Cartography
GIS
Remote Sensing

Davide Mattioli

XAI for Linear Woody Feature Segmentation

October 2025

Table of contents

1 Introduction

2 Knowledge Gap

3 Methodology

4 Results

5 Future work

What is a linear woody feature

A linear woody feature is a narrow, elongated strip of woody vegetation (shrubs and/or trees) such as hedgerows or tree lines used as field boundaries, which is operationally defined as less than 20 m wide.

Why segment linear woody features?

- Accurate hedgerow maps unlock credible biodiversity, soil protection, and climate planning across farmland.
- Generic land-cover layers miss many narrow hedges, so dedicated segmentation fills critical monitoring and policy gaps.
- Outputs guide restoration targeting and connectivity

Architecture used in pipeline

- DinoV3 Visual Transformer (ViT)
- SR: Super Resolution algorithm
- AnyUP: universal features upsampling model
- UNet: CNN for Image segmentation

Relevant work

- Ahlswede et al. (2021): CNN detect hedgerow objects in very high-resolution satellite images, object-level extraction of linear woody features.
- Huber-García et al. (2025): CNNs (DeepLab v3) on orthophotos operational hedgerow map for Bavaria,
- Muro et al. (2025): Multitemporal 3 m PlanetScope with U-Net semantic segmentation for national-scale hedgerow mapping and indicators in Germany.

Explainability

- **Transformer probing**: Analyze internal representations of DINOv3 layers
- **SHAP values**: Quantify feature importance using Shapley Additive Explanations to reveal which input pixels most influence segmentation decisions
- **Attention visualization**: Extract and visualize self-attention maps from transformer blocks to identify spatial relationships and focus regions
- **Goal**: Make black-box vision transformer decisions interpretable for agricultural monitoring and policy applications

Transformer-based Segmentation Pipeline

Pipeline stages:

- DINOv3 feature extraction
- SR: Sentinel-2 enhancement (10m $\rightarrow$ finer detail)
- AnyUP: Feature-space upsampling
- UNet with custom loss for linear structures

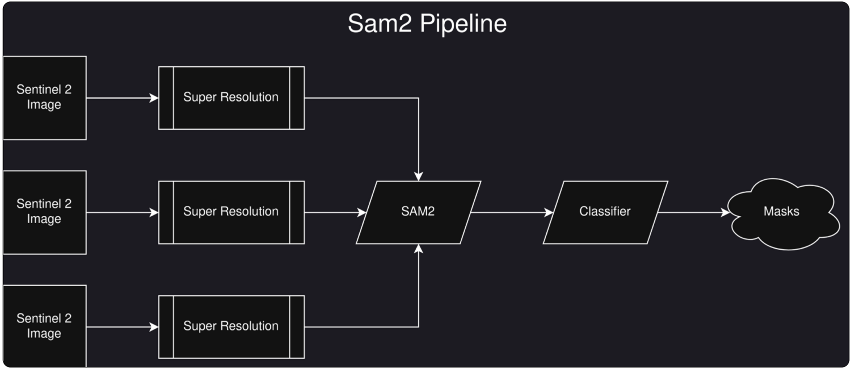
Ablation questions:

- Transformer vs. CNN encoder?
- Impact of SR on accuracy?
- Feature upsampling (FSR) contribution?

Foundation Model Generalization Capabilities

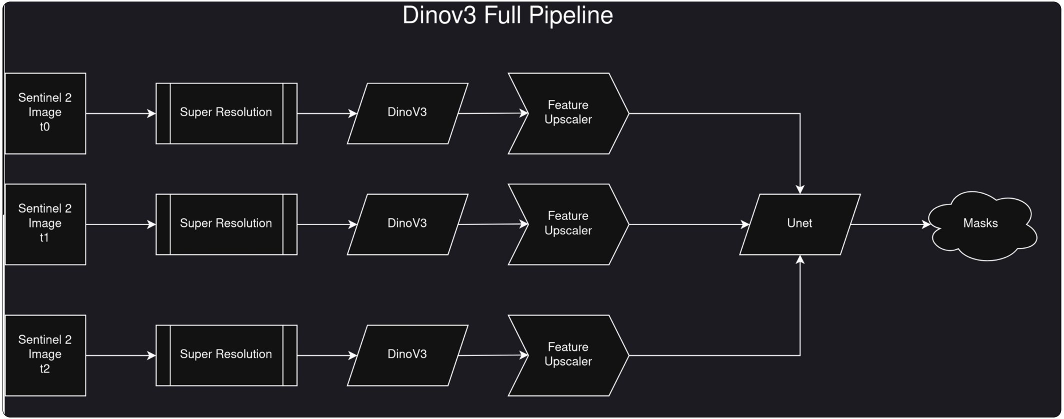
- **Geographic transfer**: Test whether models trained on European data detect linear woody features globally (e.g., Namibia, South America, Australia)
- **Terminology diversity**: Evaluate detection of shelterbelts, windbreaks, tree lines, and living fences—different names for similar landscape structures
- **Research gap**: Limited LWF mapping exists outside Europe; this work tests cross-continental applicability

Pipeline Architecture 1



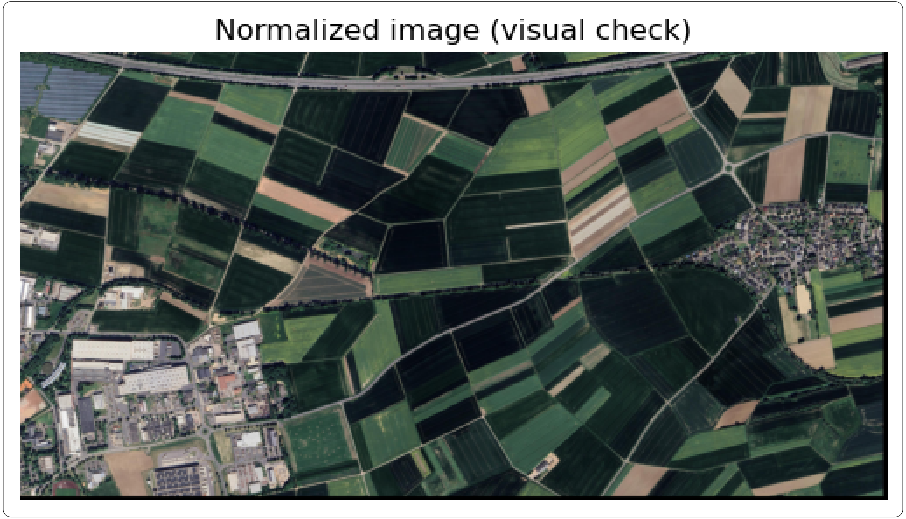
SAM2 based architecture

Pipeline Architecture

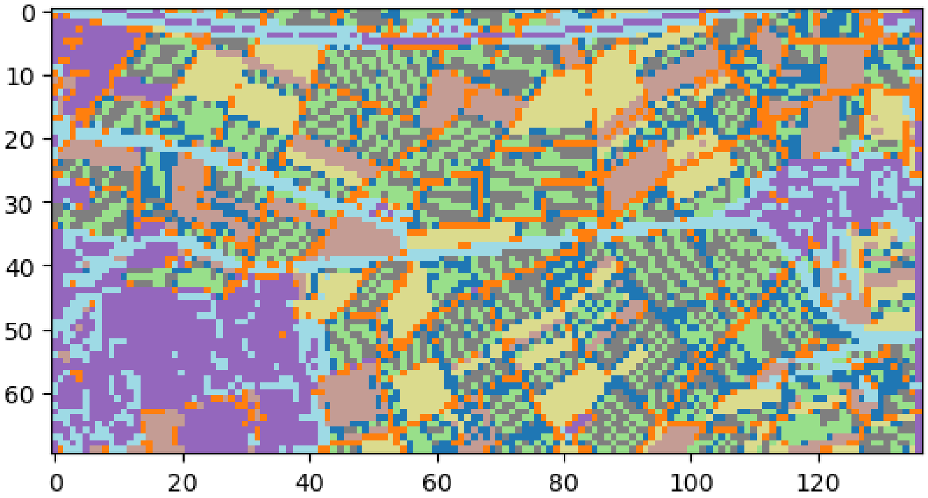


Transformer-based segmentation pipeline for linear woody features

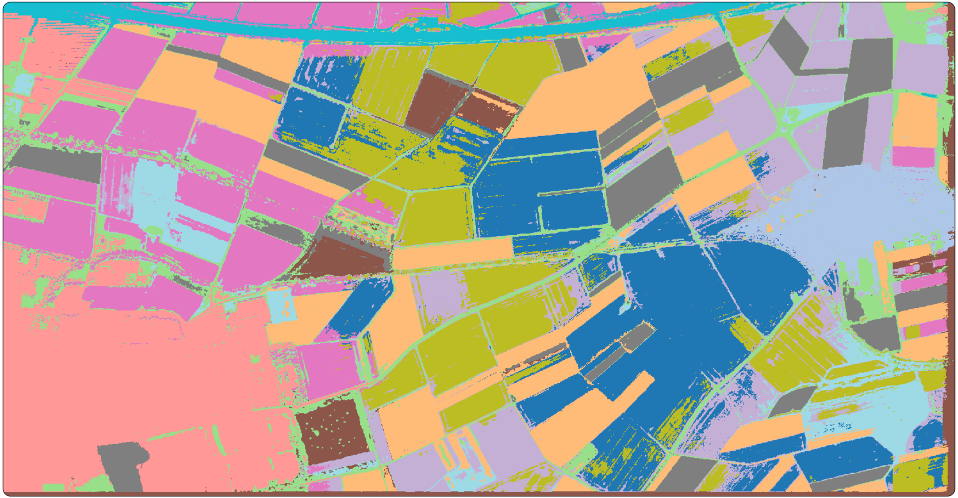
Example Image



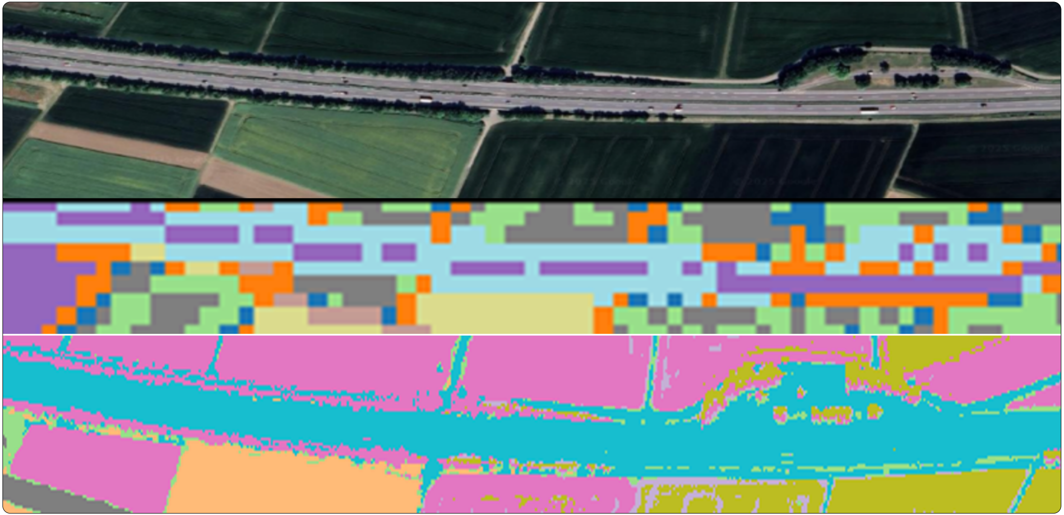
Feature Extraction with KNN



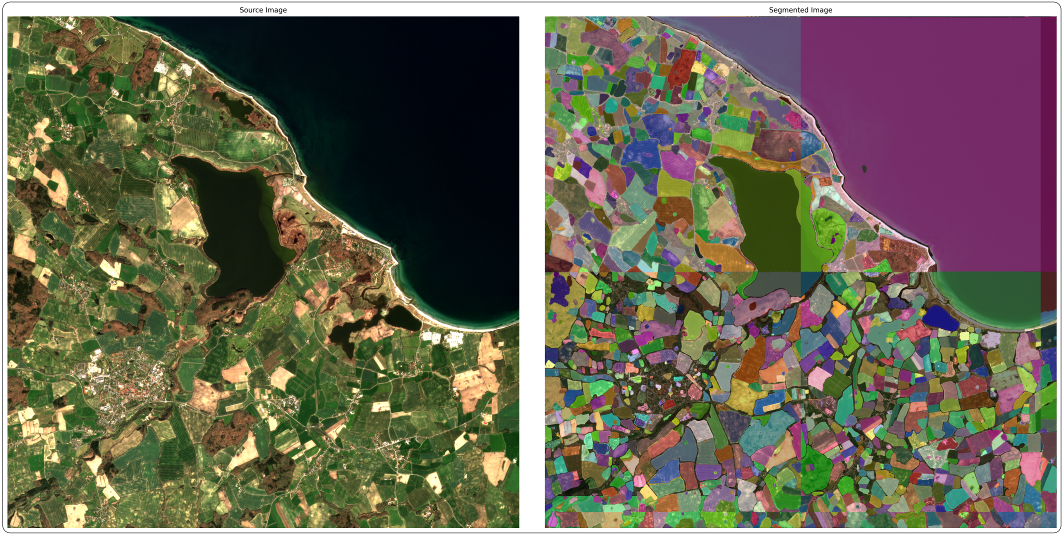
Feature Extraction with KNN Upscaled



Comparison



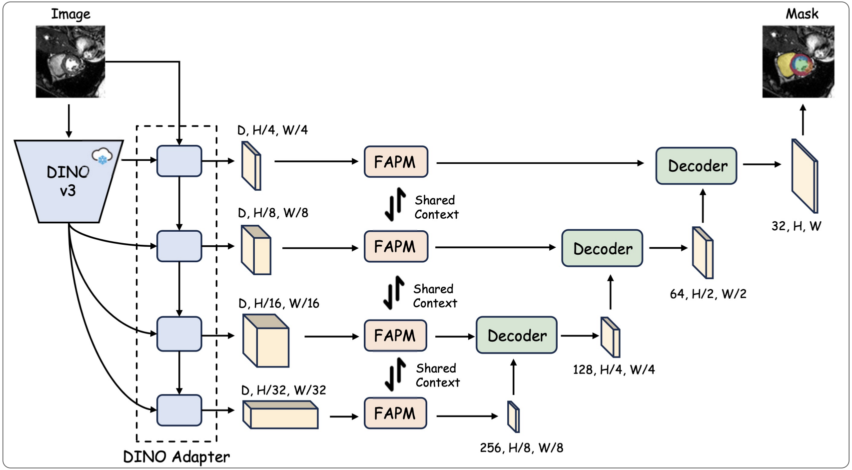
Sam2 Output



Sam2 Status

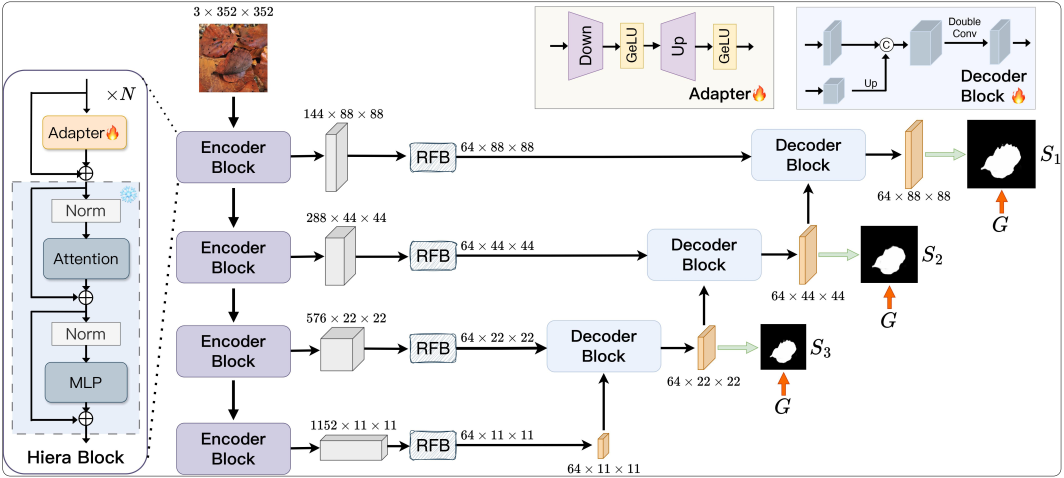
- Export SAM2-generated masks for all image tiles
- Build a classifier on top of SAM2 masks to distinguish linear woody features (LWF)
- Add an auxiliary mask covering areas not segmented by SAM2
- Validate classifier predictions against ground-truth LWF annotations

Dinov3 U-Net



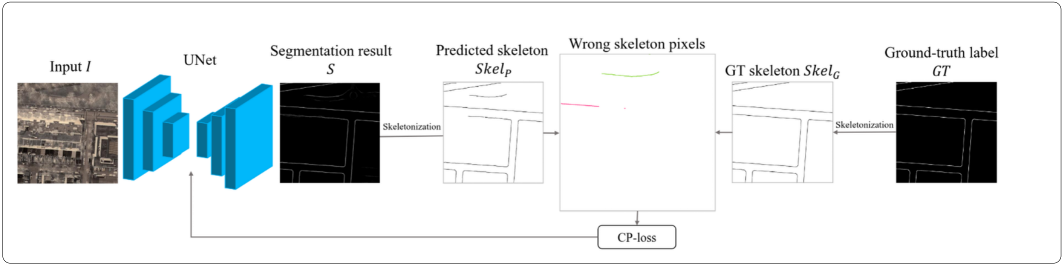
Gao et al., “Dino U-Net: Exploiting High-Fidelity Dense Features from Foundation Models for Medical Image Segmentation”

Sam2 U-Net



Xiong et al., "SAM2-UNet: Segment Anything 2 Makes Strong Encoder for Natural and Medical Image Segmentation"

CP-loss Architecture



Xu et al., “CP-Loss: Connectivity-Preserving Loss for Road Curb Detection in Autonomous Driving with Aerial Images”

CP-loss

Xu et al., "CP-Loss: Connectivity-Preserving Loss for Road Curb Detection in Autonomous Driving with Aerial Images"

Thank You!

Questions for Discussion

- **Multitemporal strategies**: How did you balance the trade-off between using all four monthly PlanetScope basemaps (higher accuracy) versus computational cost at national scale? Would a two-month combination be viable for operational monitoring?
- **Custom loss functions**: You found BCE allowed overcoming sensor artifacts through probability thresholding. Could this strategy extend to transformer-based models with attention mechanisms, or would DICE/GIOU be more suitable for feature-based architectures?

Davide Mattioli | davide.mattioli@uni-goettingen.de

References

Gao, Yifan et al. “Dino U-Net: Exploiting High-Fidelity Dense Features from Foundation Models for Medical Image Segmentation”. In: *arXiv preprint arXiv:2508.20909*. 2025. DOI: `10.48550/arXiv.2508.20909`. URL: `https://arxiv.org/abs/2508.20909`.

Xiong, Xinyu et al. “SAM2-UNet: Segment Anything 2 Makes Strong Encoder for Natural and Medical Image Segmentation”. In: *arXiv preprint arXiv:2408.08909* (2024). Technical Report. URL: `https://arxiv.org/abs/2408.08909`.

Xu, Zhenhua et al. “CP-Loss: Connectivity-Preserving Loss for Road Curb Detection in Autonomous Driving with Aerial Images”. In: *arXiv preprint arXiv:2107.11920* (2021). Accepted at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2021. DOI: `10.48550/arXiv.2107.11920`. URL: `https://arxiv.org/abs/2107.11920`.