CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations

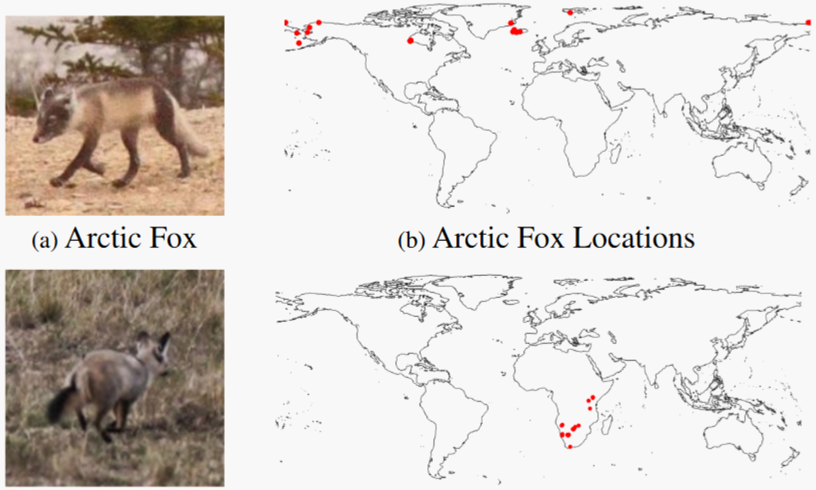
Wuhan University

Chenglong Wang

Mai G, Lao N, He Y, et al. CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations[J]. arXiv preprint arXiv:2305.01118, 2023.

Related work

The importance of geospatial information

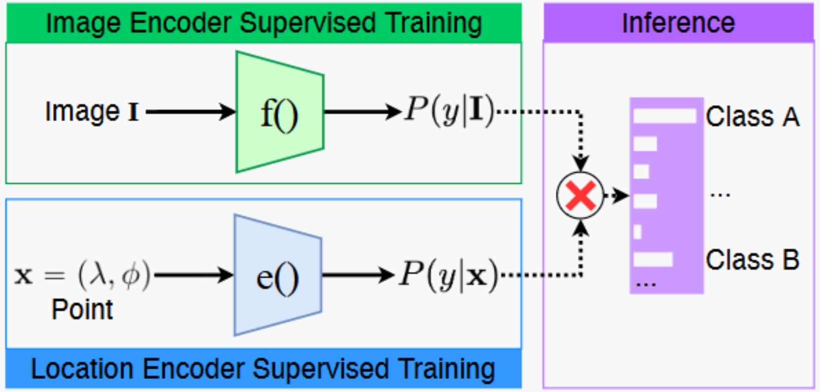


(a) Arctic Fox (b) Arctic Fox Locations

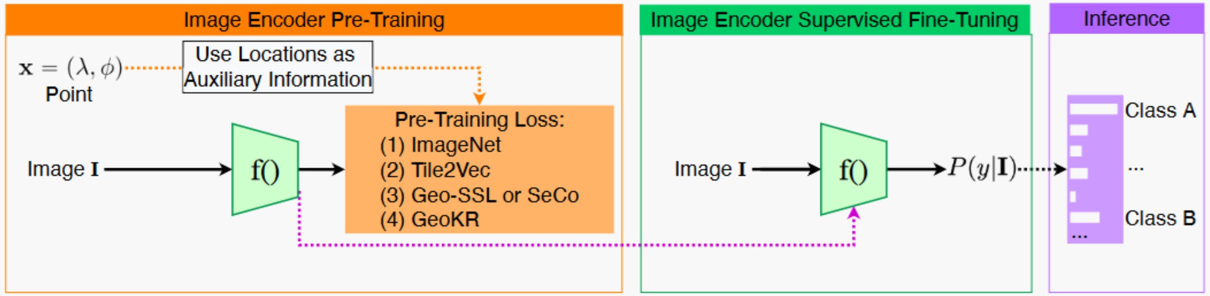
(c) Bat-Eared Fox (d) Bat-Eared Fox Locations

These two species have distinct geospatial distribution patterns, and it is very easy to tell them apart based on the geo-locations.

Geo-aware **Supervised** Learning



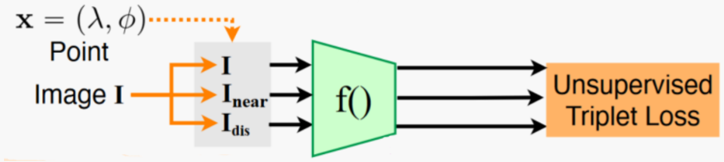
Pretrain – Finetune (using locations as auxiliary information)

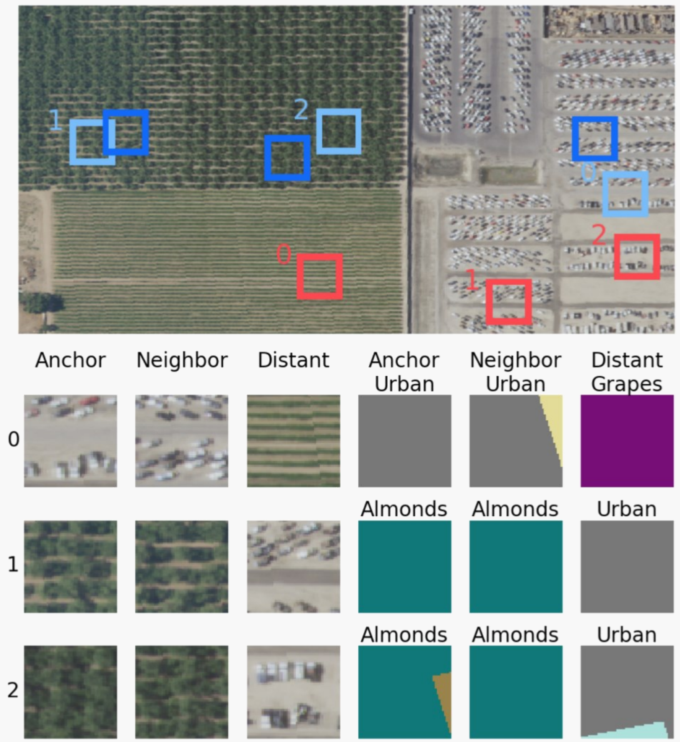


Background – Methodology (Pre-train – Fine-tune) – Future work

Related work

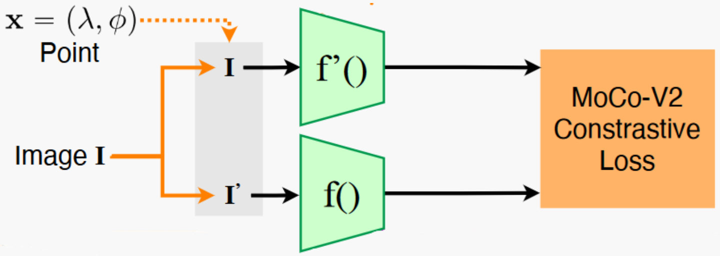
Pretrain – Finetune (Tile2Vec)

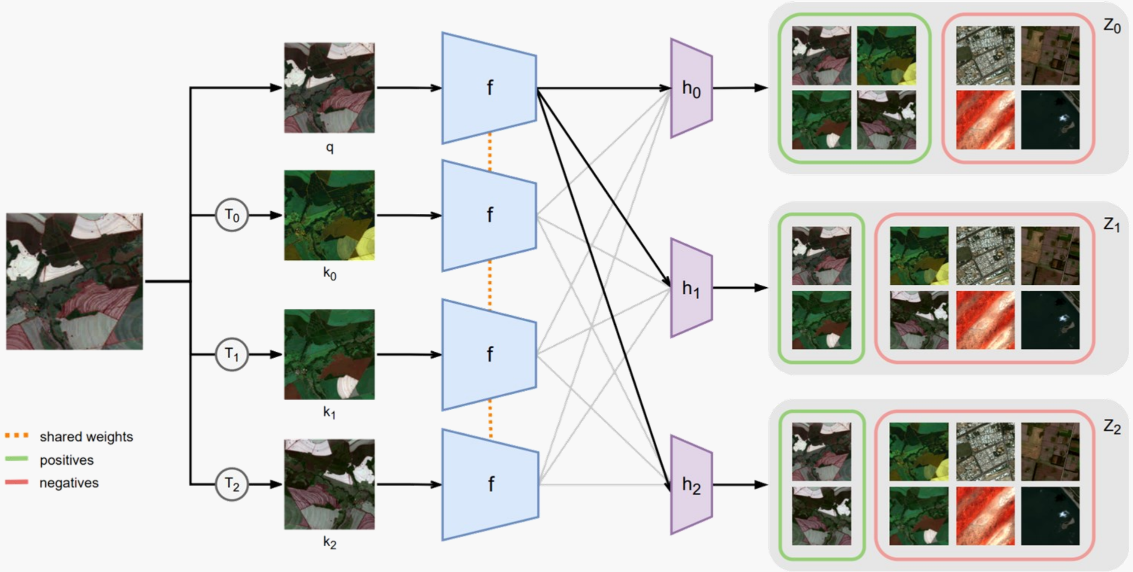




Anchor – Neighbor – Distant

Pretrain – Finetune (SeCo)



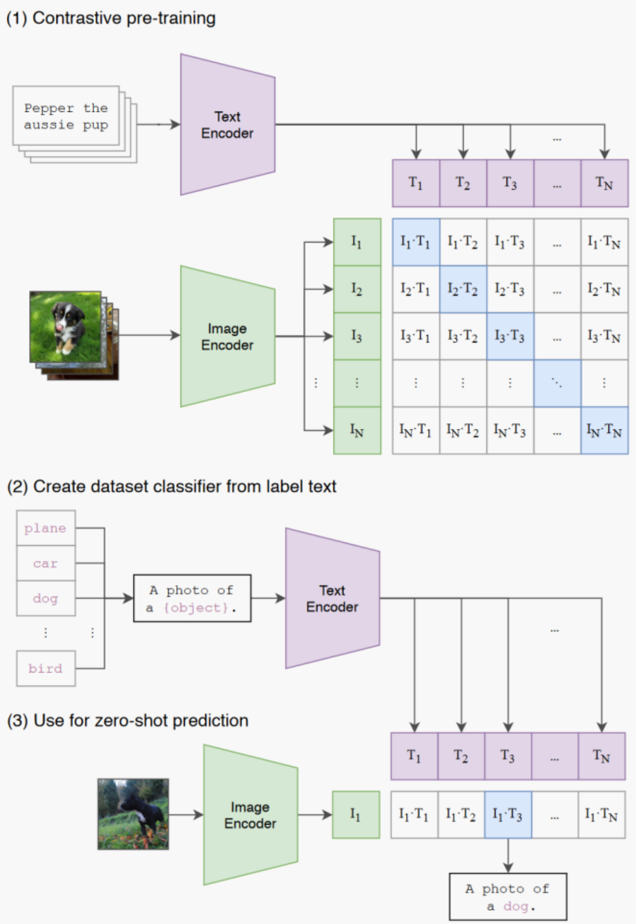


MoCo: temporal & artificial augmentations

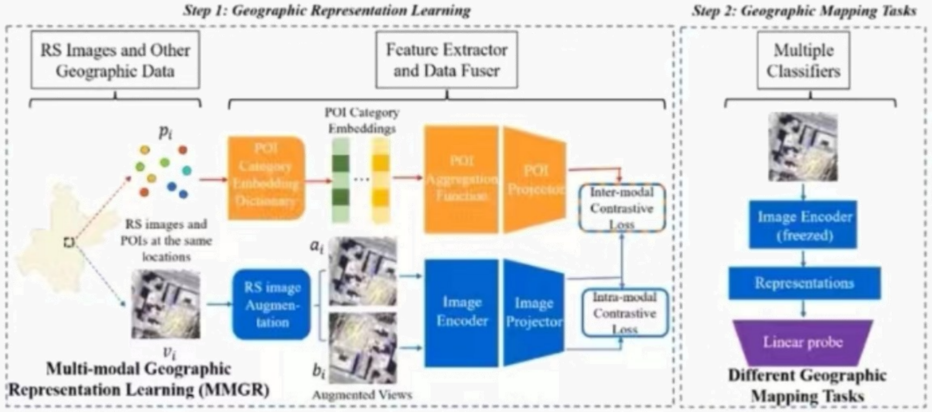
Background – Methodology (Pre-train – Fine-tune) – Future work

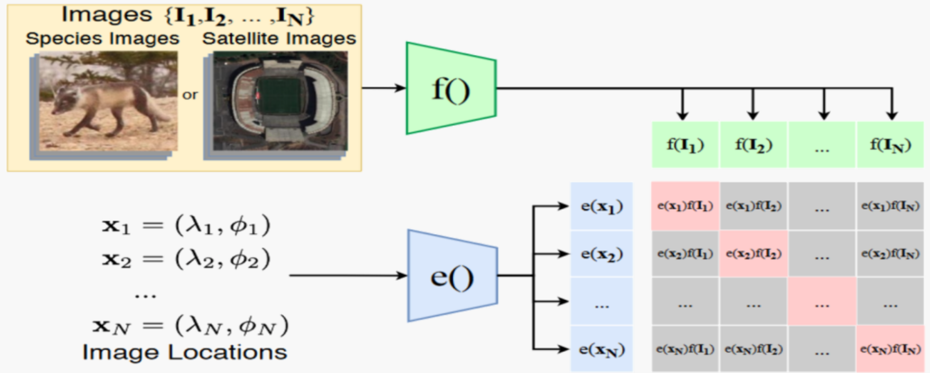
Overview （Interaction）

CLIP



CLIP in GIScience



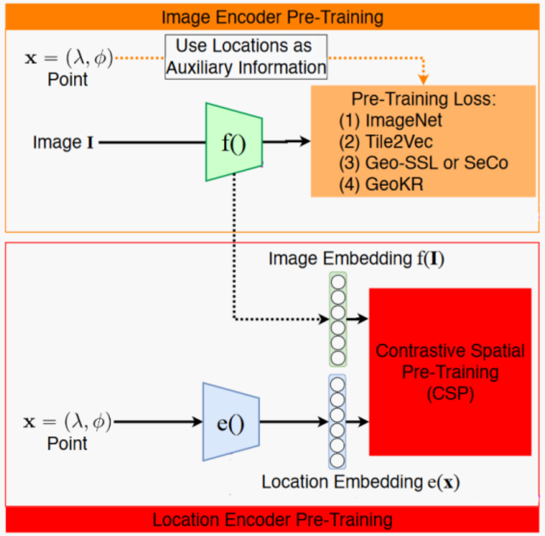


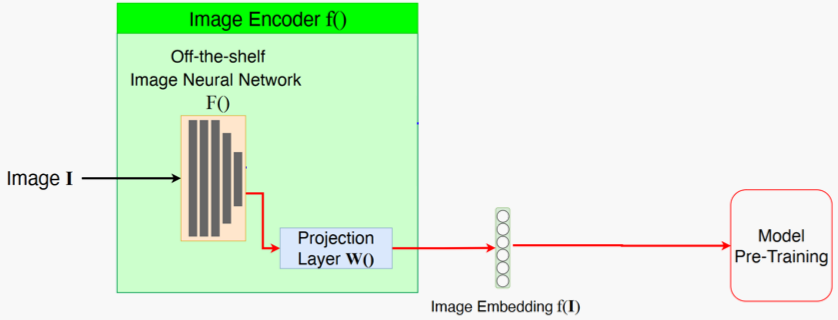
Potential problem

The number of trainable parameters of the image encoder f() is 100 times larger than that of the location encoder e()

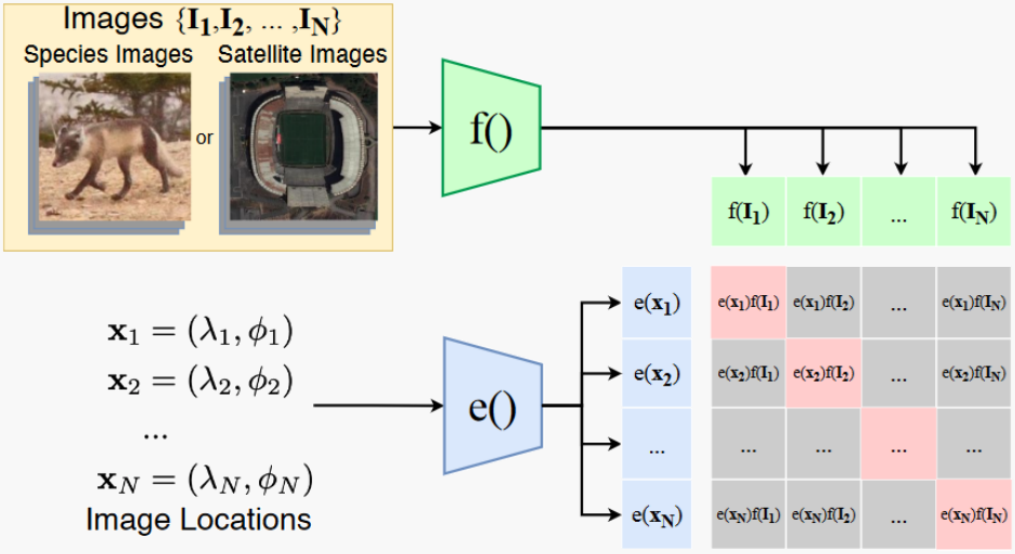
Pre-train

Architecture





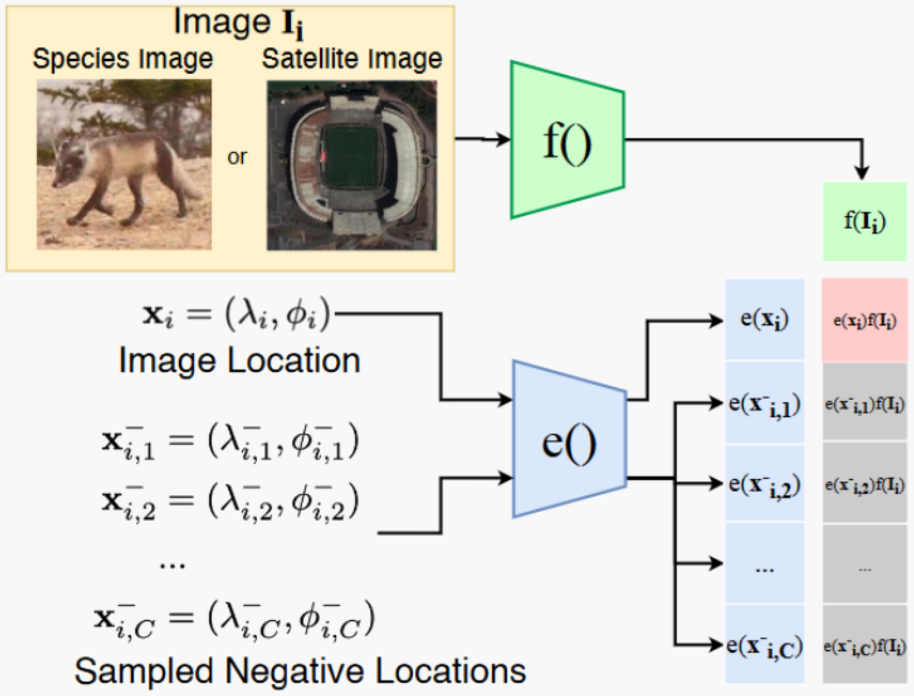
Training Pair Construction



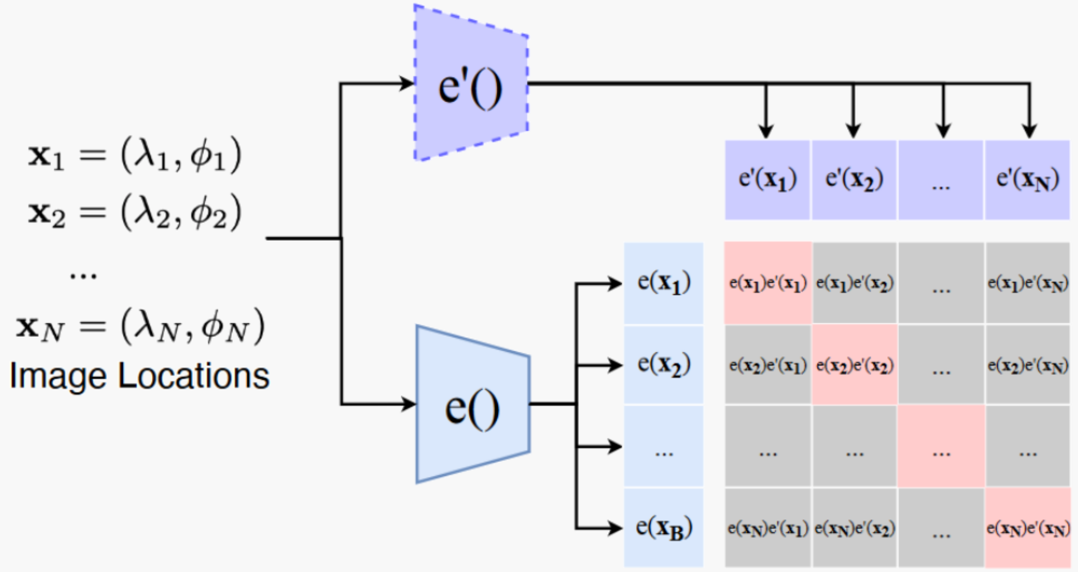
(a) In-batch negative sampling

Pre-train

- Train Pair Construction



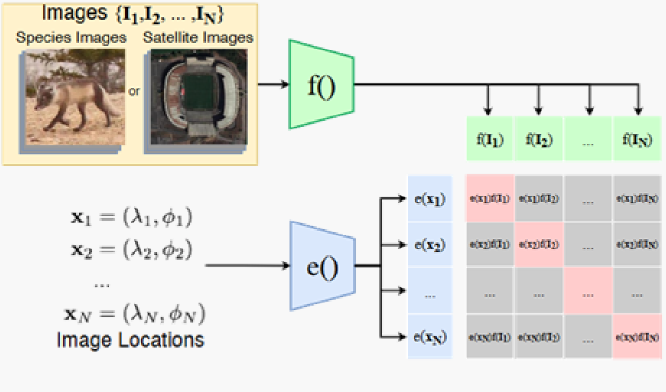
(b) Random negative location sampling

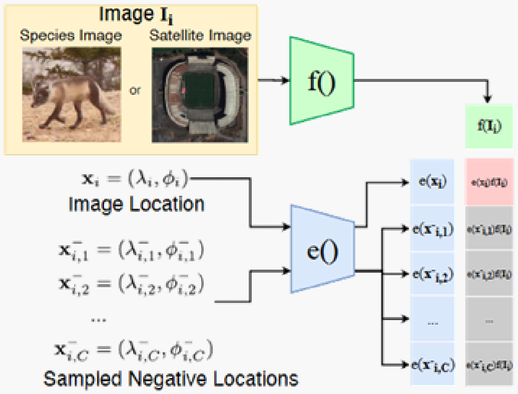


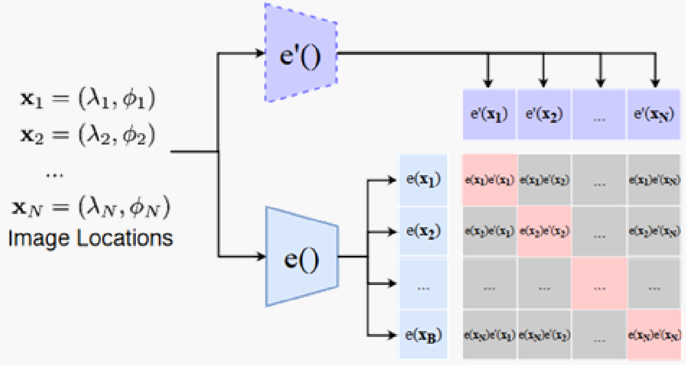
(c) SimCSE sampling

Pre-train

Contrastive Learning Objectives







1. Noise-Contrastive Estimation:

$$l_{\mathrm{NCE}}(\mathcal{P},\mathcal{N}) = -\,\mathbb{E}_{(\mathbf{a},\mathbf{b})\sim\mathcal{P}}\log\sigma(s(\mathbf{a},\mathbf{b})) \\ -\,\mathbb{E}_{(\mathbf{a},\mathbf{b}^-)\sim\mathcal{N}}\log(1-\sigma(s(\mathbf{a},\mathbf{b}^-)))$$

2. InfoNCE (MC):

$$l_{\mathrm{MC}}(\mathcal{P},\mathcal{N},\tau) \\ = \mathbb{E}_{(\mathbf{a},\mathbf{b})\sim\mathcal{P}}\frac{e^{s(\mathbf{a},\mathbf{b})/\tau}}{e^{s(\mathbf{a},\mathbf{b})/\tau}+\sum_{(\mathbf{a},\mathbf{b}^-)\in\mathcal{N}_{\mathbf{a}}}e^{s(\mathbf{a},\mathbf{b}^-)/\tau}}$$

1. self-supervised binary (NCE) loss:

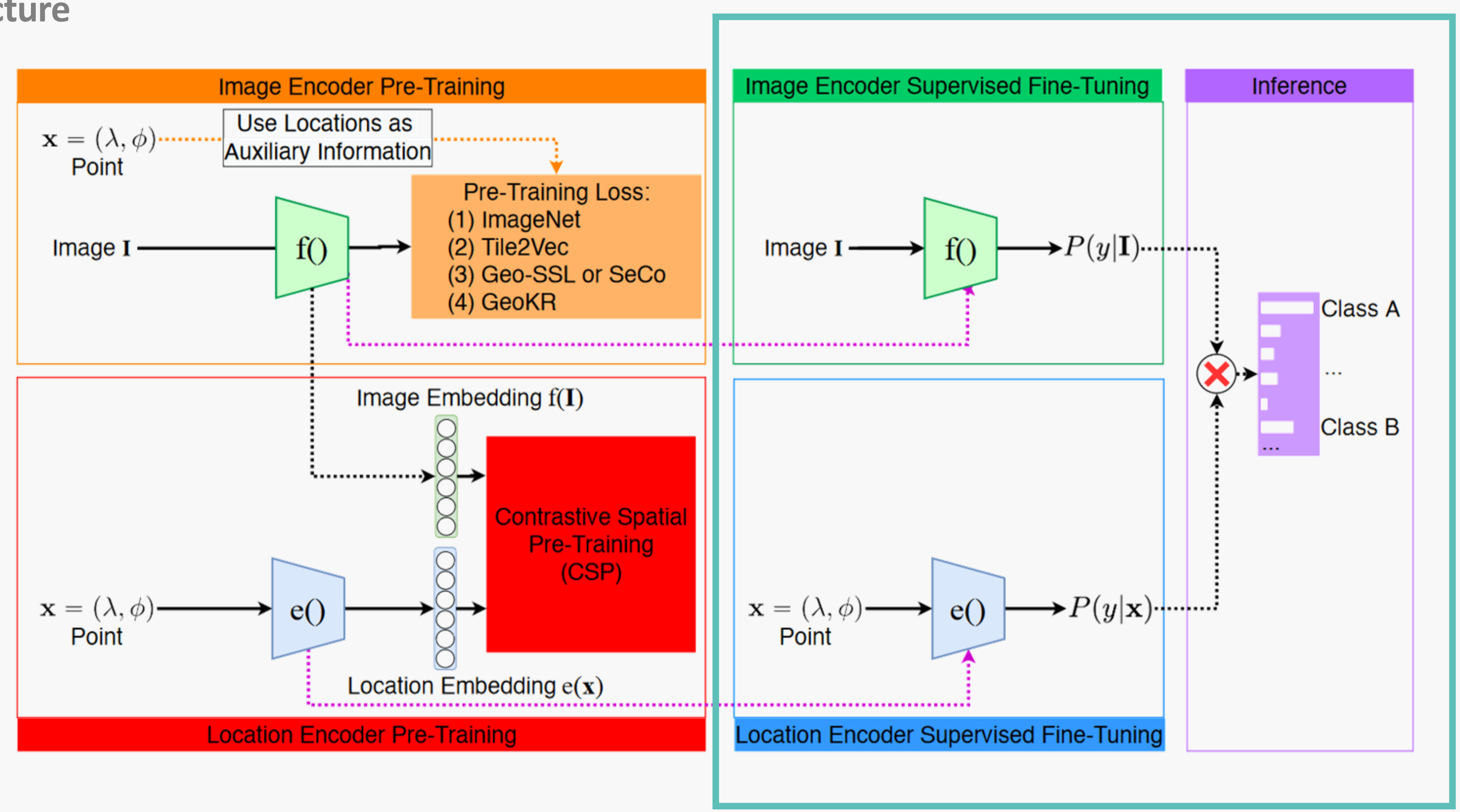
$$\begin{aligned} l_{\mathrm{NCE}}(\mathbb{X}) &= l^B_{\mathrm{NCE}}(\mathbb{X}) + \beta_1 l^L_{\mathrm{NCE}}(\mathbb{X}) + \beta_2 l^D_{\mathrm{NCE}}(\mathbb{X}) \\ &= l_{\mathrm{NCE}}(\mathcal{P}^X,\mathcal{N}^B) + \beta_1 l_{\mathrm{NCE}}(\emptyset,\mathcal{N}^L) \\ &+ \beta_2 l_{\mathrm{NCE}}(\mathcal{P}^D,\mathcal{N}^D) \end{aligned}$$

2. self-supervised multi-class (MC) loss:

$$\begin{aligned} l_{\mathrm{MC}}(\mathbb{X}) =& l^B_{\mathrm{MC}}(\mathbb{X}) + \alpha_1 l^L_{\mathrm{MC}}(\mathbb{X}) + \alpha_2 l^D_{\mathrm{MC}}(\mathbb{X}) \\ =& l_{\mathrm{MC}}(\mathcal{P}^X,\mathcal{N}^B,\tau_0) + \alpha_1 l_{\mathrm{MC}}(\mathcal{P}^X,\mathcal{N}^L,\tau_1) \\ &+ \alpha_2 l_{\mathrm{MC}}(\mathcal{P}^D,\mathcal{N}^D,\tau_2) \end{aligned}$$

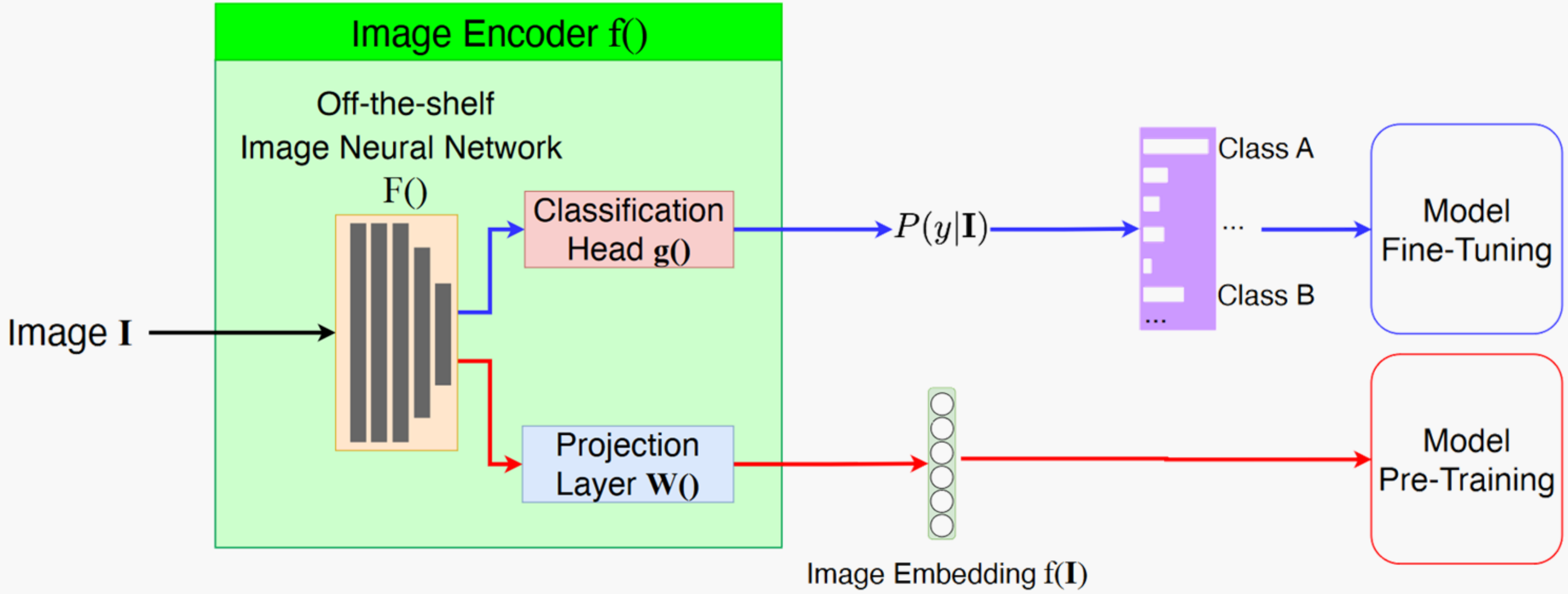
Fine-tune

- Architecture



Fine-tune

- Image Fine-tuning



Background – Methodology (Pre-train – Fine-tune) – Future work

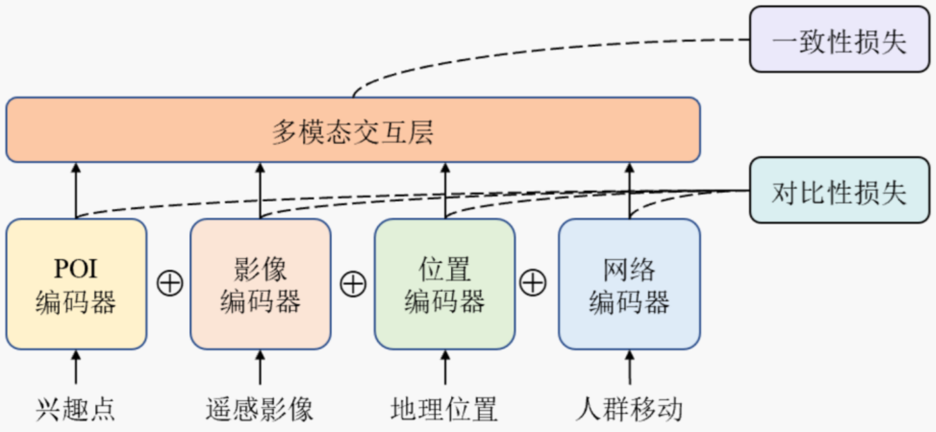
Multimodality?

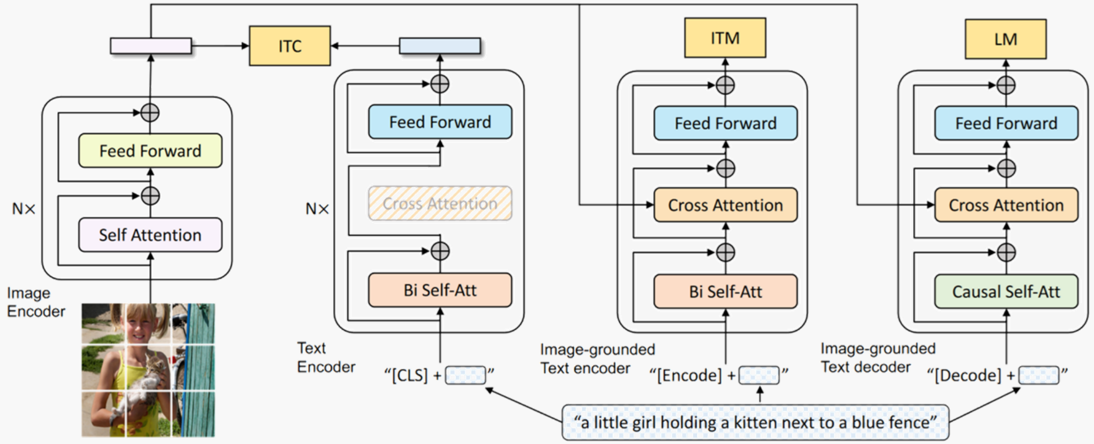
Although we only investigate the effectiveness of our CSP framework on location-image pre-training in this work, CSP can be easily extended to learn the alignment between location (or time) and data in other modalities such as text for different downstream tasks such as geo-aware text classification.

Location?

In the future, we can explore more complex geometries such as polylines (Xu et al., 2018) and polygons (Mai et al., 2023b). The proposed CSP framework can be seen as a step towards the geo-aware foundation models (Mai et al., 2022a; 2023a).

More Interaction?





BLIP(ALBEF+VLMO)

CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations

Wuhan University

Chenglong Wang

Mai G, Lao N, He Y, et al. CSP: Self-Supervised Contrastive Spatial Pre-Training for Geospatial-Visual Representations[J]. arXiv preprint arXiv:2305.01118, 2023.