Team Members: Sneha Gonipati, Zinnia Waheed, Aiah Aly, Heron Ziegel, Faisal Shaikh

Hotel-ID to Combat Human Trafficking: Project Charter II

Business Background

Human trafficking is a widespread global problem that affects millions of individuals, often occurring in concealed environments like hotel rooms. Organizations and law enforcement agencies that are dedicated to fighting human trafficking face significant challenges in identifying the locations of victims based on limited visual evidence. Frequently, photographs taken in hotel rooms serve as the only clues to identify these locations. However, manually analyzing these images is time-consuming, labor-intensive, and prone to errors. With thousands of hotels, each with unique room configurations, accurately pinpointing locations requires an automated, scalable solution. This project aims to develop a machine learning model capable of matching images of hotel rooms to their respective hotels. By providing data-driven, efficient solutions, this project will enable faster response times and improve the success rate of victim rescue operations. The client domain is human rights and crime prevention, where rapid identification of trafficking locations can assist in rescue operations. With our tools and expertise, the project leverages AI and computer vision to build an automated, scalable image recognition system that can match hotel room photos to specific hotels to tackle societal human rights issues.

Scope

Develop a computer vision model that can efficiently classify hotel images and predict the hotel ID from a new image. Deep machine learning techniques like convolutional Neural Networks (CNNs) or transformer-based models can be used to build a model using previous hotel room pictures to identify a new hotel room. Deploy an inference pipeline system that is capable of taking an image as its input and outputting the top 5 most likely hotel ID predictions. We have integrated EfficientNet and ResNet hybrid models for classification and are working to improve scores and enhance the baseline models. Incorporating Image augmentation and feature extraction to improve model robustness against variations in lighting, angles, and occlusions. Experimenting with data augmentation strategies like rotation and scaling to increase model accuracy. In addition, preprocessing strategies such as adaptive histogram equalization and contrast normalization are incorporated to enhance image consistency before feature extraction

After speaking with a member from the research team, Andrea Brandt, she recommended we integrate DINOv2 for stronger feature representations. This would be the best-performing model for this given task, as DINOv2 is considered an SOTA method for self-supervised learning and feature extraction, ultimately outperforming older models like ResNet in certain tasks. Switching from ResNet to DINOv2 could offer various benefits, like better generalizations and more enhanced feature extraction. This method learns high-quality representations that capture more details, and its embeddings can be directly compared using FAISS or cosine similarity, making DINOv2 most ideal for retrieving similar hotel images as FAISS or cosine similarity is most helpful for embedding-based retrieval. One of the model engineers also recommended cross-entropy loss with label smoothing to help prevent overconfidence in incorrect predictions

and improve generalization. Cross-entropy loss is a function used in classification problems to ensure the model generalizes well across a large number of classes while handling real-world complexities. Label smoothing enhancement prevents model overconfidence by assigning small probabilities to incorrect classes. This reduces overconfidence in misclassified predictions, which ultimately improves generalization and ensures the model does not overfit to dominant hotel classes. Since the data set is so large, the probability of misclassification is high, but with cross-entropy, model updates are proportional to misclassification confidence. This allows the training process to scale effectively without numerical instability. By leveraging DINOv2 for feature extraction and cross-entropy loss for scalability, it improves the model's ability to differentiate subtle hotel room features while also ensuring stable probability distributions across all hotel IDs. Cross-entropy pairs with DINOv2 for better feature learning. DINOv2 embeddings capture more meaningful spatial details in hotel images, while cross-entropy loss ensures these embeddings remain well-separated, enhancing training stability and scalability.

Performance metrics will be evaluated using top-k accuracy, precision-recall, and other relevant metrics. Track model performance over iterations and compare new metrics to the baseline to optimize real-world performance. Solution can be incorporated by consumers like law enforcement and NGOs can then upload an image from some source and in turn will get a list of predicted hotel ID’s. This can assist in investigations where these predictions can be used as leads to verify and locate trafficking victims. This model can also potentially be integrated with other law enforcement databases to accelerate investigations, improve resource allocation, and enhance victim recovery efforts.

Personnel

Sneha Gonipati
Roles: Model Engineer
Tasks: Managing Trello tasks, working with the EfficientNet Model, data manipulation, Github

Zinnia Waheed
Roles: Data Engineer, Model Engineer
Tasks: Improve ResNet and try to incorporate it with EfficientNet

Aiah Aly
Roles: Model Engineer
Tasks: Random assignment model, research/apply feature extraction techniques to improve model accuracy/robustness, replace ResNet embeddings with DINOv2 for stronger feature representation

Heron Ziegel
Roles: GitHub coordinator, front end if needed, ML programming
Tasks: GitHub, data manipulation, Resnet & EfficientNet hybrid model

Faisal Shaikh
Roles: Data engineer (preprocessing/EDA)
Tasks: Github, Resnet model improvement

RANDOM ASSIGNMENT

Submission and Description	Private Score	Public Score
Fork of Trial1 - Version 4 Notebook Running (after deadline) · 4m ago	0.012	0.018
Fork of Trial1 - Version 1 Submission Scoring Error (after deadline) · 32m ago		

RESNET (improvement from the addition for test time augmention)

Submission and Description	Private Score	Public Score	Selected
Resnet - inference - Version 2 Succeeded (after deadline) · 1h ago · Notebook Resnet - inference | Version 2	0.172	0.181	☐
Resnet - inference - Version 1 Succeeded (after deadline) · 1h ago · Notebook Resnet - inference | Version 1	0.156	0.166	☐

Metrics

How will we quantify the success of the project?
The performance will be evaluated based on the Mean Average Precision at 5 (MAP@5) metric, which calculates the precision of the top five predictions made by the model for every image. MAP@5 assigns higher scores to models that rank the correct label earlier in their predictions. This metric specifically aligns the model's effectiveness with the project objective of effectively detecting hotels and, hence, is highly pertinent to practical applications such as facilitating law enforcement efforts in the fight against human trafficking. We predicted and achieved an extremely low MAP@5 score for the random assignment model that we used as our baseline for comparison. We have since surpassed that score and are working to improve model accuracy and robustness. With DINOv2 we expect an additional improvement surpassing 0.20 against the leaderboard. We are tracking model success against the leaderboard, of course, but also across all model versions as we try to improve model accuracy with various features.

What are the common metrics used by others working in the domain?
On image recognition and ranking tasks, a number of measures are most commonly utilized, such as Mean Average Precision (MAP), precision and recall, top-k accuracy, and F1-score. MAP calculates precision for all the predictions and is most suitable to applications where ranking and relevance are the focus. Precision calculates the proportion of relevant results out of the total predictions, whereas recall calculates the proportion of correctly predicted relevant results out of the overall set of possible relevant results. Top-k accuracy checks whether the true label is among the top k predictions, and Top-5 accuracy is one of the standard evaluations. F1-score as the harmonic mean of precision and recall is utilized to balance between the two to

yield a balanced metric for classification tasks. In this project, MAP@5 is utilized with an emphasis on the top five predictions to guarantee an accurate and pertinent ranking to render it specifically suitable for improving hotel matches here.

What is a quantifiable metric?
The MAP@5 score is the main measurable metric that assesses the model's capacity to rank the correct hotel label in its top five predictions for every image accurately. By yielding a definite numerical score, the MAP@5 score makes it simple to compare various models and enables improvement to be tracked over time.

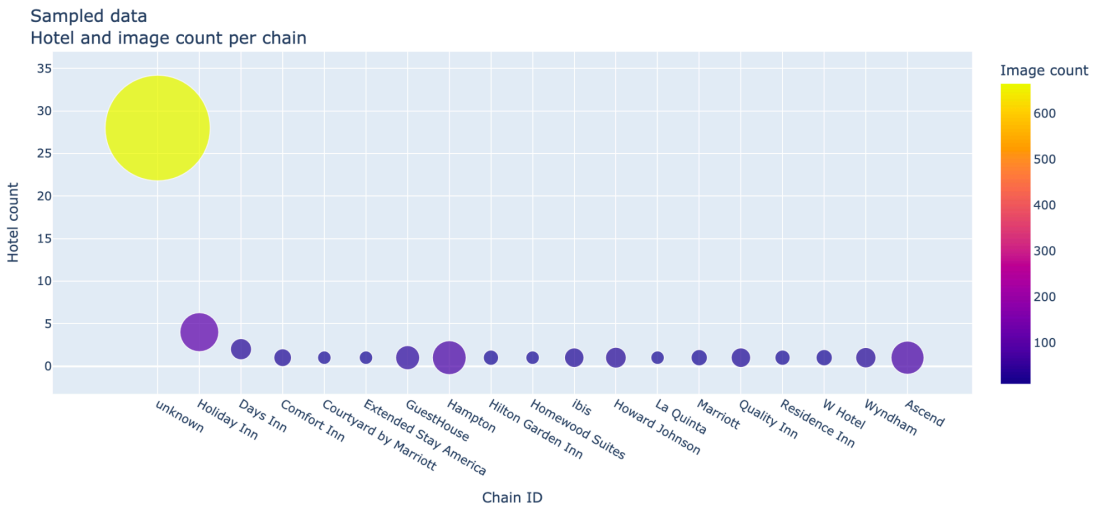
How will we measure the metric?
The MAP@5 measure will be evaluated by comparing the model's predictions on a test or validation set of images where we know the ground-truth labels. The MAP@5 score overall will be the average precision over the set of all images. Further, the model's performance will also be compared against a baseline, e.g., a naive or an already-existing model, to see if there is an improvement. The competition leaderboard establishes benchmarks, enabling direct comparison of the MAP@5 score against others and thereby guaranteeing an intensive and systematic assessment process.

Plan
Deliverable Timeline:
Project Charter revised due 3/14 *(Incorporate feedback and finalized scope)*
Phase 2 Project Demo due 3/26 *(Present improved ResNet + EfficientNet model*
Final model with no masks 3/30 *(Train DINOv2 feature extraction model)*
Bimonthly Project Report 2 due 3/31
Data Report due 4/7 *(Document data sources, preprocessing, and augmentation)*
Model + Performance Report due 4/14 *(Compare ALL model performances)*
Phase 3 Project Demo + Interactive Dashboard Demo due 4/28 *(Deploy model with masking interactive dashboard for hotel ID predictions)*

Architecture
What data do we expect?
The data was collected by taking pictures of various hotel rooms, making sure that they were relatively natural and captured most, if not all, of the room. There are a different number of pictures for each hotel room. For each hotel, there is a subfolder present on the dataset with a unique numerical identifier, and the hotel rooms consist of places from across the globe. Each picture is represented as a JPG file. The occlusions are represented as red blocks covering random portions of the picture, with transparent portions for the rest of the image. These masks are represented as PNGs and are used to simulate elements such as people, jackets, or other large objects that might cover portions of the image. Labeled hotel images dataset and unlabeled images for testing model generalization were available through the Kaggle competition. This data was not collected by personnel but instead downloaded from the “Hotel-ID to Combat Human Trafficking 2022 - FGVC9” competition from Kaggle. If the data

from this project is insufficient, the group may pull from the previous year's competition with similar data to fine-tune the model.

What do we expect and how will it be stored/operated on?
The method of storage and operation will be personnel-specific. For the personnel working on improvements to the ResNet model, the Kaggle interface and data storage are sufficient as advanced models are not yet required, and various augmentations can be quickly implemented and tested in this interface. For the personnel working on trying to implement the combination of EfficientNet and ResNet model, Google Colab is being used for their GPU functionality and the easily accessible storage. For the personnel working on the Efficient Net model, VS Code is being used to host the SSH virtual machine offered by Temple University for its incredible fast processing speeds to handle the multiprocessing required for that model. Feature extraction and embedding storage in FAISS once DINOv2 model is active. Despite the different interfaces, the code is being worked on piecemeal, so requisite pieces are being transferred between personnel as they are needed, and constant communication is occurring so that everyone is up to date on everyone else's portion of the project.
On what platform would the client/user access any deliverable services (e.g. dashboard) from the project?
As of right now, the dashboard for delivering the services has not been finalized, but the dashboard will likely include a source to upload the photo in question. Once the photograph is uploaded, a list of the hotels with the most likely hotel match will be presented to the user. Model components like data preprocessing and feature extraction using DINOv2 embeddings for improved image similarity search are being deployed alongside other classification models for hotel ID prediction. The image below represents EDA where the hotel and image count per chain is analyzed from the dataset. Currently, the hotel names are deidentified, but if a dataset with photographs labeled with the hotels in question is given, then along with the list of hotel names and addresses will be given so that the officials using the program will have all the necessary information immediately.



Communication

Trello
We have created a Trello board for the project, which can be continuously updated. There are three sections: “To Do,” “Doing,” and “Done.” As we determine specific deliverables, we will add them to the “To Do” section. When someone starts working on a deliverable, they can add their name to the card and move it to “Doing,” then finally to “Done” when it’s completed. We may want to update this with more lists such as “Needs Review” and “Deployed”, depending on the needs of the project.

Group Chat
We have created a text message-based group chat with all members of the team. We will use this to send messages to the group, plan meetings, and stay on top of deliverables.

GitHub
We will use GitHub for version control to streamline this project. We will create new branches for different elements and send pull requests to the main branch once an element is ready to be included in the project.

Meetings
We will schedule times to meet outside of class periodically during this project. Meetings could be in-person, particularly after the lab session on Tuesdays. Or they could be virtual using either Discord or Zoom.

Relevant Links
Project Presentation:
https://docs.google.com/presentation/d/1HWGOthgNwgN3X7A3k2vuHh4iXQXQJlDY11DCfTt6Vn0/edit?usp=sharing

Kaggle Project:
https://www.kaggle.com/competitions/hotel-id-to-combat-human-trafficking-2022-fgvc9/overview

Github Link:
https://github.com/hziegel/Combat-Human-Trafficking

Example Charter:
https://github.com/Azure/Azure-TDSP-ProjectTemplate/blob/master/Docs/Project/Charter.md