GT-Loc: Unifying When and Where in Images Through a Joint Embedding Space

David Shatwell[1], Ishan Dave[2], Sirnam Swetha[1], Mubarak Shah[1]
[1]University of Central Florida, [2]Adobe

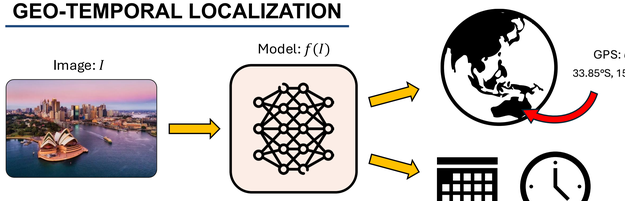
UCF Center for Research in Computer Vision

Project Page

★ ORAL

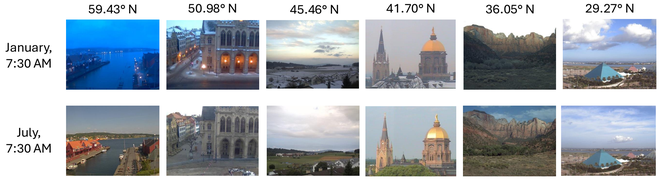
ICCV OCT 19-23, 2025 HONOLULU HAWAII

GEO-TEMPORAL LOCALIZATION



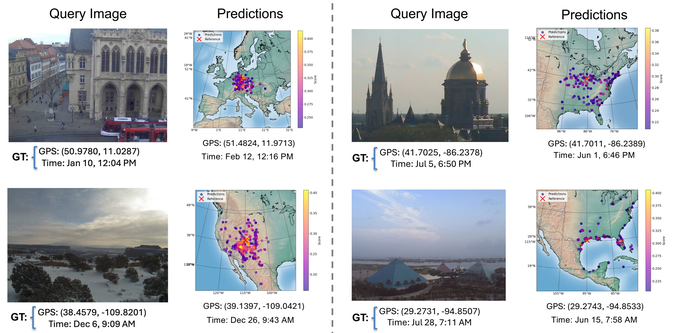
- Given an image what is the **geo-location** and **time-of capture**?
- Given a geo-location and time-of-capture, what is the **best matching image**?

CHALLENGES

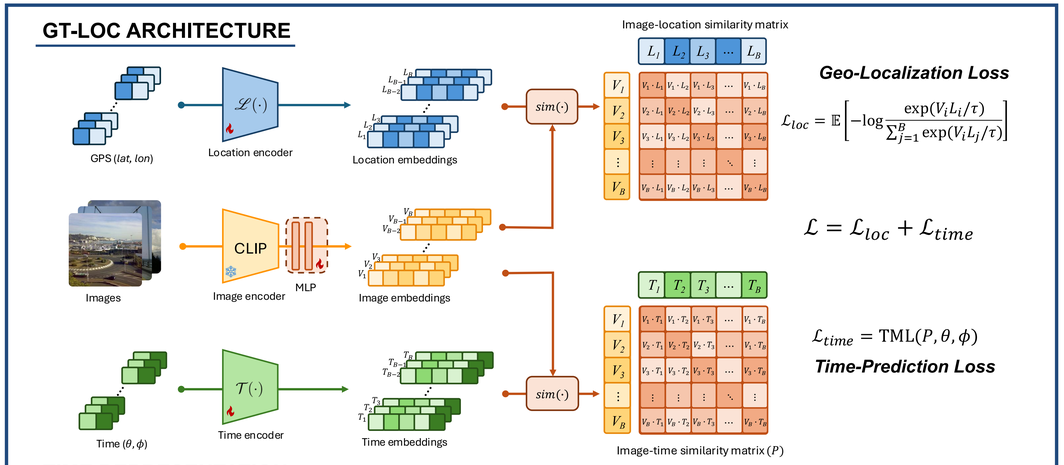


- Higher latitudes have large differences in seasons and hours of sunlight throughout the year compared to lower latitudes
- **The visual appearance of time strongly depends on the geo-location**

RESULTS: TIME PREDICTION & GEO-LOCALIZATION



GT-LOC ARCHITECTURE



Geo-Localization Loss

$$\mathcal{L}_{loc} = \mathbb{E}\left[-\log \frac{\exp(V_i L_i/\tau)}{\sum_{j=1}^{B} \exp(V_i L_j/\tau)}\right]$$

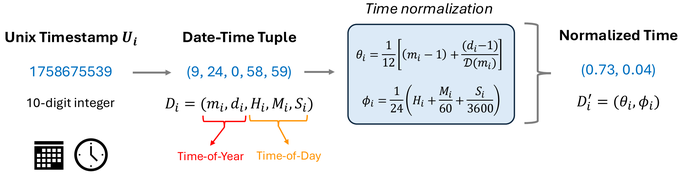
$$\mathcal{L} = \mathcal{L}_{loc} + \mathcal{L}_{time}$$

$$\mathcal{L}_{time} = \mathrm{TML}(P, \theta, \phi)$$

Time-Prediction Loss

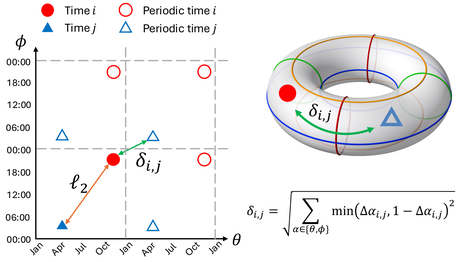
TIME REPRESENTATION

1. Timestamps are given in Unix format
2. We convert timestamps to date-time tuples
   - Time-of-year: month & days
   - Time-of-day: hours, minutes & seconds
3. Compact time representation
   - Normalized time-of-year: $\theta_i \in [0,1]$
   - Normalized time-of-day: $\phi_i \in [0,1]$



TIME DIFFERENCE ($\delta_{i,j}$)

- Months and hours are periodic
- We represent time as points in month-hour plane (torus)

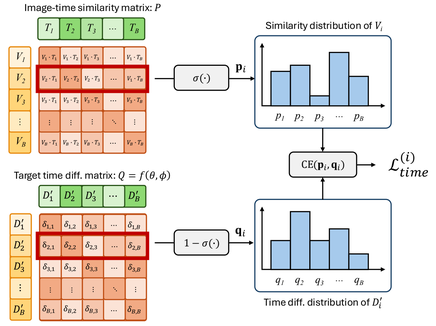


$$\delta_{i,j} = \sqrt{\sum_{\alpha \in \{\theta, \phi\}} \min\left(\Delta\alpha_{i,j}, 1 - \Delta\alpha_{i,j}\right)^2}$$

Time difference:
- ℓ_2 overestimates the time difference
- Real time difference $\delta_{i,j}$ computed between time i and the closest periodic copy of time j

TEMPORAL METRIC LEARNING (TML)

- TML: metric contrastive loss for image-time alignment
- Encourages distribution of similarities to match the distribution of time differences



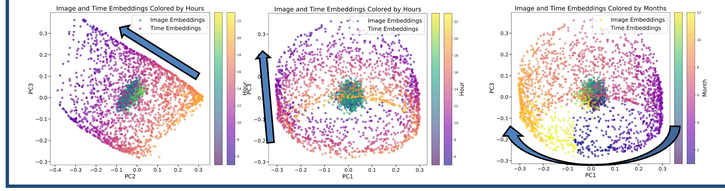
DOWNSTREAM TASK: GEO-TEMPORAL IMAGE RETRIEVAL

Query Location & Time: GPS: (38.4579, -109.8201) Time: Feb 1, 8:39 AM

EMBEDDING SPACE ANALYSIS

Time-of-day (hours): linear pattern

Time-of-year (months): circular pattern



QUANTITATIVE RESULTS

SoTA performance on 3 tasks:
- Time prediction
- Geo-localization
- Geo-temporal image retrieval

Effect of joint embedding space
- GT-Loc achieves significant improvements over baselines that use GPS as input

→ **Joint embedding is crucial for accurate time prediction!**

Time-of-Capture Prediction
Dataset: SkyFinder

Method	Month Error ↓	Hour Error ↓	TPS ↑
Zhai et al.	2.46	3.18	65.48
Padilha et al.	1.62	3.61	71.42
Salem et al.	2.72	3.05	63.25
Zhai et al. w/ CLIP	1.65	3.14	73.20
Padilha et al. w/ CLIP	1.62	2.98	74.06
Salem et al. w/ CLIP	1.56	2.87	75.02
CLIP + cls. head	1.61	3.17	73.37
CLIP + reg. head	1.55	3.06	74.33
DINOv2 + cls. head	2.24	3.74	65.61
DINOv2 + reg. head	2.12	3.53	67.49
OpenCLIP + cls. head	1.66	3.43	71.87
OpenCLIP + reg. head	1.72	3.34	71.75
TimeLoc	1.52	2.84	75.49
GT-Loc (Ours)	**1.40**	**2.72**	**77.00**

Geo-Localization (1 km)

Method	Im2GPS3k	GWS15k
[L] kNN, sigma=4	7.2	-
PlaNet	8.5	-
CPlaNet	10.2	-
ISNs	10.5	0.1
Translocator	11.8	0.5
GeoDecoder	12.8	0.7
GeoCLIP	14.11	0.6
PIGEOTTO	11.3	0.7
Img2Loc(LLaVA)	8.0	-
GT-Loc (Ours)	**14.41**	**0.88**

Geo-Temporal Image Retrieval
Dataset: SkyFinder

Method	R@1	R@5	R@10
Zhai et al.	0.91	7.81	13.61
Zhai et al. w/ CLIP	2.58	16.22	29.46
GT-Loc (Ours)	**6.69**	**24.58**	**38.54**

TAKEAWAYS

1. Shared embedding space between the image, time and location modalities
   - Enables novel downstream tasks: geo-temporal image and text retrieval
2. First retrieval method for time-of-capture prediction
3. Novel time representation over the surface of a flat torus
4. Temporal Metric Learning for image-time alignment with soft targets