AI Geolocation in Russia and Ukraine

Selena Sun
Stanford Computer Science
Vannevar Labs
selenas@stanford.edu

Charu Dwivedi
Vannevar Labs
charu.dwivedi@vannevarlabs.com

Cane Punma
Vannevar Labs
cane.punma@vannevarlabs.com

Abstract

Open source intelligence (OSINT) investigators face the challenge of verifying the location of media shared online: given an image, what are its coordinates? This "geolocation" process is labor-intensive and cannot scale with the ever-growing volume of social media content, especially in active conflict regions. To address this bottleneck, I present GeoFT, a geolocation model optimized for Russia and Ukraine. I build on the work of GeoCLIP, a global geolocation model with a CLIP backbone, by fine-tuning on region-specific datasets. By leveraging Google Street View imagery and the EyesOnRussia dataset (community-geolocated real-time images captured by civilians and journalists), our model achieves state-of-the-art inference on Russia and Ukraine, outperforming existing models by a factor of 14.77x in prediction accuracy.

The contributions in this paper are twofold. First, I present a model with significant improvements in regional accuracy for geolocation in Russia and Ukraine. GeoFT reduces the mean error distance from 3,521 km in the baseline GeoCLIP model to just 238 km on our test set. Secondly, I propose a generalizable fine-tuning methodology for regional geolocation that produces significant improvements even when using only publicly accessible Google Street View imagery (4.12x accuracy improvement, without EyesOnRussia data). Though the fine-tuned model performs better on the Russia and Ukraine regions, the fine-tuning method is extensible to other regions worldwide, enabling more reliable, automated verification of media in conflict zones and directly supporting the critical work of OSINT investigators globally.

1 Introduction

The verification of media location, or geolocation, has traditionally been a crowdsourced effort relying on human investigators using resources like Google Street View and building databases. While effective, this manual process cannot keep pace with the thousands of new images appearing daily on social media platforms. I present GeoFT, a fine-tuned model that specifically targets street-level imagery in regions of active conflict, where traditional solutions like Google Street View may be outdated or unavailable.

2 Related Work

Recent approaches to AI geolocation include GeoCLIP (Vivanco Cepeda et al., 2023), PIGEOTTO (Haas et al., 2024), and GeoDecoder (Qi et al., 2024). GeoCLIP uses contrastive learning between image and location embeddings but exhibits high average error (3,520 km). PIGEOTTO and GeoDecoder showed promising results but lack open-source implementations. Commercial solutions like GeoSpy (GeoSpy, 2024) currently rely heavily on

static image databases, limiting their effectiveness in rapidly changing environments. These approaches demonstrate the challenge of accurate geolocation at scale.

3 Methodology

3.1 Data Collection and Filtering

I curated a dataset combining two primary sources:

1. Eyes on Russia (EoR): 2,887 community-geotagged images from the conflict region (Eyes on Russia Project, 2024)
2. Google Street View (GSV): 16,159 street-level images gathered via API (Google, 2024)

Certain images inherently lack sufficient geographic identifiers, with indoor photographs (such as those of furniture) presenting significantly greater geolocation challenges than outdoor scenes containing distinctive architectural elements. To filter out the less descriptive scenes, I used GPT-4-mini as a binary classifier: "Does this image contain street features?" This preprocessing step significantly improved training data relevance by removing indoor scenes and irrelevant imagery. The final dataset contains 19,046 outdoor images.

3.2 Model Architecture

I adopt GeoCLIP's two-branch architecture consisting of a *location encoder* $\mathrm{L}(\cdot)$ and an *image encoder* $\mathrm{V}(\cdot)$. Following Vivanco Cepeda et al. (2023), GPS coordinates (φ, λ) are first converted to the Equal-Earth projection to minimise areal distortion. The projected pair is then passed through *three* Random Fourier Feature (RFF) blocks with bandwidths $\sigma \in 2^0, 2^4, 2^8$ to capture hierarchical spatial detail. Each block is followed by a 4-layer MLP (hidden $= 1,024$, output $= 512$) whose weights are trainable during fine-tuning.

We reuse the pre-trained CLIP ViT-L/14 backbone and append two linear adapters ($h_1 = 768$, $h_2 = 512$). To preserve CLIP's broadly-useful representations we *freeze* the transformer parameters and only update the adapters. Fine-tuning the model minimizes the loss function as described in Vivanco Cepeda et al. (2023).

4 Results

GeoFT exhibits a 14.77x improvement in average prediction accuracy over the baseline GeoCLIP model on the test set, a collection of 1892 images from the EyesOnRussia, Google Streetview (Russia), and Google Streetview (Ukraine).

Table 1 compares the performance of the four different models across a range of accuracy granularities (1 km, 25 km, 200 km, 750 km, and 2500 km): fine-tuned GeoCLIP, fine-tuned GeoCLIP (only Google Streetview data), baseline GeoCLIP, and GeoSpy. A prediction is accurate within X km, and all coarser accuracies, if its guess on image I_i is within X km of the ground-truth GPS label G_i.

Table 1: Geolocation Accuracy Comparison on Test Set

Distance	GeoFT	Fine-Tuned on GSV Only	Baseline GeoCLIP	GeoSpy
Mean Error (km)	**238.38**	853.39	3520.95	3304.15
2500 km	100%	92%	56%	54%
750 km	95%	54%	0%	0%
200 km	59%	7%	0%	0%
25 km	38%	6%	0%	0%
1 km	6%	0%	0%	0%

Table 2 shows sample test images from the EyesOnRussia dataset and the models' corresponding geolocation.

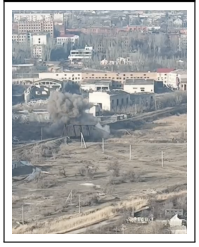
	Source	GeoFT	GeoCLIP	GeoSpy
	Eyes on Russia Dataset	**1116.09 km**	1258.15 km	1990.19 km
	Eyes on Russia Dataset	**513.07 km**	1890.84 km	5517.76 km

Table 2: Prediction errors for sample images from the Eyes on Russia dataset.

GeoFT is accurate within 1 km of the ground truth location on 6% of the test set, whereas all other models are never within 1 km. This points at the ability of GeoFT to map terrain features of newly seen images to real locations for a small subset of locations.

The results demonstrate the effectiveness of regional specialization in geolocation models. Our fine-tuned approach shows particular strength in fine-grained localization, with 38% accuracy within 25 km compared to 0% for baseline models. This improvement stems from the model learning region-specific, updated visual features (e.g., changing building appearances) previously underrepresented in public data.

5 Discussion and Future Work

GeoFT has made measurable progress in improving geolocation predictions on terrain with changing scenes, rendering AI geolocation models more useful for OSINT analysts.

In an attempt to make the model more operational, I examine the "confidences" of models' predictions. If there's a threshold for confidence above which the model's predictions become meaningfully more accurate, I could discard all predictions below this threshold. To this end, I visualize error (km) vs. confidence in Figure 1. Unfortunately, I observed no such threshold; the model is equally confident about geolocation guesses with a 1,000 km error as it is about guesses with a 0 km error.

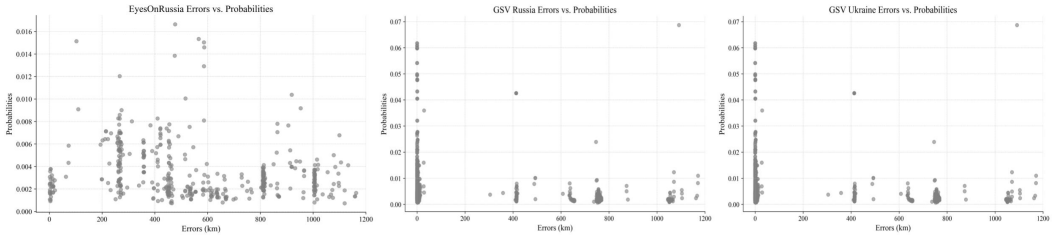


Figure 1: Error vs. Confidence for EyesOnRussia, GSV Russia, and GSV Ukraine.

Our evaluation reveals two key findings that demonstrate GeoFT's effectiveness and practical applicability.

First, when trained on the combined Eyes on Russia (EoR) and Google Street View (GSV) dataset, GeoFT demonstrates significant improvements over baseline models. Specifically,

GeoFT achieves state-of-the-art performance on Russia and Ukraine test sets respectively compared to prior work (Figure 1).

Secondly, and perhaps most significantly for practical applications, GeoFT demonstrates strong performance even when trained solely on GSV data (4.12x accuracy improvement), without requiring region-specific human-labeled datasets like EoR. This finding is crucial for extending the approach to new geographic regions where specialized datasets may not be available. Table 1 compares performance between models trained on the combined dataset versus GSV data only.

While GeoFT shows promising results, several extensions could enhance its utility:

1. Integration with production systems for continuous model improvement using new validated data
2. Expansion to video and aerial imagery analysis
3. Extension to other regions of interest with similar data collection methodology

The model can be deployed both as a standalone tool for OSINT investigators and integrated into existing intelligence platforms, providing automated first-pass location estimates for human verification.

GeoFT demonstrates that fine-tuning foundation models on carefully curated regional data can significantly improve geolocation accuracy. The success of our approach using only Google Street View data suggests that this methodology could be extended to other regions of interest, even without access to specialized OSINT datasets. This work represents a step toward scalable, automated support for OSINT investigations.

6 Acknowledgments

Thank you to Charu Dwivedi, Cane Punma, Scott Weitzner, and many more from Vannevar Labs for the opportunity to work on this project!

References

Eyes on Russia Project. Eyes on russia. `https://eyesonrussia.org`, 2024. Accessed: February 2024.

GeoSpy. Geospy: Ai-powered image geolocation. `https://geospy.ai`, 2024. Accessed: February 2024.

Google. Google street view. `https://www.google.com/streetview`, 2024. Accessed: February 2024.

Lukas Haas, Michal Skreta, Silas Alberti, and Chelsea Finn. Pigeon: Predicting image geolocations. *Preprint*, May 2024.

Feng Qi, Mian Dai, Zixian Zheng, and Chao Wang. Geodecoder: Empowering multimodal map understanding. *arXiv preprint*, Feb 2024. arXiv:2401.2401.15118.

Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *Advances in Neural Information Processing Systems*, 2023. NeurIPS 2023.