

Rank-based Geographical Regularization: Revisiting Contrastive Self-Supervised Learning for Multispectral Remote Sensing Imagery



Tom Burgert



Leonard Hackel



Paolo Rota



Begüm Demir

1. Introduction

- RS multispectral images differ from CV images due to spatial, spectral, and temporal variability, but offer zero-cost metadata like geolocation

1. Introduction

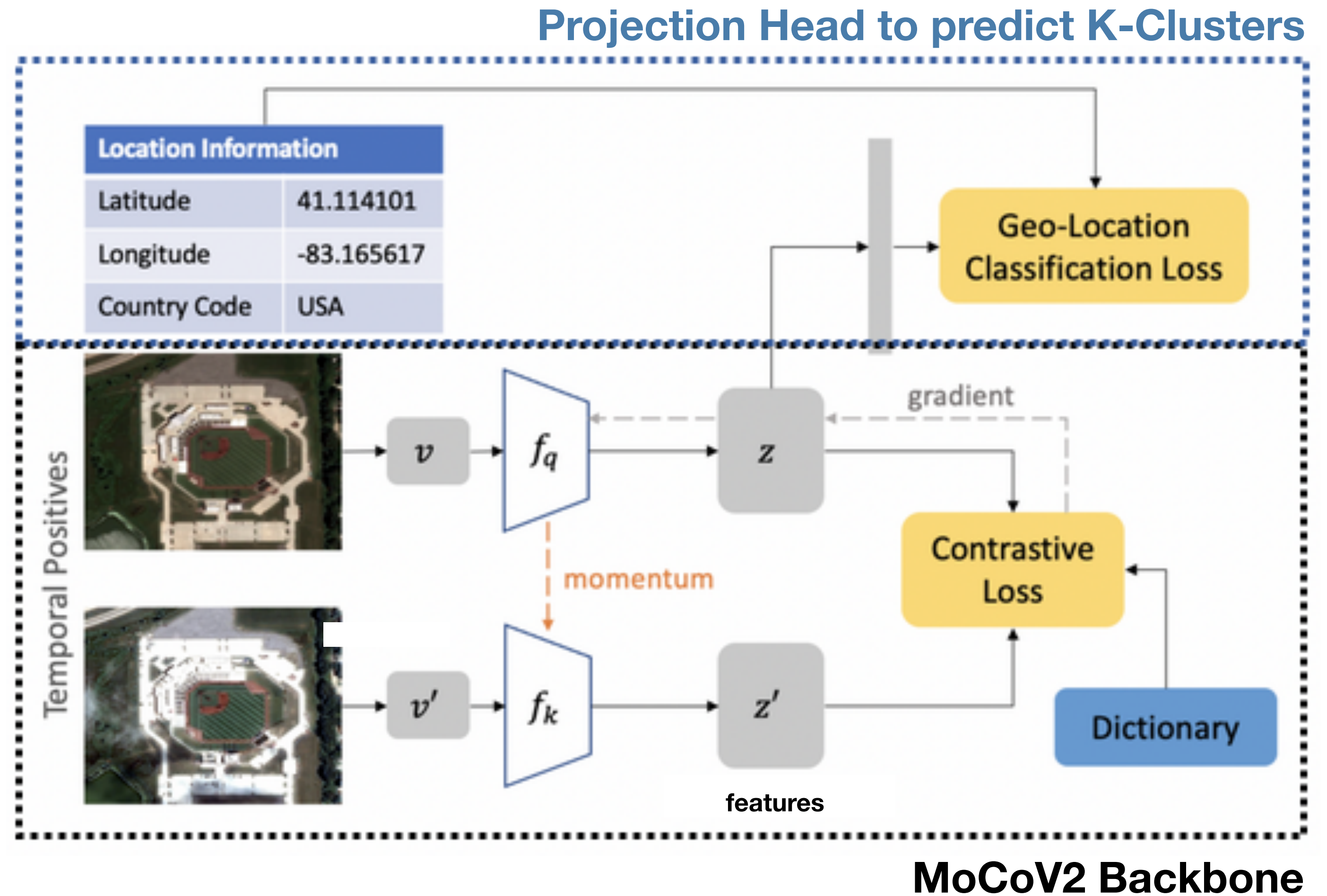
- RS multispectral images differ from CV images due to spatial, spectral, and temporal variability, but offer zero-cost metadata like geolocation
- **Contributions:**
 - We propose GeoRank, a rank-based regularization method that integrates geographical metadata as weak supervision into contrastive SSL

1. Introduction

- RS multispectral images differ from CV images due to spatial, spectral, and temporal variability, but offer zero-cost metadata like geolocation
- **Contributions:**
 - We propose GeoRank, a rank-based regularization method that integrates geographical metadata as weak supervision into contrastive SSL
 - We revisit core design decisions in RS SSL like (i) data augmentation, (ii) dataset size, (iii) image size, and (iv) temporal views, highlighting overlooked limitations and proposing improved practices

2. Related Work

Geo-Aware SSL - Ayush et al. [1]



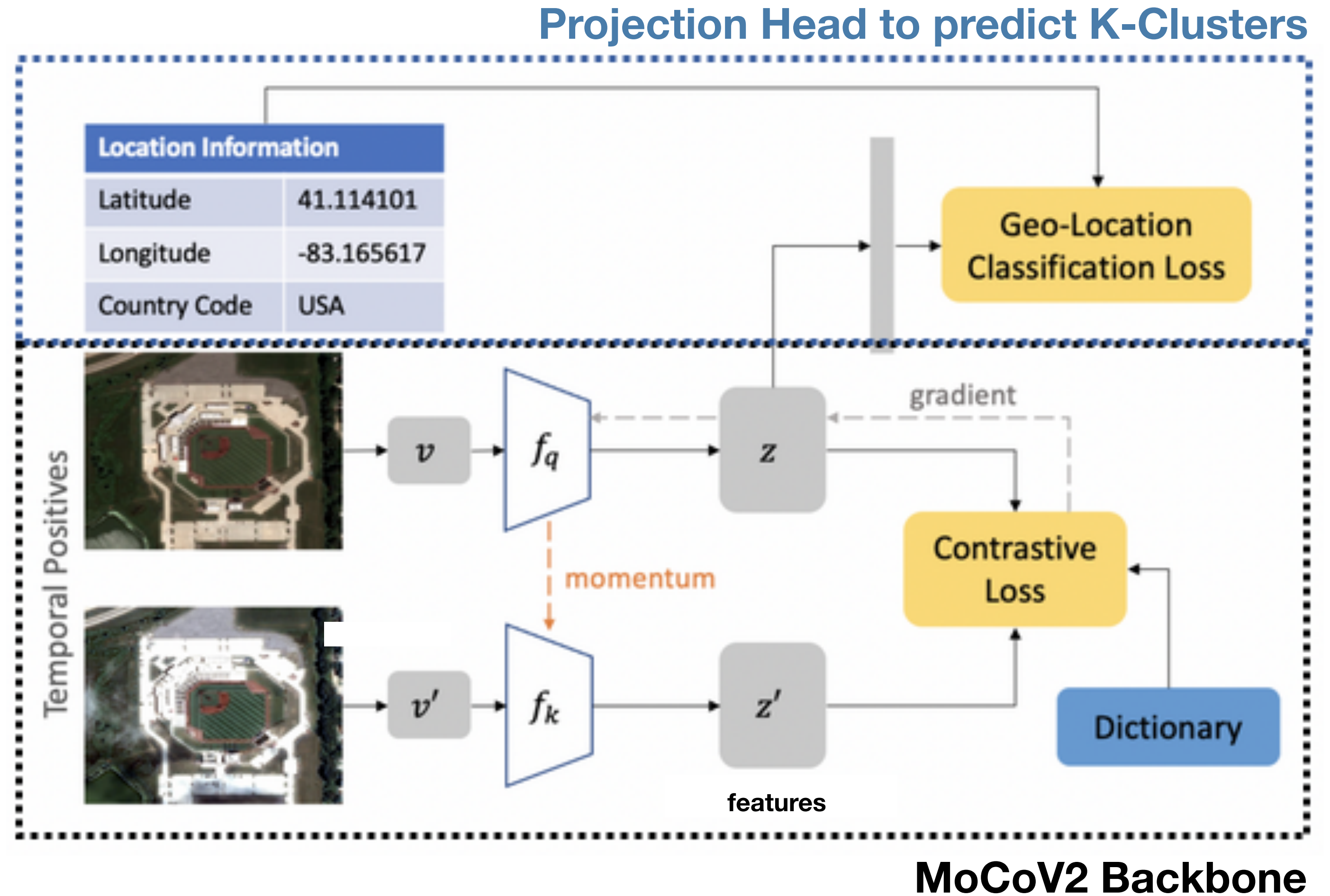
[1] K. Ayush et al., "Geography-aware self-supervised learning", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

2. Related Work

Geo-Aware SSL - Ayush et al. [1]

Limitations of SOTA:

- Rely on a two-stage approach, requiring precomputed K-Means clusters



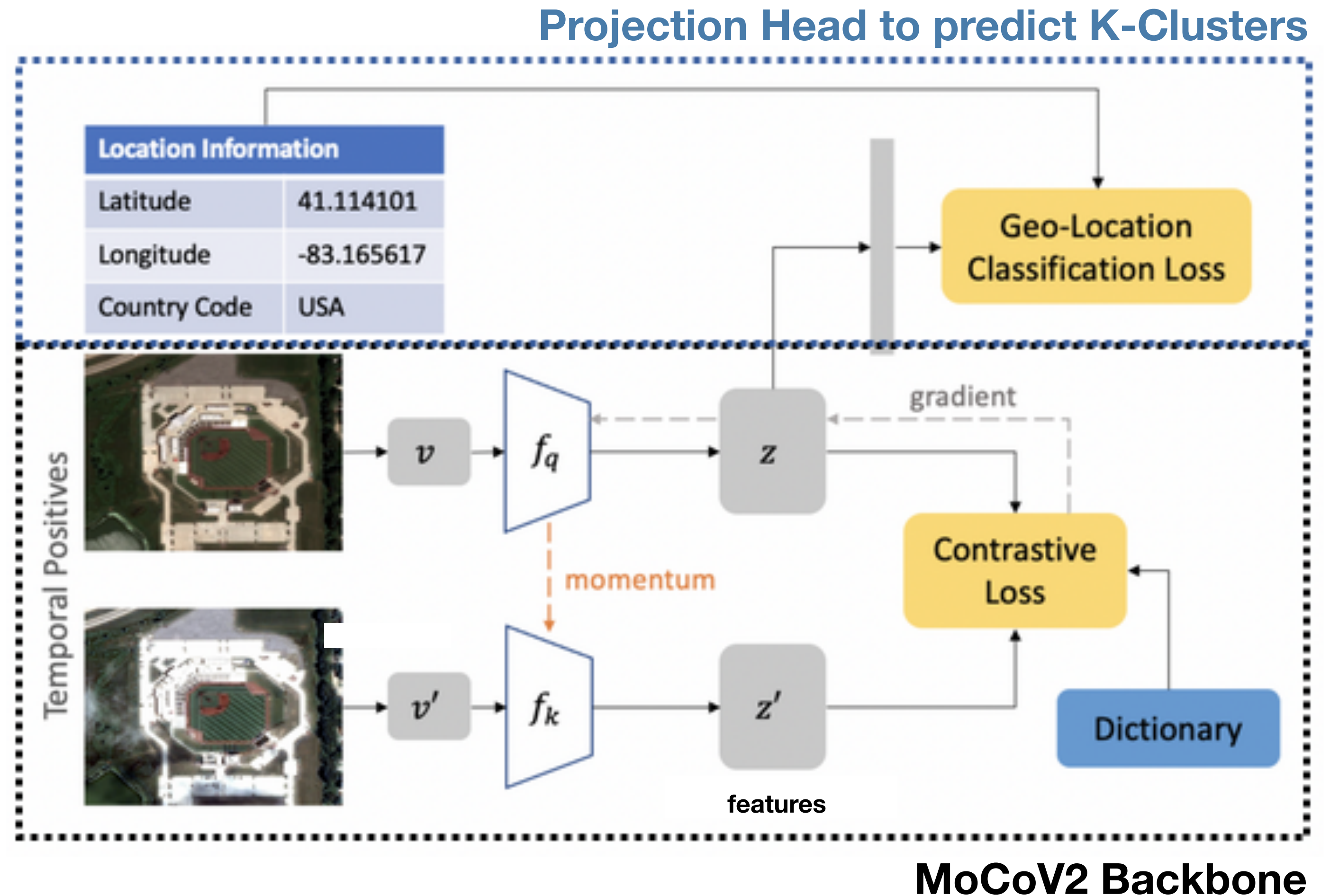
[1] K. Ayush et al., "Geography-aware self-supervised learning", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

2. Related Work

Geo-Aware SSL - Ayush et al. [1]

Limitations of SOTA:

- Rely on a two-stage approach, requiring precomputed K-Means clusters
- Use Euclidean distances, which poorly approximate Earth's spherical geometry



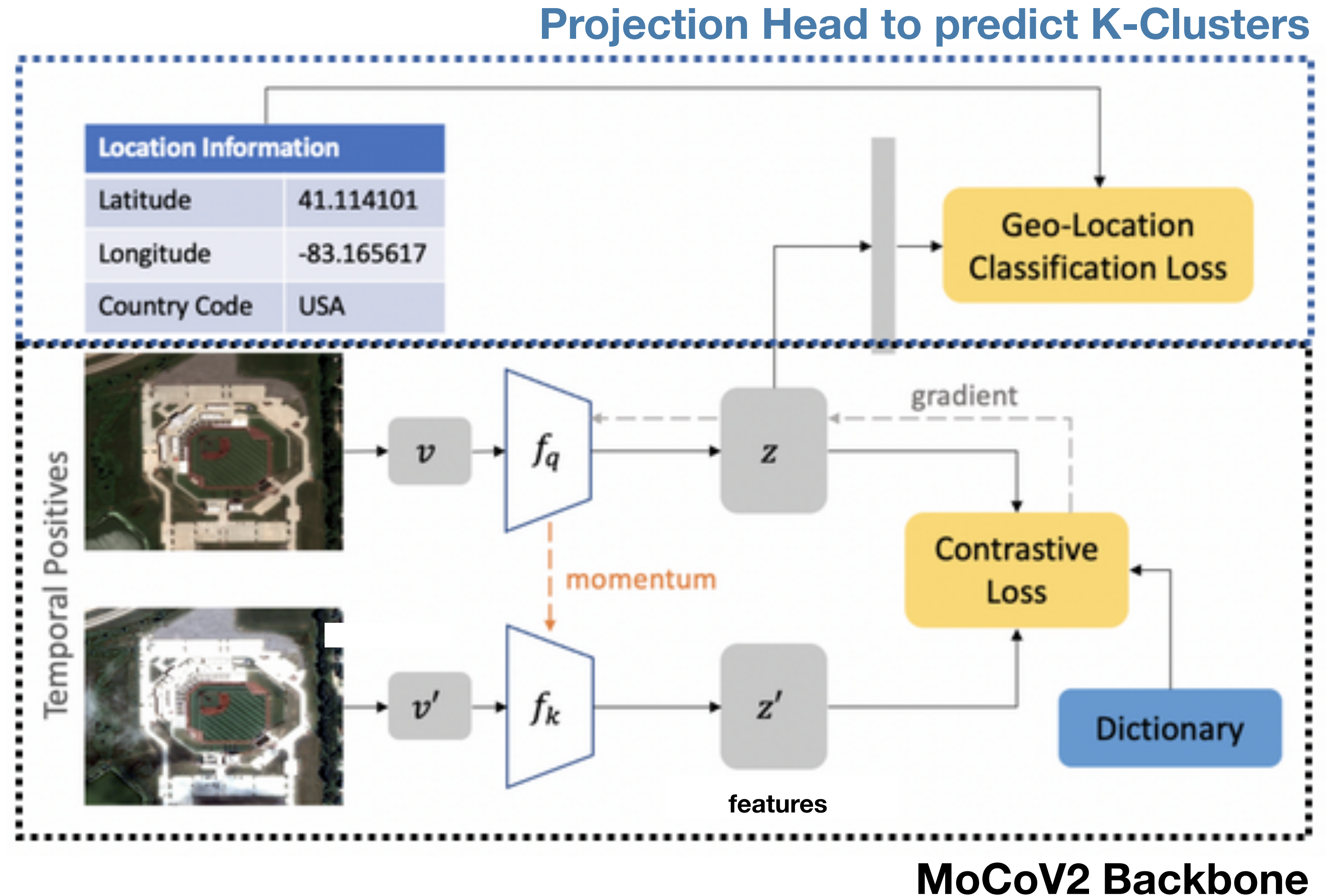
[1] K. Ayush et al., "Geography-aware self-supervised learning", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

2. Related Work

Geo-Aware SSL - Ayush et al. [1]

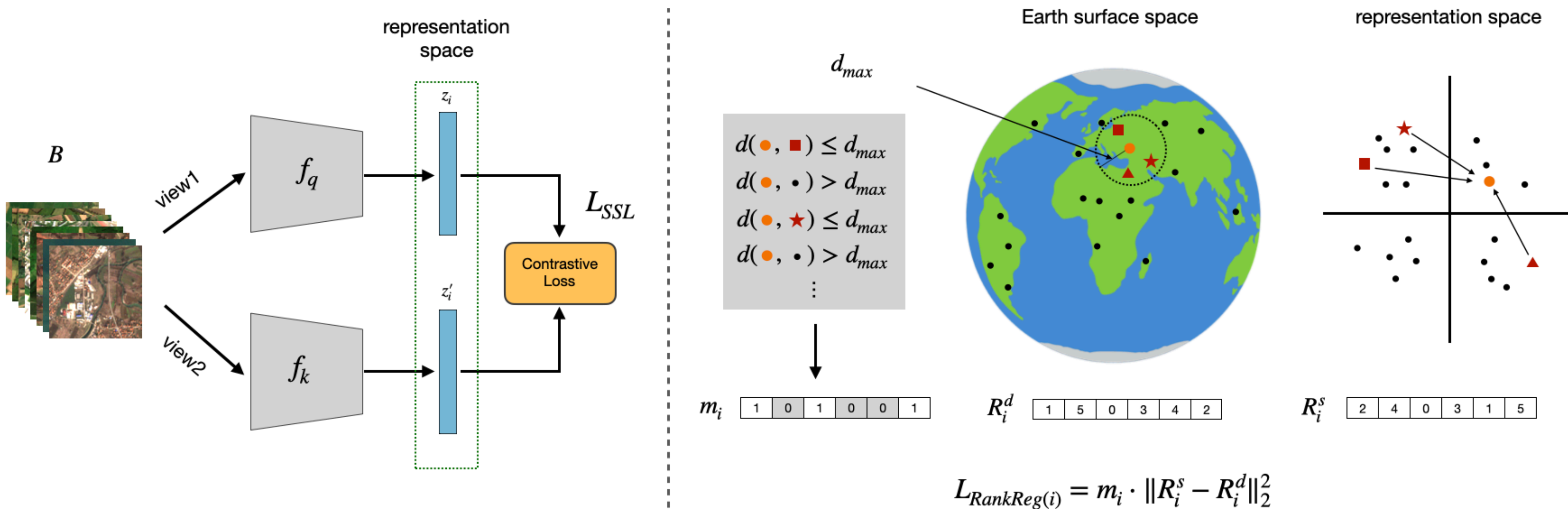
Limitations of SOTA:

- Rely on a two-stage approach, requiring precomputed K-Means clusters
- Use Euclidean distances, which poorly approximate Earth's spherical geometry
- Discretize spatial continuity via cluster assignments, ignoring fine-grained mid-range relationships



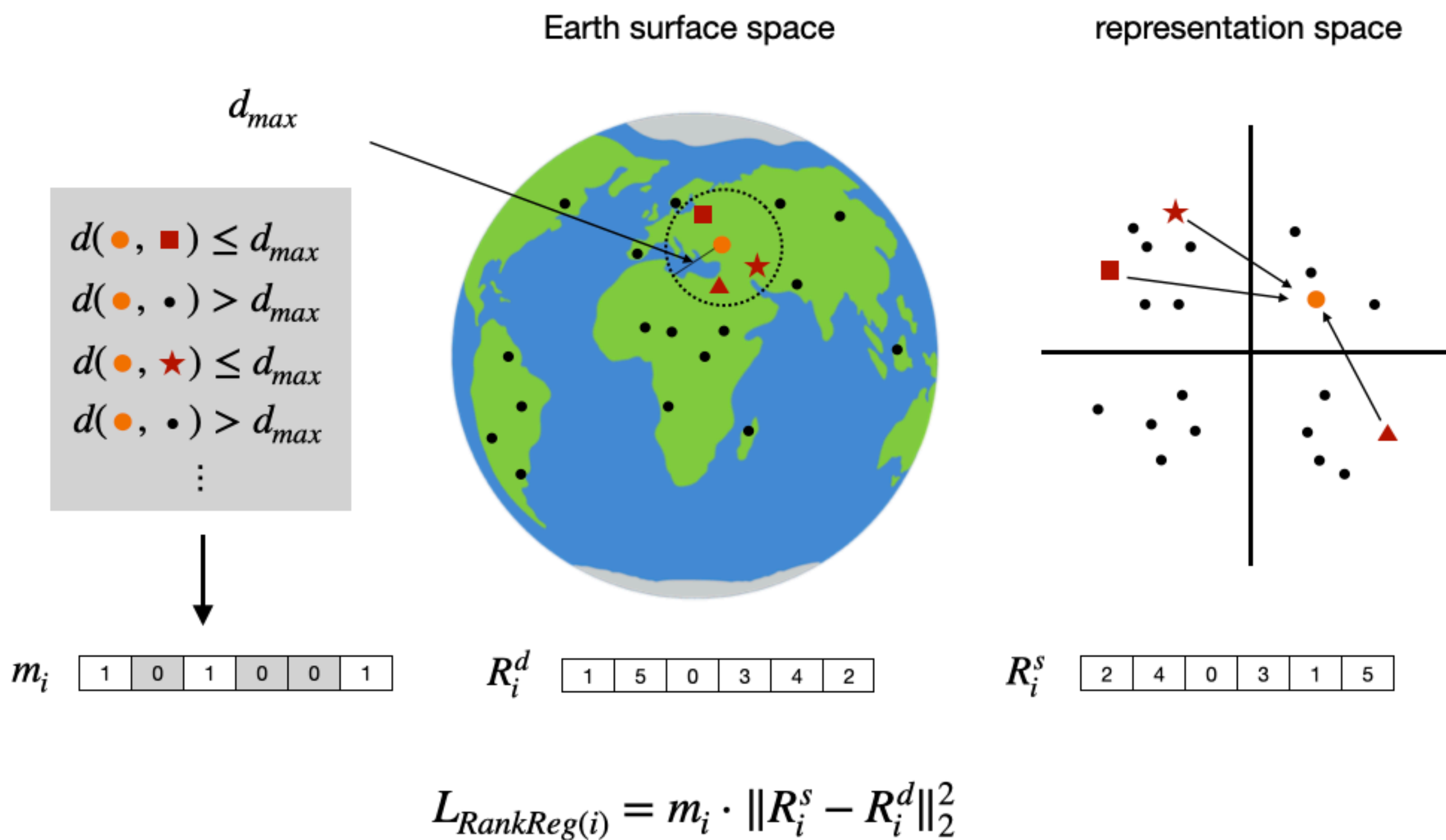
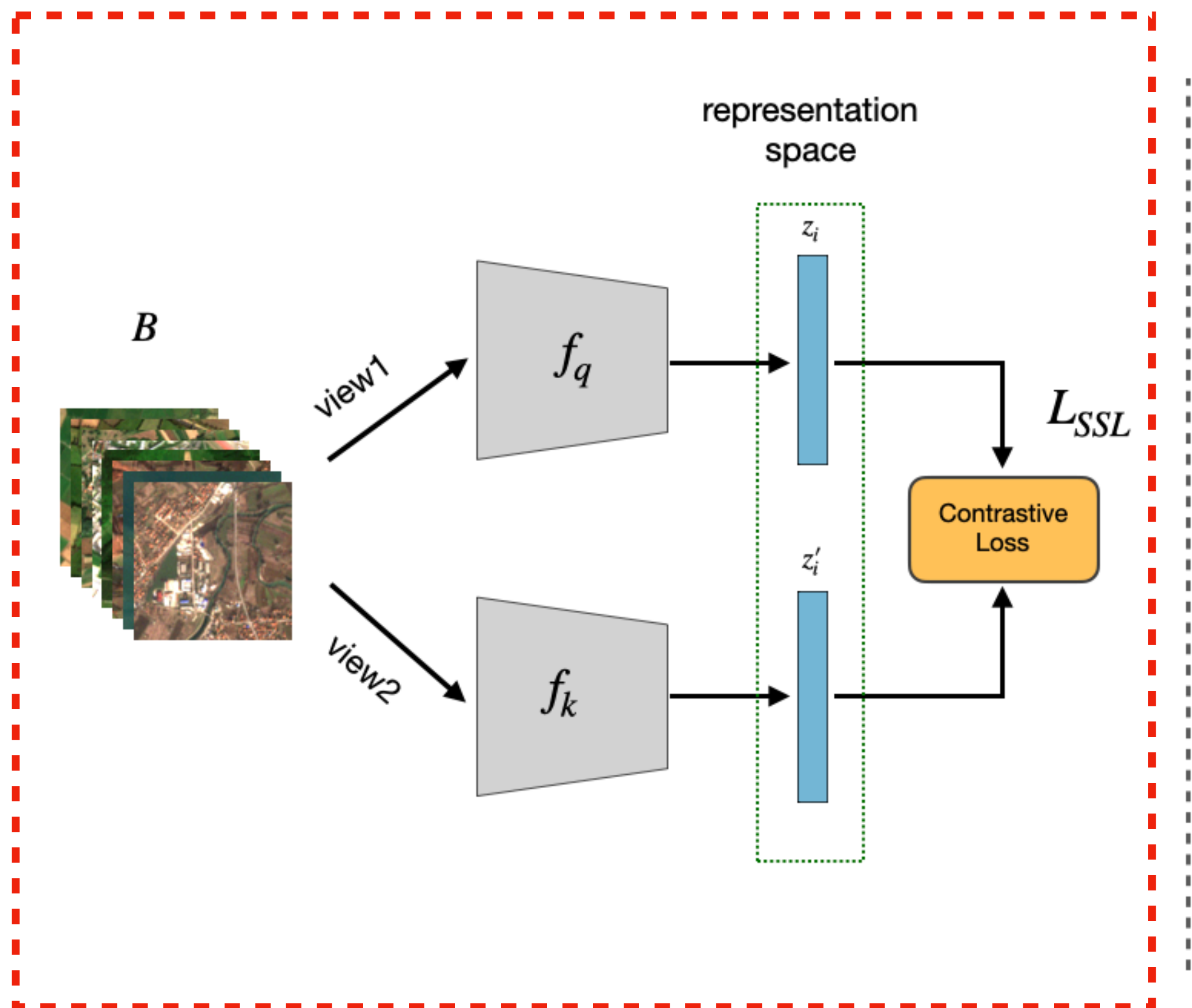
[1] K. Ayush et al., "Geography-aware self-supervised learning", in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

3. Method



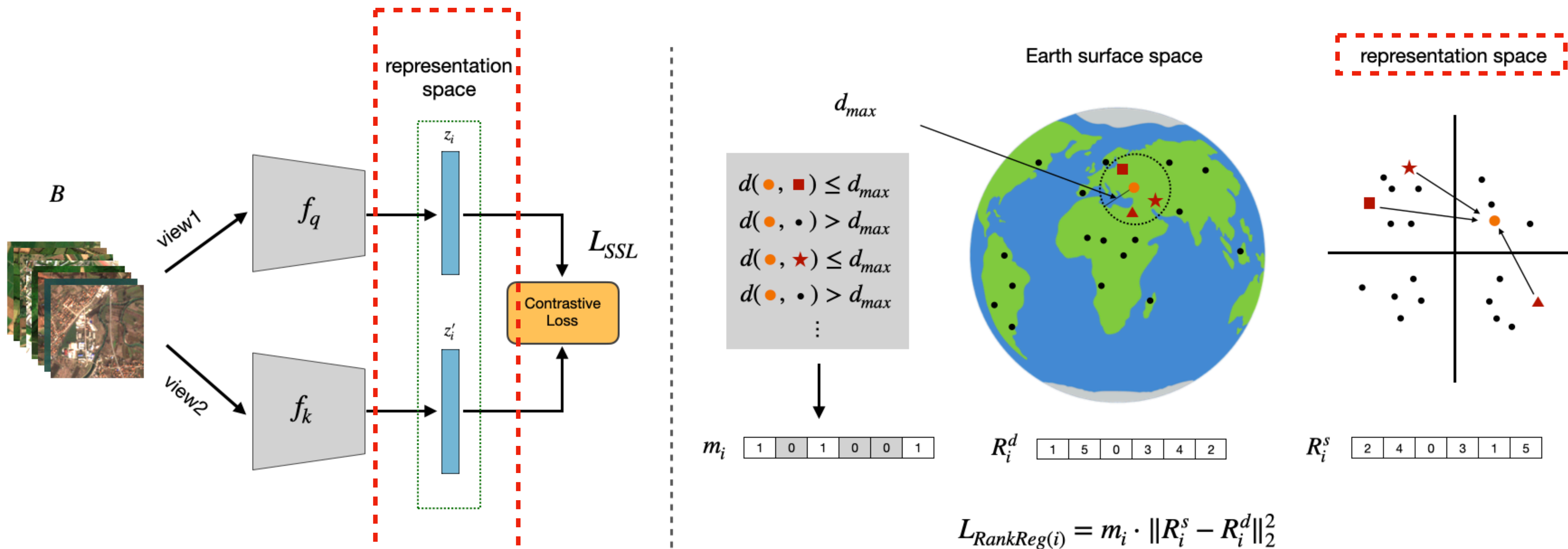
$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

3. Method



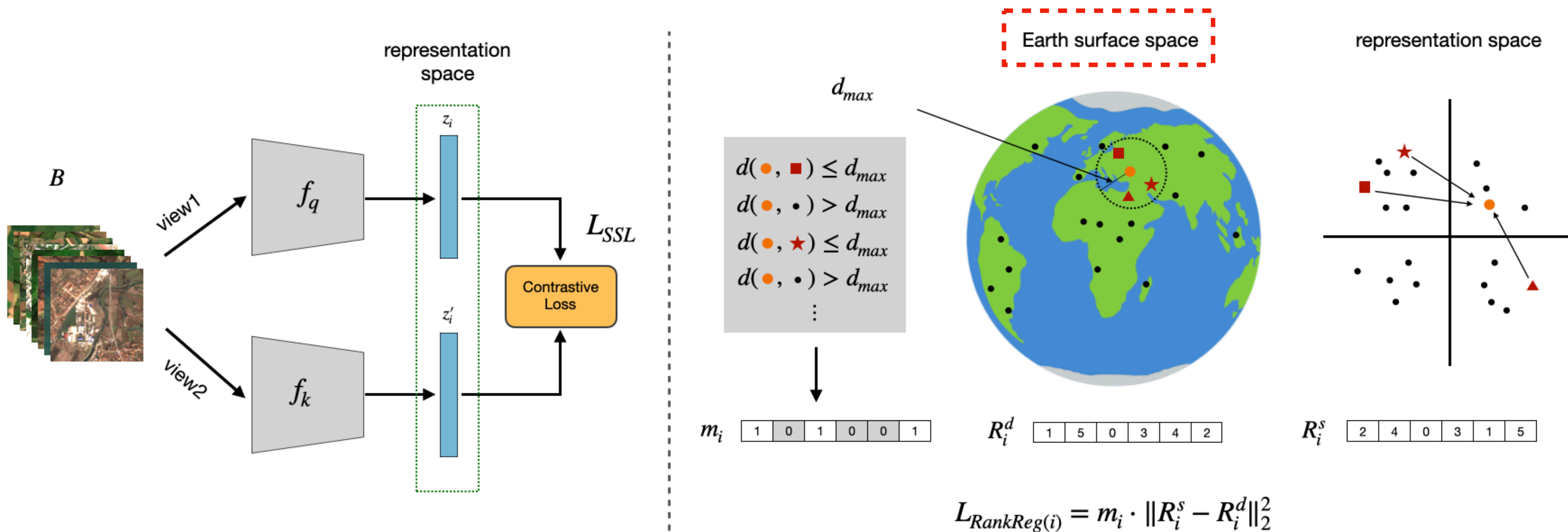
$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

3. Method



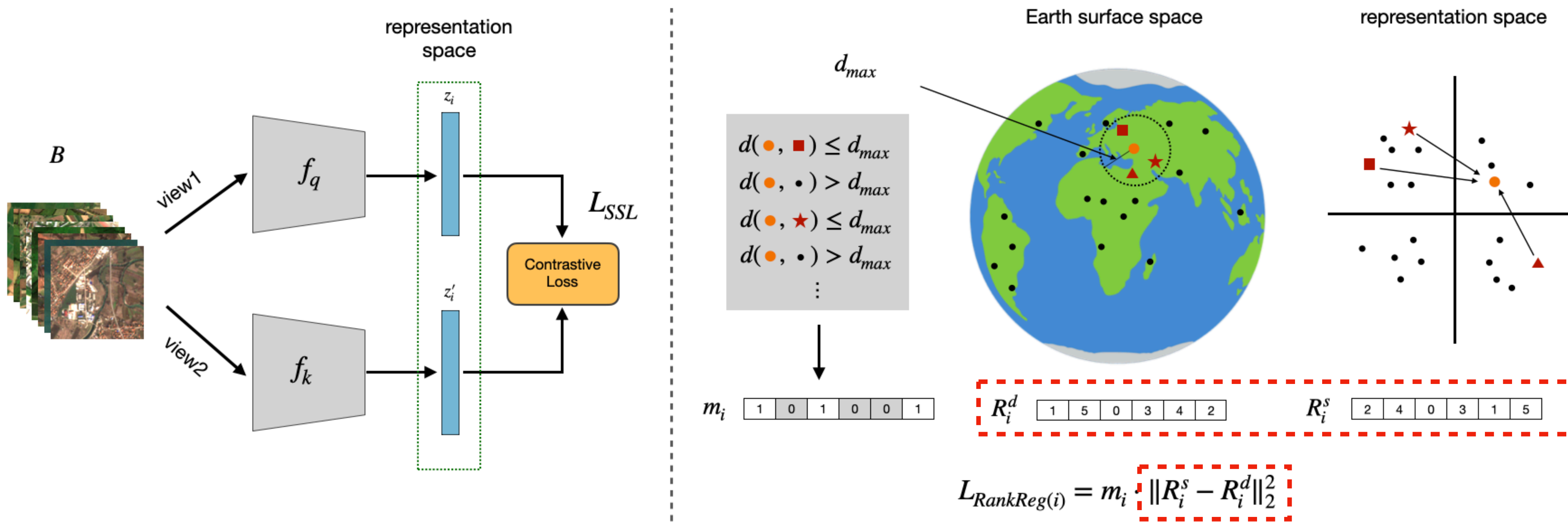
$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

3. Method



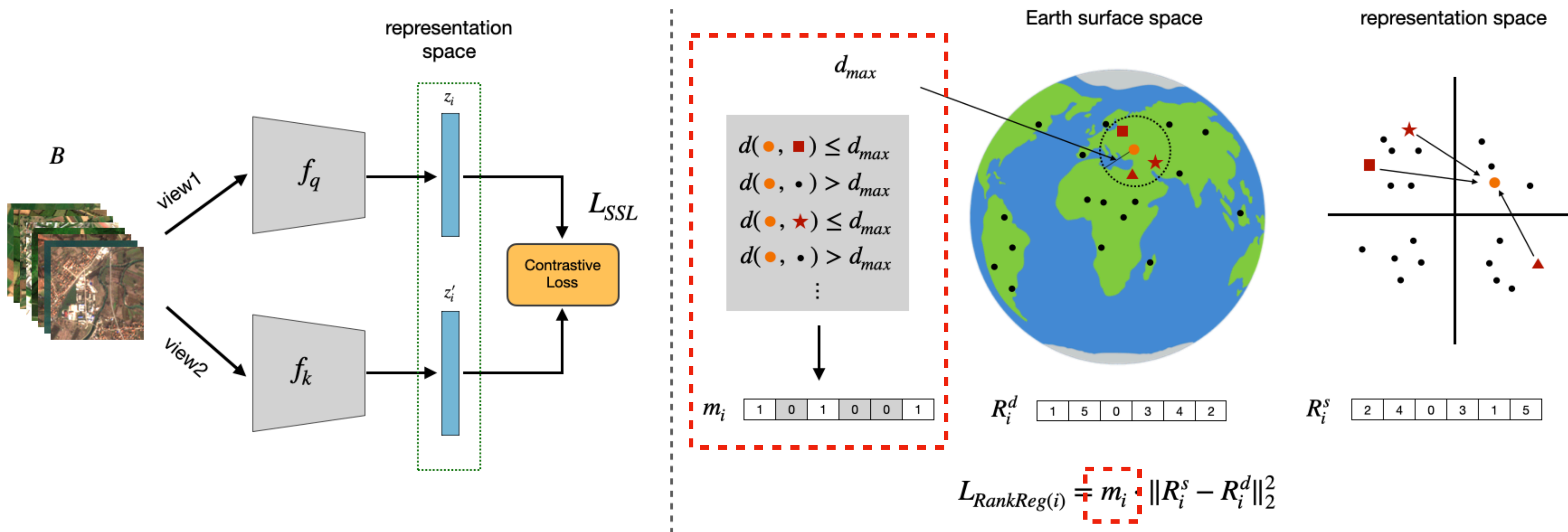
$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

3. Method



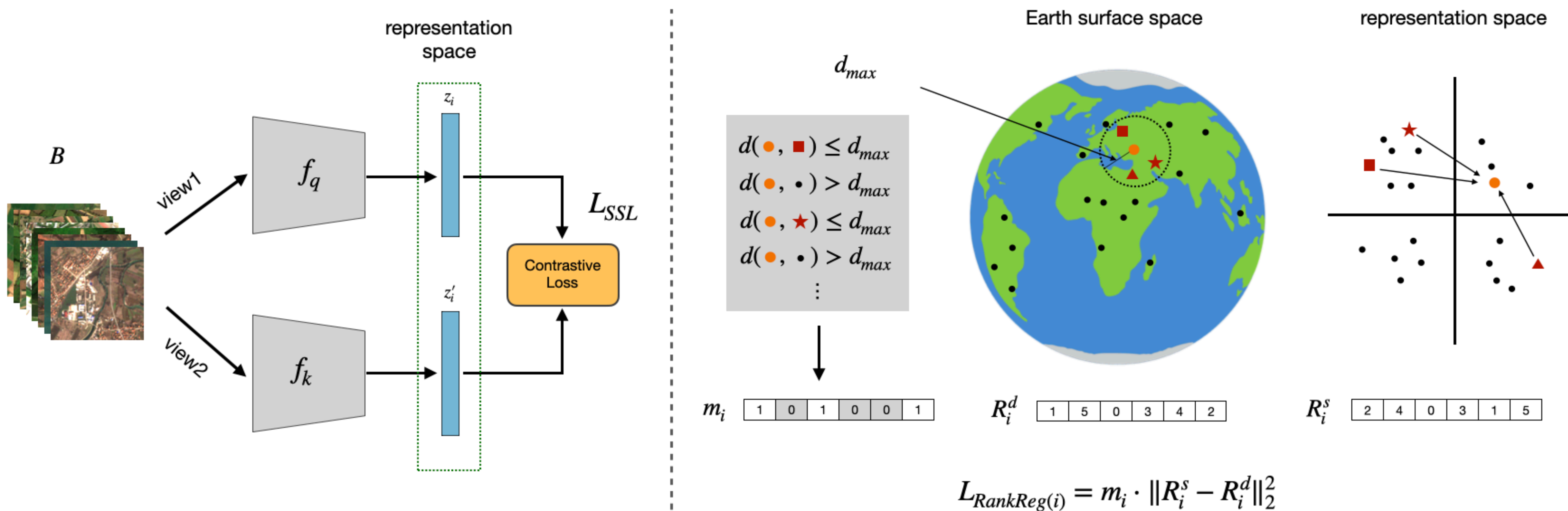
$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

3. Method



$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

3. Method



$$\mathcal{L}_{GeoRank} = \alpha \cdot \mathcal{L}_{SSL} + (1 - \alpha) \cdot \mathcal{L}_{RankReg}$$

4. Experiments: Setup

- 7 downstream tasks that include different geographical splits
- Metrics: Accuracy macro (Acc-mac) and mean average precision (AP-mac)
- k-NN evaluation protocol

- ResNet18 backbone
- MoCoV2 SSL algorithm
- Geometric augmentation only

Dataset	Task	#Images	Image Size	Location
SSL4EO [54]	-	~1000 k	264 × 264	Global
BEN-V2 [11]	MLC	~500 k	120 × 120	Europe
Sen4Agri-ML [48]	MLC	~40 k	120 × 120	Europe
EuroSAT [25]	SLC	~27 k	64 × 64	Europe
So2Sat [60]	SLC	~600 k	32 × 32	Global
CashewPlant [31]	SemSeg	~2 k	256 × 256	Africa

4.1 Experiments: GeoRank

- Compare GeoRank with a MoCoV2 baseline
- Compare GeoRank with existing contrastive SSL methods that integrate geographical metadata: Tile2Vec [2], Geoaware [1]
- Compare GeoRank with GeoBasic (only MSE instead of rank loss)
- Compare GeoRank with SOTA reconstruction-based RS SSL methods

[1] K. Ayush et al., "Geography-aware self-supervised learning", In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[2] N. Jean et al., "Tile2vec: Unsupervised representation learning for spatially distributed data", In Proceedings of the AAAI Conference on Artificial Intelligence

4.1 Experiments: GeoRank

Method	BEN-V2	EuroSAT	S4A-rand	S4A-tiles	So2Sat-rand	So2Sat-block	CashewPlant
Baseline (MoCoV2)	58.14	82.42	65.18	35.42	93.54	72.32	31.81
Tile2Vec [1]	54.95	73.39	63.10	35.20	81.67	66.14	34.40
Ayush et al. [2]	58.41	84.34	65.99	34.99	94.97	73.36	34.02
GeoBasic (ours)	57.86	82.05	65.22	35.26	93.01	71.50	32.26
GeoRank (ours)	59.19	85.09	65.91	35.17	94.95	73.46	34.94

4.1 Experiments: GeoRank

Method	BEN-V2	EuroSAT	S4A-rand	S4A-tiles	So2Sat-rand	So2Sat-block	CashewPlant
Baseline (MoCoV2)	58.14	82.42	65.18	35.42	93.54	72.32	31.81
Tile2Vec [1]	54.95	73.39	63.10	35.20	81.67	66.14	34.40
Ayush et al. [2]	58.41	84.34	65.99	34.99	94.97	73.36	34.02
GeoBasic (ours)	57.86	82.05	65.22	35.26	93.01	71.50	32.26
GeoRank (ours)	59.19	85.09	65.91	35.17	94.95	73.46	34.94

=> GeoRank outperforms or matches existing methods that integrate geographical information

4.1 Experiments: GeoRank

Method	BEN-V2	EuroSAT	S4A-rand	S4A-tiles	So2Sat-rand	So2Sat-block	CashewPlant
Baseline (MoCoV2)	58.14	82.42	65.18	35.42	93.54	72.32	31.81
Tile2Vec [1]	54.95	73.39	63.10	35.20	81.67	66.14	34.40
Ayush et al. [2]	58.41	84.34	65.99	34.99	94.97	73.36	34.02
GeoBasic (ours)	57.86	82.05	65.22	35.26	93.01	71.50	32.26
GeoRank (ours)	59.19	85.09	65.91	35.17	94.95	73.46	34.94

=> GeoRank outperforms or matches existing methods that integrate geographical information

=> Rank-based formulation improves over simple MSE formulation (GeoBasic)

4.1 Experiments: GeoRank

Method	BEN-V2	EuroSAT	S4A-rand	S4A-tiles	So2Sat-rand	So2Sat-block	CashewPlant
Baseline (MoCoV2)	58.14	82.42	65.18	35.42	93.54	72.32	31.81
Tile2Vec [1]	54.95	73.39	63.10	35.20	81.67	66.14	34.40
Ayush et al. [2]	58.41	84.34	65.99	34.99	94.97	73.36	34.02
GeoBasic (ours)	57.86	82.05	65.22	35.26	93.01	71.50	32.26
GeoRank (ours)	59.19	85.09	65.91	35.17	94.95	73.46	34.94
SatMAE	54.70	83.92	63.22	37.67	87.92	68.90	19.81
ScaleMAE	43.92	64.22	53.84	35.30	54.25	47.90	27.42
CrossScaleMAE	55.26	84.01	65.50	36.40	93.42	71.73	27.65

=> GeoRank outperforms or matches existing methods that integrate geographical information

=> Rank-based formulation improves over simple MSE formulation (GeoBasic)

=> GeoRank outperforms other reconstruction-based RS SSL methods if train and test set of downstream task have spatial proximity

4.1 Experiments: GeoRank

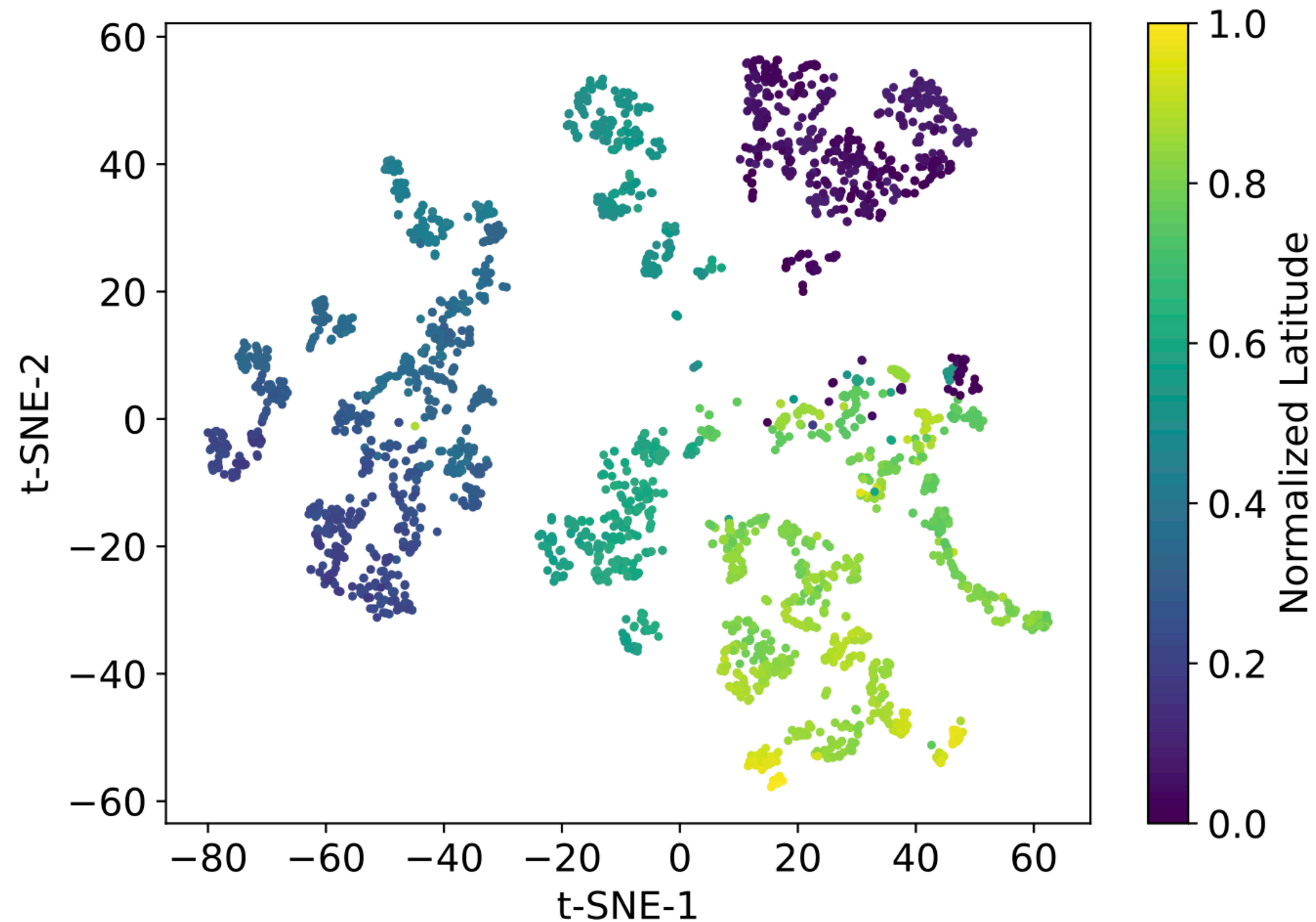
Extending different contrastive SSL algorithms

Contrastive SSL Algorithm	BEN-V2 <i>Baseline</i>	BEN-V2 <i>GeoRank</i>	EuroSAT <i>Baseline</i>	EuroSAT <i>GeoRank</i>	S4A-rand <i>Baseline</i>	S4A-rand <i>GeoRank</i>	So2Sat-rand <i>Baseline</i>	So2Sat-rand <i>GeoRank</i>
SimCLR	47.93	54.95	65.91	75.06	61.26	63.02	83.13	87.40
BYOL	50.40	57.08	67.62	79.66	59.51	64.75	77.59	81.09
SimSiam	43.53	57.53	52.14	80.10	54.42	64.51	64.20	91.35
DINO	56.61	58.22	81.49	81.87	63.72	65.67	90.54	93.63

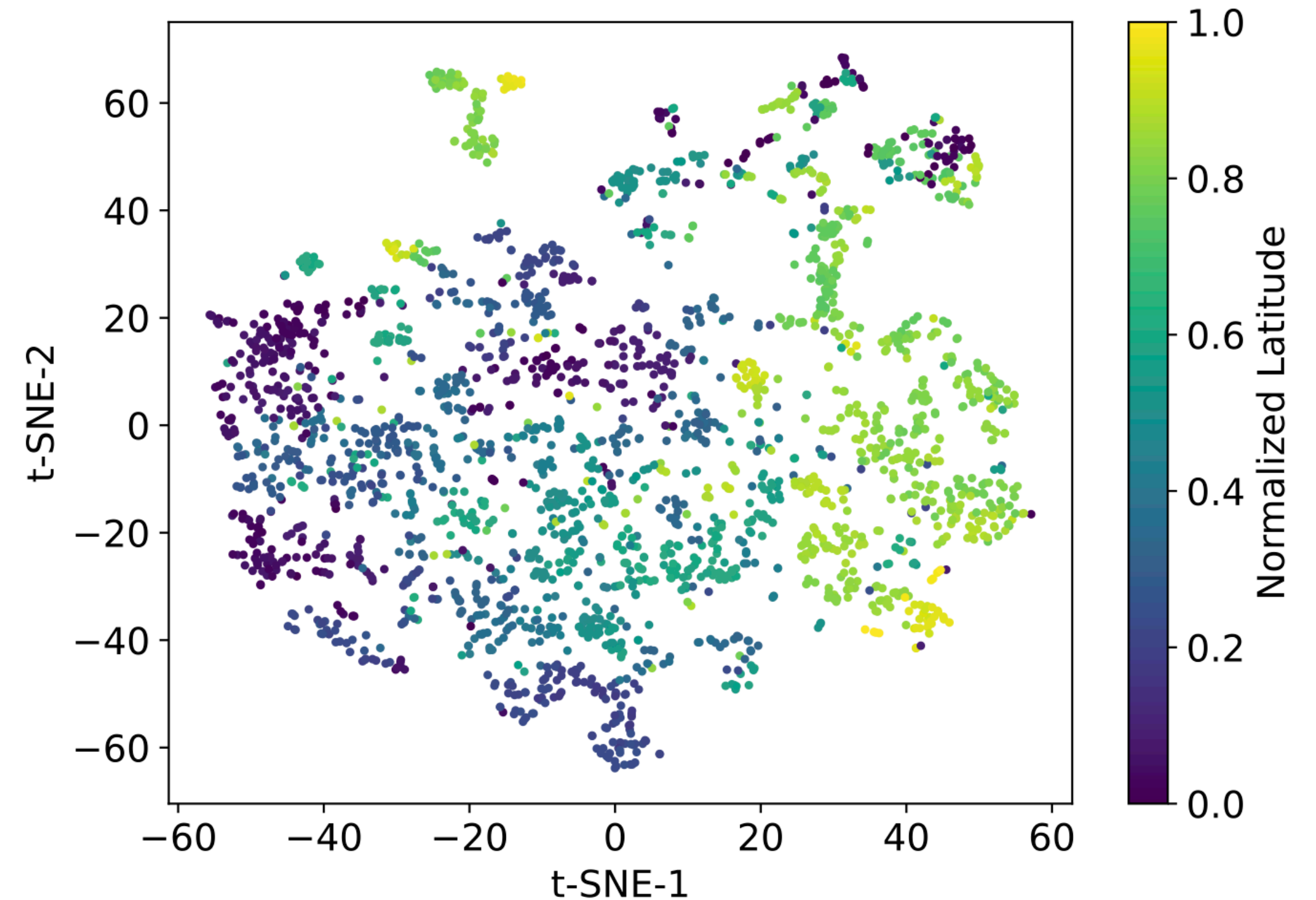
=> **GeoRank improves different contrastive SSL algorithms**

4.1 Experiments: GeoRank

a) Georank



b) Baseline



=> GeoRank separates feature space softly by geolocation

4.2 Experiments: Data Augmentation

- Common to many contrastive SSL works on multispectral RS images is the adoption of the standard set data augmentations from CV [3], [4], [5]
- We compare the standard set* of data augmentations with a simple geometric baseline**

* RandomResizeCrop, Flip, Contrast and Brightness, GaussianBlur, Grayscale

** RandomResizeCrop, RandomRotate90, Flip

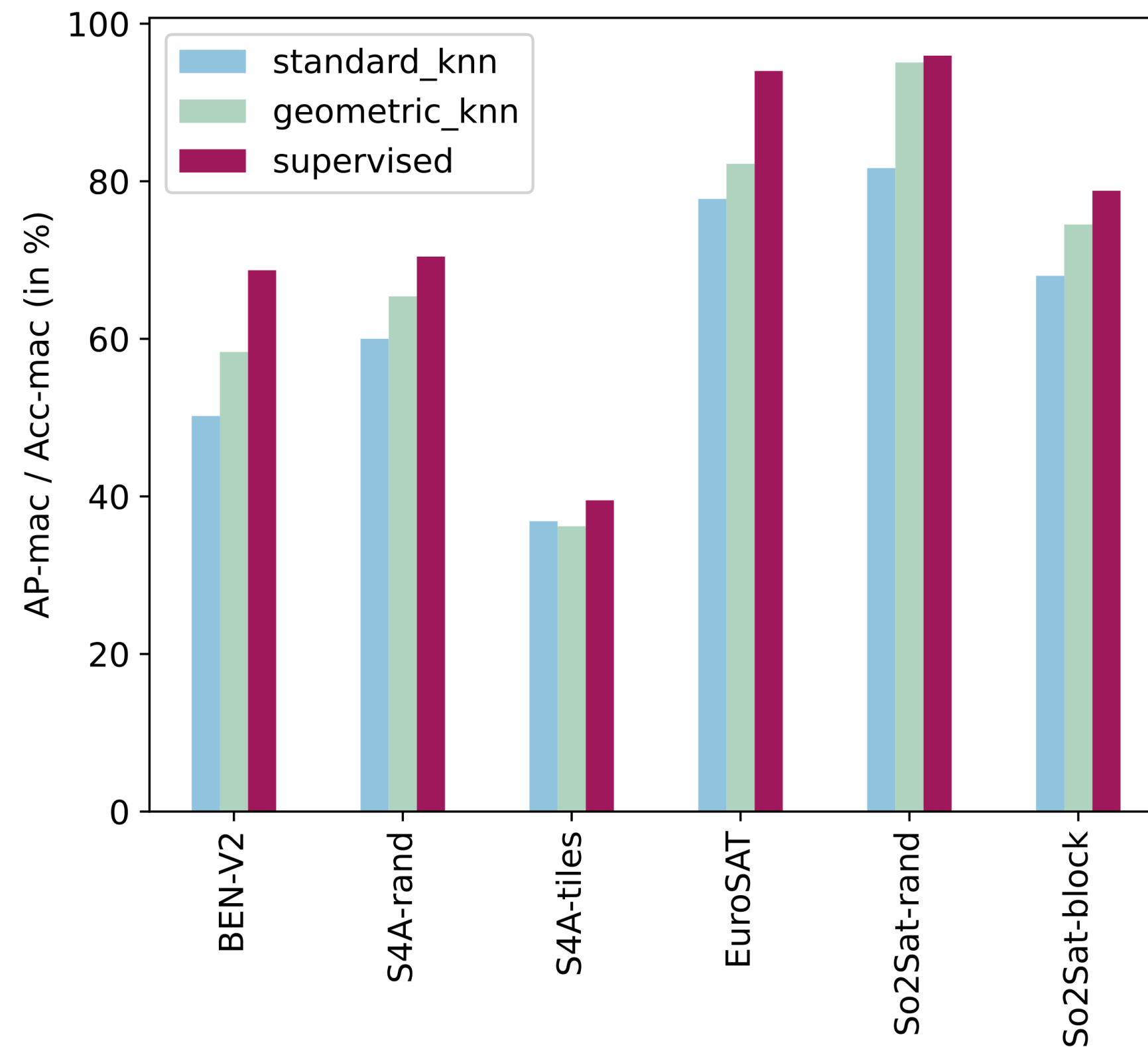
[3] H. Jung, "Contrastive Self-Supervised Learning With Smoothed Representation for Remote Sensing", IEEE Geoscience and Remote Sensing Letters, 2022.

[4] Y. Wang, "Self-supervised learning in remote sensing: A review", IEEE Geoscience and Remote Sensing Magazine, 2022.

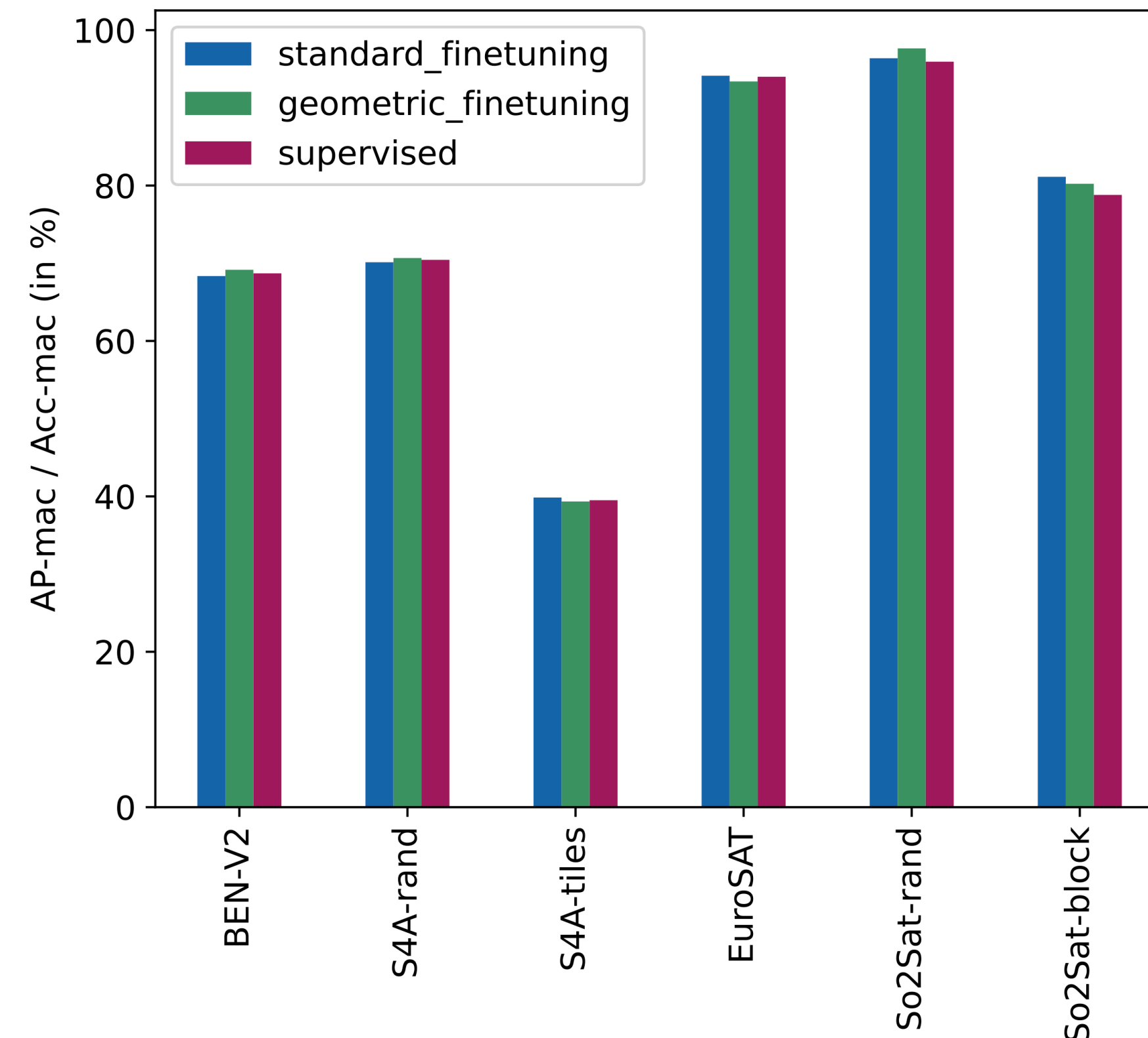
[5] Y. Wang et al., "A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation", IEEE Geoscience and Remote Sensing Magazine.

4.2 Experiments: Data Augmentation

a) k-NN



b) finetuning



=> k-NN evaluation shows large differences between models trained with standard augmentation and geometric only

4.2 Experiments: Dataset Cardinality

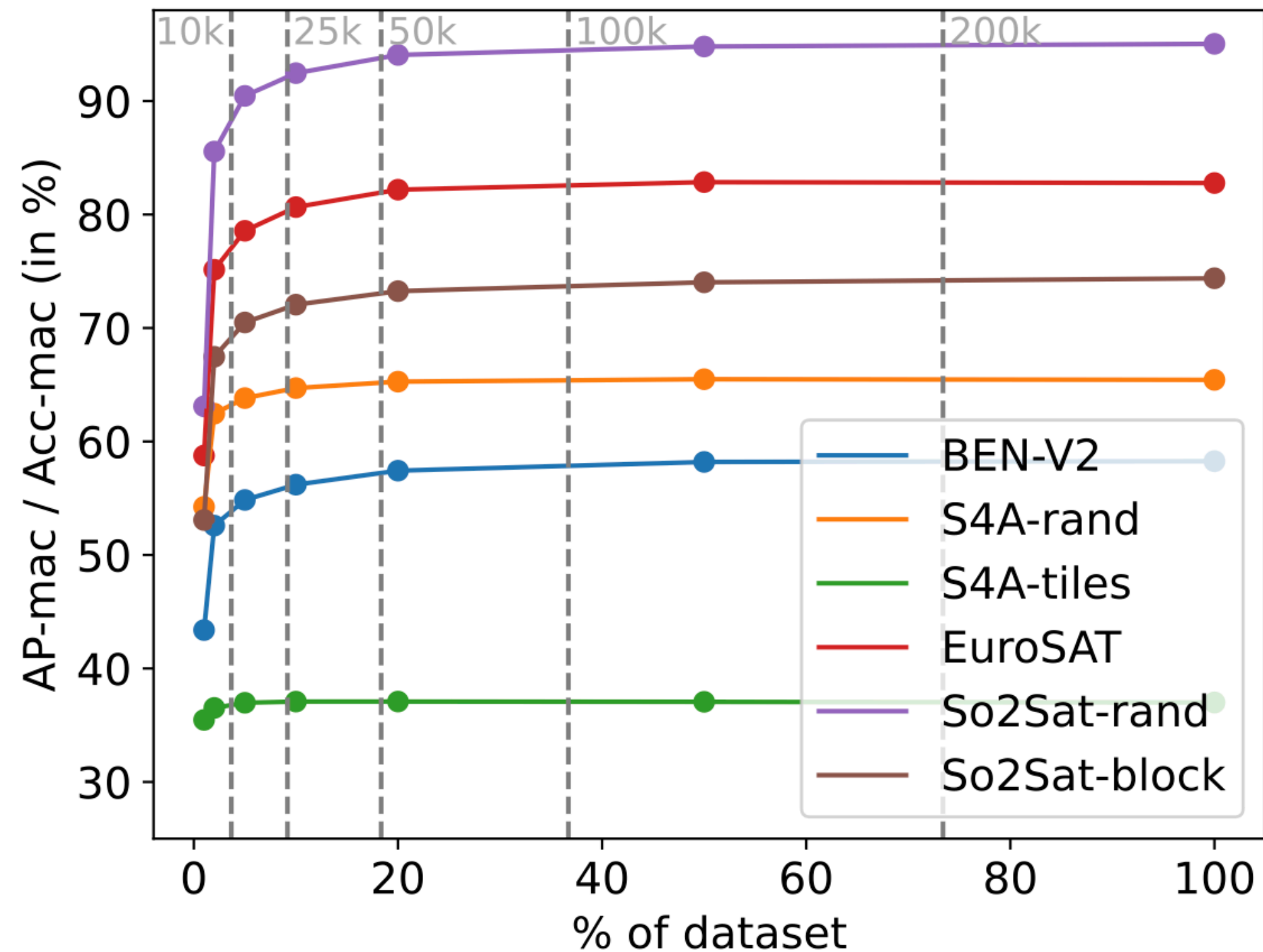
- Previous work only tested two pretraining dataset sizes (100k and 1000k) [6] or used imprecise experimental setup, e.g., different pre-processing levels between pretraining and downstream dataset [5]
- We compare downstream performance for 6 classification tasks when altering pretraining set size between 10k and 250k samples

[5] Y. Wang et al., "A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation", IEEE Geoscience and Remote Sensing Magazine.

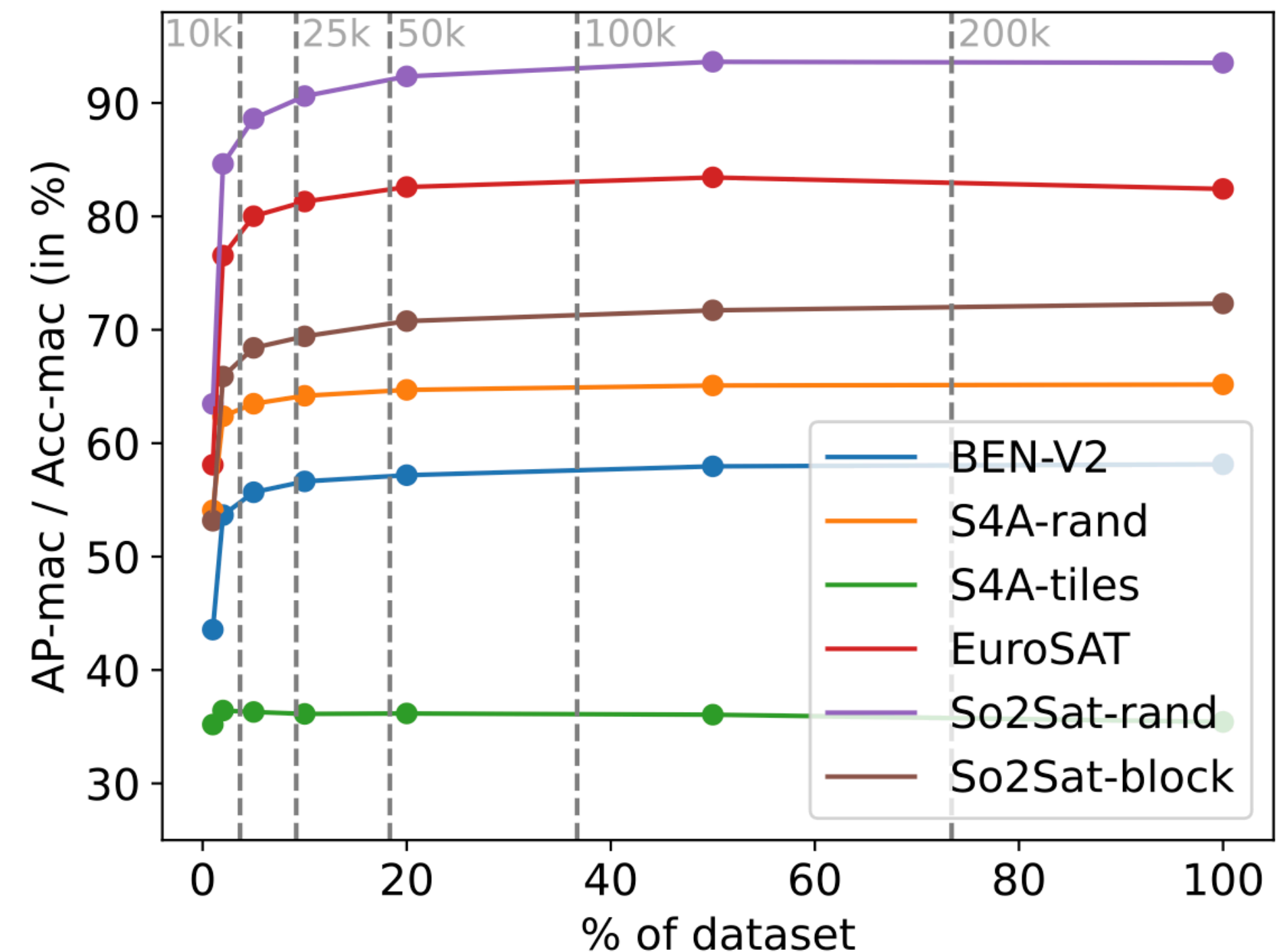
[6] O. Manas et al., "Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data", Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

4.2 Experiments: Dataset Cardinality

a) pretraining: SSL4EO



b) pretraining: BENV2



=> Downstream performance saturates around 100k pretraining samples for multispectral RS datasets

4.2 Experiments: Temporal Views & Image Size

Temporal Views

- Previous work suggests that using temporal views as augmentation universally improve contrastive SSL in RS [6]
=> we show that the experimental setup had limitations
=> we show that the effects of temporal views are task dependent

4.2 Experiments: Temporal Views & Image Size

Temporal Views

- Previous work suggests that using temporal views as augmentation universally improve contrastive SSL in RS [6]
=> we show that the experimental setup had limitations
=> we show that the effects of temporal views are task dependent

Image Size

- Existing multispectral pretraining RS datasets suggest that larger input image size (264x264) is universally beneficial following assumptions from CV [5], [6]
=> we show that pretraining input image size of 120x120 achieves similar performance while being 3x more efficient

[5] Y. Wang et al., "A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation", IEEE Geoscience and Remote Sensing Magazine.

[6] O. Manas et al., "Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data", Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

Takeaways

Takeaways

1. **Geographical Regularization:** GeoRank outperforms SOTA methods that integrate geographical metadata and improves different contrastive SSL algorithms

Takeaways

1. **Geographical Regularization:** GeoRank outperforms SOTA methods that integrate geographical metadata and improves different contrastive SSL algorithms
2. **Data Augmentation:** Standard CV augmentations are suboptimal for multispectral RS SSL

Takeaways

1. **Geographical Regularization:** GeoRank outperforms SOTA methods that integrate geographical metadata and improves different contrastive SSL algorithms
2. **Data Augmentation:** Standard CV augmentations are suboptimal for multispectral RS SSL
3. **Dataset Cardinality:** Downstream performance saturates around 100k multispectral pretraining samples

Takeaways

1. **Geographical Regularization:** GeoRank outperforms SOTA methods that integrate geographical metadata and improves different contrastive SSL algorithms
2. **Data Augmentation:** Standard CV augmentations are suboptimal for multispectral RS SSL
3. **Dataset Cardinality:** Downstream performance saturates around 100k multispectral pretraining samples
4. **Temporal Views:** Beneficial effects are task-dependent

Takeaways

1. **Geographical Regularization:** GeoRank outperforms SOTA methods that integrate geographical metadata and improves different contrastive SSL algorithms
2. **Data Augmentation:** Standard CV augmentations are suboptimal for multispectral RS SSL
3. **Dataset Cardinality:** Downstream performance saturates around 100k multispectral pretraining samples
4. **Temporal Views:** Beneficial effects are task-dependent
5. **Image Size:** Pretraining on smaller image crops retains performance with 3× faster training

Takeaways

1. **Geographical Regularization:** GeoRank outperforms SOTA methods that integrate geographical metadata and improves different contrastive SSL algorithms
2. **Data Augmentation:** Standard CV augmentations are suboptimal for multispectral RS SSL
3. **Dataset Cardinality:** Downstream performance saturates around 100k multispectral pretraining samples
4. **Temporal Views:** Beneficial effects are task-dependent
5. **Image Size:** Pretraining on smaller image crops retains performance with 3× faster training

Limitation: Benefits of GeoRank (and other geo-aware methods) under geographical domain shift are limited.

Thank You For Your Attention!



Contact: t.burgert@tu-berlin.de