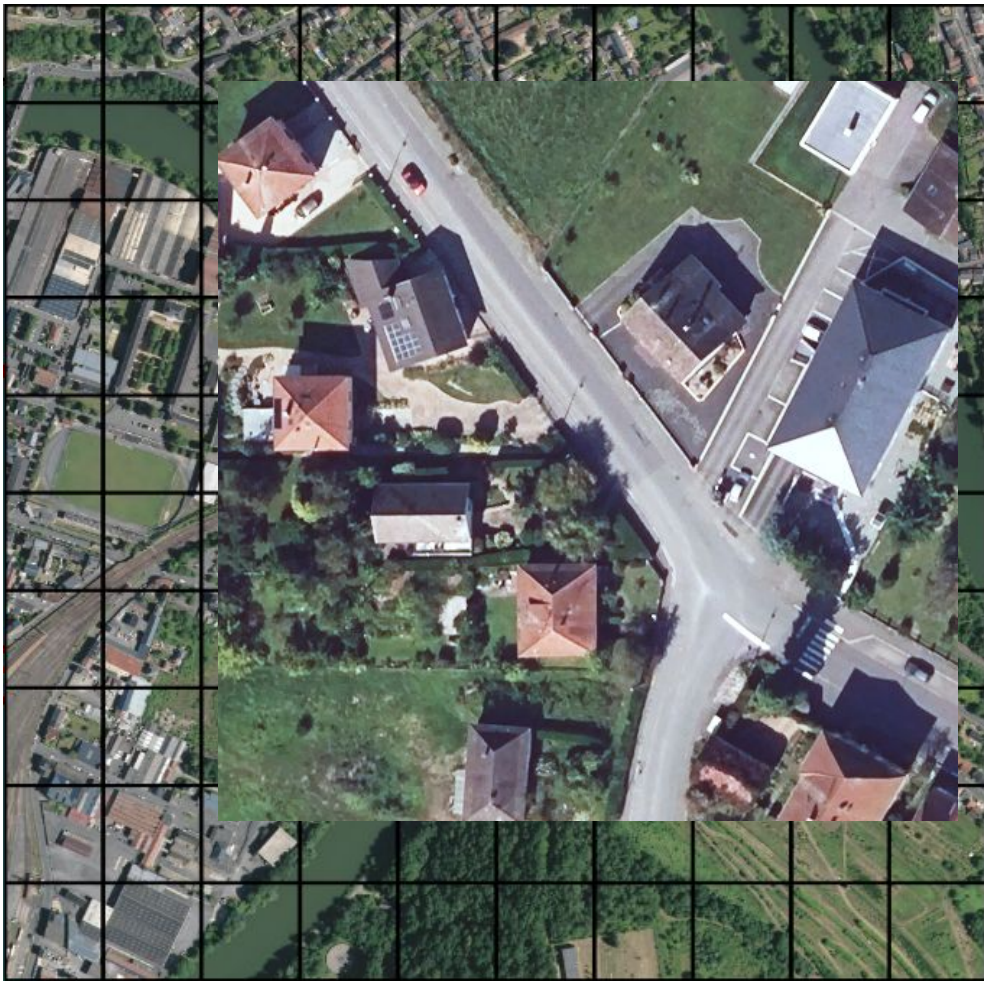


Unsupervised Multi-Domains Adaptation for Semantic Segmentation of Very High Resolution Aerial Images

Candidate:
Valerio Marsocci

13/01/2023
Institut Géographique National

Supervisors: **prof. Clement Mallet**
Nicolas Gonthier, Ph.D.
Anatol Garioud, Ph.D.
prof. Simone Scardapane



Think of a **climate mitigation**
project

What do we need?

Land cover map to avoid land consumption



What if we do not have it?

What if there is a domain shift?



Ground truth



Prediction

- building
- pervious surface
- impervious surface
- bare soil
- water
- coniferous
- deciduous
- brushwood
- vine
- grassland
- crop
- plowed land
- other

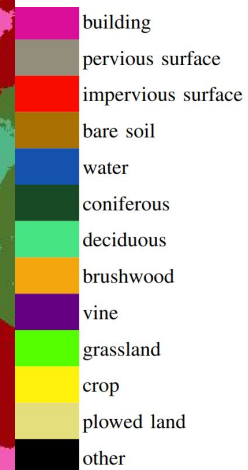
Unsupervised Domain Adaptation can help us



Ground truth



Prediction



What we will see

1. Unsupervised Domain Adaptation
2. FLAIR Dataset
3. Widespread methodologies
4. Our methodology
5. Conclusions

In this space you will find the
TAKEAWAYS!



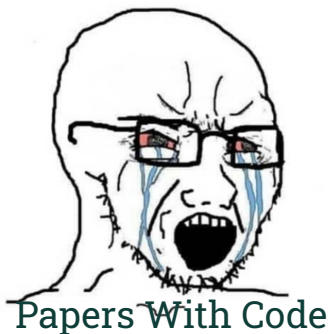
Unsupervised Domain Adaptation (UDA)

Almost what we have already seen



Definition

“Unsupervised Domain Adaptation is a learning framework to transfer knowledge learned from source domains with a large number of annotated training examples to target domains with unlabeled data only”



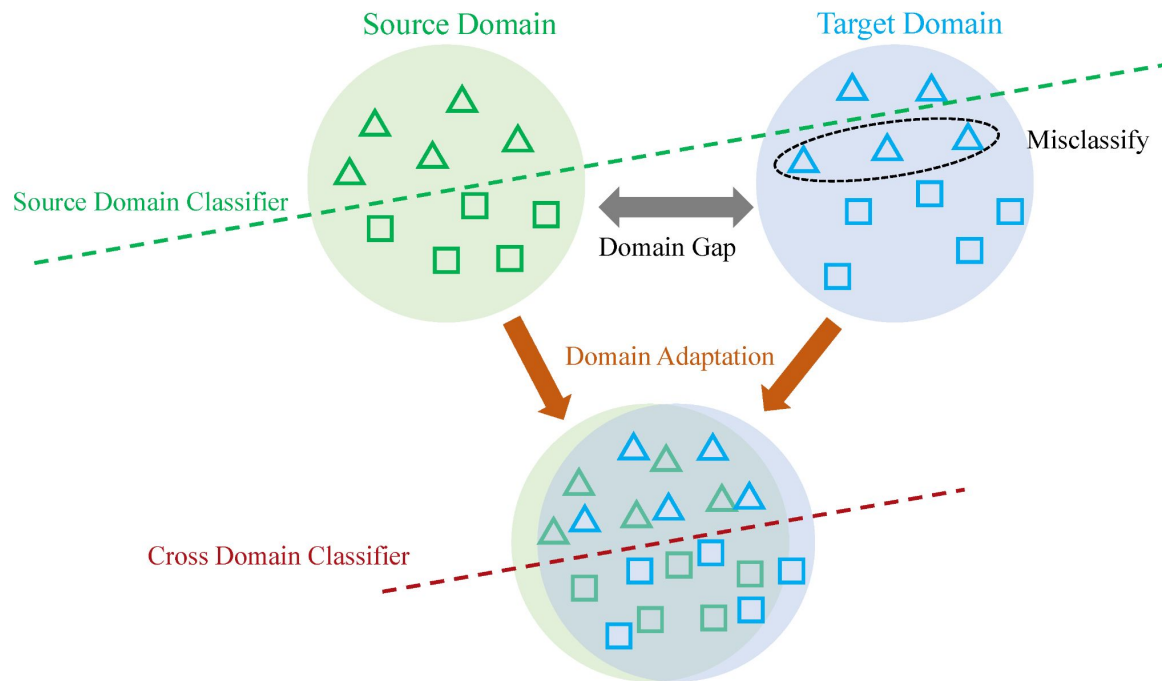
Papers With Code

UDA.



Me

What is UDA?



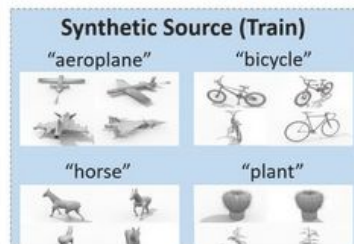
Datasets



Office-Home

classification

synthetic to real



VisDA-2017

To date no dataset for UDA for EO



Foggy Cityscapes



Cityscapes

semantic
segmentation



GTAS

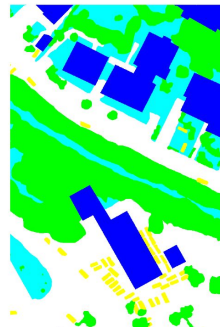
Open questions

Still a lot to do in UDA (especially for RS)

- **few** data (and few methods) for “**many domains**” UDA for semantic segmentation
- **few** proper data (and few methods) for **EO UDA**



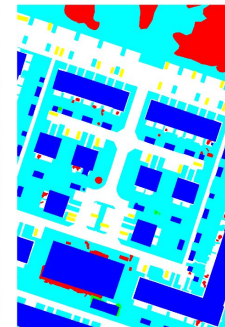
(a) Vaihingen image



(b) Ground truth



(c) Potsdam image



(d) Ground truth

FLAIR Dataset

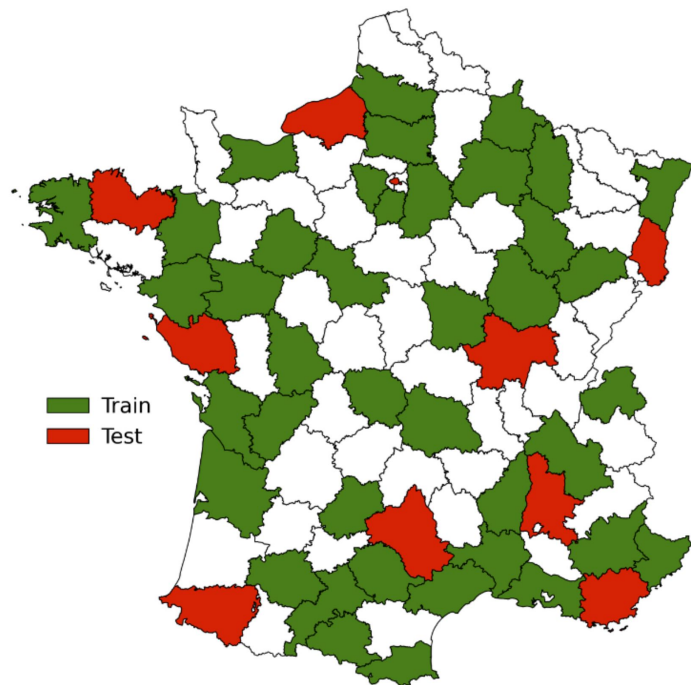
A huge multi-domains dataset
for EO semantic segmentation



French Land cover from Aerospace ImageRy

- 1725-1800 patches per each domain (50)
- 512x512 pixel per patch with 25 cm GSD
- 5 bands (RGB + IR + elevation)
- 19 classes

One of the first, huge dataset for
UDA for RS



Metadata

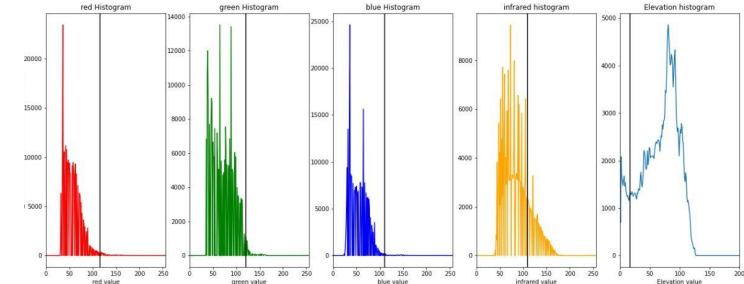
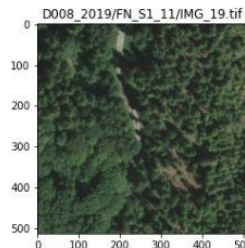
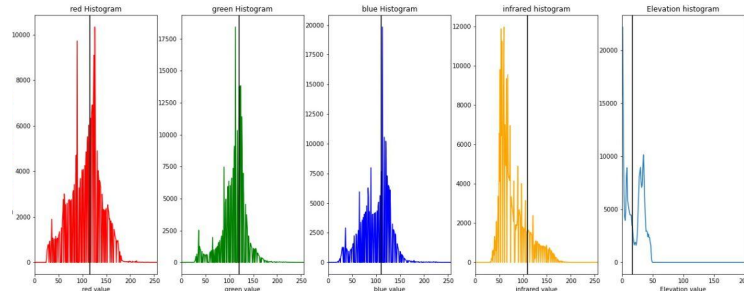
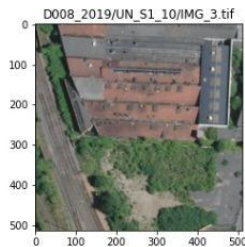
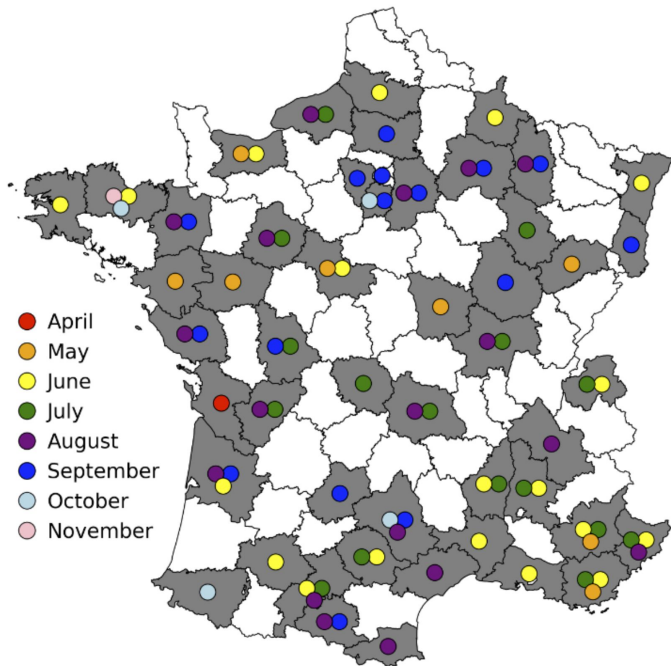
- **domain** info
- the geo **coordinates** (XY) of the centroid and the mean altitude (Z)
- the **date** and **hour** of the acquisition
- the **camera** type

CV/EO datasets are not only made of images

```
{ "IMG_000717":  
  { "domain": "D004_2021", "zone": "Z8_NF",  
  
    "patch_centroid_x": 924009.6,  
    "patch_centroid_y": 6339451.2,  
    "patch_centroid_z": 1661.9399414062,  
  
    "date": "2021-05-28",  
    "time": "08h06",  
  
    "camera": "UCE-M3-f120-s07"} },
```

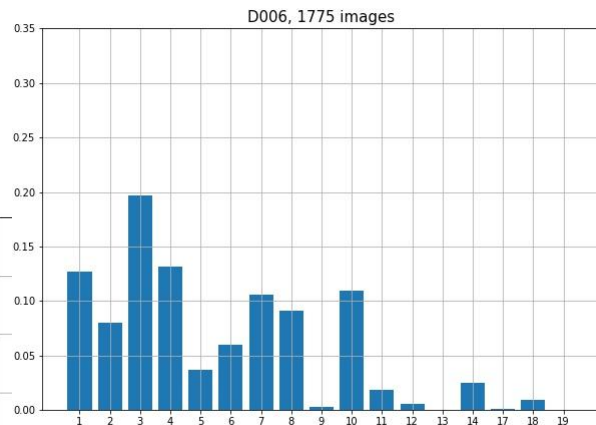
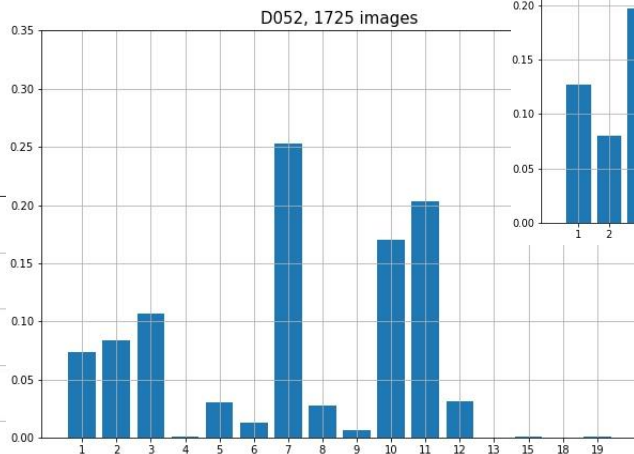
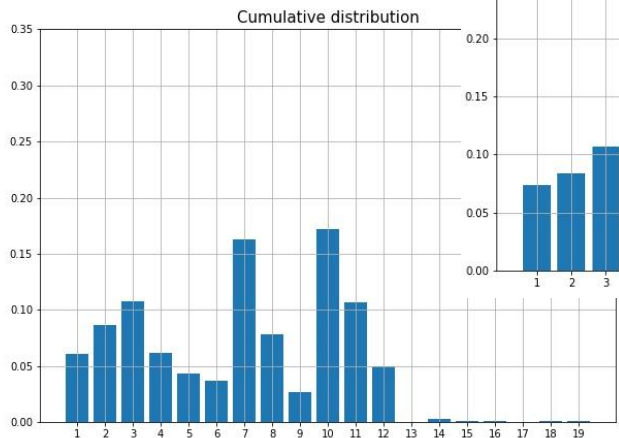
Image level analysis

High intra- and inter-domain variance at image level



Label level analysis

High diversity of label distribution



Class	MSK value
building	1
pervious surface	2
impervious surface	3
bare soil	4
water	5
coniferous	6
deciduous	7
brushwood	8
vine	9
grassland	10
crop	11
plowed land	12
other	>13

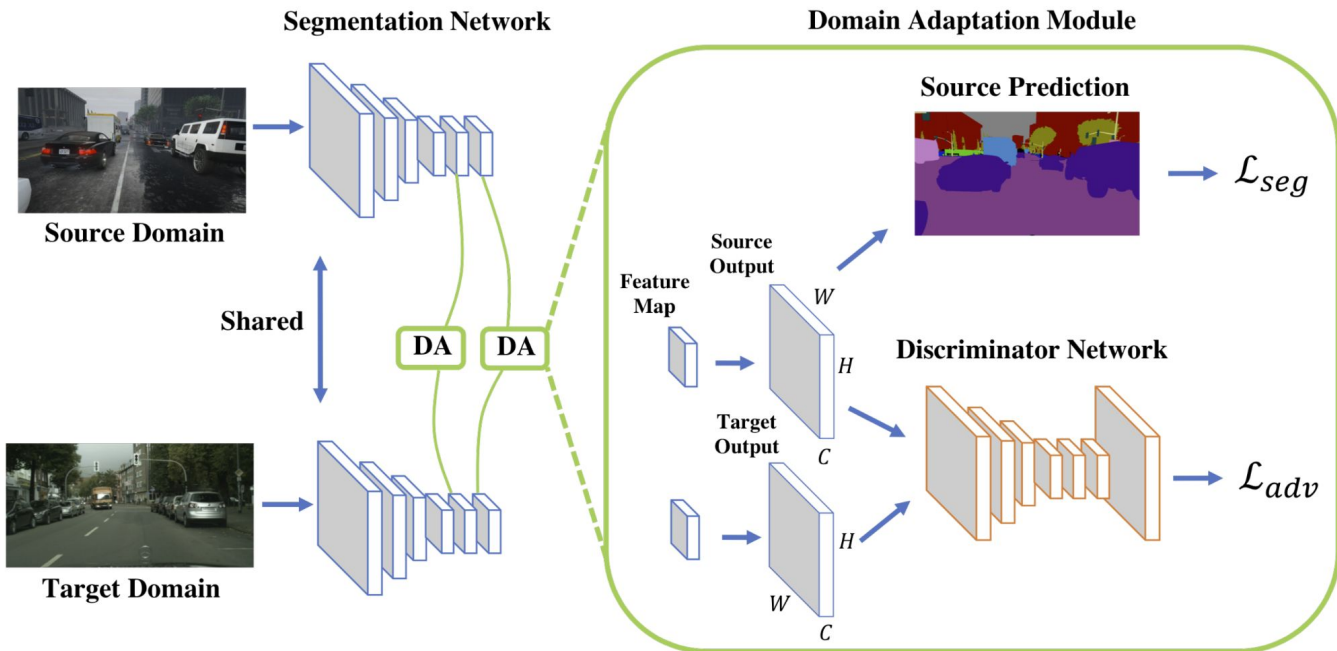
Widespread Methodologies

Spoiler: transformers are the best



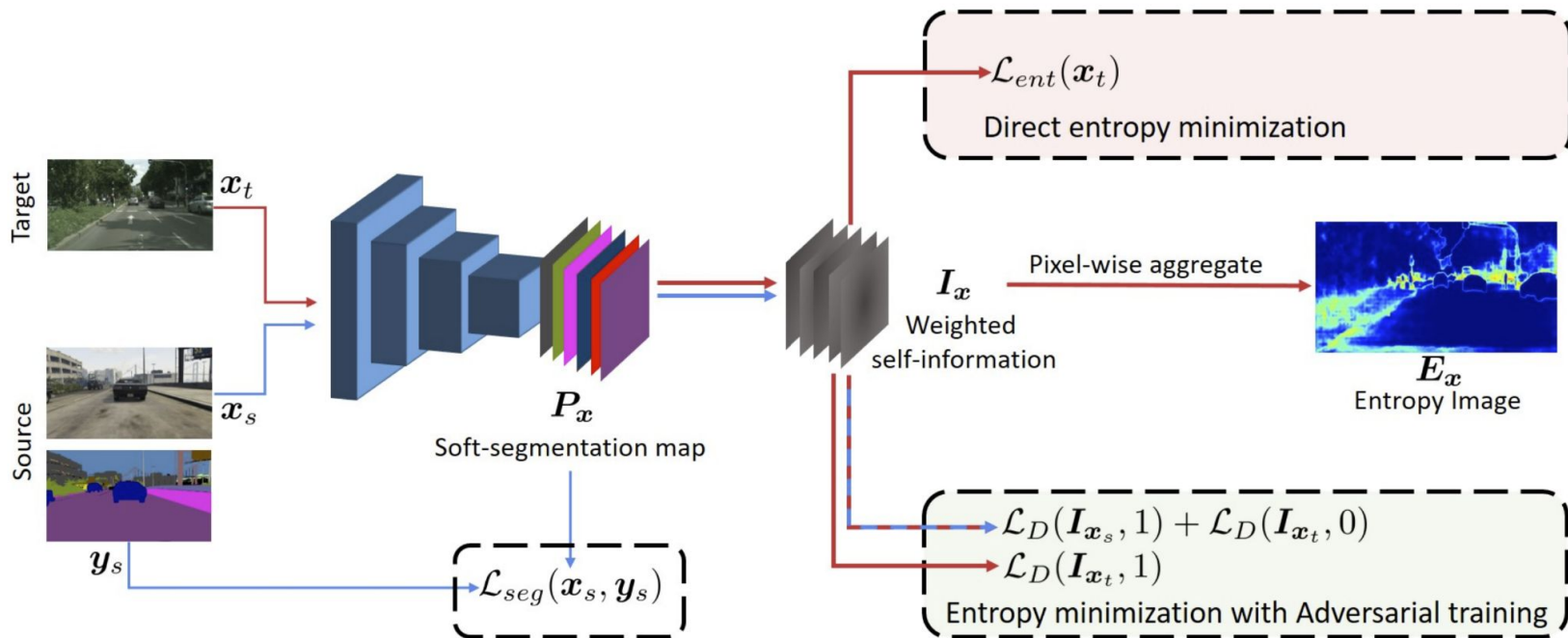
AdaptSegNet

Adversarial training reduces domain shift



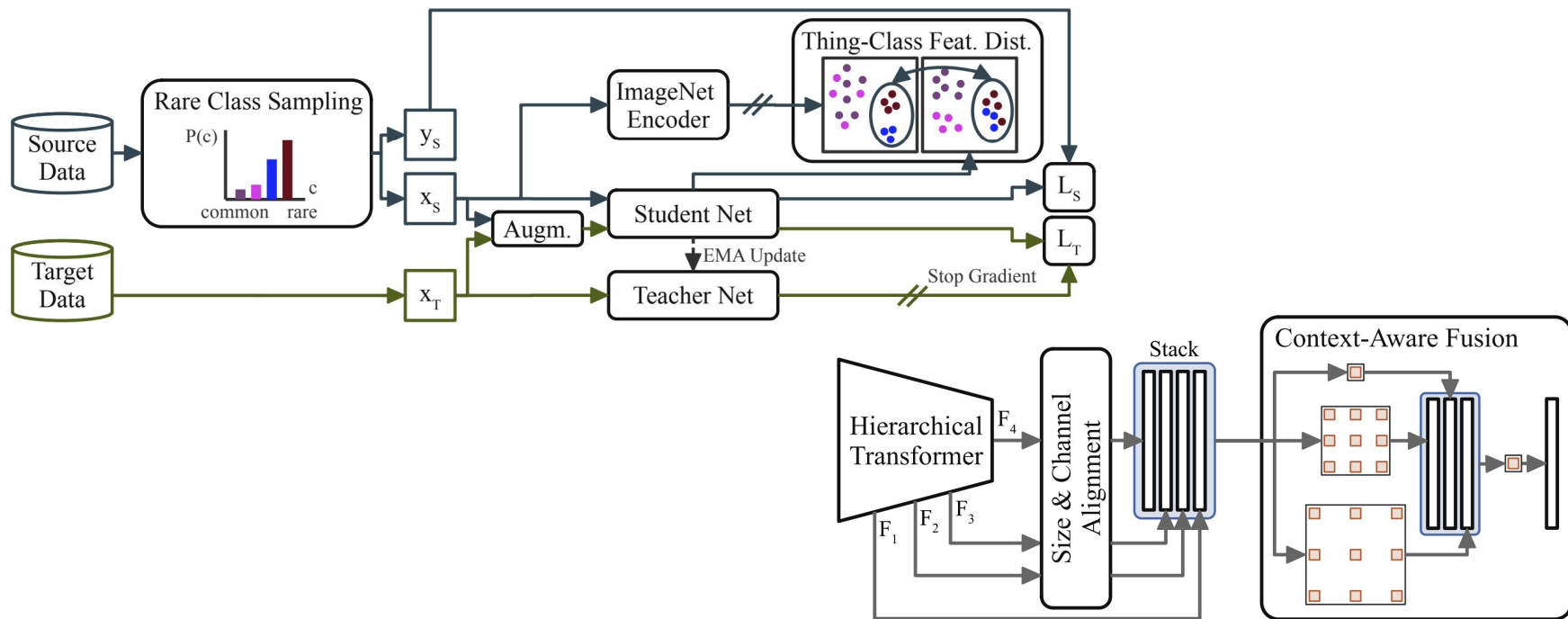
ADVENT

Entropy minimization works well for UDA



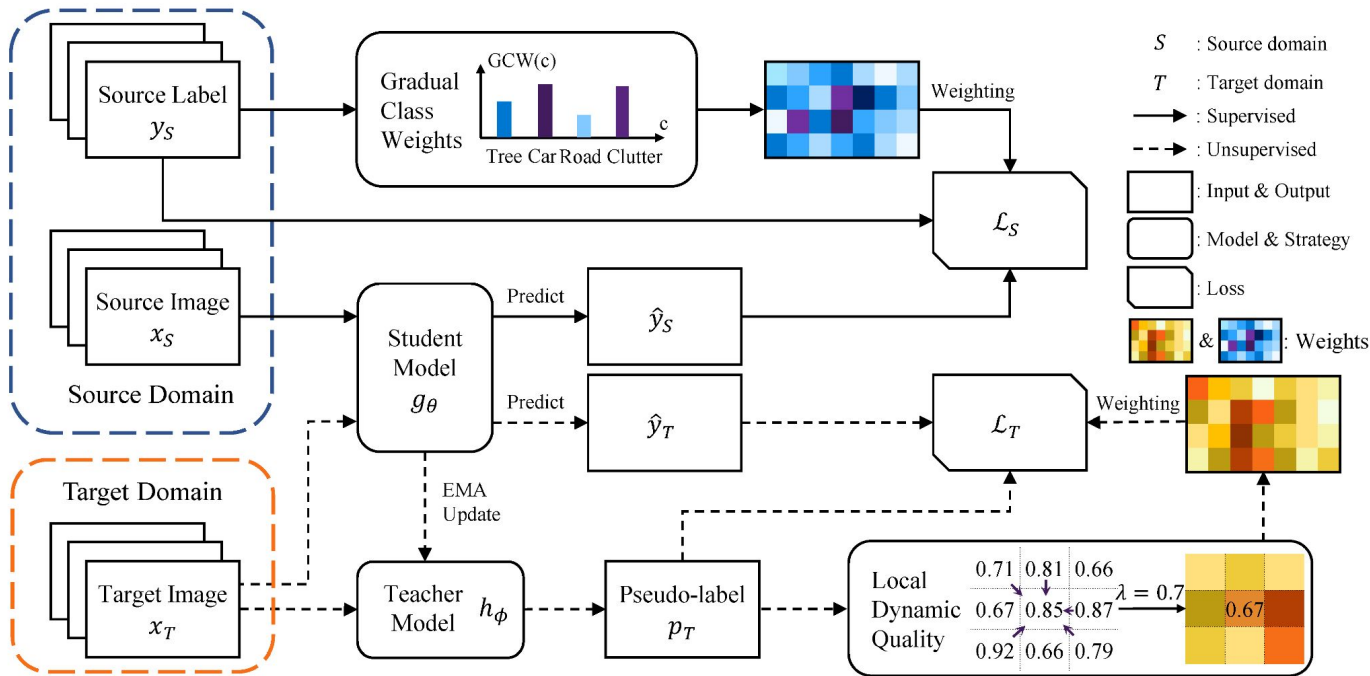
DAFormer

Sampling rare classes and using ImageNet features help



UDA for RS

EO data need their own customized models



Our methodology

or How I learnt to use geo coords

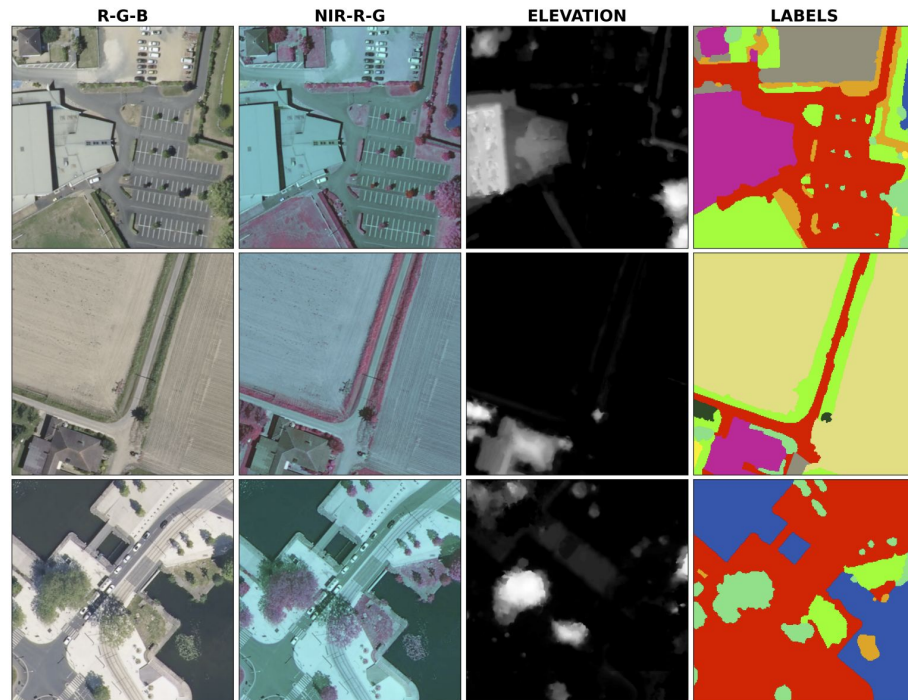


The data we used

- 10 domains for training (D06, D08, D13, D17, D23, D29, D33, D58, D67, D74)
- 3 domains for testing (D64, D68, D71)
- RGB channels

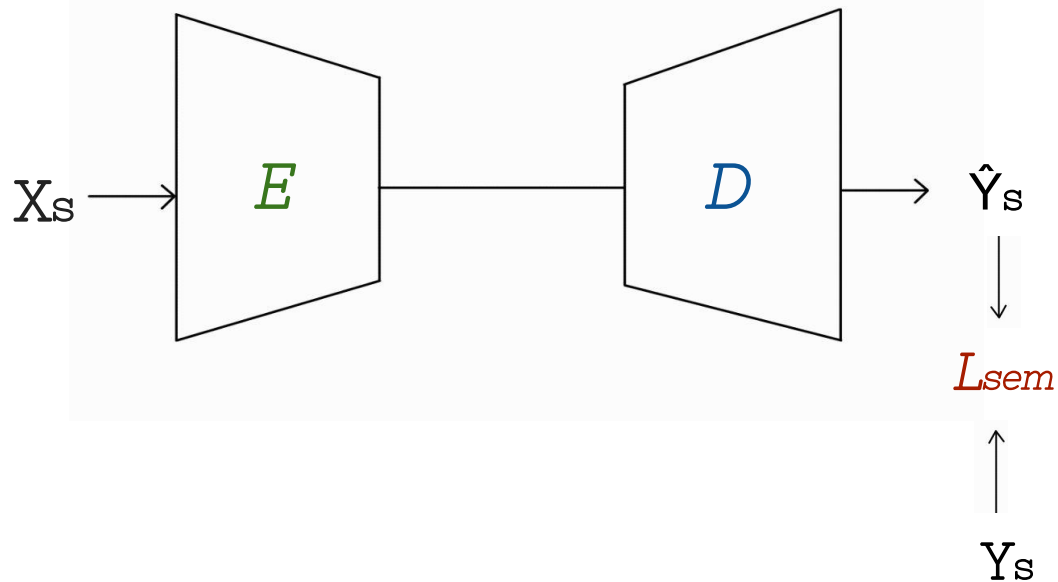
in UDA we use X_s , X_t and Y_s

we used a RGB subset



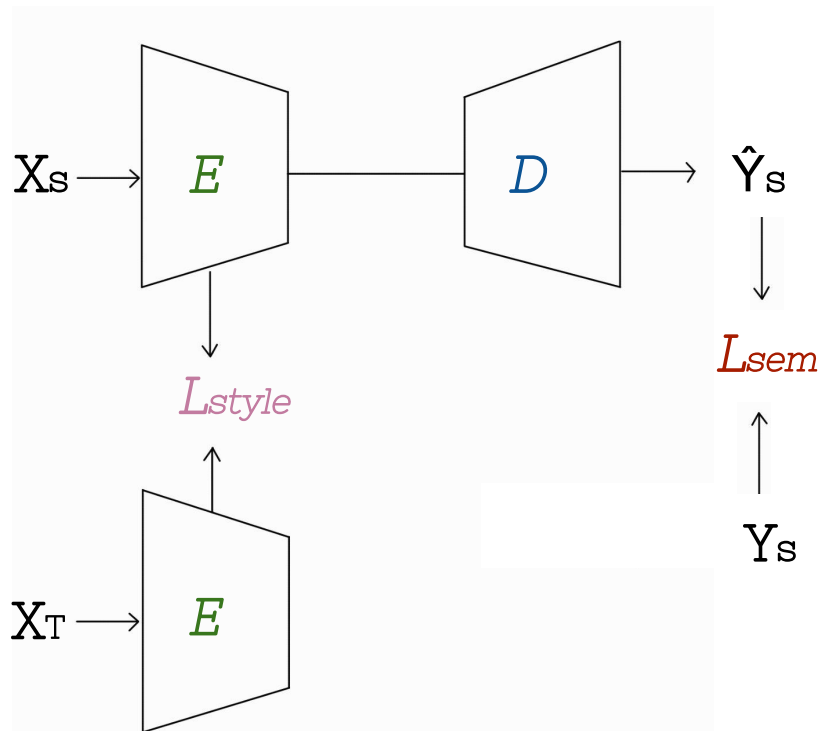
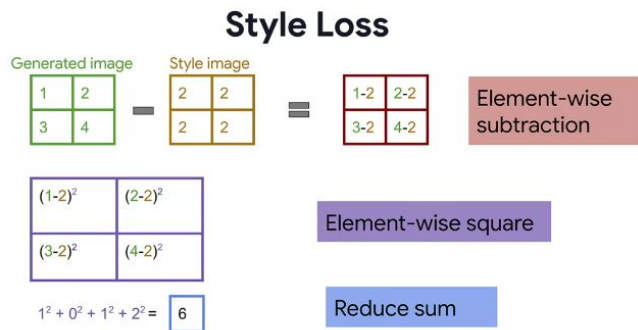
Starting with the baseline

net	mIoU (%)	params (M)
baseline	38.82	1.9



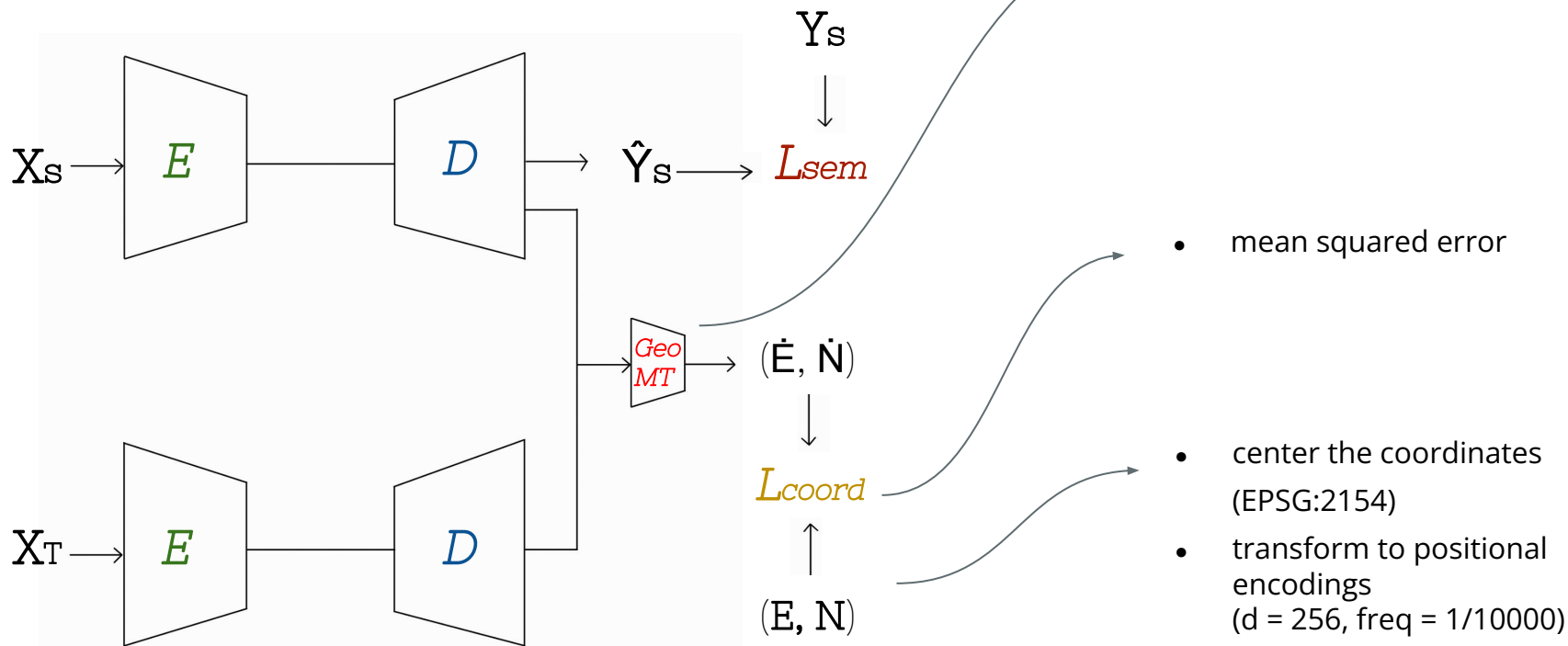
Constraining the features

some easy strategy (e.g. style loss) can improve performance



net	mIoU (%)
baseline	38.82
+style loss	39.83

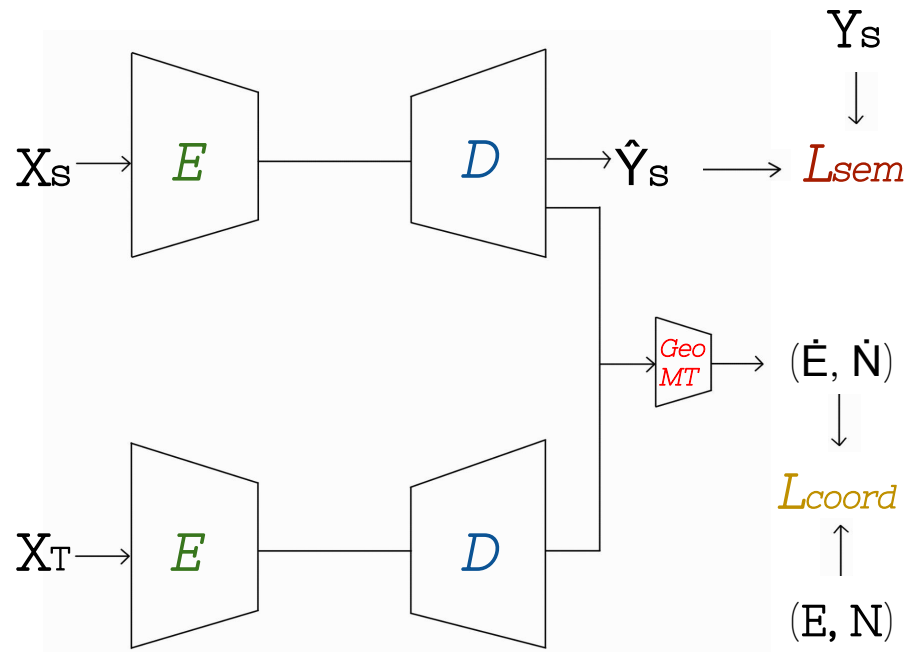
Using geo metadata



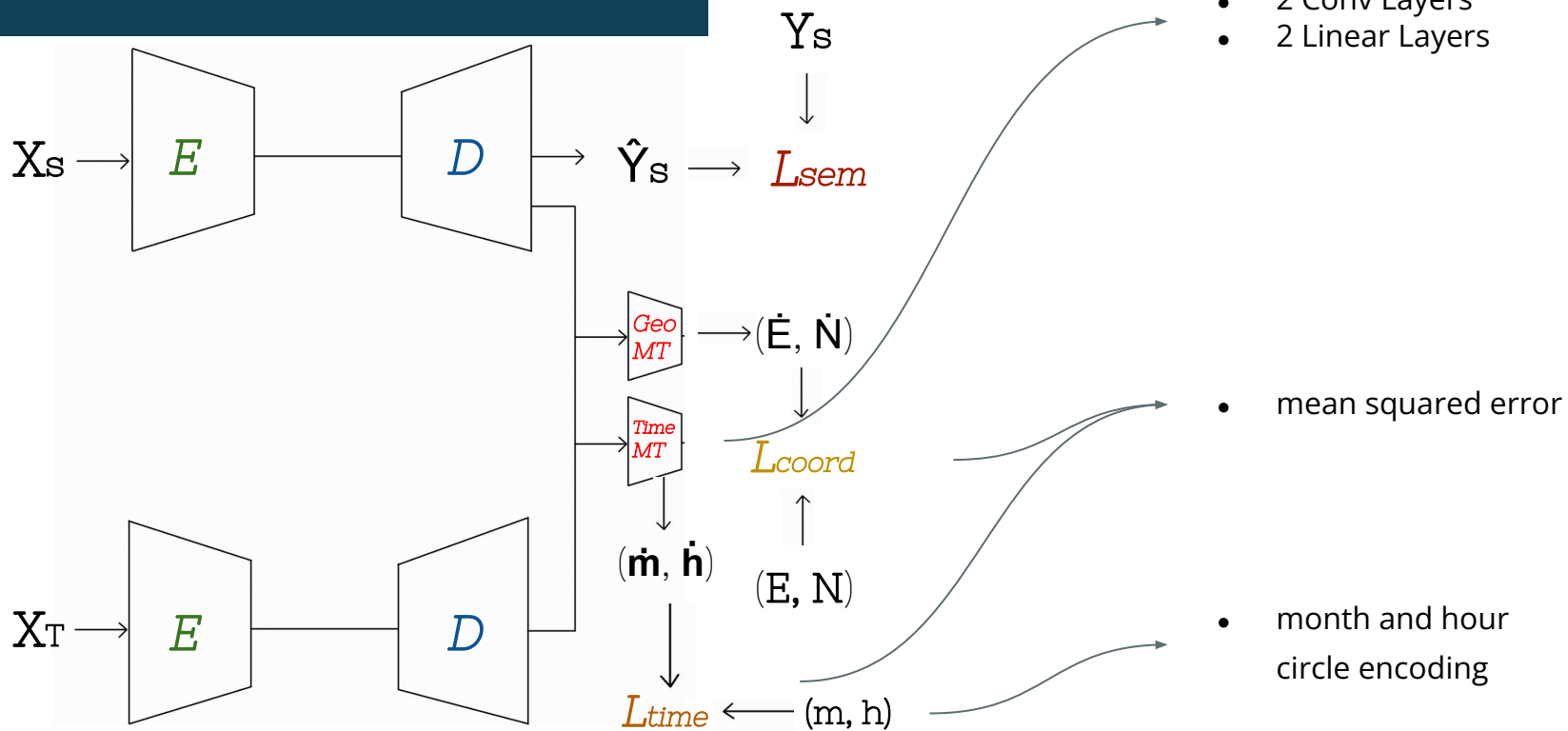
Using geo metadata

net	mIoU (%)	params (M)
baseline	38.82	1.9
+style loss	39.83	1.9
GeoMT_base	40.22 ↑	270 ↑

EO data need their own customized models



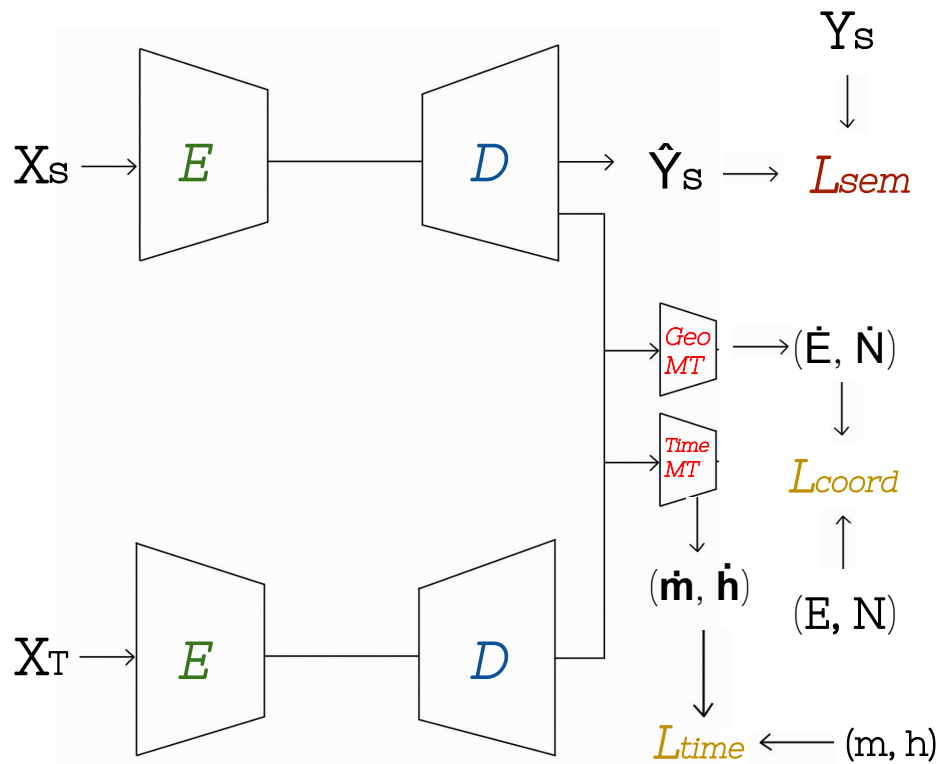
Using time metadata



Using time metadata

net	mIoU (%)	params (M)
baseline	38.82	1.9
+style loss	39.83	1.9
GeoMT_base	40.22	270
TimeGeoMT	35.25 ↓	405 ↑

more metadata ≠ better results



Geo metadata

less precise geoinfo are beneficial

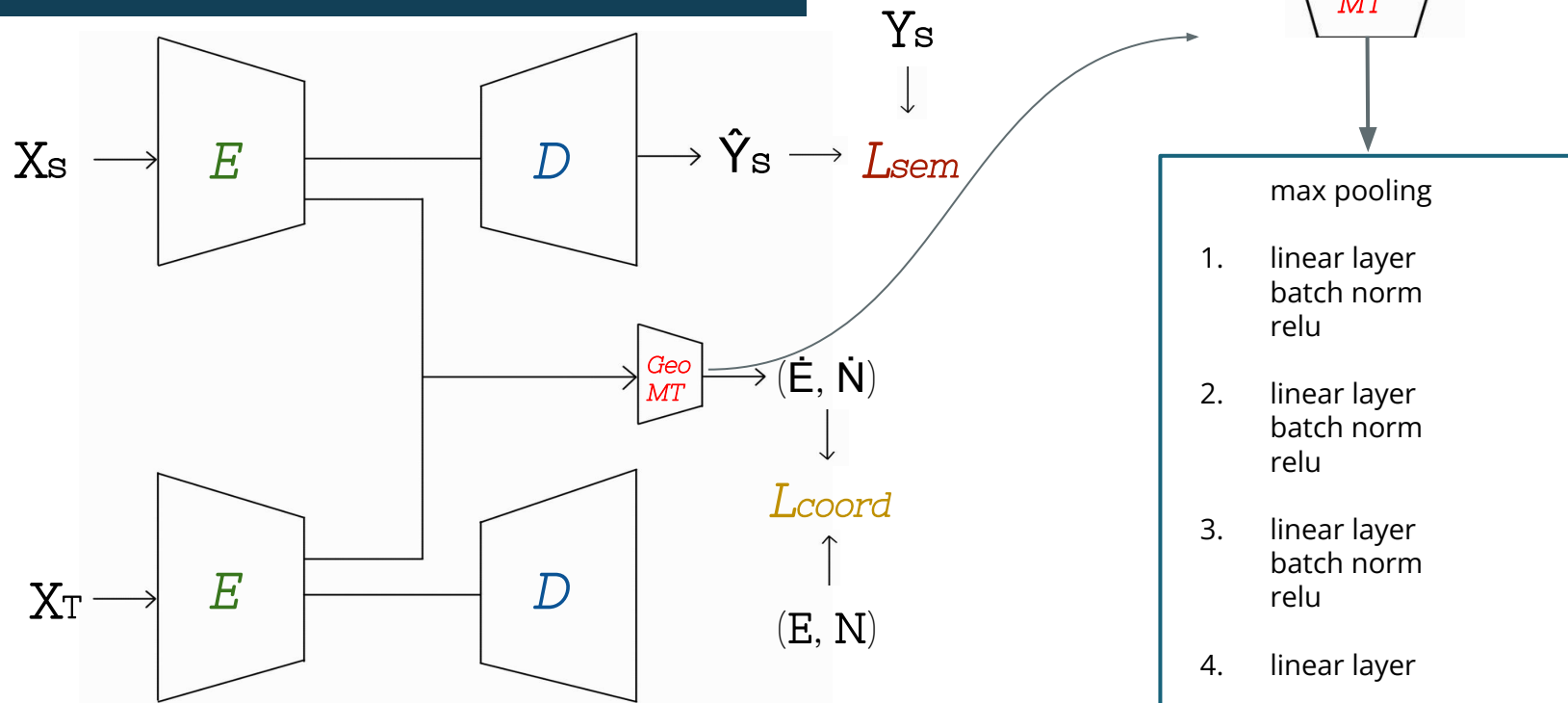
net	noise (km)	1/frequency (-)	mIoU (%)	params (M)
baseline	-	-	38.82	1.9
GeoMT_base	-	10000	40.22	270
GeoMT_noise	±30	10000	40.33	270
GeoMT_noise_lowerfrq	±30	20000	41.38	270
GeoMT_noisier_lowerfrq	±50	20000	39.4	270

(E, N)

- center the coordinates (EPSG:2154)
- **[NEW]** add noise
- transform to positional encodings (**[NEW]** with a lower frequency)

Finally shaping the multitask module

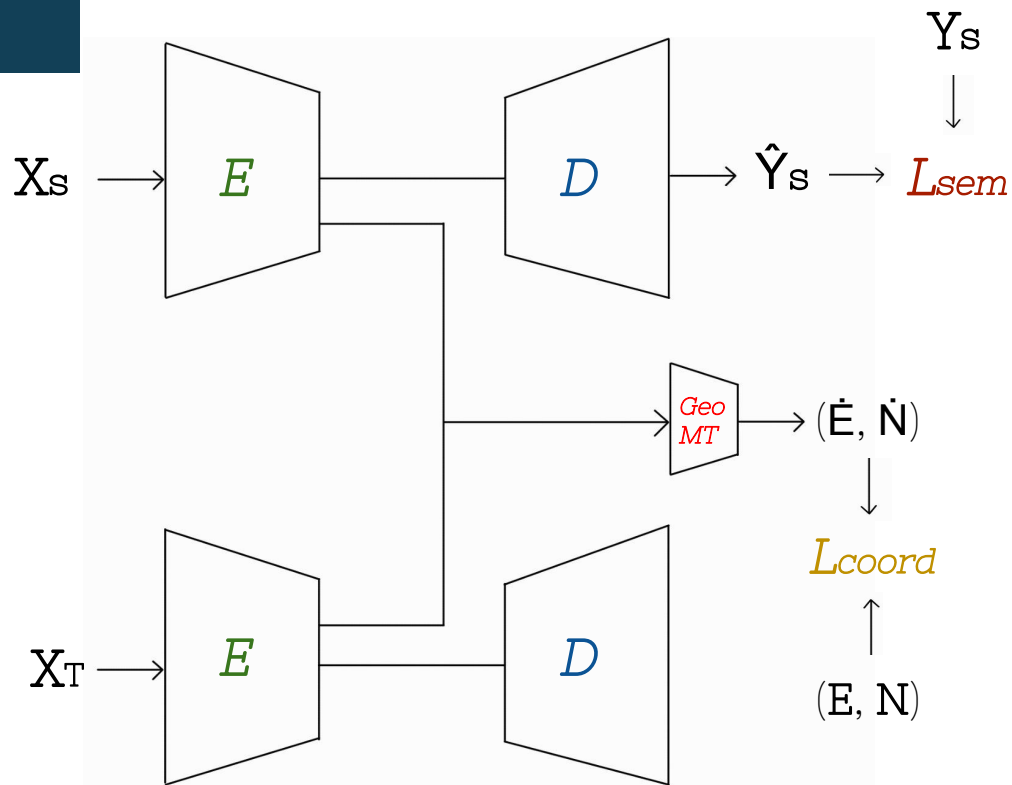
reducing the number of parameters



Finally shaping the multitask module

net	mIoU (%)	params (M)
baseline	38.82	1.9
GeoMT_noise_lowerfrq	41.38	270
GeoMT	41.26	3.3

reducing the number of parameters






Scaling up the model

size doesn't matter that much

net	mIoU (%)	params (M)
baseline	38.82	1.9
GeoMT_UNet	41.26	3.3
GeoMT_ResUNet18	43.29	32.7
GeoMT_ResUNet34	42.76	38.9
GeoMT_ResUNet50	41.03	60.1

Comparison

net	mIoU (%)	params (M)
AdaptSegNet	23.05	99
ADVENT	12.8	99
DAFormer	42.10	85
UDA_for_RS	43.41	85
ours	  43.29	33 

Conclusions

if you fell asleep, please awake now



Conclusions

- **UDA** is a really useful task, slightly **under investigated** in EO
- **FLAIR** is a huge, interesting, **real-world** EO semantic segmentation dataset
- using **metadata** in a good way could boost the model
- **scaling up** to the whole dataset with new models and idea would be fruitful

Some literature references:

Yi-Hsuan Tsai*, Wei-Chih Hung*, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang and Manmohan Chandraker, Learning to Adapt Structured Output Space for Semantic Segmentation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, Patrick Pérez, *ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019

Hoyer, Lukas and Dai, Dengxin and Van Gool, Luc, DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022

Li, W.; Gao, H.; Su, Y.; Momanyi, B.M. Unsupervised Domain Adaptation for Remote Sensing Semantic Segmentation with Transformer. *Remote Sens.* 2022, 14, 4942.
<https://doi.org/10.3390/rs14194942>

Anatol Garioud, Stephane Peillet, Sebastien Giordano, FLAIR: French Land cover from Aerial ImageRy, 2022

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.

Kumar Ayush, Burak Uz kent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021

Baudoux, Luc and Inglada, Jordi and Mallet, Clément, Toward a Yearly Country-Scale CORINE Land-Cover Map without Using Images: A Map Translation Approach, Remote Sensing, 2021

me at my wedding

me flexing 0.1% gain
of my model



Time for questions, remarks
and ideas!

Thank you!