SegDesicNet: Lightweight Semantic Segmentation in Remote Sensing with Geo-Coordinate Embeddings for Domain Adaptation

Sachin Verma, Frank Lindseth, Gabriel Kiss

Department of Computer Science,

Norwegian University of Science and Technology (NTNU), Norway

Domain & Task

- A domain denoted by **D** encompasses both the data **X** and its corresponding distribution **P(x).**

$$\mathbf{D} = \{\mathbf{X}, \mathbf{P}(\mathbf{x})\}, \forall \mathbf{x} \in \mathbf{X} \qquad (1)$$

- A task **T** characterized by the label space **Y** and prediction function **f(x)** can be interpreted as the posterior probability **p(y|x)**

$$\mathbf{T} = \{\mathbf{Y}, \mathbf{P}(\mathbf{y}|\mathbf{x})\}, \forall \mathbf{y} \in \mathbf{Y} \qquad (2)$$

Domain Adaptation

- In domain adaptation (DA) setup, we have :

  Source Domain $(\mathbf{D_S}) = \{\mathbf{X_S}, \mathbf{P_S}(\mathbf{x_S})\}, \forall \mathbf{x_s} \in \mathbf{X_S}$

  Target Domain $(\mathbf{D_T}) = \{\mathbf{X_T}, \mathbf{P_T}(\mathbf{x_T})\}, \forall \mathbf{x_T} \in \mathbf{X_T}$

  and $\mathbf{T_S}$ **,** $\mathbf{T_T}$ being their corresponding tasks

  such that ,

  1. $\mathbf{X_S} \cap \mathbf{X_T} = \emptyset$
  2. $\mathbf{T_S} = \mathbf{T_T}$
  3. $\mathbf{P}(\mathbf{x_S}) \neq \mathbf{P}(\mathbf{x_T})$, where $\mathbf{x_S} \in \mathbf{X_T}$ and $\mathbf{x_T} \in \mathbf{X_T}$ exhibit different distribution

- **DA Objective**:

  To train a model that effectively transfers knowledge from $\mathbf{D_S}$ to $\mathbf{D_T}$, despite these distributional differences

Unsupervised Domain Adaptation (UDA) operates under the assumption that there is no knowledge of the label space in the target domain ($\mathbf{Y_T}$).

Recent Work Utilizing Geographical Metadata

1. Geoinformation , at the very beginning, is used in classification task concatenated into high-level image features.
2. Inspired by neuroscience research, Space2Vec encoded absolute positions and spatial relationships of places.
3. * To our knowledge,only one specific work has addressed UDA with geo-coordinates for remote sensing, in which its DA module regresses the positional embedding of the geo-coordinates.

4. Kevin Tang, Manohar Paluri, Li Fei-Fei, Rob Fergus, and Lubomir Bourdev. Improving image classification with location context. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), December 2015.
5. M. Rußwurm, K. Klemmer, E. Rolf, R. Zbinden, and D. Tuia. Geographic Location Encoding with Spherical Harmonics and Sinusoidal Representation Networks. arXiv preprint arXiv:2310.06743, Oct 10 2023.
6. V. Marsocci, N. Gonthier, A. Garioud, S. Scardapane, and C. Mallet. Geomultitasknet: remote sensing unsupervised domain adaptation using geographical coordinates. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 2075–2085, Los Alamitos, CA, USA, jun 2023. IEEE Computer Society.

Our Contribution

1. We introduced an unsupervised DA technique for HRSI semantic segmentation aiming to mimic human comprehension effectively
2. Our proposed approach outperforms state-of-the-art methods on the widely recognized FLAIR #1 [4] data subset (benchmarked by [3], and our own custom splits). We also validated our approach using the ISPRS Potsdam[5] dataset
3. Our network has ~27% lower parameter count compared to SOTA, enabling more efficient and lightweight models without compromising performance

4. Flair n. 1: Semantic segmentation and domain adaptation. https://codalab.lisn.upsaclay.fr/competitions/8769, 2022.
5. Ahram Song and Yongil Kim. Semantic segmentation of remote-sensing imagery using heterogeneous big data: International society for photogrammetry and remote sensing potsdam and cityscape datasets. ISPRS International Journal of Geo-Information, 9(10):601, 2020.

Proposed Model Architecture

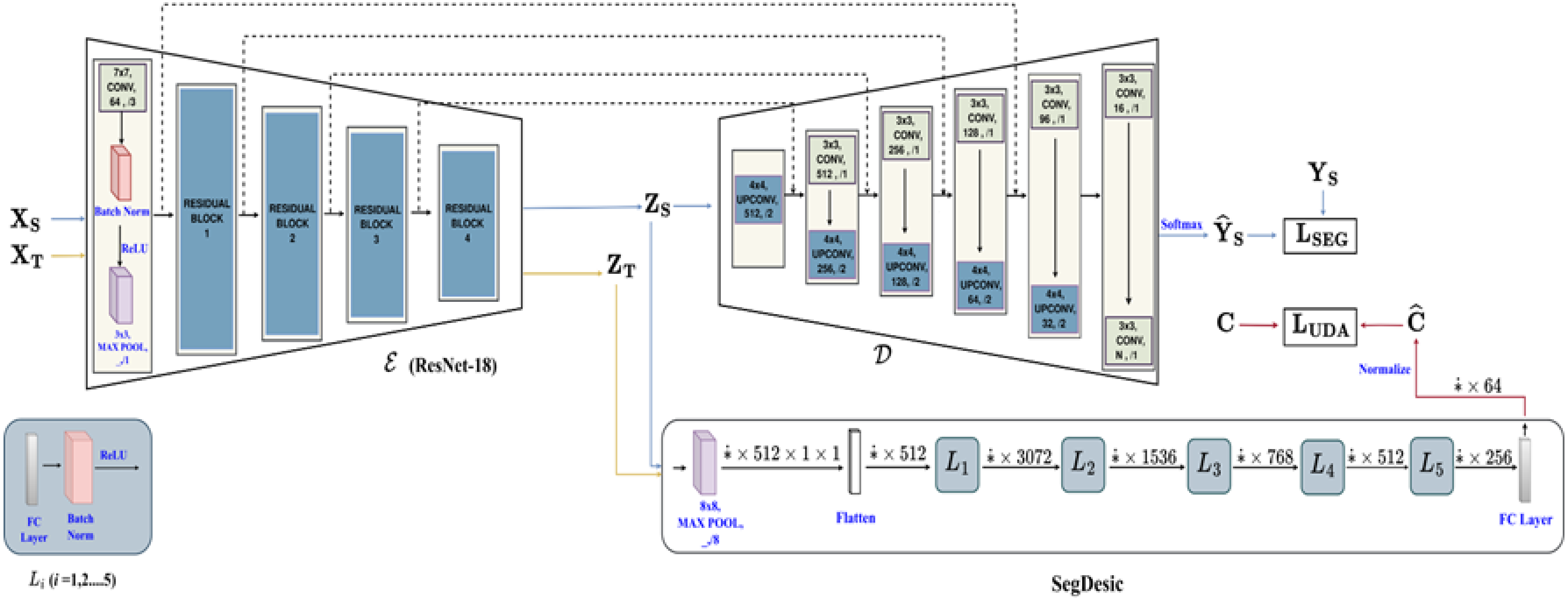


Figure 1 : UNet (ResNet-18 as Encoder $\mathcal{E}$, custom decoder $\mathcal{D}$), with SegDesic module.

Methodology : Data Representation

1. Coordinate centering to ensures that the median values of the coordinates are (0, 0)

$$C'_{\mathrm{lon}} = C_{\mathrm{lon}} - 489353.59\ \mathrm{m}$$
$$C'_{\mathrm{lat}} = C_{\mathrm{lat}} - 6587552.20\ \mathrm{m}$$

2. Transforming coordinates to EPSG:4326 from EPSG:2154, coordinate system.

$$\lambda, \phi = \mathrm{f}_{2154 \to 4326}(C'_{lon}, C'_{lat})$$

3. Multi-scale positional encoding is applied as single scale representation alone is insufficient for periodic function[6].

$$\mathrm{GRID}(\lambda, \phi) = \bigcup_{s=0}^{S-1} \left[\sin(\frac{\lambda}{\alpha_s}), \cos(\frac{\lambda}{\alpha_s}), \sin(\frac{\phi}{\alpha_s}), \cos(\frac{\phi}{\alpha_s}) \right]$$

where, $\alpha_s = \lambda_{min} \cdot g^{s/(S-1)}$ is a scaling factor that induces increasingly S high frequencies through scales s from 0 to S − 1, indicates concatenation and $\lambda_{min}, \lambda_{max}$ are the minimum and maximum grid scale with $g = \frac{\lambda_{max}}{\lambda_{min}}$.

4. By mapping the encoded vector onto the unit sphere, we preserve the inherent spherical nature of Earth's observations (EO)

$$\mathbf{C} = \frac{\mathrm{GRID}(\lambda, \phi)}{\|\mathrm{GRID}(\lambda, \phi)\|_1}$$

6. Gengchen Mai, Krzysztof Janowicz, Bo Yan, Rui Zhu,Ling Cai, and Ni Lao. Multi-scale Representation Learning for Spatial Feature Distributions using Grid Cells. arXiv preprint arXiv:2003.00824, 2020.

Loss Function

$$L = L_{seg} + \alpha \times (L_{UDA}^{S} + L_{UDA}^{T})$$

where,

$$L_{seg} = \frac{-\sum_{h=1}^{H}\sum_{w=1}^{W} y_S^{(h,w)} \cdot \log\left(h_\theta\left(x_S^{(h,w)}\right)\right)}{H \times W}$$

$$L_{\mathrm{UDA}} = 1 - \frac{\mathbf{C} \cdot \widehat{\mathbf{C}}}{\|\mathbf{C}\|_2 \cdot \|\widehat{\mathbf{C}}\|_2}$$

$$\alpha = 0.5$$

h_θ is the U-Net model with weights θ, and **H** and **W** are height and width of the images

Data Set

Data Set	SPLIT	Source Domain	Target Domain	# Train Image	# Test Image
FLAIR #1 [4]	SPLIT1 [3]	D06, D08, D13, D17, D23, D29, D33, D58, D67, D74	D64, D68, D71	16050	5350
	SPLIT2	D07, D09, D013, D30, D31, D32, D34, D46, D66, D81	D26, D64, D83	15325	5625
	SPLIT3	D07, D09, D13, D30, D31, D32, D34, D46, D66, D81	D22, D68, D71	16272	5300
ISPRS Potsdam[5]	SPLIT4	2 11, 2 12, 3 13, 3 14, 5 11, 6 15, 6 9, 7 8, 7 9	4 11, 6 13, 4 15	1296	432

Table 1 : Custom Source and Target Domain Splits

Data Set Contd..

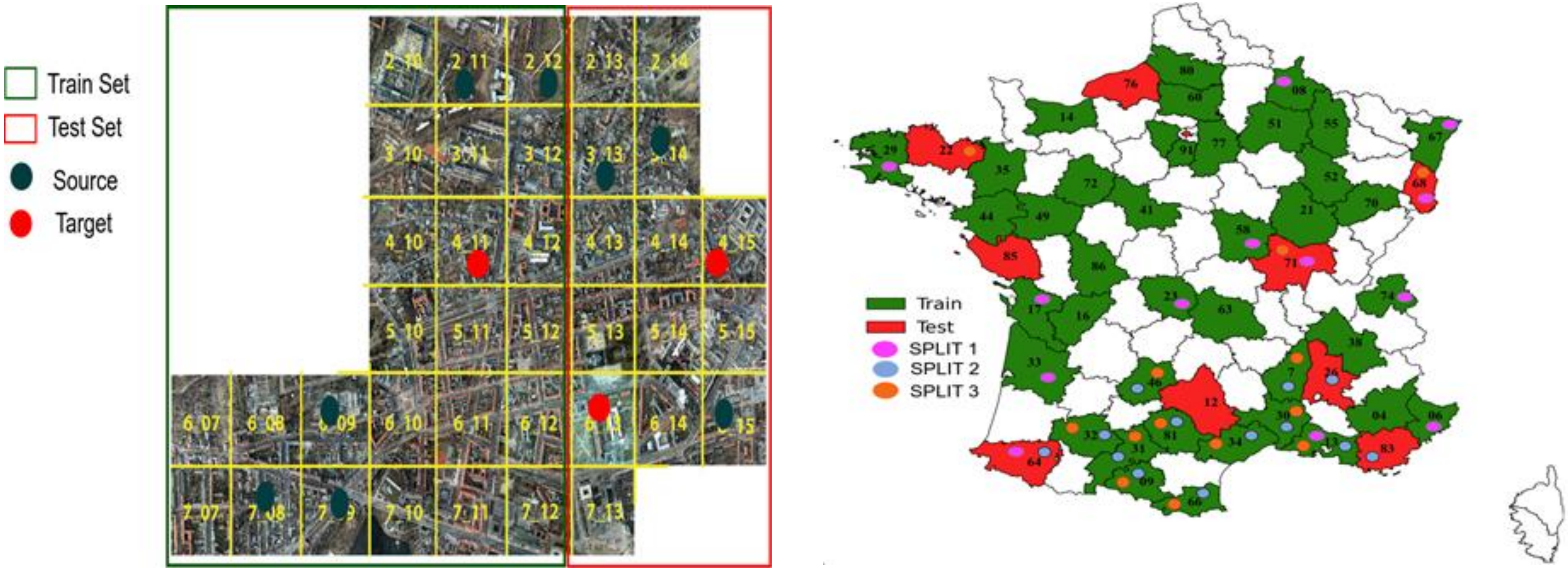


Figure 2 : Shown are the training and test set splits of the ISPRS Potsdam dataset (on the left) and FLAIR #1 dataset (right), with custom splits of the source and target domains indicated by colored dots.

Results : FLAIR #1

Method	Building	Pervious	Impervious	Bare Soil	Water	Coniferous	Deciduous	Brushwood	Vine	Grassland	Crop	Plowed Land	mIoU
CLAN	6.24	13.66	17.09	1.50	12.99	1.29	27.22	3.36	30.69	27.34	7.69	18.42	13.96
AdaptSegNet	39.98	20.75	40.23	20.36	15.25	4.93	35.37	10.99	34.51	**42.69**	11.06	23.47	24.97
ADVENT	35.79	24.38	48.82	6.85	31.98	0.00	51.65	11.79	33.33	25.76	11.46	24.29	25.51
IAST	55.67	36.43	53.71	26.95	53.33	0.00	50.67	11.56	43.24	26.28	26.31	44.27	35.70
DAFormer	67.09	45.56	61.99	55.35	65.12	8.91	54.39	**20.31**	64.39	38.79	23.74	41.83	45.62
UDA_for_RS	66.30	**48.05**	62.36	**59.28**	61.24	9.22	60.02	16.52	57.74	40.12	30.32	54.17	47.11
GeoMultiTaskNet	67.53	40.86	63.89	55.31	67.02	**13.85**	60.97	14.08	53.09	40.33	35.02	**54.79**	47.23
Ours	**67.65**	45.18	**64.13**	55.92	**71.37**	11.62	**62.65**	16.65	**74.50**	40.90	**37.83**	53.06	**50.12**

(a) SPLIT1

Method	Building	Pervious	Impervious	Bare Soil	Water	Coniferous	Deciduous	Brushwood	Vine	Grassland	Crop	Plowed Land	mIoU
CLAN	24.27	21.27	41.19	8.89	26.57	4.22	28.26	24.70	54.84	31.00	18.80	1.28	23.77
AdaptSegNet	60.07	24.06	54.22	21.29	24.31	11.90	39.02	26.31	47.25	31.56	19.95	10.56	30.88
ADVENT	56.21	29.82	55.47	27.52	44.73	15.83	41.38	22.77	55.18	41.47	12.52	30.26	36.10
IAST	68.84	**42.05**	61.19	**61.86**	**79.38**	0.00	46.27	**35.69**	69.40	35.38	34.06	**36.71**	47.57
ours	**69.60**	38.63	**65.31**	46.16	73.54	**25.52**	**46.89**	35.19	**74.00**	**44.87**	**34.65**	26.67	48.42

(b) SPLIT2

Method	Building	Pervious	Impervious	Bare Soil	Water	Coniferous	Deciduous	Brushwood	Vine	Grassland	Crop	Plowed Land	mIoU
CLAN	28.86	18.38	37.78	5.31	17.92	7.78	45.82	10.69	66.38	40.61	14.67	7.81	25.17
ADVENT	49.15	30.39	50.13	18.23	43.70	15.75	59.95	**15.37**	49.89	**53.45**	9.18	22.87	34.84
AdaptSegNet	57.95	27.68	50.17	11.78	35.25	17.96	63.73	14.28	55.64	40.86	28.26	17.81	35.11
IAST	62.31	**52.13**	**64.67**	15.04	**75.28**	0.00	61.19	12.70	69.00	30.85	**39.85**	49.28	44.36
ours	**63.33**	46.61	60.32	**18.52**	51.09	**31.25**	**66.32**	14.12	**74.58**	47.30	34.25	**52.94**	**46.72**

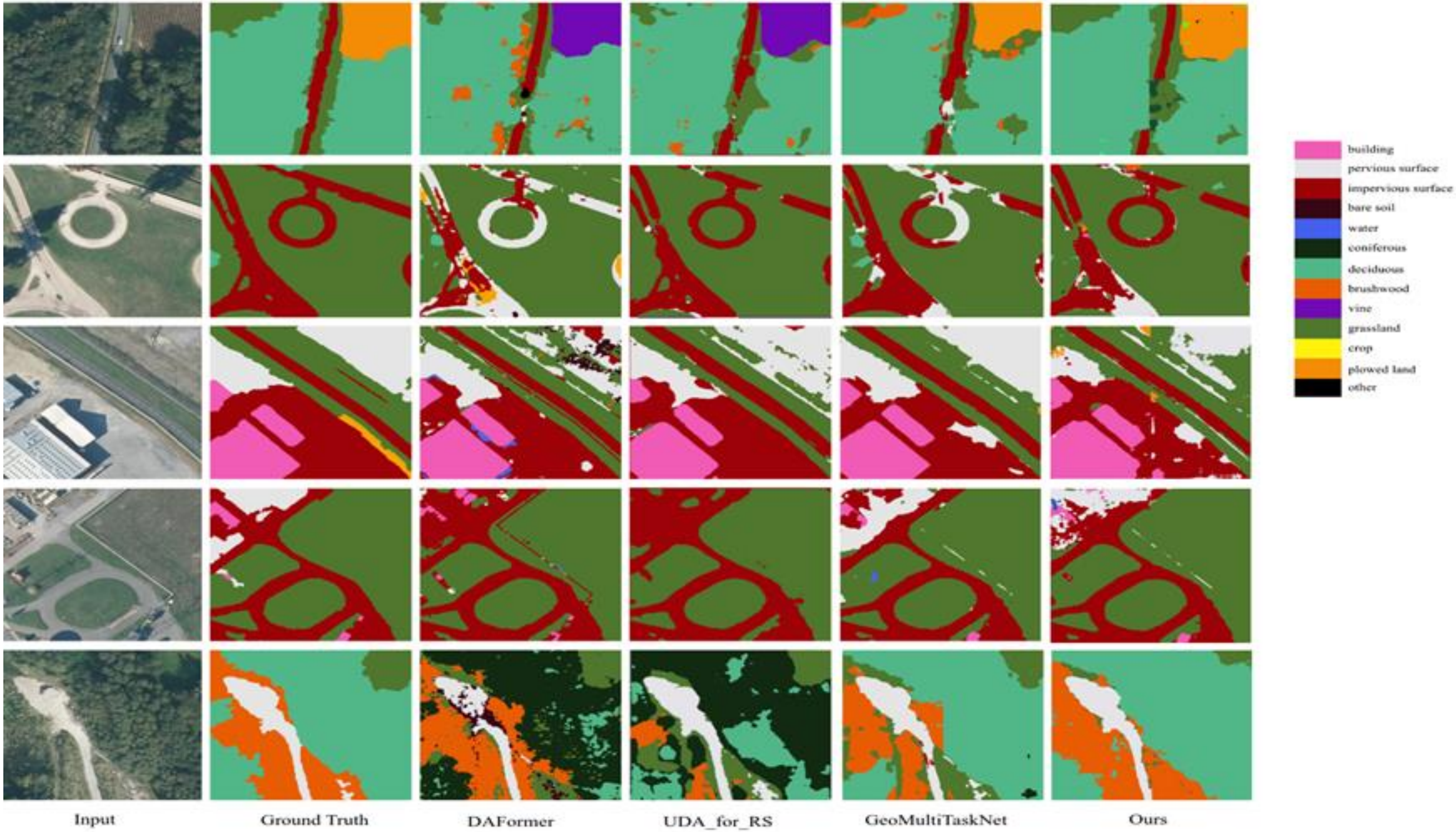
(c) SPLIT3

Table 2 : Class-wise comparison of IoU(%) for for different methods across various data domains

Results : ISPRS Potsdam

Method	Impervious	Building	Low Vege	Tree	Car	mIoU
IAST	68.42	75.86	64.77	64.46	43.72	63.46
CLAN	64.67	68.71	58.72	47.28	42.63	56.40
ADVENT	72.18	78.33	69.65	67.65	68.69	71.30
AdaptSegnet	73.40	78.89	68.14	67.69	70.65	71.75
Ours	**74.88**	**81.17**	**70.49**	**69.47**	**72.26**	**73.65**

Table 3 : Class-wise comparison of IoU(%) for for different methods across various data domains



building
pervious surface
impervious surface
bare soil
water
coniferous
deciduous
brushwood
vine
grassland
crop
plowed land
other
Input
Ground Truth
DAFormer
UDA_for_RS
GeoMultiTaskNet
Ours

Conclusion

Our proposed SegDesicNet module

- regresses the GRID positional encoding of the geo-coordinates
- we projected these feature over the unit sphere to model spherical nature of Earth's surface.
- seeks to reduce the modeling disparity between artificial neural networks and human comprehension of the physical world.

Acknowledgement

This research received funding from the PERSEUS project, a European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101034240. This paper is supported by the MoST (MobilitetsLab Stor-Trondheim) project (https://www.mobilitetslabstortrondheim.no/en)

Thank you.