Scalable and Efficient Multi-Weather Classification for Autonomous Driving with Coresets, Pruning, and Resolution Scaling

Tushar Shinde, Avinash Kumar Sharma

Indian Institute of Technology Madras Zanzibar

Introduction

- Autonomous vehicles require robust perception systems capable of operating in diverse weather conditions.
- Performance degradation under extreme weather conditions (rain, snow, fog) poses significant challenges, impairing both visibility and sensor readings.
- Deploying deep learning models on edge devices in autonomous vehicles is challenging due to computational and memory constraints.

Motivation

- Traditional deep learning models struggle with balancing computational efficiency and accuracy across varied weather conditions.
- A scalable approach is needed to: 1) Reduce training complexity, 2) Optimize model size, and 3) Maintain high classification accuracy.

Objectives

The present study investigates the following objectives:

- Develop a scalable framework for multi-weather classification in autonomous driving using coreset selection, resolution scaling, and model compression.
- Evaluate the impact of different coreset selection methods and dataset fractions on classification performance.
- Assess model robustness under low-resolution settings.
- Demonstrate how adaptive pruning combined with 8-bit quantization can reduce model size while maintaining competitive accuracy.

Methodology Overview

Our framework introduces three complementary techniques to enhance multi-weather classification performance and efficiency:

1. Coreset Selection

- Reduces training complexity while preserving classification accuracy
- Identifies most informative samples from original training set $D = \{(x_i, y_i)\}_{i=1}^{N}$
- Selects a representative subset $C \subseteq D$ to minimize computational overhead
- We evaluate random and margin-based selection at different data fractions (1.0, 0.75, 0.5, 0.25, 0.1, 0.05)

2. Resolution Scaling

- Images resized to three resolutions: 224×224, 112×112, and 56×56
- Investigates trade-off between computational efficiency and accuracy
- Allows flexible deployment based on hardware constraints
- We analyze the impact of resolution reduction on feature preservation and model performance

3. Layer-wise Adaptive Pruning and Quantization

Layer Importance Metric (LIM):

- Incorporates Layer Parameter Ratio and Layer Sparsity Ratio
- Parameter ratio of layer i:

$$P_i = \frac{\text{Number of parameters in layer } i}{\text{Total parameters in model}}$$

- Sparsity ratio of layer i:

$$S_i = \frac{\text{Number of near-zero parameters in layer } i}{\text{Total parameters in layer } i}$$

- LIM computation: $\text{LIM}_i = \beta \cdot P_i + (1 - \beta) \cdot S_i$

Model Compression Techniques:

- **Adaptive Pruning:** Removes less important weights based on LIM
- **8-bit Quantization:** Reduces weight precision
- These techniques enable up to 85% model size reduction while maintaining competitive accuracy

Experimental Setup

Dataset: WEDGE dataset (3,360 synthetic images, 16 distinct weather conditions)

Platform: All experiments conducted on Kaggle using NVIDIA Tesla P100/G4 GPUs with PyTorch

Models: ResNet18 and EfficientNetB0 architectures

Training: Learning rate 0.001, batch size 64, 25 epochs with Adam optimizer

Resolution Variations: 224×224, 112×112, 56×56

Pruning Levels: $k_i = \{2, 1.5, 1, 0.5\}$ (95%, 86%, 68%, 38% pruning rates)

Metrics: Classification accuracy and compression ratio

Results & Analysis

Baseline Performance: ResNet18: 75.60%, EfficientNetB0: 77.83% (224×224)

Impact of Resolution Scaling:

- 224×224 to 112×112: Slight performance drop
- 56×56: Significant degradation but still usable

Effect of Coreset Selection:

- Margin-based selection outperforms Random selection

Table 1. Performance Comparison of Multi-Weather Classification Across Models, Resolution Levels, and Coreset Selection Methods.

	ResNet18						EfficientNetB0					
Fraction	224x224		112x112		56x56		224x224		112x112		56x56	
	Random	Margin	Random	Margin	Random	Margin	Random	Margin	Random	Margin	Random	Margin
1.00	75.60	-	73.21	-	64.88	-	77.83	-	73.21	-	68.01	-
0.75	73.51	72.62	68.01	69.49	62.50	62.95	76.93	75.74	72.92	69.05	63.10	63.69
0.50	70.83	70.98	68.45	64.58	61.76	60.57	73.36	71.88	71.43	69.64	60.27	63.54
0.25	65.48	65.33	61.46	61.46	54.46	55.21	68.90	68.60	64.73	63.54	54.02	55.51
0.10	53.72	56.10	48.81	46.58	42.26	46.58	59.97	59.08	53.57	54.76	43.75	41.82
0.05	38.54	42.71	33.48	36.01	32.44	25.89	45.68	47.47	40.18	44.20	29.17	33.04

Model Compression Results:

- LAP + 8-bit Q: 76.64% accuracy with 5.5× compression ratio
- Standard pruning (P level 0.5): 69.64% accuracy with 1.7× compression

Table 2. Comparison of Accuracy and Compression Ratio (CR) across resolutions, coreset methods, and P levels on ResNet18 model.

Resolution	224x224				112x112				56x56			
Coreset	Random		Margin		Random		Margin		Random		Margin	
Method	Acc.	CR	Acc.	CR	Acc.	CR	Acc.	CR	Acc.	CR	Acc.	CR
P level 2	6.25%	21.3	6.25%	21.3	6.25%	21.2	6.25%	21.2	6.25%	20.8	6.25%	20.7
P level 1.5	6.55%	8.3	6.40%	8.3	7.74%	8.3	6.70%	8.3	6.70%	8.3	8.33%	8.3
P level 1	9.52%	3.5	15.92%	3.5	24.55%	3.5	27.53%	3.5	29.91%	3.6	21.13%	3.6
P level 0.5	69.64%	1.7	63.10%	1.7	69.49%	1.7	68.45%	1.7	64.43%	1.7	61.90%	1.7
Baseline	76.49%	1.0	74.55%	1.0	72.62%	1.0	71.88%	1.0	66.82%	1.0	64.29%	1.0
LAP	76.51%	1.2	74.40%	1.3	73.07%	1.5	70.09%	1.9	65.48%	1.4	63.84%	2.4
LAP + 8-bit Q	76.64%	5.5	74.11%	6.4	74.26%	8.3	72.02%	8.2	66.07%	6.7	65.48%	7.3

Conclusion

- Successfully integrated coreset selection, resolution scaling, and model compression
- Achieved up to 85% model compression while maintaining competitive accuracy
- Our framework offers favorable trade-offs between accuracy and efficiency
- Results demonstrate that our approach is well-suited for deployment in autonomous vehicles under adverse weather conditions

Future Work

- Adapting the model to dynamic environments
- Exploring mixed precision quantization
- Implementing dynamic pruning for improved real-time performance
- Extending our approach to other autonomous driving perception tasks

References

[1] 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*, pp. 770–778, 2016.

[2] C. Tang, K. Ouyang, Z. Wang, Y. Zhu, W. Ji, Y. Wang, and W. Zhu, "Mixed-precision neural network quantization via learned layer-wise importance," in *European Conference on Computer Vision*, pp. 259–275, Springer, 2022.

[3] T. Shinde, "Adaptive quantization and pruning of deep neural networks via layer importance estimation," in *Workshop on Machine Learning and Compression, NeurIPS 2024*, 2024.

Acknowledgment

This work was in part supported by the Walmart Center of Technical Excellence (IIT Madras) Project Grant Award.