Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Panoramic Glass Image Segmentation Network

Guanlin Pan, Yan Cui*, Qingling Chang, Kelin Li, Yangtao Ou, Haohui Yu

School of Electronic and Information Engineering, Wuyi University, Jiangmen, Guangdong, 529000, China,

**Author to whom correspondence should be addressed.*

Abstract: In panoramic images, the geometric distortion caused by wide-angle lenses makes traditional semantic segmentation methods difficult to accurately segment the glass areas. To address the challenges of capturing spatial features and integrating context information, we propose the Panoramic Glass Image Segmentation Network (PGISNet). This network integrates the Matrix Decomposition Base Module (MDBM), the Transparent Perception Consistency Module (TACM), the Context and Texture Compensation Module (CTCM), and the Multi-scale Gated Context Attention Module (MGCA), constructing a progressive feature processing flow. Experimental results on the PanoGlassV2 benchmark test show that PGISNet achieved 90.03% IoU, 94.76% F-score, and 94.0% PA, significantly outperforming existing methods, verifying its effectiveness and advancement in the panoramic image glass segmentation task.

Keywords: Glass segmentation; Machine learning; Panoramic segmentation

Online publication: December 16, 2025

1. Introduction

The segmentation of glass in panoramic images is highly challenging due to the transparency, reflectivity and wide-angle distortion effects of the objects: transparency causes confusion between the foreground and background, reflection introduces interfering information, and geometric distortion further increases the difficulty of segmentation. To address this, this paper proposes an innovative panoramic glass segmentation network, PGISNet, whose core consists of four collaboratively operating modules MDBM, TACM, CTCM and MGCA, forming a progressive feature processing flow. MDBM is responsible for feature separation and purification, TACM enhances the perception and consistency of transparent areas, CTCM performs difference compensation and texture restoration, and MGCA integrates multi-scale context to generate discriminative segmentation features. This architecture gradually addresses these challenges and ultimately achieves precise glass region segmentation.

2. Methodology

2.1. Overview

We proposed PGISNet, a network based on the encoder-decoder architecture, with SegNext as the backbone. It

constructs a progressive feature optimization process through four core modules: MDBM, TACM, CTCM, and MGCA, significantly improving the segmentation accuracy of glass areas in panoramic images (**Figure 1**).

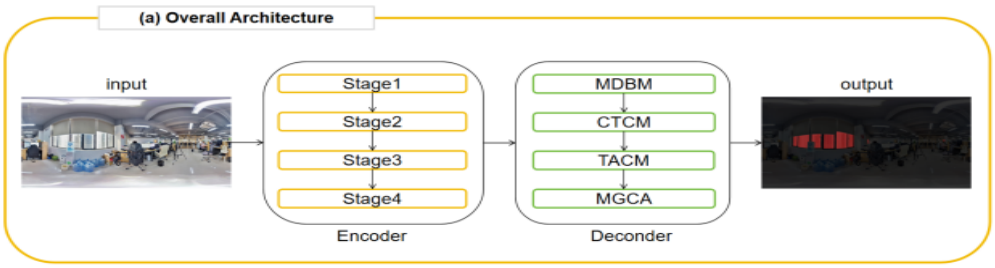

Figure 1. The architecture diagram of PGISNet.

2.2. Matrix decomposition base module

In the task of transparent object segmentation, the underlying features often contain texture noise and background interference, which affect the model's perception of the transparent regions. To address this issue, we designed MDBM, based on the idea of non-negative matrix factorization (NMF), and introduced a dual transposed convolution architecture to achieve efficient feature decomposition and reconstruction. This module first enhances the feature expression ability through transposed convolution, and then uses the learnable NMF process to decompose the features into base matrices and coefficient matrices, extracting the common features of the transparent objects. Another independent transposed convolution branch fuses the original input and the output of the NMF path to generate the final result. This dual-path design improves the resolution and the discriminative feature reconstruction ability, while effectively suppressing noise and obtaining purer and more robust feature representations (**Figure 2**).

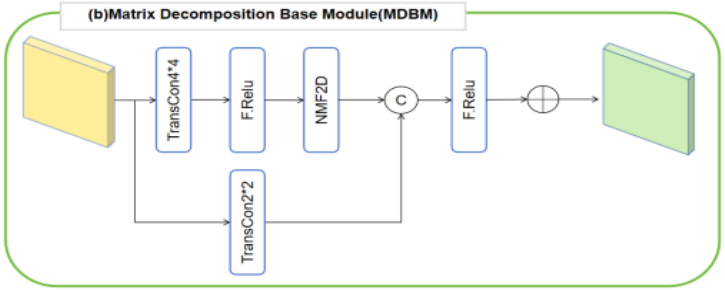

Figure 2. MDBM structure diagram.

2.3. Context and texture compensation module

The CTCM module employs an attention-guided dual-branch differential compensation mechanism to address the contradiction between context dependence and texture destruction in glass recognition. This module first generates a soft attention mask to estimate the probability of glass regions: The context-enhancement branch captures long-range context through 3×3 standard convolution and 5×5 dilated convolution, and uses the mask to focus on semantic reasoning of transparent regions; The texture-compensation branch employs lightweight convolution (including 1×1 dimension reduction and 3×3 convolution) to enhance the texture details of non-glass regions, and repairs the occluded background through the reverse mask. The outputs of the dual branches are concatenated and then processed through a fusion unit containing convolution, batch normalization, and ReLU to achieve adaptive

integration, generating more discriminative enhanced features (**Figure 3**).

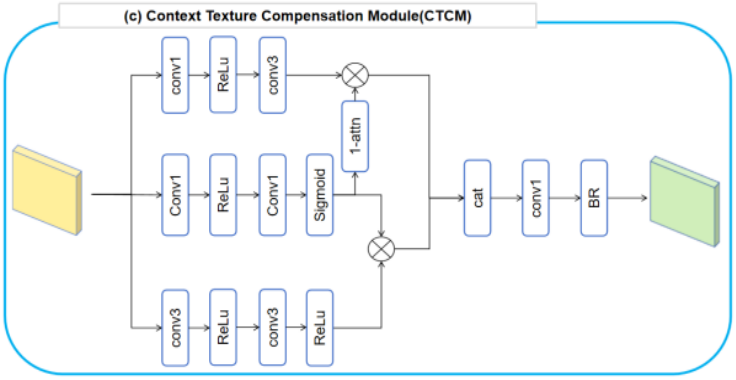

Figure 3. CTCM structure diagram.

2.4. Transparency-aware consistency module

In the segmentation of transparent objects, the features of the transparent regions often become weak and inconsistent due to background fusion. To address this issue, we propose the TACM, which synchronously improves local saliency and global semantic consistency through a dual-path supervision mechanism. This module first uses lightweight gated attention to enhance the response of the transparent regions point by point, then extracts semantic vectors through the global context branch to suppress semantic drift; finally, it fuses the local response with the global prior to generate dynamic weights, and achieves robustness enhancement through adaptive feature adjustment and residual connection (**Figure 4**).

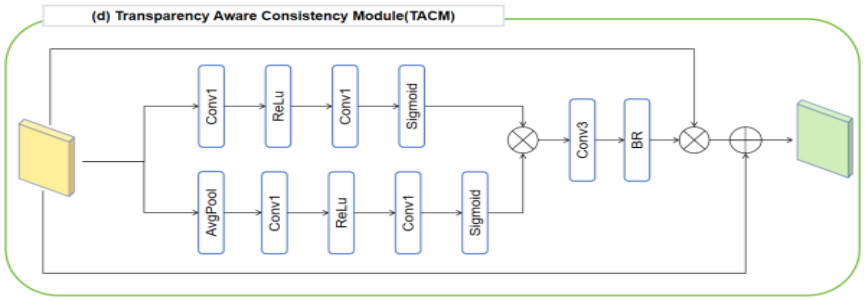

Figure 4. TACM structure diagram.

2.5. Multi-scale gated context attention

To address the challenge of multi-scale fusion of glass objects, we propose the MGCA module, a lightweight ASPP gated fusion unit. This module adopts a five-branch multi-scale perception and dynamic gating fusion strategy: after dimensionality reduction through 1×1 convolution, five complementary branches are constructed in parallel (1×1 convolution maintains the resolution, three depth-wise separable dilated convolutions capture multi-scale context, and global pooling extracts global semantics). The branch attention mechanism is introduced, and the outputs of the five branches are concatenated and then processed through point convolution to generate 5-channel weights. Through Softmax normalization, a spatially adaptive weight distribution is formed. Finally, the weighted features are projected back to the original channels through 1×1 convolution, and combined with learnable parameters to output refined features through residual connections (**Figure 5**).

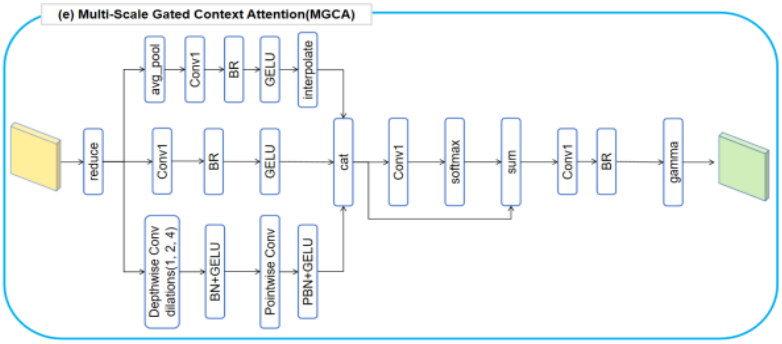

Figure 5. MGCA structure diagram

3. Results and discussion

3.1. Details of experiment

In order to comprehensively evaluate the effectiveness of PGISNet, this experiment strictly followed the training and testing procedures adopted by SegNext [1]. The experimental data was sourced from the PanoGlassV2 dataset, which is specifically designed for the glass segmentation task in panoramic images [2]. We used 1512 images for training and 353 images for testing. All experiments were conducted under the hardware/software configuration specified in **Table 1**.

Table 1. Experimental detail

Configuration	Version
Operating system	Ubuntu 22.04
Framework	Pytorch 2.1.2
GPU	NVIDIA RTX 3090
CDUA	11.8
Language	Python 3.8
Optimizer	AdamW
Batch size	4
Epochs	160,000
Learning rate	6×10-4
Image size	512×512

3.2. Evaluation

We referred to the evaluation metrics of RGB-T and Trans10K-v2, and conducted performance evaluations on the selected methods and our model on the PanoGlassV2 dataset [2–4]. The core metric mainly adopted was the Intersection over Union (IoU), which is widely used in the fields of semantic segmentation and glass segmentation. Its calculation definition is as follows:

$$IoU = \frac{TP}{TP+FP+FN} \tag{1}$$

Among them, TP, FP and FN represent the number of true positive, false positive and false negative pixels respectively. Besides IoU, we also use pixel accuracy (PA) as an evaluation metric, and its calculation formula is as follows:

$$\mathrm{PA} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{2}$$

Here, TN represents the number of true negative pixels. Additionally, we also use the F-score as another evaluation metric. This metric is the harmonic mean of the average accuracy and the average recall rate, and its calculation formula is as follows:

$$\mathrm{Fscore} = \frac{(1+\beta^2)\ \mathrm{Precision}\times\mathrm{Recall}}{\beta^2\times\mathrm{Precision}+\mathrm{Recall}} \tag{3}$$

When β is set to 1, the precision and recall rate are defined as $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$ respectively .

3.3. Quantitative evaluation

To ensure fairness, apart from RGB-T, Trans2Seg and TransLab, we used the segmentation tool to conduct pre-evaluation on all the models [3–5]. PGISNet was compared with 16 advanced semantic segmentation methods, 3 glass segmentation methods, and 3 panoramic image segmentation methods, including semantic segmentation methods such as PSPNet, SegNext, etc., glass segmentation methods such as RGB-T, etc., and panoramic segmentation methods such as PanoGlassNet, etc. [1,6,7]. The experimental results are shown in **Table 2**. PGISNet achieved the best performance on the PanoGlassV2 dataset that IoU of 90.03%, PA of 94%, F-score of 94.76%, and all indicators surpassed the existing comparison methods [2].

Table 2. The comparison of quantification for each model

Methods	Backbone	IoU↑	PA↑	Fscore↑	Param(M)	Flops(G)
Glass						
TransLab [5]	ResNet50 [5]	80.23	84.8	83.35	41.13	62.7
RGB-T [3]	ResNet50 [3]	84.3	89.47	93.45	86.3	84
Trans2Seg [4]	ResNet50 [5]	86.52	92.2	94.3	328.55	222.29
Semantic						
Vit [8]	Vit [8]	56.68	66.05	72.35	145.64	391.72
BiSeNetv2 [9]	BiSeNetv2 [9]	62.45	81.11	77.25	13.23	11.02
STDC [10]	STDCNet [10]	74.58	82.32	85.44	8.275	8.461
ResNeSt [11]	ResNet [11]	84.77	89.88	91.76	69.9	263.64
CCNet [12]	ResNet50 [12]	86.81	91.4	92.51	47.592	201
Twins [13]	PCPVT [13]	87.75	92.95	93.48	134.46	281.08
SegNext [2]	MSCAN [2]	88.19	93.63	93.72	27.77	31.24
Swin [14]	Swin [14]	87.85	92.12	93.53	232.97	407.64
ConvNext [15]	ConvNeXt [15]	89.17	93.6	94.29	390.91	830.72
Panoramic						
360BEV [16]	trans4pass [17]	68.65	84.07	81.41	27.32	515.72
Trans4Pass [17]	trans4pass [17]	69.03	84.52	81.61	29.97	27.3
PanoGlassNet [7]	ConvNeXt [15]	89.28	93.61	94.31	427.5	581
Ours						
-	MSCAN [1]	90.03	94	94.76	41.814	211

3.4. Ablation experiment

The ablation experiments conducted on the PanoGlassV2 dataset in this section demonstrate that using MDBM alone can improve feature purity, but the overall improvement is limited [2]. Introducing TACM alone can enhance perceptual consistency but lacks texture compensation. Using only CTCM can improve detail restoration but weakens feature consistency. Using MGCA alone can enhance multi-scale context fusion, but due to the lack of support from the preceding modules, the performance is still limited. The theoretical analysis and experimental results indicate that the four modules have complementary functions and must work together to achieve the optimal segmentation performance.

Table 3. The ablation experiments

Model	IoU↑	PA↑	Fscore↑
Base	88.19	93.63	93.72
Base + MDBM	89.67	94.07	94.55
Base + TACM	88.68	93.55	94.0
Base + CTCM	87.85	92.8	93.53
Base + MGCA	88.63	93.1	93.97
PGISNet	90.03	94	94.76

Figure 6. Comparative analysis of sample data. The relevant parts are highlighted in yellow.

4. Conclusion

The main contributions of this paper lie in the development of a series of modules that collectively enhance transparent object segmentation. We introduce the MDBM, which leverages non-negative matrix factorization to extract shared features of transparent objects, and the TACM, which employs a dual-path collaborative mechanism to strengthen feature consistency within transparent regions. We further construct the CTCM, which performs context reasoning for transparent areas and texture restoration for non-transparent areas through an attention-guided differential compensation strategy. In addition, we propose the MGCA, which achieves efficient multi-scale context fusion using a five-branch perceptual structure with dynamic gated integration. Despite these advancements, the serial combination of MDBM, TACM, CTCM, and MGCA inevitably increases model parameters and inference latency, reflecting a key limitation of the current design. Future work will therefore focus on structural optimization and model lightweighting, aiming to develop a transparent object segmentation framework that achieves both high accuracy and high efficiency.

Funding

This research was supported by the National Key Research and Development Program of China under Grant No. 2022YFA1602003, entitled "Intelligent Monitoring of Taishan Neutrino Detector".

Disclosure statement

The authors declare no conflict of interest.

References

[1] Guo M, Lu C, Hou Q, et al., 2022, Rethinking Convolutional Attention Design for Semantic Segmentation. ArXiv. https://doi.org/10.48550/arXiv.2209.08575

[2] Chang Q, Meng X, Hong Z, et al., 2024, ProgressiveGlassNet: Glass Detection with Progressive Decoder. In: 2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), 917–925.

[3] Huo D, Wang J, Qian Y, et al., 2023, Glass Segmentation with RGB-Thermal Image Pairs. IEEE Trans Image Process, 2023(32): 1911–1926.

[4] Xie E, Wang W, Wang W, et al., 2021, Segmenting Transparent Object in the Wild with Transformer. ArXiv. https://doi.org/10.48550/arXiv.2101.08461

[5] Xie E, Wang W, Wang W, et al., 2020, Segmenting Transparent Objects in the Wild. ArXiv. https://doi.org/10.48550/arXiv.2003.13948

[6] Zhao H, Shi J, Qi X, et al., 2017, Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6230–6239.

[7] Chang Q, Liao H, Meng X, et al., 2024, PanoglassNet: Glass Detection with Panoramic RGB and Intensity Images. IEEE Trans Instrum Meas, 2024(99): 1.

[8] Dosovitskiy A, Beyer L, Kolesnikov A, 2020, An Image is Worth 16x16 Words; Transformers for Image Recognition at Scale. ArXiv. https://doi.org/10.48550/arXiv.2010.11929

[9] Yu C, Gao C, Wang J, 2020, BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. ArXiv. https://doi.org/10.48550/arXiv.2004.02147

[10] Fan M, Lai S, Huang J, et al., 2021, Rethinking BiSeNet for Real-Time Semantic Segmentation. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9711–9720.

[11] Zhang H, Wu C, Zhang Z, 2022, Split-Attention Networks. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2735–2745.

[12] Huang Z, Wang X, Huang L, et al., 2019, CCNet: Criss-Cross Attention for Semantic Segmentation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 603–612.

[13] Chu X, Tian Z, Wang Y, 2021, Twins: Revisiting the Design of Spatial Attention in Vision Transformers. Advances in Neural Information Processing Systems (NeurIPS 2021), 9355–9366.

[14] Liu Z, Lin Y, Cao Y, 2021, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992–10002.

[15] Liu Z, Mao H, Wu C, 2022, A ConvNet for the 2020s. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11966–11976.

[16] Teng Z, Zhang J, Yang K, 2022, 360BEV: Panoramic Semantic Mapping for Indoor Bird's Eye View. ArXiv. https://doi.org/10.48550/arXiv.2303.11910

[17] Zhang J, Yang K, Ma C, 2022, Bending Reality: Distortion-Aware Transformers for Adapting to Panoramic Semantic Segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16917–16927.