Problem Statement:

Despite rapid advances in computer vision, real-time vision systems such as autonomous vehicles, surveillance, and robotics still face major challenges in low-light and adverse weather conditions (night, fog, rain, or snow). Images captured under these scenarios suffer from poor visibility, noise, and degraded features, leading to a significant drop in the performance of downstream perception tasks like detection, tracking, and recognition. Most current methods either optimize for traditional image quality metrics (like PSNR/SSIM)—which do not always correlate with task performance—or are too computationally intensive for real-time deployment on edge devices.

Key Gaps Identified:

1. **Task-Driven Optimization:** Most enhancement models are not tuned for improving downstream perception (detection/recognition), optimizing instead for perceptual quality metrics.
2. **Real-Time Edge Deployment:** Many recent methods, especially those using deep models, require large computation and cannot run efficiently on real-time, low-power edge devices.
3. **Multi-Weather Unified Architecture:** Existing approaches often handle only specific weather scenarios or require multiple separate models, preventing unified deployment.
4. **Domain Adaptation & Generalization:** Current methods struggle to generalize across different cameras and unseen conditions, and require extensive retraining or paired data.
5. **Multi-Modal Sensor Fusion:** There is little practical integration of complementary sensors (e.g., RGB, thermal, event cameras), which can provide key information in extremely adverse settings.
6. **Evaluation Framework:** Most benchmarks ignore real-world, safety-critical requirements, do not test multi-weather, cross-domain performance, or do not account for temporal consistency.

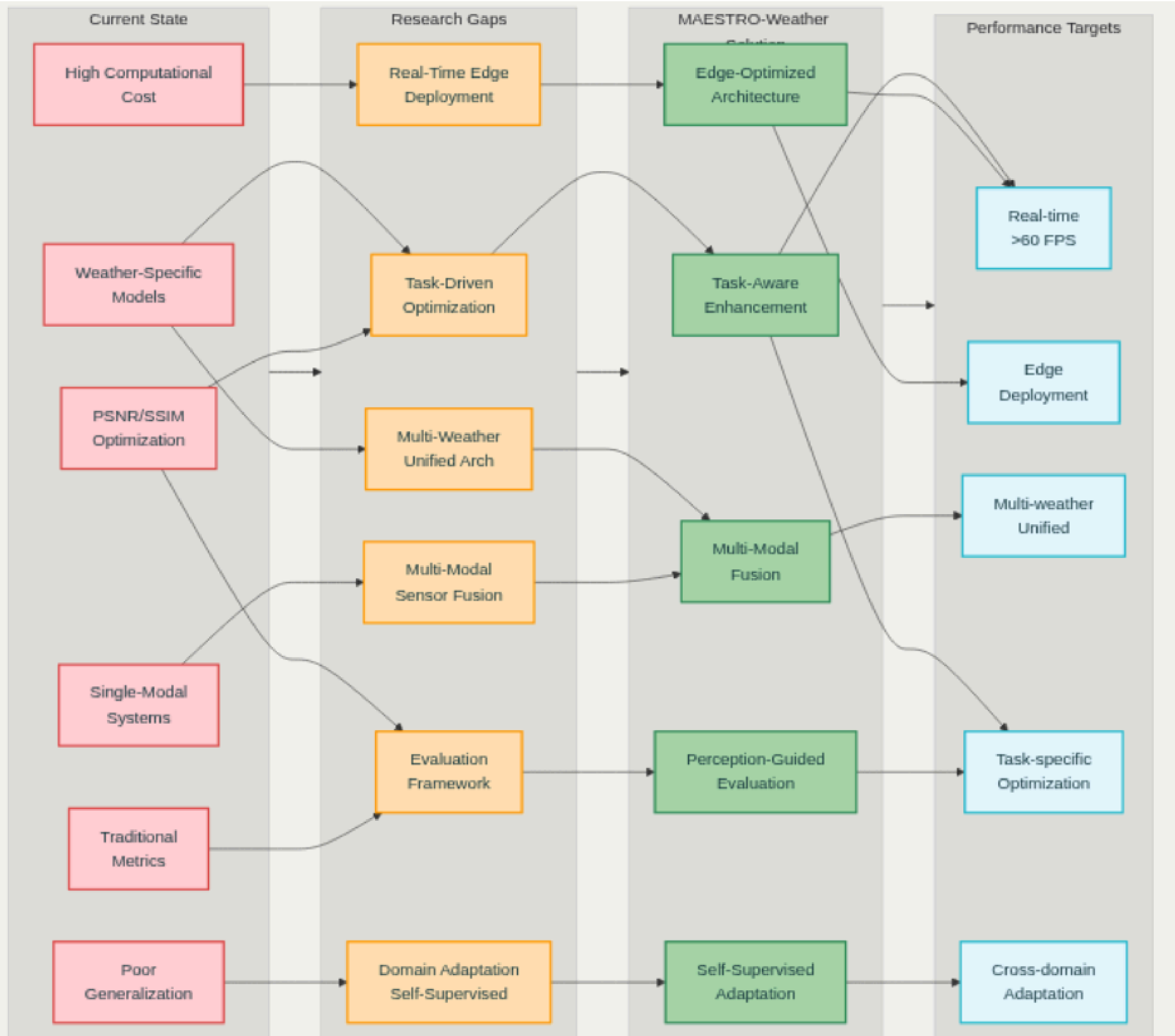
Proposed Solution (MAESTRO-Weather):

A unified, Task-Aware Multi-Modal Enhancement and Perception Optimization framework for Real-Time Edge Vision Systems:

- Task-Aware Enhancement: Jointly optimize both image enhancement and perception tasks (object detection, segmentation, tracking) using multi-task learning and perception-guided loss.
- Multi-Modal Attention Fusion: Integrate RGB, thermal, and event camera data via adaptive, attention-based fusion mechanisms to leverage complementary strengths in all conditions.
- Self-Supervised Domain Adaptation: Use synthetic weather transformations and contrastive learning to adapt to new domains and weather without the need for paired data.

- Perception-Guided Evaluation: Move beyond PSNR/SSIM—use the actual performance of perception tasks (mAP, IoU) and real-world benchmarks to evaluate enhancement.
- Unified Real-World Benchmarking: Validate across diverse datasets (multi-weather, cross-camera, night/fog scenarios) and on realistic edge devices.

This approach directly addresses all the key gaps and delivers a robust, deployable solution for low-light and adverse weather image enhancement that is tightly aligned with real-world, safety-critical vision needs



Literature Survey:

Overview of Existing Studies

The field of low-light and adverse weather image enhancement for real-time vision systems has evolved significantly from traditional physics-based approaches to sophisticated deep learning methods. Classical approaches include Retinex theory-based methods and atmospheric scattering models that explicitly model illumination and weather effects [1]. These methods remain valuable for interpretable corrections and low-compute scenarios but often struggle with complex real-world conditions.

Deep learning approaches have dominated recent research, with CNN-based end-to-end enhancers showing strong perceptual results [4,7]. Vision transformers and hybrid architectures have emerged to capture global illumination cues, particularly for automotive applications [8]. Advanced generative methods, including conditional diffusion models, achieve state-of-the-art restoration quality for multi-weather scenarios [6].

Lightweight architectures specifically designed for real-time deployment include linear RGB mapping approaches (FLIME) [3], image-to-curve models (LIMPID) [4], and re-parameterizable networks for maritime surveillance [9]. These methods prioritize computational efficiency while maintaining acceptable enhancement quality.

Key Applications and Performance

Current applications span autonomous driving with vehicle-focused enhancement systems [8], UAV tracking using transformer-guided enhancement [2], surveillance systems for ultra-high-definition transportation monitoring [7], and maritime applications with specialized lightweight networks [9]. Performance evaluation typically relies on PSNR/SSIM metrics, though recent works emphasize downstream task performance for perception systems.

Major Research Gaps and Limitations

Several critical gaps persist in current research:

Task-Driven Optimization: Most methods optimize for perceptual quality metrics (PSNR/SSIM) rather than downstream perception tasks (detection, tracking, recognition). This misalignment between enhancement objectives and real-world applications limits practical effectiveness [7,8].

Real-Time Performance Trade-offs: Current real-time systems sacrifice restoration quality for latency requirements. The balance between computational efficiency and enhancement quality remains suboptimal for safety-critical applications [3,9].

Domain Adaptation and Generalization: Methods often fail to generalize across different cameras, weather conditions, and deployment scenarios without retraining or fine-tuning [5]. Self-supervised adaptation approaches show promise but remain limited [5].

Standardized Evaluation: Lack of task-oriented, real-world benchmarks that include varied optics, motion blur, and mixed weather conditions. Current evaluation methods inadequately assess safety-critical deployment readiness [11,12].

Sensor Fusion Integration: Limited exploration of hybrid approaches combining RGB enhancement with thermal, radar, or event cameras for extreme conditions. Multi-sensor strategies remain underutilized despite their potential for robustness [1].

Temporal Consistency: Video-based enhancement methods struggle with maintaining temporal coherence while processing real-time feeds, particularly in dynamic weather conditions [5,10].

The field requires fundamental shifts toward task-aware optimization, robust cross-domain adaptation, and comprehensive evaluation frameworks that prioritize safety-critical performance over traditional image quality metrics.

References

[1] Miclea et al., "Visibility Enhancement and Fog Detection: Solutions for Mobile Systems," Sensors, 2021.

[2] "HighlightNet: Real-Time UAV Tracking," arXiv:2208.06818, 2022.

[3] Sun, "FLIME—Fast Low Light Enhancement for Real-Time Environments," AISC, 2023.

[4] Wu et al., "LIMPID: Lightweight Image-to-Curve Model," Photonics, 2023.

[5] Azizi & Kuo, "SALVE: Self-supervised Adaptive Video Enhancement," arXiv:2212.11484, 2022.

[6] Laidlaw et al., "Patch-Based Denoising Diffusion for Adverse Weather," arXiv:2207.14626, 2022.

[7] Qu et al., "Double Domain Guided Real-time Enhancement," arXiv:2309.08382, 2023.

[8] Wang et al., "VELIE: Vehicle-based Low-light Enhancement," Sensors, 2024.

[9] Lu et al., "Rep-Enhancer: Re-parameterizing for Maritime Surveillance," EUC, 2022.

[10] Liu et al., "Rain Streaks Removal in Low-light Images," CCC, 2022.

[11] Berviller et al., "Road User Estimation via Deep Learning," Sensors, 2023.

[12] Han & Yang, "Deep Learning Bridge Detection," J. Advanced Transportation, 2022.