Machine Learning for fine-grained Class-agnostic Visual Counting

Viet Bach Tran
Supervised by: Prof. Minh Hoai Nguyen

What?

- Visual counting is a big problem in the Computer Vision.
- Building a comprehensive text-only dataset consisting of 1,000,000 images of more than 50 categories, focusing on fine-grained visual counting.

Dataset Design

- Number of categories: 50
- Sub-categories: stacked, spreaded, mono-color, multi-color, multi-shape, etc.
- Background: different backgrounds for different categories
- Example list of categories:
    - Food: rice grains, peas, M&Ms, cornflakes
    - Fun: poker chips, dices, coins
    - Office: pins, paper clips, sticky notes
    - Household: nails, screws, bolt nuts, marbles
- Each category, 10 counts_total
- Each counts_total, 10 scene
- Each scene, 100 images of different angles

Why?

Image ID: 1114, Pred: 147, GT: 28
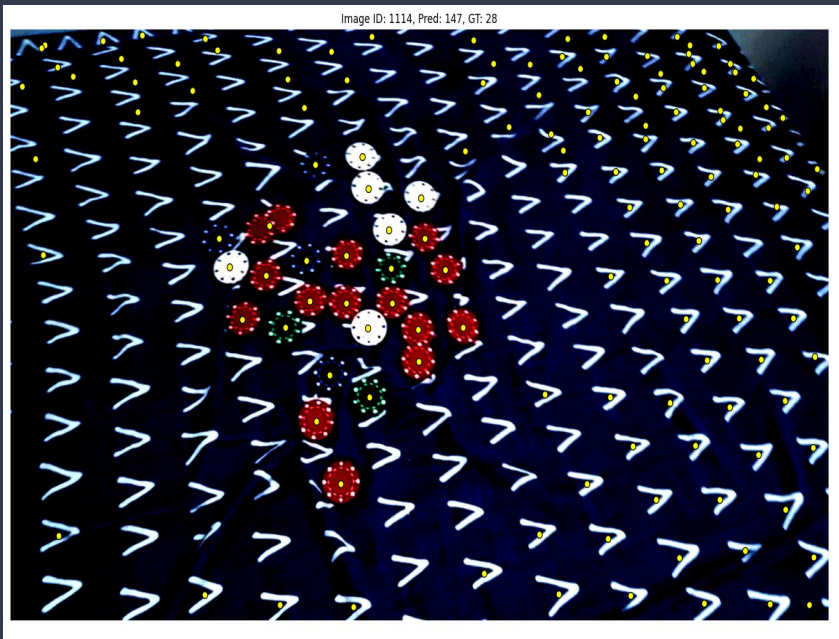
(Object: "poker chips", background: my bed)
Pred: 147, GT: 28

- Real-world applications
    - Count the number of apples in a basket and differentiate between fresh and rotten apples
    - Count the number of people in the crowd who are looking at you
    - Count the number of Swifties in a group of Uni students

- Scientific gaps
    - Current SOTA models don't perform well in fine-grained counting based on the object's attributes (e.g. color, shape, etc.)
    - A lack of high-quality dataset for evaluation, especially for class agnostic fine-grained counting (the most common dataset right now only has more than 6,000 images)

First results

- Have taken 5,000 images of poker chips in different environments
- Have shown really low accuracy (MAE and RMSE) in SOTA models

First challenges

- The images' quality is not very high
- The time to build the dataset is very long (~200 hrs)
- The money to buy all the objects need is very high (~500$)
- The number of objects per category is pretty small (~max 10 apples)

Goal

- Dataset: Have a robust and diverse dataset
- Code: Have a ready-to-run codebase for testing all the SOTA models
- Progress: Have 30% of a CVPR-tier paper

Project Evaluation

- Dataset: Check for diversity, robustness, and practicality
- Progress: Assess with supervisor for current progress
- Experiment: Show low accuracy (MAE & RMSE) in SOTA models for the dataset

Thanks for listening!

Feel free to ask me any questions