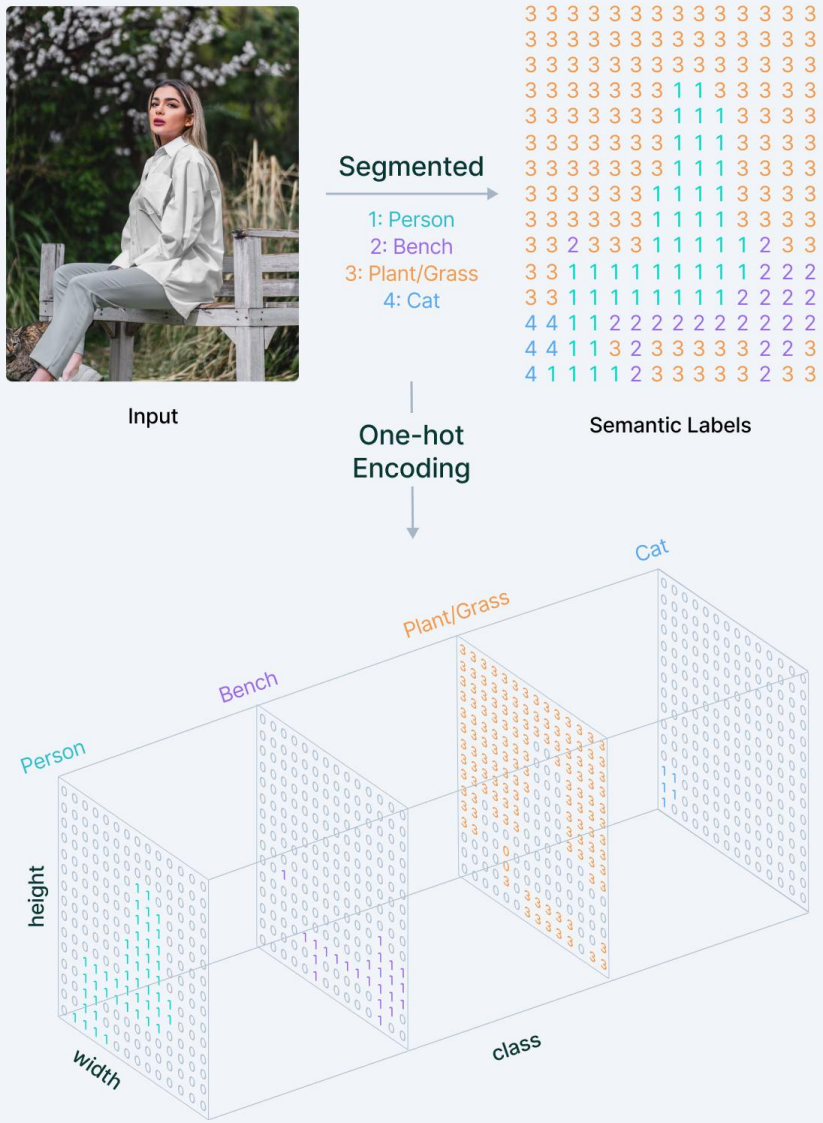
Comparative Evaluation of Segmentation vs Specialized Models for Off-Road Traversability Estimation

Mrinank Sivakumar – a1936624
Supervisor: Dr. Feras Dayoub
University of Adelaide – Research Project A

THE UNIVERSITY of ADELAIDE

Project Introduction

- Problem:
  - Off-road environments are unstructured.
  - Autonomous robots struggle to identify traversable vs non-traversable areas.



Project Introduction

- Research Gap:
  - **Segmentation models**
    - Good at recognizing visual classes
    - Not tailored for traversability
  - **Specialized models**
    - Incorporate geometry & robot-specific footprints
    - Are more complex

THE UNIVERSITY of ADELAIDE

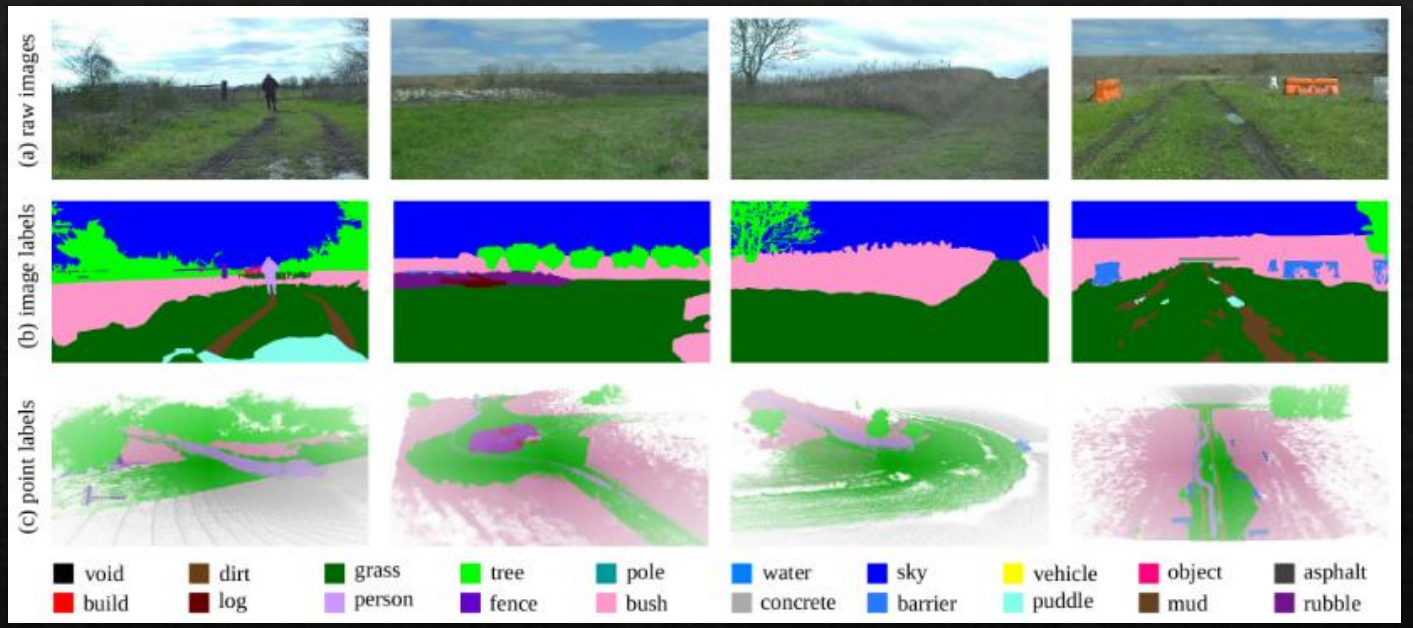
Project Introduction

- Research Question:
  - How effective is a segmentation-based model compared to a specialized traversability model on off-road datasets?
- Objective:
  - Evaluate Grounded-SAM as a baseline and compare with Follow the Footprints.

THE UNIVERSITY of ADELAIDE

Dataset Overview

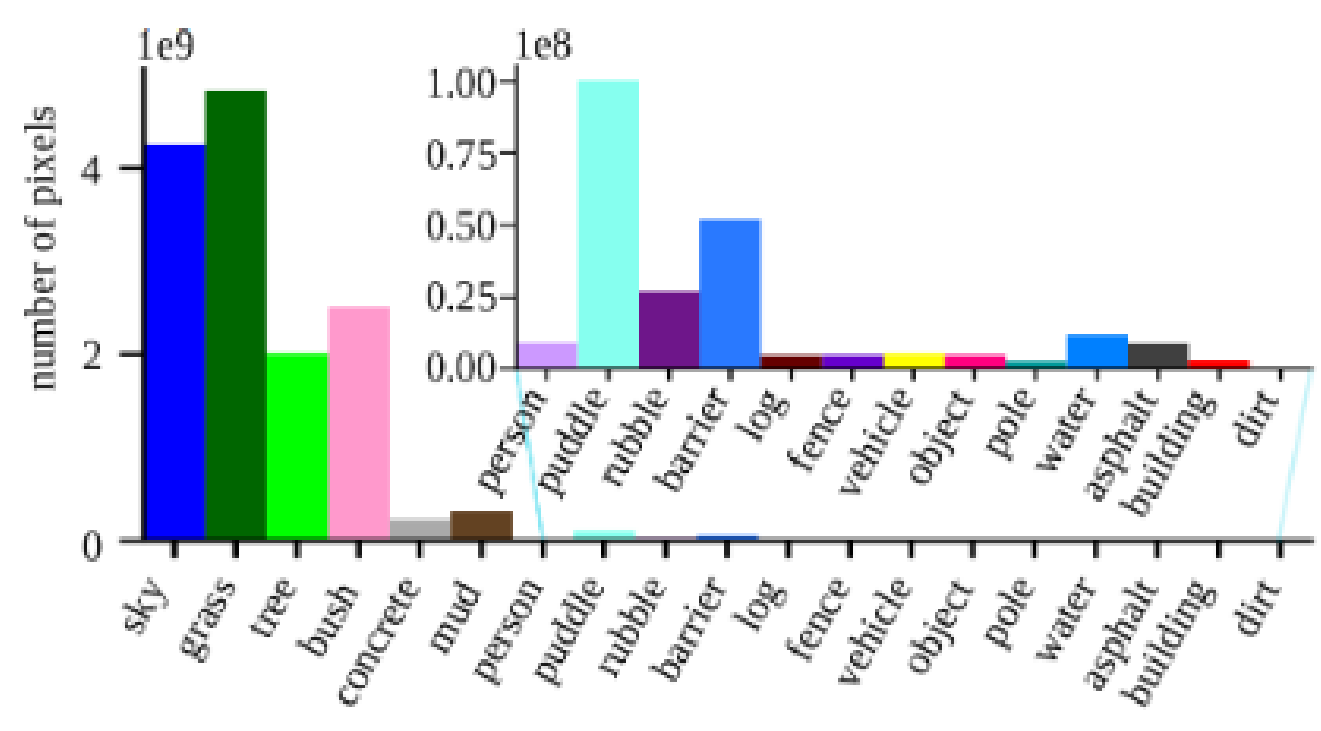
- RELLIS-3D
  - Outdoor off-road robot dataset
  - Modalities: RGB, Semantic labels, LiDAR, GPS/IMU
  - We use: **RGB + Labels only**



Preprocessing & Current Status

- Preprocessing plan:
  - Retain labels for EDA & per-class analysis
  - Grounded-SAM: Prompt, Mask, Post-Map
  - FtFoot: RGB, Traversability Map
- Current Status
  - Dataset downloaded, EDA in progress
  - Model setup in progress

Data Analysis: Current Work



- Exploratory Data Analysis (EDA)
  - Reviewing image-label pairs
  - Analyzing class distribution stats
  - Visualizing sample frames
- Model Setup
  - Grounded-SAM
    - Building environment, installing packages, and writing dataset-to-model pipeline
  - Follow the Footprints
    - Prepare WSL Ubuntu and other required applications/packages

Data Analysis: Planned Evaluation

- Comparison Plan:
  - Run Grounded-SAM and FtFoot on dataset
  - Evaluate qualitative performance (visual alignment)
  - Evaluate quantitative metrics (IoU, Precision, Recall)
- Key Evaluation Points:
  - How well Grounded-SAM semantic segmentation aligns with traversability
  - How accurately FtFoot maps match label-based regions
  - Differences in behaviour across terrain types
- Feedback for:
  - Evaluation metrics
  - Prompt design and fusion strategies
  - Better visual comparability and other suggestions

Model Architectures

Grounded-SAM

- Modular architecture
  - Grounding DINO: text-conditioned object detector
  - SAM: general-purpose segmentation on detected boxes
- Input: RGB image + text prompt
- Output: segmentation mask per prompt
- Not trained for traversability & requires external input to map classes
- Strengths: Open-vocabulary & Generalization
- Weaknesses: No geometry & No robot awareness

Follow the Footprints

- TE network composed of two core modules
  - Guide Filter Network (GFN)
    - Fuses RGB-D features with surface normal using guide filter layers
    - Dynamically conditioned convolutions per pixel
  - Footprint Supervision Module (FSM)
    - Uses robot's prior trajectory as weak supervision
    - Applies random walk to propagate traversability from traversed areas
- Output: Traversability heatmap
- Strengths: Self-supervised, Robot-aware, & Multimodal
- Weaknesses: Requires trajectory data & Is more complex to deploy

Thank You for Listening

References

Grounded-SAM Paper:
Li, X, Li, X, Qin, H, Wang, Y, Wang, H, He, C, Zhao, T & Shao, J 2024, 'Grounded segment anything', *arXiv preprint*, arXiv:2401.14159.

Follow the Footprints (FtFoot) Paper:
Jeon, Y, Lee, D, Kim, H, Kim, J, Choi, JW & Kim, HJ 2024, 'Follow the footprints: Learning geometry-oriented traversability from RGB video through semantic footprints', *arXiv preprint*, arXiv:2402.15363.

RELLIS-3D Dataset Paper:
Jiang, P, Osteen, P, Wigness, M & Saripalli, S 2022, 'RELLIS-3D dataset: Data, benchmarks and analysis', *arXiv preprint*, arXiv:2011.12954v4.