Technical University of Munich

Recent Trends, Challenges, and Limitations of Explainable AI in Remote Sensing

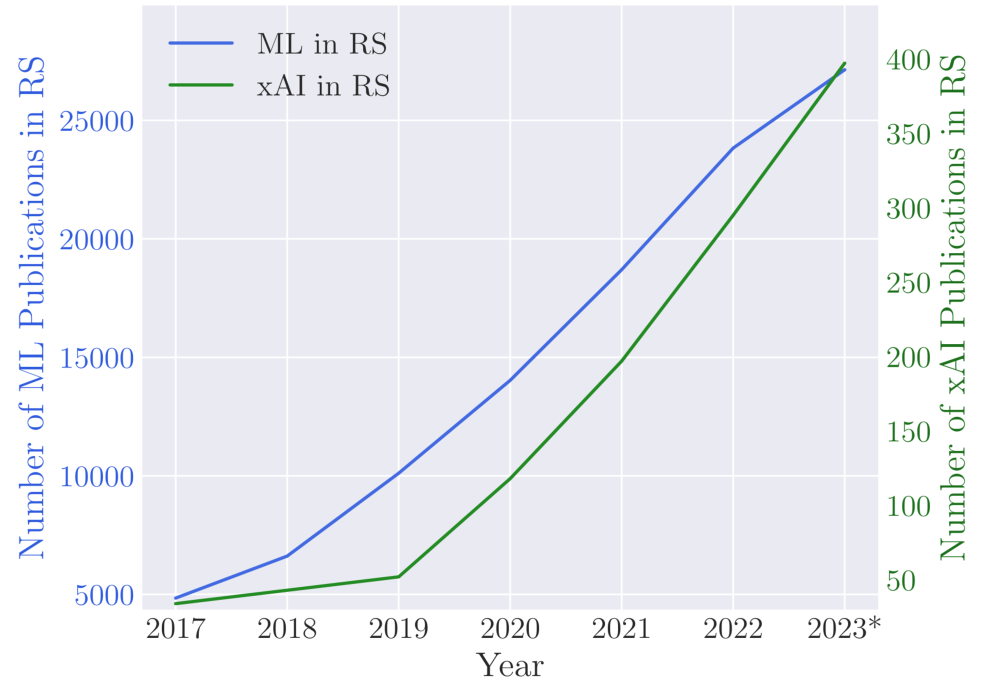
Adrian Höhl[1*], **Ivica Obadic**[1,2*], Miguel-Ángel Fernández-Torres[3], Dario Oliveira[4], and Xiao Xiang Zhu[1,2]

[1]Data Science in Earth Observation, Technical University of Munich (TUM), [2]Munich Center for Machine Learning (MCML), [3]Image Processing Laboratory (IPL), Universitat de València, [4]School of Applied Mathematics, Getulio Vargas Foundation, Brazil, *shared first authorship

CVPR XAI4CV

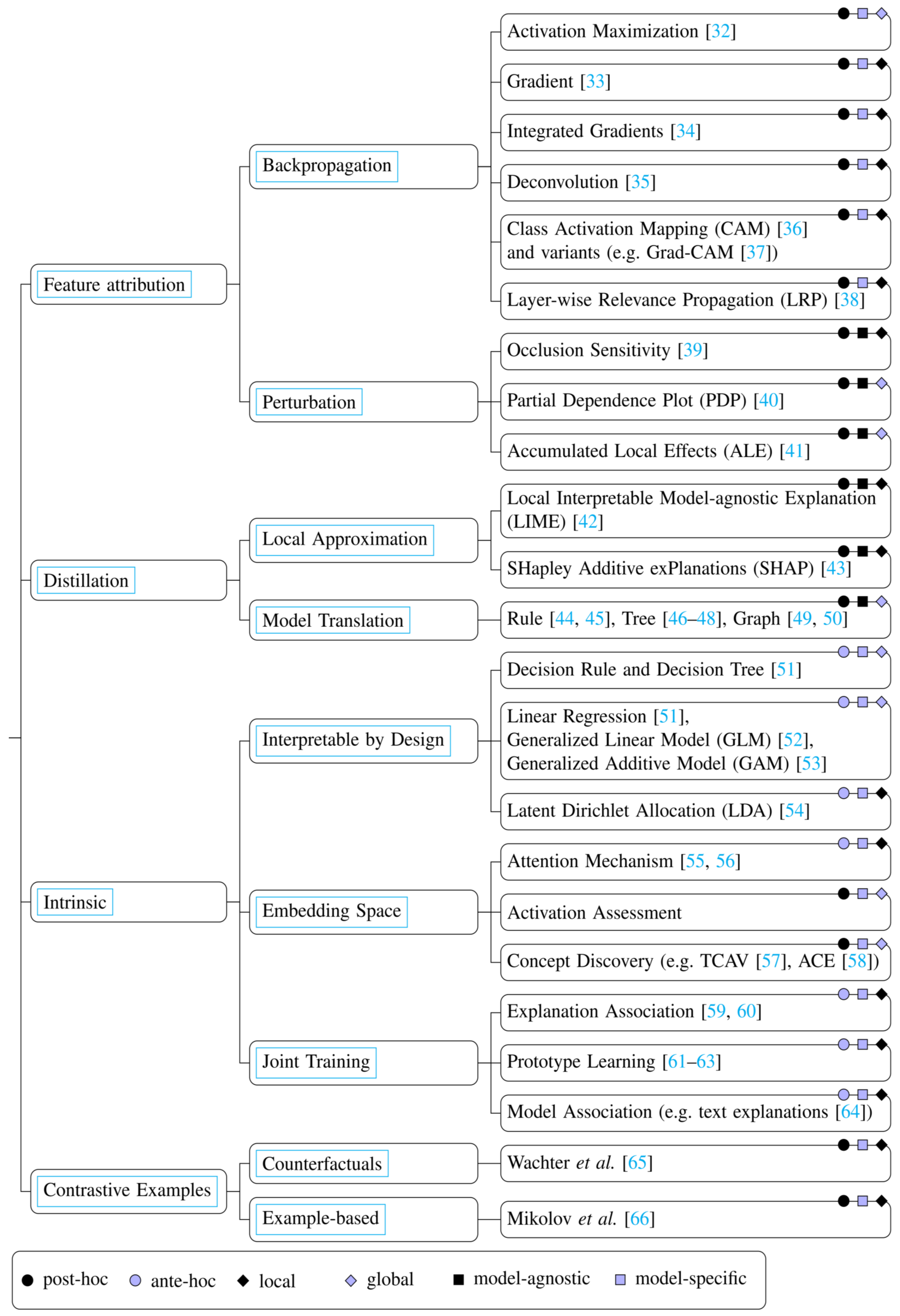
Motivation and Method

- Explainable AI (xAI) is increasingly used in remote sensing (RS)
- Other reviews[1,2] do not extensively cover the usage of xAI across RS nor reflect on the recent challenges in the integration of the two fields
- Transparent and reproducible review by following the PRISMA guidelines[3]
- Search queries in IEEE, Scopus, and Springer databases to cover the literature from 2017 until 2023
- Results: 207 papers included after filtering out the 1075 papers recieved



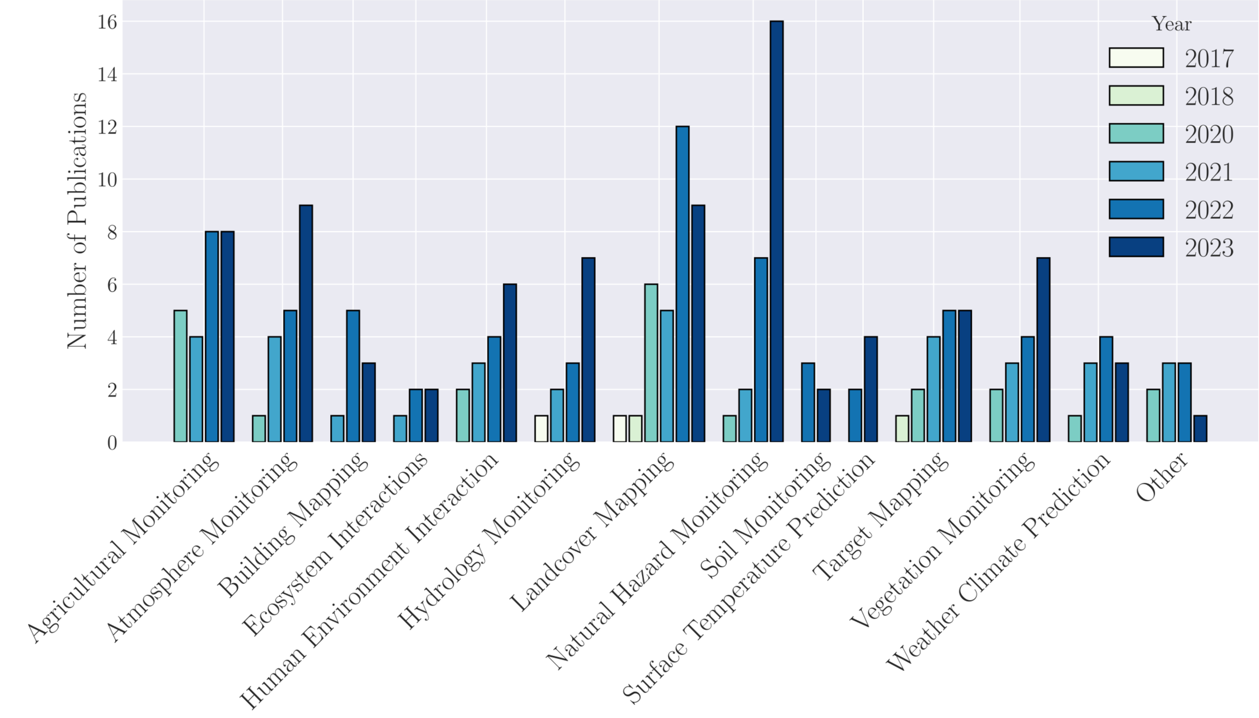
xAI Methods Overview

The categorization from Ras et al. [4] was extended to include the recent developments in xAI and cover the approaches used in RS.



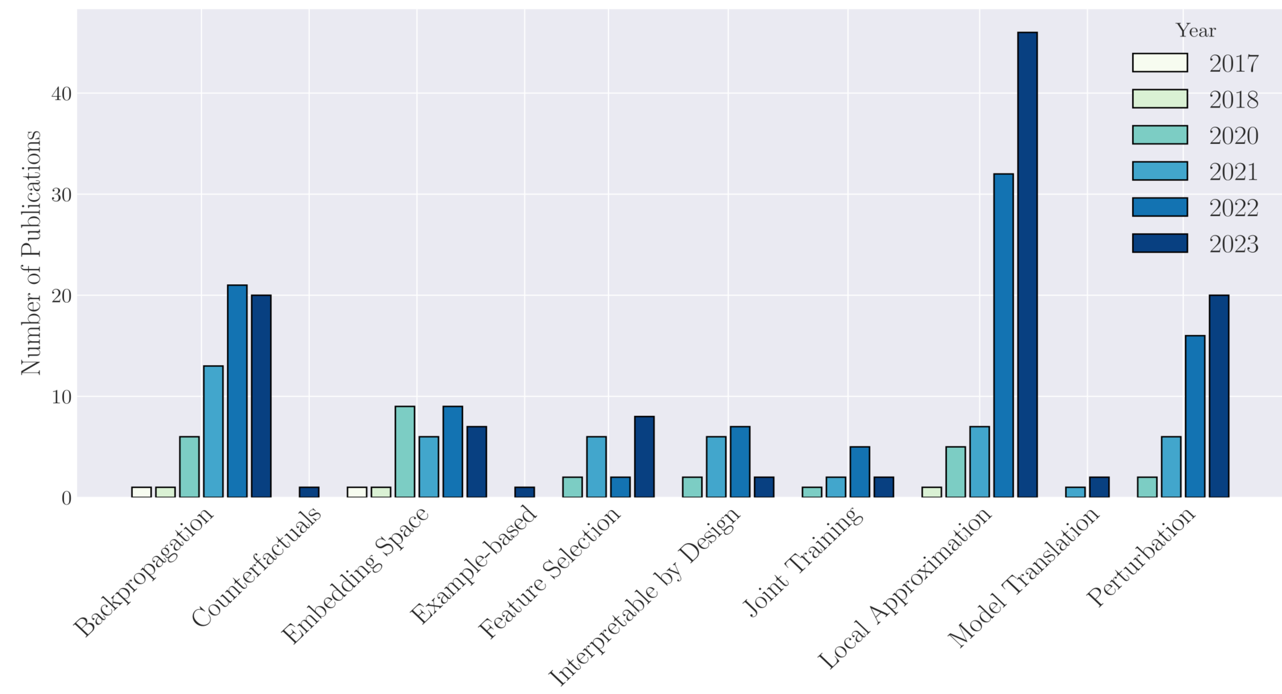
Results and Trends

For which EO applications has xAI been used recently?



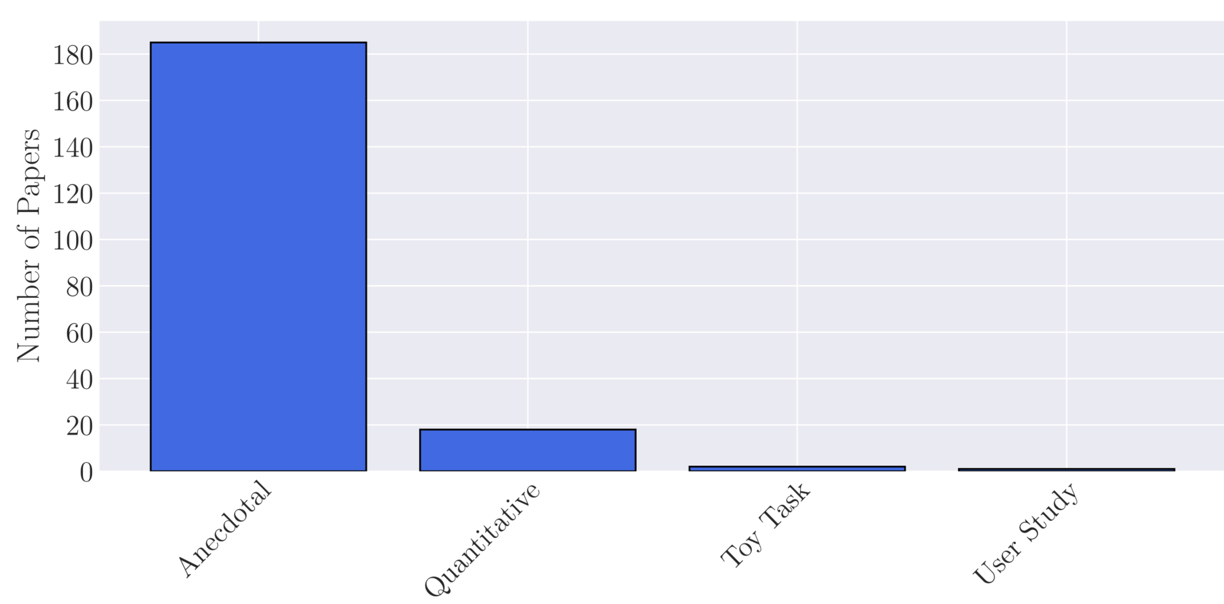
- xAI is increasingly used for EO applications related to natural hazards and atmosphere
- Consistently high utilization of xAI for traditonal EO applications like landcover mapping and agricultural monitoring

What are the popular xAI methods in RS?



- High usage of post-hoc xAI approaches
- Local approximation (LIME and SHAP) and backpropagation (Grad-CAM) methods are particularly used

How are xAI findings evaluated?



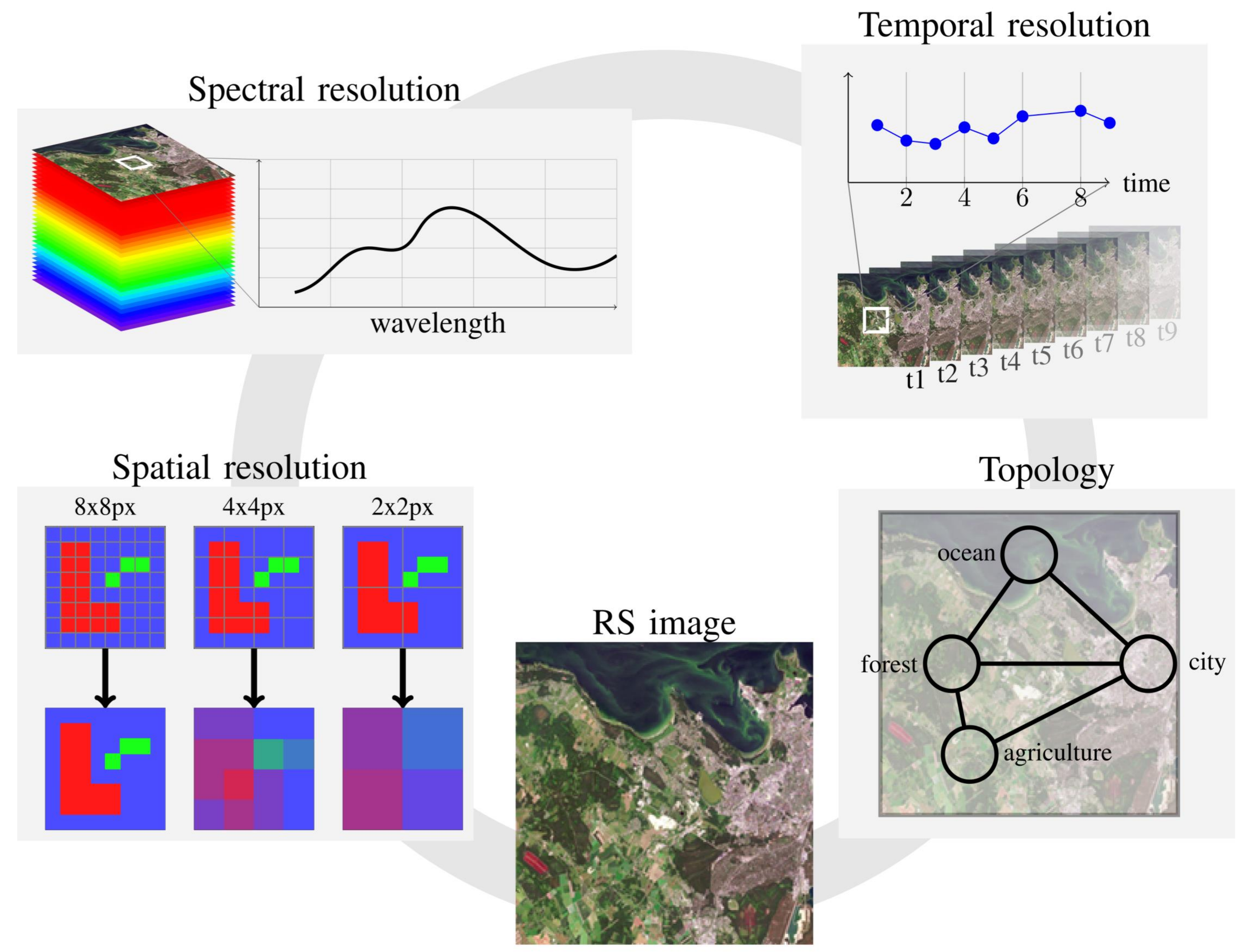
- Mostly anecdotal evaluation, although quantitative metrics can provide more objective evaluation
- User studies are essential to evaluate the expert's ability for using the xAI findings

Challenges and Limitations

RS image properties

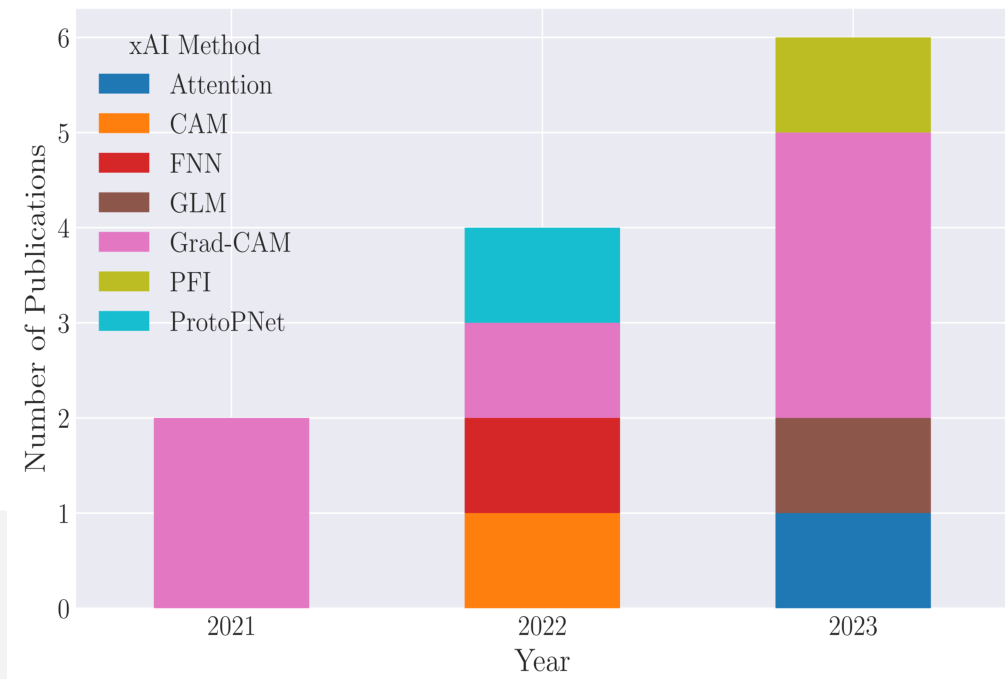
xAI approaches used in RS should consider the properties of RS data:

- Scale: Sharp object boundaries, spatial and spectral resolution (e.g., target objects can be quite small)
- Topology: Geographic confounders and teleconnections are hard to model
- Time series: Most xAI methods do not account for temporal dependencies, e.g. is it appropriate to use SHAP to explain time-series models?



Adapting xAI methods to RS data

- Increasing number of adapted methods proposed in the last year
- Highest focus is on the CAM methods for the accurate localization of the small objects present in RS imagery
- Permutation feature importance (PFI) adaptation [5] incorporates the importance of spatial distances
- ProtoPNet modification [6] considers the location of the features

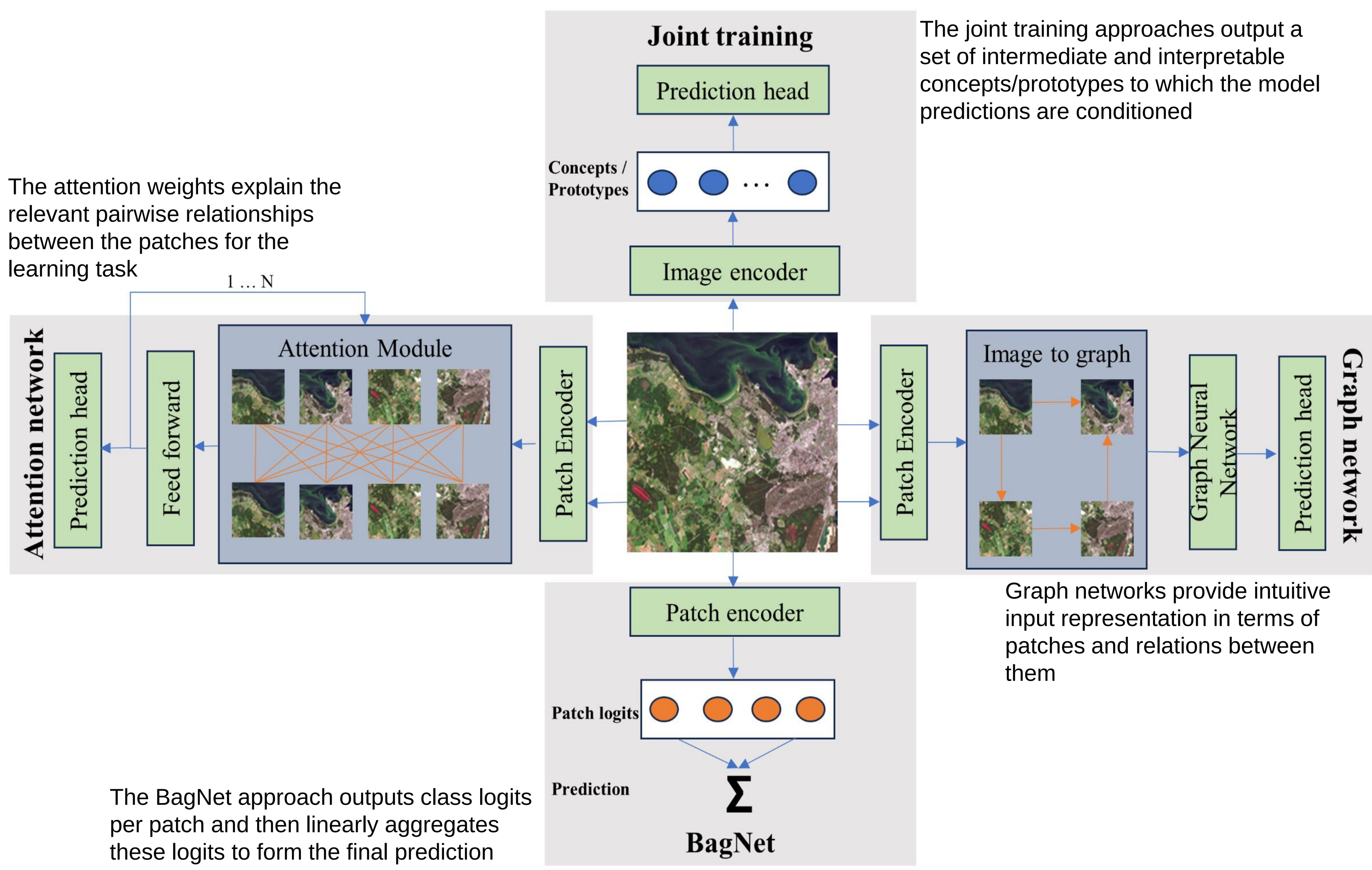


Towards Interpretable Neural Networks

- Most post-hoc methods for DNNs provide only saliency maps
- Intrinsic methods for DNNs can provide more intuitive insights than raw feature importances
- Only a few works introduce interpretable DNNs for EO tasks. Used are joint training approaches, attention networks, graph networks and BagNet models.

The joint training approaches output a set of intermediate and interpretable concepts/prototypes to which the model predictions are conditioned

The attention weights explain the relevant pairwise relationships between the patches for the learning task



Graph networks provide intuitive input representation in terms of patches and relations between them

The BagNet approach outputs class logits per patch and then linearly aggregates these logits to form the final prediction

Technical University of Munich
Department of Aerospace and Geodesy
Chair of Data Science in Earth Observation
Bundesministerium für Wirtschaft und Klimaschutz
ML4Earth
MCML Munich Center for Machine Learning
International FUTURE LABS

References:
1. Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. IEEE Access, 8, 42200-42216.
2. Gevaert, C. M. (2022). Explainable AI for Earth observation: A review including societal and regulatory perspectives. International Journal of Applied Earth Observation and Geoinformation, 112, 102869.
3. Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews BMJ 2021; 372 :n71 doi:10.1136/bmj.n71
4. Ras, G., Xie, N., Van Gerven, M., & Doran, D. (2022). Explainable deep learning: A field guide for the uninitiated. Journal of Artificial Intelligence Research, 73, 329-396.
5. Brenning, A. (2023). Spatial machine-learning model diagnostics: a model-agnostic distance-based approach. International Journal of Geographical Information Science, 37(3), 584-606.
6. Barnes, E. A., Barnes, R. J., Martin, Z. K., & Rader, J. K. (2022). This looks like that there: Interpretable neural networks for image tasks when location matters. Artificial Intelligence for the Earth Systems, 1(3), e220001.