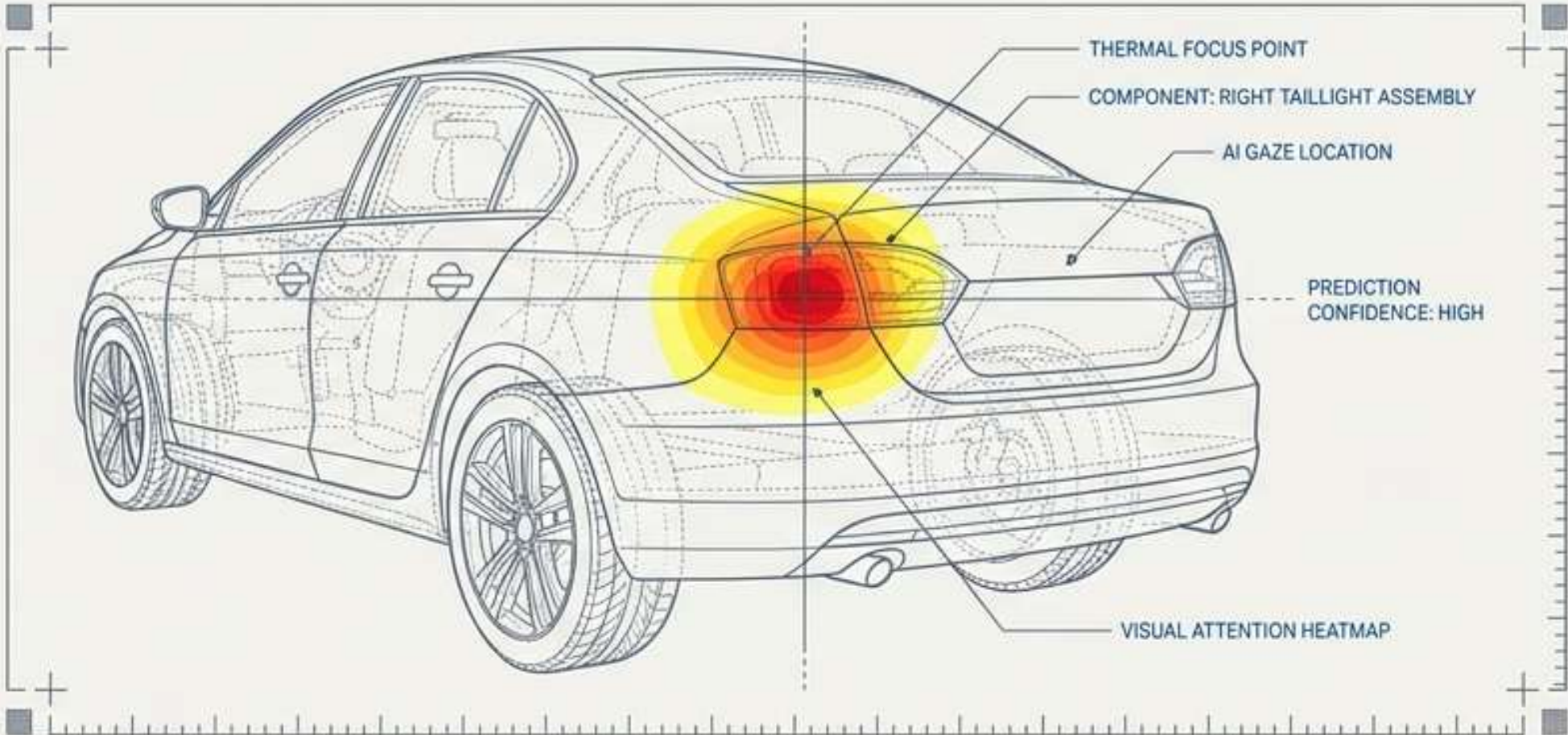
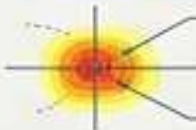
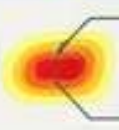


Illuminating the Black Box

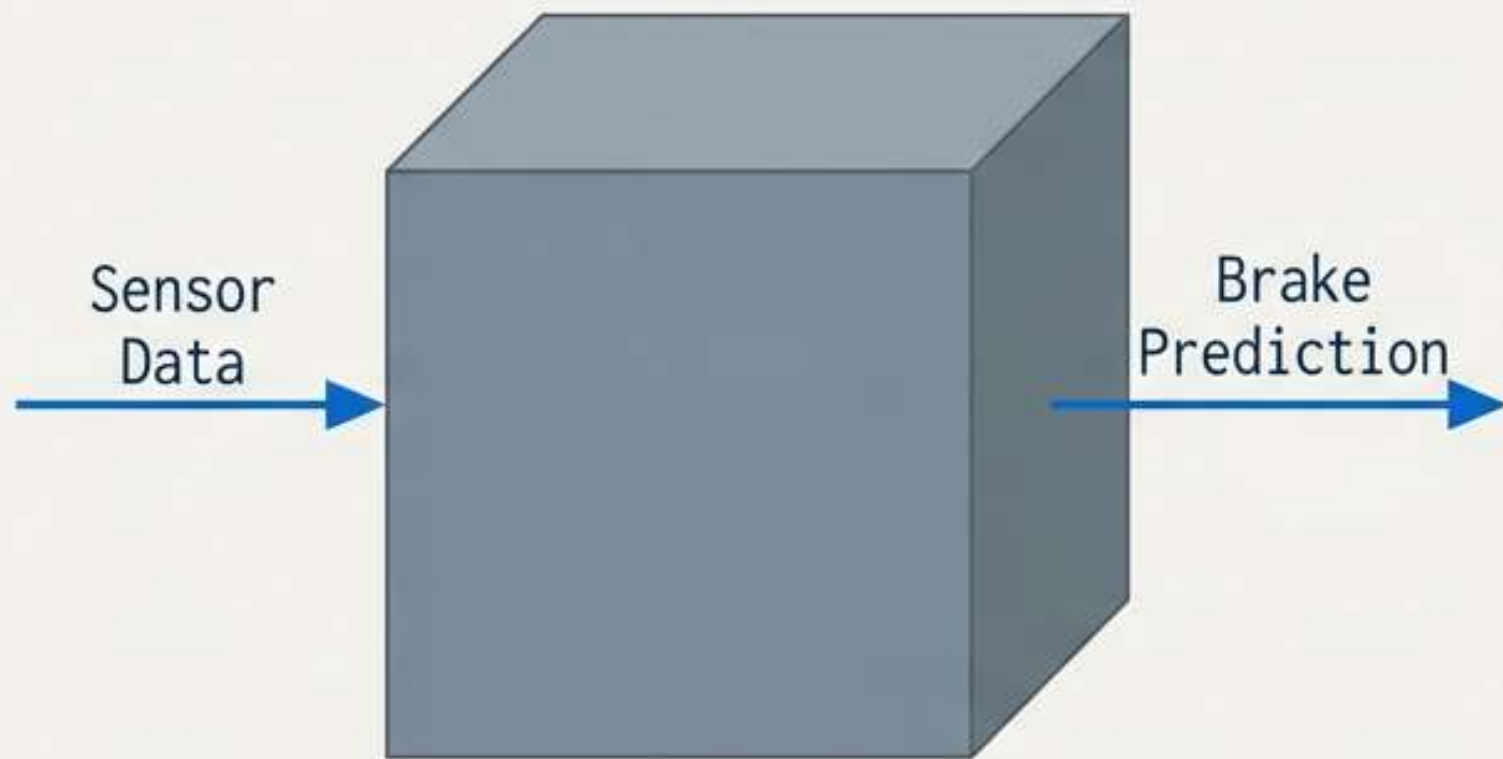
Real-Time Explainable AI for YOLO-Based
Brake Brake Prediction in
Autonomous Vehicles



YOLO-Based Detection	Feature Activation	XAI Explanation Methodology	Real-Time Output
OBJECT: BRAKE LIGHT CONFIDENCE: 98.7%	 OBJECT: BRAKE LIGHT CONFIDENCE: 98.7% SHAP VALUE: +0.85	SHAP VALUE: +0.85  SHAP VALUE SHAP VALUE	ACTION: DECELERATE PREDICTION



The Transparency Bottleneck in SAE Level 2 & 3 Autonomy



High Capability:

- Deep learning models perfectly **process dynamic multi-modal sensor data**.

Zero Transparency:

- The **reasoning behind critical decisions** ⚠ (e.g., detecting a vehicle braking ahead) is **completely hidden**. ⚠

The Trust Deficit:

- Lack of interpretability **blocks regulatory compliance, hinders error diagnosis, and prevents public trust** in intelligent transportation.

Accuracy is no longer enough. **Safety-critical perception requires real-time explainability.**

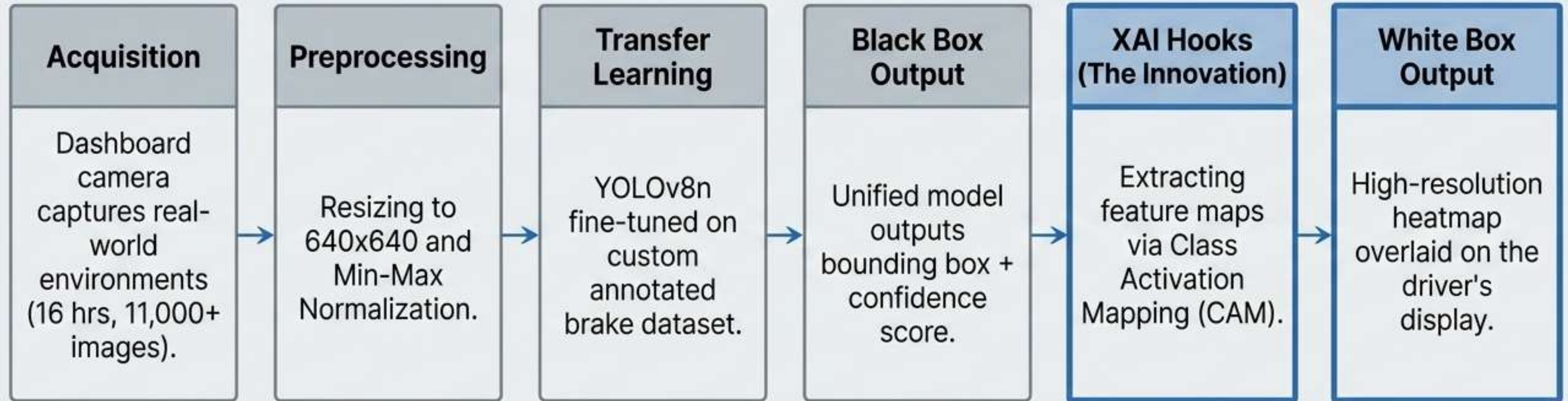
The XAI Showdown: Speed vs. Spatial Accuracy

Methodology	Performance Profile	The Verdict
LIME & SHAP	High fidelity for CNNs, but severe computational overhead (0.28s - 3.8s latency).	Too slow for real-time AV inference.
Traditional CAM (VGG16/ResNet)	Good for image classification, but requires slow back-propagation.	Lacks precise spatial localization for object detection.
The YOLO Architecture Challenge	Non-sequential feature extraction pipelines do not natively store activation details for individual objects.	A new framework is required.

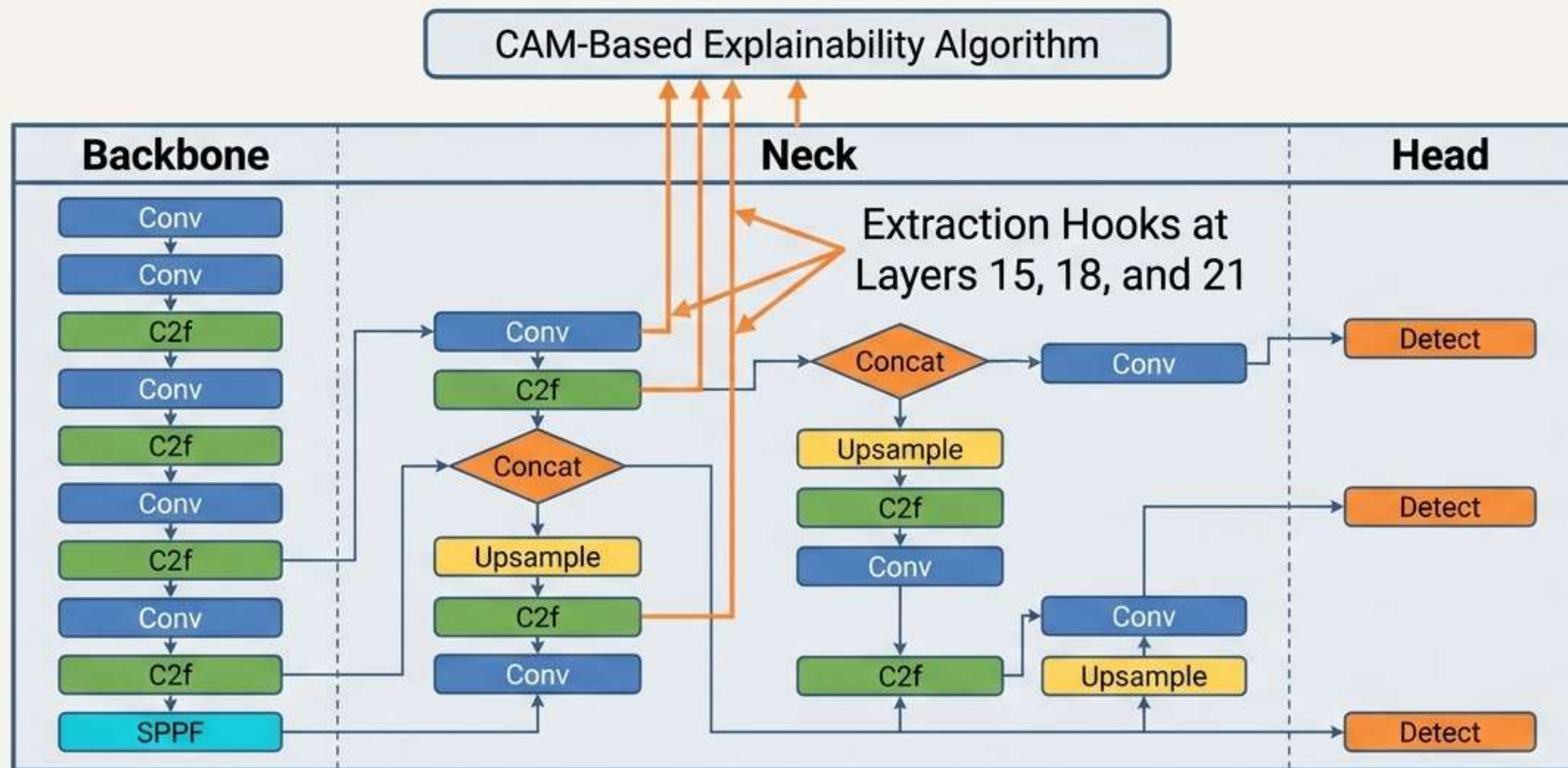
The Required Solution:

A **gradient-efficient CAM integration** that hooks directly into YOLOv8 without breaking its **real-time object detection speed**.

The Data-to-Decision Pipeline



Anatomy of an Explanation: Instrumenting YOLOv8



Hooks capture activations and gradients from intermediate convolutional layers, bypassing the opaque final detection process to trace exactly which pixels triggered the prediction.

Baseline Validation: Real-Time Accuracy Maintained

Precision  0.638 0.726 (Brake Light Off)	Recall  0.762 0.853 (Brake Light On)	Contextual Readout YOLOv8n was selected for its ultra-low parameter count (3.2M) and FLOPs (8.7G). It optimizes the balance between edge device inference speed and detection accuracy across diverse conditions. <table><tr><td></td><td></td><td></td></tr><tr><td></td><td></td><td></td></tr></table>						

Visual Evidence: From Opacity to True Transparency

The Black Box Output



The model correctly detects the vehicle and outputs a 'BrakeOn' classification, but we must blindly trust its reasoning.

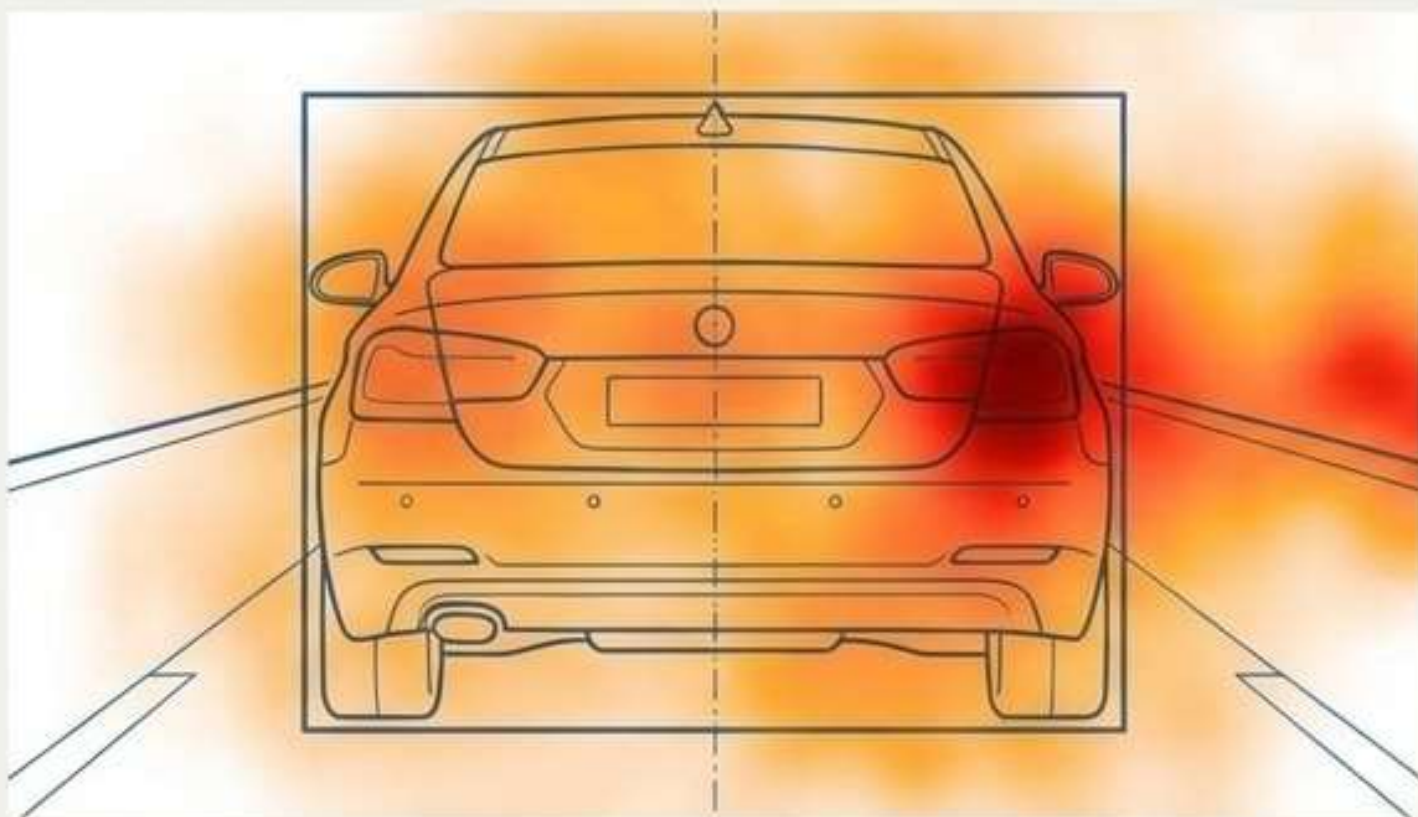
The White Box Output



The model is transparent. We can visually verify that it triggered the 'BrakeOn' decision specifically because it detected the physical illuminated taillights.

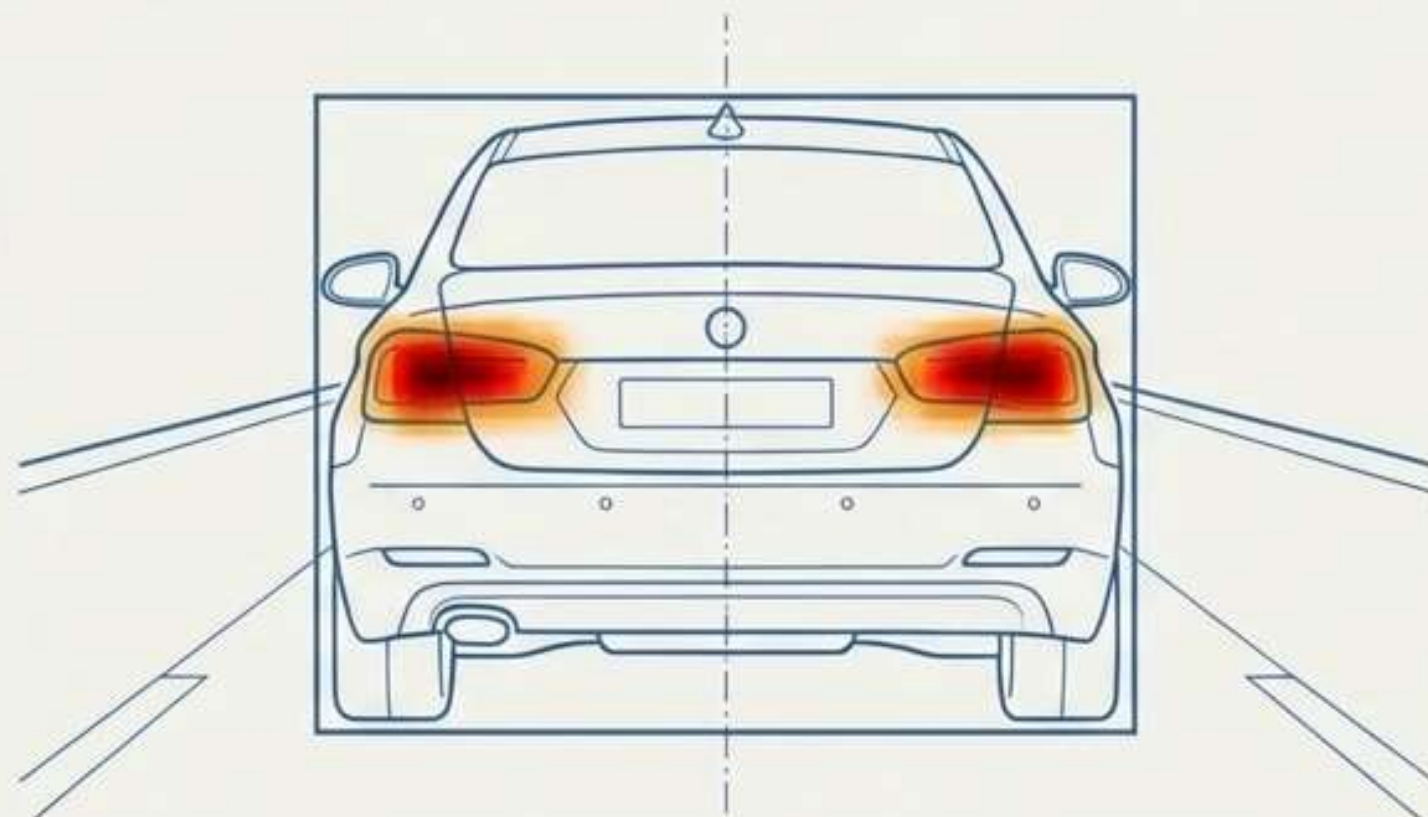
The XAI Diagnostic: Why HiResCAM Wins

Failing Grade: Background Bleed



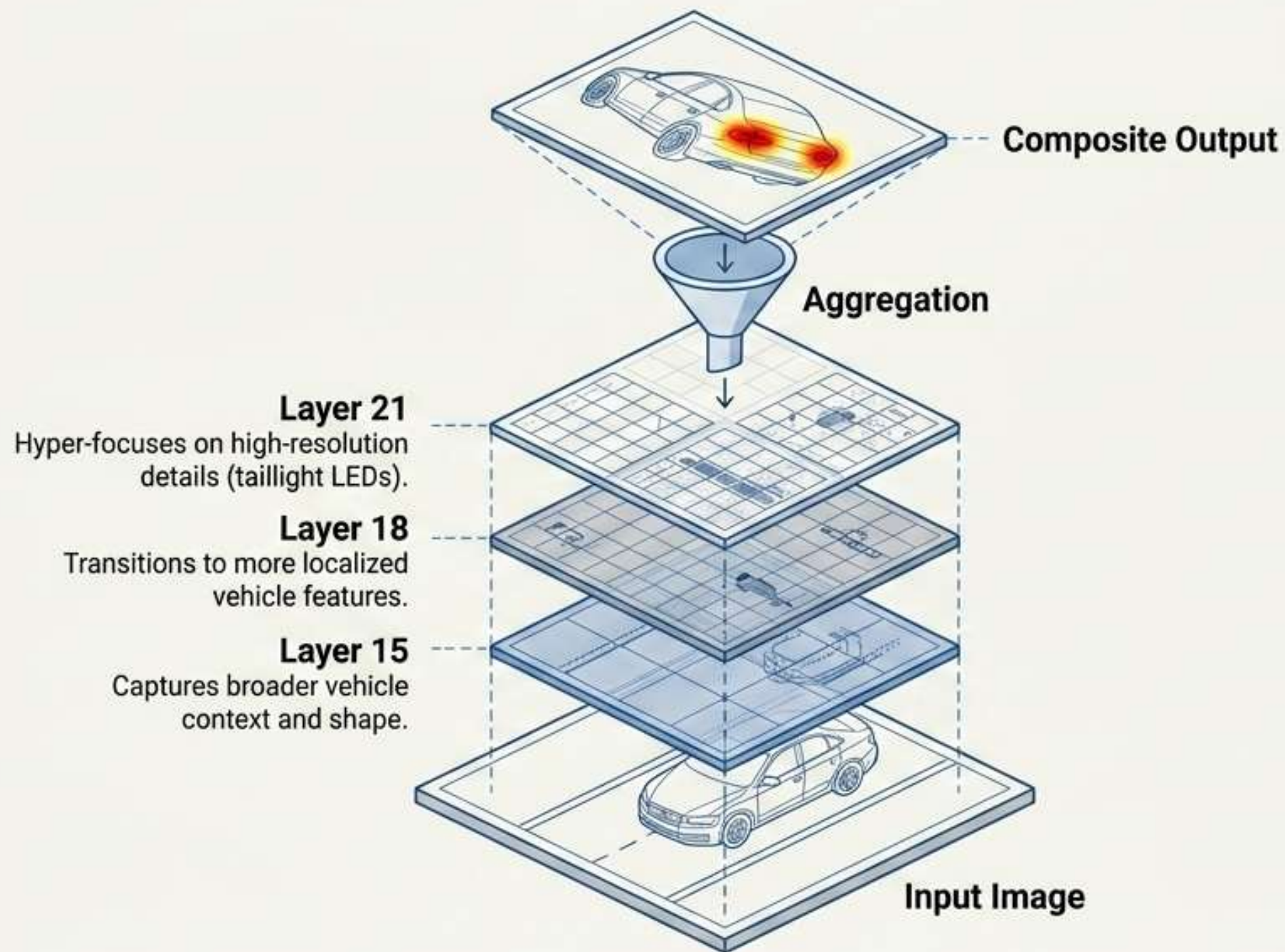
Techniques like GradCAM and EigenCAM generate activations that extend beyond the detection area, failing to provide precise localization at deeper layers.

Winning Grade: Precise Localization



HiResCAM refines the gradient averaging process with element-wise multiplication, successfully capturing class-discriminative features tightly localized to the detected object.

Multi-Scale Feature Aggregation



A single large feature manifests as multiple smaller features at different network scales. Aggregating layers 15, 18, and 21 guarantees the XAI accurately reflects the model's multi-resolution perspective.

Trust Through Transparency: The Path to Level 3+ Autonomy

System Validation



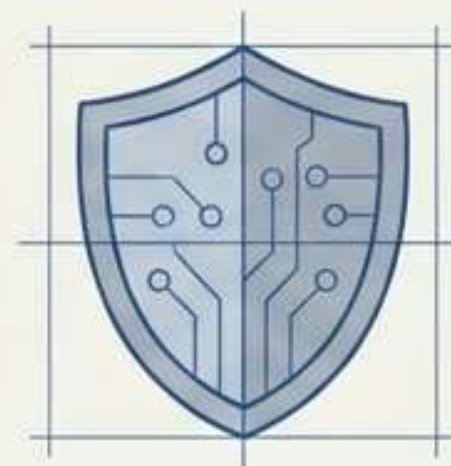
Empowers engineers to run fine-grained diagnostics and debug edge cases by seeing exactly what the model sees.

Driver Assurance



Bridges the gap between opaque AI and human reasoning. Drivers can visually verify that Adaptive Cruise Control is braking for the right reasons.

Regulatory Viability



Sets a new benchmark for model accountability, aligning AI-driven perception systems with strict safety requirements for intelligent transportation.

Real-time explainability is no longer a research luxury—it is the foundational prerequisite for safe, autonomous mobility.