

G^3 : Geolocation via Guidebook Grounding

Grace Luo*, Giscard Biamby*, Trevor Darrell, Daniel Fried, Anna Rohrbach



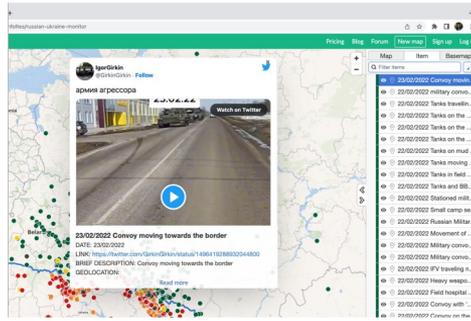
University of California, Berkeley

Why Geolocation?

Protest in **Berlin** today against the CV Lockdown and tyranny...



1. Fact Checking. This photo shows a street parade in **Switzerland**...
([AFP News Fact Check](#)).



2. Event Monitoring. A map of photos and videos manually fact-checked for the Russia-Ukraine war. ([Center for Info Resilience](#))



3. Event Reconstruction. An overlay of video showing Charlotte police kettling on a 3D reconstruction of the street it was taken. ([SITU Research](#))

GEOGUESSR



4. GeoGuessr. An example play-along demonstrating how geolocation has become a popular pastime amongst the general public. ([Youtube](#))

Let's play! Where was this image taken?



Let's play! Where was this image taken?



Guidebook

A key point to remember in GeoGuessr is that if you see a **red house** or several red houses then you are probably in a Nordic country (**Sweden**, Norway, Finland, Iceland or Denmark).

Sweden often has **white dashes on the sides of its roads.**

These dashes are different from the Norwegian dashes as **they are shorter than the space between the dashes.**

Let's play! Where was this image taken?



Guidebook

A key point to remember in GeoGuessr is that if you see a **red house** or several red houses then you are probably in a Nordic country (**Sweden**, Norway, Finland, Iceland or Denmark).

Sweden often has **white dashes on the sides of its roads.**

These dashes are different from the Norwegian dashes as **they are shorter than the space between the dashes.**

Overview

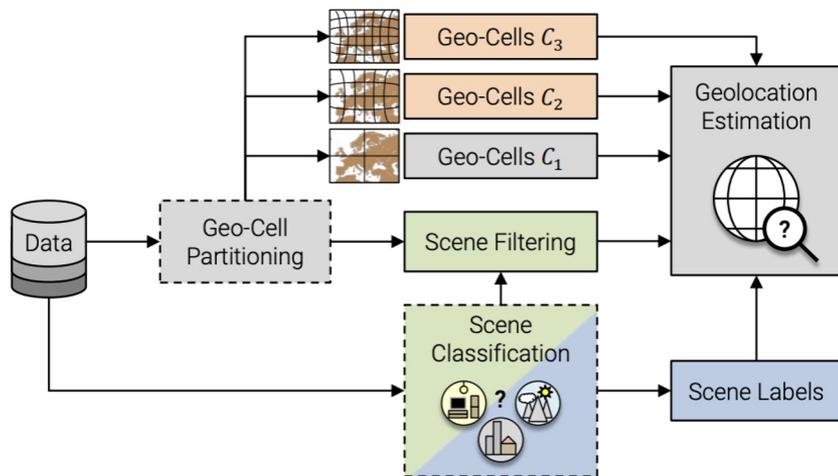
Why not transform the **traditionally vision-centric** task of geolocation into a **multimodal** one with **relevant web text** from the game GeoGuessr?

- Geolocation via Guidebook Grounding: predict the location of an image *given relevant textual knowledge in the form of a guidebook*.
- Our approach G^3 : Uses knowledge from this human-written guidebook to improve over an image-only state-of-the-art method by over **5%** in country classification accuracy.

Related Work

ISN: Trained CNN on millions of (Flickr image, GPS) pairs to predict which geo-cell an image lies in.

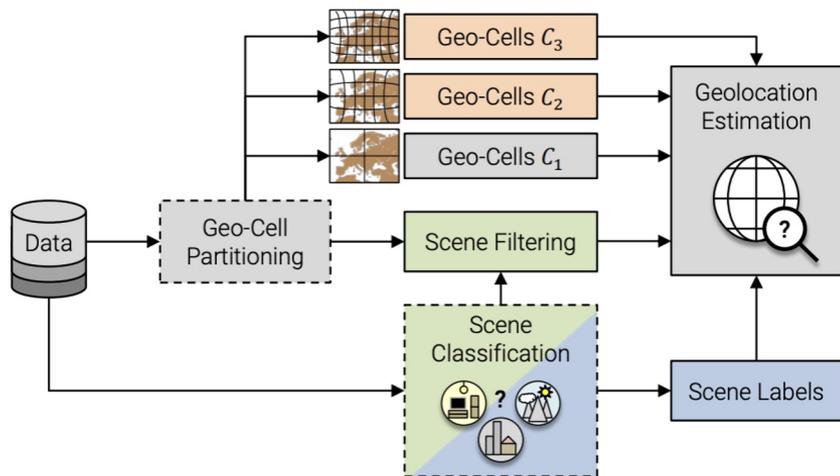
Muller-Budack et. al., ECCV 2018.



Related Work

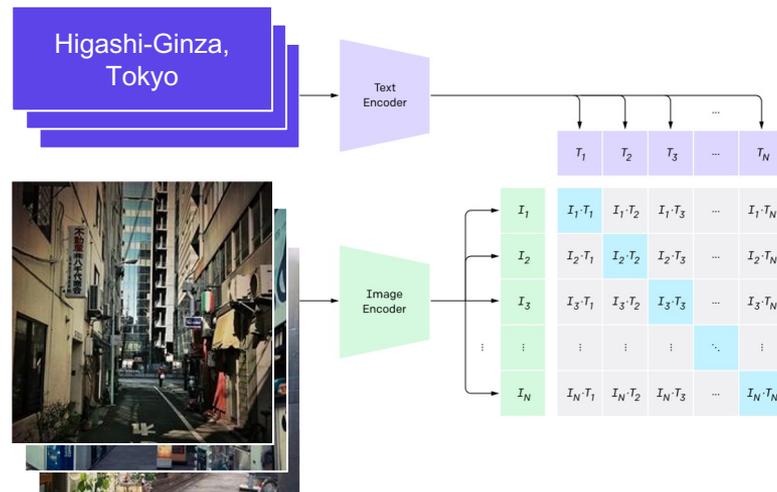
ISN: Trained CNN on millions of (Flickr image, GPS) pairs to predict which geo-cell an image lies in.

Muller-Budack et. al., ECCV 2018.



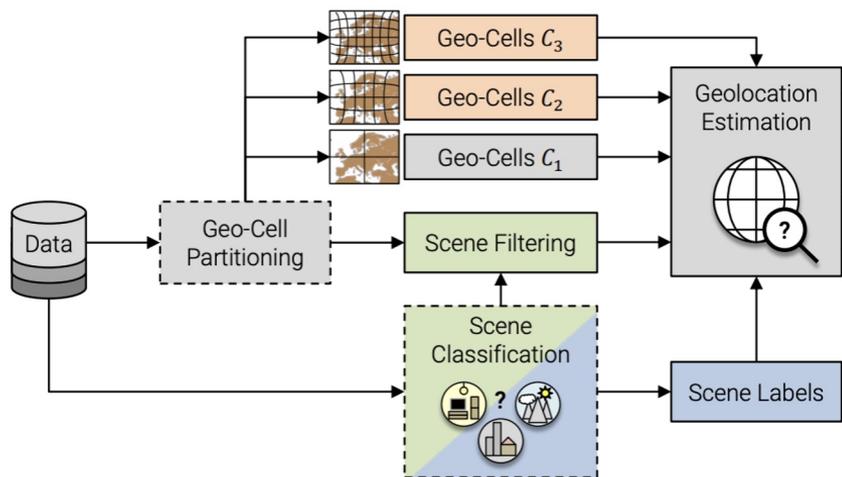
CLIP: Trained joint image-text embedding on millions of (web image, web caption) pairs to maximize cosine similarity of matching pairs.

Radford et. al., Arxiv 2021.

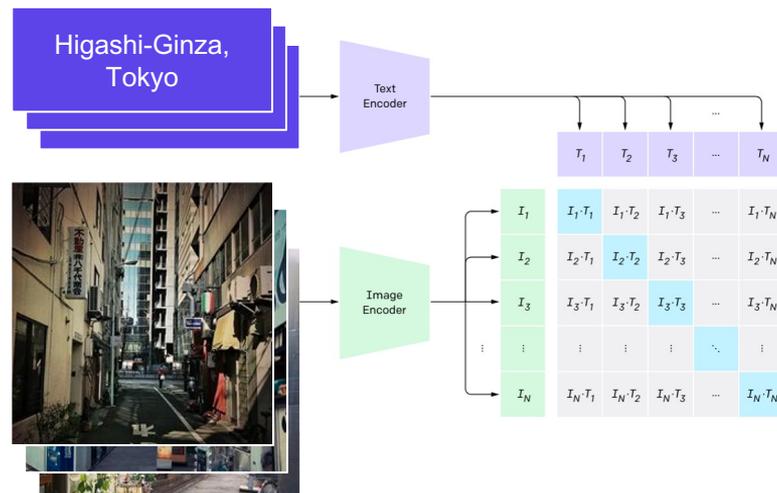


Related Work

ISN: Trained CNN on millions of (Flickr image, GPS) pairs to predict which geo-cell an image lies in.
Muller-Budack et. al., ECCV 2018.



CLIP: Trained joint image-text embedding on millions of (web image, web caption) pairs to maximize cosine similarity of matching pairs.
Radford et. al., Arxiv 2021.



Ours: Use the knowledge from these embeddings and **relevant clues retrieved from a guidebook.**

Dataset

Images + Locations + Relevant Textual Knowledge

300k+ StreetView Images



3k+ GeoGuessr Guidebook Clues

Always be sure to look at both sides of **Thai** road markers and remember that **Thai** drivers drive on the left when they see the marker information.

Most **Thai** roads will have some form of yellow central line as well as continuous white edge lines.

Many **Thai** houses can also look quite affluent and be two-storey and fairly large.

Dataset

Images + Locations + Relevant Textual Knowledge

300k+ StreetView Images



3k+ GeoGuessr Guidebook Clues

Always be sure to look at both sides of **Thai** road markers and remember that **Thai** drivers drive on the left when they see the marker information.

Most **Thai** roads will have some form of yellow central line as well as continuous white edge lines.

Many **Thai** houses can also look quite affluent and be two-storey and fairly large.

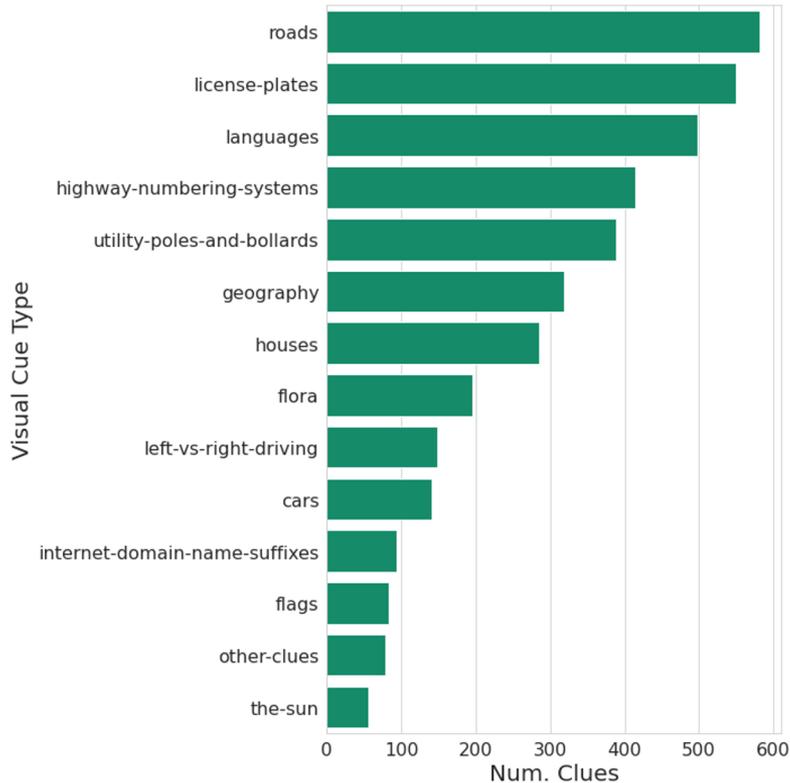
StreetView Images



Geo-diverse StreetView images from 90 countries.



Guidebook Clues



Roads “Most of North and South America have some iteration of yellow centre lines.”

License Plates “Argentina also has plates with a blue horizontal line on top.”

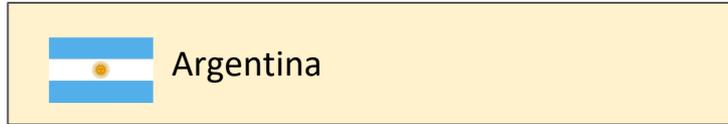
Languages “French: This language is spoken across France, ...Switzerland, ...Belgium and...Canada...”

Highway Numbering Systems “New Zealand also has red shields throughout the country that denote their highways.”

Utility poles and bollards “In the above image, there are vertical yellow and black stripes on the utility pole which are Japanese.”

...

Pseudo Labels



Roads “Most of North and South America have some iteration of yellow centre lines.”

License Plates “Argentina also has plates with a blue horizontal line on top.”

Languages “French: This language is spoken across France, ...Switzerland, ...Belgium and...Canada...”

Highway Numbering Systems “New Zealand also has red shields throughout the country that denote their highways.”

Utility poles and bollards “In the above image, there are vertical yellow and black stripes on the utility pole which are Japanese.”

...

Approach G^3



Approach G^3



Argentina also has plates with a blue horizontal line on top.

In the above image, there are vertical yellow and black stripes on the utility pole which are Japanese.

...

Swedish directional signs are blue with white lettering...

Approach G^3



Argentina also has plates with a blue horizontal line on top.

In the above image, there are vertical yellow and black stripes on the utility pole which are Japanese.

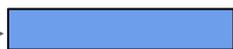
...

Swedish directional signs are blue with white lettering...

Text Encoder



0.76



0.08

...



0.00

Clue Embeddings

L_{attn}

Attention Layer



Approach G^3

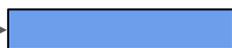


Argentina also has plates with a blue horizontal line on top.

In the above image, there are vertical yellow and black stripes on the utility pole which are Japanese.

...

Swedish directional signs are blue with white lettering...



0.76



0.08

...



0.00

Clue Embeddings

Attention Layer

L_{attn}



Approach G^3



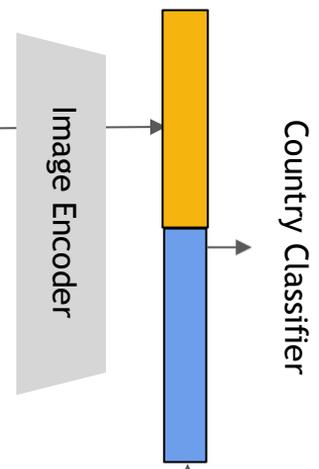
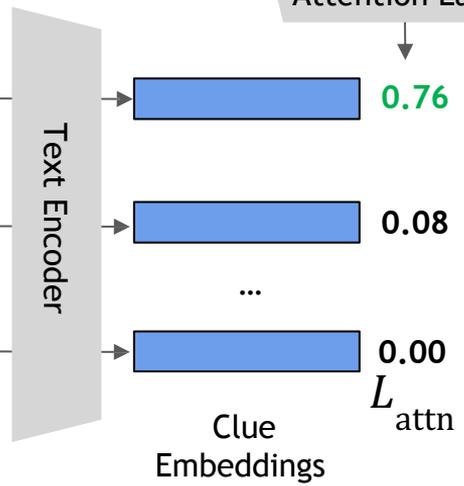
- 0.98 **Argentina**
 - 0.01 Chile
 - 0.01 Uruguay
 - ...
 - 0.00 Sweden
 - 0.00 Netherlands
 - 0.00 Nigeria
- L_{country}

Argentina also has plates with a blue horizontal line on top.

In the above image, there are vertical yellow and black stripes on the utility pole which are Japanese.

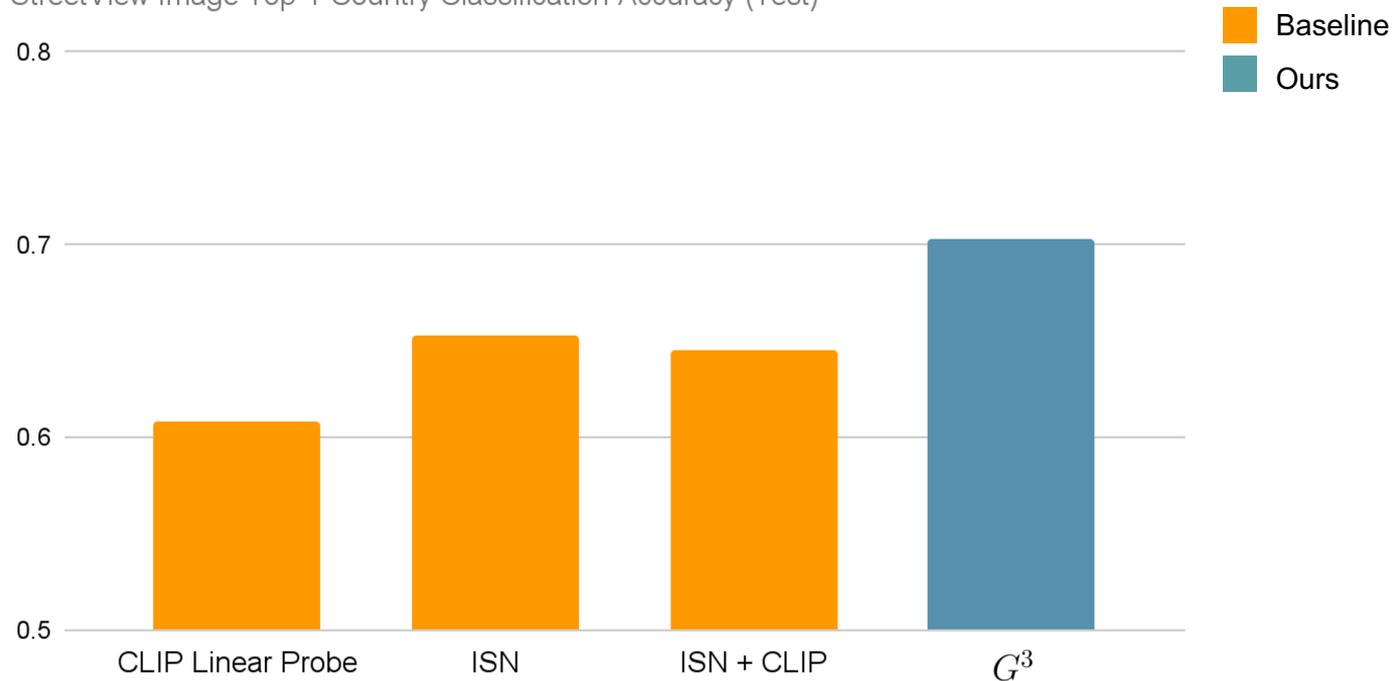
...

Swedish directional signs are blue with white lettering...

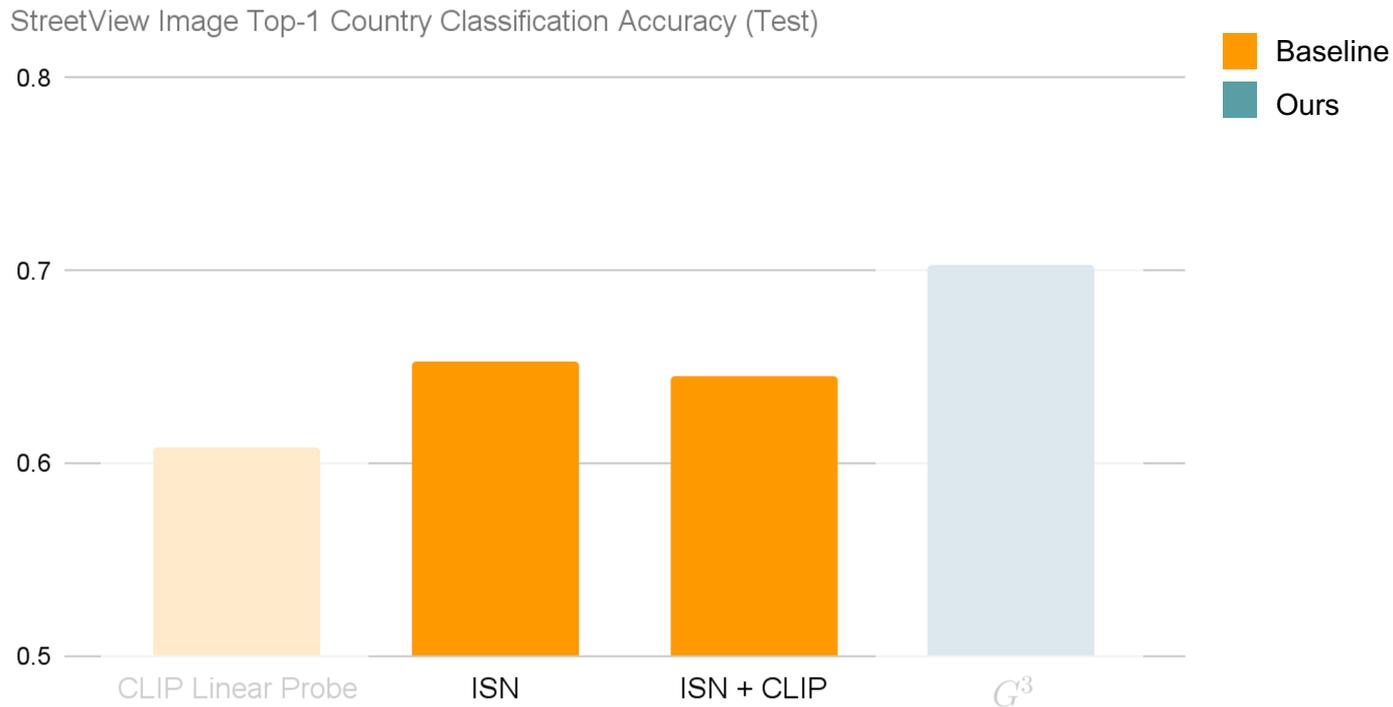


Experiments

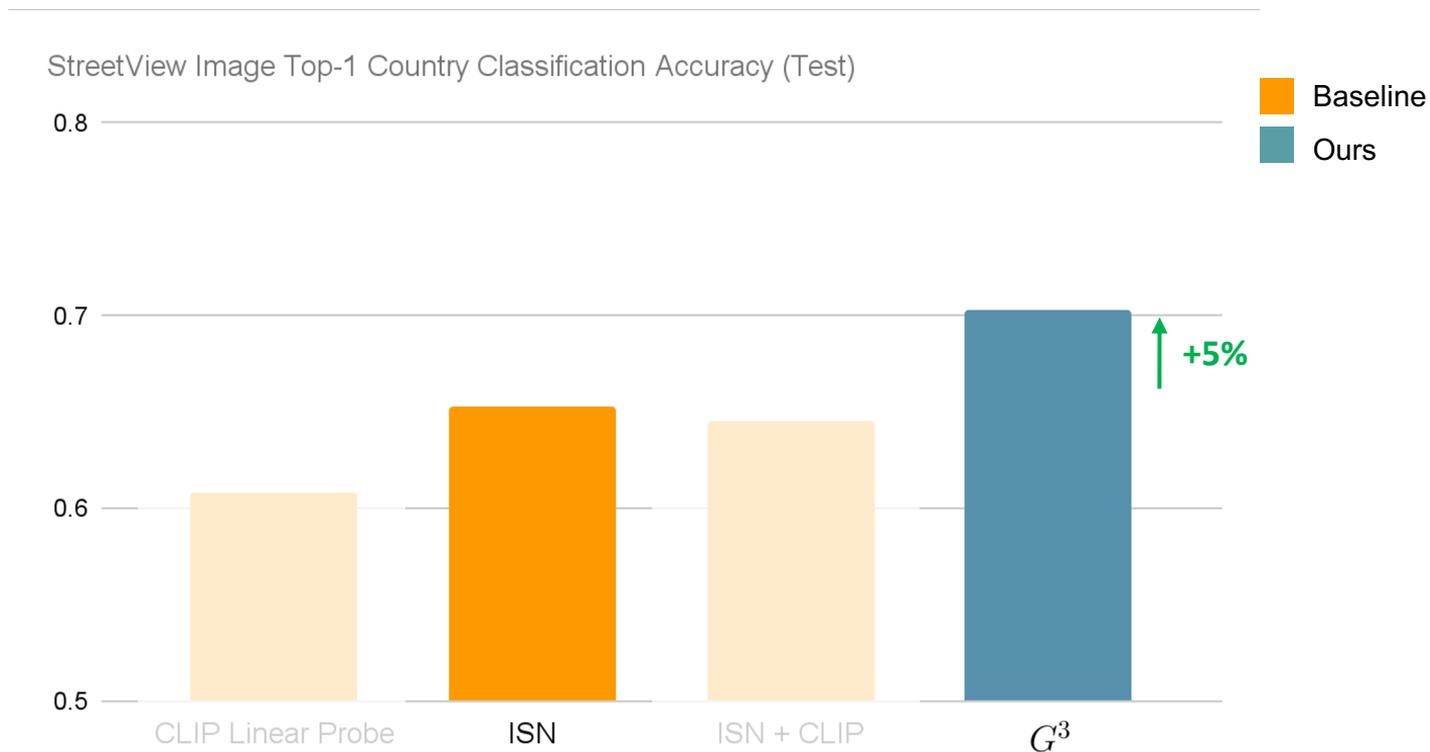
StreetView Image Top-1 Country Classification Accuracy (Test)



Experiments



Experiments



Qualitative Results

GT Location: Botswana

ISN Prediction: Kenya

ISN + CLIP Prediction: UAE

G³ Prediction: Botswana



Top-10 Retrieved Guidebook Clues	
0.9410	As large amounts of the UAE Street View coverage centres around Dubai, you may be able to see the city's skyscrapers in the background.
	...
0.9211	Botswana typically has high quality main roads and the environment is a blend of desert and savanna.
0.9204	Botswana's main roads are in excellent condition, looking almost brand new.
0.9173	The soil is typically not as red as the Ugandan and Kenyan soils.

Qualitative Results

GT Location: Albania

ISN Prediction: Albania

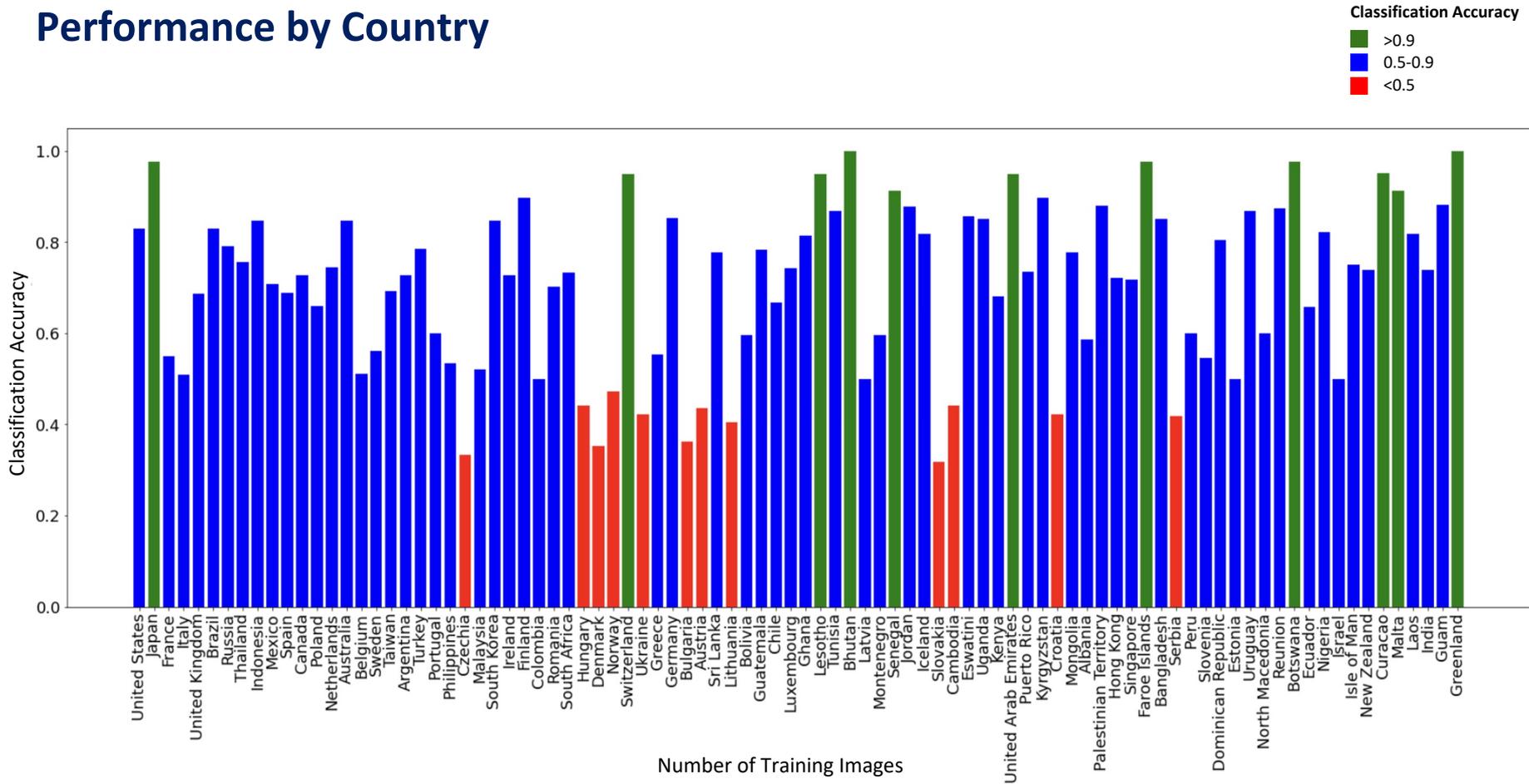
ISN + CLIP Prediction: Albania

G^3 Prediction: Turkey



Top-10 Retrieved Guidebook Clues	
0.9264	The Albanian language stands out for having lots of q, j and k letters.
0.9129	Black and white corner signs are a common sight in Albania.
	...
0.9104	Turkish roads on GeoGuessr are often wide and I find them to be among the widest roads in the world.
0.9011	Turkey uses either blue or green coloured directional signs.

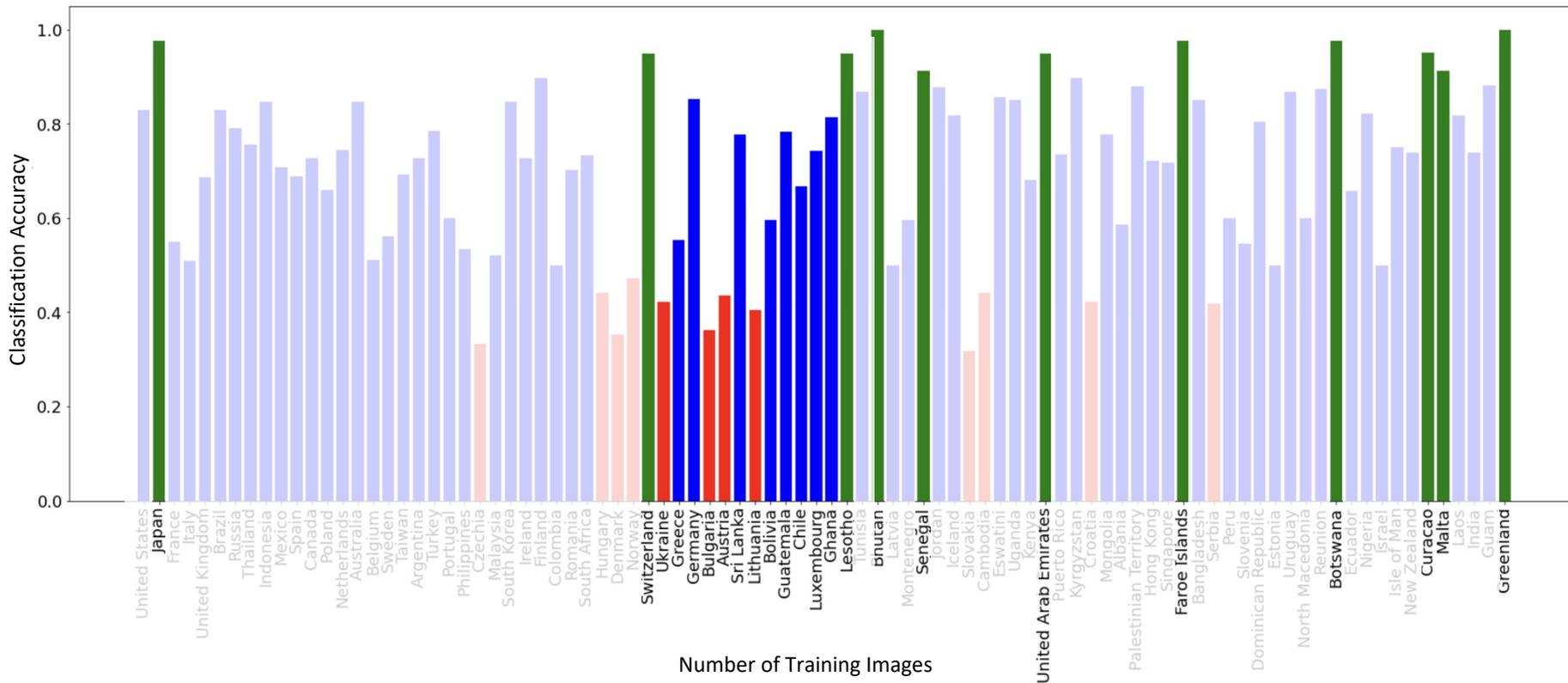
Performance by Country



Performance by Country

Classification Accuracy

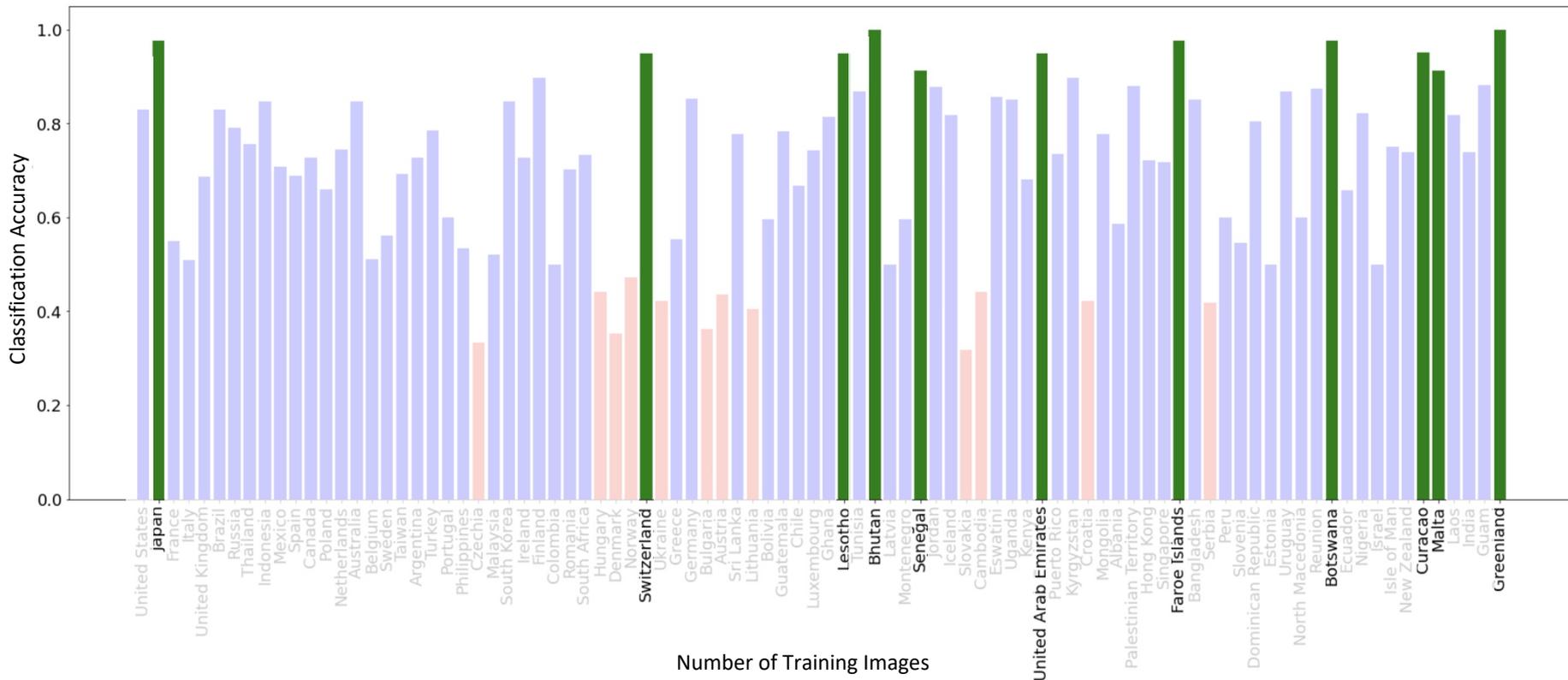
- >0.9
- 0.5-0.9
- <0.5



Performance by Country

Classification Accuracy

- >0.9
- 0.5-0.9
- <0.5



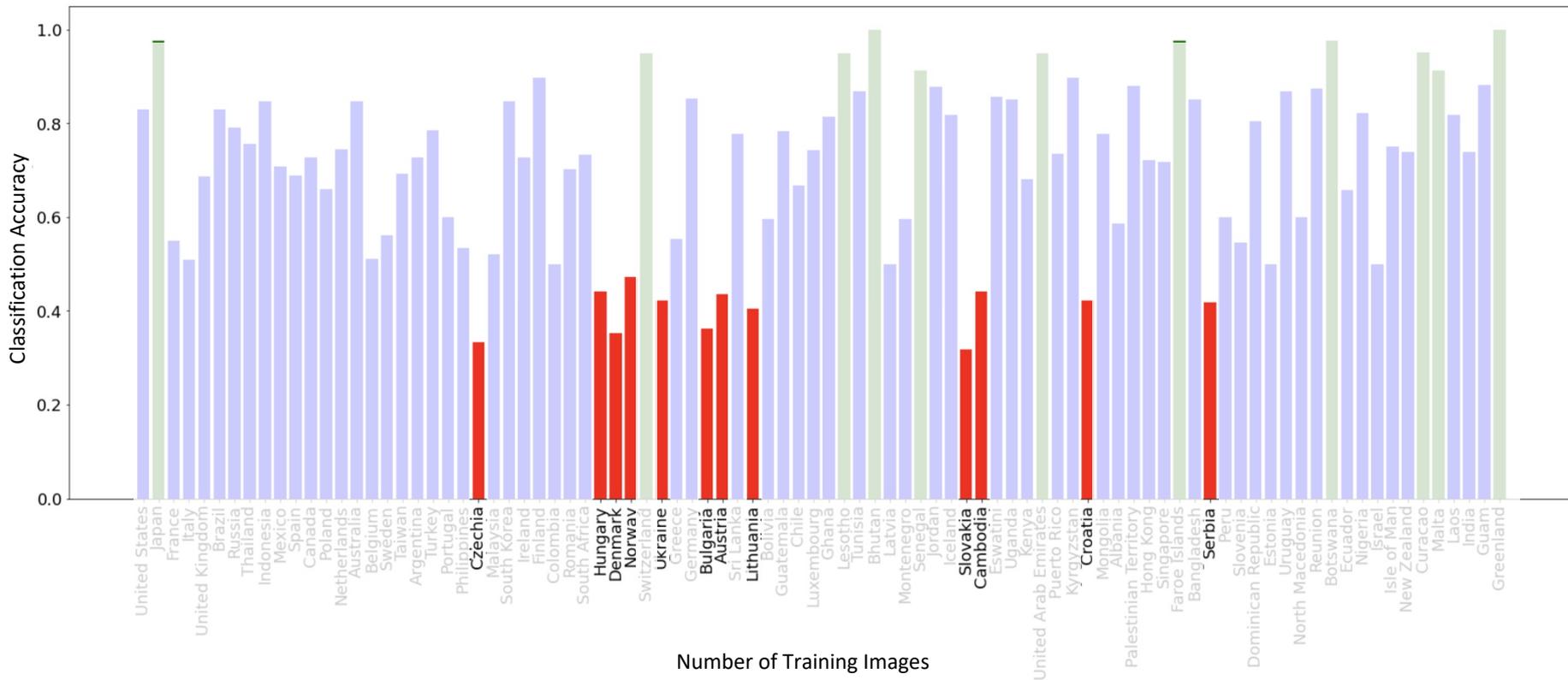
Performance by Country

Classification Accuracy

>0.9

0.5-0.9

<0.5



Summary

- We present the new multimodal task of Geolocation via Guidebook Grounding and an accompanying method G^3 .
- While we collect the StreetView images and guidebook clues *independently, with no manual labeling* of which clues are relevant, because they relate to the same domain of street level imagery our method is able to learn how to bridge the two modalities.
- We observe that our method still struggles with distinguishing between neighboring and visually similar countries, and that this seems to be a stronger indicator of performance than representation in the training set.

Thank you!

https://github.com/g-luo/geolocation_via_guidebook_grounding

