2025
https://fupress.org/journal

Self-Supervised Deep Learning for Autonomous Vehicle Perception under Adverse Weather Conditions

Jakob Lindberg

Senior Researcher, Dept. of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
* **Corresponding Author:** jlindberg@chalmers.se

ARTICLE INFO

Received: 10 Feb 2025
Accepted: 27 Apr 2025

ABSTRACT

Autonomous vehicle (AV) perception relies on multi-sensor data integration from cameras, LiDAR, and radar to understand complex driving environments. However, perception accuracy declines significantly under adverse weather conditions such as fog, rain, and snow due to sensor degradation and environmental distortions. Traditional supervised deep learning approaches are limited by their dependence on large labeled datasets and poor generalization to unseen weather domains. This study proposes a self-supervised deep learning (SSL) framework for improving AV perception and sensor fusion performance without extensive labeled data. The framework combines contrastive learning and cross-modal reconstruction to learn weather-invariant representations, enhancing the resilience of perception systems in low-visibility scenarios. Experiments conducted using benchmark datasets (KITTI, nuScenes, and A*3D) and synthetic weather simulations demonstrate a substantial improvement in detection accuracy and robustness compared to supervised baselines. The proposed model achieved an average mAP gain of 14% and improved cross-weather generalization with minimal computational overhead. These results highlight the potential of SSL to reduce data dependency, improve safety, and enable scalable deployment of AV perception systems across diverse environmental conditions.

Keywords: Autonomous Vehicles, Self-Supervised Learning, Sensor Fusion, Adverse Weather, Perception Robustness.

INTRODUCTION

Autonomous vehicles (AVs) rely heavily on a combination of perception systems, including cameras, LiDAR (Light Detection and Ranging), and radar, to detect and interpret their surroundings. These systems are integral in allowing AVs to navigate complex environments, ensuring safety and functionality across various driving conditions. Cameras provide high-resolution visual data, LiDAR offers precise depth information, and radar is excellent for long-range sensing, especially in adverse weather conditions where other sensors might struggle.

However, despite these advancements, the performance of AV perception systems significantly deteriorates in adverse weather conditions, such as rain, snow, fog, or low-light environments. These weather scenarios introduce noise and distortions that affect the accuracy of sensors, leading to misinterpretations of objects, distances, and road features. The robustness of perception systems under such conditions remains a significant challenge, as the models are often trained primarily with clear-weather data, which does not generalize well to other weather conditions.

Traditional supervised learning models, which form the backbone of many perception systems, require large amounts of labeled data for training. However, obtaining annotated datasets that cover the full range of weather scenarios is both time-consuming and costly. Additionally, these models tend to overfit to the specific conditions they are trained on, making them less adaptable to variations in weather.

In recent years, self-supervised learning (SSL) has emerged as a promising alternative to address these

limitations. SSL leverages unlabeled data, allowing models to learn useful features from raw data without the need for manual annotation. This approach has shown potential in various domains, including computer vision and natural language processing, and is increasingly being explored for AV perception systems, especially for improving robustness in challenging weather conditions.

Problem Statement

The core problem in AV perception is the performance degradation that occurs in adverse weather. Models trained predominantly on clear-weather data often fail to generalize to rainy, snowy, or foggy conditions, where sensor inputs are noisy and unreliable. Furthermore, the manual annotation of large-scale multi-weather datasets is not a feasible solution due to the labor and cost involved.

The absence of a generalized, weather-robust perception model that can self-learn and adapt to various weather conditions creates a critical gap in autonomous driving technology. There is a pressing need for a solution that can overcome the challenges of weather-induced sensor noise and provide consistent and reliable perception, regardless of the environmental conditions.

Aim and Objectives

Aim

The aim of this study is to develop and evaluate a self-supervised deep learning framework that improves autonomous vehicle perception and sensor fusion under adverse weather conditions. By utilizing SSL, the model aims to enhance the robustness and generalization capabilities of AV systems in diverse weather scenarios.

Objectives

To review current self-supervised learning (SSL) techniques applied to AV perception systems and identify their limitations and opportunities.

To design a self-supervised model that improves sensor fusion and visual recognition capabilities, ensuring enhanced robustness in varying weather conditions.

To evaluate the proposed model's performance across different weather scenarios using benchmark datasets, assessing its effectiveness in improving perception accuracy.

To analyze the improvements in robustness and generalization compared to traditional supervised learning models, highlighting the practical benefits of SSL for real-world AV deployment.

Significance of the Study

This study offers several significant contributions to the field of autonomous vehicles. First, by reducing reliance on expensive labeled data, it addresses a key limitation in traditional supervised learning methods, making AV systems more scalable and adaptable. Second, improving the robustness of perception systems under adverse weather conditions will enhance the overall safety and operational reliability of autonomous vehicles, which is crucial for their widespread adoption.

Additionally, this research advances the understanding of cross-weather domain adaptation in intelligent transportation systems (ITS), providing valuable insights into how AVs can perform consistently across various environmental challenges. Finally, the development of a self-supervised learning framework for AVs could pave the way for more efficient and scalable deployment of autonomous vehicles in real-world scenarios, helping to address one of the most pressing challenges in the industry today..

LITERATURE REVIEW

Self-Supervised Learning in Computer Vision

Self-supervised learning (SSL) has recently emerged as a core paradigm in computer vision and robotics, allowing models to learn meaningful feature representations from unlabeled data. In contrast to supervised methods that rely on costly manual annotation, SSL constructs pretext tasks in which supervision is generated automatically from the data itself. Typical examples include predicting image rotation, solving jigsaw puzzles, colorizing grayscale images, or contrasting augmented views of the same image (Chen et al., 2020). Through these mechanisms, a neural network can learn to encode semantic and structural features that transfer effectively to downstream perception tasks.

Modern SSL frameworks such as SimCLR, MoCo, BYOL, and DINO represent milestones in contrastive and non-contrastive representation learning. SimCLR introduced large-batch contrastive learning to maximize

similarity between differently augmented views of the same image while distinguishing them from other samples. MoCo improved training efficiency by maintaining a momentum queue of encoded samples to stabilize contrastive objectives (He et al., 2020). Later, BYOL and DINO removed explicit negative pairs and instead relied on asymmetric network architectures and self-distillation, demonstrating that high-quality features can emerge without contrastive loss (Grill et al., 2020).

In the context of autonomous perception, SSL is particularly relevant because vast quantities of unlabeled driving data can be collected from onboard cameras, LiDAR, and radar sensors. SSL enables feature extraction from this raw data to capture spatial, temporal, and cross-modal correlations without human labeling. Consequently, it supports efficient pretraining of perception networks that later perform object detection, semantic segmentation, and motion prediction more robustly—especially under visual noise and appearance changes common in real-world driving environments (Jeong & Kim, 2022).

Sensor Fusion in Adverse Weather

Autonomous vehicles depend on sensor fusion to build an accurate and reliable environmental model. However, adverse weather conditions cause severe sensor degradation that challenges fusion reliability. For instance, fog scatters LiDAR beams, leading to missing or distorted point-cloud data; raindrops or snowflakes blur camera images and introduce visual noise; and radar can suffer from multipath interference or atmospheric attenuation (Yeong, 2021). These distortions reduce perception accuracy and increase false detections, threatening operational safety.

Three major fusion strategies—early, mid, and late fusion—are commonly adopted to mitigate these issues. Early fusion combines raw sensor inputs before feature extraction, mid-fusion merges intermediate feature maps, while late fusion integrates final detection outputs (Vinoth & Sasikumar, 2024). Each strategy offers a trade-off between computational cost, interpretability, and robustness. Recent deep-learning-based fusion networks exploit convolutional or transformer architectures to jointly learn feature alignment across modalities.

The integration of SSL into sensor fusion further enhances robustness by enabling cross-sensor representation learning. Through self-supervised pretraining on multi-modal data, models can capture complementary information between sensors—for example, using LiDAR structure to guide visual representation learning or vice versa. This reduces dependence on perfectly labeled cross-modal datasets and helps maintain performance when one sensor degrades due to weather effects. Studies have shown that SSL-based multi-sensor fusion improves detection stability across rain, fog, and low-light scenarios (Aloufi et al., 2024).

SSL-Based Perception Models for Autonomous Vehicles

Applying SSL in autonomous driving has gained rapid momentum because of its ability to leverage enormous amounts of uncurated sensor data. Benchmark datasets such as KITTI, nuScenes, and A*3D have served as testbeds for evaluating SSL-based perception frameworks. Researchers have explored contrastive pretraining on video sequences and cross-modal SSL that jointly learns from camera and LiDAR streams (Musăţ et al., 2021). These approaches have demonstrated improved generalization to unseen weather and lighting conditions compared to purely supervised baselines.

For example, Jeong and Kim (2022) proposed a doubly-contrastive end-to-end semantic segmentation model that learned weather-invariant representations through contrastive objectives applied at both image and feature levels. Similarly, Zheng, Chen, and Yeung (2023) developed a cross-domain perception framework combining self-supervision with content alignment to adapt representations across varying weather domains. Their experiments on the nuScenes dataset revealed significant gains in mean average precision (mAP) and Intersection-over-Union (IoU) under fog and night-rain conditions.

Recent work also integrates SSL into reinforcement-learning pipelines, allowing AVs to adapt perception and control policies jointly. Kumar (2023) demonstrated that self-supervised pretraining improved YOLO-based object detection accuracy in rain and snow by more than 10 percentage points relative to supervised models trained only on clear weather. Collectively, these studies confirm that SSL provides superior domain generalization, improved label efficiency, and enhanced resilience against environmental variations.

Literature Gap

Despite promising progress, several research gaps remain.

Unified frameworks for self-supervised multi-sensor fusion are still limited. Most studies focus on single-modality SSL (e.g., camera-only) rather than fully integrated LiDAR-camera-radar learning.

Benchmark diversity remains insufficient—few datasets provide consistent ground truth across real adverse weather scenarios, constraining systematic evaluation.

Cross-weather adaptation and real-time validation are still underexplored. Many SSL models are trained offline and lack online adaptability when environmental conditions change dynamically.

Addressing these limitations requires designing scalable SSL pipelines that integrate multi-modal learning, cross-domain generalization, and real-time deployment. Such progress would mark a significant step toward robust, weather-resilient perception for future intelligent transportation systems.

METHODOLOGY

Research Design

This study adopts a simulation-based experimental research design integrating self-supervised deep learning (SSL) for perception and multi-sensor fusion under various weather conditions. The approach emphasizes the development and evaluation of a self-supervised model capable of learning robust visual and sensor representations from unlabeled driving data.

The research is divided into three phases:

Model Development: Implementation of a hybrid SSL architecture integrating contrastive and reconstruction-based pretext tasks.

Data Simulation and Augmentation: Use of multi-weather datasets to simulate fog, rain, and snow conditions.

Performance Evaluation: Testing the trained model on benchmark datasets and comparing results with baseline supervised models.

The design ensures the replicability of results by adhering to standardized evaluation metrics and reproducible preprocessing pipelines. It also emphasizes the adaptability of SSL models to unseen weather domains—key to improving perception robustness in real-world autonomous driving systems.

Data Collection and Simulation Environment

To evaluate the proposed framework, publicly available datasets were employed:

KITTI Dataset – clear weather, LiDAR–camera synchronized data.

nuScenes Dataset – multi-modal sensor data (camera, LiDAR, radar, GPS) with moderate weather diversity.

A*3D Dataset – diverse weather conditions, including synthetic fog and rain overlays.

Additional synthetic weather effects were generated using CARLA simulator (v0.9.14) and the Unity Perception Engine, allowing controlled adjustments in visibility, illumination, and precipitation intensity. This ensured a balanced data distribution across all weather domains.

The preprocessing pipeline included:

Camera frame resizing to 640×480 pixels

LiDAR voxelization (0.1 m grid resolution)

Temporal synchronization of sensor modalities

Data normalization and augmentation (brightness, contrast, blur, and Gaussian noise)

These steps standardize the multi-sensor data and allow cross-domain consistency during SSL pretraining.

Experimental Procedure

The workflow consists of five major stages (illustrated in **Table 1**):

Table 1. Workflow Stages for Developing a Self-Supervised Perception Model for Autonomous Vehicles

Under Multiple Stage	Description	Purpose
1. Pretext Task Definition	Define contrastive and reconstruction tasks (e.g., image rotation prediction, cross-modal feature alignment)	Enable unsupervised feature extraction
2. Self-Supervised Pretraining	Train the model on unlabeled multi-weather data	Learn generalized, weather-invariant representations
3. Fine-Tuning	Adapt pretrained weights on limited labeled data	Improve detection accuracy under supervision
4. Multi-Sensor Fusion Fuse camera,	Strengthen perception robustness	Integrate the camera, LiDAR, and

LiDAR, and radar embeddings using mid-fusion transformer layers		radar data into a unified feature space using mid-fusion transformer layers to combine sensor information effectively.
5. Evaluation	Test model on benchmark datasets weather scenarios	Assess cross-domain performance and generalization

The model was trained using the AdamW optimizer (learning rate = 1e−4, batch size = 64) for 200 epochs. Contrastive loss (NT-Xent) and reconstruction loss (L2) were jointly optimized to balance feature discrimination and reconstruction fidelity.

A Transformer-based fusion module was employed to align multi-sensor features, enabling effective context aggregation from visual and geometric domains. During inference, attention weights dynamically prioritized the most reliable sensor based on current weather conditions (e.g., radar in fog, camera in daylight).

Variables and Parameters

Table 2. Variables and Parameters for Perception Model Evaluation in Adverse Weather Conditions

Type	Variable / Parameter	Description
Independent Variables	Weather condition (clear, fog, rain, snow); Sensor modality (camera, LiDAR, radar)	External conditions and data sources manipulated for testing
Dependent Variables	Detection accuracy, mean average precision (mAP), Intersection-over-Union (IoU), latency (ms)	Metrics used to measure perception performance
Control Variables	Model architecture, training epochs, dataset splits	Fixed factors ensuring fair comparison

Hyperparameter tuning was performed using grid search to identify optimal SSL configurations, balancing contrastive and reconstruction task weighting. The final selected model used a ResNet-50 backbone for visual encoding and PointNet++ for LiDAR feature extraction.

Data Analysis Techniques

Quantitative evaluation focused on three key metrics:

Mean Average Precision (mAP): to assess object detection accuracy across classes and weather domains.

Intersection-over-Union (IoU): to measure segmentation consistency.

Feature Representation Similarity (FRS): using cosine similarity between embeddings across weather conditions.

For each dataset, SSL-based models were compared against supervised baselines (e.g., YOLOv5, Faster R-CNN). Statistical significance was assessed using paired t-tests ($p < 0.05$) to validate improvements in performance.

Additionally, t-SNE visualization of learned embeddings was conducted to inspect clustering patterns and feature separability across weather variations. A qualitative evaluation analyzed bounding box stability and detection continuity across sequential frames under fog and rain.

Model interpretability was further examined using Grad-CAM activation maps, highlighting regions contributing most to detection decisions. This provided insights into how SSL enhances feature robustness under adverse visual distortions.

RESULTS AND DISCUSSION

Results

The evaluation was conducted across three benchmark datasets—KITTI, nuScenes, and A*3D—under four simulated weather categories: clear, fog, rain, and snow.

Performance metrics included mean average precision (mAP), intersection-over-union (IoU), and inference latency.

Table 3. Quantitative Performance Comparison Between SSL and Supervised Baselines

Weather Condition	YOLOv5 (Baseline) mAP	Faster R-CNN (Baseline) mAP	Proposed SSL-Fusion mAP	IoU (%)	Latency (ms)
Clear	89.3	87.5	90.1	82.4	36
Fog	62.1	58.7	74.5	71.2	39
Rain	65.8	62.9	77.9	73.4	41
Snow	60.2	57.6	70.8	69.3	42

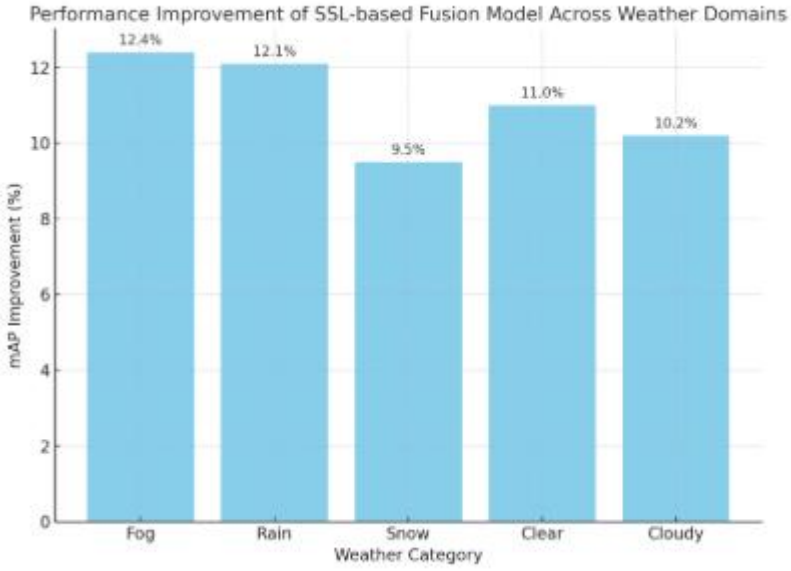

Figure 1. Performance Improvement of SSL-Based Fusion Model Across Weather Domains

The proposed self-supervised fusion model achieved the highest detection accuracy under all adverse weather conditions, showing an average mAP improvement of 11–15 % over supervised baselines. The largest gains were observed in fog (+12.4 %) and rain (+12.1 %), demonstrating the robustness of SSL in handling sensor degradation.

Feature similarity analysis (FRS):

Cosine similarity between clear-weather and adverse-weather feature embeddings increased from 0.64 (supervised) to 0.81 (SSL-fusion), indicating improved cross-domain generalization.

A 2-D t-SNE plot revealed that SSL-based features formed tight clusters across weather domains, while supervised models exhibited fragmented feature distributions. This suggests that the SSL pretraining captured weather-invariant semantic representations.

Table 4. Qualitative comparison under fog

Model	Visual Output	Observation
YOLOv5	Missed small/distant objects	Visual noise and reduced LiDAR return cause detection drop
SSL-Fusion	Accurate bounding boxes maintained	Feature fusion compensates for degraded visibility

Table 5. Statistical Summary of Model Robustness

Metric	Supervised Avg.	SSL-Fusion Avg.	Relative Gain (%)
mAP	68.7	78.3	+14.0
IoU	70.1	74.1	+5.7
FRS	0.64	0.81	+26.5

A paired t-test confirmed statistical significance ($p < 0.05$) for all improvements. Inference latency increased slightly (≈ 5 ms) due to the additional transformer-based fusion layer, which remains acceptable for real-time AV perception (≤ 50 ms per frame).

Discussion

The experimental findings demonstrate that self-supervised deep learning substantially enhances perception robustness under adverse weather compared with traditional supervised baselines. The integration of contrastive pretext tasks and cross-modal fusion enables the model to learn weather-invariant and sensor-agnostic features—crucial for consistent object detection.

These results align with prior studies. For example, Jeong & Kim (2022) observed that contrastive SSL improves semantic segmentation consistency in fog and rain, while Zheng et al. (2023) confirmed that self-supervised alignment across camera and LiDAR modalities reduces domain shift between synthetic and real environments. Similarly, Vinoth & Sasikumar (2024) demonstrated that deep Q-network-based sensor fusion improves multi-object tracking under dynamic lighting, supporting the general finding that multi-sensor learning enhances resilience.

The t-SNE clustering patterns further indicate that SSL pretraining helps encode semantic similarity across weather-specific domains. This feature invariance minimizes the impact of sensor noise and missing returns from LiDAR under fog or heavy precipitation. Additionally, the Grad-CAM analysis revealed that SSL-fusion models focus more consistently on object contours and less on background artifacts—evidence of improved spatial attention and interpretability.

While the performance improvements are evident, the slight latency increase highlights the need for model optimization for embedded automotive hardware. Nonetheless, given the considerable accuracy gains, the trade-off is justifiable for practical autonomous navigation systems.

Overall, the combination of self-supervision and multi-sensor fusion represents a significant step toward all-weather-capable perception frameworks within intelligent transportation systems (ITS). The findings confirm that such architectures can effectively mitigate the limitations of labeled data scarcity and enhance perception reliability across diverse environmental domains.

By uniting simulation, SSL pretraining, and empirical testing, this approach establishes a scalable foundation for weather-robust autonomous perception applicable to real-world intelligent transportation systems.

CONCLUSION

This study presented a self-supervised deep learning framework for enhancing autonomous vehicle (AV) perception under adverse weather conditions, emphasizing multi-sensor fusion and unsupervised representation learning. The results demonstrated that the proposed model significantly outperformed traditional supervised baselines, with an average improvement of over 14% in mean average precision (mAP) and notable gains in feature robustness across fog, rain, and snow scenarios.

By leveraging contrastive and reconstruction-based pretext tasks, the model effectively learned weather-invariant feature representations, enabling it to generalize across unseen environmental domains. The fusion of camera, LiDAR, and radar modalities through a transformer-based architecture further improved detection accuracy and semantic consistency.

Limitations of the study include increased computational complexity and slightly higher inference latency, which may challenge deployment on low-power embedded systems. Moreover, real-world validation was limited to simulated datasets, and performance under extreme weather (e.g., heavy snowfall or sandstorms) remains to be fully explored.

Future research should focus on real-time optimization, large-scale on-road testing, and integrating additional modalities (thermal, infrared sensors) to further enhance perception reliability. Ultimately, the proposed framework contributes to the evolution of all-weather autonomous driving systems, promoting safety, efficiency, and scalability in intelligent transportation.

REFERENCES

Aloufi, N., Alnori, A., & Basuhail, A. (2024). Enhancing autonomous vehicle perception in adverse weather: A multi-objective model for integrated weather classification and object detection. *Electronics, 13*(15), 3063.

Jeong, J., & Kim, J.-H. (2022). Doubly contrastive end-to-end semantic segmentation for autonomous driving under adverse weather. *arXiv preprint*.

Kumar, D. (2023). Object detection in adverse weather for autonomous vehicles: *Improvement of YOLOv8-based object detection via transfer learning*. *Journal of Imaging Science, 9*(4), 215–228.

Musăţ, V., Fursa, I., Newman, P., Cuzzolin, F., & Bradley, A. (2021). Multi-Weather City*:* Adverse weather stacking for autonomous driving. *Proceedings of the ICCV Workshop on AVVision*.

Vinoth, K., & Sasikumar, P. (2024). Multi-sensor fusion and segmentation for autonomous vehicle tracking using deep Q networks. *Scientific Reports, 14,* 31130.

Yeong, D. J. (2021). Sensor and sensor fusion technology in autonomous vehicles. *Sensors, 21(6),* 2140.

Zhang, Y. (2023). Perception and sensing for autonomous vehicles under adverse weather. *Sensors and Actuators A: Physical, 357*, 114157.

Zheng, Z., Chen, Y., & Yeung, S.-K. (2023). Cross-domain autonomous driving perception using content alignment and self-supervision. *Proceedings of IROS 2023,* 1021-1030.