INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	SAQUIB MAZHAR
Roll Number	:	186102113
Programme of Study	:	Ph.D.

Thesis Title: Lightweight Deep Learning Architectures for Semantic Scene Segmentation for Applications in Autonomous Driving

Name of Thesis Supervisor(s)	:	Prof. M.K. Bhuyan and Prof. Shaik Rafi Ahamed
Thesis Submitted to the Academic Division	:	Electronics and Electrical Engineering Department
Date of completion of Thesis Viva-Voce Exam	:	29/05/2025
Key words for description of Thesis Work	:	Semantic Segmentation, Scene Understanding, Real-time Processing, Lightweight CNNs, Deep Learning, Autonomous Driving

---

SHORT ABSTRACT

Semantic segmentation, which assigns a class label to each pixel in an image, is fundamental to autonomous driving systems that must perceive and understand complex environments in real time. Achieving high segmentation accuracy while maintaining computational efficiency remains a major challenge, particularly for embedded systems with limited resources. This thesis addresses these challenges by proposing novel methods that enhance both segmentation performance and efficiency. A key contribution is a composite loss function that integrates cross-entropy, boundary, and region-based losses to improve accuracy around object edges and better handle small and infrequent classes. This leads to improved segmentation results, especially for critical but underrepresented objects such as pedestrians and traffic lights. To support real-time operation, this work introduces lightweight network architectures. The Inverse Residual Dilation Pyramid Network (IRDPNet) employs efficient bottlenecks and multi-scale dilation to significantly reduce model size while preserving accuracy. The Block Attention Network (BANet) further enhances contextual understanding by integrating a modified attention mechanism that captures long-range dependencies with minimal overhead. Additionally, the Context-Guided Multi-scale Attention Network (CGMANet) is proposed to combine both low-level spatial features and high-level semantic cues through a hybrid attention mechanism. This architecture effectively balances detail preservation and context awareness, making it highly suitable for deployment on embedded platforms. Collectively, these contributions offer practical solutions for real-time, accurate semantic segmentation in autonomous driving scenarios.