

Thermal-Visible Image Fusion for Robust Autonomous Driving Perception Under Adverse Weather Conditions

Kerttu Pennanen

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.

kerttu.work@missouri.edu

Callan M. Segersun

Department of Computer Science, University of Houston, Houston, TX, USA.

callanmsegersun920@uh.edu

Joline Bohman

Department of Computer Science, University of New Hampshire, Durham, NH, USA.

jolinebohman223@unh.edu

Abstract

Autonomous driving perception must remain dependable when rain, fog, snow, dust, glare, and low illumination degrade visible-band imagery. Thermal-visible image fusion offers a multimodal strategy for combining complementary sensing physics, yet its value depends less on pixel-level enhancement than on system-level integration, architectural trade-offs, and deployment constraints. This paper examines thermal-visible fusion as a socio-technical perception infrastructure for robust autonomous driving under adverse weather. It analyzes sensing modality complementarity, fusion architectures, robustness limits, sensor placement, computational sustainability, and governance implications. The discussion emphasizes that fusion cannot eliminate ambiguity; it redistributes uncertainty across sensors, models, and operational domains. Effective designs therefore require calibrated degradation awareness, graceful degradation policies, traceable confidence estimation, and infrastructure support for maintenance and validation. The paper also considers fairness in pedestrian detection, policy questions around safety certification, and cross-domain lessons from rescue robotics, rail, and aviation. It argues that thermal-visible fusion should be treated as an accountable layer within a broader mobility system rather than as a standalone algorithmic remedy. The conclusion identifies research priorities in uncertainty-aware fusion, embedded efficiency, adverse-weather datasets, and regulatory frameworks that reward transparent robustness rather than benchmark overfitting.

Keywords

thermal-visible fusion; autonomous driving; adverse weather; multimodal perception; system architecture; robustness; governance; infrastructure; fairness; sustainability.

1. Introduction

Autonomous driving has advanced from early modular systems toward increasingly integrated perception and planning pipelines, but robust operation under adverse weather remains a persistent boundary condition [1]. Visible cameras provide rich texture, color, lane marking,

and semantic detail, yet their performance collapses in fog, heavy rain, snow, and nighttime glare. Thermal infrared cameras detect emitted heat and can preserve pedestrian, animal, and vehicle signatures when visible contrast is poor. Combining these modalities through thermal-visible image fusion is therefore attractive for autonomous driving perception under adverse weather. The underlying premise is that complementary physical information can compensate for modality-specific failure, producing a representation that is more stable than any single sensor stream.

However, the translation from fusion algorithm to deployable perception system is not straightforward. Benchmark datasets such as KITTI have shaped visual perception research around clear-weather urban driving, and their limitations reveal how evaluation routines can privilege conditions that are not representative of global operations [2]. A vehicle that performs well on sunny highways may fail in freezing fog, unlit rural roads, or dust-choked construction zones. Thermal-visible fusion must therefore be assessed not only by segmentation accuracy or detection precision, but by its contribution to system-level dependability, including failure detection, degraded-mode planning, maintainability, and regulatory accountability.

This paper adopts a system-level perspective. It treats thermal-visible fusion as an infrastructural capability embedded in vehicles, fleets, maps, maintenance regimes, and public policy. The discussion first reviews the operational context and requirements of adverse-weather perception. It then examines thermal and visible sensing physics, fusion architectures, robustness failure modes, deployment trade-offs, sustainability, fairness, and governance. Case illustrations from rescue robotics, rail, aviation, and smart-city infrastructure are used to clarify cross-domain lessons. The central argument is that thermal-visible fusion is necessary but insufficient: it can improve perception under adverse weather only when coupled with uncertainty-aware architecture, lifecycle governance, and honest operational domain definitions.

2. Operational Context and Requirements

Adverse weather perception is not a single technical problem but a family of operational conditions with distinct physical causes and safety consequences. Fog scatters light and reduces contrast, rain introduces droplets, reflections, and motion blur, snow creates occlusion and camouflage, and dust produces scattering and sensor fouling. Low illumination reduces visible-band signal while thermal sensing remains comparatively stable, though thermal contrast can invert when objects and backgrounds reach similar temperatures. These variations mean that a fusion model trained on one weather pattern may not transfer to another without explicit adaptation or multimodal redundancy.

The requirements for autonomous driving perception include detection, classification, localization, tracking, free-space estimation, lane understanding, and hazard prediction. Under adverse weather, these tasks become coupled. A missed pedestrian is not merely a detection error; it affects trajectory prediction, risk assessment, and emergency braking. A false positive may cause unnecessary braking, reducing trust and traffic efficiency. Thermal-visible fusion must therefore support calibrated confidence rather than merely maximize average accuracy. The system must know when fusion is reliable, when one modality dominates, and when the vehicle should enter a minimal-risk condition.

Operational design domains must be defined with weather, road type, speed, illumination, and sensor health in mind. A thermal-visible fusion pipeline designed for low-speed campus

shuttles may not satisfy high-speed highway requirements where detection range and latency are strict. Similarly, a fleet operating in desert environments must address dust accumulation and thermal saturation, while a fleet in northern winters must address snow buildup, salt spray, and ice on sensor housings. The perception stack is thus inseparable from maintenance logistics, cleaning systems, calibration schedules, and remote monitoring.

3. Thermal and Visible Sensing Modalities

Visible cameras remain the primary source of semantic richness in autonomous driving. They capture color, texture, signs, signals, and lane markings that thermal sensors cannot reliably provide. Their high spatial resolution and low cost make them attractive for dense perception. Yet visible sensing depends on reflected light, so fog, rain, snow, and night conditions reduce signal-to-noise ratio and contrast. Headlights and streetlights can also create glare and blooming, further complicating visual perception.

Thermal cameras detect long-wave infrared radiation associated with object temperature. They can reveal pedestrians, animals, and vehicles against cooler backgrounds, especially at night or in smoke and dust. Their relative independence from visible illumination is a major advantage. However, thermal imagery often lacks texture, may suffer from low contrast when thermal crossover occurs, and can be affected by environmental factors such as sun heating, wet surfaces, and sensor noise. Thermal sensors are also more expensive, less standardized, and sometimes subject to export controls, which affects fleet-scale deployment.

Thermal-visible fusion seeks to exploit complementarity rather than replace one modality with another. Surveys of infrared and visible image fusion describe a progression from spatial-domain and transform-domain methods to deep learning approaches that learn joint representations [3]. Classical techniques such as guided filtering can preserve edges and transfer details from visible to thermal imagery [4]. Deep learning methods can learn end-to-end fusion, but they require large, diverse datasets and careful regularization to avoid hallucinating structures that do not exist in either modality [5]. More recent surveys emphasize that fusion quality must be evaluated in downstream tasks, not only through visual metrics [6].

4. Fusion Architectures and Structural Trade-offs

Fusion architectures can be broadly grouped by where fusion occurs: early or pixel-level, intermediate or feature-level, and late or decision-level. Early fusion creates a combined image or tensor that downstream networks consume. It can preserve fine detail but is sensitive to misregistration and modality-specific noise. Intermediate fusion combines feature maps from separate encoders, allowing modality-specific processing before interaction. This approach often balances flexibility and performance, but it introduces complexity in training, calibration, and explainability. Late fusion keeps modality pipelines separate until decisions are combined, which can improve modularity and fault isolation but may lose cross-modal context.

For adverse weather, the choice of fusion level interacts with sensor placement and temporal alignment. If thermal and visible cameras are not spatially calibrated, pixel-level fusion may blur meaningful boundaries or create ghosting. If they operate at different frame rates or exposure settings, temporal fusion may combine stale and current information. Sensor placement studies show that field of view, mounting height, vibration, and occlusion strongly influence multimodal perception quality [10]. A fusion algorithm cannot compensate for a thermal camera mounted where it is frequently covered by road spray or snow.

A further trade-off concerns adaptive versus static fusion. Static fusion applies a fixed combination rule, which is simple but brittle when one modality degrades. Adaptive fusion estimates modality reliability from weather, illumination, motion, or learned uncertainty and adjusts contribution weights. Adaptive systems can be more robust, but they require confidence estimation that is itself reliable. In safety-critical settings, an incorrect reliability estimate can be more dangerous than a fixed conservative policy. Recent work has explored diffusion-driven semantic segmentation with multimodal image fusion for enhanced perception [12], illustrating how generative and discriminative components can be integrated. Such methods expand the design space, but they also raise questions about computational cost, interpretability, and failure modes that are not captured by average accuracy.

5. Robustness and Failure Modes in Adverse Weather

Robustness under adverse weather is often framed as a dataset problem, but it is also a failure-mode problem. Foggy scene understanding research has used synthetic data to improve semantic segmentation, showing that simulation can supplement scarce real-world annotations [8]. However, synthetic fog may not reproduce the optical complexity of wet roads, spray, and mixed illumination. Deep multimodal sensor fusion studies in unseen adverse weather demonstrate that radar, LiDAR, thermal, and visible sensors fail in different ways and that fusion must account for correlated as well as independent failures [9]. If rain blinds both visible and thermal cameras through a shared housing, redundancy collapses.

Thermal-visible fusion can improve pedestrian detection at night, but it can also introduce new failure modes. Thermal crossover can make a pedestrian disappear against a warm background. Wet clothing can reduce thermal signature. Visible modality may detect reflective clothing but fail in darkness. A fusion model trained to trust thermal cues may become overconfident when thermal contrast is low. Conversely, a model trained to trust visible edges may hallucinate lane markings in fog. Robustness therefore requires explicit degraded-mode logic, not only better fusion weights.

Failure modes also include temporal inconsistency. If fusion alternates between modalities across frames, object appearance may flicker, confusing tracking and prediction. Multi-LiDAR and multimodal tracking studies highlight the importance of temporal consistency for object detection and tracking [16]. In autonomous driving, flicker can cause unstable trajectories or unnecessary braking. A robust fusion system should maintain temporal coherence, detect sensor degradation, and communicate uncertainty to planning. It should also support fallback behaviors such as reduced speed, increased headway, or route diversion when perception confidence falls below a safe threshold.

6. Deployment, Infrastructure, and Sustainability

Deployment of thermal-visible fusion is not limited to vehicle hardware. It involves calibration bays, clean rooms, test tracks, weather chambers, data pipelines, over-the-air update systems, and fleet maintenance. Thermal cameras require periodic calibration and lens cleaning. Visible cameras require exposure and lens maintenance. Fusion models require retraining as sensor hardware, mounting positions, or operational domains change. These lifecycle costs are often underestimated in research prototypes, yet they determine whether a system remains reliable after months of service.

Infrastructure also shapes fusion performance. Smart intersections, roadside units, high-definition maps, and weather monitoring can provide external context that complements onboard sensing. For example, a road weather information system can inform the vehicle that

a bridge is icy or that fog is dense ahead, allowing the perception stack to adjust confidence thresholds before degradation occurs. However, such infrastructure introduces governance questions about data sharing, latency, cybersecurity, and equitable deployment. Vehicles operating in regions without smart infrastructure may require more onboard redundancy, increasing cost and energy consumption.

Sustainability is a further concern. Thermal sensors and high-performance computation add weight, power draw, and cooling requirements. Embedded automotive platforms must balance perception quality against thermal envelope and battery range [17]. A fusion model that improves accuracy by a small margin but doubles compute may be undesirable for electric fleets. Sustainable design may favor event-driven fusion, model compression, and selective high-resolution processing. It may also favor modular upgrades, where thermal modules can be added to vehicles in safety-critical routes rather than across entire fleets. The environmental cost of manufacturing and maintaining sensors should be considered alongside operational energy use.

7. Governance, Fairness, and Policy Implications

Governance of thermal-visible fusion must address safety certification, liability, privacy, and fairness. Safety certification traditionally relies on traceable requirements, validation, and verification. Deep fusion models complicate certification because their internal representations are difficult to interpret and their behavior may shift across weather conditions. A formal model of safe and scalable self-driving cars provides a useful reference for specifying safe fallback behavior [18]. Thermal-visible fusion should be integrated into such frameworks through explicit assumptions, monitored conditions, and verifiable degradation policies.

Privacy is another concern. Thermal imagery can reveal body heat signatures and may be perceived as more intrusive than visible imagery. Although thermal cameras do not capture identifiable facial detail in the same way as visible cameras, they can still detect presence, movement, and occupancy. Fleet operators must define retention policies, anonymization procedures, and access controls. Cross-border operations may face different legal regimes for infrared sensing and data localization. These issues affect public acceptance and regulatory approval.

Fairness is also relevant. Perception systems can perform unevenly across skin tone, clothing, age, disability, and mobility aids. Thermal-visible fusion may improve detection for some groups but not others. For example, pedestrians in heavy clothing may have reduced thermal contrast, while cyclists in reflective gear may be easier for visible cameras but harder for thermal ones. Fairness auditing should therefore include weather, illumination, and demographic proxies. Policy frameworks should require disaggregated performance reporting rather than only aggregate metrics. This is especially important for vulnerable road users who bear disproportionate risk in traffic collisions.

8. Case Illustrations and Cross-Domain Comparisons

Rescue robotics offers a useful comparison. Fusion of radar, LiDAR, and thermal information for hazard detection in low-visibility environments demonstrates how heterogeneous sensors can support operations in smoke, dust, and darkness [7]. Rescue robots often operate at lower speeds and with human supervision, which permits different risk tolerances than highway driving. Nevertheless, the architectural lesson is relevant: fusion must be tied to mission-

specific decisions such as stopping, mapping, or searching. A fused image is not an end in itself.

Rail and aviation provide additional lessons. Rail systems use track circuits, signals, and weather monitoring to maintain safe separation even when visibility is poor. Aviation uses redundant instruments, procedures, and air traffic control to manage adverse weather. These domains do not rely solely on improved cameras; they embed perception within layered safety systems. Autonomous driving should similarly treat thermal-visible fusion as one layer among many, including radar, LiDAR, maps, vehicle-to-everything communication, and operational rules.

Cross-domain comparison also highlights the role of datasets and benchmarks. The AI City Challenge and similar efforts show how large-scale multimodal benchmarks can drive progress in detection and tracking [15]. Yet benchmarks can create tunnel vision if they overrepresent clear weather or specific cities. End-to-end driving models such as DeepDriving and end-to-end learning for self-driving cars illustrate the appeal and opacity of learned perception-to-control pipelines [13], [14]. MultiNet demonstrates the value of joint semantic reasoning for real-time perception [11]. These works suggest that thermal-visible fusion should be evaluated within integrated perception systems, not as an isolated image enhancement task.

9. Future Directions

Future research should prioritize uncertainty-aware fusion that distinguishes aleatoric noise from model uncertainty and from sensor degradation. This requires architectures that can represent modality reliability, detect out-of-distribution weather, and communicate calibrated confidence to planning. It also requires datasets that capture rain, fog, snow, dust, glare, thermal crossover, and sensor fouling across geographies and seasons. Synthetic data can help, but it must be validated against real-world edge cases [8].

Embedded efficiency is another priority. Thermal-visible fusion must run within automotive power, latency, and cost budgets. Research should explore event-driven fusion, sparsity, quantization, and hardware-software co-design. It should also examine maintenance-aware learning, where models account for gradual sensor degradation and cleaning cycles. Sustainability metrics should be reported alongside accuracy, including energy per inference and lifecycle impact.

Finally, policy and governance research should develop certification methods for multimodal perception. Regulators need ways to audit fusion behavior under adverse weather, verify fallback policies, and assess fairness across road user groups. Standards for thermal calibration, data retention, and cybersecurity would reduce fragmentation. Public engagement is also necessary, because thermal sensing raises privacy expectations that cannot be resolved by technical accuracy alone. The most robust system will be one that is technically capable, institutionally accountable, and socially legitimate.

10. Conclusion

Thermal-visible image fusion offers a compelling pathway toward robust autonomous driving perception under adverse weather, but its promise depends on system-level design and governance. Visible and thermal sensors provide complementary information, yet both fail in structured ways that fusion cannot automatically resolve. Early, intermediate, and late fusion architectures each present trade-offs in accuracy, modularity, calibration, and explainability.

Adaptive and diffusion-driven methods expand the design space, but they also introduce new risks around compute cost, interpretability, and overconfidence. Deployment requires attention to sensor placement, maintenance, infrastructure, energy use, and lifecycle sustainability. Governance must address safety certification, privacy, fairness, and public accountability. The central conclusion is that thermal-visible fusion should not be treated as a standalone algorithmic fix. It should be embedded within layered perception, degraded-mode planning, transparent validation, and inclusive policy frameworks. By doing so, autonomous vehicles can move closer to dependable operation across the weather conditions that real mobility systems inevitably face.

References

1. Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., Kolter, J. Z., Langer, D., Pink, O., Pratt, V., Sokolsky, M., Stanek, G., Stavens, D., Teichman, A., Werling, M., & Thrun, S. (2011). Towards fully autonomous driving: Systems and algorithms. In 2011 IEEE Intelligent Vehicles Symposium (IV) (pp. 163–168). IEEE. <https://doi.org/10.1109/IVS.2011.5940562>
2. 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 (pp. 3354–3361). IEEE. <https://doi.org/10.1109/CVPR.2012.6248074>
3. Ma, J., Ma, Y., & Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, 45, 153–178. <https://doi.org/10.1016/j.inffus.2018.02.004>
4. Li, S., Kang, X., & Hu, J. (2013). Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7), 2864–2875. <https://doi.org/10.1109/TIP.2013.2244222>
5. Liu, Y., Chen, X., Wang, Z., Wang, Z. J., Ward, R. K., & Wang, X. (2018). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42, 158–173. <https://doi.org/10.1016/j.inffus.2017.10.007>
6. Zhang, H., Xu, H., Tian, X., Jiang, J., & Ma, J. (2021). Image fusion meets deep learning: A survey and perspective. *Information Fusion*, 76, 323–336. <https://doi.org/10.1016/j.inffus.2021.06.008>
7. Fritsche, P., Zeise, B., Hemme, P., & Wagner, B. (2017). Fusion of radar, LiDAR and thermal information for hazard detection in low visibility environments. In 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR) (pp. 96–101). IEEE. <https://doi.org/10.1109/SSRR.2017.8088146>
8. Sakaridis, C., Dai, D., & Van Gool, L. (2018). Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9), 973–992. <https://doi.org/10.1007/s11263-018-1072-8>
9. 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 (CVPR) (pp. 11682–11692). IEEE. <https://doi.org/10.1109/CVPR42600.2020.01170>

10. Pfeuffer, A., & Dietmayer, K. (2018). Optimal sensor placement for multi-sensor vehicle perception. *IEEE Sensors Journal*, 18(17), 7196–7206.
<https://doi.org/10.1109/JSEN.2018.2857830>
11. Teichmann, M., Weber, M., Zoellner, M., Cipolla, R., & Urtasun, R. (2018). MultiNet: Real-time joint semantic reasoning for autonomous driving. In *2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1013–1020). IEEE.
<https://doi.org/10.1109/IVS.2018.8500504>
12. Wang Z, He D, Zhao L, et al. Difusionseg: Diffusion-driven semantic segmentation with multi-modal image fusion for enhanced perception[J]. *Knowledge-Based Systems*, 2025: 114481.
13. Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015). DeepDriving: Learning affordance for direct perception in autonomous driving. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 2722–2730). IEEE.
<https://doi.org/10.1109/ICCV.2015.312>
14. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., & Zieba, K. (2016). End to end learning for self-driving cars. *arXiv*. <https://arxiv.org/abs/1604.07316>
15. Naphade, M., Chang, M.-C., Sharma, A., Anastasiu, D. C., Jagarlamudi, V., Chakraborty, P., Huang, T., Wang, S., Liu, M.-Y., Chellappa, R., Hwang, J.-N., & Lyu, S. (2020). The 2019 AI City Challenge. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 452–460). IEEE.
<https://doi.org/10.1109/CVPRW.2019.00062>
16. Sualeh, M., & Kim, G.-W. (2019). Dynamic multi-LiDAR based multiple object detection and tracking. *Sensors*, 19(6), 1474. <https://doi.org/10.3390/s19061474>
17. Kocić, J., Jovičić, N., & Drndarević, V. (2018). An end-to-end deep neural network for autonomous driving designed for embedded automotive platforms. *Sensors*, 18(7), 2120.
<https://doi.org/10.3390/s18072120>
18. Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. *arXiv*. <https://arxiv.org/abs/1708.06374>