2.
- [3] S. Y. Han, H. J. Kwon, Y. Kim, and N. I. Cho, "Noise-robust pupil center detection through CNN-based segmentation with shape-prior loss," *IEEE Access*, vol. 8, pp. 64739-64749, 2020.

- [4] S. Gowroju, Aarti, and S. Kumar, "Robust Pupil Segmentation using UNET and Morphological Image Processing," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, Cairo, Egypt, 2021, pp. 105-109.
- [5] N. Sokolova, M. Taschwer, S. Sarny, D. Putzgruber-Adamitsch, and K. Schoeffmann, "Pixel-Based Iris and Pupil Segmentation in Cataract Surgery Videos Using Mask R-CNN," in *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, Iowa City, IA, USA, 2020, pp. 1-4.
- [6] W. Hackbusch, "Tensor spaces and numerical tensor calculus," in *Springer Series in Computational Mathematics*, Springer, Germany, 2012.
- [7] J. A. Bengua, H. N. Phien, H. D. Tuan, and M. N. Do, "Efficient Tensor Completion for Color Image and Video Recovery: Low-Rank Tensor Train," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2466-2479, May 2017.
- [8] T. N. Le, D. B. Giap, J. -W. Wang, and C. -C. Wang, "Tensor-Compensated Color Face Recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 3339-3354, 2021.
- [9] D. B. Giap, T. Ngoc Le, J. -W. Wang, and C. -N. Wang, "Wavelet Subband-Based Tensor for Smartphone Physical Button Inspection," *IEEE Access*, vol. 9, pp. 107399-107415, 2021.
- [10] T. G. Kolda and B. W. Bader, "Tensor decomposition and applications," *SIAM Review*, vol. 51, no. 3, pp. 455-500, Sep. 2009.
- [11] T. G. Kolda, "Multilinear operators for higher-order decompositions," *Tech. Rep. No. SAND2006-2081*, Sandia National Laboratories, 2006.
- [12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251-1258.
- [13] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep High-Resolution Representation Learning for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349-3364, Oct. 2021.
- [14] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Proc. European Conference on Computer Vision*, 2018, pp. 801-818.
- [15] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936-944.
- [16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241.
- [17] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *IEEE Visual Communications and Image Processing*, 2017, pp. 1-4.
- [18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1-14.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248-255.
- [20] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, no.1, vol. 6, pp. 80-83, 1945.