Far Away in the Deep Space: Dense Nearest-Neighbor-Based Out-of-Distribution Detection

Silvio Galesso Max Argus Thomas Brox

galessos, argusm, brox@cs.uni-freiburg.de

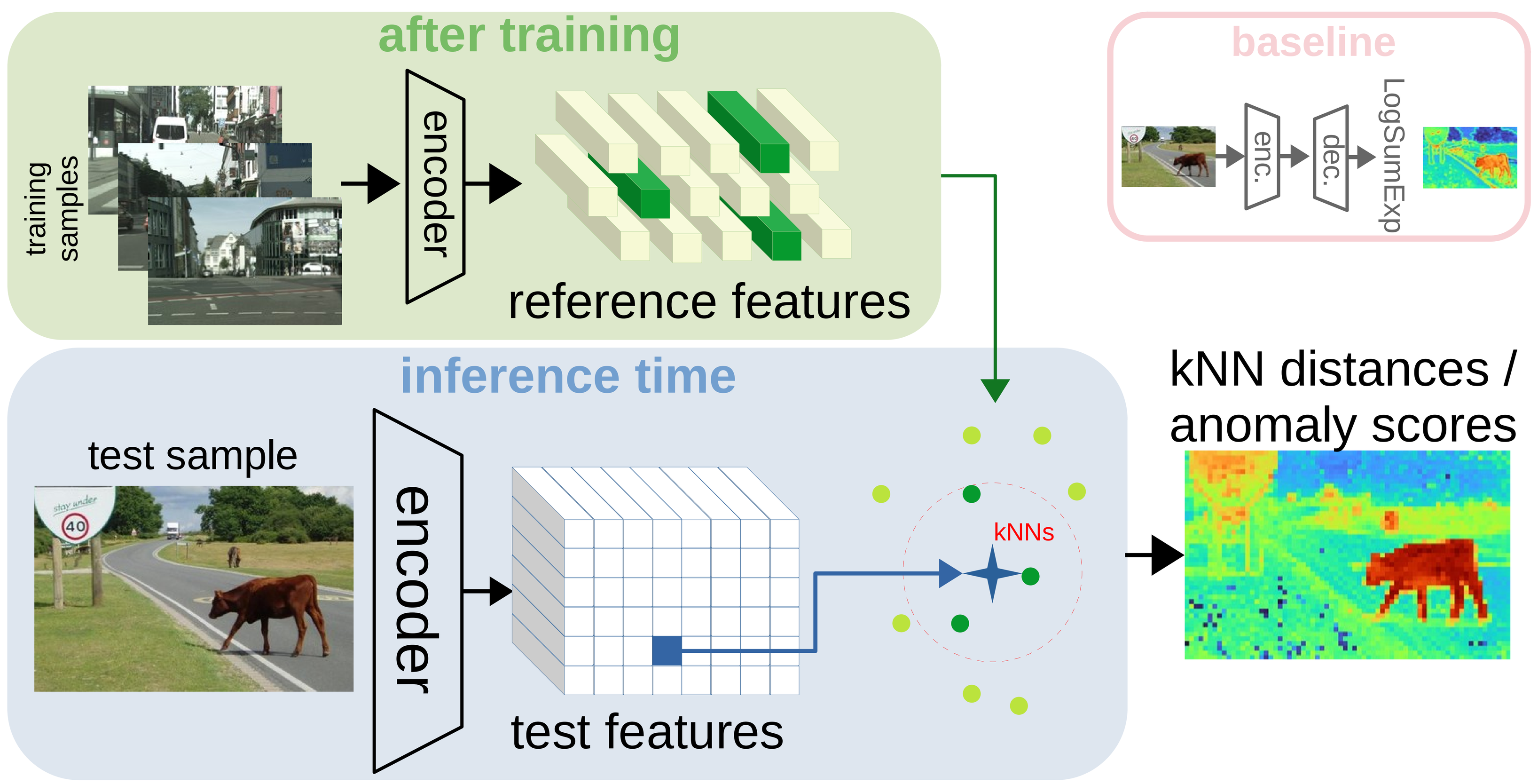
ICCV23

Dense Out-of-Distribution Detection

Goal: detect out-of-distribution objects in semantic segmentation data. Challenges:

- Complex in-distribution data, few classes but diverse appearance.
- Data is not object-centric, entities interact in the scene.
- Need for fast and accurate dense predictions.

Dense OoD Detection with Deep Nearest Neighbors



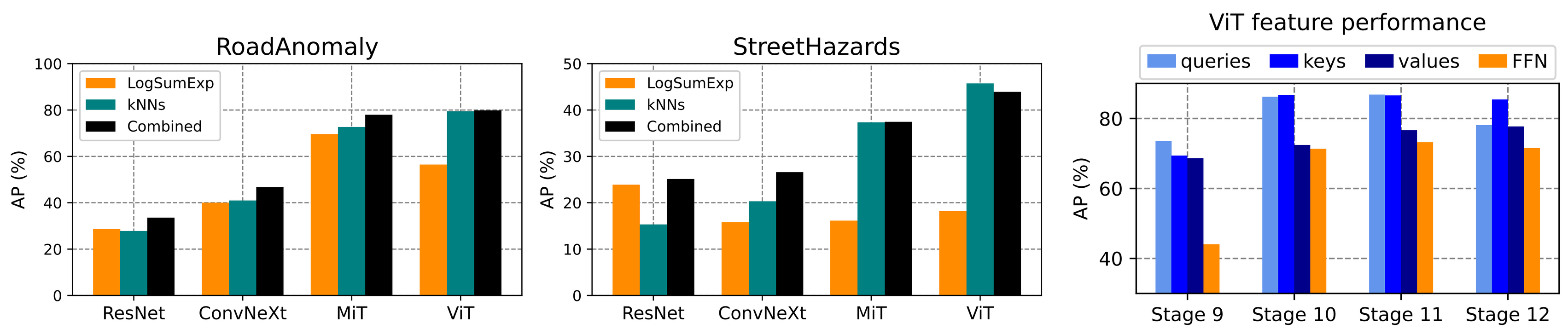
1. Select a trained semantic segmentation model.
2. Use the model to collect a set of N local **reference features** $\mathbf{R} \in \mathbb{R}^{N\times C}$ from in-distribution samples (training dataset). Each feature vector encodes an image patch.
3. At inference time, collect features for the test sample: $\mathbf{T} \in \mathbb{R}^{H\times W\times C}$.
4. For each test feature, compute the expected L_2 distance to its nearest neighbors. The resulting distance map will serve as the **anomaly scores**.

As **baseline** we use model uncertainty, computed as $\mathrm{LogSumExp}(\mathrm{logits})$.

Selecting Good Features

We compare representations from **convolutional** (ResNet, ConvNeXt) and **attention-based** (MiT, ViT) models.

- All networks are trained for supervised semantic segmentation.
- No out-of-distribution data is used.

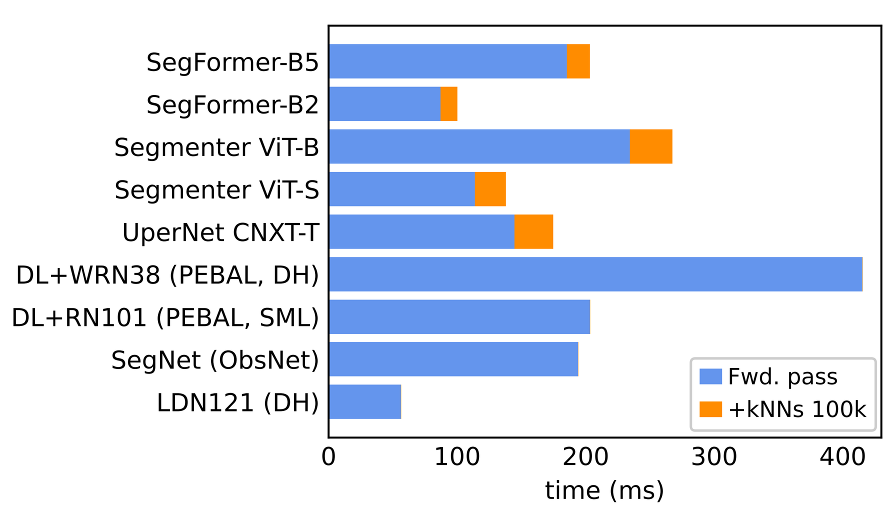


Key findings:

- kNN distances are overall better than model uncertainty.
- **Transformer features greatly outperform CNN ones.**
- Attention features (e.g. keys) perform best.
- kNN anomaly scores can be combined with parametric model uncertainty.

Runtime

- Overall and added runtime depends on the architecture (feature size and resolution).
- Search time for kNNs with 100k reference features (as for other results) is a small fraction.



SOTA Comparison

Method	OE	RoadAnomaly AP	RoadAnomaly FPR_{95}	StreetHazards AP	StreetHazards FPR_{95}
DML		37.0	37.0	14.7	17.3
MOoSe		43.6	32.1	15.2	17.6
PEBAL	✓	62.4	28.3	-	-
DenseHybrid	✓	-	-	30.2	13.0
M2F-EAM		66.7	13.4	-	-
RbA		78.5	11.8	-	-
Ours-Seg.ViT-B		85.6	**9.8**	**46.2**	**14.9**
Ours-SETR-L		**85.9**	13.8	-	-

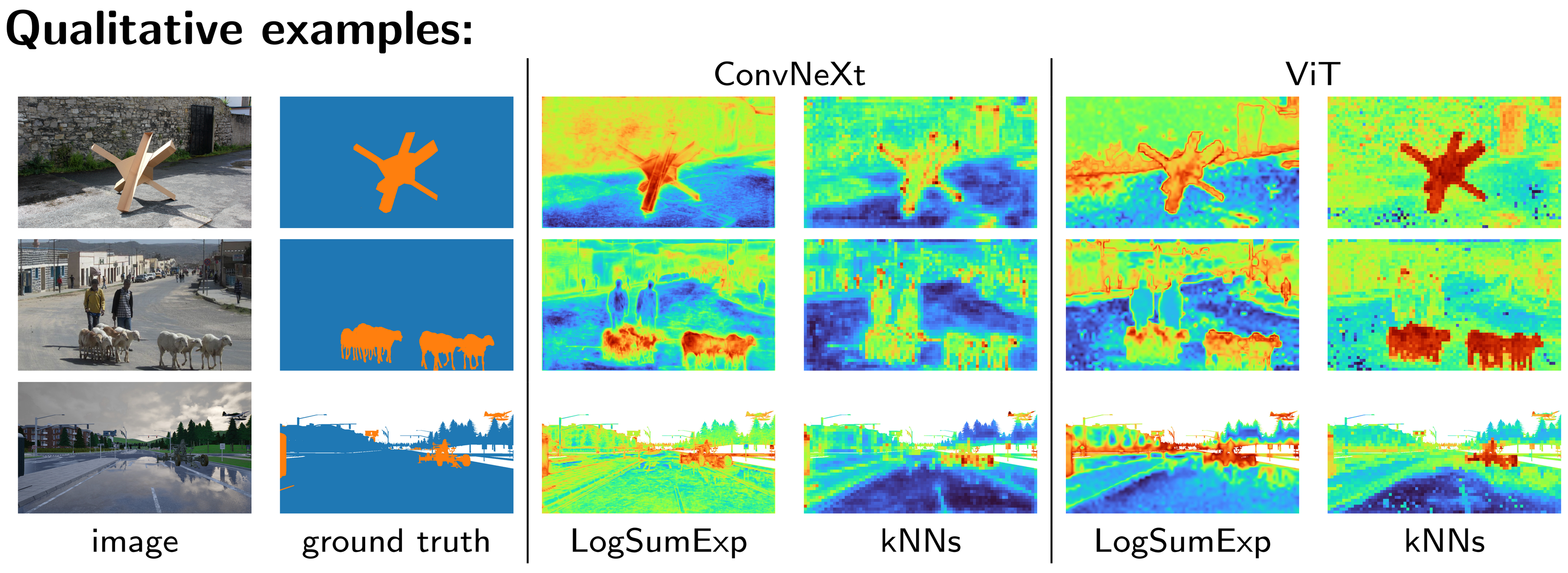
Method	OE	SMIYC-Anomaly AP	SMIYC-Anomaly FPR_{95}
Resynth.		52.3	25.9
PEBAL	✓	49.1	40.8
NFlowJS		56.9	34.7
ObsNet		75.4	26.7
M2F-EAM		76.3	93.9
DenseHybrid	✓	78.0	9.8
RbA		86.1	15.9
Ours-Seg.ViT-B		**88.9**	**11.4**

- On RoadAnomaly, StreetHazards, and SMIYC-Anomaly, kNNs reports best Average Precision (AP), and False Positive Rate at 95% True Positive Rate (FPR_{95}).
- Our approach outperforms methods that make use of out-of-distribution data during training (Outlier Exposure - OE).

Method	OE	FS-Lost&Found AP	FS-Lost&Found FPR_{95}
M2F-EAM		9.4	41.5
NFlowJS		39.4	9.0
DenseHybrid	✓	43.9	6.2
PEBAL	✓	44.2	7.6
FlowEneDet		50.2	**5.2**
GMMSeg		55.6	6.6
Ours-Seg.ViT-B		**62.2**	8.9
Ours-Seg.ViT-B	✓	69.8	7.5

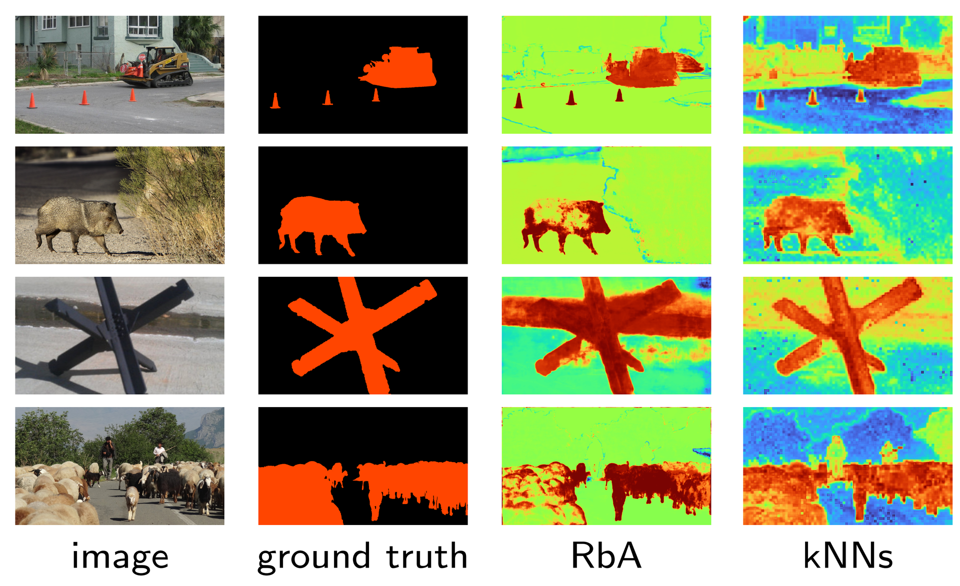
- On Fishyscapes Lost & Found, the approach shows better AP.
- FPR_{95} is high due to the **small anomalous objects** in the data (kNNs work at lower resolution).
- kNNs benefit from outlier exposure.

Qualitative examples:



State-of-the-Art comparison:

- **RbA** [1] uses the model uncertainty of **MaskFormer**.
- Separate mask prediction gives smoother anomaly scores (row 1).
- It also worsens false positives/negatives (rows 2-4).



Analysis

What makes attention features well suited for OoD detection with kNNs? Hypotheses:

- Self-attention is an implicit RBF kernel machine, similar to SVMs [2, 3].
- The multi-head architecture reduces the effective feature dimensionality.

Limitations of our approach:

- kNNs resolution is typically lower than model uncertainty and depends on the architecture. This harms performance on very small objects.
- Performance is tied to a specific class of models (transformers).
- Additional time and memory requirements, model dependent.

References

[1] Nayal et al. Rba: Segmenting unknown regions rejected by all. In *ICCV*, 2023.

[2] Song et al. Implicit kernel attention. *AAAI*, 2021.

[3] Tarzanagh et al. Transformers as support vector machines, 2023.

Acknowledgment: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 417962828