Applying Computer Vision to Geo-Locate Imagery

Jamie Luna[1], Rachel Roggenkemper[2], Matteo Shafer[2], Dr. Hunter Glanz[2], Dr. Jonathan Ventura[1]
[1] Computer Science and Software Engineering, Cal Poly, [2]Department of Statistics, Cal Poly

CAL POLY
Bailey College *of* Science & Mathematics

Introduction

Human trafficking affects approximately 21 million victims and generates an estimated $50 billion in revenue for traffickers each year. Global Emancipation Network (GEN), a 501(c)(3) aiming to disrupt human trafficking networks globally, used Meta AI's Segment Anything Model (SAM) and OpenAI's Contrastive Language-Image Pretraining (CLIP) model to predict the origin of a given photograph. We applied Zero Shot Instance Segmentation on top of this model to identify and label objects within the given photograph. Our work explains the predictions made by the SAM and CLIP models. This approach provides insight into how the location prediction was made and the influential factors in making that prediction. This approach allows victims to be identified and found more effectively, aiding GEN in their mission.

Methods

We use several models layered sequentially to achieve our goal.

1. Segment Anything Model to break images into smaller segments.
2. Contrastive Language-Image Pretraining model for image geolocalization based on the prompt.
3. Zero Shot Instance Segmentation for object recognition

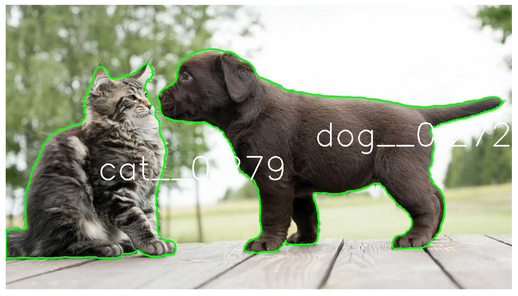

Figure 1: Zero Shot Instance Segmentation example output

Results

To improve explainability, we added several features to the our applet:

- An editable list of possible location outputs (default options: 50 states or all countries)
- An editable input prompt to the model
- Outputting the top 5 location predictions for the input image
- The probability of each location prediction in the output

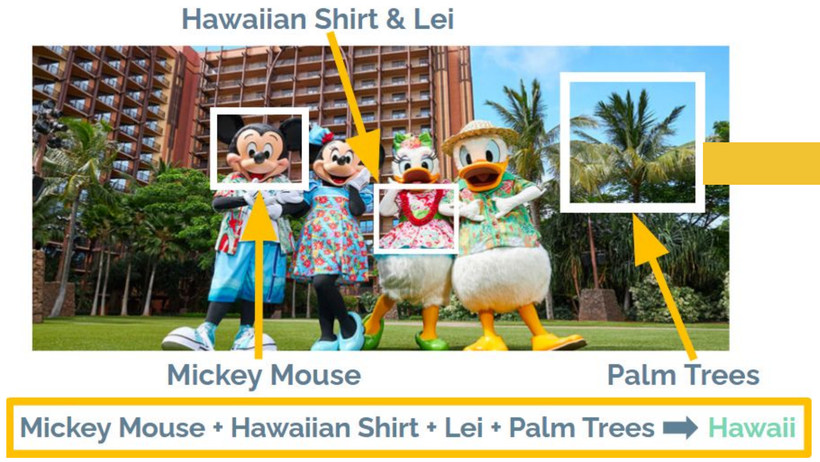

Figure 2: Ideal scenario - Relevant segments and object recognition leading to a correct location prediction

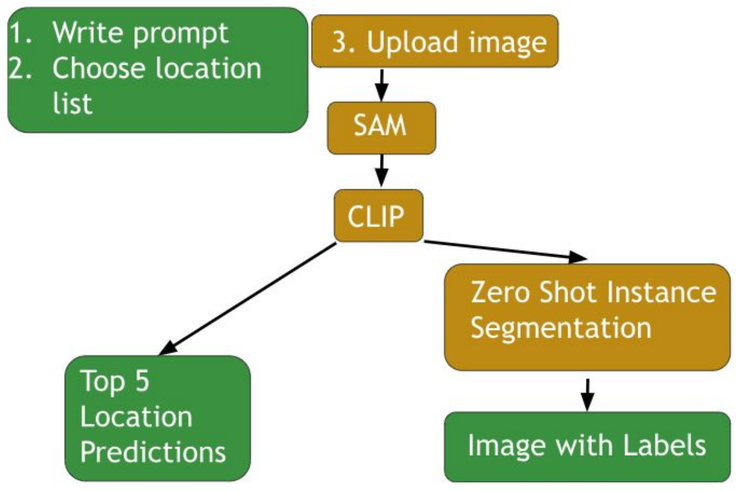

Figure 3: System flowchart for our applet

Future Work

With the current implementation, objects are not consistently recognized correctly. There are several steps we intend to implement to improve consistency:

- Increase the probability of the top location predicted
- Applying image filters to increase contrast, saturation, and exposure
- Identify influential image segments to label and weight them

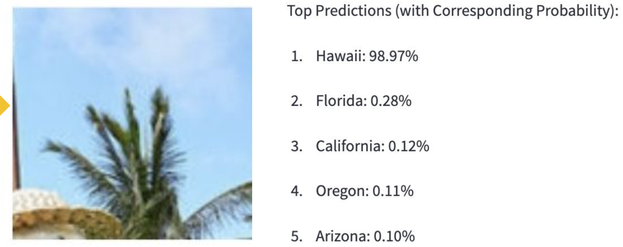

Figure 4: Example output of location predictions for a segment

References

"Meta Ai's Segment Anything Model (Sam) Explained: The Ultimate Guide." *Encord*, encord.com/blog/segment-anything-model-explained/. Accessed 30 Apr. 2024.

Passarelli, John. "Jvpassarelli/Sam-Clip-Segmentation: Image Instance Segmentation - Zero Shot - Openai's Clip + Meta's Sam." *GitHub*, github.com/jvpassarelli/sam-clip-segmentation. Accessed 30 Apr. 2024.

Acknowledgements

Sherrie Caltagirone, Global Emancipation Network

Fred Lichtenstein, Camera Forensics