# TAS-NIR: A VIS+NIR Dataset for Fine-grained Semantic Segmentation in Unstructured Outdoor Environments

13th Workshop on Planning, Perception and Navigation for Intelligent Vehicles

Peter Mortimer & Hans-Joachim Wünsche

Institute for Autonomous Systems Technology (TAS)
Department of Aerospace Engineering
Bundeswehr University Munich

October 23, 2022

# Introduction to Near-Infrared (NIR) Imaging

A RGB/VIS image of vegetation
and a Near-Infrared (NIR, 770-1100nm)
capture of the same scene (from [1]).

[1] Colouring the Near-Infrared, Clement Fredembach and Sabine Süsstrunk, Color Imaging Conference 2008.

# TAS-NIR: A VIS+NIR Dataset

# TAS-NIR: A VIS+NIR Dataset

Example images from the TAS-NIR Dataset (https://mucar3.de/iros2022-ppniv-tas-nir).

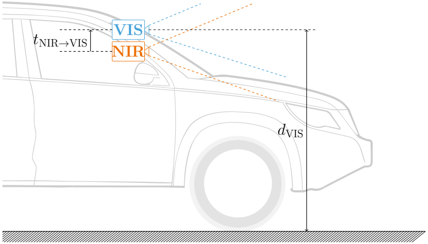
# VIS+NIR Semantic Segmentation Datasets

| Dataset | No. Scenes | No. Classes | Resolution | Scene Type |
|---|---|---|---|---|
| EPFL Semantic Segmentation Dataset [2] | $370^{\dagger}$ | 11 | $1024 \times 768$ px | Outdoor Photography |
| HyKo2 [3] | $78^{\ddagger}$ | 11 | $214 \times 417$ px | Outdoor Driving |
| Freiburg Forest [4] | 366 | 7 | $1024 \times 768$ px | Outdoor Driving |
| **TAS-NIR (ours)** | 209 | 23 | $1200 \times 480$ px | Outdoor Driving |

†: The outdoor scenes from the EPFL Semantic Segmentation Dataset are only compared here.

‡: Only scenes taken with the MQ022HG NIR camera are considered in the HyKo2 dataset.

# Camera Setup



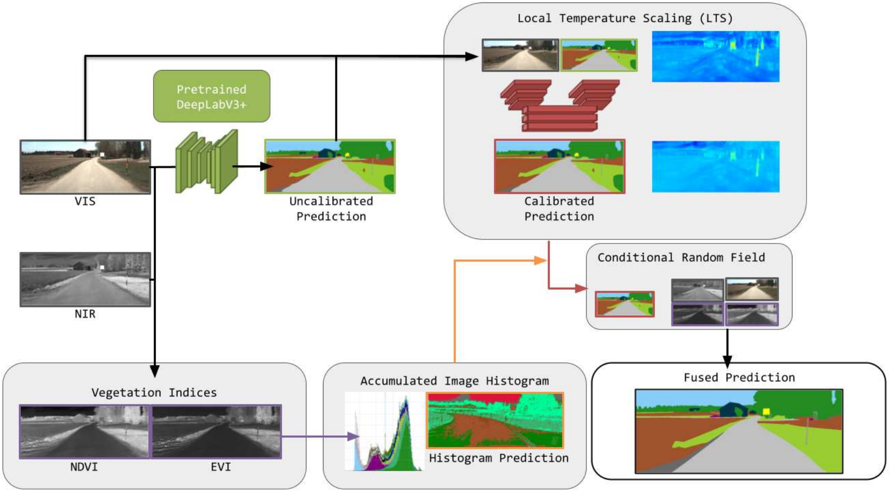

RGB image from a Basler acA2440-20gc.

NIR image from a Basler acA1300-60gmNIR
with a 765nm longpass filter.

# Projecting NIR $\rightarrow$ VIS

By assuming a flat ground plane, we warp
the NIR camera perspective to match that of the color camera.

# VIS+NIR Semantic Segmentation Benchmark




Overview of the VIS+NR Semantic Segmentation Benchmark Components

# Vegetation Indices

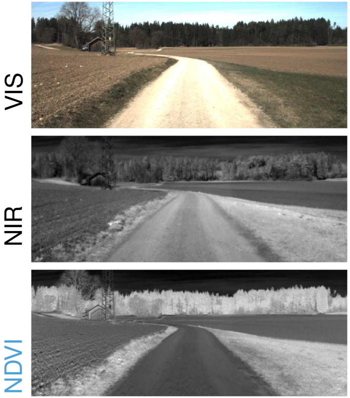
## Normalized Difference Vegetation Index

$$\text{NDVI} = \begin{cases} \dfrac{\text{NIR} - \text{R}_{\text{VIS}}}{\text{NIR} + \text{R}_{\text{VIS}}} & \text{otw.} \\ 0 & \text{if NIR} = \text{R}_{\text{VIS}} = 0 \end{cases}$$

## Enhanced Vegetation Index

$$\text{EVI} = \begin{cases} \dfrac{2 \cdot (\text{NIR} - \text{R}_{\text{VIS}})}{\text{NIR} + C_1 \cdot \text{R}_{\text{VIS}} - C_2 \cdot \text{B}_{\text{VIS}}} & \text{otw.} \\ 0 & \text{if NIR} = \text{R}_{\text{VIS}} = \text{B}_{\text{VIS}} = 0 \end{cases}$$

where $C_1 = 6.0$ and $C_2 = 7.5$



# Vegetation Indices

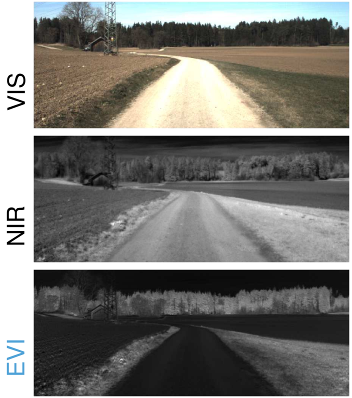
## Normalized Difference Vegetation Index

$$\text{NDVI} = \begin{cases} \dfrac{\text{NIR} - \text{R}_{\text{VIS}}}{\text{NIR} + \text{R}_{\text{VIS}}} & \text{otw.} \\ 0 & \text{if NIR} = \text{R}_{\text{VIS}} = 0 \end{cases}$$

## Enhanced Vegetation Index

$$\text{EVI} = \begin{cases} \dfrac{2 \cdot (\text{NIR} - \text{R}_{\text{VIS}})}{\text{NIR} + C_1 \cdot \text{R}_{\text{VIS}} - C_2 \cdot \text{B}_{\text{VIS}}} & \text{otw.} \\ 0 & \text{if NIR} = \text{R}_{\text{VIS}} = \text{B}_{\text{VIS}} = 0 \end{cases}$$

where $C_1 = 6.0$ and $C_2 = 7.5$



# VIS+NIR Semantic Segmentation Benchmark

VIS

NIR

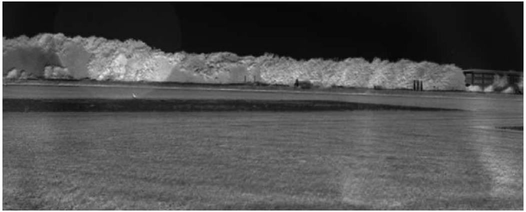
NDVI

EVI

VIS

NIR

NDVI



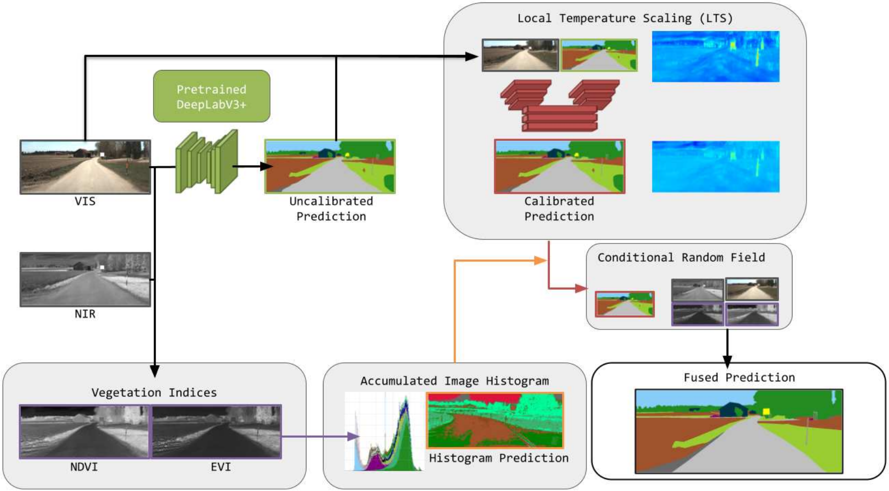
EVI

VIS

NIR

NDVI

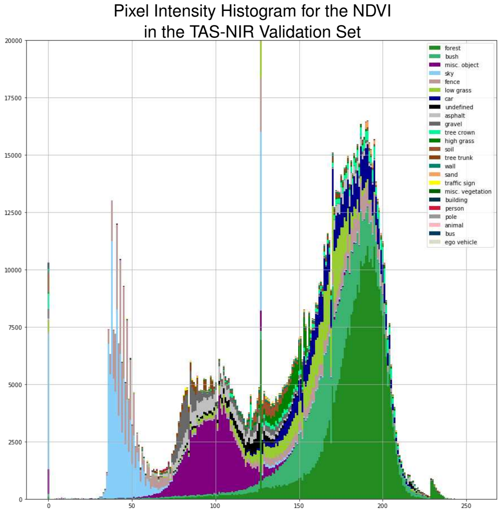
EVI

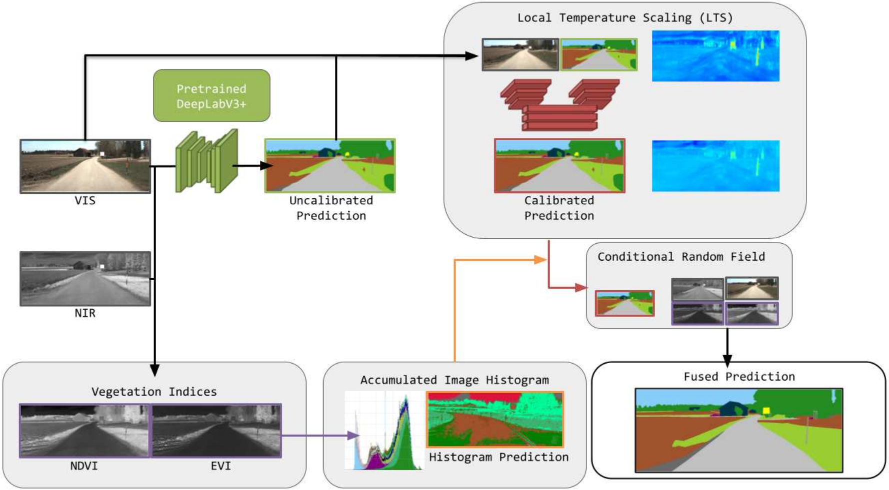



Overview of the VIS+NR Semantic Segmentation Benchmark Components

# Accumulated Image Histogram

For both vegetation indices NDVI and EVI a image histogram is calculated across all images in the validation set. The intensity values are binned and added as normalized weights for each pixel in the late-fusion of the semantic segmentation prediction.

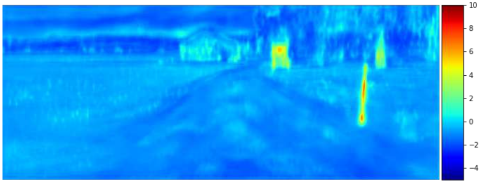




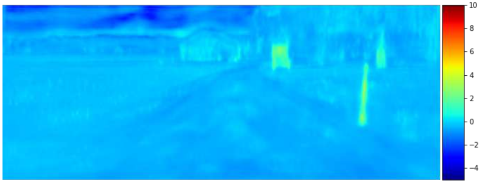

Overview of the VIS+NR Semantic Segmentation Benchmark Components

# Calibrated Pixel Predictions using LTS

Local Temperature Scaling (LTS) [5] uses an additional neural network, that is trained to produce temperature map $T$ to calibrate the outputs of the previously overconfident VIS-only Semantic Segmentation model.

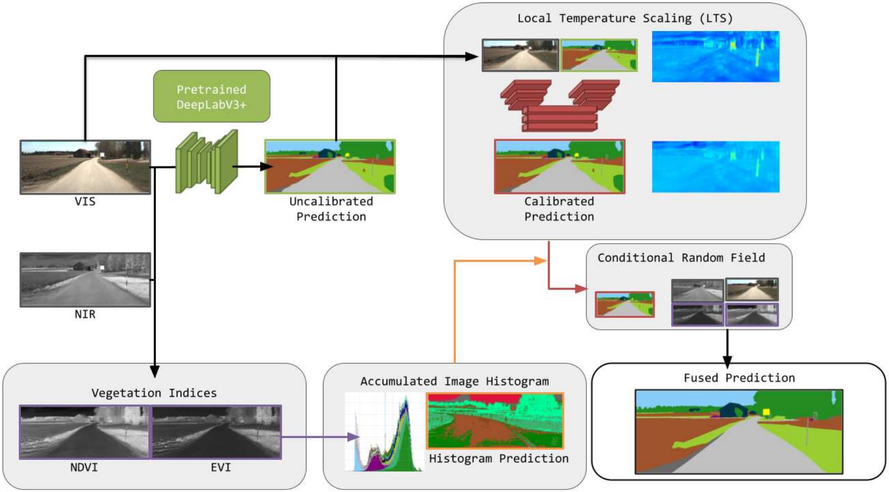



Uncalibrated Activations for the Semantic Class **Pole** (without LTS)




Calibrated Activations for the Semantic Class **Pole** (with LTS)

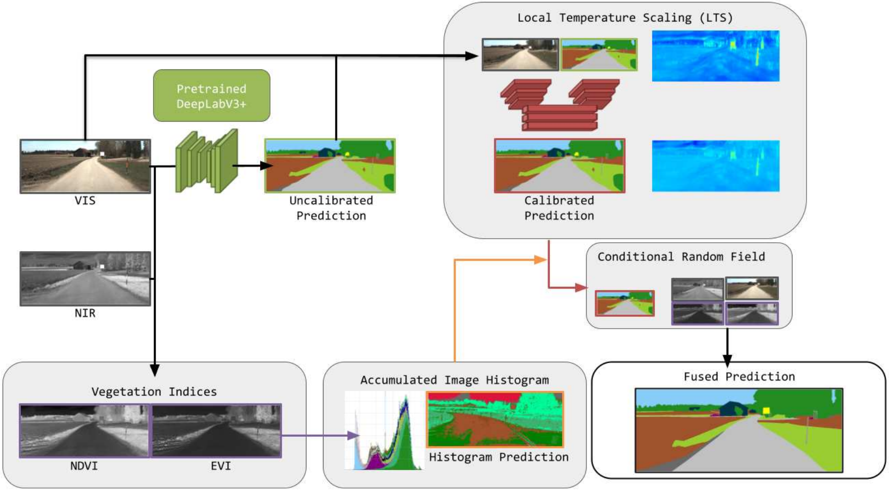
# VIS+NIR Semantic Segmentation Benchmark




Overview of the VIS+NR Semantic Segmentation Benchmark Components

# Conditional Random Fields

Predictions are smoothed using a Conditional Random Field (CRF) [6]. We compare the performance of the final CRF label assignment for different structural input images (VIS, NDVI, EVI).



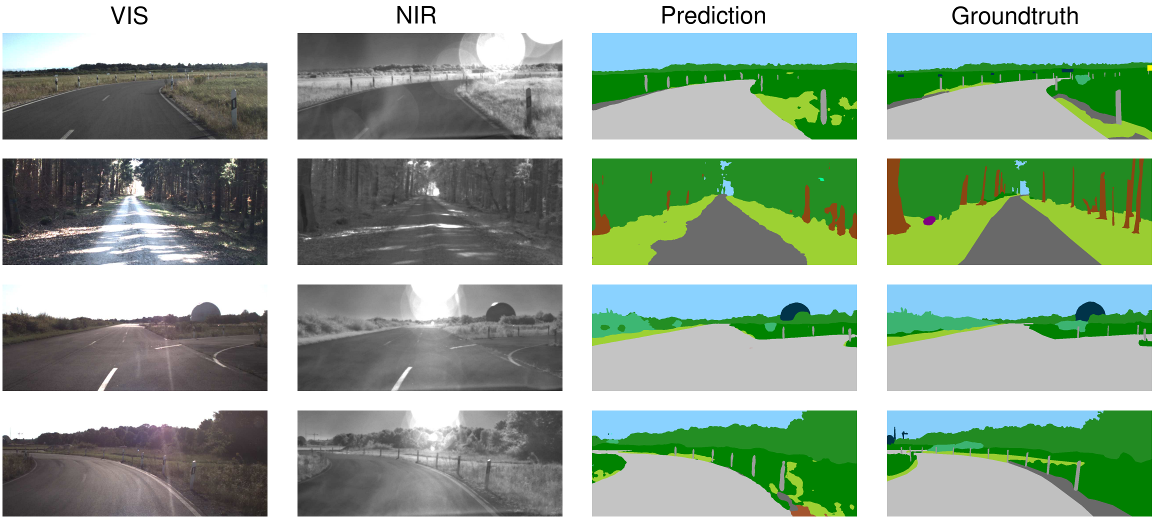

Overview of the VIS+NR Semantic Segmentation Benchmark Components

# Results

| network | LTS | Histogram | CRF | mIoU | gravel | soil | low grass | high grass | bush | tree crown | tree trunk | forest |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DeepLabv3+ | × | × | × | 30.61 | 34.14 | 2.55 | 56.86 | 30.98 | 10.14 | 7.43 | 10.80 | 72.38 |
| | Yes | NDVI | × | 30.56 | 34.17 | 2.50 | 56.75 | 30.80 | 10.26 | 7.42 | 10.74 | 72.39 |
| | Yes | EVI | × | 30.62 | 34.13 | 2.54 | 56.88 | 30.97 | 10.13 | 7.44 | 10.76 | 72.40 |
| | Yes | NDVI | NDVI | 47.97 | 61.02 | 23.11 | **62.48** | 46.93 | 37.78 | 46.34 | 30.65 | **73.12** |
| | Yes | EVI | EVI | 50.16 | 62.08 | 30.40 | 62.17 | **48.43** | 38.59 | 56.98 | **32.64** | 72.88 |
| | Yes | × | VIS | **52.19** | **70.50** | **43.10** | 60.09 | 47.99 | **41.08** | **59.89** | 29.23 | 72.60 |

The Intersection over Union (IoU) in percent for each semantic class of interest and the mean Intersection over Union (mIoU) over all semantic classes of interest. We use local temperature scaling (LTS) to calibrate the output of DeepLabV3+. The addition of a image histogram-based approach alone shows no significant improvement. We observe a significant improvement from the fully-connected conditional random field. In the CRF we denote if and which image type we used as structural input for the CRF.

# VIS+NIR Semantic Segmentation Benchmark



# Conclusions

- The TAS-NIR Dataset is too small for an end to end learning approach.
- A simple fusion of global pixel intensities of the vegetation indices is insufficient.
- CRF Post-processing of the network output greatly improves the overall segmentation performance. The vegetation indices as structural input also lead to significant improvements.
- The CRF output can lead to issues in detecting thin obstacles like poles or fences.

# Future Work


We are currently collecting data using a VIS+NIR prism camera for our **GOOSE Dataset**. It will contain over 10.000 annotated 3D point clouds and VIS+NIR image pairs in unstructured outdoor environments.

Thank you for your attention!

For more info, please visit the TAS-NIR project page:
https://mucar3.de/iros2022-ppniv-tas-nir/

[1] Clément Fredembach and Sabine Süsstrunk. Colouring the Near-Infrared. pages 176–182, 2008.

[2] Neda Salamati, Diane Larlus, Gabriela Csurka, and Sabine Süsstrunk. Semantic Image Segmentation Using Visible and Near-Infrared Channels. In *Proc. European Conf. Comput. Vision (ECCV)*, 2012.

[3] Christian Winkens, Florian Sattler, Veronika Adams, and Dietrich Paulus. HyKo: A Spectral Dataset for Scene Understanding. In *Proc. IEEE Int. Conf. Comput. Vision Workshops (ICCVW)*, pages 254–261, 2017.

[4] Abhinav Valada, Johan Vertens, Ankit Dhall, and Wolfram Burgard. AdapNet: Adaptive Semantic Segmentation in Adverse Environmental Conditions. In *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pages 4644–4651. IEEE, 2017.

[5] Zhipeng Ding, Xu Han, Peirong Liu, and Marc Niethammer. Local Temperature Scaling for Probability Calibration. In *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, pages 6889–6899, 2021.

[6] Philipp Krähenbühl and Vladlen Koltun. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In *Conf. on Neural Inf. Processing*, 2011.