High-performance panoramic annular lens design for real-time semantic segmentation on aerial imagery

Jia Wang[a], **Kailun Yang**[b], **Shaohua Gao**[a], **Lei Sun**[a], **Chengxi Zhu**[a], **Kaiwei Wang**[a], **and Jian Bai**[a,*]

[a]Zhejiang University, College of Optical Science and Engineering, Hangzhou, China
[b]Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract. As the visual perception window of the drone system, the lens provides great help for obtaining visual information, detection, and recognition. However, traditional lenses carried on drones cannot have characteristics of a large field of view (FoV), small size, and low weight at the same time. To meet the above requirements, we propose a panoramic annular lens (PAL) system with 4K high resolution, a large FoV of (30 deg to 100 deg) × 360 deg, an angular resolution of 12.2 mrad of aerial perspective, and great imaging performance. We equip a drone system with our designed PAL to collect panoramic image data at an altitude of 100 m from the track and field and obtain the first drone-perspective panoramic scene segmentation dataset Aerial-PASS, with annotated labels of track and field. We design an efficient deep architecture for aerial scene segmentation. Trained on Aerial-PASS, the yielded model accurately segments aerial images. Compared with the ERF-PAPNet and SwiftNet semantic segmentation networks, the network we adopted has higher recognition accuracy with the mean IoU greater than 86.30%, which provides an important reference for the drone system to monitor and identify specific targets in real-time in a large FoV. © *2022 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.OE.61.3.035101]

Keywords: panoramic annular lens; semantic segmentation; aerial imagery; drone system.

Paper 20211441G received Dec. 9, 2021; accepted for publication Feb. 15, 2022; published online Mar. 1, 2022.

1 Introduction

Drones have characteristics of flexibility, low cost, and lightweight, so they are widely used in military, environmental surveillance, remote sensing, and other fields.[1] The camera on the drone provides an indispensable visual aid for target recognition and monitoring. People's demand for wide-angle and high-resolution images makes lens design more complex. In general, to obtain a panoramic image, drones need a rotating servo mechanism to control the position and attitude of the lens, which causes the whole system to be complex and heavy. Image splicing will bring about problems such as exposure difference and stitching discontinuity, which will reduce the accuracy of recognition and monitoring.[2] Therefore, a more stable and simple optical system is desirable to increase the detection capability of the drone system.

Optical systems with a large field of view (FoV), such as fisheye lens,[3] catadioptric panoramic optical system,[4] and PAL imaging system,[5] can achieve real-time staring imaging without stitching. Compared with the general catadioptric panoramic system and fisheye lens, the PAL system shows clear advantages, which has a compact structure, better image quality, and smaller negative distortion. The PAL system was first proposed by Greguss.[6] In recent decades, the PAL system has received extensive attention and has been greatly improved. Many researchers are committed to increasing the focal length, the FoV, and the resolution parameters of the PAL system.[7–9] In addition, dual-channel solutions based on the dichroic film, polarizer, semitransparent and semi-reflective mirror, etc. were proposed to tackle the inherent blind area problem,[10–12] which improved the imaging performance[13] and increased the FoV[14,15] of the PAL. Compared with the traditional

*Address all correspondence to Jian Bai, Bai@zju.edu.cn

PAL lens, our designed PAL system can provide a large FoV, a high resolution, compact structure, lightweight, and better image quality, which is suitable for aerial scene perception.

Based on a fully convolutional network, images can be semantically segmented at the pixel level in an end-to-end fashion.[16] The higher the accuracy of the network, the higher its computational complexity, and it is not suitable for auto-driving cars and aerial scene segmentation. To achieve fast and accurate semantic segmentation, a large number of lightweight networks have emerged, such as SwiftNet,[17] AttaNet,[18] RFNet,[19] and DDRNet.[20] In the previous work of our team, Yang et al.[21] proposed a general semantic segmentation framework based on a panoramic annular system, which verified that a robust panoramic segmentation is feasible. The resolution of the PAL used in Yang et al.'s work is only 2K. However, the resolution of the PAL system proposed in this paper is 4K. The higher the resolution, the clearer and more delicate the imaging effect. In addition, a high-performance panoramic annular lens (PAL) system is desired for real-world scene perception. In this paper, we propose an Aerial-PASS system with an efficient deep architecture designed for aerial scene segmentation and explore the superiority of the panoramic view in security monitoring applications.

This paper is organized as following. Section 2 presents the optical design of the PAL system on drones. Section 3 introduces the system prototype and database establishment. Section 4 shows semantic segmentation test results. Section 5 summarizes our work in this paper.

2 Optical Design of the PAL

2.1 *Principle of the PAL System*

The PAL system follows the imaging principle of flat cylinder perspective (FCP),[22] that is, its FoV is similar to a cylindrical surface, and the image is in the flat annular areas shown in Fig. 1. The PAL system is composed of a block-structured panoramic front unit (BSPFU) and a relay lens system (RLS). The BSPFU can turn the light from the side large FoV 360 deg around the optical axis into a small FoV. After the light enters the RLS, it images on the two-dimensional annular image plane.

The imaging law of general optical systems follows the principle of object-image similarity. However, for large FoV imaging systems, the principle of object-image similarity is no longer applicable. As the FoV (ω) increases, when it approaches 90 deg, tan ω will tend to infinity,

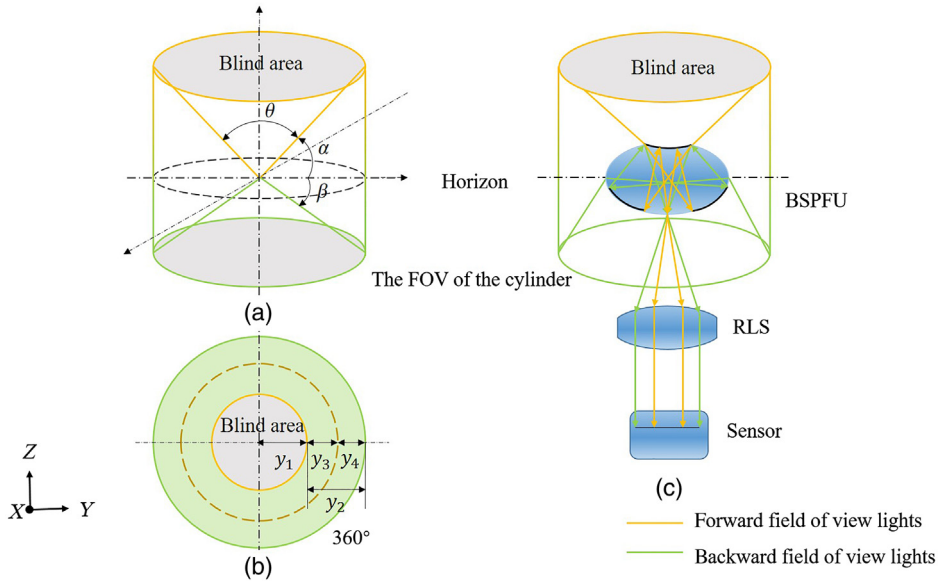


Fig. 1 Principle of FCP and ray-tracing in PAL. (a) Cylindrical imaging model. (b) Schematic diagram of a two-dimensional image plane. The green area is the imaging part, and the gray area is the blind area; (c) PAL composition and imaging schematic model.

which means that the sensor will not be able to receive the infinite image formed by the system. Therefore, only by introducing negative distortion into the optical design can the image size be controlled. The commonly used characterization method is F-Theta distortion, to control the image height, as shown in Eq. (1):

$$y = f' \cdot \omega, \tag{1}$$

where y represents the image height, f' represents the focal length, and ω represents the angle of view. Therefore, the range of the blind area and imaging part on the detector are determined by the FoV and the focal length, as shown in Fig. 1:

$$\begin{cases} y_1 = f' \cdot \theta \\ y_2 = y_3 + y_4 \\ y_3 = f' \cdot \alpha \\ y_4 = f' \cdot \beta \end{cases}, \tag{2}$$

where y_1 represents the radius of the circular blind area, y_2 represents the height of the annular imaging area, y_3 represents the height of the annular imaging area formed by the α FoV, y_4 represents the height of the annular imaging area formed by the β FoV, θ is the blind spot angle, α is the field angle above the horizontal axis, and β is the field angle below the horizontal axis.

The aberrations have a great influence on the PAL with field curvature, distortion, and chromatic aberration because of its special imaging method.[23] The special shape and material selection of the BSPFU play a crucial role in the correction of aberrations. The crown glass and flint glass glued together can correct chromatic aberrations well. In this work, we adopt BSPFU with the glued form of H-LAK8A and H-LAF50B. These two glass materials have a high refractive index and a high Abbe number, which can effectively deflect the optical path and reduce the chromatic aberration of the system. The follow-up lens group plays a vital role in the second imaging and correction of aberrations of the panoramic system. In this paper, the subsequent lens group is adopted with a deformed double Gaussian structure. The double Gaussian symmetric structure can well correct coma and curvature of field and reduce the impact of residual aberration on the overall image quality.

2.2 *Optical Design Process and Results*

Considering that the drone system has a light structure and can realize functions such as terrain survey and target recognition, the size, weight, and resolution of the lens are the focus of several indicators. The number of pixels of the sensor and the frame size determine the imaging capabilities of the lens. An image sensor is chosen as the design sample, whose resolution is $20.8\text{MP} - 5280(H) \times 3956(V)$, and the size of a single-pixel is $3.4\ \mu\text{m} \times 3.4\ \mu\text{m}$. Thus, the size of the image plane can be calculated to be $18.0\ \text{mm} \times 13.5\ \text{mm}$.

After determining the image sensor parameters, working spectral range, the incident FoV, size, and other technical parameters, it is necessary to clarify the specific steps of the optical design. From Eq. (1), it can be known that the focal length of the system can be determined according to the image height and the FoV. In this design, the image height is 13.5 mm, the maximum FoV is 100 deg, and the focal length can be calculated to be 3.86 mm. After careful consideration, the major design parameters for our PAL system were selected, as shown in Table 1.

The optical design software Zemax OpticStudio is used to design a miniaturized high-resolution PAL system according to the above design specifications. In the process of design optimization, we check the balance of various aberrations and design a system with good imaging quality by adjusting the structural parameters of the optical system in the merit function editor. The optimized optical structure is shown in Fig. 2. It is composed of 10 lenses in six groups, all of which are standard spherical surfaces. The designed PAL system has a total length of 79.5 mm, a maximum diameter of 44 mm, and a back focal length of 13.5 mm, and it can be easily applied to other sensors of similar resolution.

Table 1 Specifications of the PAL system.

Parameter	Specification
Working spectrum	0.486 to 0.0656 μm
FoV	(30 deg to 100 deg) × 360 deg
F-number	4.7
Effective focal length	3.86 mm
F-Theta distortion	<5%
Size	<80 mm (*L*) × 50 mm (*D*)

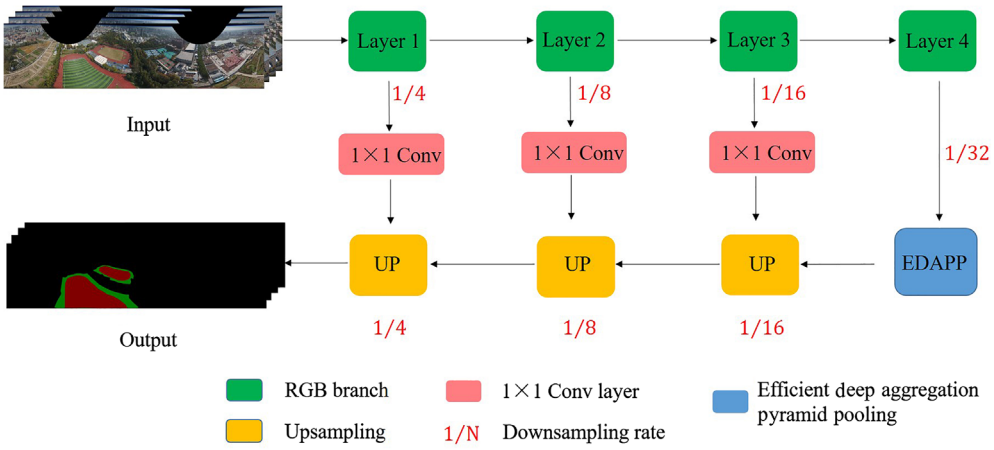


Fig. 2 The layout of the structure of the PAL system.

The whole PAL system has a telecentric design in the image space, and the chief ray angle incident on the image surface is less than 3 deg, which is beneficial to obtain uniform illumination of the image surface and high imaging quality. When the angle between the light and the image plane is too large, if the protective glass is not added in the design process, it will have a great impact on the image quality. To make the design more accurate, the filter of the detector is added at the end of the system structure. The material of the filter is H-K9L and the thickness is 0.8 mm.

2.3 *Image Quality Analysis*

The modulation transfer function (MTF), spot diagram, and distortion diagram characterize the image quality of the optical system. The single-pixel size of the sensor used is 3.4 μm, and the Nyquist frequency can be calculated to be 147 lp/mm. As shown in Fig. 3, the MTF curves of all the fields of view of the PAL system are greater than 0.4 at the cut-off frequency, indicating that the system has good imaging quality.

The spot diagram of the five FoV of the system, as shown in Fig. 4. The black circle is the Airy disk. The maximum root mean square radius under 100 deg FoV is 2.519 μm, which is smaller than the single-pixel size of the sensor, indicating that wide-FoV panoramas can be clearly imaged.

The field curvature and F-Theta distortion are shown in Fig. 5. It shows that the field curvature is less than 0.1 mm, and the distortion is less than 1% in the entire FoV. The distortion at the edge and center of the captured image is small and uniform.

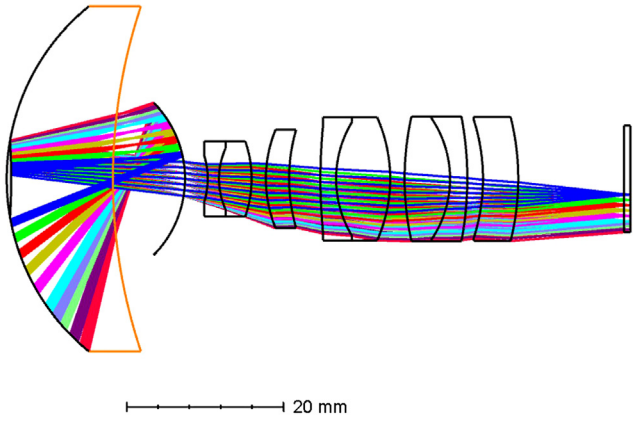


Fig. 3 MTF diagram.

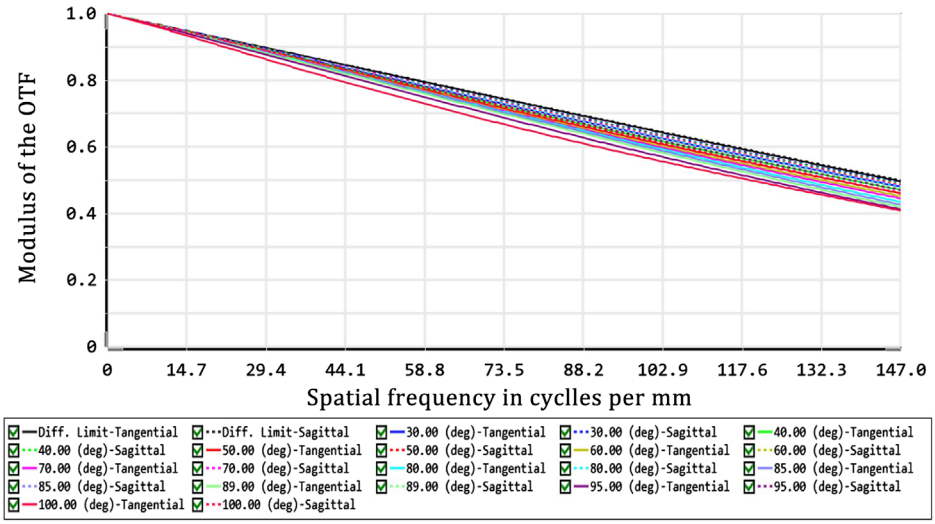


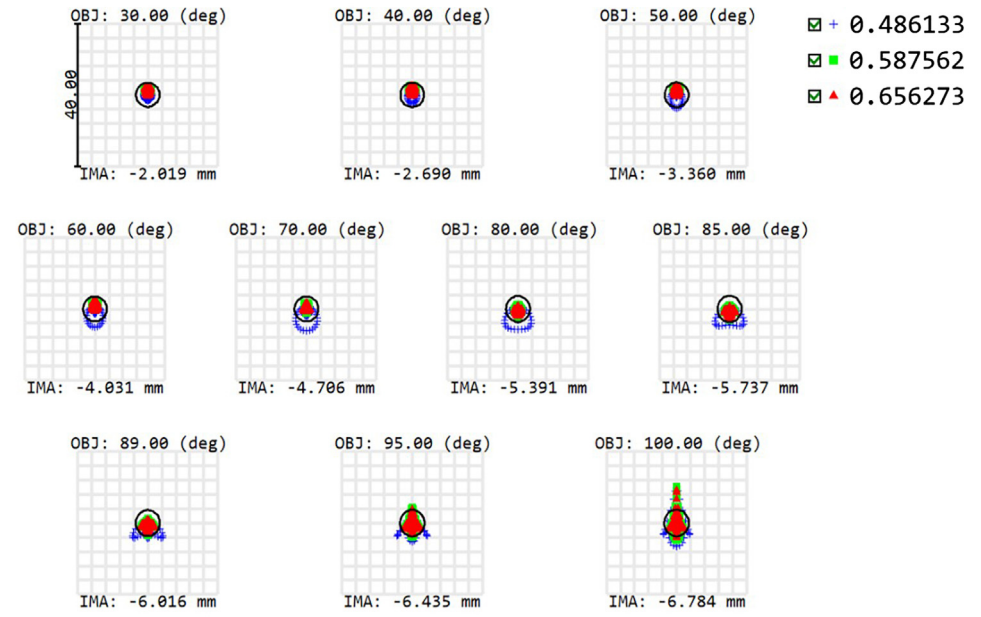
Fig. 4 Spot diagram.

The relative illumination reflects the brightness and darkness distribution of the system imaging results. The relative illumination is greater than 0.86, as shown in Fig. 6. It indicates that this system can maintain high energy and brightness both in the center and edge FoV.

2.4 *Tolerance Analysis*

A good image quality evaluation of the optical design does not mean that the performance of the system is excellent. Before using this system, the tolerance analysis is an indispensable step, which can evaluate the difficulty of system manufacturing and assembly. The tolerance ranges of surfaces and elements we set is shown in Table 2. The tolerance of material index and Abbe number are 0.001 and 1.

The evaluation standard is the average diffraction MTF, and the analysis method is the Monte Carlo method. After 20 sets of Monte Carlo iterations, the results are shown in Table 3. 90% of the MTF value is greater than 0.33 at the Nyquist frequency. The wide tolerance range and excellent analysis results allow the mechanical mechanism to be designed in a low-cost straight-tube package.



Units are μm. Airy Radius: 3.356 μm. Legend items refer to Wavelengths

Field	1	2	3	4	5	6	7	8	9	10
RMS radius :	1.110	1.214	1.225	1.217	1.304	1.473	1.531	1.553	1.701	2.519
GEO radius :	2.472	2.710	3.754	5.129	6.125	6.043	5.928	5.876	5.610	10.241

Scale bar : 40 Reference : Chief Ray

Fig. 5 Field curvature and F-Theta distortion diagram.

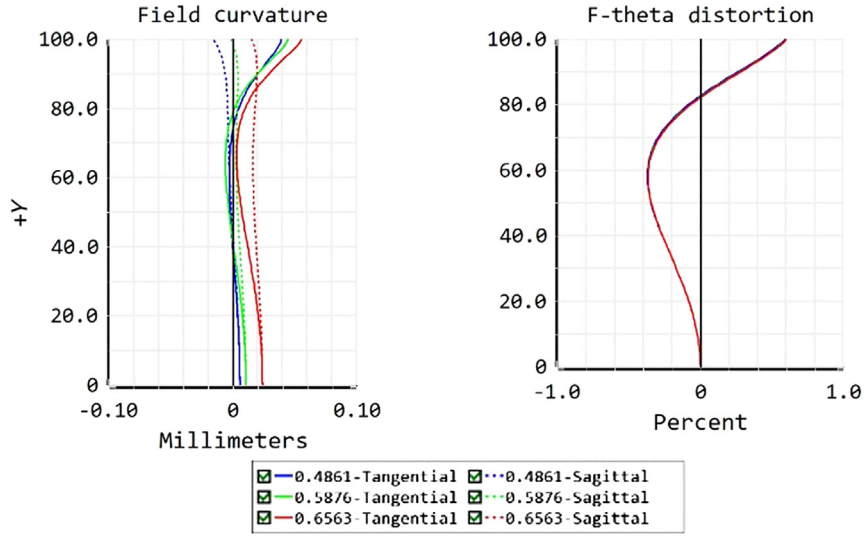


Fig. 6 Relative illumination.

3 Acquisition of Aerial Images

3.1 *PAL Prototype*

A good image quality evaluation of the optical design does not mean that the performance of the system is excellent. Before using this system, the tolerance analysis is an indispensable step, which can evaluate the difficulty of system manufacturing and assembly. The tolerances we give are relatively loose, with thickness tolerances of ± 0.02 mm, and decenter and tilt tolerances of

Table 2 Tolerance data.

Parameter	Surface tolerances	Element tolerances
Radius (fringes)	3	—
S + A irregularity (fringes)	0.3	—
Thickness (mm)	±0.02	—
Decenter *X* (mm)	±0.01	±0.01
Decenter *Y* (mm)	±0.01	±0.01
Tilt *X* (deg)	±0.02	±0.02
Tilt *X* (deg)	±0.02	±0.02

Table 3 Monte Carlo tolerance analysis results.

Probability of MTF@147 lp/mm	Value
90%	0.33143490
80%	0.34092705
50%	0.37283590
20%	0.40125212
10%	0.41765449

±0.01 mm. The evaluation standard is the average diffraction MTF, and the analysis method is the Monte Carlo method. After 20 sets of Monte Carlo iterations, the results show that 90% of the MTF curve is greater than 0.34 at the cut-off frequency. The mechanical structure of the PAL system is shown in Fig. 7. Since the aperture stop surface is located at the last transmission surface of the PAL block, it is not necessary to add a spacer to act as a mechanical stop. The drone system has strict requirements on the quality of the carried optical lens. The actual weight of this PAL is 140 g, which meets the load-bearing requirements of drones.

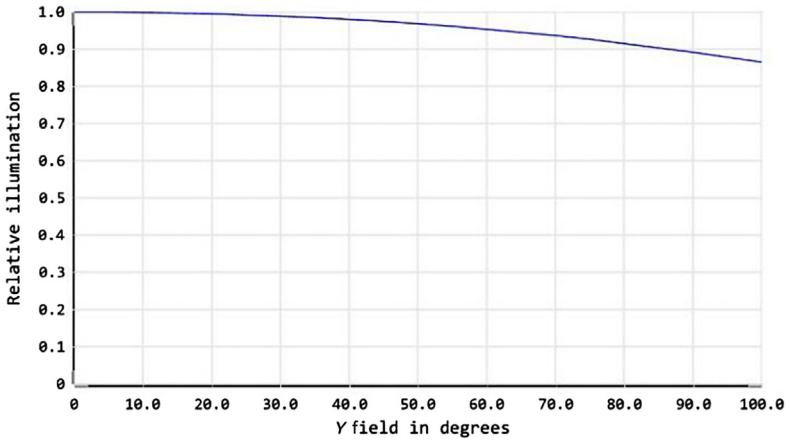


Fig. 7 Mechanical structure and lens prototype.

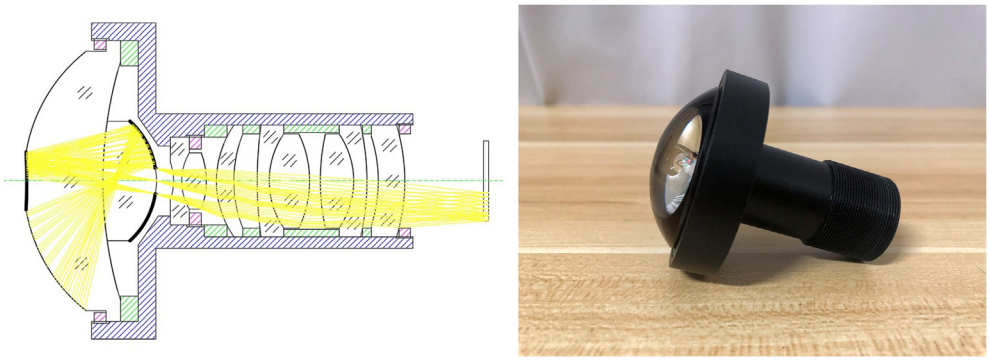


Fig. 8 The Aerial-PASS system and the data collection process.

3.2 *Establishment of the Database*

To verify the performance of our PAL system for semantic segmentation, we choose the flyable area on the campus-the track and field to collect data. We build the aerial panoramic dataset, which we term Aerial-PASS.

There are 10 race tracks in the standard track and field, and each track has a width of 1.22 m. The PAL system is held on DJI's drone system named "inspire 2," and the PAL block is placed vertically downward to cover a wider FoV on the ground. The stray light caused by the direct light source such as the sun is reduced, which is conducive to obtaining clearer image data. Since the drone has a limited flying height of 120 m, it is remotely controlled to collect image data in multiple directions at a height of 100 m above the ground. The experimental schematic diagram is shown in Fig. 8. The yellow area in the middle is the blind area, and the green area is the part of the FoV that can participate in imaging. The total FoV is 360 deg $\times$ 70 deg, which is divided into the FoV of 10 deg upward and 60 deg downward relevant to the horizon. From the collected original image data, it can be seen that the PAL can clearly distinguish a single runway at a height of 100 m from the ground, with an angular resolution of 12.2 mrad. To sum up, the advantages of the PAL that can be applied to semantic segmentation are as follows: (1) the panoramic camera has high resolution; (2) the vertical FoV is large; (3) the lens is light in weight and easy to carry.

4 Panoramic Semantic Segmentation

4.1 *Lightweight Semantic Segmentation Model*

For aerial scene understanding applications, the computation budget is often restricted. Thereby, we need to design a lightweight semantic segmentation model. In this paper, we design a lightweight U-Net-like model with multiscale receptive fields.[24] The proposed network architecture is shown in Fig. 9.

Fig. 9 The architecture of the proposed segmentation model.

To maintain a real-time inference speed, we adopt the lightweight ResNet-18[25] as the backbone of our model. The feature maps from each layer of ResNet are introduced to the upsampling module with a skip connection implemented via a 1×1 convolution. Inspired by the deep aggregation pyramid pooling module in DDRNet, to enlarge the receptive field of objects of different sizes, we add an efficient deep aggregation pyramid pooling module at the end of ResNet to ensure a multiscale receptive field for panoramic images. We replace 3×3 convolution with 1×3 and 3×1 convolution for better inference speed while maintaining a large receptive field. The upsampling module consists of three efficient upsampling blocks. The feature maps are upsampled and fused with the feature maps from the skip connections, via elementwise addition and a convolution layer. Through skip connection, the high-resolution feature maps of the target objects are sufficiently merged with the feature maps with rich semantic information, and the identification efficiency of the target objects has been improved.

4.2 *Panoramic Images Expansion*

We calibrated the PAL system using the interface provided by the omnidirectional camera toolbox.[26] Before training semantic segmentation models, the annular image collected by the PAL system is unfolded into a rectangular image. The expansion equations are described as follows:

$$\begin{cases} i = \frac{r - r_1}{r_2 - r_1} \times h \\ j = \frac{\theta}{2\pi} \times w \end{cases}, \tag{3}$$

where i and j denote the index of x, y and axis of the unfold image, respectively. r_1 and r_2 are the inner and outer radii of the annular image. h and w are the height and width of the expanded rectangular image, respectively. We unfolded the annular image to a rectangular image whose resolution is 2048×512. Figure 10 shows the unfolding process of the PAL image. Since the output format of the sensor is 16:9, a part of the top and bottom of the panoramic image is cropped. There are two semicircular black areas in the expanded image, but this does not affect the image clarity and resolution.

4.3 *Model Training*

We cut out 462 images from the panoramic video taken as our aerial image semantic segmentation dataset, which has been made publicly available to foster aerial panoramic scene

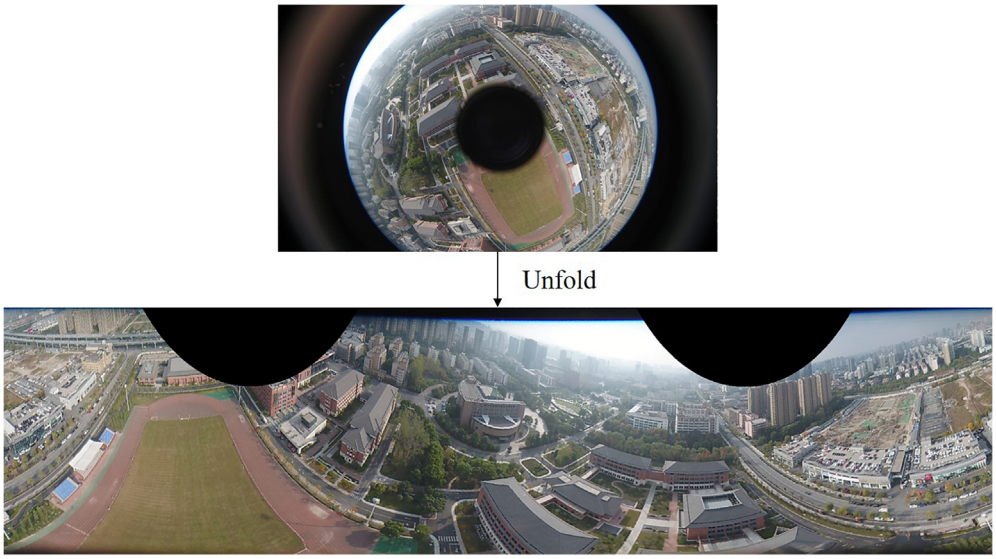


Fig. 10 The unfolding process of the PAL image.

segmentation.[27] Pixel-level labels were annotated in two critical classes for playground aerial scene understanding: the track and the central lawn. After that, we randomly chose 42 images as the test set.

We use Adam[28] for optimization, and the learning rate is 5×104. Cosine annealing learning rate adjustment strategy was adopted and the minimum value of the previous epoch is 5×104. The weight decay was set to 1×104. The ResNet-18 backbone was initialized with pretrained weights from ImageNet,[29] and the rest of the model was initialized with the Kaiming initialization method.[30] We update the pretraining parameters with four times smaller learning rate and weight decay rate. The model was trained for 100 epochs and the batch size was 6. We use the standard "intersection over union" (IoU) for evaluation:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{4}$$

where TP denotes true positive, FP denotes false positive, and FN denotes false negative at the pixel level.

4.4 *Semantic Segmentation Results*

Based on our collected data, we created a benchmark to compare our proposed network with other two competitive real-time semantic segmentation networks, which are the single-scale SwiftNet and ERF-PSPNet.[31] We keep the same training strategy for the three models and the same testing set of the database. Table 4 shows the performance comparison of the three networks on the test set. The network we adopted in our work outperforms the other two in all three categories and mean IoU. Figure 11 visualizes the panoramic aerial scene parsing results of three semantic segmentation networks.

Green represents the track and red represents the field. Here, we qualitatively evaluate the semantic segmentation of track and field, which verifies the reliable performance of our system

Table 4 Comparison of the IoU of our method and the other two networks on the aerial dataset.

Network	Track	Field	Others	Mean
ERF-PAPNet	64.16%	97.67%	92.02%	84.62%
SwiftNet	63.15%	98.76%	91.63%	84.52%
Ours	67.67%	99.06%	92.16%	86.30%

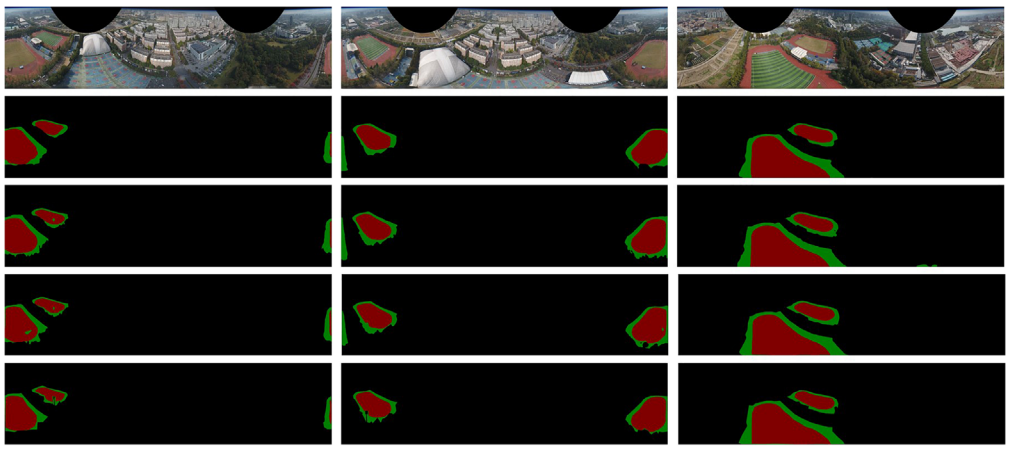


Fig. 11 Qualitative semantic segmentation results. From top to bottom row: RGB input image, ERF-PSPNet, SwiftNet, our method, and ground truth.

for panoramic aerial understanding. To compare with state-of-the-art networks, we train and verify our network on the Cityscapes.[32] It is a large-scale RGB dataset that focuses on semantic segmentation of urban street scenes. Through verification, our proposed method achieves a balance between accuracy and speed and reaches the state-of-the-art performance on the Cityscapes dataset with 39.4 Hz in full resolution on a single 2080Ti GPU processor. The optical system we designed has a high resolution and a good imaging quality at the edge FoV, and the segmentation algorithm has good robustness. We aim to further incorporate smaller-feature objects such as buildings and cars in the future.

5 Conclusion

Aerial photography has strict requirements for the lightness and portability of the lens. Under the limit of size, the observable range of a single traditional lens is often not large. In this paper, we discussed the imaging principle of FCP and designed a lightweight U-Net-like model for panoramic semantic segmentation. Also, we designed a 4K high-resolution PAL with a large FoV $(30 \text{ deg to } 100 \text{ deg}) \times 360 \text{ deg}$, where the total length is 79.5 mm, the maximum diameter is 44 mm, and the weight is 140 g, making it easily deployable on the drone system. A large number of aerial images are collected on the track and field, and pixel-level semantic segmentation is performed. Compared with two competitive real-time semantic segmentation networks, our proposed network has better performance while maintaining a fast inference speed of 39.4 fps. The recognition effect in average IoU is better than 86.30%. Through the verification of this paper, the aerial view imaging and semantic segmentation of the PAL system brings a broader application scenario to the fields of investigation, security monitoring, and so on. In the future, we plan to design a multispectrum, multifocal, multiview PAL system with the aerial scene semantic segmentation model to achieve more functions and applications.

Acknowledgments

The authors thank the Science Challenge Project (No. TZ2016006-0502-02) and National Natural Science Foundation of China (NSFC, No. 61875173).

References

1. B. Custers, *Future of Drone Use*, Springer, Hague, Netherlands (2016).
2. L. Kong and H. Yu, "Research on aerial image splicing technology of UAV," in *IEEE Int. Conf. Mechatron. and Autom.*, IEEE, pp. 376–380 (2018).
3. Y. Xiong and K. Turkowski, "Creating image-based VR using a self-calibrating fisheye lens," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 237–243 (1997).
4. J. Fabrizio, J. P. Tarel, and R. Benosman, "Calibration of panoramic catadioptric sensors made easier," in *Proc. IEEE Workshop Omnidirectional Vision*, IEEE, pp. 45–52 (2002).
5. I. Powell, "Panoramic lens," *Appl. Opt.* **33**, 7356–7361 (1994).
6. P. Greguss, "Panoramic security," *Proc. SPIE* **1509**, 55–66 (1991).
7. S. Niu et al., "Design of a panoramic annular lens with a long focal length," *Appl. Opt.* **46**, 7850–7857 (2007).
8. K. Zhang et al., "Design of a panoramic annular lens with ultrawide angle and small blind area," *Appl. Opt.* **59**, 5737–5744 (2020).
9. J. Wang et al., "Design of high resolution panoramic annular lens system," *Proc. SPIE* **11338**, 113382I (2019).
10. S. Gao, E. A. Tsyganok, and X. Xu, "Design of a compact dual-channel panoramic annular lens with a large aperture and high resolution," *Appl. Opt.* **60**, 3094–3102 (2021).
11. A. Amani, J. Bai, and X. Huang, "Dual-view catadioptric panoramic system based on even aspheric elements," *Appl. Opt.* **59**, 7630–7637 (2020).
12. Y. Luo et al., "Non-blind area PAL system design based on dichroic filter," *Opt. Express* **24**, 4913–4923 (2016).

13. Q. Zhou et al., "Design and implementation of a high-performance panoramic annular lens," *Appl. Opt.* **59**, 11246–11252 (2020).
14. Y. Huang et al., "Design of a compact two-channel panoramic optical system," *Opt. Express* **25**, 27691–27705 (2017).
15. W. Song et al., "Design and assessment of a 360 panoramic and high-performance capture system with two tiled catadioptric imaging channels," *Appl. Opt.* **57**, 3429–3437 (2018).
16. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3431–3440 (2015).
17. M. Orsic et al., "In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 12607–12616 (2019).
18. Q. Song, K. Mei, and R. Huang, "AttaNet: attention-augmented network for fast and accurate scene parsing," in *Proc. AAAI Conf. Artif. Intell.*, AAAI, pp. 2567–2575 (2021).
19. L. Sun et al., "Real-time fusion network for rgb-d semantic segmentation incorporating unexpected obstacle detection for road-driving images," *IEEE Robot. Autom. Lett.* **5**, 5558–5565 (2020).
20. Y. Hong et al., "Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes," http://arXiv:2101.06085 (2021).
21. K. Yang et al., "PASS: panoramic annular semantic segmentation," *IEEE Trans. Intell. Transp. Syst.* **21**, 4171–4185 (2020).
22. Z. Huang, J. Bai, and X. Y. Hou, "Design of panoramic stereo imaging with single optical system," *Opt. Express* **20**, 6085–6096 (2012).
23. X. Zhou et al., "Comparison of two panoramic front unit arrangements in design of a super wide angle panoramic annular lens," *Appl. Opt.* **55**, 3219–3225 (2016).
24. L. Sun et al., "Aerial-PASS: panoramic annular scene segmentation in drone videos," in *Proc. IEEE Eur. Conf. on Mobile Robots (ECMR)*, IEEE, pp. 1–6 (2021).
25. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 770–778 (2016).
26. D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *IEEE/RSJ Int. Conf. Intell. Rob. and Syst.*, IEEE, pp. 5695–5701 (2006).
27. K. Wang, "PASS Dataset," http://wangkaiwei.org/downloadeg.html (2021).
28. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proc. Int. Conf. on Learn. Represent.* (2015).
29. O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision* **115**, 211–252 (2015).
30. K. He et al., "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vision*, IEEE, pp. 1026–1034 (2015).
31. K. Yang et al., "Unifying terrain awareness through real-time semantic segmentation," in *IEEE Intell. Veh. Symp.*, IEEE, pp. 1033–1038 (2018).
32. M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3213–3223 (2016).

Jia Wang received her BS degree in optical engineering from Changchun University of Science and Technology, China, in 2018. She is currently pursuing her PhD with the College of Optical Science and Engineering, Zhejiang University. Her current research interests include optical design and optical simulation.

Kailun Yang received his PhD in information sensing and instrumentation from Zhejiang University, China, in 2019. He is a postdoctoral researcher at Computer Vision for Human–Computer Interaction (CV:HCI) Lab, Karlsruhe Institute of Technology (KIT), Germany. His research is in real-time computer vision for intelligent vehicles and real-world navigation assistance systems for visually impaired people.

Shaohua Gao received his BS and MS degrees in optical engineering from Changchun University of Science and Technology, China, in 2018 and 2021, respectively. He also received

the MS degree in optical engineering from ITMO University, Russia, in 2020. He is currently pursuing his PhD with the College of Optical Science and Engineering, Zhejiang University. His current research interests include optical design and computational vision.

Lei Sun received his BS degree in optical engineering from Beijing Institute of Technology, China, in 2018. He is currently pursuing his PhD with the College of Optical Science and Engineering, Zhejiang University. His current research interests include computational vision and semantic segmentation.

Chengxi Zhu received his BS degree in optical engineering from Zhejiang University, China, in 2020. He is currently pursuing his MS degree with the College of Optical Science and Engineering, Zhejiang University. His current research interests include optical design and optical simulation.

Kaiwei Wang received his BS degree in 2001 and his PhD in 2005 from Tsinghua University. He started postdoctoral research at the Center of Precision Technologies of Huddersfield University, funded by the Royal Society International Visiting Postdoctoral Fellowship and the British Engineering Physics Council. He joined Zhejiang University in February 2009 and has been mainly researching on Intelligent Detection Device for Passive Fiber Components and Visual Assisting Technology for the Visually Impaired since then. He had published 34 patents and more than 100 research papers in both domestic and foreign journals and professional conferences. He is the deputy director of the National Optical Instrument Engineering Technology Research Center of Zhejiang University.

Jian Bai is a professor at the Zhejiang University. He received his BS degree in computer science and Technology from Zhejiang University, China, in 1989. He received his MS and PhD degrees in optical engineering from Zhejiang University, China, in 1992 and 1995, respectively, and received his postdoctoral degree from Osaka University in 2000. He mainly researches the design of panoramic annular optical system, the measurement of long focal length lens, and the accelerometer of micro-opto-electromechanical system.