Abstract

Modern visual recognition systems often fail where reliability matters most. For example, in safety-critical environments characterized by rare events, changing conditions, strict latency, and privacy constraints. This thesis addresses **three fundamental obstacles** that limit such systems: **Data Scarcity**, **Domain Drift**, and **Embedded Deployability**. Each is tackled through a dedicated contribution. (i) a Diffusion-based Data Augmentation Framework, (ii) a Video Stream–aware Online Student–Teacher Distillation Paradigm, and (iii) an Edge AI Deployable Stereo Surveillance Solution.

First, **we propose The Controllable Image Augmentation (CIA) Framework**. A Stable Diffusion-based generative pipeline that enriches scarce datasets with semantically faithful, auto-labeled synthetic samples. CIA complements classical augmentations by preserving spatial and semantic integrity, achieving up to a 10% improvement in mAP, and significantly enhancing generalization under data scarcity.

Second, **we contrast CIA generated data with physically rendered simulation data from a game engine**, revealing a precision–recall complementarity. CIA mixes improve precision through photometric realism, while simulation increases recall via structural diversity. Small ($\sim$10%) real-synthetic (generated or simulated) blends, yield the best trade-off. Naive Tri-source combinations (real/generated/simulated) lead to bias interference, motivating a curriculum mixing strategy that introduces data sources progressively by reliability and domain similarity.

Third, to mitigate domain drift under real-world variability, **we introduce the Stream-Based Active Distillation (SBAD) paradigm**. It's an online, budget-aware learning methodology for auto-adaptive edge deployment. SBAD scores incoming frames by utility, performing micro-batched updates with temporal coherence verification, to enable continual adaptation under bandwidth and latency constraints.

Finally, **we present INFRASECURE, an NVIDIA Orin–based Stereo Vision surveillance solution**, that operates in real time and ensures privacy by design. Instead of transmitting full video, it processes locally, generating event-level alerts and automatic safety reports.

Collectively, our contributions materialize within the lifecycle of the INFRASECURE Edge AI system designed for INFRABEL. They establish a coherent data centric framework that governs every stage of deployment, from addressing data scarcity through controlled augmentation, to ensuring real-time adaptability and privacy-preserving operation directly at the edge.