A Progressive Framework for Illuminant Estimation using Tiny Neural Networks: From RGB to Multispectral Fusion

Abstract

The accurate estimation and normalization of ambient lighting is a fundamental challenge in computer vision, crucial for ensuring the robustness of downstream tasks like object recognition and tracking. However, existing methods for Ambient Lighting Normalization (ALN) often assume simplified lighting conditions and are computationally too intensive for real-time applications on resource-constrained platforms. This proposal outlines a research plan to develop a novel, progressive, and lightweight framework for illuminant estimation using tiny neural networks. The research will commence by designing a highly efficient model using **standard RGB data**. If deemed relevant and beneficial, the framework will be **extended to incorporate Near-Infrared (NIR) data**, leveraging its unique properties to enhance robustness. Finally, the potential for a full **multispectral fusion** will be explored. The core architecture will be a flexible hybrid model that can be adapted to fuse data streams using a dynamic gating mechanism. This will ensure that additional spectral information is integrated intelligently, improving performance without unnecessary computational cost. The proposed models, optimized for deployment on edge devices, will be trained and evaluated on complex, multi-source colored lighting datasets like CL3AN [5], aiming to set a new state-of-the-art in efficient and scalable illuminant estimation.

1. Introduction and Motivation

Color constancy—the ability to perceive a consistent color for an object despite changes in the illuminating light source—is a cornerstone of robust visual understanding. In digital imaging, this translates to the task of illuminant estimation and Ambient Lighting Normalization (ALN), which aims to recover an image as if it were captured under uniform, neutral lighting. Achieving this is critical for a wide range of applications, from smartphone photography and 24-hour surveillance to autonomous driving and precision agriculture, where consistent visual input is paramount.

However, real-world illumination is far from simple. It often involves complex interactions of multiple colored light sources, occlusions, and diverse material properties, leading to intricate shadows, color shifts, and glare. Current state-of-the-art ALN methods often fail in these scenarios because they are trained on datasets with oversimplified assumptions, such as single white light sources. Furthermore, leading deep learning models, while powerful, are typically large and computationally expensive, making them unsuitable for

deployment on edge devices, or even directly in the sensor, where real-time performance and low computation are essential.

This research proposes to bridge the gap between high-accuracy ALN and the demands of edge computing by adopting a progressive approach. We will first establish a new performance baseline with a "tiny" neural network operating on standard **RGB data**. From there, we will systematically investigate the performance-to-cost trade-off of incorporating additional spectral bands, such as **Near-Infrared (NIR)**, which is less affected by color variations and provides robust structural details [2, 4]. This methodical progression will allow us to build a robust and truly efficient solution for illuminant estimation under complex, real-world lighting conditions.

2. Problem Statement and Research Gap

The central problem this research addresses is the lack of computationally efficient and robust illuminant estimation models capable of operating in real-time on edge devices under complex, multi-source colored lighting. The current research landscape, as evidenced by recent studies [1, 2, 4, 5], reveals several key gaps:

- **Oversimplified Lighting Models:** Most existing research focuses on single white light sources or uniform color fields, failing to address the non-homogeneous, multi-colored lighting common in real-world scenarios [5].

- **High Computational Cost:** State-of-the-art models for image restoration and ALN are too resource-intensive for edge deployment. As noted in [1], many models remain too computationally intensive for resource-constrained platforms.

- **Unexplored Progressive Fusion Strategies:** While data fusion (e.g., RGB-NIR) is a known technique, there is a lack of research that systematically evaluates the trade-offs of adding spectral complexity. It is unclear at what point the performance gain from additional bands like NIR is outweighed by the increased computational cost and potential for artifacts [2].

This PhD project will tackle these challenges by developing a novel, lightweight, and scalable architecture that performs illuminant estimation, starting with a minimal RGB-only approach and progressively incorporating more spectral information in a principled manner.

3. Proposed Methodology

To address the outlined problem, we propose the development of a lightweight, adaptable model architecture. The research will proceed in distinct phases, allowing for a thorough analysis at each stage of spectral complexity.

3.1. Phase 1: Foundational RGB Model

The initial focus will be on creating a state-of-the-art tiny neural network for illuminant estimation using only RGB data. The core of the model will be a hybrid of **Convolutional Neural Networks (CNNs)** and possibly **attention layers**, balancing the efficiency of CNNs for local feature extraction with the ability to capture long-range dependencies [1]. The network's objective will be to decompose the input scene into its constituent illumination (L) and reflectance (R) components, as described by the **Retinex theory** [5]. This RGB-only model will serve as the fundamental baseline for all subsequent work.

3.2. Phase 2: Extension to RGB-NIR Fusion

Once the RGB model is established, we will extend the architecture into a dual-stream model to incorporate NIR data. A key innovation will be the **dynamic fusion mechanism** connecting the two modalities. Instead of simple concatenation, the model will learn to dynamically weigh the features from the RGB and NIR streams at multiple scales. This adaptive approach will allow the model to:

- Rely more on the **NIR stream for structural guidance** when the RGB image is noisy or contains ambiguous textures [2].

- Prioritize the **RGB stream for color information** while suppressing potential structural artifacts from the NIR input.

- Enhance robustness to sensor noise or modality dropouts.

This phase will critically evaluate whether the inclusion of NIR data provides a significant enough improvement in accuracy and robustness to justify the increased model size and latency.

3.3. Phase 3: Potential Extension to Multispectral Data

If the RGB-NIR fusion proves highly effective, the framework's scalability will be tested by extending it to a broader range of multispectral data, such as **Short-Wave Infrared (SWIR)**. This would involve adapting the architecture into a multi-stream version, where each stream is a specialized sub-network for a specific spectral band. This extension could offer enhanced material disambiguation, as different materials may have distinct signatures in non-visible bands. The dynamic gating mechanism would become even more crucial here,

learning to intelligently select and weigh the most informative features from each spectral stream based on the scene's context.

4. Research Plan and Evaluation

This research will be conducted over three years, with the following key phases:

- **Year 1: Literature Review and Foundational RGB Model.** Conduct a comprehensive review of illuminant estimation and tiny neural networks. Implement baseline models and develop the initial **lightweight RGB-only architecture**. Train and evaluate this model on the CL3AN dataset [5] to establish a strong benchmark.

- **Year 2: RGB-NIR Extension and Fusion Analysis.** Develop the **dual-stream architecture and dynamic fusion mechanism** for RGB-NIR data. This will involve addressing the challenge of limited RGB-NIR datasets. Conduct extensive experiments to compare the performance and efficiency of the RGB-NIR model against the RGB-only baseline.

- **Year 3: Optimization, Multispectral Extension, and Thesis Finalization.** Refine the fusion model for optimal performance and computational efficiency. If promising, **prototype a multispectral extension**. Complete all experiments, finalize results, and write the PhD thesis. Submit findings to top-tier computer vision conferences.

5. Expected Contributions

This research is expected to make the following novel contributions to the field of computer vision:

1. **A Progressive and Lightweight Framework:** The first tiny neural network framework designed to be progressively scaled from RGB to RGB-NIR and potentially multispectral data for robust illuminant estimation.

2. **An Advanced Fusion Strategy:** A novel dynamic fusion mechanism that intelligently combines cross-spectral information, along with a rigorous analysis of the performance-vs-complexity trade-offs of such fusion.

3. **Edge-Ready Solution for ALN:** A practical, high-performance model for ALN that is computationally efficient enough for real-time deployment on resource-constrained platforms.

4. **New Benchmarks:** Establishing new state-of-the-art performance benchmarks for lightweight, progressive illuminant estimation on challenging datasets like CL3AN [5].

By successfully completing this research, we will advance the capabilities of computational color constancy and enable a new class of real-time, on-device applications that can operate reliably in uncontrolled and complex lighting environments.

6. References

[1] Galymzhankyzy, Z., & Martinson, E. (2025). Lightweight Multispectral Crop-Weed Segmentation for Precision Agriculture. *arXiv preprint arXiv:2505.07444*.

[2] Jin, S., Yu, B., Jing, M., Zhou, Y., Liang, J., & Ji, R. (2023). Dark VisionNet: Low-Light Imaging via RGB-NIR Fusion with Deep Inconsistency Prior. *Proceedings of the AAAI Conference on Artificial Intelligence, 37*(1), 1104-1112.

[3] Nanduri, A., Huang, S., & Chellappa, R. (2025). Multi-Domain Biometric Recognition using Body Embeddings. *arXiv preprint arXiv:2503.10931*.

[4] Wang, Y., Wang, H., Wang, L., Wang, X., Zhu, L., Lu, W., & Huang, H. (2025). Complementary Advantages: Exploiting Cross-Field Frequency Correlation for NIR-Assisted Image Denoising. *arXiv preprint arXiv:2412.16645*.

[5] Vasluianu, F. A., Seizinger, T., Wu, Z., & Timofte, R. (2025). After the Party: Navigating the Mapping From Color to Ambient Lighting. *arXiv preprint arXiv:2508.02168*.