Geoguessr Classification Proposal

Aaditya Reddy Anugu, Justin Kang, Nathaniel Alexander Koehler, Patrick Soo, Zhixuan Wang

1. Intro/Background

Geoguesser is a game in which players are randomly placed somewhere in Google Street View and need to guess what their exact location is.

Previous literature [3][5] discusses using techniques like CNNs and transfer learning to analyze other image datasets to identify pneumonia in x-rays and find skin lesions and skin cancer from pictures. These papers have used these techniques to identify features that can help models classify an image. However, one of the shortcomings previous literature emphasizes is a lack of sufficient data in training models.

The following dataset, GeoLocation - Geoguessr Images (50K), found through Kaggle, contains 50,000 streetview images of the world, with every image belonging to 1 of 150+ countries. The data itself is not uniform as there are more images within certain countries compared to others, but we plan to combine datasets and prune folders with insufficient data.

2. Problem Definition and Motivation

We are interested in seeing if we can train a model to accurately perform this task of identifying key objects that belong to only specific parts of the world, and correctly identifying which country the street view image is from.

This brings us to our problem - there may be certain circumstances in which it would be helpful to determine a relative location given a set of images, such as crime investigations. Thus, our motivation towards a potential solution to this is to start by using the Geoguessr dataset found through Kaggle, and train the dataset to determine which country it is in.

3. Methods

The first preprocessing method we will employ is standardization [2], which is defined by $X' = \frac{X-\mu}{\sigma}$ and will help reduce differences in lighting and exposure. The second preprocessing method will be min-max scaling [2], which normalizes the 0-255 pixel range to 0-1 for easier model calculation, and is defined by $X' = \frac{X-X_{min}}{X_{max}-X_{min}}$. Lastly, the third preprocessing method will be log scaling, which essentially applies the log function to every data point [4]. Other than scaling the image array values, we will also clean our data by omitting countries (classes) with less than 100 images to prevent overfitting of some classes.

As it is an image classification problem, our first model will be a CNN – a convolutional neural network. This will be a quite simple model, consisting of layers such as conv2D, pooling, dropout, and fully connected layers. The second model will be a transfer learning model, fine tuning an already trained model (e.g. DenseNet121). The third model will be a KNN (K-Nearest Neighbors) model that we learned in class.

4. (Potential) Results and Discussion

An expected result of this project would be providing the program a picture(s) of a specific place, and it will guess what country the image(s) are from. Although we understand that the program will make mistakes on labeling the country, our hope is that it will be more informed and more accurate compared to the average uninformed person.

5. References

[1] R. K., "Geolocation - Geoguessr images (50k)," Kaggle, https://www.kaggle.com/datasets/ubitquitin/geolocation-geoguessr-images-50k (accessed Feb. 20, 2024).

[2] A. Bhandari, "Feature scaling: Engineering, Normalization, and standardization (updated 2024)," Analytics Vidhya, https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/ (accessed Feb. 20, 2024).

[3] Kim, H.E., Cosa-Linan, A., Santhanam, N. et al. Transfer learning for medical image classification: a literature review. BMC Med Imaging 22, 69 (2022). https://doi.org/10.1186/s12880-022-00793-7
2
[4] "Normalization | Machine learning | Google for developers," Google, https://developers.google.com/machine-learning/data-prep/transform/normalization (accessed Feb. 20, 2024).

[5] Weyand, T., Kostrikov, I., Philbin, J. (2016). PlaNet - Photo Geolocation with Convolutional Neural Networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9912. Springer, Cham. https://doi.org/10.1007/978-3-319-46484-8_3