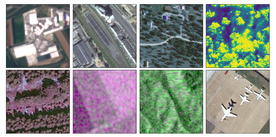
Parameter Efficient Self-Supervised Geospatial Domain Adaptation

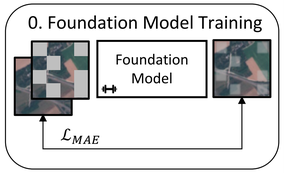
Linus Scheibenreif[1] Michael Mommert[2,1] Damian Borth[1]
[1]University of St.Gallen, Switzerland [2]Stuttgart University of Applied Sciences, Germany

Challenges

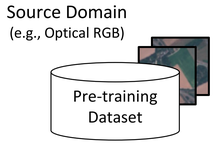
Remote sensing imagery (satellite/aerial):
- Large amounts of unlabeled data (TBs/day)
- Small labeled datasets
- Wide variety of heterogeneous sensors
  - Multispectral
  - Hyperspectral
  - SAR, …



Visual foundation models:
pretrained for specific modality, usually RGB
- MAE (ImageNet) [1]
- SatMAE [2]
- ScaleMAE [3]



Adaptation to unseen modalities and tasks:
- Poor zero-shot capabilities
- Fine-tuning is computationally expensive
- Often infeasible with small labeled datasets



Contributions

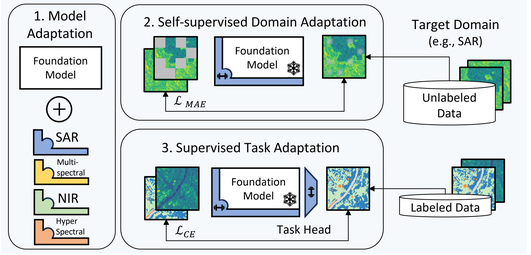
Scaled Low Rank Adapters (SLR) for visual foundation models:
1. Parameter efficient adaptation to unseen modalities
2. Self-supervised continual pre-training framework
3. Reduced memory footprint, improved linear evaluation, fine-tuning, and few-shot performance

References

[1] He, Kaiming, et al. "Masked Autoencoders are Scalable Vision Learners." *CVPR* 2022.
[2] Cong, Yezhen, et al. "SatMAE: Pre-training Transformers for Temporal and Multi-spectral Satellite Imagery." *NeurIPS* (2022): 197-211.
[3] Reed, Colorado J., et al. "Scale-MAE: A Scale-aware Masked Autoencoder for Multiscale Geospatial Representation Learning." *ICCV* 2023.
[4] Hu, Edward J., et al. LoRA: Low-Rank Adaptation of Large Language Models. *ICLR* 2022.

Method

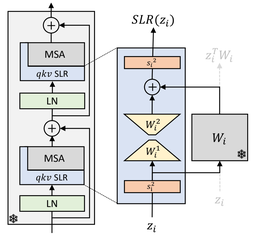
Goal: Adapt pre-trained RGB foundation model to a target task using a small labeled dataset of potentially unseen modality



1. Model Adaptation

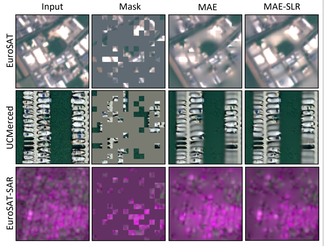
- Introduce parameters for new modality
- Adds 1-2% of model parameters
- Scaled low-rank adapters:
  Linear transform $f(z_i) = z_i W_i$ becomes

$$f_{\text{ada}}(\boldsymbol{z}_i) = \boldsymbol{s}_i^2 \left[(\boldsymbol{s}_i^1 \odot \boldsymbol{z}_i)\boldsymbol{W}_i + ((\boldsymbol{s}_i^1 \odot \boldsymbol{z}_i)\boldsymbol{W}_i^1)\boldsymbol{W}_i^2\right]$$



2. Self-Supervised Domain Adaptation

- Parameters of visual foundation model are fixed.
- **SLR** adapters trained on target modality
- Model learns modality characteristics with self-supervised reconstruction task



3. Supervised Task Adaptation

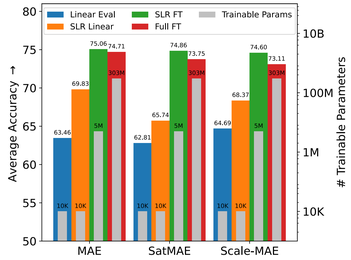
Model is adapted for the task of interest (classification, segmentation) with supervised training of task-head (**SLR Linear**) or task head and adapters (**SLR FT**).

Results

Linear-eval & Finetuning Accuracy

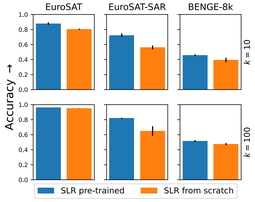
SLR Linear boosts linear probing accuracy, **SLR FT** is on par with fine-tuning using only 1-2% of trainable parameters.

Dataset	MAE				SatMAE				ScaleMAE			
	Linear	SLR Lin.	SLR FT	FT	Linear	SLR Lin.	SLR FT	FT	Linear	SLR Lin.	SLR FT	FT
EuroSAT	93.27	96.61	98.66	98.82	92.00	94.53	98.21	97.79	94.52	95.69	98.65	98.73
RESISC45	77.90	87.08	93.84	95.16	81.80	84.02	92.57	93.39	86.28	87.40	92.87	95.12
FireRisk	37.89	41.78	52.23	49.17	38.27	38.14	50.80	51.21	41.05	41.86	52.63	51.45
TreeSatAI	23.05	38.69	57.66	53.78	21.33	29.55	55.56	50.99	23.48	37.15	53.97	52.58
EuroSAT-SAR	77.95	84.22	87.00	86.46	71.83	79.66	87.17	86.86	73.53	82.73	86.44	78.49
BENGE-S1-C	34.81	42.14	42.80	46.82	35.23	35.60	45.14	44.77	35.35	37.02	45.59	45.07
BENGE-S1-S	68.06	69.74	69.84	69.00	66.57	68.45	70.63	68.27	68.10	68.44	69.05	67.23
UCMerced	94.74	98.35	98.43	98.50	95.44	95.95	98.81	96.65	95.18	96.67	97.60	96.20
Average	63.46	69.83	75.06	74.71	62.81	65.74	74.86	73.74	64.69	68.37	74.60	73.11



Self-supervised Adaptation

Self-supervised training of **SLR** adapters improves performance on new modalities



Few-shot Learning

Adapters significantly improve few-shot performance over conventional fine-tuning.

Method	Params	$k = 10$	$k = 100$
Linear Eval.	10k	75 ± 0.5	89 ± 0.5
SLR Linear	10k	74 ± 0.2	92 ± 0.5
SLR Scale	0.5M	87 ± 0.6	$\mathbf{96 \pm 0.1}$
SLR FT	7.3M	$\mathbf{88 \pm 2.0}$	$\mathbf{96 \pm 0.1}$
Fine-tune	304M	82 ± 2.0	95 ± 0.4

Table 3. Few-shot results with SatMAE on EuroSAT.

Method	Params	$k = 10$	$k = 100$
Linear Eval.	10k	63 ± 0.8	63 ± 0.2
SLR Linear	10k	71 ± 2.9	75 ± 0.1
SLR Scale	0.5M	$\mathbf{74 \pm 3.0}$	77 ± 0.3
SLR FT	7.3M	72 ± 3.0	$\mathbf{82 \pm 1.0}$
Fine-tune	303M	64 ± 1.6	77 ± 3.0

Table 4. Few-shot results with MAE on EuroSAT-SAR.

Parameter Efficiency

SLR adapters introduce $2 \cdot D \cdot r + 2 \cdot D$ additional parameters per linear layer

For example, **ViT-L** (303M):
+3.9M ($r = 8$), or +7.1M ($r = 16$)
new parameters.

Adapter Design

SLR adapters provide strong performance relative to other parameter efficient fine-tuning methods and facilitate few-shot learning.

Method	Accuracy
BitFit [48]	80.34 ± 2.4
$(IA)^3$ [20]	76.72 ± 3.5
Norm tuning [9]	79.00 ± 3.1
LoRA [17]	85.86 ± 0.3
SLR (ours)	$\mathbf{87.14 \pm 0.1}$

Table 5. Performance of different parameter efficient fine-tuning methods when adapting an ImageNet MAE to SAR data (EuroSAT-SAR).

linus.scheibenreif@unisg.ch

github.com/HSG-AIML/GDA