GCVPN: A Graph Convolutional Visual Prior-Transform Network for Actual Occluded Image Recognition

Lei Wang*
Tianjin University
Tianjin, China
wanglei2019@tju.edu.cn

Nannan Wu*
Tianjin University
Tianjin, China
nannan.wu@tju.edu.cn

Huaming Wu†
Tianjin University
Tianjin, China
whming@tju.edu.cn

Wei Yu
Tianjin University
Tianjin, China
weiyu@tju.edu.cn

Fan Zhang
Sino-Singapore Tianjin Eco-City
Tianjin, China
zhangf@eco-city.gov.cn

Shuo Chen
Tianjin Shenglian Intelligent
Technology Development Co., Ltd.
Tianjin, China
18630631289@163.com

Abstract

Image recognition plays a critical role in urban security, traffic management, and environmental monitoring, yet achieving high accuracy in obstructed scenes remains a challenge. To address this, we propose a Graph Convolutional Visual Prior-Transform Network (GCVPN), which significantly improves recognition accuracy and efficiency in complex environments. GCVPN introduces an image prior slicing and topology transformer to convert image data into graph-structured slice features, integrating domain overlap sampling and planar mapping to handle symmetry and enable precise, rapid anomaly detection. By combining a traditional VGG backbone with graph convolutional layers, GCVPN jointly captures topological relationships and feature semantics, while maintaining real-time efficiency with continuous recognition at 30 video frames per second. Extensive experiments demonstrate its effectiveness in photovoltaic panel anomaly detection and face occlusion recognition, highlighting strong potential for applications in intelligent surveillance and autonomous driving.

CCS Concepts

• **Computing methodologies** → **Computer vision tasks**; **Artificial intelligence**; **Neural networks**.

Keywords

Occluded Image Recognition; Image Occlusion Recognition; Prior Segmentation; Graph Relational Network

ACM Reference Format:

Lei Wang, Nannan Wu, Huaming Wu, Wei Yu, Fan Zhang, and Shuo Chen. 2025. GCVPN: A Graph Convolutional Visual Prior-Transform Network for Actual Occluded Image Recognition. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3746252.3761518

*Both authors contributed equally to this research.
†Corresponding author.

CIKM '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3761518

1 Introduction

Urban development drives societal progress, with image recognition technology enhancing urban safety, face recognition [23], traffic management [1], and environmental monitoring [3, 40]. However, complex scenarios pose challenges like occlusion, uneven lighting, and spatial alignment issues, with occlusion causing 40% of detection failures [37] due to obscured objects, such as vehicles hidden by trees or pedestrians, faces blocked by masks, etc. This results in incomplete target appearance, blurred edges, and background confusion, complicating classification and localization, thus reducing system accuracy.

Traditional algorithms like ResNet [14] and Auto-Encoder [19] struggle with occluded images, often misinterpreting environmental factors such as ambient light variations, adverse weather, and shadows, leading to erroneous feature learning. They fail to differentiate occluded from visible regions [12], overlook missing information, and struggle in complex environments, compromising visual system reliability. This study develops lightweight, robust image recognition techniques to address occlusion, improving precision and performance across diverse urban applications, including real-time traffic flow optimization, enhanced security surveillance, and environmental monitoring.

Recent advancements in occluded image recognition, driven by deep learning techniques like attention mechanisms [13], Generative Adversarial Networks (GANs) [11], and Graph Neural Networks (GNNs) have led to advanced methods such as multi-scale attention, non-modal segmentation-based information completion, and adaptive filter updates. Large-scale models like Stable Diffusion [29] further enhance occlusion completion by leveraging object shape priors, significantly improving accuracy and reducing consumption.

The occluded image recognition methods face challenges due to high computational complexity [20], especially in large-scale models, making them impractical for real-time deployment on resource-constrained platforms like mobile devices or edge systems. While lightweight methods reduce computational demands, they often

compromise precision and generalizing ability, particularly in complex occlusion scenarios, as seen in simplified enhanced feature approaches. These limitations hinder performance in diverse real-world urban applications, such as real-time traffic monitoring, autonomous vehicle navigation, and efficient edge-based surveillance.

Furthermore, most existing occluded image recognition studies rely on datasets such as KITTI [10] and ILSVRC2012 [6], which fail to capture diverse scenarios involving extreme illumination, adverse weather, or dynamic occlusion. As a result, current models often exhibit limited robustness in real-world environments [35], restricting their effectiveness for end-to-end detection in resource-constrained settings. Addressing this gap is essential to advancing applications in autonomous driving, real-time urban surveillance, and edge-based environmental monitoring.

This study presents the Graph Convolutional Visual Prior-Transform Network (GCVPN), an optimized architecture for occluded image recognition in complex environments, significantly improving accuracy, efficiency, and robustness for applications in intelligent systems like autonomous navigation and urban surveillance. The main contributions of this paper are as follows:

- This study introduces a novel topological transformation method using axisymmetric properties to enhance occluded image recognition. By leveraging image symmetry and spatial correlations, it pioneers a prior topological module, improving global feature inference for occluded images and boosting accuracy in processing complex and incomplete data for urban applications like surveillance and traffic monitoring.
- GCVPN innovatively harnesses prior knowledge of symmetry and spatial correlations inherent in natural images, seamlessly integrating a topological transformation module with a linear convolution module, which enables simultaneous processing of topological relationships and feature analysis of image data while increasing the inference efficiency to enable continuous recognition of 30 frames Real-time video, exhibits low computational complexity and robust generalization capabilities.
- GCVPN has been trained and validated on complex real-world datasets encompassing multi-resolution scenarios, extreme illumination conditions, and dynamic occlusion, demonstrating exceptional adaptability and robust generalization. The lightweight architecture surpasses the most advanced approaches in occluded image recognition, rendering it exceptionally suited for real-time edge computing in intelligent transportation systems, urban surveillance, and autonomous vehicles.

2 Related Work

Recent advancements in occluded image detection methods leverage deep learning techniques, including Convolutional Neural Networks (CNNs) [21], Graph Neural Networks (GNNs) [30], and Spatio-temporal Graph Networks [2]. However, these approaches grapple with environmental disturbances, biased feature extraction and insufficient exploitation of shape priors, thereby constraining their efficacy in traffic monitoring and surveillance applications. Innovations in attention mechanisms [34] and GANs [11], such as multi-scale attention and non-modal segmentation, combined with pre-trained models like Stable Diffusion [28, 29], improve occlusion handling and thereby advance autonomous driving and urban monitoring.

2.1 Methods based on CNNs

CNNs excel in object recognition, facial recognition and medical image analysis, such as classifying vulnerable plaques in OCT images for cardiovascular disease diagnosis, improving accuracy and reducing doctors' workload. Models like ResNet [14], DenseNet [17], and Inception [32] enhance recognition accuracy and stability. ShortCut3-ResNet [33] improves ResNet by learning diverse image features, while DenseNet with MHGSO [8, 17] excels in plant leaf disease detection, demonstrating CNNs' versatility in image recognition tasks.

Despite their strengths, CNNs' spatially localized operators struggle to capture distant spatial relationships and limit their modeling capabilities in complex scenes [14]. Adapting CNNs to image data necessitates additional processing steps that elevate training complexity and risk information loss, especially in demanding image recognition tasks.

2.2 Methods based on GNNs

GNNs excel in behavior recognition, gait recognition and multi-view image processing by capturing node relationships. ST-GCN [38] enhances gait recognition accuracy from blurry images via cluster analysis. Fusion-GCN [7] improves multi-modal behavior recognition by integrating sensor data. In multi-view tasks such as forest fire recognition, GNNs improve accuracy by reducing false alarms and enhancing the reliability of urban surveillance and environmental monitoring. However, the complex design and training requirements of GNNs demand substantial computational resources, presenting optimization challenges. Despite this, the flexibility and performance of GNNs across diverse scenarios highlight their potential to advance image recognition and stimulate research into more resource-efficient solutions.

2.3 Methods based on Transformer

The Transformer architecture has been adapted for computer vision with impressive results. Vision Transformer [18] processes image patch sequences and excels in image classification benchmarks. Swin Transformer [24] enhances performance with multi-scale feature extraction and local window attention, which improves the stability and accuracy for high-resolution images and complex scenes. These advancements enable applications like real-time video surveillance, autonomous driving scene analysis and urban traffic monitoring.

Recurrent Attention Multi-Scale Transformer [16] captures fine-grained local motion in dynamic scenes and enhances video action classification by modeling spatiotemporal dependencies. AdaViT [27] enhances efficiency through adaptive computation and dynamic labeling tailored to varying image complexities. While Transformers excel at capturing global features, they require large amounts of data and computational resources, increasing the risk of overfitting on small datasets and complicating deployment in resource-constrained settings such as edge-based surveillance.

2.4 Methods for Image Generation and Completion

The image generation and complement techniques aim to reconstruct occluded objects' shapes, which restore visible and invisible regions with precision. LLE-CNNs [9] evolve from encoder-decoder designs to multi-stage task decomposition and feature space repair. SDAmodal [39] uses shape priors for high-quality and annotation-free completion with robust zero-shot generalization. These advancements enhance applications like autonomous vehicle perception, real-time urban monitoring and medical image reconstruction.

CNN-GAN hybrids [5] leverage 3D shape priors and multi-frame data for enhanced accuracy, while the Occlusion-aware Spatial Attention Transformer [15] uses commonsense and linguistic priors for robust reconstruction. High computational complexity limits their use on resource-constrained devices, which impacts real-time traffic surveillance and portable diagnostics.

These methods face challenges due to high computational complexity, especially in large-scale models. They make real-time deployment on resource-constrained platforms like mobile devices or edge systems impractical. However, lightweight methods that use less resource-intensive struggle to balance precision and efficiency, thereby limiting their generalization in complex occlusion scenarios and real-world applications. Therefore, we propose the GVCPN method to effectively address these issues.

3 Methodology

3.1 Problem Definition

In the research of image recognition, image data is represented as a 3D matrix $X_o \in \mathbb{R}^{W\times H\times C}$, where W, H and C denote width, height, and color channels. RGB images are typically represented with $C = 3$, corresponding to three color channels: red (R), green (G), and blue (B). Initial convolutional layers resize images into standardized feature matrices. After topology converters segment the images into regions, features from each region form a node set $V = \{v_1, v_2, \ldots, v_N\}$, where N is the number of features. The Inter-region relationships create an edge set E which could form a graph $G = (V, E)$. The GCNs update image-regions' node feature representations $H^{(l)} \in \mathbb{R}^{N\times F}$, where F is the feature dimension. The mathematical framework defines image dimensions and provides a theoretical basis for structured image processing.

3.2 Proposed Framework

To tackle the occluded image recognition, we propose the GCVPN method, which leverages image priors through graph convolution while integrating topology transformers and geometric graph convolution to enhance urban surveillance and traffic analysis. The framework of the proposed method is illustrated in Fig. 1, and its core modules are as follows:

- **Adjustment Initialization**: This module preprocesses the acquired image data via two parallel 3×3 convolutions, and uses pooling to reduce the dimensionality of its RGB feature components and ensure compatibility with the subsequent algorithmic processing.
- **Prior-Cutting GCN**: This module includes the *Horizontal Partitioning Topology Transformer* and *Horizontal & Vertical Partitioning Topology Transformer*, which preprocess the images into 7 parts for convenient topology transformation. And then it extracts the spatial structural relationships of each partition through the GCNs, and has been followed by two layers of 3×3 convolution layers that absorb global data attributes.
- **Linear Decoder** : This module analyzes features from RGB channel data and subsequently performs the recognition task through the stack of three fully connected layers interspersed with two ReLU activations.

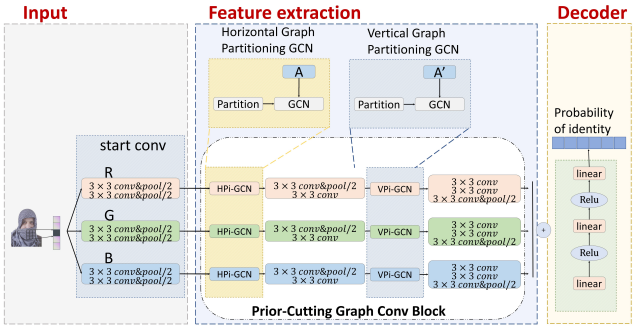

Figure 1: The proposed GCVPN framework leverages visual geometric priors to segment occluded images and employs GCNs to analyze the intrinsic topological relationships within the segmented images.

3.2.1 Adjustment Initialization. This module preprocesses images to a standard $224\times 224\times 3$ size, generating feature maps by applying convolution and pooling operations on the RGB color channels. Specifically, the input image data $X \in \mathbb{R}^{W\times H\times 3}$. The GCVPN model then generates feature maps from the RGB channels via two stages, each comprising a 3×3 convolution, ReLU activation, and 2×2 max pooling. The process extracts and compresses features from each color channel for further processing in occluded image recognition tasks.

3.2.2 Prior-Cutting Graph Convolution Network. In order to enhance the robustness of image recognition systems in complex environments, we adopt a data topology transformer to convert image data into graph-structured data efficiently. The transformer consists of three main components: image topology prior segmentation module, topology relationship representation module, and global feature extraction module.

GCVPN performs prior horizontal and vertical segmentation of image data. Experiments show that the optimal number of segmentation slices for typical facial recognition and photovoltaic panel datasets is seven. In real-world image recognition tasks involving occlusion and shadow, the image prior segmentation and topology relationship modules effectively capture the symmetric properties of the image data.

Image Topology Prior Segmentation Module. This module processes the RGB features of the color channel in image data as $C_{\text{in}} \in \mathbb{R}^{W\times H\times F}$. The feature extraction process for each channel includes two stages: Horizontal Partition GCN (HPi-GCN) and Vertical Partition GCN (VPi-GCN).

To improve the performance of GCN in handling complex expressions or partial occlusion, the HPi-GCN module partitions the image using natural horizontal feature lines, such as in facial images or photovoltaic panel images. The segmentation process is as follows:

- *Define Horizontal Regions.* For example, in facial images, the image is divided into seven main horizontal regions: top (hair, eyes, forehead), middle (nose, mouth), and bottom (chin, neck). Similar segmentation can be applied to photovoltaic panel images.
- *Region Division.* Based on predefined facial feature markers, map these regions onto the feature map. Let $y_1, y_2, \ldots, y_7$ be the vertical position markers after slicing the image.

In the segmentation process, an edge-overlapping window scheduling strategy is employed to ensure appropriate overlap between segmentation windows, reducing information loss and edge effects. The starting and ending coordinates of the window are:

$$W_{\text{end}}^{(i)} = i \times (1 - r_W) \times W_s + W_s, \tag{1}$$

where r_W is the horizontal overlap ratio. For each region of the image, extract its corresponding subset of features x_i^{horiz}.

In the design of the VPi-GCN, using facial images as an example, we can define natural partitions as follows: left ear, left cheek, left eye, nose, right eye, right cheek, and right ear. Similar segmentation can be applied to photovoltaic panel images. The starting and ending coordinates of the vertical segmentation window are:

$$H_{\text{end}}^{(j)} = j \times (1 - r_H) \times H_s + H_s, \tag{2}$$

where r_H is the vertical overlap ratio. For each region of the image, extract its corresponding subset of features x_j^{vert}.

Topology Relationship Representation Module. For the feature vectors x_i,x_j of each sub-region slice, we employ GCN to handle hidden relationships between the slices, where the dynamic adjacency matrix represents the intrinsic relationships of the image slice features. The specific steps are as follows:

The model performs graph convolution on the dynamic adjacency matrix A_i and feature matrix X_i:

$$X_i' = \sigma((\hat{A}_i)X_iW_i), \tag{3}$$

where $(\hat{A}_i)$ is the normalized adjacency matrix for each region, W_i is the weight matrix for each region, σ is the activation function.

For each segmented window, we first extract its feature vector through a convolution operation:

$$X_i^{\text{horiz}} = \text{conv}^{\text{horiz}}(\text{Horiz\_GCN}(x_i^{\text{horiz}})) \in \mathbb{R}^F, \tag{4}$$

$$X_j^{\text{vert}} = \text{conv}^{\text{vert}}(\text{Vert\_GCN}(x_j^{\text{vert}})) \in \mathbb{R}^F. \tag{5}$$

Then, we reassemble the processed sub-region feature maps $X_i$horiz, X_j^{vert} into a unified feature representation X^{horiz}, X^{vert} for combining feature maps.

Global Feature Extraction Module. After the horizontal and vertical graph convolution operations, the obtained feature representations $H_{\text{horiz}}^{\text{final}}$ and $H_{\text{vert}}^{\text{final}}$ respectively represent the global features of horizontal and vertical slices. We integrate these features:

$$H_{\text{final}} = H_{\text{horiz}}^{\text{final}} + H_{\text{vert}}^{\text{final}}. \tag{6}$$

By utilizing natural image features to partition regions, we can more effectively capture the characteristics of different parts of the image horizontally and vertically. This strategy is particularly useful for handling images with local occlusions or heterogeneity, enhancing the model's ability to represent image features. The Prior-Partitioning Graph Convolution Block can better handle image recognition tasks in complex environments.

3.2.3 Linear Decoder. We use a linear decoder to classify the aggregated features:

$$Y = \text{softmax}(W_{\text{out}}H_{\text{final}} + b_{\text{out}}), \tag{7}$$

where W_{out} and b_{out} are the weights and biases of the fully connected layer. The input RGB features are effectively transformed into an output vector representing identity probabilities through combining the linear layer and ReLU activation function. This process involves not only feature fusion and transformation but also accomplishes the recognition task. The final loss function and the algorithm workflow can be obtained as follows:

$$L(Y, \hat{Z}) = -[Y \log(\hat{Z}) + (1 - Y) \log(1 - \hat{Z})]. \tag{8}$$

Algorithm 1 Feature Extraction and GCN Processing

Require: Input image $X \in \mathbb{R}^{W \times H \times 3}$, number of horizontal regions P_{horiz}, number of vertical regions P_{vert}
Ensure: Processed feature maps H_{final}
  procedure Feature Extraction GCN($X, P_{\text{horiz}}, P_{\text{vert}}$)
    $C_1 = \text{Pool}(\text{ReLU}(\text{Conv}(X, K_1)))$ ▷ 1st conv.
    $C_2 = \text{Pool}(\text{ReLU}(\text{Conv}(C_1, K_2)))$ ▷ 2nd conv.
    $H_{\text{Horiz}} = Horiz\_Graph\_Part\_GCN(C_2, P_{\text{horiz}})$
    $H_{\text{Verti}} = Verti\_Graph\_Part\_GCN(C_2, P_{\text{vert}})$
    $H_{\text{final}} = (H_{\text{Horiz}} + H_{\text{Verti}})$
    return H_{final}
  end procedure

The GCVPN algorithm standardizes images, applies convolutions and pooling to generate feature maps C_1, C_2, and partitions them into regions processed by GCNs (Algorithm 1). Outputs combine into H_{final}, converted to the output features Y for recognition. It excels in occlusion recognition for edge-based urban applications like surveillance and traffic analysis.

3.3 Complexity Analysis

3.3.1 Time Complexity Analysis. The time complexity of the GCVPN algorithm is determined by its core modules:

- **Adjustment Initialization**: Preprocesses images with two convolutions ($O(W \times H)$), pooling ($O\left(\frac{W}{2} \times \frac{H}{2}\right)$), and ReLU ($O(W \times H)$), totaling $O(W \times H)$.
- **Prior-Cutting GCN**: Performs horizontal and vertical segmentation ($O(7 \times W \times H)$) and graph convolution ($O(P \times F)$), totaling $O(7 \times P \times F)$.
- **Linear Decoder**: Fully connected layer & Softmax is $O(F)$.

The overall complexity is $O(W \times H + 7 \times W \times H + 7 \times P \times F + F)$, it's suitable for edge computing and scalable to larger datasets.

3.3.2 Space Complexity Analysis. The algorithm's space complexity is primarily influenced by three key modules:

- **Adjustment Initialization Module**: This module generates an output feature map of size $\frac{W}{2} \times \frac{H}{2} \times 3$, resulting in a space complexity of $O\left(\frac{W}{2} \times \frac{H}{2} \times 3\right)$.
- **Prior-Cutting GCN Module**: Each region's feature matrix size is $P \times F$, contributing to a total space complexity of $O(7 \times P \times F)$.
- **Linear Decoder Module**: The space complexity of the output vector is $O(F)$.

The overall space complexity of the algorithm can be expressed as $O\left(\frac{W}{2} \times \frac{H}{2} \times 3 + 7 \times P \times F + F\right)$.

The algorithm's low space complexity, resulting from processing image feature representations instead of storing raw data, makes it well-suited for memory-constrained edge computing. Its modular design also ensures scalability, allowing it to handle complex image data with minimal additional memory requirements.

3.3.3 Scalability of the Algorithm. The GCVPN algorithm achieves a balance between efficiency and scalability, making it suitable for both edge and large-scale computing. For edge computing, its low time and space complexity allows rapid processing on resource-constrained devices. In large-scale computing scenarios, the modular and hierarchical design facilitates parallel processing of extensive datasets and complex tasks, ensuring flexible deployment across diverse environments.

4 Experiments

4.1 Experimental Settings

4.1.1 Setting and Dataset. We used an NVIDIA GeForce RTX 3090 GPU cluster for model training. Python 3.8 and PyTorch 2.1.1 were selected for development due to their portability and maintainability, facilitating rapid model iteration and performance enhancement. The datasets used in this study are as follows:

- **Photovoltaic Panel**: The dataset comprises 10,000 high-altitude aerial images of photovoltaic panels, capturing anomalies such as occlusions (leaves, debris, dust), aging, and damages (missing surfaces, peel-off) under diverse lighting, weather, viewing angles, shadows, and rainwater conditions. Annotated for anomalies, it provides a valuable resource for developing robust computer vision models for obstructed solar panel inspection.
- **Face Recognition**: The dataset comprises about 495,000 facial images from approximately 10,575 individuals, capturing diverse lighting, expressions, poses, clothing, and image sizes for complexity. It also expanded about 30,000 images of 525 individuals, including occluded (masks, hair, lighting variations) and non-occluded faces, enabling experiments comparing occlusion conditions.

4.1.2 Baselines. We evaluate GCVPN's performance in photovoltaic panel obstruction, face, and obscuring facial recognition tasks, comparing it with PCA [4], LDA [31], ResNet [33], VGG [26], ConvNeXt [36], and Efficient-Former [22]. Experiments used the same test set, with 40%, 60%, and 80% of the dataset for training, and 10% each for testing and validation to ensure fairness. Baseline methods are detailed below:

- **PCA**: PCA reduces dataset dimensionality by projecting data onto a new space, maximizing variance, often used to denoise and simplify models.
- **LDA**: LDA reduces dimensionality in supervised learning, maximizing class separation and minimizing within-class variance, effective for classification.
- **ResNet**: ResNet uses residual blocks to solve vanishing/exploding gradients, enabling deeper networks for complex feature learning.
- **VGG**: VGG networks have simple convolutional and pooling layers that excel in ImageNet image classification tasks.
- **ConvNeXt**: ConvNeXt blends convolutional and transformer features with larger kernels, inverted bottlenecks, and layer normalization.
- **Efficient-Former**: Efficient-Former optimizes efficiency using latency-driven slimming and lightweight MLP modules, balancing speed and accuracy for resource-constrained devices, but struggles with occluded image recognition.

4.1.3 Evaluation Metrics. We evaluated image recognition algorithms using standardized metrics as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{9}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{10}$$

$$TPR = \frac{TP}{TP + FN}, \tag{11}$$

$$FPR = \frac{FP}{FP + TN}, \tag{12}$$

where *TP*, *TN*, *FP*, and *FN* denote true positives, true negatives, false positives, and false negatives. The F1 Score is the harmonic mean of precision and recall. *Roc_Auc* measures binary classifier performance via the ROC curve, plotting *TPR* vs. *FPR* [25], with values from 0 to 1 (higher is better). These metrics comprehensively evaluate model performance.

4.2 Experimental Results

This section evaluates the GCVPN model's performance in photovoltaic panel obstruction recognition, for recognition and obscuring facial recognition tasks, which are compared with PCA, LDA, ResNet, VGG, ConvNeXt, and Efficient-Former.

4.2.1 Photovoltaic Panel Obstruction Recognition Task. In photovoltaic panel obstruction recognition, GCVPN outperforms traditional methods such as PCA, LDA, ResNet, and VGG, which often struggle with complex anomalies. The model also surpasses Efficient-Former and ConvNeXt in accuracy while maintaining computational efficiency, as shown in Table 1. GCVPN excels at detecting anomalies and reducing false alarms, making it highly effective for urban monitoring applications.

4.2.2 Facial Recognition Task. In the facial recognition task, GCVPN demonstrates strong competitiveness, as shown in Table 1. The model effectively extracts facial features and maintains high recognition accuracy under complex backgrounds and varying lighting

Table 1: Results of three recognition tasks: photovoltaic panel obstruction, facial recognition, and obscured facial recognition.

Dataset	Model	40% Train Dataset			60% Train Dataset			80% Train Dataset		
		F1	Acc	Roc_Auc	F1	Acc	Roc_Auc	F1	Acc	Roc_Auc
Photovoltaic Panel Obstruction Recognition	PCA	0.002	0.002	0.471	0.003	0.002	0.494	0.002	0.001	0.580
	LDA	0.820	0.705	0.610	0.924	0.860	0.573	0.554	0.420	0.589
	Resnet	0.631	0.912	0.939	0.687	0.935	0.949	0.489	0.824	0.964
	VGG	0.653	0.927	0.918	0.721	0.944	0.957	0.981	0.963	0.959
	ConvNeXt	0.858	0.756	0.550	0.848	0.753	0.562	0.851	0.746	0.566
	EfficientFormer	0.503	0.518	0.537	0.522	0.520	0.491	0.536	0.525	0.553
	GCVPN	**0.870**	**0.935**	**0.929**	**0.879**	**0.959**	**0.966**	**0.984**	**0.971**	**0.989**
Facial Recognition	PCA	0.024	0.701	0.716	0.589	0.988	0.714	0.128	0.757	0.769
	LDA	0.464	0.522	0.523	0.416	0.523	0.523	0.463	0.522	0.522
	Resnet	0.829	0.843	0.907	0.833	0.844	0.910	0.826	0.841	0.906
	VGG	0.828	0.828	0.907	0.831	0.833	0.901	0.830	0.834	0.922
	ConvNeXt	**0.924**	**0.876**	**0.952**	**0.927**	**0.887**	**0.957**	**0.929**	**0.888**	**0.958**
	EfficientFormer	0.801	0.730	0.845	0.819	0.799	0.855	0.866	0.825	0.822
	GCVPN	0.832	0.849	0.908	0.845	0.847	0.915	0.918	0.838	0.926
Obscuring Facial Recognition	PCA	0.005	0.485	0.573	0.002	0.520	0.602	0.218	0.713	0.736
	LDA	0.323	0.588	0.508	0.463	0.555	0.529	0.518	0.551	0.604
	Resnet	0.757	0.747	0.827	0.760	0.744	0.833	0.791	0.794	0.871
	VGG	0.772	0.749	0.804	0.764	0.733	0.823	0.790	0.791	0.870
	ConvNeXt	**0.884**	0.814	0.851	**0.889**	0.819	0.866	0.887	0.824	0.880
	EfficientFormer	0.590	0.519	0.571	0.522	0.544	0.544	0.522	0.526	0.556
	GCVPN	0.868	**0.823**	**0.853**	0.887	**0.827**	**0.869**	**0.896**	**0.840**	**0.897**

conditions. In contrast, traditional methods such as PCA, LDA, ResNet, and VGG show highly unstable performance, especially when trained on only 40% of the dataset. While ConvNeXt achieves performance comparable to GCVPN, its inference time is significantly longer, and Efficient-Former performs worse than our method on the same dataset.

4.2.3 Obscuring Facial Recognition Task. In the occluded facial recognition task, GCVPN not only effectively identifies facial features but also adapts to the loss of facial information caused by occlusion, maintaining high recognition accuracy. In contrast, traditional methods such as PCA, LDA, ResNet and VGG perform poorly in face recognition tasks involving obstruction, with their performance significantly declining as the data volume decreases, as detailed in Table 1. Meanwhile, despite requiring higher computational time, ConvNeXt does not achieve superior performance and Efficient-Former's performance remains slightly inadequate.

In summary, GCVPN's superiority lies in leveraging prior topological relationships of image patches to recognize graph structure relationships between image blocks, particularly excelling in occlusion scenarios by understanding dependencies and symmetrical relationships among features. This enhances its capability to handle occluded image recognition, which reduces inference time and improves efficiency. The model maintains high accuracy and robustness even with increasing data volume.

4.3 Ablation Study

This section assesses the GCVPN model's key components in photovoltaic panel obstruction, face recognition, and obscuring facial recognition tasks via ablation experiments, evaluating their contributions to overall performance.

Photovoltaic Panel Obstruction Recognition Task. In photovoltaic panel anomaly detection, as shown in Table 2. The dual GCN structure could capture global features better than a single GCN layer. Despite partial compensation by fully connected layers, removing GCN layers and short links significantly reduces the performance of global feature processing. Only utilizing fully connected layers severely degrades performance.

Table 2: Ablation study.

Model Configuration	F1	Acc	Auc
Photovoltaic Panel Obstruction Recognition			
GCVPN	**0.984**	**0.971**	**0.988**
Single GCN + Single Short Link	0.699	0.939	0.893
Double GCN + Single Short Link	0.684	0.933	0.958
No GCN + No Short Links	0.886	0.811	0.885
Fully Connected, No Convolution	0.779	0.667	0.789
Only GCN, No CNN	-	-	-
Facial Recognition			
GCVPN	**0.837**	**0.837**	**0.926**
Single GCN + Single Short Link	0.833	0.835	0.924
Double GCN + Single Short Link	-	-	-
No GCN + No Short Links	0.825	0.827	0.917
Fully Connected, No Convolution	-	-	-
Only GCN, No CNN	0.696	0.666	0.721
Obscuring Facial Recognition			
GCVPN	**0.782**	**0.759**	**0.837**
Single GCN + Single Short Link	0.771	0.752	0.832
Double GCN + Single Short Link	-	-	-
No GCN + No Short Links	0.778	0.755	0.833
Fully Connected, No Convolution	-	-	-
Only GCN, No CNN	0.605	0.593	0.629

Facial and Obscuring Facial Recognition Task. In face recognition tasks (Table 2), GCVPN demonstrates strong competitiveness, effectively extracting facial features and maintaining high accuracy under complex backgrounds, varied lighting, and occlusions. When key components such as GCN are removed, the model retains some recognition capability, but its accuracy is significantly reduced.

The GCVPN model's GCN, shortcut connections, and convolutional modules are vital for high-performance image recognition, enabling global feature processing, effective feature propagation, and basic spatial feature extraction.

4.4 Parametric Sensitivity Analysis

To evaluate GCVPN's parameter sensitivity in occluded image recognition, we tested slice number, dropout rate, and hyperparameter γ as shown in Table 3.

Table 3: Parameter sensitivity experiments.

No. of Slices	F1	Acc	Auc
2	0.825	0.827	0.917
4	0.823	0.824	0.916
7	**0.837**	**0.837**	**0.926**
14	0.821	0.825	0.916
28	0.826	0.827	0.917
Dropout Rate			
0.01	**0.984**	**0.971**	**0.988**
0.03	0.740	0.945	0.977
0.05	0.659	0.920	0.967
hyperparameters γ			
0.01	**0.984**	**0.971**	**0.988**
0.3	0.982	0.968	0.978
0.5	0.740	0.945	0.977

4.5 Complexity Analysis

We evaluate the complexity and efficiency of GCVPN by comparing its parameter count, training time, and inference speed against representative baselines on the face recognition task. As summarized in Table 4, PCA and LDA involve neither learnable parameters nor measurable training time. For deep learning baselines such as ResNet, VGG, ConvNeXt, and EfficientFormer, GCVPN consistently achieves superior accuracy (AUC and Acc) while maintaining high efficiency. Notably, GCVPN ranks second in parameter count and delivers the highest AUC and Acc. The model reduces inference time by nearly half compared to the best-performing baseline, and GCVPN could handle 30 frames per second of continuous video. The inference speed's result demonstrates strong suitability for real-time occluded image recognition on edge devices.

Table 4: Complexity analysis in obscured facial recognition.

Methods	Model size	Tr time	Inf Time	F1	Acc	Roc_Auc
GCVPN	136.21M	1024.79	**0.0033**	**0.896**	**0.840**	**0.897**
PCA	-	-	-	0.218	0.713	0.736
LDA	-	-	-	0.518	0.551	0.604
resnet	40.84M	1166.12	0.0810	0.791	0.794	0.871
VGG	168.92M	1081.41	0.0062	0.790	0.791	0.870
ConvNeXt	27.82M	13.83	0.0086	0.887	0.824	0.880
Eff-Former	11.92M	68.13	0.0054	0.522	0.526	0.556

4.6 Case Study

4.6.1 Occlusion Case Study. This section assesses GCVPN's robustness in detecting photovoltaic panel anomalies under occlusions like shadows or leaves. Compared to the original GCVPN, simulations with bottom-right occlusion show accurate anomaly detection with preserved features, while top-left occlusion reduces accuracy due to signal loss (Fig. 2). GCVPN effectively handles minimal or non-critical occlusions, ensuring reliable performance in complex urban applications like real-time monitoring.

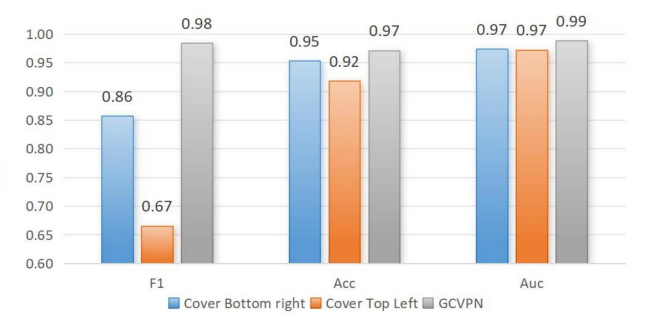

Figure 2: Occlusion study in the abnormal detection task.

4.6.2 Error Analysis Case Experiment. GCVPN shows significant advantages in occluded and blurry face recognition. As shown in Fig. 3, GCVPN could effectively handle complex backgrounds and lighting variations, mitigate interference and adapt to diverse conditions. It also recognizes occluded face images more efficiently, demonstrating superiority in challenging face recognition tasks.

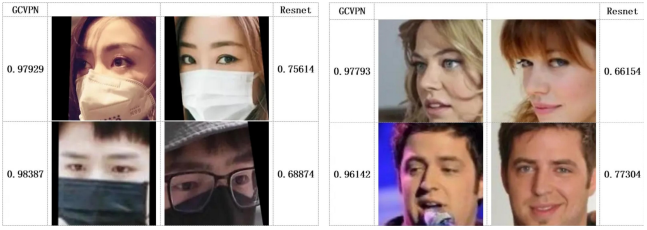

Figure 3: Obvious and unobtrusive facial recognition.

5 Conclusions

This study introduces GCVPN, which combines topology transformers and geometric graph convolution with VGG to improve both accuracy and speed in occluded image recognition. It employs a prior topology transformer, GCN and co-alignment sampling for creating the geometric graph structures to achieve efficient processing. New datasets for photovoltaic panel anomalies and face recognition address data scarcity, thereby enhancing urban surveillance and traffic management. Experiments show GCVPN excels in detecting photovoltaic panels and facial obstructions, enhancing intelligent video surveillance, autonomous driving, and traffic flow prediction. It boosts urban security, optimizes road use, and supports sustainable environmental monitoring.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (31400), the Tianjin Intelligent Manufacturing Special Fund Project (20211093 & 22ZYQYSN00110), and the Emerging Frontiers Cultivation Program of Tianjin University Interdisciplinary Center.

GenAI Usage Disclosure

The authors disclose that Generative Artificial Intelligence (GenAI) tools were used only to polish this manuscript.

References

[1] Fengping An, Jianrong Wang, and Ruijun Liu. 2024. Road Traffic Sign Recognition Algorithm Based on Cascade Attention-Modulation Fusion Mechanism. *IEEE Transactions on Intelligent Transportation Systems* 25, 11 (2024), 17841–17851. doi:10.1109/TITS.2024.3439699

[2] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lucic, and Cordelia Schmid. 2021. ViViT: A Video Vision Transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 6816–6826. doi:10.1109/iccv48922.2021.00676

[3] J Bharadiya. 2023. Convolutional neural networks for image classification. *International Journal of Innovative Science and Research Technology* 8, 5 (2023), 673–677. doi:10.5281/zenodo.8020781

[4] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma. 2015. PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing* 24, 12 (2015), 5017–5032. doi:10.1109/TIP.2015.2475625

[5] Hao Chen, Yaohui Wang, Benoit Lagadec, Antitza Dantcheva, and Francois Bremond. 2021. Joint Generative and Contrastive Learning for Unsupervised Person Re-identification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2004–2013. doi:10.1109/CVPR46437.2021.00204

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 248–255. doi:10.1109/CVPR.2009.5206848

[7] Michael Duhme, Raphael Memmesheimer, and Dietrich Paulus. 2021. *Fusion-GCN: Multimodal Action Recognition Using Graph Convolutional Networks*. Springer International Publishing, 265–281. doi:10.1007/978-3-030-92659-5_17

[8] Mohammed A. El-Shorbagy, Anas Bouaouda, Hossam A. Nabwey, Laith Abualigah, and Fatma A. Hashim. 2024. Advances in Henry Gas Solubility Optimization: A Physics-Inspired Metaheuristic Algorithm With Its Variants and Applications. *IEEE Access* 12, 1 (2024), 26062–26095. doi:10.1109/ACCESS.2024.3365700

[9] Shiming Ge, Jia Li, Qiting Ye, and Zhao Luo. 2017. Detecting Masked Faces in the Wild with LLE-CNNs. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 426–434. doi:10.1109/CVPR.2017.53

[10] A. Geiger, P. Lenz, and R. Urtasun. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3354–3361. doi:10.1109/cvpr.2012.6248074

[11] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. 2021. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2021), 3313–3332. doi:10.1109/TKDE.2021.3130191

[12] Baicang Guo, Hongyu Zhang, Huanhuan Wang, Xinwei Li, and Lisheng Jin. 2024. Adaptive occlusion object detection algorithm based on OL-IoU. *Scientific Reports* 14, 1 (Nov. 2024). doi:10.1038/s41598-024-78959-2

[13] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R. Martin, Ming-Ming Cheng, and Shi-Min Hu. 2022. Attention mechanisms in computer vision: A survey. *Computational Visual Media* 8, 3 (Sept. 2022), 331–368. doi:10.1007/s41095-022-0271-y

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778. doi:10.1109/CVPR.2016.90

[15] Jiseong Heo, Yooseung Wang, and Jihun Park. 2022. Occlusion-aware spatial attention transformer for occluded object recognition. *Pattern recognition letters* 159 (2022), 70–76. doi:10.1016/j.inffus.2024.102229

[16] Yunqing Hu, Xuan Jin, Yin Zhang, Haiwen Hong, Jingfeng Zhang, Yuan He, and Hui Xue. 2021. RAMS-Trans: Recurrent Attention Multi-scale Transformer for Fine-grained Image Recognition. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4239–4248. doi:10.1145/3474085.3475561

[17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2261–2269. doi:10.1109/CVPR.2017.243

[18] Yingzi Huo, Kai Jin, Jiahong Cai, Huixuan Xiong, and Jiacheng Pang. 2023. Vision transformer (Vit)-based applications in image classification. In *2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity)*. IEEE, 135–140. doi:10.1109/BigDataSecurity-HPSC-IDS58521.2023.00033

[19] Walaa N. Ismail, Hessah A. Alsalamah, Mohammad Mehedi Hassan, and Ebtesam Mohamed. 2023. AUTO-HAR: An adaptive human activity recognition framework using an automated CNN architecture design. *Heliyon* 9, 2 (Feb. 2023), e13636. doi:10.1016/j.heliyon.2023.e13636

[20] Madhavi Kondapally, K. Naveen Kumar, Chalavadi Vishnu, and C. Krishna Mohan. 2024. Towards a Transitional Weather Scene Recognition Approach for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems* 25, 6 (2024), 5201–5210. doi:10.1109/TITS.2023.3331882

[21] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324. doi:10.1109/5.726791

[22] Yanyu Li, Geng Yuan, Yang Wen, Ju Hu, Georgios Evangelidis, Sergey Tulyakov, Yanzhi Wang, and Jian Ren. 2022. Efficientformer: Vision transformers at mobilenet speed. *Advances in Neural Information Processing Systems* 35 (2022), 12934–12949. doi:10.5555/3600270.3601210

[23] Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. 2024. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research* 21, 1 (2024), 104–135. doi:10.1007/s11633-023-1459-z

[24] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 9992–10002. doi:10.1109/ICCV48922.2021.00986

[25] Jayawant N. Mandrekar. 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* 5, 9 (Sept. 2010), 1315–1316. doi:10.1097/jto.0b013e3181ec173d

[26] Gaurav Meena, Krishna Kumar Mohbey, Ajay Indian, and Sunil Kumar. 2022. Sentiment Analysis from Images using VGG19 based Transfer Learning Approach. *Procedia Computer Science* 204 (2022), 411–418. doi:10.1016/j.procs.2022.08.050

[27] Lingchen Meng, Hengduo Li, Bor-Chun Chen, Shiyi Lan, Zuxuan Wu, Yu-Gang Jiang, and Ser-Nam Lim. 2022. AdaViT: Adaptive Vision Transformers for Efficient Image Recognition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12299–12308. doi:10.1109/CVPR52688.2022.01199

[28] Daniel Neimark, Omri Bar, Maya Zohar, and Dotan Asselmann. 2021. Video Transformer Network. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 3156–3165. doi:10.1109/ICCVW54120.2021.00355

[29] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), 10684–10695. doi:10.1109/CVPR52688.2022.01042

[30] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks* 20, 1 (2009), 61–80. doi:10.1109/TNN.2008.2005605

[31] Xiaoyan Sun and Xinhao Wang. 2021. *LDA-Enhanced Federated Learning for Image Classification with Missing Modality*. Springer Singapore, 227–241. doi:10.1007/978-981-16-5188-5_17

[32] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9. doi:10.1109/CVPR.2015.7298594

[33] Youhui Tian. 2020. Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. *IEEE Access* 8 (2020), 125731–125744. doi:10.1109/ACCESS.2020.3006097

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010. doi:10.5555/3295222.3295349

[35] Lei Wang, Deke Guo, Huaming Wu, Keqiu Li, and Wei Yu. 2024. TC-GCN: Triple cross-attention and graph convolutional network for traffic forecasting. *Information Fusion* 105 (2024), 102229. doi:10.1016/j.inffus.2024.102229

[36] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. 2023. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16133–16142. doi:10.1109/CVPR52729.2023.01548

[37] Sheng Xu, Qinghua Su, Kaizheng Wan, Xiangyu Qi, and Zhichao Zhang. 2025. Pedestrian Occlusion Target Detection by Integrating Global and Local Features. *Computer Science and Application* 15 (2025), 28. doi:10.1038/s41598-025-05213-8

[38] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence* (New Orleans, Louisiana, USA) *(AAAI'18/IAAI'18/EAAI'18)*. AAAI Press, Article 912, 9 pages.

[39] Guanqi Zhan, Chuanxia Zheng, Weidi Xie, and Andrew Zisserman. 2024. Amodal Ground Truth and Completion in the Wild. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 28003–28013. doi:10.1109/CVPR52733.2024.02645

[40] Jingyu Zhang, Ao Xiang, Yu Cheng, Qin Yang, Liyang Wang, et al. 2024. Research on Detection of Floating Objects in River and Lake Based on AI Image Recognition. *Journal of Artificial Intelligence Practice* 7, 2 (2024), 97–106. doi:10.23977/jaip.2024.070213