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 RD Project

- **Beneficiaries:** Urban planners, municipal authorities, geospatial analysts, emergency responders, and hackathon participants will benefit from real-time, low-cost aerial image analysis tools.
- **Quantified Benefits:** The proposed model may achieve inference times ¡ 0.5s per 10,000×10,000 px image with ¡8 GB GPU memory (e.g., RTX 4070), enabling training within 6–8 hours. Downsampling the 1.1 TB dataset to 70 GB RGB format accelerates processing, making deployment feasible for mid-range hardware and hackathon settings.

4 Related Work

4.1 Survey of Related Work

- **Prior Approaches:** Deep learning models like U-Net, DeepLabV3+, and FCN have been widely used for shadow-casting object detection from high-resolution aerial imagery, achieving high accuracy but requiring heavy computational resources [4]. This project extends their findings by focusing on efficient architectures suitable for resource-limited environments.

- **High Computational Requirements:** Models like DeepLabV3+ and Faster R-CNN require powerful GPUs and are impractical for real-time low-power deployment [4, 5].
- **Dependence on Proprietary Data:** Leading models rely on multispectral or commercial datasets like NIR and LiDAR, limiting accessibility in resource-constrained areas [6].
- **Occlusion and Illumination Variability:** Shadow detection suffers under occlusions and changing sunlight, complicating accurate segmentation [3, 7].
- **Manual Annotation Bottleneck:** State-of-the-art approaches demand extensive manual labeling, hindering scalability across urban landscapes [1, 5].

5 Problem Statement

- **Deficits Addressed:** Most existing models for shadow detection are computationally intensive, require large annotated datasets, and are unsuitable for real-time low-power applications [4, 5]. Additionally, the high-resolution GeoTIFF format of the Bonn dataset demands significant pre-processing.
- **Proposed Approach:** We aim to develop a lightweight, efficient deep learning model for semantic segmentation of shadow-casting objects using downsampled RGB GeoTIFF imagery of Bonn. The model leverages crowd-annotated data from the AI & Shadows Hackathon and incorporates fast architectures with targeted augmentation to enhance generalization.
- **Evaluation Strategy:** Our model will be benchmarked against standard segmentation architectures based on accuracy (IoU, F1), inference/training time, and GPU usage, focusing on performance-efficiency trade-offs suitable for hackathon environments.

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.