Real-Time Detection of Hazardous Road Conditions: An Efficient High-Resolution Framework with EfficientNetV2

Minsoo Kim, Taesoo Jun
Pervasive Intelligent Computing Laboratory, Department of Software Engineering
Kumoh National Institute of Technology, Gumi, South Korea
{kms991022, taesoo.jun}@kumoh.ac.kr

***Abstract*—Accurate classification of road surface conditions is critical for autonomous driving safety. However, differentiating between hazardous conditions like black ice and wet roads is challenging due to high visual similarity and the loss of fine-grained textures in low-resolution inputs. To address this, we propose a robust framework utilizing EfficientNetV2-S, architecturally optimized for a native high-resolution input of 384×384. By leveraging this design, we capture subtle surface details without the inefficiency of upscaling older architectures. We further integrate CBAM and Class-Balanced Loss to refine feature selection. Experimental results on the RSCD dataset demonstrate that our method achieves 93.01% accuracy within just 15 training epochs. Notably, our approach improves the recall of the hardest class (*dry_concrete_slight*) by 15.53% compared to ResNet-50, while maintaining the highest Ice Recall of 99.49% among all tested models.**

***Index Terms*—Road Surface Classification, EfficientNetV2, Attention Mechanism, High-Resolution Processing, Deep Learning.**

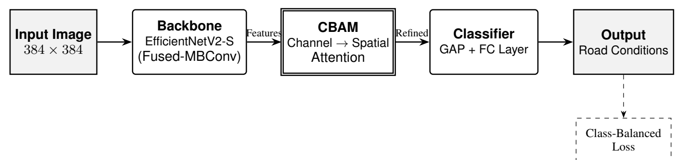

Fig. 1. Overall Architecture. The proposed framework processes high-resolution inputs through EfficientNetV2-S. The extracted features are refined by the CBAM module (Channel and Spatial Attention) before final classification into road surface conditions. Training is guided by Class-Balanced Loss.

I. Introduction

Real-time recognition of road surface conditions is essential for active safety systems [1]. Hazardous conditions like "Black Ice" are often invisible to the human eye but pose severe risks. Misjudging such slippery surfaces as dry asphalt can lead to catastrophic accidents.

Reliable classification is hindered by the loss of **fine-grained texture details** (e.g., slight cracks or surface glint) in standard CNNs designed for low-resolution (224×224) inputs. Conventional approaches often fail to capture these minute features, leading to high false-negative rates. Furthermore, simply upscaling inputs for older models like ResNet-50 results in excessive computational costs suitable for real-time applications.

To address these challenges, we propose a framework optimized for both **detail preservation** and **efficiency**. The main contributions are:

- We demonstrate that **high-resolution input** (384×384) is indispensable for distinguishing fine-grained road textures.
- We adopt **EfficientNetV2-S** [2] as the backbone. Unlike ResNet, V2-S is **natively designed for** 384×384, allowing it to process high-resolution data with optimal efficiency by minimizing memory access overhead [2].
- We integrate **CBAM** [3] and **Class-Balanced Loss** [5] to explicitly focus on informative texture regions and effectively handle the long-tail distribution inherent in road datasets.

II. Related Work

Existing research often faces a trade-off between speed and accuracy. Standard models like **ResNet-50** provide strong feature extraction but are optimized for 224×224 inputs. Scaling them to 384×384 significantly increases FLOPs without guaranteeing proportional accuracy gains due to inflexible receptive fields [4].

Conversely, lightweight models like **MobileNetV3** prioritize speed but often lack the capacity to encode complex high-frequency texture features. **EfficientNetV2-S** bridges this gap by employing **Fused-MBConv** layers specifically tuned to handle larger resolutions (384×384) efficiently, making it ideal for texture-centric tasks where both detail and speed are paramount.

III. Proposed Method

Our framework consists of three integrated stages: high-resolution feature extraction, attention-based feature refinement, and loss re-weighting. Figure 1 illustrates the overall architecture.

A. *EfficientNetV2 Backbone (Native 384)*

We employ EfficientNetV2-S as our feature extractor. A key advantage of V2-S is that its architecture is officially optimized for an input resolution of 384×384. This contrasts with older models designed for 224×224, where increasing resolution is an inefficient adaptation. V2-S utilizes **Fused-MBConv** blocks to efficiently handle this higher pixel count without memory bottlenecks.

TABLE I
PERFORMANCE COMPARISON (384×384 INPUT)

Model	Params	FLOPs	Ice Recall	Hard Recall*	Acc
ResNet-50	25.6 M	12.11 G	98.47%	63.62%	87.88%
MobileNetV3	**5.4 M**	**0.6 G**	97.91%	74.38%	88.41%
EffNetV2 (Vanilla)	21.5 M	8.37 G	98.30%	67.23%	90.54%
Ours (Proposed)	21.7 M	8.38 G	**99.49%**	**79.15%**	**93.01%**

* *Hard Recall* refers to *dry_concrete_slight*, the most challenging class.

B. Attention Mechanism (CBAM)

To further refine the extracted features, we insert the Convolutional Block Attention Module (CBAM). CBAM applies Channel (M_c) and Spatial (M_s) Attention sequentially to the feature map F:

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F' \tag{1}$$

where $\otimes$ denotes element-wise multiplication. This dual attention strategy allows the network to effectively focus on relevant texture patterns (”what”) and their locations (”where”) while suppressing background noise.

C. Class-Balanced Loss

We adopt Class-Balanced Loss to address data imbalance. It re-weights the Cross-Entropy (CE) loss based on the effective number of samples (n_y):

$$Loss(p, y) = \frac{1 - \beta}{1 - \beta^{n_y}} \mathcal{L}_{CE}(p, y) \tag{2}$$

We set $\beta = 0.999$. This ensures that rare but critical classes like Black Ice contribute more significantly to gradient updates, preventing bias towards the majority class.

IV. EXPERIMENTS

A. Setup

We used the **RSCD dataset** [6]. Models were implemented in PyTorch and trained for **15 epochs** using **AdamW** to ensure fast convergence. We evaluated all baselines at the same 384×384 resolution to ensure a fair comparison of architectural capabilities.

B. Results and Analysis

Table I presents the comparison results.

1) Superiority on Hard Samples: Our model demonstrates exceptional capability on fine-grained textures. As shown in Table I, ResNet-50 achieved only **63.62%** recall on the hardest class (*dry_concrete_slight*). This performance drop implies that standard down-sampling in older architectures discards high-frequency texture information. In contrast, our method, leveraging V2-S and CBAM, achieved **79.15%**, marking a substantial **+15.53% improvement**.

2) Safety-Critical Reliability: For autonomous driving, avoiding false negatives on hazardous classes is paramount. Our method achieved a **99.49%** Ice Recall, the highest among all tested models. This highlights the effectiveness of using an architecture natively optimized for high-resolution texture analysis, even within a limited training duration of **15 epochs**.

V. CONCLUSION

We proposed a framework that leverages the **native high-resolution capability** (384×384) of EfficientNetV2 combined with CBAM. Achieving 93.01% accuracy and a 15.53% improvement on hard classes within just **15 epochs**, our results confirm that aligning model architecture with input resolution is crucial for robust road surface analysis. Future work will investigate further performance gains with extended training iterations and validate real-time inference robustness on actual embedded automotive platforms.

ACKNOWLEDGMENT

This work was supported by the Innovative Human Resource Development for Local Intellectualization (IITP-2025-RS-2020-II201612) and the ITRC program (IITP-2025-RS-2024-00438430) through the Institute of Information & Communications Technology Planning & Evaluation (IITP) funded by the Ministry of Science and ICT (MSIT); the Basic Science Research Program (2018R1A6A1A03024003) through the National Research Foundation of Korea (NRF); and the regional innovation system & education (RISE)-(Specialized University) program through the Gyeongbuk RISE Center, funded by the Ministry of Education (MOE) and the Gyeongsangbuk-do, Republic of Korea (2025-rise-15-105).

REFERENCES

[1] D. K. Dewangan and S. P. Sahu, “RCNet: Road classification convolutional neural networks for intelligent vehicle system,” *Intell. Service Robotics*, vol. 14, pp. 199–214, 2021.

[2] M. Tan and Q. V. Le, “EfficientNetV2: Smaller models and faster training,” in *Proc. ICML*, 2021.

[3] S. Woo, J. Park, J. Lee, and I. S. Kweon, “CBAM: Convolutional block attention module,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[4] L.-Y. Gao, Z. Qu, S.-Y. Wang, and S.-F. Xia, “A lightweight neural network model of feature pyramid and attention mechanism for traffic object detection,” *IEEE Trans. Intell. Veh.*, vol. 9, no. 2, pp. 3422–3435, 2024.

[5] Y. Cui et al., “Class-balanced loss based on effective number of samples,” in *Proc. CVPR*, 2019.

[6] T. Zhao, J. He, J. Lv, D. Min, and Y. Wei, “A comprehensive implementation of road surface classification for vehicle driving assistance: Dataset, models, and deployment,” *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8361–8374, 2023.