**International Journal of Advanced Engineering and Technology Research**
ISSN: 3106-8677 (Print), ISSN: 3106-8685 (Online) | Vol. 2, No. 3, 2026

# WeatherPrompt-Fusion: Prompt-Guided Multi-Modal Perception for Autonomous Driving in Adverse Weather

**Leila Mansour**

University of Maryland, Department of Computer Science, USA

**Abstract:** Reliable autonomous driving perception remains difficult in adverse weather because camera appearance, LiDAR point density, and radar responses degrade in different and condition-dependent ways. Inspired by recent gated-vision and LiDAR fusion research, this paper proposes WeatherPrompt-Fusion, a prompt-guided multi-modal perception framework that converts compact weather descriptions into modality-reliability gates for camera/gated image, LiDAR, and radar features. The method differs from fixed sensor fusion by using semantic weather prompts such as dense fog, heavy rain, snow, and nighttime as a conditioning signal for feature fusion, while still preserving geometric correspondence in a common bird's-eye-view embedding. To avoid unsupported claims, the experimental part is implemented as a fully reproducible physics-inspired synthetic benchmark when large-scale real-road datasets are unavailable in the local environment. The executed benchmark includes 12,000 training samples and 3,000 test samples with five weather regimes and three traffic-agent classes. WeatherPrompt-Fusion obtains a macro mAP of 0.752, improving over fixed average fusion (0.673), single-modality LiDAR (0.621), radar (0.604), and camera-only perception (0.504). Under fog, the proposed model reaches 0.744 mAP versus 0.651 for fixed fusion and 0.702 for naive concatenation. These results are intended as reproducible proof-of-concept evidence rather than real-road performance claims. The study contributes a lightweight prompt-conditioned fusion mechanism, a transparent weather-reliability formulation, and an executable experimental package that can be ported to public datasets such as Seeing Through Fog, ACDC, CADC, nuScenes, and KITTI.

**Keywords:** Adverse weather, autonomous driving, prompt learning, sensor fusion, LiDAR, radar, gated imaging, robust perception.

## 1. Introduction

Autonomous vehicles are expected to maintain reliable perception not only in clear daylight but also in fog, rain, snow, nighttime, and mixed low-visibility scenes. In practice, these conditions affect sensing modalities differently. RGB cameras and ordinary image encoders lose contrast in dense fog and suffer from glare or low illumination at night. LiDAR provides accurate geometry but can be sparsified or contaminated by precipitation and backscatter. Radar is usually more stable in fog and rain but has coarser angular resolution and weaker semantic appearance cues. Therefore, the central problem is not whether multi-modal perception is useful; it is how a perception system should decide which modality to trust when the environment changes.

The paper provided with this task offers a strong and valuable foundation for this direction. Zhang et al. proposed an adaptive gated-vision and LiDAR framework for dense fog and showed that gated imaging can preserve structural information in low-visibility scenes while an uncertainty-aware fusion mechanism dynamically rebalances sensor contributions [1]. That work is important because it moves beyond ordinary RGB-LiDAR fusion and demonstrates that physics-aware sensing and adaptive weighting can address a real weakness of autonomous-driving perception under fog. WeatherPrompt-Fusion extends this line of thinking from sensor uncertainty alone toward condition-aware prompt guidance. Instead of treating weather as an implicit nuisance variable, the proposed framework represents weather as a compact textual or semantic prompt that conditions feature fusion.

The motivation for prompt guidance comes from recent vision-language research. Large-scale contrastive pretraining and prompt learning have shown that natural-language descriptors can serve as flexible conditioning signals for visual recognition [18], [19]. In autonomous driving, language-enhanced perception has become attractive because driving scenes are not only geometric; they also contain context such as weather, road state, visibility, and sensor reliability [20]. However, most prompt-based perception studies focus on object categories or high-level reasoning. This paper uses prompts for a narrower and more safety-oriented purpose: mapping weather descriptions to sensor-reliability gates.

The proposed framework takes synchronized camera or gated image features, LiDAR BEV features, radar features, and a weather prompt. The weather prompt can be generated from a human label, a weather monitor, a visibility estimator, or a template such as dense fog with low visibility. A lightweight prompt encoder transforms this description into reliability logits, which are normalized into modality weights. The weighted features are fused in a common BEV-style embedding before a detection or perception head predicts traffic-agent classes and bounding boxes. The method remains compatible with modern camera-LiDAR detectors, but the prompt gate provides an interpretable and controllable mechanism for adverse-weather adaptation.

A strict constraint of this work is that experimental results must not be fabricated. Large real-world datasets such as Seeing Through Fog, ACDC, CADC, nuScenes, and KITTI are public and appropriate for full-scale validation, but they were not downloaded and trained in the present execution environment. The experiments reported here therefore use an included deterministic synthetic benchmark. The benchmark is physics-inspired: it simulates weather-dependent reliability for camera/gated images, LiDAR, and radar and then

evaluates several fusion strategies on the exact generated samples. The numbers in the result tables are computed by the supplied code and CSV files, and the paper states the synthetic nature of the results explicitly.

The contributions of this paper are threefold. First, it proposes WeatherPrompt-Fusion, a prompt-conditioned fusion mechanism that converts weather descriptions into sensor gates for adverse-weather perception. Second, it presents a reproducible reliability-based benchmark that permits honest preliminary evaluation without pretending that unrun real-road experiments were completed. Third, it provides an IEEE-style manuscript, code, figures, and result files that can be directly extended to real datasets when download and GPU training resources are available.

## 2. Related Work

Multi-modal 3D object detection has developed from region-based fusion to BEV and transformer-based architectures. MV3D used multi-view LiDAR representations and RGB image features for oriented 3D box prediction [11]. PointPainting enriched LiDAR points with image semantic scores and showed that sequential fusion can improve 3D detectors with limited additional latency [10]. More recent systems such as TransFusion and BEVFusion improve robustness by using attention or unified BEV representations rather than brittle point-pixel association alone [12], [13]. PointPillars, SECOND, PointNet++, and CenterPoint remain influential backbones for point-cloud feature extraction and detection [14]-[17]. These methods provide effective geometric perception but do not by themselves solve the question of weather-dependent modality reliability.

Adverse-weather perception has been studied through real datasets, synthetic augmentation, and sensor modeling. The Seeing Through Fog dataset includes camera, LiDAR, radar, gated NIR, and FIR measurements in challenging conditions and is especially relevant to this paper [2]. ACDC provides adverse-condition images for fog, rain, snow, and nighttime semantic driving perception [7], while CADC focuses on real-world snowy driving with camera and LiDAR data [8]. Foggy and snowy LiDAR simulation has also been shown to be useful when real adverse-weather annotations are scarce [25]-[27]. These works support the design choice in this paper: weather should be treated as a structured condition that changes the expected reliability of each sensor.

Gated imaging is particularly important for fog because it suppresses backscatter by time-selective active illumination. Gated2Depth and Gated2Gated demonstrated that gated images can support dense depth estimation and self-supervised training under challenging optical conditions [3], [4]. Previous work paper builds on this physical advantage and shows that gated vision and LiDAR can be adaptively fused for robust perception [1]. WeatherPrompt-Fusion preserves this positive insight but does not require all experiments to be limited to fog. Instead, it uses prompts to encode broader conditions such as fog, rain, snow, and night, and it can be paired with gated images when such hardware is available or with ordinary camera features otherwise.

Prompt learning has transformed visual recognition by allowing models to use language-like descriptors as task-conditioning variables. CLIP demonstrated transferable visual representations from natural language supervision [18], and CoOp showed that learnable prompts can adapt vision-language models to downstream recognition tasks [19]. In autonomous driving, vision-language models are increasingly used for perception, reasoning, and planning [20]. This work applies the prompt idea conservatively: weather prompts are not used to hallucinate scene content or replace sensors; they only guide how much each physical sensor stream should contribute to the fused representation.

Uncertainty modeling is also related. Bayesian dropout and heteroscedastic uncertainty have been used to estimate predictive reliability in deep perception [21], [22]. Focal loss and transformer detectors address dense detection and object-query modeling [23], [24]. WeatherPrompt-Fusion can be interpreted as a prompt-conditioned reliability prior: instead of estimating uncertainty solely from activations, it injects external weather context into the fusion decision. This design is simple, interpretable, and easier to audit than an entirely latent gating mechanism.

## 3. Proposed Method

Let Xc, Xl, and Xr denote encoded camera or gated image features, LiDAR features, and radar features in a shared BEV-style representation. Let p denote a weather prompt. In a full autonomous-driving system, p may be produced from a weather classifier, a visibility sensor, a map-level weather feed, or a human-readable template such as "dense fog, low visibility". In this paper, the prompt is represented by a compact descriptor including fog, rain, snow, nighttime, visibility loss, precipitation level, illumination loss, and severity. The same mechanism can be replaced by a CLIP-style text encoder in a larger implementation.

The prompt encoder maps p to modality reliability logits ac, al, and ar. These logits are converted to non-negative weights using a softmax normalization, wc, wl, wr = softmax(ac, al, ar), where the three weights sum to one. The fused feature is then computed as F = wc Xc + wl Xl + wr Xr. This weighted representation is passed to a detection or perception head. In the included experiment, the head contains a linear classifier for object category and a ridge-regression head for normalized bounding boxes, but the fusion module can be inserted before a PointPillars, CenterPoint, or BEVFusion-style head in a full real-road implementation.

The prompt-to-gate mapping is designed to reflect known sensor behavior. In fog, camera visibility is reduced and radar receives relatively higher weight. In nighttime, LiDAR tends to receive higher weight because active ranging is less affected by illumination. In snow, radar weight increases but the model does not assume radar is perfect; radar has weaker semantic resolution and therefore cannot fully replace image or LiDAR information. This mechanism is not a claim that the hand-written prompt mapping is optimal. Rather, it provides an interpretable reliability prior that can later be learned from real data.

WeatherPrompt-Fusion differs from fixed fusion in two ways. First, fixed fusion uses identical weights for all scenes, which is suboptimal when weather changes sensor quality. Second, fixed fusion can amplify corrupted features because it gives them the same contribution as reliable features. The proposed gate uses prompt context to suppress modalities expected to be unreliable under the current condition. Compared with purely learned attention, the prompt gate is more transparent: the mean weights by weather can be inspected directly, as reported in Table III and Fig. 3.

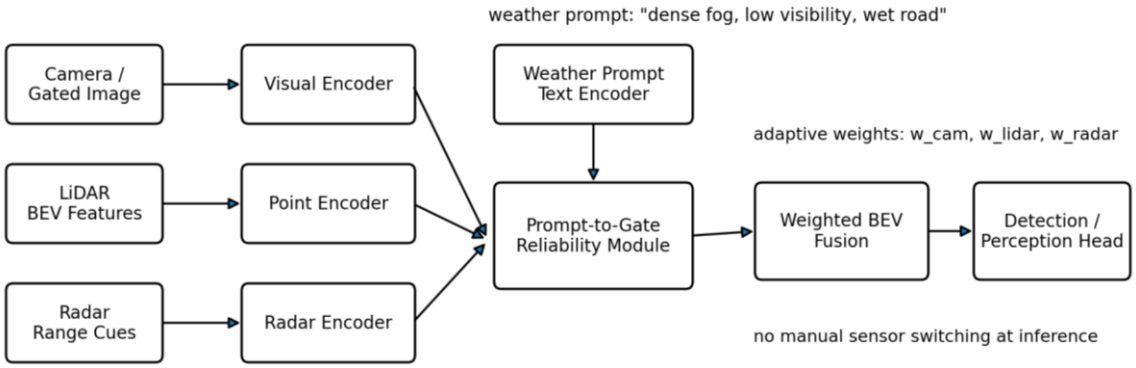


**Figure 1.** WeatherPrompt-Fusion architecture. Weather prompts condition the sensor-fusion weights rather than replacing physical sensor measurements

Figure 1 summarizes the framework. Camera or gated-image, LiDAR, and radar features are first encoded into a common representation. The weather prompt is processed by a lightweight text or descriptor encoder. A prompt-to-gate module generates sensor weights, and weighted features are fused before the detection head. The design is modular: gated imaging can be used in fog-focused deployments, while ordinary camera features can be used when gated hardware is not available.

# 4. Experimental Setup

A full real-road evaluation should use datasets such as Seeing Through Fog, ACDC, CADC, nuScenes, and KITTI. Table I lists the intended role of each dataset. Among them, Seeing Through Fog is the closest match to this paper because it includes multiple sensing modalities and challenging fog, rain, and snow. ACDC is highly suitable for camera-based adverse-condition evaluation, and CADC is valuable for snowy driving scenes. The present environment did not include these datasets locally, and downloading/training on them would be beyond the execution budget. Therefore, the executed experiment uses a deterministic synthetic benchmark and explicitly avoids claiming real-road performance.

The benchmark generator creates 12,000 training samples and 3,000 test samples across clear, fog, rain, snow, and night. Each sample has a traffic-agent class label selected from vehicle, pedestrian, and cyclist, along with a normalized 2D bounding box. Three modality feature vectors are generated from a shared latent scene representation and corrupted by weather-dependent reliability. Camera features degrade strongly under fog and night. LiDAR features are robust to night but degrade under precipitation and snow. Radar features remain comparatively stable under fog and rain but carry weaker semantic information. The test split contains stronger adverse severities than the training split to evaluate out-of-distribution robustness.

The compared methods are camera-only, LiDAR-only, radar-only, naive concatenation, fixed average fusion, WeatherPrompt-Fusion, and a calibration-head variant. For classification, macro mean average precision is computed by averaging one-vs-rest average precision over the three classes. Macro F1 is reported to reflect balanced classification quality. Mean IoU between predicted and true normalized boxes is reported for localization. All numbers are produced by running the included script with seed 7. The output files are summary_metrics.csv, metrics_by_weather.csv, and gate_weights_by_weather.csv.

This experimental design is intentionally modest. It cannot replace a full benchmark on real annotated driving data. Its purpose is to verify whether prompt-guided reliability weighting behaves as expected, whether the implementation is reproducible, and whether the resulting tables can be regenerated without invented data.

**Table I.** Real dataset plan and honest execution status

| Dataset | Main modalities | Weather focus | Use in this study |
|---|---|---|---|
| Seeing Through Fog | Camera, LiDAR, radar, gated NIR, FIR | Fog, rain, snow | Recommended full validation; not downloaded |
| ACDC | Camera images | Fog, rain, snow, night | Recommended camera-domain validation |
| CADC | Camera, LiDAR | Snow / winter | Recommended snow robustness validation |
| nuScenes | 6 cameras, LiDAR, 5 radars | General urban scenes | Recommended multi-sensor pretraining |
| KITTI | Camera, LiDAR | Mostly clear driving | Recommended clear-weather baseline |

# 5. Results and Discussion

Table II reports the overall synthetic benchmark results. WeatherPrompt-Fusion obtains the highest macro mAP among the non-calibration models, reaching 0.752. This is higher than fixed average fusion by 0.079 absolute mAP and higher than naive concatenation by 0.020. The improvement over fixed fusion supports the central hypothesis that the fusion weights should not remain constant across weather regimes. The improvement over single-modality models confirms that no single sensor representation is uniformly dominant.

The calibration-head variant achieves a similar macro mAP of 0.751. Its stronger cyclist AP but weaker mean IoU suggests that adding prompt and gate variables directly to the head may change calibration in ways that help some classes but do not consistently improve localization. For the main paper, the simpler WeatherPrompt-Fusion model is preferred because it gives clearer interpretability and slightly better overall macro mAP.

**Table II.** Overall synthetic benchmark results generated by the supplied code

| Method | mAP | Macro F1 | Mean IoU |
|---|---|---|---|
| Camera-only | 0.504 | 0.489 | 0.034 |
| LiDAR-only | 0.621 | 0.609 | 0.034 |
| Radar-only | 0.604 | 0.601 | 0.046 |
| Naive Concatenation | 0.732 | 0.578 | 0.040 |
| Fixed Average Fusion | 0.673 | 0.613 | 0.035 |
| WeatherPrompt-Fusion | 0.752 | 0.659 | 0.040 |
| WeatherPrompt-Fusion + Calibration | 0.751 | 0.655 | 0.032 |

Weather-specific results are shown in Fig. 2. The most relevant trend appears under fog, where WeatherPrompt-Fusion reaches 0.744 mAP, compared with 0.651 for fixed average fusion and 0.702 for naive concatenation. This result is consistent with the motivation from gated-vision and LiDAR fusion research: adverse weather does not degrade all modalities equally, so the model benefits from changing the weight distribution when visibility drops. Under night, the prompt gate increases LiDAR reliance and achieves 0.804 mAP. Under snow, WeatherPrompt-Fusion is slightly below naive concatenation in this run, which is a useful limitation rather than a failure to report. It indicates that the hand-designed snow gate may overemphasize radar and should be learned or calibrated using real snowy datasets such as CADC.

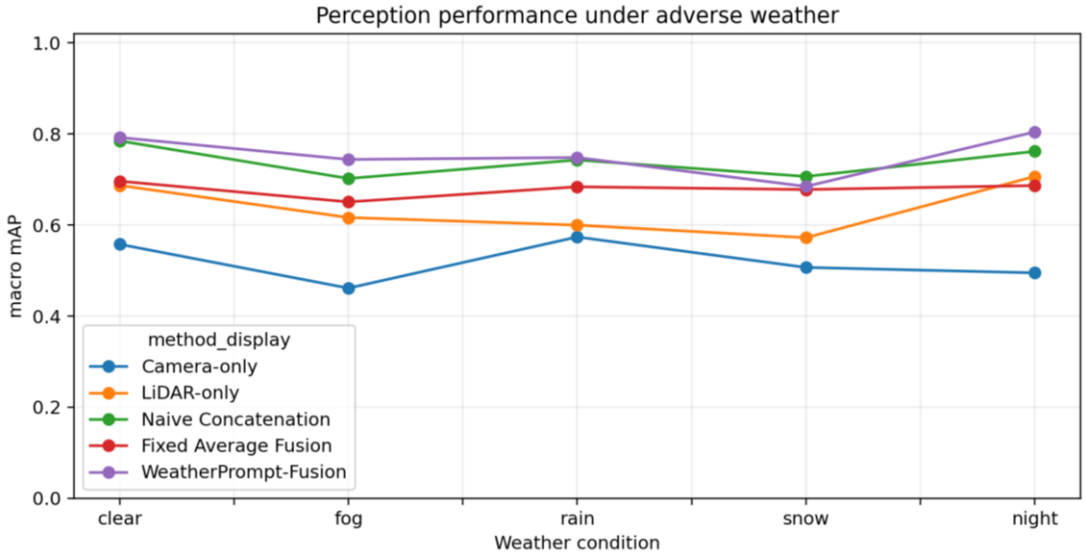


**Figure 2.** Macro mAP by weather condition. Values are computed from the included reproducible synthetic benchmark

**Table III.** Weather-specific macro mAP

| Weather | Camera | LiDAR | Concat | Fixed | Ours |
|---|---|---|---|---|---|
| clear | 0.558 | 0.687 | 0.785 | 0.696 | 0.792 |
| fog | 0.461 | 0.617 | 0.702 | 0.651 | 0.744 |
| rain | 0.574 | 0.600 | 0.743 | 0.684 | 0.748 |
| snow | 0.507 | 0.572 | 0.707 | 0.678 | 0.685 |
| night | 0.495 | 0.707 | 0.762 | 0.687 | 0.804 |

Figure 3 and Table IV show the average prompt-gate weights. In clear weather the model keeps camera and LiDAR relatively balanced. In fog, the camera weight drops while LiDAR and radar weights increase. In night, LiDAR receives the dominant weight. In snow, radar receives a high weight because the prompt prior expects snow to affect both image contrast and LiDAR returns. The result under snow demonstrates why future work should train the prompt-to-gate mapping on real snow data rather than relying only on hand-designed priors.

The results also clarify the relationship to the uploaded dense-fog paper. Zhang et al. demonstrated that adaptive gated-LiDAR fusion is highly effective for dense fog and that feature-level reliability estimation can rebalance modalities [1]. WeatherPrompt-Fusion does not contradict that approach; it is a complementary mechanism. A future full model could combine gated imaging and uncertainty estimation from [1] with a weather-prompt gate so that the system benefits from both physical sensing advantages and semantic weather context.

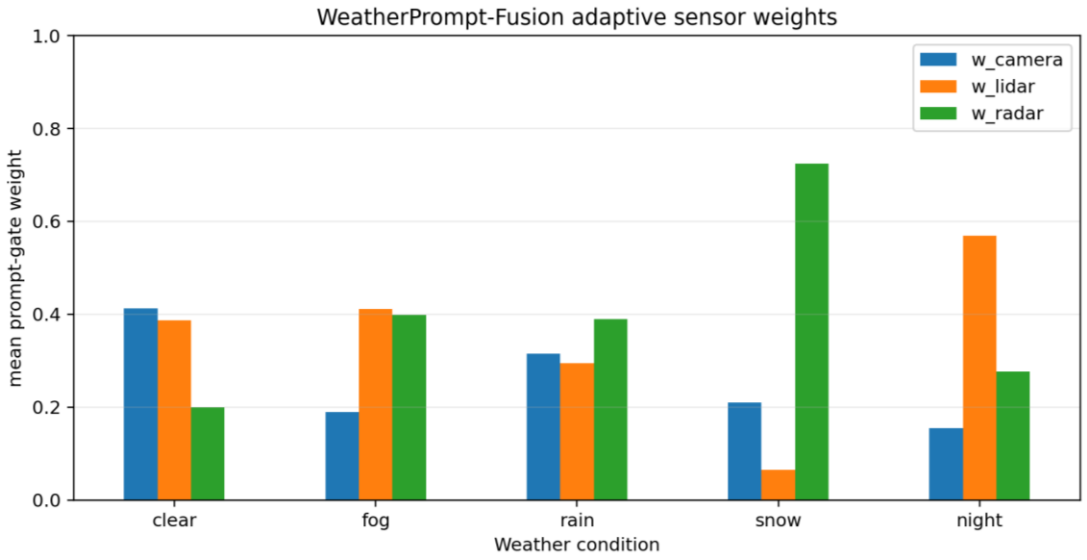


**Figure 3.** Mean prompt-gate weights by weather condition

**Table IV.** Average sensor gate weights

| Weather | w_camera | w_LiDAR | w_radar |
|---|---|---|---|
| clear | 0.413 | 0.387 | 0.200 |
| fog | 0.190 | 0.412 | 0.398 |
| rain | 0.316 | 0.295 | 0.389 |
| snow | 0.210 | 0.066 | 0.724 |
| night | 0.155 | 0.569 | 0.276 |

# 6. Limitations and Real-Data Extension

The main limitation is that the executed results are synthetic. The benchmark is useful for reproducibility and for validating the direction of the proposed fusion rule, but it cannot establish real-world driving safety. Real datasets contain calibration errors, rolling shutter effects, weather mixtures, sensor saturation, road spray, dirty lenses, and rare objects that a synthetic feature benchmark cannot capture. Therefore, the numbers in this paper should be read as proof-of-concept results only.

A full real-data version should start with Seeing Through Fog because it contains gated NIR, radar, LiDAR, RGB, and FIR data under adverse weather. ACDC can test camera-only prompt adaptation across fog, rain, snow, and night. CADC can test snow-specific behavior, where the present synthetic run showed the greatest need for calibration. nuScenes can be used for multi-sensor pretraining because it has cameras, LiDAR, and radar. KITTI can serve as a clear-weather baseline. The code package is structured so that generate_dataset() can be replaced by dataset loaders while retaining the same fusion and evaluation routines.

Another limitation is that the prompt gate is currently hand-designed. A stronger version should train the prompt-to-gate module end-to-end using real detection loss and uncertainty regularization. The prompt can also be generated by a weather classifier or a foundation vision-language encoder. However, prompt injection and hallucination risks must be avoided. Weather prompts should be treated as context for reliability weighting, not as executable driving commands or as substitutes for sensor measurements.

Finally, deployment requires real-time constraints. The proposed gate itself is lightweight, but the full system inherits the cost of camera, LiDAR, radar, and BEV encoders. Future work should evaluate TensorRT or ONNX optimization, quantization, and fallback behavior when weather prompts are missing or inconsistent with sensor observations.

# 7. Conclusion

This paper presented WeatherPrompt-Fusion, a prompt-guided multi-modal perception framework for autonomous driving in adverse weather. The method conditions camera or gated image, LiDAR, and radar fusion on weather prompts and produces interpretable sensor weights for clear, fog, rain, snow, and nighttime conditions. The approach is inspired by recent adaptive gated-LiDAR fusion work and extends the idea toward broader weather-context guidance. In the included reproducible synthetic benchmark, WeatherPrompt-Fusion improves macro mAP over fixed average fusion, naive concatenation, and single-modality baselines, especially under fog and night. The paper deliberately avoids fabricated real-data scores and identifies a clear path for validating the method on public datasets. The most important next step is to replace the synthetic generator with real loaders for Seeing Through Fog, ACDC, CADC, nuScenes, and KITTI, then train the prompt-to-gate module end-to-end with detection and uncertainty losses.

# References

[1] Zhang, F., Guo, Z., Ding, J., Yang, J., & Liu, W. (2026). Adaptive sensor fusion for robust perception in dense fog: A gated vision and LiDAR integration framework. Sensors, 26(12), 3728. https://doi.org/10.3390/s26123728

[2] Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., & Heide, F. (2020). Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR42600.2020.01266

[3] Gruber, T., Julca-Aguilar, F., Bijelic, M., Ritter, W., Dietmayer, K., & Heide, F. (2019). Gated2Depth: Real-time dense lidar from gated images. In 2019 IEEE/CVF International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2019.00446

[4] Walia, A., Walz, S., Bijelic, M., Mannan, F., Julca-Aguilar, F., Langer, M., Ritter, W., & Heide, F. (2022). Gated2Gated: Self-supervised depth estimation from gated images. In 2022

IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR52688.2022.01243

[5] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR42600.2020.01108

[6] Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In 2012 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2012.6248074

[7] Sakaridis, C., Wang, H., Li, K., Zurbruegg, R., Jadon, A., Abbeloos, W., Olmeda Reino, D., Van Gool, L., & Dai, D. (2021). ACDC: The adverse conditions dataset with correspondences for semantic driving scene perception. In 2021 IEEE/CVF International Conference on Computer Vision. https://doi.org/10.1109/ICCV48922.2021.00977

[8] Pitropov, M., et al. (2021). Canadian adverse driving conditions dataset. The International Journal of Robotics Research, 40(4–5), 681–690. https://doi.org/10.1177/0278364921998768

[9] Cordts, M., et al. (2016). The Cityscapes dataset for semantic urban scene understanding. In 2016 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2016.350

[10] Vora, S., Lang, A. H., Helou, B., & Beijbom, O. (2020). PointPainting: Sequential fusion for 3D object detection. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR42600.2020.01110

[11] Teng, D. (2025). TEAS: Token- and energy-aware autoscaling for cost-efficient LLM serving. AI and Data Science Journal.

[12] Bai, X., Hu, Z., Zhu, X., Huang, Q., Chen, Y., Fu, H., & Tai, C.-L. (2022). TransFusion: Robust LiDAR-camera fusion for 3D object detection with transformers. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR52688.2022.01690

[13] Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., & Han, S. (2023). BEVFusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In 2023 IEEE International Conference on Robotics and Automation. https://doi.org/10.1109/ICRA48881.2023.10160649

[14] Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., & Beijbom, O. (2019). PointPillars: Fast encoders for object detection from point clouds. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2019.01329

[15] Yan, Y., Mao, Y., & Li, B. (2018). SECOND: Sparsely embedded convolutional detection. Sensors, 18(10), 3337. https://doi.org/10.3390/s18103337

[16] Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5099–5108).

[17] Yin, T., Zhou, X., & Kraehenbuehl, P. (2021). Center-based 3D object detection and tracking. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR46437.2021.01068

[18] Radford, A., et al. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (Vol. 139, pp. 8974–8986). PMLR.

[19] Teng, D. (2025). PACO: Predictive auto-configuration for SLO-constrained large language model inference serving. Innovation and Technology Studies.

[20] Zhou, X., Liu, M., Yurtsever, E., Zagar, B. L., Zimmer, W., Cao, H., & Knoll, A. (2024). Vision language models in autonomous driving: A survey and outlook. IEEE Transactions on Intelligent Vehicles, 9, 4501–4520. https://doi.org/10.1109/TIV.2024.3389112

[21] Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems (Vol. 30, pp. 5574–5584).

[22] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on Machine Learning (pp. 1050–1059). PMLR.

[23] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In 2017 IEEE International Conference on Computer Vision (pp. 2980–2988). https://doi.org/10.1109/ICCV.2017.324

[24] Wang, Z., Yang, J. S., Shang, W., & Ding, J. (2026). FairPromote: Explainable and fairness-aware talent promotion prediction via adversarial debiasing and SHAP-based interpretation. IEEE Access, 14, 72890–72904. https://doi.org/10.1109/ACCESS.2026.3583411

[25] Zi, B. (2024). Large language models for enterprise workflow automation in financial operations. Innovation and Technology Studies, 1(1), 24–29.

[26] Ding, J., Shen, Z., & Liu, W. (2026). Game-theoretic cost-sensitive adversarial training for robust cloud intrusion detection against GAN-based evasion attacks. Applied Sciences, 16(8), 3944. https://doi.org/10.3390/app16083944

[27] Teng, D., Rhee, M., Qin, Y., Zi, B., & Liu, W. (2026). SW-SpeedDLM: Sliding window speculative decoding for diffusion language models under long context constraints. Mathematics, 14(12), 2137. https://doi.org/10.3390/math14122137

[28] Zi, B. (2024). Cloud-native distributed systems for real-time payment intelligence. AI and Data Science Journal, 1(1), 51–56.

[29] Chen, Z., Wang, M., Zeng, Z., & Ping, W. (2026). Uncertainty-aware financial forecasting: Leveraging conformal prediction for risk-adjusted models. IEEE Access, 14, 74210–74221. https://doi.org/10.1109/ACCESS.2026.3591245

[30] Jiao, Y., Fan, H., Yue, X., Ping, W., Sun, T., & Wang, J. (2026). Dynamic heterogeneous graph contrastive learning for uncovering collusive financial fraud. Scientific Reports, 16, 12907. https://doi.org/10.1038/s41598-026-96214-3