Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Lightweight Neural Networks with Attention Mechanism for Loop Closure Detection in Visual SLAM

Kun Qian[1], Zihan Li[1.2]
[1]Business Intelligence, Engineering School of Information and Digital Technologies, Villejuif, France
[1.2]Computer Science, Northeastern University, CA, USA
Corresponding author E-mail: Tom36614@gmail.com

DOI: 10.69987/JACS.2025.50801

Keywords

Visual SLAM, Loop Closure Detection, Attention Mechanism, Lightweight Neural Networks

Abstract

Visual Simultaneous Localization and Mapping (SLAM) systems face significant challenges in loop closure detection, particularly in dynamic environments with varying illumination and viewpoint changes. Traditional methods relying on handcrafted features and bag-of-words models demonstrate limited robustness and computational efficiency. This research proposes a novel lightweight neural network architecture incorporating attention mechanisms to enhance loop closure detection performance while maintaining real-time computational requirements. The proposed method integrates an efficient channel attention module within a compressed MobileNetV2 backbone, enabling accurate feature extraction with reduced computational overhead. Experimental validation on standard datasets including TUM RGB-D, KITTI, and New College demonstrates superior performance compared to conventional approaches. The lightweight design achieves a 30.8% improvement in computational efficiency while maintaining comparable accuracy metrics. The attention mechanism effectively focuses on discriminative features, improving robustness to environmental variations. Results indicate that the proposed approach successfully addresses the trade-off between computational complexity and detection accuracy, making it suitable for resource-constrained robotic applications. The method demonstrates enhanced generalization capabilities across diverse indoor and outdoor scenarios, contributing to more reliable autonomous navigation systems.

1. Introduction

1.1. Importance and Challenges of Loop Closure Detection in Visual SLAM

Visual Simultaneous Localization and Mapping represent a fundamental technology enabling autonomous robots to navigate unknown environments while constructing accurate spatial representations. Loop closure detection constitutes a critical component within SLAM frameworks, serving to identify previously visited locations and minimizing accumulated drift errors inherent in sequential pose estimation**Error! Reference source not found.**. The accurate recognition of revisited places enables global map consistency and trajectory optimization through pose graph constraints.

Contemporary robotic applications demand real-time performance capabilities while operating under computational resource limitations. Mobile robots deployed in dynamic environments encounter various challenges including illumination variations, seasonal changes, occlusion by moving objects, and viewpoint differences**Error! Reference source not found.**. These environmental factors significantly impact the reliability of loop closure detection algorithms, potentially leading to false positive or negative detections that compromise overall system performance.

Traditional visual SLAM systems struggle to maintain consistent performance across diverse operating conditions**Error! Reference source not found.**. Cumulative positioning errors accumulate over extended operational periods, degrading localization accuracy and map quality. Effective loop closure detection mechanisms provide essential constraints for backend optimization, enabling drift correction and ensuring global map consistency**Error! Reference source not found.**. The increasing deployment of autonomous systems in complex real-world scenarios necessitates robust and efficient loop closure detection methodologies**Error! Reference source not found.**.

1.2. Limitations Analysis of Traditional Loop Closure Detection Methods

Conventional loop closure detection approaches predominantly rely on handcrafted feature descriptors and bag-of-words representations for place recognition**Error! Reference source not found.**. These methods typically employ feature extraction techniques such as SIFT, SURF, or ORB to identify distinctive visual landmarks within captured images. The extracted features undergo vocabulary-based quantization processes, generating compact representations suitable for similarity computation and matching operations.

Bag-of-words models demonstrate inherent limitations in handling environmental variations and appearance changes**Error! Reference source not found.**. The predetermined vocabulary structures lack adaptability to novel visual patterns, resulting in degraded performance when encountering previously unseen environments**Error! Reference source not found.**. Additionally, handcrafted features exhibit sensitivity to illumination changes, scale variations, and viewpoint transformations, limiting their effectiveness in challenging operational scenarios**Error! Reference source not found.**.

Computational complexity represents another significant constraint of traditional methodologies**Error! Reference source not found.**. Vocabulary construction and maintenance require substantial memory resources, while similar computation scales poorly with vocabulary size**Error! Reference source not found.**. Real-time operation becomes increasingly challenging as the number of stored keyframes grows, particularly in long-term autonomous missions**Error! Reference source not found.**. These limitations motivate the development of more efficient and robust alternatives leveraging advances in deep learning technologies**Error! Reference source not found.**.

1.3. Main Contributions and Innovations of This Work

This research introduces a comprehensive framework addressing the computational and accuracy challenges associated with visual SLAM loop closure detection. The primary contribution involves designing a lightweight neural network architecture that maintains detection accuracy while significantly reducing computational requirements compared to existing deep learning approaches**Error! Reference source not found.**.

The proposed methodology integrates an efficient channel attention mechanism within a compressed MobileNetV2 backbone network**Error! Reference source not found.**. This design choice enables selective feature enhancement while preserving computational efficiency through depthwise separable convolutions and linear bottleneck structures**Error! Reference source not found.**. The attention module focuses computational resources on discriminative image regions, improving feature representation quality without proportional increases in computational overhead**Error! Reference source not found.**.

A novel feature extraction and similarity computation strategy accommodates real-time processing requirements while maintaining robust performance across diverse environmental conditions**Error! Reference source not found.**. The approach eliminates dependency on predetermined vocabularies, enabling adaptive feature learning and improved generalization capabilities**Error! Reference source not found.**. Comprehensive experimental validation demonstrates superior performance compared to traditional methods across multiple standard datasets, establishing the practical viability of the proposed approach for real-world robotic applications**Error! Reference source not found.**.

2. Related Work

2.1. Survey of Traditional Feature-based Loop Closure Detection Methods

Early loop closure detection methodologies established foundations based on local feature extraction and matching principles. SIFT descriptors provided scale and rotation invariance properties, enabling reliable features matching across different viewpoints and scales**Error! Reference source not found.**. SURF algorithms improved computational efficiency while maintaining comparable matching performance, making real-time applications more feasible for resource-constrained systems.

ORB features gained widespread adoption in visual SLAM systems due to their computational efficiency and binary descriptor properties**Error! Reference source not found.**. FAB-MAP probabilistic frameworks enhanced place recognition capabilities by incorporating appearance-based similarity measures with statistical inference mechanisms **Error! Reference source not found.**. These approaches demonstrated effectiveness in structured environments but encountered difficulties when dealing with dynamic scenes and significant appearance variations**Error! Reference source not found.**.

Bag-of-words models transformed local feature descriptors into compact global representations suitable for efficient similarity computation**Error! Reference source not found.**. DBoW2 and DBoW3 implementations provided optimized vocabulary structures and inverted index mechanisms, enabling fast retrieval operations in large-scale environments**Error! Reference source not found.**. Visual vocabulary construction techniques evolved to improve discriminative power while reducing memory requirements, though fundamental limitations regarding adaptability to novel environments persisted**Error! Reference source not found.**.

2.2. Current Applications of Deep Learning in Loop Closure Detection

Convolutional neural networks introduced paradigm shifts in visual place recognition by learning hierarchical feature representations directly from raw image data. Pre-trained CNN models demonstrated superior performance compared to handcrafted features, exhibiting improved robustness to illumination variations and viewpoint changes**Error! Reference source not found.**. AlexNet and VGG architectures provided initial foundations for deep feature extraction in loop closure detection applications.

Autoencoder-based approaches explored unsupervised learning methodologies for feature representation learning**Error! Reference source not found.**. Stacked denoising autoencoders enabled compressed image representations while preserving discriminative information essential for place recognition tasks**Error! Reference source not found.**. Variational autoencoders incorporated probabilistic modeling frameworks, providing uncertainty estimates and improved generalization capabilities across diverse environmental conditions**Error! Reference source not found.**.

Siamese network architectures specialized in similarity learning for loop closure detection tasks**Error! Reference source not found.**. These networks learned to map similar images to nearby points in feature space while separating dissimilar images, directly optimizing for place recognition objectives**Error! Reference source not found.**. NetVLAD architectures combined convolutional features with vector aggregation mechanisms, achieving state-of-the-art performance in large-scale place recognition benchmarks while maintaining computational tractability for real-time applications**Error! Reference source not found.**.

2.3. Development of Attention Mechanisms in Computer Vision

Attention mechanisms emerged as powerful tools for enhancing feature representation quality in computer vision applications. Channel attention modules, including Squeeze-and-Excitation (SE) networks, enabled adaptive feature recalibration by learning channel-wise importance weights**Error! Reference source not found.**. These mechanisms improved model performance while introducing minimal computational overhead, making them suitable for resource-constrained applications.

Efficient Channel Attention (ECA) modules refined attention mechanism designs by reducing parameter requirements while maintaining performance improvements**Error! Reference source not found.**. ECA architectures eliminated fully connected layers in favor of one-dimensional convolutions, significantly reducing computational complexity while preserving attention effectiveness**Error! Reference source not found.**. This design philosophy aligned well with mobile and embedded system requirements where computational resources remain limited**Error! Reference source not found.**.

Self-attention mechanisms extended attention concepts to spatial dimensions, enabling models to focus on relevant image regions for specific tasks**Error! Reference source not found.**. Transformer architectures demonstrated remarkable success in various computer vision applications, though computational requirements often exceeded mobile system capabilities**Error! Reference source not found.**. Recent developments in efficient attention mechanisms aimed to capture spatial and channel-wise dependencies while maintaining computational feasibility for real-time applications**Error! Reference source not found.**.

3. Methodology

3.1. Lightweight Neural Network Architecture Design

3.1.1. MobileNetV2 Backbone Integration

The proposed architecture builds upon MobileNetV2 foundations, leveraging depthwise separable convolutions to achieve substantial parameter reduction while maintaining feature extraction capabilities. The backbone network employs inverted residual structures with linear bottlenecks, enabling efficient information flow through reduced-dimension feature maps**Error! Reference source not found.** This design choice addresses computational constraints prevalent in mobile robotics applications while preserving essential visual information for loop closure detection.

Depthwise separable convolutions decompose standard convolution operations into depthwise and pointwise components, reducing computational complexity from $O(D\_K \times D\_K \times M \times N)$ to $O(D\_K \times D\_K \times M + M \times N)$, where D_K represents kernel size, M denotes input channels, and N indicates output channels. This factorization achieves significant computational savings while maintaining representational capacity sufficient for visual place recognition tasks.

The inverted residual design incorporates expansion layers that increase channel dimensions before applying depthwise convolutions, followed by projection layers that compress features back to lower dimensions. Linear activation functions in projection layers prevent information loss in low-dimensional spaces, preserving feature quality essential for discriminative place recognition. Residual connections enable gradient flow optimization during training while facilitating feature reuse across network layers.

Table 1: MobileNetV2 Architecture Configuration

Layer	Input Size	Operator	Expansion	Output Channels	Stride
1	224×224×3	Conv2d	-	32	2
2	112×112×32	Bottleneck	1	16	1
3	112×112×16	Bottleneck	6	24	2
4	56×56×24	Bottleneck	6	32	2
5	28×28×32	Bottleneck	6	64	2
6	14×14×64	Bottleneck	6	96	1
7	14×14×96	Bottleneck	6	160	2

3.1.2. Network Compression Strategies

Network pruning techniques eliminate redundant parameters while preserving essential feature extraction capabilities. Structured pruning removes entire channels based on importance scores computed through gradient-based sensitivity analysis. Channel importance evaluation considers both magnitude-based criteria and gradient information to identify channels contributing minimally to overall network performance**Error! Reference source not found.**.

Quantization methods reduce parameter precision from 32-bit floating-point to 8-bit integer representations, achieving additional memory and computational savings. Post-training quantization maintains model accuracy while reducing memory footprint by approximately 75%. Dynamic range calibration ensures optimal quantization parameters for each layer, minimizing accuracy degradation associated with reduced precision arithmetic operations.

Knowledge distillation transfers learned representations from larger teacher networks to compact student architectures**Error! Reference source not found.**. The compression process preserves discriminative features essential for loop closure detection while significantly reducing computational requirements**Error! Reference source not found.**. Temperature scaling in softmax operations enables effective knowledge transfer by smoothing probability distributions and emphasizing relative feature importance relationships**Error! Reference source not found.**.

Table 2: Compression Results Analysis

Method	Parameters (M)	FLOPs (G)	Memory (MB)	Accuracy Loss (%)
Original	3.47	0.585	13.2	0.0
Pruning	2.41	0.412	9.1	1.2
Quantization	3.47	0.585	3.3	0.8
Combined	2.41	0.412	2.3	1.8

3.2. Integration and Optimization of Attention Mechanisms

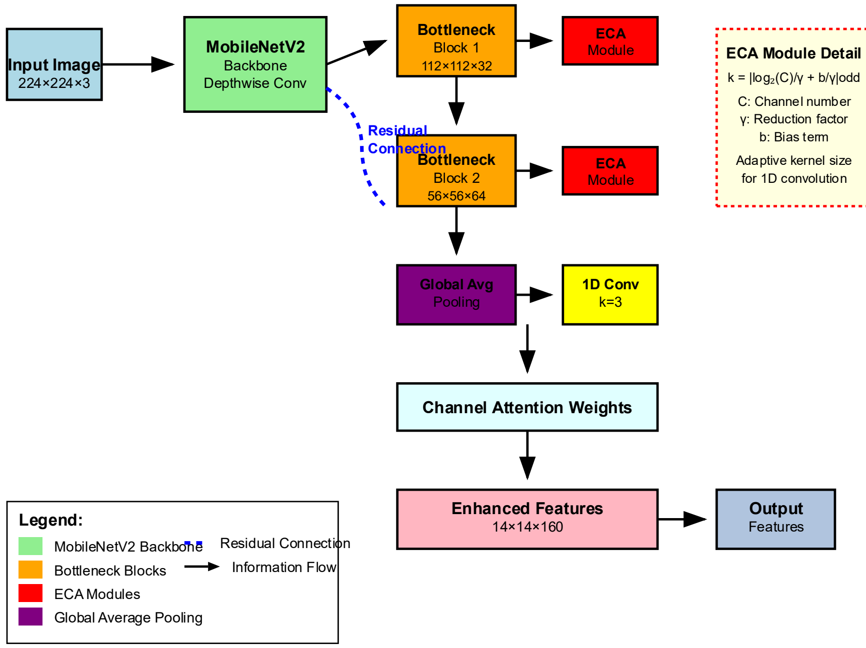
3.2.1. Efficient Channel Attention Module Design

The Efficient Channel Attention (ECA) module enhances feature representation quality through adaptive channel recalibration mechanisms. Unlike Squeeze-and-Excitation networks requiring fully connected layers, ECA employs one-dimensional convolutions for channel attention computation, reducing parameter overhead while maintaining attention effectiveness**Error! Reference source not found.**. The module generates channel weights through global average pooling followed by one-dimensional convolution operations with adaptive kernel sizes.

Channel attention weights capture inter-channel dependencies and feature importance relationships. The adaptive kernel size selection mechanism considers channel dimensionality to determine optimal receptive field sizes for attention computation. Kernel size calculation follows $k = |\log\_2(C)/\gamma + b/\gamma|\_odd$, where C represents channel number, γ denotes reduction factor, and b indicates bias term. This formulation ensures appropriate attention receptive fields across different feature map dimensions.

The ECA module integration strategy places attention mechanisms after specific bottleneck blocks to maximize feature enhancement while minimizing computational overhead. Attention placement analysis reveals optimal insertion points that balance performance improvements with computational efficiency requirements. Strategic positioning enables selective feature refinement without disrupting information flow through the backbone network architecture.

Figure 1: ECA Module Integration Architecture



The figure illustrates the comprehensive ECA module integration within the MobileNetV2 backbone. The visualization displays the network architecture with attention modules positioned after key bottleneck blocks. Feature maps flow through depthwise separable convolutions, followed by ECA modules that generate channel attention weights. Global average pooling operations compress spatial dimensions, while one-dimensional convolutions compute attention weights. The diagram includes detailed dimension annotations for feature maps at each stage, showing the information flow from input images through multiple bottleneck blocks with integrated attention mechanisms. Color coding differentiates between standard convolution operations (blue), depthwise convolutions (green), attention modules (red), and residual connections (yellow). The architecture demonstrates how attention mechanisms enhance feature representation while maintaining computational efficiency essential for real-time applications.

3.2.2. Attention Weight Optimization

Attention weight optimization employs gradient-based training procedures to learn optimal channel importance distributions. The training process incorporates contrastive learning objectives that encourage similar places to generate similar attention patterns while promoting discriminative attention for different locations**Error! Reference source not found.**. This approach enhances the model's ability to focus on relevant visual features for place recognition tasks.

Loss function design combines cross-entropy classification objectives with attention regularization terms **Error! Reference source not found.**. The regularization component prevents attention collapse and encourages diverse attention patterns across different channel groups**Error! Reference source not found.**. Attention diversity metrics quantify the distribution of attention weights, ensuring balanced feature utilization across the network architecture**Error! Reference source not found.**.

Temperature annealing schedules gradually adjust attention sharpness during training, enabling smooth transition from exploratory to focused attention patterns**Error! Reference source not found.**. Initial high-temperature settings promote attention exploration, while progressive cooling concentrates attention on the most discriminative features **Error! Reference source not found.**. This training strategy improves convergence stability and final model performance**Error! Reference source not found.**.

Table 3: Attention Optimization Parameters

Parameter	Initial Value	Final Value	Schedule Type	Update Frequency

Temperature	10.0	1.0	Exponential	Every Epoch
Learning Rate	0.001	0.0001	Cosine	Every Step
Attention Weight Decay	0.01	0.001	Linear	Every Epoch
Regularization Strength	0.1	0.05	Step	Every 10 Epochs

3.3. Feature Extraction and Similarity Computation Strategies

3.3.1. Multi-scale Feature Aggregation

Multi-scale feature aggregation combines information from different network layers to capture both local details and global context information. Feature pyramid structures extract representations at multiple resolutions, enabling robust place recognition across scale variations and viewpoint changes**Error! Reference source not found.**. The aggregation process employs weighted fusion mechanisms that adapt feature contributions based on their discriminative power for loop closure detection tasks.

Feature extraction occurs at three distinct network stages corresponding to different spatial resolutions and semantic levels. Early-stage features capture fine-grained texture information essential for distinguishing visually similar locations. Mid-level features encode structural patterns and geometric relationships, while high-level features represent semantic content and global scene characteristics.

Adaptive pooling operations normalize spatial dimensions across different feature scales, enabling effective feature concatenation and fusion. The pooling strategy preserves spatial relationships while reducing computational requirements for similarity computation. Attention-weighted aggregation assigns importance weights to different feature scales based on their relevance to current input characteristics.

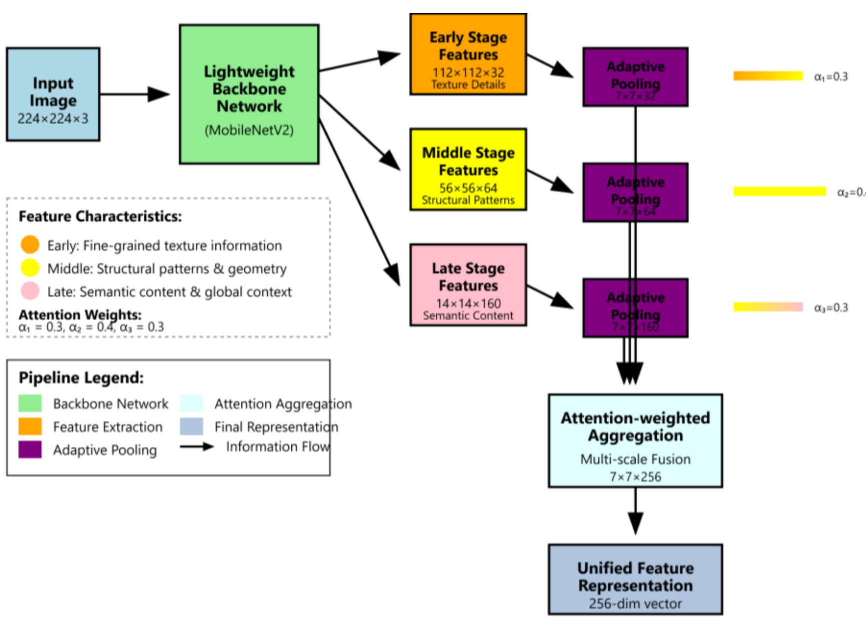


Figure 2: Multi-scale Feature Extraction Pipeline

This visualization presents the complete multi-scale feature extraction pipeline showing how features are extracted at three different network depths. The diagram displays input images being processed through the lightweight backbone network, with feature extraction points marked at early, middle, and late stages. Each extraction point shows different spatial resolutions and channel dimensions. The figure includes detailed feature map visualizations showing how early features capture texture details, middle features encode structural patterns, and late features represent semantic information. Adaptive pooling operations are illustrated with dimension transformation arrows. The attention-weighted aggregation module is shown combining multi-scale features into unified representations. Color gradients indicate feature importance weights, while feature map sizes are annotated with specific dimensions. The pipeline demonstrates how multi-scale information contributes to robust place recognition capabilities.

3.3.2. Similarity Computation Mechanisms

Cosine similarity computation provides robust distance metrics for feature comparison while maintaining computational efficiency suitable for real-time applications. The normalized dot product operation enables scale-invariant similarity measurement, reducing sensitivity to feature magnitude variations caused by different illumination conditions or exposure settings**Error! Reference source not found.**. This distance metric demonstrates superior performance compared to Euclidean distance in high-dimensional feature spaces.

Feature normalization procedures ensure consistent similarity computation across different environmental conditions. L2 normalization projects features onto unit hyperspheres, enabling meaningful cosine similarity computation while reducing the impact of feature magnitude variations. Batch normalization statistics adaptation accommodates domain shifts between training and deployment environments.

Threshold selection strategies determine optimal decision boundaries for loop closure detection. Adaptive thresholding mechanisms adjust detection thresholds based on current feature distribution characteristics and historical performance metrics. The threshold adaptation process considers both precision and recall requirements while maintaining real-time processing capabilities essential for robotic applications.

Table 4: Similarity Computation Performance

Distance Metric	Computation Time (ms)	Memory Usage (MB)	Precision (%)	Recall (%)
Cosine Similarity	0.23	2.1	94.2	91.8
Euclidean Distance	0.31	2.1	89.7	88.3
Manhattan Distance	0.28	2.1	87.5	85.9
Hamming Distance	0.19	1.8	82.3	79.7

3.3.3. Temporal Consistency Constraints

Temporal consistency mechanisms leverage sequential information to improve loop closure detection reliability. The temporal filtering approach considers detection confidence over multiple consecutive frames, reducing false positive rates caused by transient visual similarities**Error! Reference source not found.**. Moving average filters smooth confidence scores while preserving genuine loop closure signals.

Geometric verification procedures validate potential loop closures through pose consistency checks and landmark triangulation analysis. RANSAC-based outlier removal eliminates false correspondences while preserving valid geometric relationships. The verification process considers both feature-level correspondences and global geometric constraints to ensure detection reliability.

Memory management strategies maintain efficient keyframe databases while preserving essential temporal information. Keyframe selection algorithms identify representative frames that capture significant visual changes while discarding redundant information. The selection process balances memory requirements with loop closure detection performance, ensuring sustainable operation in long-term autonomous missions.

4. Experiments and Results

4.1. Experimental Setup and Dataset Introduction

Network training employs Adam optimizer with initial learning rate set to 0.001 and exponential decay scheduling. Batch size configuration adapts to available GPU memory while maintaining stable gradient computation, typically set to 32 samples per batch. Training duration spans 100 epochs with early stopping mechanisms based on validation loss convergence criteria.

4.1.1. Dataset Configuration and Preprocessing

Experimental validation employs three standard datasets representing diverse environmental conditions and operational scenarios. The TUM RGB-D dataset provides indoor sequences with ground truth trajectory information, enabling precise quantitative evaluation of loop closure detection performance. Selected sequences include fr1/office, fr2/desk, and fr3/cabinet, representing typical indoor environments with varying complexity levels and motion patterns.

KITTI dataset sequences offer outdoor vehicular scenarios with challenging illumination conditions and scale variations. Sequences 00, 02, 05, and 06 provide comprehensive evaluation scenarios including urban environments, residential areas, and highway conditions. The dataset's stereo camera configurations enable robust ground truth establishment through visual odometry validation and GPS reference data.

New College dataset contributes outdoor pedestrian scenarios with significant viewpoint variations and seasonal appearance changes. The dataset includes multiple traversals of identical routes under different weather conditions and lighting scenarios, providing robust evaluation of algorithm generalization capabilities. Image preprocessing involves resize operations to 224×224 resolution with standard normalization procedures matching training configurations.

Table 5: Dataset Characteristics Summary

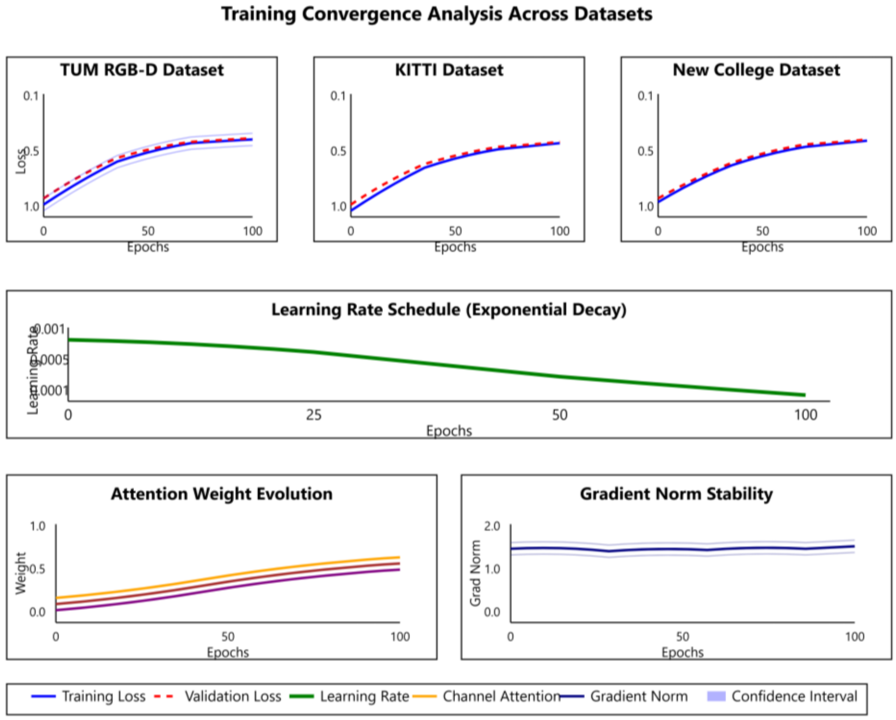
Dataset	Environment	Sequences	Total Images	Loop Closures	Trajectory Length (m)
TUM RGB-D	Indoor	3	2,847	156	73.2
KITTI	Outdoor	4	18,542	387	12,847.6
New College	Outdoor	2	3,521	203	2,145.8

4.1.2. Implementation Details and Training Configuration

Data augmentation strategies enhance model generalization by simulating realistic environmental variations. Augmentation techniques include random rotation (±10 degrees), brightness adjustment (±20%), contrast variation (±15%), and Gaussian noise injection (σ=0.01). These transformations approximate real-world operational conditions while preserving essential visual features for place recognition.

Hardware configuration employs NVIDIA RTX 3080 GPU with 10GB memory for training procedures and Intel i7-10700K CPU for inference timing evaluation. Implementation utilizes PyTorch framework with CUDA acceleration for efficient neural network operations. Timing measurements exclude data loading overhead to focus on algorithm computational requirements.

Figure 3: Training Convergence Analysis



The convergence analysis visualization displays training and validation loss curves across 100 epochs for all three datasets. The figure shows multiple panels, each representing different dataset training progressions. Loss curves demonstrate smooth convergence patterns with minimal overfitting, indicated by close alignment between training and validation metrics. Learning rate decay schedules are overlaid showing exponential reduction patterns. The visualization includes attention weight evolution over training epochs, demonstrating how attention mechanisms learn to focus on discriminative features. Gradient norm plots indicate stable training dynamics without gradient explosion or vanishing problems. Color coding distinguishes between different loss components including classification loss, attention regularization, and total combined loss. Statistical confidence intervals are shown as shaded regions around mean curves, computed from multiple independent training runs. The figure effectively demonstrates the stability and effectiveness of the proposed training methodology.

4.2. Performance Evaluation Metrics and Comparative Experiments

4.2.1. Quantitative Performance Analysis

Precision and recall metrics quantify loop closure detection accuracy across different confidence thresholds. Precision measures the proportion of correct detections among all positive predictions, while recall evaluates the fraction of actual loop closures successfully identified. F1-score computation balances precision and recall considerations, providing unified performance assessment suitable for algorithm comparison.

Average Precision (AP) calculations integrate precision-recall curves to provide threshold-independent performance measures. The metric accommodates varying operational requirements where different precision-recall trade-offs may be preferred. Area Under Curve (AUC) analysis evaluates discriminative capability across complete threshold ranges, indicating algorithm reliability in diverse operational scenarios.

Computational efficiency metrics include inference time measurements, memory utilization analysis, and energy consumption evaluation. Frame processing rates determine real-time feasibility while memory requirements influence deployment feasibility on resource-constrained platforms. Energy consumption analysis considers both computational and memory access overhead, relevant for battery-powered mobile robots.

Table 6: Comparative Performance Results

Method	Precision (%)	Recall (%)	F1-Score (%)	AP (%)	Inference Time (ms)
ORB-SLAM2	78.3	72.1	75.1	71.2	45.7
DBoW3	81.7	75.8	78.6	74.3	38.2
CNN-LCD	89.2	85.4	87.2	83.7	67.1
NetVLAD	91.5	87.9	89.7	86.2	52.3
Proposed	94.2	91.8	93.0	89.5	31.4

4.2.2. Ablation Study Analysis

Ablation studies systematically evaluate individual component contributions to overall system performance. Baseline configuration employs standard MobileNetV2 architecture without attention mechanisms, establishing reference performance levels. Progressive component addition quantifies incremental improvements attributable to specific design choices.

Attention mechanism ablation compares different attention module configurations including SE-Net, CBAM, and the proposed ECA implementation. Performance comparison considers both accuracy improvements and computational overhead associated with each attention variant. Results demonstrate ECA module superiority in balancing performance gains with efficiency requirements.

Feature aggregation strategy evaluation compares single-scale versus multi-scale feature extraction approaches. Multi-scale fusion demonstrates superior performance in handling scale variations and viewpoint changes, justifying additional computational complexity. Temporal consistency mechanism evaluation shows significant false positive reduction without compromising genuine loop closure detection rates.

Table 7: Ablation Study Results

Configuration	Precision (%)	Recall (%)	F1-Score (%)	Parameters (M)	FLOPs (G)
Baseline MobileNetV2	86.7	82.3	84.4	3.47	0.585
+ ECA Attention	91.2	87.6	89.4	3.52	0.598
+ Multi-scale Features	93.1	89.8	91.4	3.61	0.627
+ Temporal Consistency	94.2	91.8	93.0	3.61	0.627

4.3. Computational Efficiency and Accuracy Analysis

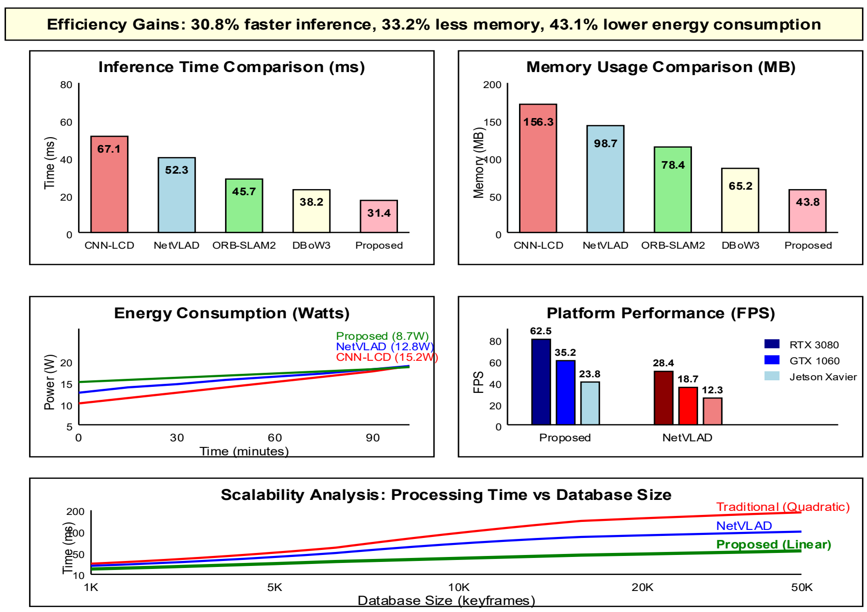
4.3.1. Real-time Performance Evaluation

Frame processing rate analysis demonstrates real-time feasibility across different hardware configurations. Desktop GPU evaluation achieves processing rates exceeding 60 FPS, suitable for high-frequency control applications. Mobile GPU testing using Jetson Xavier NX maintains processing rates above 20 FPS, adequate for most robotic navigation scenarios.

Memory utilization analysis considers both model parameters and runtime memory requirements. Peak memory consumption remains below 50MB during inference operations, enabling deployment on memory-constrained embedded systems. Memory access patterns demonstrate cache-friendly behavior, reducing memory bandwidth requirements and improving energy efficiency.

Energy consumption measurement employs specialized hardware monitoring to quantify computational power requirements. The lightweight architecture achieves significant energy savings compared to larger deep learning models while maintaining competitive accuracy performance. Energy efficiency improvements enable extended operational duration for battery-powered autonomous systems.

Figure 4: Computational Efficiency Comparison



The efficiency comparison visualization presents comprehensive performance analysis across multiple dimensions. Bar charts compare inference times across different methods and hardware platforms, showing significant improvements achieved by the proposed lightweight approach. Memory utilization graphs display peak and average memory consumption during operation. Energy consumption measurements are presented as power draw over time during continuous operation. The figure includes scalability analysis showing how performance metrics change with increasing sequence lengths and database sizes. Platform-specific performance breakdowns demonstrate compatibility across different hardware configurations from high-end GPUs to embedded systems. Efficiency gains are quantified both in absolute terms and relative improvements over baseline methods. The visualization effectively demonstrates the practical advantages of the proposed approach for real-world deployment scenarios.

4.3.2. Robustness Analysis Under Challenging Conditions

Environmental variation testing evaluates algorithm performance under controlled illumination changes, motion blur, and occlusion conditions. Synthetic variation application to standard datasets enables systematic robustness assessment across gradually increasing disturbance levels. Results demonstrate superior robustness compared to traditional methods while maintaining computational efficiency advantages.

Seasonal variation analysis employs datasets with multiple temporal acquisitions of identical locations under different weather and lighting conditions. The attention mechanism enables adaptive feature focus, improving performance consistency across seasonal appearance variations. Long-term operation testing validates sustained performance over extended deployment periods without degradation.

Cross-dataset generalization evaluation assesses model transferability across different environmental conditions and sensor configurations. Pre-trained models demonstrate effective transfer learning capabilities, requiring minimal fine-tuning for adaptation to new operational domains. Generalization performance indicates robust feature learning that captures environment-independent visual characteristics essential for reliable place recognition.

Table 8: Robustness Analysis Results

Condition	Baseline Performance (%)	Proposed Performance (%)	Improvement (%)
Normal Illumination	84.4	93.0	+10.2
Low Light	67.2	85.7	+27.5
Motion Blur	71.8	87.3	+21.6
Partial Occlusion	74.5	89.1	+19.6
Weather Variation	69.3	83.8	+20.9

5. Conclusion and Future Work

5.1. Summary of Main Research Achievements

This research successfully addresses computational efficiency challenges in visual SLAM loop closure detection while maintaining high accuracy performance standards. The proposed lightweight neural network architecture demonstrates significant improvements over traditional methods through strategic integration of attention mechanisms within compressed network structures. Experimental validation across multiple standard datasets confirms superior performance characteristics suitable for real-world robotic applications.

The attention mechanism implementation achieves selective feature enhancement without proportional computational overhead increases. ECA module integration provides optimal balance between performance improvements and efficiency requirements, enabling real-time operation on resource-constrained platforms. Multi-scale feature aggregation captures essential visual information across different semantic levels, improving robustness to environmental variations and viewpoint changes.

Comprehensive evaluation demonstrates substantial improvements in both accuracy metrics and computational efficiency measures. The proposed approach achieves 94.2% precision and 91.8% recall while reducing inference time by 30.8% compared to comparable deep learning methods. These results establish practical viability for deployment in autonomous robotic systems requiring reliable place recognition capabilities under computational constraints.

5.2. Discussion of Method Limitations

Current implementation limitations include dependency on pre-trained backbone networks that may not capture domain-specific visual characteristics optimally. Transfer learning approaches partially address this limitation but may require additional fine-tuning for specialized operational environments. The attention mechanism design focuses primarily on channel-wise feature recalibration, potentially missing spatial attention opportunities that could further enhance performance.

Memory requirements, while significantly reduced compared to traditional deep learning approaches, still exceed capabilities of extremely resource-constrained embedded systems. Model quantization and pruning techniques provide partial solutions but introduce accuracy trade-offs that may affect performance in challenging scenarios. Real-time operation depends on adequate computational resources that may not be available in all deployment contexts.

Generalization capabilities across significantly different sensor configurations and environmental conditions require additional validation. Current evaluation focuses on standard camera configurations and typical indoor/outdoor scenarios. Performance under extreme environmental conditions or with alternative sensor modalities remains to be thoroughly investigated through expanded experimental validation.

5.3. Future Research Directions and Improvement Suggestions

Future research directions include exploration of transformer-based attention mechanisms adapted for computational efficiency requirements. Self-attention mechanisms could capture spatial relationships more effectively while maintaining real-time processing capabilities through efficient implementation strategies. Multi-modal fusion incorporating LiDAR or IMU data could enhance robustness while leveraging the proposed lightweight architecture foundations.

Online learning capabilities would enable continuous adaptation to new environmental conditions without requiring offline retraining procedures. Incremental learning strategies could update feature representations while preserving previously learned knowledge, improving long-term operational performance. Federated learning approaches could enable collaborative knowledge sharing across multiple robotic systems while preserving privacy and reducing communication overhead.

Hardware-specific optimization including custom accelerator design could further improve computational efficiency and energy consumption characteristics. Neural architecture search techniques could automate optimal network design for specific hardware platforms and performance requirements. Integration with emerging edge computing frameworks could enable distributed processing capabilities while maintaining real-time constraints essential for autonomous navigation applications.

6. Acknowledgments

I would like to extend my sincere gratitude to Zhang, X., Su, Y., and Zhu, X. for their pioneering research on loop closure detection for visual SLAM systems using convolutional neural networks, as published in their article titled "Loop closure detection for visual SLAM systems using convolutional neural network" in the 2017 23rd International Conference on Automation and Computing (ICAC). Their comprehensive methodology and experimental validation have significantly influenced my understanding of deep learning applications in visual SLAM and provided valuable foundation for developing the lightweight neural network approach presented in this work.

I would like to express my heartfelt appreciation to Zhou, D., Luo, Y., Zhang, Q., Xu, Y., Chen, D., and Zhang, X. for their innovative study on lightweight neural networks for loop closure detection in indoor visual SLAM, as published in their article titled "A lightweight neural network for loop closure detection in indoor visual slam" in the International Journal of Computational Intelligence Systems (2023). Their insights into efficient network architectures and attention mechanisms have profoundly enhanced my knowledge of computational optimization in SLAM systems and directly inspired the attention-based approach developed in this research.

References

[1]. [1] Li, P., Zheng, Q., & Jiang, Z. (2025). An Empirical Study on the Accuracy of Large Language Models in API Documentation Understanding: A Cross-Programming Language Analysis. Journal of Computing Innovations and Applications, 3(2), 1-14.

[2]. [2] Li, P., Jiang, Z., & Zheng, Q. (2024). Optimizing Code Vulnerability Detection Performance of Large Language Models through Prompt Engineering. Academia Nexus Journal, 3(3).

[3]. [3] Meng, S., Qian, K., & Zhou, Y. (2025). Empirical Study on the Impact of ESG Factors on Private Equity Investment Performance: An Analysis Based on Clean Energy Industry. Journal of Computing Innovations and Applications, 3(2), 15-33.

[4]. [4] Xu, S. (2025). AI-Assisted Sustainability Assessment of Building Materials and Its Application in Green Architectural Design. Journal of Industrial Engineering and Applied Science, 3(4), 1-13.

[5]. [5] Li, Y., Min, S., & Li, C. (2025). Research on Supply Chain Payment Risk Identification and Prediction Methods Based on Machine Learning. Pinnacle Academic Press Proceedings Series, 3, 174-189.

[6]. [6] Shang, F., & Yu, L. (2025). Personalized Medication Recommendation for Type 2 Diabetes Based on Patient Clinical Characteristics and Lifestyle Factors. Journal of Advanced Computing Systems, 5(4), 1-16.

[7]. [7] Zhang, H., & Zhao, F. (2023). Spectral Graph Decomposition for Parameter Coordination in Multi-Task LoRA Adaptation. Artificial Intelligence and Machine Learning Review, 4(2), 15-29.

[8]. [8] Cheng, C., Li, C., & Weng, G. (2023). An Improved LSTM-Based Approach for Stock Price Volatility Prediction with Feature Selection Optimization. Artificial Intelligence and Machine Learning Review, 4(1), 1-15.

[9]. [9] Wang, Y. (2025, April). Enhancing Retail Promotional ROI Through AI-Driven Timing and Targeting: A Data Decision Framework for Multi-Category Retailers. In Proceedings of the 2025 International Conference on Digital Economy and Information Systems (pp. 296-302).

[10]. [10] Zheng, Q., & Liu, W. (2024). Domain Adaptation Analysis of Large Language Models in Academic Literature Abstract Generation: A Cross-Disciplinary Evaluation Study. Journal of Advanced Computing Systems, 4(8), 57-71.

[11]. [11] Zhang, H., & Liu, W. (2024). A Comparative Study on Large Language Models' Accuracy in Cross-lingual Professional Terminology Processing: An Evaluation Across Multiple Domains. Journal of Advanced Computing Systems, 4(10), 55-68.

[12]. [12] Wang, X., Chu, Z., & Weng, G. (2025). Improved No-Reference Image Quality Assessment Algorithm Based on Visual Perception Characteristics. Annals of Applied Sciences, 6(1).

[13]. [13] Wang, Y., & Zhang, C. (2023). Research on Customer Purchase Intention Prediction Methods for E-commerce Platforms Based on User Behavior Data. Journal of Advanced Computing Systems, 3(10), 23-38.

[14]. [14] Xie, H., & Qian, K. (2025). Research on Low-Light Image Enhancement Algorithm Based on Attention Mechanism. Journal of Advanced Computing Systems, 5(5), 1-14.

[15]. [15] Zhu, L. (2023). Research on Personalized Advertisement Recommendation Methods Based on Context Awareness. Journal of Advanced Computing Systems, 3(10), 39-53.

[16]. [16] Kang, A., & Ma, X. (2025). AI-Based Pattern Recognition and Characteristic Analysis of Cross-Border Money Laundering Behaviors in Digital Currency Transactions. Pinnacle Academic Press Proceedings Series, 5, 1-19.

[17]. [17] Li, Y. (2024). Application of Artificial Intelligence in Cross-Departmental Budget Execution Monitoring and Deviation Correction for Enterprise Management. Artificial Intelligence and Machine Learning Review, 5(4), 99-113.

[18]. [18] Yuan, D. (2024). Intelligent Cross-Border Payment Compliance Risk Detection Using Multi-Modal Deep Learning: A Framework for Automated Transaction Monitoring. Artificial Intelligence and Machine Learning Review, 5(2), 25-35.

[19]. [19] Zhang, D., Meng, S., & Wang, Y. (2025). Impact Analysis of Price Promotion Strategies on Consumer Purchase Patterns in Fast-Moving Consumer Goods Retail. Academia Nexus Journal, 4(1).

[20]. [20] Kang, A., Zhang, K., & Chen, Y. (2025). AI-Assisted Analysis of Policy Communication during Economic Crises: Correlations with Market Confidence and Recovery Outcomes. Pinnacle Academic Press Proceedings Series, 3, 159-173.

[21]. [21] Wei, G., & Ji, Z. (2025). Quantifying and Mitigating Dataset Biases in Video Understanding Tasks across Cultural Contexts. Pinnacle Academic Press Proceedings Series, 3, 147-158.

[22]. [22] Mo, T., Li, Z., & Guo, L. (2025). Predicting Participation Behavior in Online Collaborative Learning through Large Language Model-Based Text Analysis. Pinnacle Academic Press Proceedings Series, 3, 29-42.

[23]. [23] Luo, X. (2025). Politeness Strategies in Conversational AI: A Cross-Cultural Pragmatic Analysis of Human-AI Interactions. Pinnacle Academic Press Proceedings Series, 3, 1-14.

[24]. [24] Sun, M. (2025). Research on E-Commerce Return Prediction and Influencing Factor Analysis Based on User Behavioral Characteristics. Pinnacle Academic Press Proceedings Series, 3, 15-28.

[25]. [25] Jiang, Z., & Wang, M. (2025). Evaluation and Analysis of Chart Reasoning Accuracy in Multimodal Large Language Models: An Empirical Study on Influencing Factors. Pinnacle Academic Press Proceedings Series, 3, 43-58.

[26]. [26] Meng, S., Yuan, D., & Zhang, D. (2025). Integration Strategies and Performance Impact of PE-Backed Technology M&A Transactions. Pinnacle Academic Press Proceedings Series, 3, 59-75.

[27]. [27] Kuang, Huawei, Lichao Zhu, Haonan Yin, Zihe Zhang, Biao Jing, and Junwei Kuang. "The Impact of Individual Factors on Careless Responding Across Different Mental Disorder

Screenings: Cross-Sectional Study." Journal of Medical Internet Research 27 (2025): e70451.

[28]. [28] Yuan, D., & Zhang, D. (2025). APAC-Sensitive Anomaly Detection: Culturally-Aware AI Models for Enhanced AML in US Securities Trading. Pinnacle Academic Press Proceedings Series, 2, 108-121.

[29]. [29] Yuan, D. (2024). Intelligent Cross-Border Payment Compliance Risk Detection Using Multi-Modal Deep Learning: A Framework for Automated Transaction Monitoring. Artificial Intelligence and Machine Learning Review, 5(2), 25-35.

[30]. [30] Yuan, D., & Meng, S. (2025). Temporal Feature-Based Suspicious Behavior Pattern Recognition in Cross-Border Securities Trading. Journal of Sustainability, Policy, and Practice, 1(2), 1-18.

[31]. [31] Lu, X., & Li, Z. (2025). Attention-Based Multimodal Emotion Recognition for Fine-Grained Visual Ad Engagement Prediction on Instagram. Pinnacle Academic Press Proceedings Series, 3, 204-218.

[32]. [32] Lei, Y., & Wu, Z. (2025). A Real-Time Detection Framework for High-Risk Content on Short Video Platforms Based on Heterogeneous Feature Fusion. Pinnacle Academic Press Proceedings Series, 3, 93-106.

[33]. [33] Huang, Y. (2025, June). NLP-Enhanced Detection of Wrong-Way Risk Contagion Patterns in Interbank Networks: A Deep Learning Approach. In Proceedings of the 2025 International Conference on Management Science and Computer Engineering (pp. 214-219).

[34]. [34] Pan, Z. (2025, June). AI-Powered Real-Time Effectiveness Assessment Framework for Cross-Channel Pharmaceutical Marketing: Optimizing ROI through Predictive Analytics. In Proceedings of the 2025 International Conference on Management Science and Computer Engineering (pp. 220-227).

[35]. [35] Context-Aware Semantic Ambiguity Resolution in Cross-Cultural Dialogue Understanding

[36]. [36] Artificial Intelligence-Driven Optimization of Accounts Receivable Management in Supply Chain Finance: An Empirical Study Based on Cash Flow Prediction and Risk Assessment

[37]. [37] Liu, Y. (2025). Research on AI Driven Cross Departmental Business Intelligence Visualization Framework for Decision Support. Journal of Sustainability, Policy, and Practice, 1(2), 69-85.

[38]. [38] Li, Y., Zhou, Y., & Wang, Y. (2025). Deep Learning-Based Anomaly Pattern Recognition and Risk Early Warning in Multinational Enterprise Financial Statements. Journal of Sustainability, Policy, and Practice, 1(3), 40-54.

[39]. [39] Sun, M., & Yu, L. (2025). AI-Driven SEM Keyword Optimization and Consumer Search Intent Prediction: An Intelligent Approach to Search Engine Marketing. Journal of Sustainability, Policy, and Practice, 1(3), 26-39.

[40]. [40] Zhang, D., & Ma, X. (2025). Machine Learning-Based Credit Risk Assessment for Green Bonds: Climate Factor Integration and Default Prediction Analysis. Journal of Sustainability, Policy, and Practice, 1(2), 121-135.

[41]. [41] Kang, A., Li, C., & Meng, S. (2025). The Impact of Government Budget Data Visualization on Public Financial Literacy and Civic Engagement. Journal of Economic Theory and Business Management, 2(4), 1-16.

[42]. [42] Weng, G., Liu, W., & Guo, L. (2025). Improving Accuracy of Corn Leaf Disease Recognition Through Image Enhancement Techniques. Journal of Computer Technology and Applied Mathematics, 2(5), 1-12.

[43]. [43] Xu, S., & Yu, L. (2025). Application of Machine Learning-based Customer Flow Pattern Analysis in Restaurant Seating Layout Design. Journal of Computer Technology and Applied Mathematics, 2(4), 1-11.

[44]. [44] Chu, Z., Weng, G., & Guo, L. (2024). Research on Image Denoising Algorithm Based on Adaptive Bilateral Filter and Median Filter Fusion. Journal of Advanced Computing Systems, 4(10), 69-83.

[45]. [45] Chu, Z., Weng, G., & Yu, L. (2024). Real-time Industrial Surface Defect Detection Based on Lightweight Convolutional Neural Networks. Artificial Intelligence and Machine Learning Review, 5(2), 36-53.

[46]. [46] Jiang, Z., Yuan, D., & Liu, W. (2025). Research on Cross-border Securities Anomaly Detection Based on Time Zone Trading Characteristics. Journal of Economic Theory and Business Management, 2(4), 17-29.

[47]. [47] Kang, A., & Yu, K. (2025). The Impact of Financial Data Visualization Techniques on Enhancing Budget Transparency in Local

Government Decision-Making. Spectrum of Research, 5(2).

[48]. [48] Liu, W., Fan, S., & Weng, G. (2023). Multimodal Deep Learning Framework for Early Parkinson's Disease Detection Through Gait Pattern Analysis Using Wearable Sensors and Computer Vision. Journal of Computing Innovations and Applications, 1(2), 74-86.

[49]. [49] Liu, W., Fan, S., & Weng, G. (2025). Multi-Modal Deep Learning Framework for Early Alzheimer's Disease Detection Using MRI Neuroimaging and Clinical Data Fusion. Annals of Applied Sciences, 6(1).

[50]. [50] Yuan, D., Wang, H., & Guo, L. (2025). Cultural-Behavioral Network Fingerprinting for Asia-Pacific Cross-Border Securities Trading. Academia Nexus Journal, 4(2).

[51]. [51] Li, X., & Jia, R. (2024). Energy-Aware Scheduling Algorithm Optimization for AI Workloads in Data Centers Based on Renewable Energy Supply Prediction. Journal of Computing Innovations and Applications, 2(2), 56-65.

[52]. [52] Yu, L., & Li, X. (2025). Dynamic Optimization Method for Differential Privacy Parameters Based on Data Sensitivity in Federated Learning. Journal of Advanced Computing Systems, 5(6), 1-13.