

RESEARCH ARTICLES

A Lightweight Traffic Sign Recognition Network for Autonomous Driving: Integrating Dual-Branch Attention and Weather-Robust Augmentation

Ting Wang[†] and Zhihao Weng^{†‡}

[†]Changsha Medical University, Changsha, Hunan, China

[‡]Zhejiang Automotive Engineering Advanced Research Institute, Hangzhou, Zhejiang, China

[†]Corresponding author. Email: 1144842572@qq.com

Abstract

Traffic sign recognition is a key perception task in autonomous driving and advanced driver-assistance systems, where accuracy and real-time performance directly affect road safety. In real-world scenes, fog, rain, and low illumination reduce image contrast and introduce noise, degrading deep models, while in-vehicle platforms impose strict limits on model size and computation. We propose CAMSNet, a lightweight traffic sign recognition network. Its architecture uses a lightweight residual backbone based on depthwise separable convolutions and a Dual-branch Attention Calibration Module (DACM), whose parallel channel- and position-attention branches model channel dependencies and spatial location. Their weights are fused elementwise and applied through residual reweighting to enhance sign features adaptively. During training, a weather-robust augmentation strategy based on atmospheric scattering randomly adds synthetic fog, Gaussian noise, and brightness changes. This augmentation is applied only to training data, leaving inference as a lightweight, one-way network. On the German Traffic Sign Recognition Benchmark (GTSRB), CAMSNet achieves 98.19% top-1 accuracy with 0.458 million parameters and 35.88 million multiply-accumulate operations (MACs), exceeding ResNet-18, which has about 24 times as many parameters, and MobileNetV2, which has about five times as many. CAMSNet with weather augmentation (CAMSNet+W) reaches 98.50% accuracy and 96.34% under severe synthetic fog. This is 36.98 percentage points above ResNet-18 and 36.12 points above CAMSNet without weather augmentation. CAMSNet+W also shows the smallest accuracy drops under severe noise and low brightness. Ablation and controlled comparisons indicate that attention improves robustness under degraded conditions while maintaining similar clean-domain accuracy. DACM is competitive with ECA and CBAM and offers a useful balance between parameter count and robustness. In this single run, weather augmentation improved robustness without an observed loss in clean-test accuracy. Robustness conclusions are based on synthetic corruptions; generalization to real adverse-weather scenes remains to be established.

Keywords: Traffic Sign Recognition; Autonomous Driving; Lightweight Network; Attention Mechanism; Weather Robustness; Depthwise Separable Convolution

1. Introduction

1.1 Background and motivation

Autonomous driving relies on environmental perception, planning, and control. Traffic signs encode speed limits, prohibitions, and warnings; errors in their recognition can affect vehicle behavior and safety. Deep learning has substantially improved recognition, with CNNs exceeding human-level performance on standard benchmarks [1]. A forward-facing camera supplies road images

to perception modules, whose detection and classification outputs inform speed control and route decisions [2]. Missing a stop sign or confusing speed limits can create serious risk. Vehicle compute is shared among perception tasks, so the recognition model must also meet tight latency and memory budgets.

Sign images vary with viewpoint, scale, illumination, weather, and occlusion. Signs are often small. Backlighting and shadows cause uneven brightness, while fog and rain lower contrast and blur details; water or dirt on a lens adds further degradation [3]. Such effects challenge generalization. This study uses the German GTSRB benchmark, while recognizing that sign appearance and class vocabularies vary across regions. In deployment, high-capacity models may exceed embedded-platform budgets, while adverse weather impairs the features on which those models rely. Restoration stages can add computation and propagate errors. We therefore study a lightweight classifier that addresses efficiency and synthetic-weather robustness within one framework.

1.2 Prior work and research gap

Early traffic sign recognition used hand-crafted features such as HOG, LBP, or SIFT with SVMs or random forests. These methods can work in controlled settings but are vulnerable to illumination change, occlusion, and weather. CNNs became dominant as AlexNet demonstrated deep visual representation learning [4]; ResNet introduced residual connections that facilitate optimization of deeper networks [5]. Their capacity can be costly on vehicles.

MobileNet decomposes standard convolution into depthwise and pointwise operations to reduce cost [6]. MobileNetV2 adds inverted residual blocks and linear bottlenecks [7]. Attention modules improve feature selection at modest cost: SE models channel dependencies [8]; CBAM applies channel and spatial attention sequentially [9]; ECA uses one-dimensional convolution for local cross-channel interaction [10]; and coordinate attention encodes horizontal and vertical position information [11].

Lightweight structures and attention have been applied to traffic-sign detection, including efficient vision Transformer/YOLOv5 with CBAM [12] and embedded YOLOv8 [13]. Detection adds localization to classification, while classification remains valuable as a perception component. E-MobileViT is a lightweight recognition model [14]. For adverse weather, HACTNet combines CNN and Transformer components [15]; other work compares real-weather training, synthetic augmentation, and denoising [16]; A-BDD provides synthetic weather and illumination augmentations [17]. Three broad strategies address degraded images: restoration before recognition, joint degradation-recognition learning, and training-time augmentation. Restoration and joint models can increase inference complexity; training-time augmentation can leave inference unchanged, although its effectiveness depends on whether simulated and real degradations match. This study evaluates synthetic fog, noise, and brightness changes in a lightweight classifier.

1.3 Contributions

We propose CAMSNet with four contributions: (1) a depthwise-separable residual backbone with three feature stages, expanding from 48 to 384 channels while reducing resolution from 24×24 to 6×6 ; (2) DACM, whose parallel channel and position branches fuse weights elementwise and apply residual reweighting; (3) online training augmentation using synthetic fog, Gaussian noise, and brightness changes without inference-time restoration; and (4) evaluation on GTSRB covering efficiency, accuracy, ablation, corruption robustness, and attention visualization. Section 2 reviews related work; Section 3 describes the method; Section 4 presents experiments; Section 5 discusses findings and limitations; and Section 6 concludes.

2. Related Work

2.1 Benchmarks and CNN classification

GTSRB, introduced by Stallkamp et al., has 43 classes and more than 50,000 real-world German sign images, including 39,209 training and 12,630 test images [1]. It includes variation in illumination, weather, and occlusion, with an imbalanced class distribution. The benchmark spans speed-limit, prohibition, warning, priority, and directional signs; some warning classes have relatively few examples. CNNs reached and exceeded human-level results in the 2011 IJCNN competition [1]. Belgian traffic sign datasets and TT100K are other resources, the latter mainly supporting detection. We use GTSRB for its common use in classification and direct comparability.

CNNs use local receptive fields and shared weights to learn hierarchical features. AlexNet combined convolutional and fully connected layers with ReLU and dropout [4]; ResNet's identity shortcuts support optimization and gradient flow [5]. Traffic sign recognition requires fine-grained discrimination among symbols, numerals, and colors across varying scales. Shallow high-resolution features can preserve details, while deeper features encode global shape and class semantics. Models that rely strongly on color may fail under weather-induced shifts, motivating robust shape and structure features.

2.2 Lightweight networks and attention

MobileNet decomposes convolution into depthwise and pointwise operations [6]; MobileNetV2 adds expansion, depthwise filtering, projection, and linear bottlenecks [7]. Depthwise convolution alone does not mix channels, so pointwise convolution and residual design help maintain representation capacity. Parameter count determines storage, MACs approximate compute, and actual latency also depends on memory access and operator implementation. Traffic-sign classification has a fixed class set and small inputs, permitting a compact task-specific representation. E-MobileViT combines MobileViT with efficient local and convolutional attention for traffic-sign recognition [14].

SE recalibrates channels from global pooling [8]; CBAM sequentially applies channel and spatial attention [9]; ECA models local channel interaction with a one-dimensional convolution [10]; coordinate attention encodes features along height and width [11]. SE lacks explicit position modeling; sequential processing can compress features repeatedly; ECA is compact but channel-focused. DACM uses independent parallel channel and position branches, multiplies their weights elementwise, and applies residual reweighting. Its performance is compared with ECA and CBAM under the same backbone and protocol.

2.3 Degradation and robust recognition

Atmospheric scattering combines attenuated scene radiance and airlight, reducing contrast in fog [3]. Rain can cause local occlusion; noise and low illumination reduce signal quality. Restoration-first approaches add computation and may require paired data; joint learning increases model complexity; training-time augmentation teaches recognition models to tolerate simulated degradation without changing inference. Fog, noise, and brightness shifts are common corruption settings [3, 18]. Recent studies include HACTNet for signs under fog, night, rain, and snow [15]; comparisons of real-weather training, synthetic augmentation, and denoising for detection [16]; and the A-BDD synthetic augmentation resource [17]. Such methods broaden training distributions, but synthetic and real weather differ. Augmentation must be sufficiently strong to help without shifting the training distribution too far from clean conditions.

3. Method

3.1 Problem and overall framework

We formulate recognition as single-image classification: for input x , the model predicts a distribution over $C = 43$ classes. The goal is to retain accuracy while reducing parameters and computation and

improving robustness to fog, noise, and low illumination. CAMSNet comprises preprocessing, a lightweight backbone, attention calibration, and a classification head. Images are resized to 48×48 , normalized, passed through depthwise-separable blocks, calibrated by DACM after each stage, and classified using global average pooling, dropout, and a fully connected layer. Weather augmentation affects training only. At inference, one forward pass produces 43 scores; the highest-scoring class is returned. Figure 1 shows the framework.

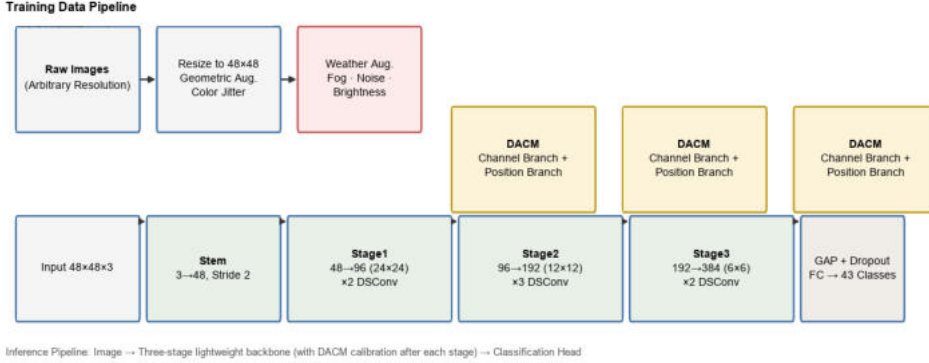


Figure 1. Overall architecture of CAMSNet.

3.2 Lightweight backbone

A depthwise-separable convolution applies a spatial convolution to each channel, then a 1×1 pointwise convolution mixes channels. A standard convolution has $C_{in}C_{out}k^2$ parameters; the separable form has $C_{in}k^2 + C_{in}C_{out}$. Their ratio is

$$\frac{k^2 + C_{out}}{C_{out}k^2} = \frac{1}{C_{out}} + \frac{1}{k^2},$$

approaching one ninth for a 3×3 kernel and large C_{out} . Each block uses depthwise convolution, batch normalization, and pointwise convolution. When channels match and stride is one, an identity shortcut is added before ReLU [5].

An initial stride-2 convolution maps 48×48 input to 24×24 with 48 channels. Stage 1 has two depthwise-separable blocks and outputs 96 channels at 24×24 , extracting edges and color regions. Stage 2 has three blocks; its first downsamples to 12×12 and expands to 192 channels. Stage 3 has two blocks; its first downsamples to 6×6 and expands to 384 channels. DACM follows each stage. This keeps shallow detail while increasing deeper semantic capacity. The model has fewer than 0.5M parameters, about one fifth of MobileNetV2 and one fortieth of ResNet-18.

3.3 Dual-branch Attention Calibration Module

The channel branch follows ECA [10]: global average pooling reduces feature map F to a channel descriptor, a one-dimensional convolution of kernel size 5 models local channel interactions, and sigmoid produces channel weights. The position branch is inspired by coordinate attention [11]. Features are averaged along horizontal and vertical axes; each directional descriptor passes through a 1×1 convolution with reduction ratio $r = 8$ and intermediate width $\max(8, \lfloor C/8 \rfloor)$. Encodings are expanded, concatenated, and fused by a 1×1 convolution into a C -channel spatial weight map followed by sigmoid.

The branch weights are multiplied to form $A \in [0, 1]$, and the output is $F' = F \times (1 + A)$. High-attention regions are amplified while original features remain available. Unlike sequential channel-spatial processing in CBAM [9], the parallel design avoids successive compression. Each module adds five channel-branch parameters. The position branch has approximately $4C \lfloor C/8 \rfloor$ parameters. At $C = 96, 192, 384$, intermediate widths are 12, 24, and 48, with position-branch counts 4,608, 18,432, and 73,728. Across three modules, DACM adds 96,783 parameters (about 21% of the 0.458M total). Modules follow every stage because shallow features encode color and outline, mid-level features encode symbols and numerals, and deep features carry class semantics.

3.4 Weather-robust augmentation

Synthetic fog follows the atmospheric-scattering model [3]:

$$I = tJ + (1 - t)A,$$

where J is scene radiance, t transmission, and A atmospheric light. We use a uniform approximation $t = \exp(-\beta)$, sampling fog density β during training and setting A to a bright sky-like value. This reduces contrast and shifts the image toward the airlight color. Additive Gaussian noise approximates sensor degradation; its standard deviation is sampled and values are clipped to range. Brightness augmentation applies a random scale to approximate overcast, dusk, or low-light scenes. The three perturbations are applied independently online during training, with clean and degraded examples both retained. The maximum fog coefficient is 1.2 (transmission about 0.3), maximum noise standard deviation is 0.06, and minimum brightness factor is 0.6. The total degraded-sample share remains below half to balance clean accuracy and robustness.

3.5 Training and implementation

We use stochastic gradient descent with momentum 0.9, weight decay 10^{-4} , initial learning rate 0.1, cosine annealing, batch size 128, and 25 epochs. Training augmentation includes random rotation, translation, scaling, and color jitter; horizontal flips are excluded because mirrored speed signs can change meaning. Batch normalization is used; dropout probability is 0.2. PyTorch default initialization is used without pretrained weights. The best validation checkpoint is evaluated on the test set; hyperparameters are fixed before testing. All experiments use seed 2024. Each configuration was trained once because of CPU and time constraints, so multi-seed estimates and significance tests are unavailable. Logs and intermediate outputs were saved for reproducibility.

4. Experiments and Results

4.1 Dataset, metrics, and setup

GTSRB has 43 classes, 39,209 training images, and 12,630 test images [1]. Images range from 15×15 to 250×250 and vary in illumination, weather, occlusion, and viewpoint. We use the official split and reserve 5% of training data for validation with a fixed seed. Images are resized to 48×48 by bilinear interpolation and normalized to $[0, 1]$. Test data receive no random transformations. Metrics are top-1 accuracy, trainable parameters, MACs, mean and standard deviation of single-image CPU latency, and accuracy drop under corruption.

Experiments used Windows, an AMD Ryzen 5 5600XT six-core/twelve-thread CPU, 32 GB RAM, and PyTorch. Configurations were Baseline CNN, ResNet-18 [5], MobileNetV2 [7], CAMS-Backbone without DACM, CAMSNet, and CAMSNet+W. Latency used a $3 \times 48 \times 48$ input, 20 warm-up runs, and 100 timed inferences with default framework threading and no operator-level optimization. Training augmentation used fog probability 0.25 with $\beta \sim U(0.4, 1.2)$, noise probability 0.15 with standard deviation $U(0.02, 0.06)$, and brightness probability 0.15 with scale $U(0.6, 1.0)$. Validation, corruption generation, and testing used fixed seeds. Best validation epochs ranged from 19 to 25.

4.2 Efficiency and clean accuracy

Table 1 reports efficiency. ResNet-18 has the largest parameter and MAC counts. CAMSNet uses 0.458M parameters; weather augmentation does not alter the model, so CAMSNet and CAMSNet+W have the same parameter count and MACs. Baseline CNN remains competitive in MACs and latency. Actual vehicle latency depends on hardware and operator optimization.

Table 1. Model parameters, computation, and CPU inference latency.

Model	Parameters (M)	MACs (M)	CPU latency (ms, mean $\pm$ SD)
Baseline CNN	1.284	24.41	0.76 ± 0.10
ResNet-18	11.199	97.98	6.72 ± 0.66
MobileNetV2	2.279	16.62	5.68 ± 0.39
CAMSNet	0.458	35.88	3.39 ± 0.26
CAMSNet+W	0.458	35.88	3.39 ± 0.27

CAMSNet achieves 98.19% top-1 test accuracy with 0.458M parameters and 35.88M MACs, exceeding ResNet-18 (97.21%; 11.199M) and MobileNetV2 (95.53%; 2.279M). Baseline CNN reaches 98.32% with 1.284M parameters and 24.41M MACs, 0.13 points above CAMSNet; this single-run difference cannot establish statistical significance. CAMSNet+W reaches 98.50%, the highest result. No clean-test accuracy decrease was observed, though stability needs multi-seed testing.

Table 2. Top-1 accuracy on the GTSRB test set.

Model	Accuracy (%)
Baseline CNN	98.32
ResNet-18	97.21
MobileNetV2	95.53
CAMSNet	98.19
CAMSNet+W	98.50

4.3 Ablation and attention comparison

Removing DACM gives 98.40% clean accuracy and 51.77% severe-fog accuracy; CAMSNet gives 98.19% and 60.22%, an 8.45-point fog gain. CAMSNet+W reaches 96.34% under severe fog, 36.12 points above CAMSNet, while slightly exceeding its clean accuracy. Under the same backbone and protocol, ECA obtains 98.57% clean and 69.49% severe-fog accuracy, above DACM in this single run; CBAM obtains 97.65% and 46.34%. DACM therefore does not outperform every attention module. Branch ablations show 57.95% severe-fog accuracy for channel-only, 60.72% for position-only, and 60.22% for full DACM. The position branch contributes more in this setting, and the full module did not exceed position-only. Repeated runs are required to assess stability.

Table 3. Attention and weather-augmentation ablation.

Configuration	Clean (%)	Severe fog (%)
CAMS-Backbone (no attention)	98.40	51.77
CAMSNet (with attention)	98.19	60.22
CAMSNet+W	98.50	96.34

Table 4. Attention modules under a common backbone and protocol.

Configuration	Parameters (M)	MACs (M)	Clean (%)	Severe fog (%)
CAMS-Backbone	0.361	31.51	98.40	51.77
CAMS+ECA	0.361	31.51	98.57	69.49
CAMS+CBAM	0.387	31.64	97.65	46.34
CAMS+DACM	0.458	35.88	98.19	60.22

Table 5. DACM branch-level ablation.

Configuration	Parameters (M)	MACs (M)	Clean (%)	Severe fog (%)
Backbone, no attention	0.361	31.51	98.40	51.77
Channel branch only	0.361	31.51	98.05	57.95
Position branch only	0.458	35.88	98.33	60.72
Full DACM	0.458	35.88	98.19	60.22

4.4 Robustness

Test corruptions are light/severe fog ($\beta = 0.8/1.6$), light/severe Gaussian noise (standard deviations 0.05/0.15), and low brightness (scale 0.6). CAMSNet+W reaches 96.34% under severe fog (2.16-point drop from clean), 85.22% under severe noise (13.28-point drop), and 98.00% under low brightness (0.50-point drop). Baseline CNN reaches 47.64% in severe fog and 59.52% under severe noise; ResNet-18 reaches 59.36% and 76.15%. CAMSNet+W is 36.98 points above ResNet-18 in severe fog. Fog reduces contrast and shifts brightness; noise disrupts local cues; low brightness compresses color separability. These results suggest transfer among the tested synthetic corruptions, but do not establish performance under real rain, uneven fog, night glare, or lens droplets.

Table 6. Top-1 accuracy (%) under clean and perturbed conditions.

Model	Clean	Light fog	Severe fog	Light noise	Severe noise	Low brightness
Baseline CNN	98.32	78.31	47.64	90.13	59.52	97.93
ResNet-18	97.21	83.17	59.36	94.89	76.15	96.39
MobileNetV2	95.53	72.55	47.74	91.12	67.99	94.87
CAMSNet	98.19	87.39	60.22	94.34	68.80	97.74
CAMSNet+W	98.50	98.30	96.34	97.29	85.22	98.00

4.5 Visualization and error analysis

Grad-CAM compares CAMSNet with the attention-free backbone. It weights final convolutional maps using gradients of the target score and upsamples them to image size. The backbone activations

generally cover the sign but are more diffuse; DACM activations are more concentrated within sign boundaries. For slippery-road signs, the CAMSNet activation covers internal graphics more evenly. This is a qualitative illustration, not causal evidence.

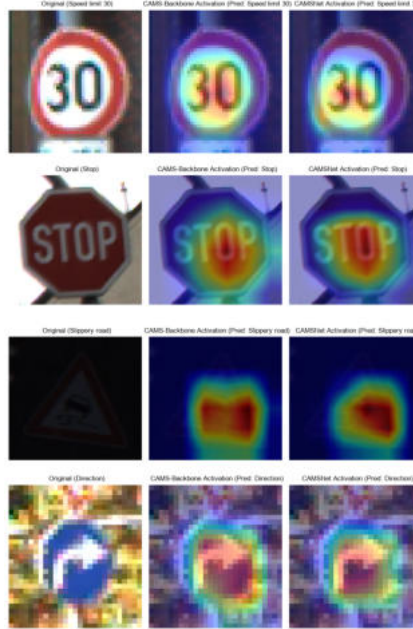


Figure 2. Grad-CAM activation comparison between CAMSNet and CAMS-Backbone.

CAMSNet+W misclassifies 189 of 12,630 clean test images (1.50%). The most frequent classes are Bumpy Road (30), Roadworks (25), Other Danger (20), and Right Lane Ends (17), together 48.7% of errors. Roadworks (class 22) and Left Curve (25) are confused 24 times; Bumpy Road (27), Slippery Road (29), and Uneven Road (28) are mutually confused 21 times; speed-limit 60 and 80 signs (classes 3 and 5) are confused seven times. Right Lane Ends (class 30; 240 training examples) is misclassified 17 times, including nine predictions as No Overtaking by Trucks (class 20). Small, tilted, or occluded signs also lose details. Such errors reflect visual similarity and class imbalance; temporal evidence or joint detection and classification may help [1].

5. Discussion

5.1 Interpretation

The results show competitive clean accuracy and improved performance on the tested synthetic corruptions. DACM improves severe-fog accuracy over the attention-free backbone, although ECA performs better on clean accuracy and fog in the single controlled run. Weather augmentation exposes the network to degraded examples without adding inference-time restoration. The backbone coordinates depthwise and pointwise operations with residual links, channel expansion, resolution, and attention placement. Attention can reweight features under tight model budgets; augmentation is useful when simulated corruptions match task vulnerabilities in contrast, color, and pixel detail. Gains may diminish if real degradation differs substantially from simulations. A dehazing stage could be tested as a supplement, with added inference cost. Actual deployment latency depends on vehicle hardware and operator optimization.

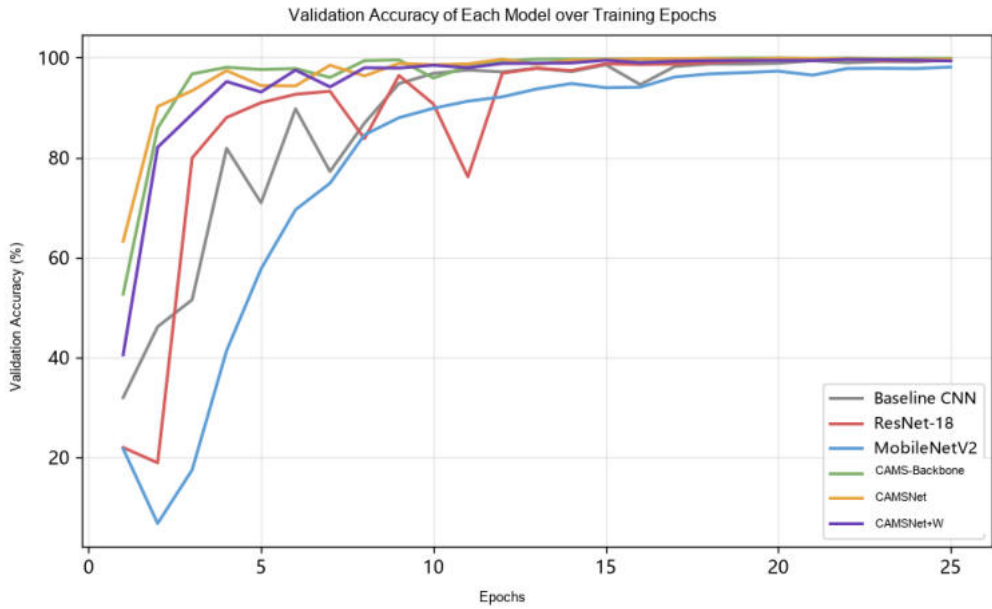


Figure 3. Validation accuracy by epoch for evaluated models.

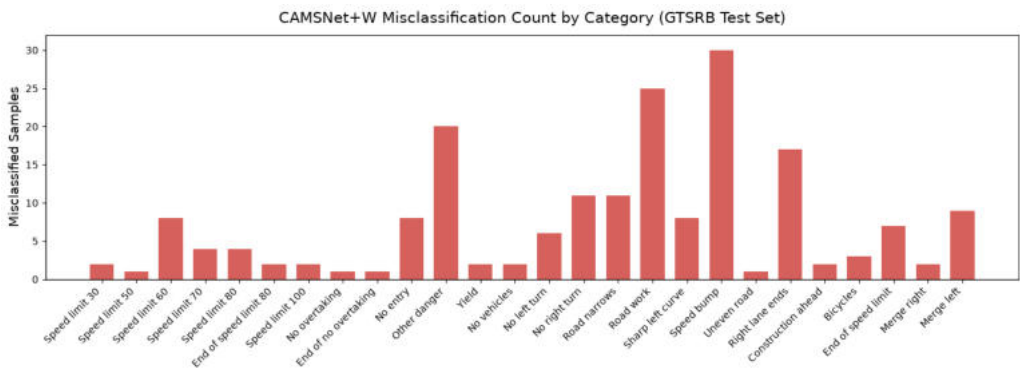


Figure 4. Misclassified test images by class.

5.2 Limitations and future work

The study uses only GTSRB, so cross-country transfer remains untested. Synthetic corruptions do not fully reproduce uneven fog or raindrops on lenses. Each configuration was trained once, without multi-seed uncertainty or significance tests. Comparisons include ECA and CBAM but not SE, coordinate attention, or a broader module set. The task is classification and does not localize signs. Future work should evaluate additional national datasets and real-weather images, repeat runs with statistical analysis, expand attention comparisons, and extend the method to detection or joint detection-classification. Quantization and knowledge distillation, including 8-bit quantization, may further reduce deployment requirements.

6. Conclusion

We proposed CAMSNet, combining a depthwise-separable residual backbone, parallel dual-branch attention, and training-time augmentation based on atmospheric scattering. On GTSRB, CAMSNet achieves 98.19% with 0.458M parameters and 35.88M MACs, compared with 97.21% for ResNet-18 and 95.53% for MobileNetV2. DACM raises severe-fog accuracy from 51.77% to 60.22%. CAMSNet+W reaches 98.50% clean accuracy and 96.34% in severe synthetic fog, with gains under tested noise and low brightness; no clean accuracy loss was observed in this single run. The combination offers a potential low-cost path to robust recognition because augmentation adds no inference computation. Further work should test repeated seeds, more datasets, real weather, and joint detection-classification.

Author Contributions

Ting Wang (First author): Investigation, Writing–original draft, Writing–review & editing. Zhihao Weng (Corresponding author): Conceptualization, Funding acquisition. Both authors read and approved the final manuscript.

Compliance with Ethical Standards

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This work was financially supported solely by Zhihao Weng (the corresponding author). No external funding agency or grant was involved.

Competing Interest Declaration

The authors declare that they have no competing interests.

Acknowledgments

The authors would like to thank Changsha Medical University and the Zhejiang Automotive Engineering Advanced Research Institute for providing the research environment and facilities.

References

- [1] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition,” *Neural Networks*, vol. 32, pp. 323–332, 2012.
- [2] S. Grigorescu, B. Trasnea, T. Cocias, et al., “A survey of deep learning techniques for autonomous driving,” *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, 2020.

- [3] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*. Lake Tahoe: Curran Associates, 2012, pp. 1097–1105.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016, pp. 770–778.
- [6] A. G. Howard, M. Zhu, B. Chen, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.
- [7] M. Sandler, A. Howard, M. Zhu, et al., "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018, pp. 4510–4520.
- [8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018, pp. 7132–7141.
- [9] S. Woo, J. Park, J. Y. Lee, et al., "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision*. Munich: Springer, 2018, pp. 3–19.
- [10] Q. Wang, B. Wu, P. Zhu, et al., "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020, pp. 11534–11542.
- [11] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021, pp. 13713–13722.
- [12] G. Zeng, Z. Wu, L. Xu, et al., "Efficient vision transformer YOLOv5 for accurate and fast traffic sign detection," *Electronics*, vol. 13, no. 5, article 880, 2024.
- [13] Y. Luo, Y. Ci, S. Jiang, et al., "A novel lightweight real-time traffic sign detection method based on an embedded device and YOLOv8," *Journal of Real-Time Image Processing*, 2024.
- [14] S. Song, X. Ye, and S. Manoharan, "E-MobileViT: A lightweight model for traffic sign recognition," *Industrial Artificial Intelligence*, vol. 3, no. 1, article 3, 2025.
- [15] M. S. Devgan, G. Singh, P. Sharma, et al., "Attention-guided lightweight CNN-transformer fusion for real-time traffic sign recognition in adverse environments: HACTNet," *IET Intelligent Transport Systems*, 2026, doi: 10.1049/itr2.70167.
- [16] H. Gupta, O. Kotlyar, H. Andreasson, et al., "Robust object detection in challenging weather conditions," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE, 2024, pp. 7508–7517.
- [17] F. Assion, F. Gressner, N. Augustine, et al., "A-BDD: Leveraging data augmentations for safe autonomous driving in adverse weather and lighting," arXiv:2408.06071, 2024.
- [18] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in *Proceedings of the 7th International Conference on Learning Representations*, New Orleans, 2019.