Where We Are and What We’re Looking At:
Query Based Worldwide Image Geo-Localization Using Hierarchies and Scenes
Brandon Clark*  Alec Kerrigan*  Parth Parag Kulkarni  Vicente Vivanco Cepeda  Mubarak Shah
Center for Research in Computer Vision, University of Central Florida

Classification Approach to Geolocalization

Objective - "To determine the location of an input image, by classifying it into its correct geographical cell" [1]

Contributions
❖ The first Transformer Decoder for worldwide image geo-localization.
❖ The first model to produce multiple sets of features for an input image, and the first model capable of extracting scene-specific information without needing a separate network for every scene.
❖ A new testing dataset that reduces landmark bias and reduces biases created by social media.
❖ A significant improvement over previous SOTA methods on all datasets.
❖ A qualitative analysis of the features our model learns for every hierarchy and scene query.

Google-World-Streets-15K Dataset

Distribution of images across the world for the proposed dataset compared to existing validation datasets

Goal - eliminate biases towards commonly visited locations or landmarks and have a more even distribution across the Earth
14,955 images covering 193 countries

Collection steps:
1. sample a country
2. select a random city
3. sample a GPS coordinate within a 5 km radius of the center of the city

GeoDecoder Architecture

Novel decoder-based architecture designed to learn unique sets of features for every scene and every hierarchy

Components:
- Transformer Encoder
- Hierarchy Independent Decoder
  - SWIN [2]
  - Hierarchy Independent Decoder
  - Feature Selection Using Scene Confidence
  - Hierarchy Independent Decoder

Goal
- eliminate biases towards commonly visited locations or landmarks and have a more even distribution across the Earth

Reference