

# GeoSURGE: Geo-localization using Semantic Fusion with Hierarchy of Geographic Embeddings

## Supplementary Material

### 1. Ablation Studies for IM2GPS, IM2GPS3k, and YFCC4k

Below we provide results from hierarchy depth, semantic fusion, and geographic representation ablation studies for the remaining benchmark datasets which did not fit within the main body of the article. These include IM2GPS, IM2GPS3k, and YFCC4k.

**Hierarchy depth ablations** for IM2GPS, IM2GPS3k, and YFCC4k are provided in Table 1. See Section ?? for details of the hierarchy depth ablations. We observed similar trends for the benchmarks presented here to the benchmarks presented in the main body of the article. Deeper hierarchies improve performance, stemming from smaller geocells with more precise locations as well as more diversity among geographical features at different scales.

**Semantic fusion ablations** for IM2GPS, IM2GPS3k, and YFCC4k are provided in Table 2. See Section ?? for details of the semantic fusion ablations. We observed similar trends for the benchmarks presented here to the benchmarks presented in the main body of the article. Accuracy tends to increase with number of fusion blocks.

**Geographic representation ablations** for IM2GPS, IM2GPS3k, and YFCC4k are provided in Tables 3 and 4. See Section ?? for details of the geographic representation ablations. We observed similar trends for the benchmarks presented here to the benchmarks presented in the main body of the article. In general, we observe that both hierarchy and geographic embeddings are complementary components of GeoSURGE’s core representation. Both are essential to achieving best overall performance.

### 2. Frozen CLIP Ablations

In GeoSURGE, all but the last CLIP Vision Transformer encoder blocks are kept frozen during training as described in Section ???. In this ablation study, we freeze all CLIP Vision Transformer encoder blocks to gauge the contribution of finetuning the last CLIP Vision Transformer encoder block for improved alignment between embeddings. All other variables are set to defaults and held constant across the GeoSURGE ablations. The results of the ablation are shown in Table 5. While this ablation shows strong performance, fine-tuning the last CLIP layer to better merge with later layers gives further gains by improving the alignment between visual and geographic embeddings.

Table 1. Hierarchy Depth Ablations

IM2GPS GCD accuracy; higher is better					
Hierarchy Levels	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
7	27.0	54.4	70.0	84.4	93.2
5	19.4	48.9	67.9	82.3	92.4
3	18.1	48.5	66.7	82.7	92.4
1	18.1	45.6	67.1	84.0	92.8

  

IM2GPS3k GCD accuracy; higher is better					
Hierarchy Levels	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
7	17.2	42.5	58.1	74.6	87.6
5	13.5	40.2	56.7	74.1	86.9
3	12.4	39.0	56.5	73.6	86.5
1	11.0	37.8	56.5	73.2	86.0

  

YFCC4k GCD accuracy; higher is better					
Hierarchy Levels	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
7	19.9	33.6	48.7	67.4	82.0
5	12.2	31.8	48.2	66.3	80.9
3	10.7	31.0	46.9	65.4	81.1
1	8.4	29.5	47.1	65.3	80.9

Table 2. Semantic Fusion Ablations

IM2GPS GCD accuracy; higher is better					
Fusion Blocks	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
3	27.0	54.4	70.0	84.4	93.2
2	22.4	49.8	68.8	83.1	93.2
1	21.5	50.2	67.1	83.5	91.1
None	19.8	49.8	68.3	83.5	93.2

  

IM2GPS3k GCD accuracy; higher is better					
Fusion Blocks	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
3	17.2	42.5	58.1	74.6	87.6
2	16.1	42.0	58.2	75.0	86.8
1	16.8	41.9	57.2	73.6	86.1
None	15.3	41.5	57.0	74.0	86.3

  

YFCC4k GCD accuracy; higher is better					
Fusion Blocks	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
3	19.9	33.6	48.7	67.4	82.0
2	15.1	33.6	48.9	66.8	81.8
1	15.8	33.6	48.7	66.0	81.0
None	14.4	31.7	47.8	65.5	80.8

Table 3. Hierarchical Geographic Representation Ablations

IM2GPS GCD accuracy; higher is better					
Geographic Embeddings?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	27.0	54.4	70.0	84.4	93.2
No	22.8	54.0	71.3	83.5	92.8
IM2GPS3k GCD accuracy; higher is better					
Geographic Embeddings?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	17.2	42.5	58.1	74.6	87.6
No	17.4	43.5	59.1	75.6	87.6
YFCC4k GCD accuracy; higher is better					
Geographic Embeddings?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	19.9	33.7	49.4	67.4	82.0
No	15.7	33.6	48.7	67.4	81.5

Table 4. Flat Geographic Representation Ablations

IM2GPS GCD accuracy; higher is better					
Geographic Embeddings?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	22.4	51.5	67.5	83.4	91.1
No	22.8	53.2	67.1	84.0	93.2
IM2GPS3k GCD accuracy; higher is better					
Geographic Embeddings?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	15.9	41.8	56.8	73.2	86.5
No	15.9	41.4	56.9	73.4	85.5
YFCC4k GCD accuracy; higher is better					
Geographic Embeddings?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	15.2	32.9	47.4	65.6	80.5
No	14.9	32.6	47.5	64.9	80.5

Table 5. Frozen CLIP Ablations

IM2GPS GCD accuracy; higher is better					
Finetune CLIP?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	27.0	54.4	70.0	84.4	93.2
No	24.1	52.3	70.5	84.0	93.2
IM2GPS3k GCD accuracy; higher is better					
Finetune CLIP?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	17.2	42.5	58.1	74.6	87.6
No	16.7	41.6	57.3	74.4	86.4
YFCC4k GCD accuracy; higher is better					
Finetune CLIP?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	19.9	33.6	48.7	67.4	82.0
No	19.3	33.5	47.3	65.5	80.9
YFCC26k GCD accuracy; higher is better					
Finetune CLIP?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	17.8	31.5	45.1	64.3	79.3
No	15.7	29.5	43.4	61.3	77.2
GWS15k GCD accuracy; higher is better					
Finetune CLIP?	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
Yes	1.0	4.6	21.9	54.7	80.8
No	0.9	4.8	22.9	54.0	81.1