Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

DLRV Project Proposal

Shadow Casting Object Segmentation

Kai Glasenapp

Sai Mukkundan Ramamoorthy

Shrikar Nakhye

Supervised by

Prof. Dr. Sebastian Houben

June 2025

1 Shadow Casting Object Segmentation

1.1 Model Evaluation and Objectives

The goal of this project is to compare different deep learning models for the semantic segmentation in aerial images. Special attention is given to the trade-off between segmentation accuracy and inference speed, as both are critical for real-world applications in the Hackathon.

To identify suitable architectures, various models will be trained and evaluated using the prepared dataset. During this process, hyperparameter tuning is applied to optimize model performance with respect to both pixel-wise accuracy and computational efficiency. This includes adjustments to learning rates, batch sizes, and loss function weighting.

Given the practical constraints of the project, all models are expected to be trained and tested within a limited time frame on the day of the hackathon. Furthermore, the models will be deployed to perform inference on real aerial images from a selected urban area, requiring fast processing and reliable segmentation results.

This setup reflects a realistic scenario in which trained models must generalize well to unseen data while remaining computationally lightweight enough for deployment on typical hardware. The project thereby explores not only the technical capabilities of different architectures but also their applicability in time-constrained, resource-limited environments such as smart city applications.

1.1.1 Evaluation Metrics

To assess and compare the performance of different semantic segmentation models, several quantitative metrics are used:

- **Training Time:** The total time required to train the model on the training dataset. This metric helps estimate the feasibility of training within a limited time frame, such as during the hackathon.
- **Inference Time:** The time a model needs to generate predictions for a single input image. Faster inference times are crucial for real-world applications

like city-wide segmentation.

- **GPU Memory Usage:** The amount of VRAM consumed during training and inference. Models with lower memory requirements are easier to deploy on hardware with limited resources.
- **Intersection over Union (IoU):** Measures the overlap between the predicted segmentation and the ground truth. It is calculated as the ratio of the intersection to the union of predicted and actual segments. IoU is a standard metric for semantic segmentation accuracy.
- **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure of a model's ability to correctly identify relevant pixels. Especially useful when dealing with class imbalance.

2 Dataset Description

The dataset used in this project consists of high-resolution aerial imagery provided by the City of Bonn. The images were captured in 2024 using an *IGI UrbanMapper-2* camera mounted on a *Cessna 404 TITAN* aircraft. These *Color Infrared (CIR) True Ortho TIFFs* include four spectral channels: red, green, blue, and near-infrared (NIR), with the NIR channel covering wavelengths between 700 and 850 nm. As true ortho images, all objects appear distortion-free and are geometrically corrected, making them particularly suitable for geospatial analysis.

Each image measures 10,000 × 10,000 pixels (250 m × 250 m) with a ground resolution of 2.5 cm per pixel. All images are georeferenced using associated worldfiles (TFW), allowing precise positioning in geographic information systems. The complete dataset comprises 2,736 TIFF files, occupying over 1.1 TB of storage.

To comply with data protection regulations, the original resolution of 2.5 cm per pixel was downsampled to 10 cm per pixel. This step ensures that individual persons or vehicles cannot be identified, thus addressing privacy concerns in accordance with legal and ethical standards for handling aerial imagery.

As a result of this resolution reduction and the exclusion of the near-infrared channel from the final training set, the dataset size was significantly reduced. The

downsampled RGB dataset now requires only approximately 70 GB of storage, making it more manageable for training and evaluation in standard machine learning environments while still preserving sufficient detail for the task of semantic segmentation.

3 Hardware

- Operating System: Linux (Kernel version 6.8.0)
- Processor: Intel Core i5-12400 (6 cores, 12 threads, up to 4.4 GHz)
- GPU: NVIDIA GeForce RTX 4070 (16 GB, CUDA 12.5)

3.1 Relevance of This R&D Project

- **Who will benefit from the results of this R&D project?**

  The primary beneficiaries of this project are urban planners, municipal authorities, and geospatial analysts involved in smart city initiatives. The models developed and evaluated here can also benefit emergency response teams, infrastructure maintenance units, and environmental monitoring agencies that require rapid, accurate aerial image analysis. Additionally, participants and organizers of hackathons focused on real-time geospatial AI will find the findings highly valuable.

- **What are the benefits? Quantify the benefits with concrete numbers.**

  The optimized models could provide a balance between accuracy and efficiency, potentially enabling near real-time inference with minimal computational overhead. For example, inference times may be reduced to under 0.5 seconds per 10,000 × 10,000 pixel image, and GPU memory usage could be kept below 8 GB, allowing deployment on mid-range GPUs such as the NVIDIA GeForce RTX 4070. Training time may be optimized to fit within a single hackathon day (typically 6–8 hours), facilitating rapid prototyping and iteration. Furthermore, by downsampling the original 1.1 TB dataset to a 70 GB RGB

dataset, training and evaluation could become significantly faster without a substantial loss in segmentation quality. This may lower the entry barrier for researchers and practitioners working in resource-constrained environments. Collectively, these efficiencies could support the development of scalable AI solutions for city-wide analysis and real-time decision-making.

4 Related Work

4.1 Survey of Related Work

- **What have other people done to solve the problem?** Several studies have employed deep learning techniques to detect shadow-casting objects from remote sensing data. In the literature, convolutional neural networks (CNNs) have been widely used for semantic segmentation tasks involving object detection based on shadow cues.

  A comprehensive review by Liu et al. (2023) [4] shows that models such as U-Net, DeepLabV3+, and FCN are frequently applied to high-resolution satellite and aerial imagery for shadow-related object detection. These models typically yield high segmentation accuracy but demand substantial computational resources, limiting their deployment on edge devices.

  This project builds upon these findings by developing and benchmarking efficient segmentation models specifically tailored to run in resource-constrained environments for the purpose of shadow-casting object detection.

4.2 Limitation and Deficits in the State of the Art

- **High Computational Requirements.** Many state-of-the-art deep learning models for urban vegetation mapping, such as DeepLabV3+ and Faster R-CNN, are computationally demanding. They require high-end GPUs and extensive training time, often making them impractical for deployment on low-power platforms like the NVIDIA Jetson Nano. For instance, Velasquez-Camacho et al. showed that although Faster R-CNN achieved good precision, its inference time and recall were significantly lower compared to more

lightweight models [5]. Similarly, Liu et al. highlighted that mainstream models trained on high-resolution imagery typically require substantial computational resources, making them less suitable for real-time or embedded use cases [4].

- **Reliance on Multispectral or Proprietary Data.** A number of leading models rely on non-RGB inputs such as near-infrared (NIR), thermal, or high-resolution LiDAR data to detect shadow-casting objects effectively. However, such datasets are not always publicly available or easy to collect in real time, especially for urban planning in resource-constrained regions. For instance, Zhang et al. developed an object-based shadow index using high-resolution satellite images with illumination intensity data, demonstrating the reliance on multispectral commercial data [6].

- **Inadequate Handling of Occlusions and Illumination Variability.** Shadow-casting object detection is highly sensitive to occlusions and changes in sunlight angles across time and space. Tall buildings, tree canopies, and overhangs can obstruct light paths and create complex shadow patterns. Additionally, seasonal and diurnal changes in sun elevation introduce significant variability in shadow geometry. Zhu and Woodcock [7] reported that cloud shadows and variations in illumination intensity complicated automated detection processes in satellite imagery. Similarly, Chen et al. highlighted challenges in distinguishing cast shadows from dark objects in cluttered urban scenes using traditional object segmentation methods [3].

- **High Dependence on Manual Annotation.** Most training datasets in this domain require meticulous manual labeling, including bounding boxes, masks, or canopy outlines. This process is labor-intensive and not scalable for city-wide or cross-city applications. Velasquez-Camacho et al. reported manually labeling over 40,000 trees for their deep learning pipeline, which represents a significant upfront investment in human resources [5]. Branson et al. also relied on supervised labeling for their fine-grained catalog of street trees, highlighting the barrier that annotation poses for automation and expansion [1].

5 Problem Statement

- **Which of the deficits are you going to solve?**

  Urban planning currently lack scalable, low-cost tools for accurately identifying Shadow-Casting Areas in Urban Aerial Data in real time. Most existing semantic segmentation models are too computationally intensive to be trained and require extensive annotation processing. And the GeoTIFF format of the Bonn dataset is of high dimensions and often require extensive pre-processing, before the image segmentation step. We leverage the crowd sourced annotation data from AI & Shadows Hackathon to use as a ground truth to train a less resource intensive image segmentation model to classify shaded regions in the city of Bonn.

- **What is your intended approach?**

  This project proposes to design and train a compact, efficient deep learning model tailored for semantic segmentation of objects which cast shadows based on high-resolution GeoTIFF imagery of Bonn. Key strategies include selecting a fast response and less data intensive architecture (satisfying the time and resource constraints of hackathon environment), training on downsampled RGB aerial imagery, and targeted data augmentation to improve generalizability.

- **How will you compare your approach with existing approaches?**

  The proposed model will be benchmarked against standard semantic segmentation architectures in terms of key performance metrics: accuracy (IoU, F1 Score), inference time, training time, and GPU memory usage. Comparisons will be conducted using the same dataset under controlled conditions. Particular focus will be given to the trade-off between performance and efficiency, evaluating which models are feasible for real-time deployment and inference, thereby satisfying the time constraints and resource constraints in a hackathon setting.

5.1 Deliverables

Minimum Viable

- To perform image segmentation on Bonn city dataset to identify the different types of objects such as buildings and trees which specifically cast shadows. We intent to use latest YOLO image segmentation model and use the hackathon participants annotation data as the ground truth.

Expected

- To geo-reference the segmented mask data (shadow casting object data) and also to compare and evaluate the performance with other state of art image segmentation models.

Desirable

- Quantize the segmentation model and adapt it to resource constrained environment like Nvidia Jetson to validate the real time inference of the model using GARRULUS[2] project drones (*only if time permits).

Please note that the final grade will not only depend on the results obtained in your work, but also on how you present the results.

References

[1] Steve Branson, Jan Dirk Wegner, David Hall, Nikita Lang, Konrad Schindler, and Pietro Perona. From google maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135:13–30, 2018. doi: 10.1016/j.isprsjprs.2017.11.008.

[2] Hochschule Bonn-Rhein-Sieg. Garrulus. URL `https://www.h-brs.de/de/garrulus`. Accessed: 2025-06-03.

[3] Andres Sanin, Conrad Sanderson, and Brian C Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Pattern Recognition*, 45 (4):1684–1695, 2012. doi: 10.1016/j.patcog.2011.09.017.

[4] Luisa Velasquez-Camacho, Maddi Etxegarai, and Sergio de Miguel. Implementing deep learning algorithms for urban tree detection and geolocation with high-resolution aerial, satellite, and ground-level images. *Computers, Environment and Urban Systems*, 105:102025, 2023. ISSN 0198-9715. doi: https://doi.org/10.1016/j.compenvurbsys.2023.102025. URL `https://www.sciencedirect.com/science/article/pii/S0198971523000881`.

[5] Luisa Velasquez-Camacho, Maddi Etxegarai, and Sergio de Miguel. Implementing deep learning algorithms for urban tree detection and geolocation with high-resolution aerial, satellite, and ground-level images. *Computers, Environment and Urban Systems*, 105:102025, 2023. doi: 10.1016/j.compenvurbsys.2023.102025.

[6] Tian Zhang, Hong Fu, and Chuan Sun. Object-based shadow index via illumination intensity from high resolution satellite images over urban areas. *Sensors*, 20(4):1077, 2020. doi: 10.3390/s20041077.

[7] Zhe Zhu and Curtis E Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012. doi: 10.1016/j.rse.2011.10.028.