CVPR JUNE 3-7, 2026 DENVER COLORADO

EARTH VISION

UniGeoCLIP: Unified Geospatial Contrastive Learning

Guillaume Astruc, Eduard Trulls, Jan Hosang,
Loic Landrieu, Paul-Edouard Sarlin

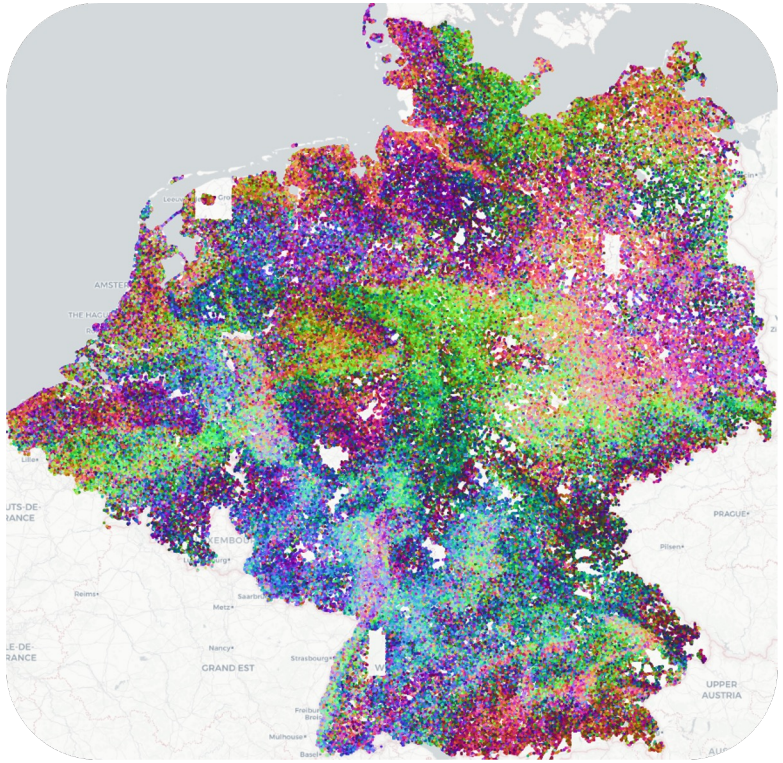
EarthVision Workshop, CVPR 2026

Google™ LaSTI cnes CENTRE NATIONAL D'ÉTUDES SPATIALES IGN INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORESTIÈRE ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES IP PARIS

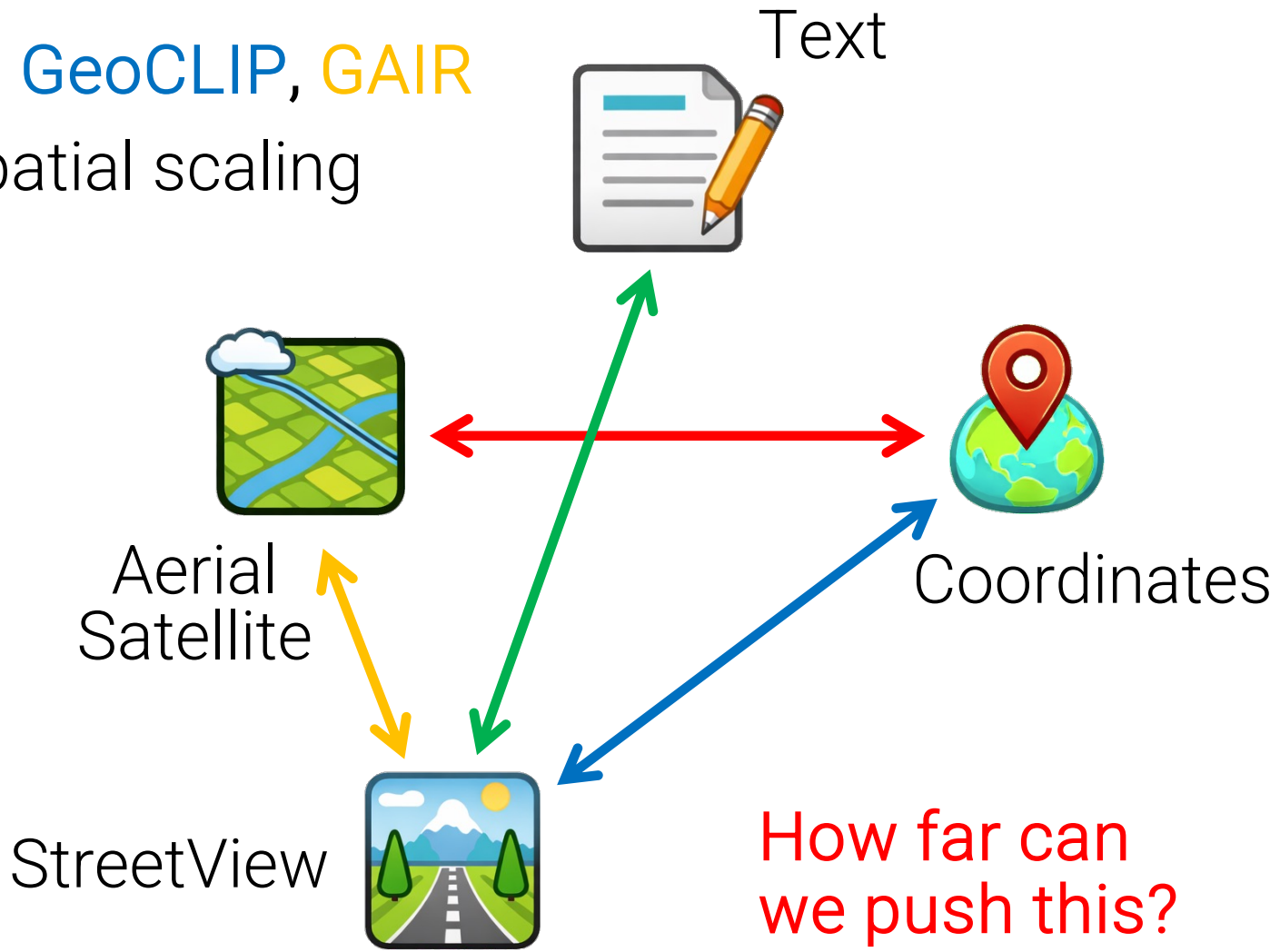
Representation learning for geospatial data

Past works: CLIP, SatCLIP, GeoCLIP, GAIR

Semantics emerge from spatial scaling



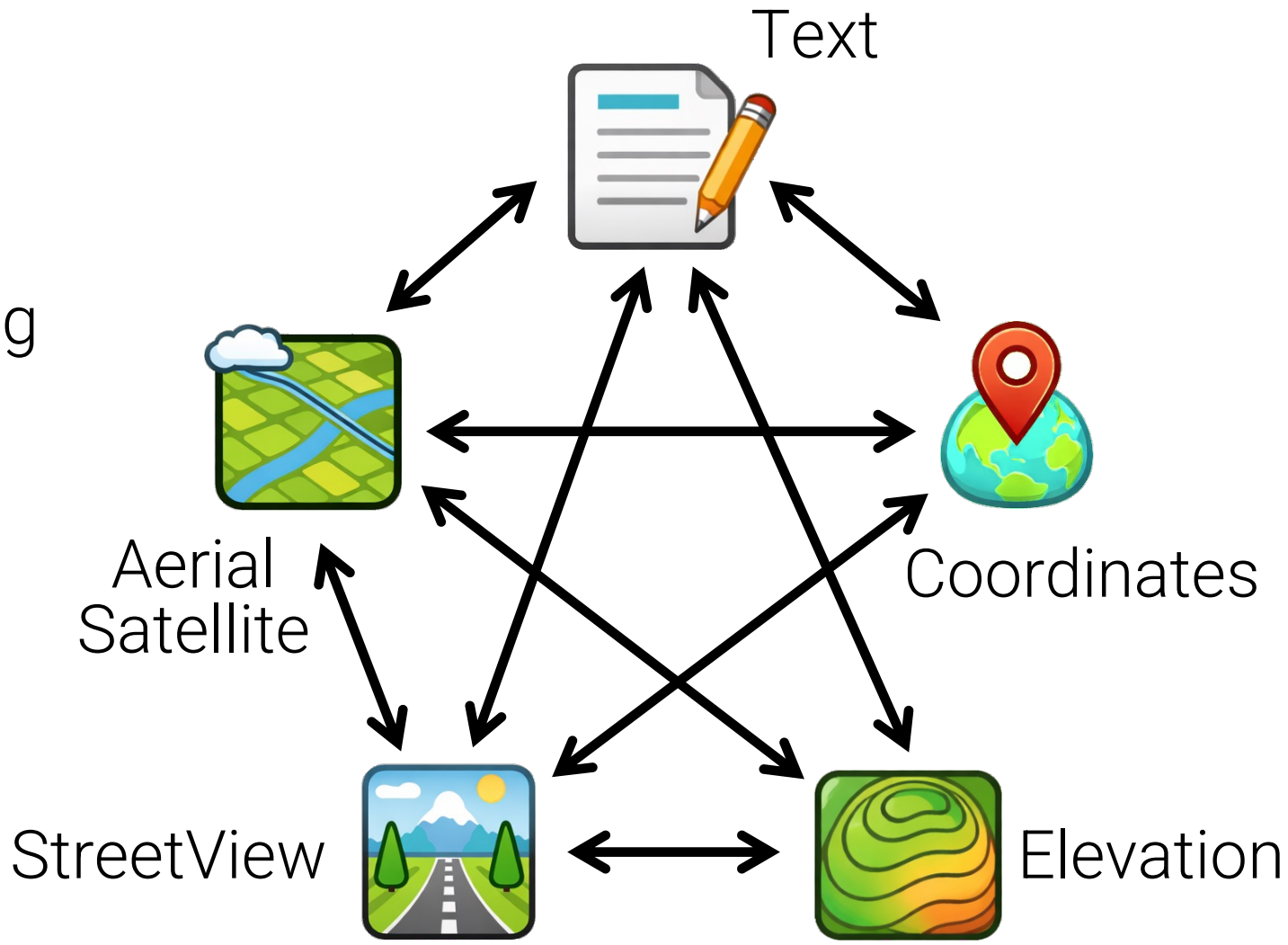
[Scaling Geoloc, NeurIPS 2025]



UniGeoCLIP

Massively multimodal
all-to-all contrastive learning

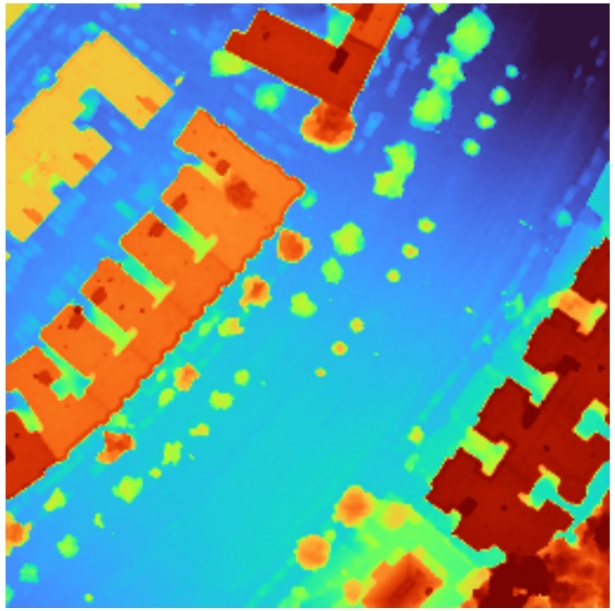
Joint embedding space
for 5 modalities



Training data

(a) Aerial Image.

(b) Street View.



(c) Surface Model.

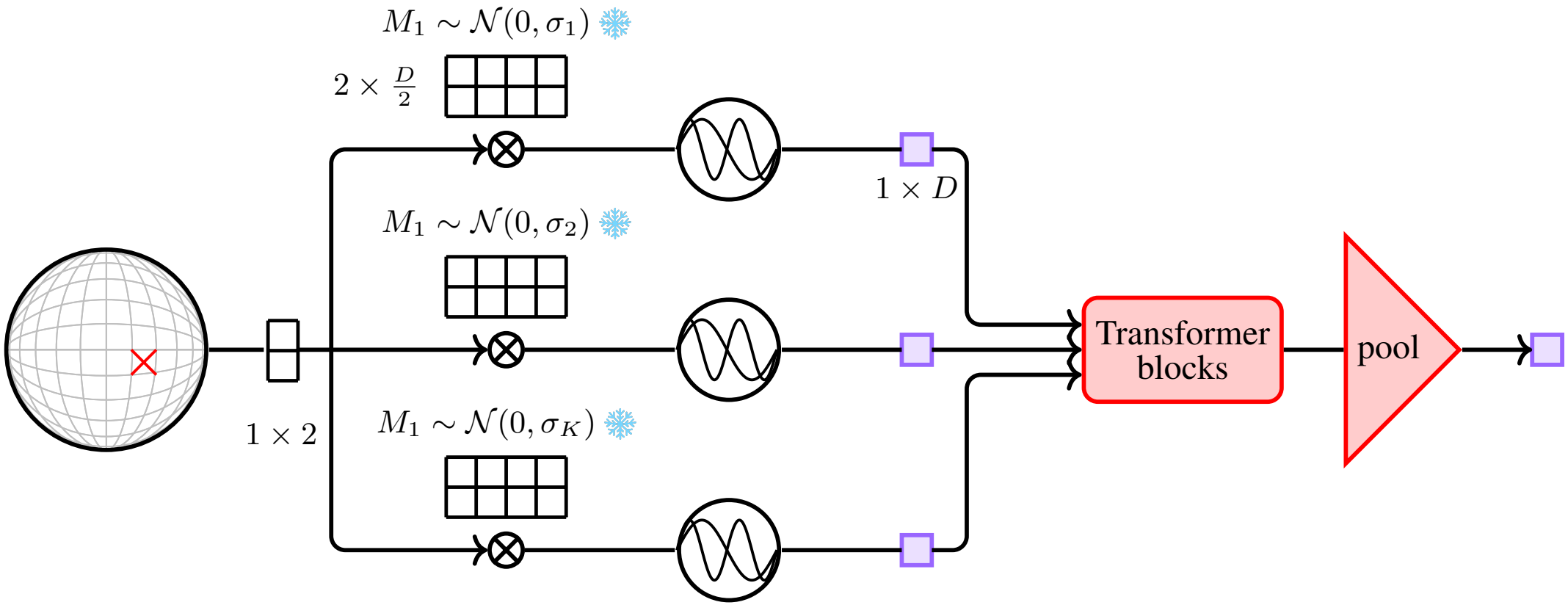
(d) Location.

Situated in Boston's lively Brighton neighborhood, this area is a convenient urban base close to Boston College and with accessible parks. Everyday conveniences include multiple grocery options, including Whole Foods Market and Star Market, as well as diverse restaurants ranging from Spanish tapas at Barcelona Wine Bar to Korean BBQ at Naksan. For nightlife, there's the welcoming atmosphere at Harry's Bar & Grill and The Publick House, and late-night dining at Barcelona Wine Bar.

(e) Text Description.

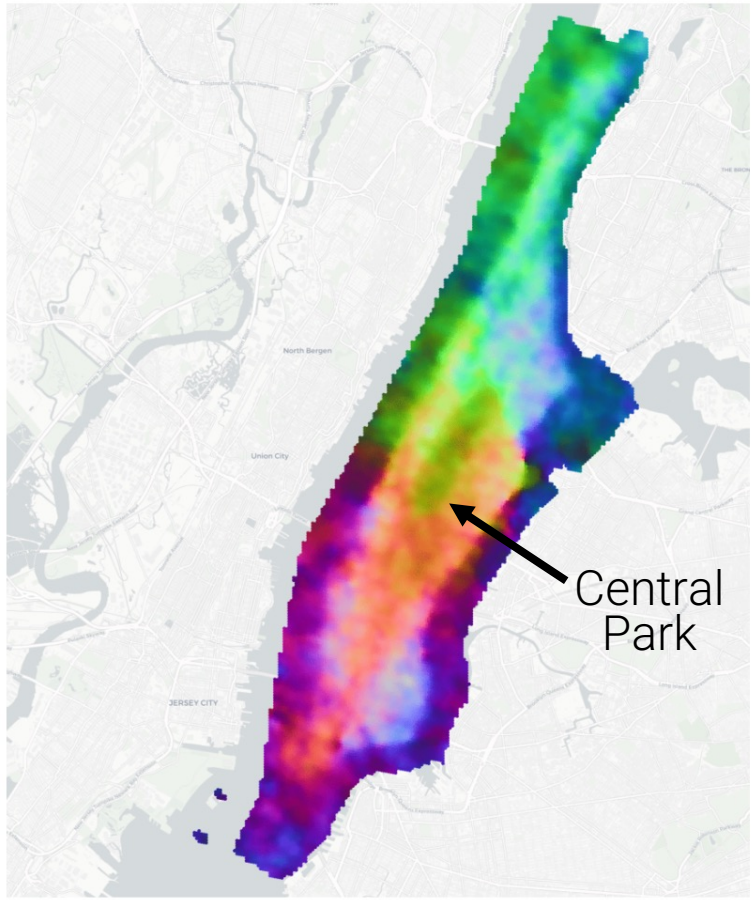
800k locations in 14 US metros, ~60M StreetView images

Scaling coordinate encoding

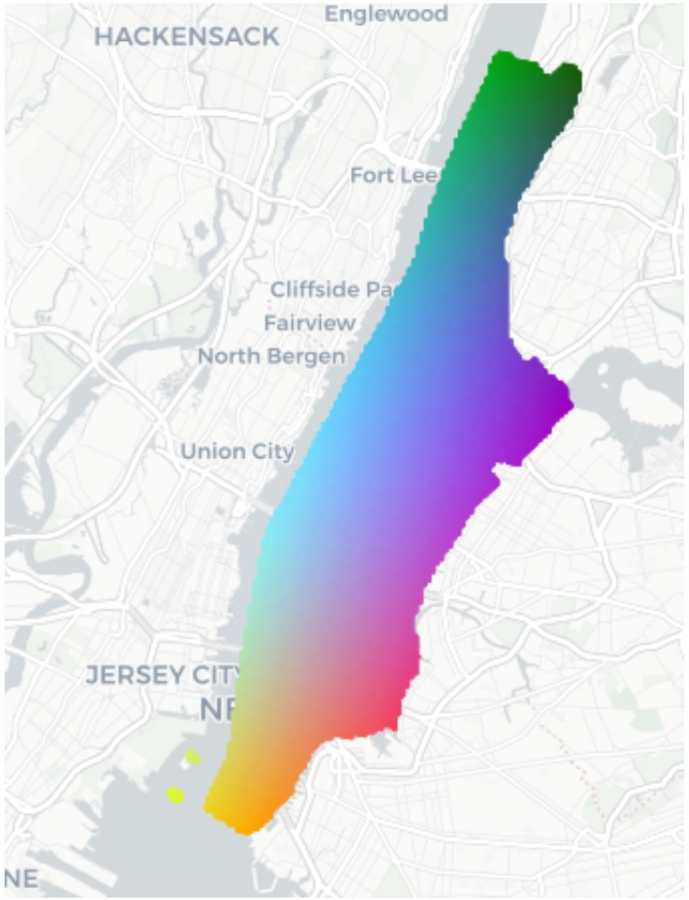


Multi-scale architecture for high resolution and capacity

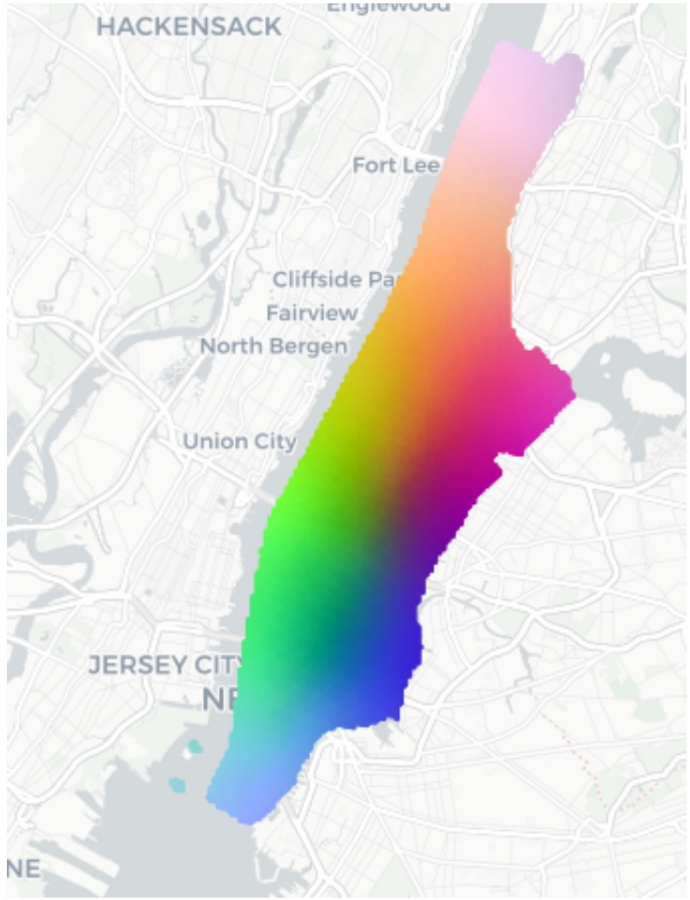
Learned representations



UniGeoCLIP



SatCLIP

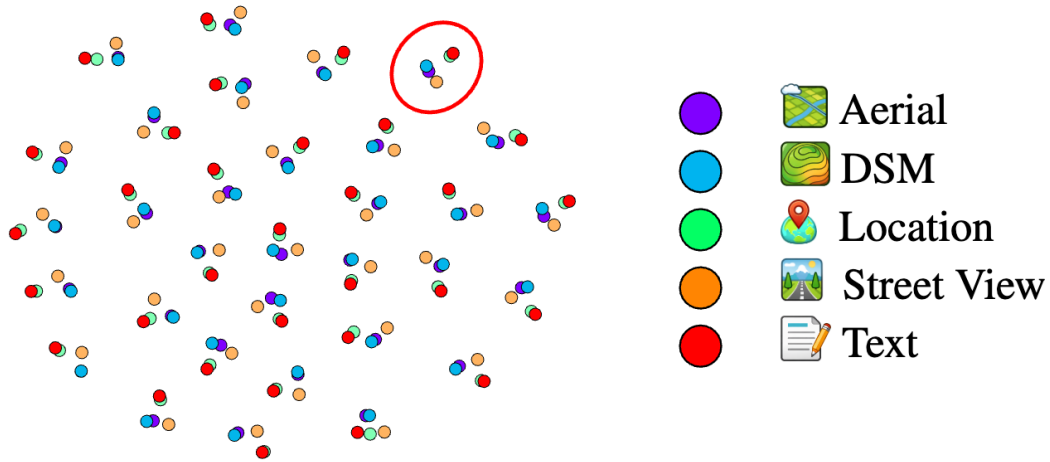


GeoCLIP

Extensive evaluations

Tasks:

- Cross-modality retrieval
- Semantic probing of each encoder
- Socioeconomic indicators



Key results:

- Scaling: each modality improves the representations
- Competitive with specialized, larger models

See you at poster #12 for more details!

Thank you!

gastruc.github.io/unigeoclip