GeoBot: Pattern Recognition and Leveraging CNNs for Accurate Country Identification in Geolocation

Michael Salton
Department of Science
Western University
msalton@uwo.ca

Matthew Tennant
Department of Science
Western University
mtennan7@uwo.ca

Jiaqi Yuan
Department of Science
Western University
jyuan233@uwo.ca

Emily Grant
Department of Science
Western University
egrant47@uwo.ca

Benjamin Oliveria
Department of Science
Western University
boliveri@uwo.ca

***Abstract*—GeoBot is an advanced machine learning model designed to excel in the challenging task of country prediction from images. Leveraging a convolutional neural network (CNN), GeoBot has been meticulously trained on a diverse dataset of 50,000 images extracted from Google Street View, encompassing various global locations. These images encapsulate a rich spectrum of geographical features, including landmarks, street signs, road markings, license plates, foliage, and landscapes, providing the model with a comprehensive set of cues for accurate country identification.**

This research undertaking encompasses the critical phases of dataset curation, model training, and fine-tuning to optimize the AI's precision and computational efficiency in geographical prediction. GeoBot is equipped with state-of-the-art tools, incorporating Yolov8 for image classification, ensuring the model's competence in handling intricate visual data.

The outcomes of this study promise to illuminate the challenges and potentials of image classification, shedding light on its capabilities for country recognition in the context of diverse geographical landscapes. GeoBot represents a significant leap towards harnessing the power of artificial intelligence for geolocation tasks and demonstrates the growing synergy between machine learning and computer vision in real-world applications.

AI : Artificial Intelligence
CNNs : Convolutional Neural Networks
GPS : Global Positioning System

I. Introduction

Over the years there has been significant advancement in AI with developments in how machine vision can be used to interpret and interact with visual information from the world. One application of machine vision that has gained significant prominence is the task of identifying the location where a given image was taken, called geo-localization. This application of machine vision impacts a wide range of fields such as augmented reality, robotics, and self-driving vehicles [1]. In pursuit of understanding and constructing methods to implement machine vision, various problems need to be considered. One of the fundamental problems in computer vision is image classification, where the objective is to have the application analyze a test image and subsequently assign the image to a specific class from a given set [2]. Additionally, there is also the task of optimizing the image classification process so that it is accurate and can classify positional images of different spacial regions, giving it the potential to be scalable [3].

A. Motivation

Image classification and pattern recognition is a widely used method in machine learning to solve many modern-day problems, especially in interdisciplinary research, where the old traditional methods are considered ineffective, costly, and obsolete [4]. In the evolving field of machine learning, it is within this context that our research finds its motivation to contribute to the interconnected field of machine vision and image classification.

Our goal is to create a bot that is efficient in recognizing patterns and location pointers that hint at the exact country where the image was taken. In most cases, patterns that are commonly analyzed are vegetation, road markings, signs, languages on road signs, or other man-made objects, such as vehicles or infrastructures [4]. Pattern recognition for machine learning is often based on pre-existing datasets and learned features [5]. This reliance is also essential to enhance GeoBot's ability to extract meaningful information so its learned patterns can increase the accuracy of classifying various location snapshots.

Developing a bot that can correctly classify images to country names has various potential applications outside of research, such as tourism to help users identify location spots or organizing user-created content online based on the location of images posted.

B. Related Works

Our research is grounded at the intersection of geo-localization and image classification, two approaches that have already been extensively researched within the field of machine learning [1]. To understand how these two approaches can be applied together to solve machine vision problems it is important to understand what research is already done, and to explore any potential discussions which emerged from the results of said research.

The conference paper, written by E. Müller-Budack et al., titled "Geolocation Estimation of Photos using a Hierarchical Model and Scene Classification", looks into using different deep learning methods to predict GPS coordinates based on a given image. Similarly to the topic of our research, this paper also focuses on the problem of using a model to identify the location in which a picture was captured. While our research addresses this problem by using YOLO trained on ImageNet for object detection, Müller-Budack et al.'s study makes use of the ResNet model that has been trained on ImageNet [7]. It is important to note that there are research gaps between Müller-Budack et al.'s study and our own. For instance, the image set used in their study consists of 4,723,695 images which greatly surpasses the size of our own set of training images. We can attribute this difference to our limited resources for our study. The findings of this particular study present a CNN approach that integrates hierarchical knowledge to demonstrate efficiency without using complex methods that rely on a specific reference dataset [7].

C. Problem Definition

This research wishes to reveal both the challenges and possibilities within the field of image classification and geo-localization. To achieve this objective, we take advantage of the recent advancements in the YOLO model, specifically YOLOv8, which is the latest released version of the model as of the time of this report. The YOLO model has had significant advancement through improved performance and newly added features present in YOLOv8 [8]. However, despite the upgrade in quality from the previous model version, YOLOv7, there are few comprehensive studies or reports that highlight the potential limitations, challenges, and versatility of YOLOv8 for image classification and geo-localization. Therefore, through practical testing and fine-tuning, this research strives to fill this gap to further understand the applications of state-of-the-art tools such as YOLOv8 when combined with the computer vision problem of geo-localization.

From solving the challenges which present themselves in YOLOv8 and image classification, we address the central problem of this research which is to achieve accurate prediction of a country, based on an image using only the visual information contained within it. The problem is complex due to the diverse range of geographical features and cues that must be considered, including landmarks, street signs, road markings, license plates, foliage, landscapes, language, people, and other visual elements. The task involves image classification intending to identify the country associated with the image.

D. Objectives

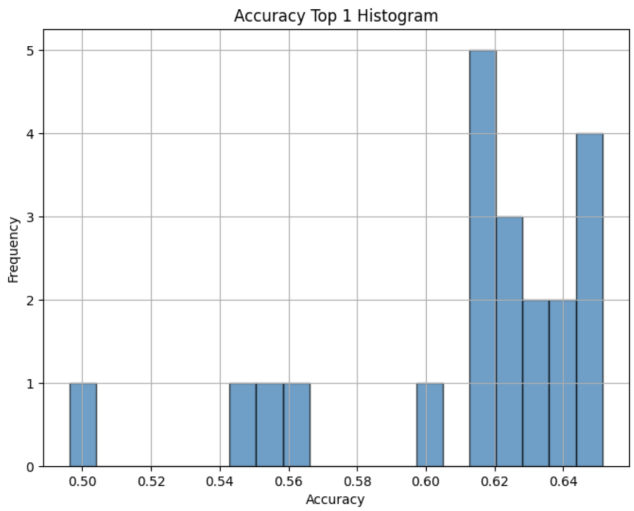
In pursuit of advancing the capabilities of machine learning in the realm of geo-localization, this research encompasses several key objectives. Firstly, we aim to develop an advanced machine learning model, GeoBot, which harnesses the power of CNNs and cutting-edge image classification techniques to achieve precise country prediction solely from images. To enable the effective training and refinement of GeoBot, we meticulously curate a vast and diverse dataset comprising 50,000 images sourced from Google Street View. This dataset spans various global locations, providing GeoBot with a rich and comprehensive set of visual cues for accurate country identification (fig.1). Furthermore, we place a significant emphasis on optimizing GeoBot's precision and computational efficiency, ensuring its competence in processing intricate visual data. Finally, we explore the potential real-world applications of GeoBot, ranging from aiding tourism by helping users identify location spots to enhancing content organization based on image location information. Through these endeavors, our research seeks to illuminate the possibilities and challenges within the intersection of machine learning, computer vision, and geo-localization.

Fig. 1. Validation Batch.

E. Challenges

Several critical challenges must be addressed in the development of GeoBot for accurate country prediction from images. First and foremost, the model must contend with the inherent variability in visual cues present in images captured across different countries. This variability encompasses factors such as language on road signs, variations in vegetation types, diversity in road markings, and differences in architectural styles. To ensure the model's effectiveness, it is imperative that it can adeptly capture and analyze these diverse visual cues, a fundamental requirement for precise country prediction. Additionally, scalability is a crucial consideration, as GeoBot must be capable of classifying images from a wide array of spatial regions (fig. 1), accommodating the extensive range of geographical landscapes and features found across the globe. Furthermore, the availability of a substantial and diverse dataset is vital for the model's robust training, although resource constraints may pose limitations on the dataset's size. Lastly, achieving the delicate balance between accuracy and computational efficiency in real-time image classification, particularly when dealing with large datasets, presents a formidable challenge that demands innovative solutions. These challenges collectively shape the core focus of our research as we endeavor to develop GeoBot into a highly capable tool for country recognition from images.

Fig. 2. Histogram of Accuracy.



F. Significance

The successful development of GeoBot and the resolution of the aforementioned challenges have the potential to revolutionize the field of geo-localization and image classification. This research contributes to the growing synergy between machine learning and computer vision, offering insights into the capabilities and limitations of image classification for country recognition. Moreover, the practical applications of GeoBot extend beyond research, with potential benefits in tourism, content organization, and other areas where image location identification is valuable.

G. Findings

Our final model achieved a top-1 accuracy of 65% and a top-5 accuracy of 85% on a model with 57.4 million parameters after 21 epochs (fig.2).

II. Methodology

In this section, we describe the methodology employed for training and fine-tuning the image classification model used in our research as well as details and processes we took when selecting and preparing the dataset. We adopted the YOLOv8 architecture as the foundation for our model, and this section outlines the key decisions made during the training process.

A. Dataset Selection

We relied on an existing dataset sourced from Kaggle, consisting of a collection of over 50,000 images. This dataset served as the foundation for our model development and training process. Nevertheless, as we delved into the dataset, we encountered a significant challenge related to data imbalance, a common issue in machine learning and image classification tasks.

The source of this data imbalance stemmed from the inherent variations in Google Street View coverage across different countries. In an ideal scenario, we aimed to have a uniform representation of images from all countries to ensure a fair and accurate training process for our model. Our primary objective was to secure a minimum of 100 images for each country in our dataset, as this threshold was deemed necessary to achieve a robust and comprehensive training set.

However, the reality of Google Street View's global coverage introduced complexities in achieving this goal. Some countries had an abundance of images available, while others were relatively underrepresented. This imbalance in the dataset posed challenges, as our model's performance could be adversely affected if certain countries were over-represented, leading to a bias in country prediction.

To address this issue, we embarked on data curation efforts, seeking to balance the dataset by augmenting underrepresented countries with additional images and potentially reducing the number of images from overrepresented regions. Our aim was to create a more equitable dataset that would enable GeoBot to learn and generalize effectively across all countries, ensuring that its country prediction capabilities were both accurate and unbiased. This process of data curation and balancing was a crucial step in the development of the model and played a pivotal role in its overall performance.

B. Data Pre-Processing

Our foremost challenge revolved around the uneven distribution of Google Street View coverage across various countries. Notably, countries like Canada, France, Luxembourg (fig.4), Spain, and the United States featured abundant coverage, yielding an excess of 500 images already present in our dataset. In contrast, countries like Costa Rica (fig.3), China, Pakistan, Vietnam, and several others exhibited limited Google Street View representation, offering fewer than 10 accessible images. Meeting the target of acquiring 100 distinct images for these countries posed a considerable challenge. To address this issue, we found it necessary to adapt our image count objectives for these nations as a pragmatic solution.

We encountered another minor challenge related to the geographical size of specific countries. Despite their relatively comprehensive Google Street View coverage, some countries proved too vast to practically collect 100 images. Conversely, for smaller nations, reaching a count of 100 images appeared both challenging and excessive. To address this issue, we opted to revise the image requirements for smaller countries such as Andorra and Luxembourg to approximately double the initial image count. To effectively tackle these challenges, we categorized countries with fewer than 100 images into three distinct groups based primarily on the extent of Google Street View coverage

Firstly, the easily accessible countries. These are countries with robust Google Street View coverage, where obtaining 100 images posed no significant obstacles. Secondly, the moderate coverage countries. Countries with reasonably good Google Street View coverage, allowing us to double the initially

Fig. 3. Country with low coverage (Costa Rica).

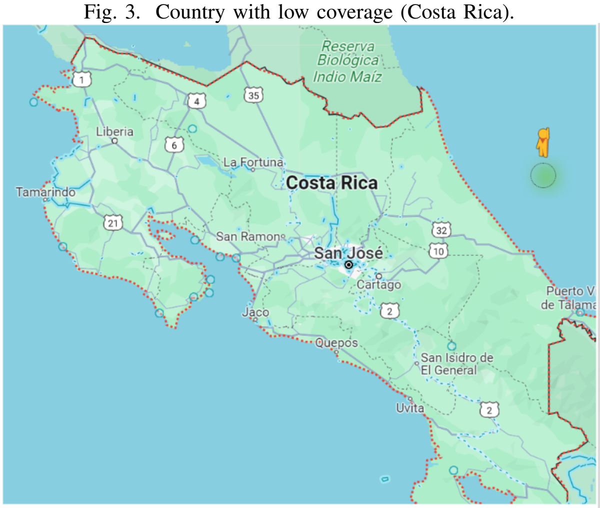


Fig. 4. Country with high coverage (Luxembourg).

planned image count to meet our target. Lastly, the substantial challenging countries. For countries with limited Google Street View coverage, gathering even a few unique images proved to be quite challenging. Consequently, we decided to exclude countries falling into this category from our dataset, as well as those with no Google Street View coverage at all.

C. Model Selection

We initially began with a pre-trained image classification model derived from YOLOv8, a well-established object detection framework. Our decision to use YOLOv8 as the base model was driven by its reputation for efficiency and effectiveness in object detection tasks.

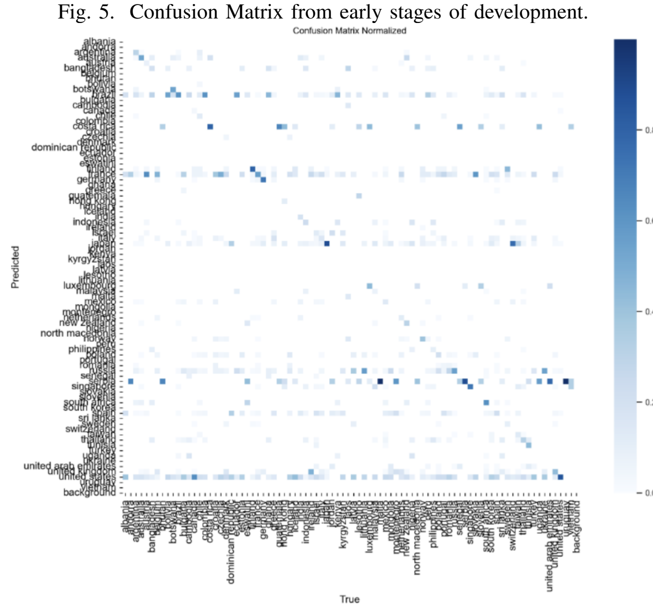
TABLE I
TRAINING SPLIT FOR THE YOLOV8 MODEL.

Training	Validation	Testing	Total
41147	5144	5144	51435

D. Image Size

The choice of image size plays a crucial role in the performance of the model. We initially started training with an image size of 64x64 pixels. However, early results indicated that this image size yielded sub-optimal performance(fig. 5). To address this limitation, we decided to increase the image size to 224x224 pixels. This adjustment allowed the model to capture more intricate details and contextual information, ultimately leading to improved accuracy in our classification task.

Fig. 5. Confusion Matrix from early stages of development.



E. Model Scaling

Our initial model choice, YOLOv8n-cls, had 2.7 million parameters and demonstrated a top-1 accuracy of 66.6% on the pre-trained model. However, during the training process, we observed that the model plateaued at an accuracy of approximately 35% after only 5 epochs. To overcome this performance bottleneck, we decided to adopt a larger model variant, YOLOv8x-cls, which boasts 57.4 million parameters. This model demonstrated a top-1 accuracy of 78.4% in its pre-trained state. We believed that this larger model could capture more complex features and patterns within our dataset.

F. Training Duration

Training deep learning models requires an optimal balance between computational resources and training time. With the YOLOv8x-cls model, we embarked on an extended training journey that spanned 21 epochs. This decision was made to ensure that the model had ample time to converge and adapt to the characteristics of our specific classification task. After about 18 epochs the model's accuracy started to plateau so we thought fit to end training soon after.

In summary, our methodology encompassed the selection of a powerful base model, careful consideration of image size, a transition to a larger model variant, and an extended training duration. These methodological choices were driven by a commitment to achieving the highest possible accuracy in our image classification task while maintaining computational feasibility. These decisions collectively formed the foundation of our research approach and paved the way for the results and findings presented in this study.

III. Experimental Results

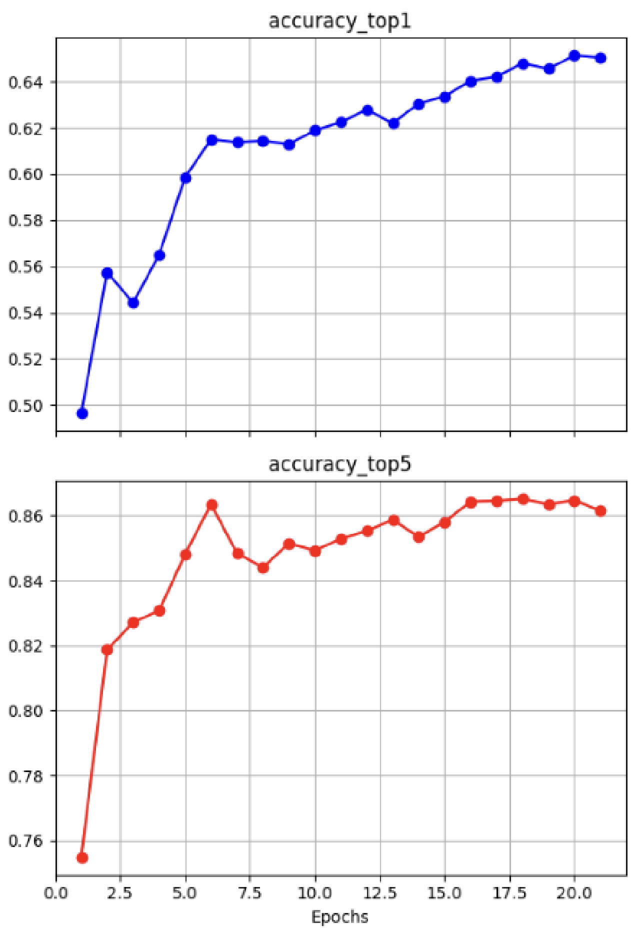


Fig. 6. Model Accuracy.

Figure 6 displays the accuracy of GeoBot when attempting to classify country images over 22 epocs. It is able to correctly classify images to country names over 60% of the time with both graphs showing the accuracy rate increasing over 22 epocs. While the accuracy_top1 rate increased over time reaching over 64% accuracy, the accuracy_top5 rate reaches a maximum of just over 84% between epocs 5 and 7.5 and then stabilizes at that value after 15 epocs.

Figure 7 displays the training and validation loss when training the data, both of which decrease dramatically to

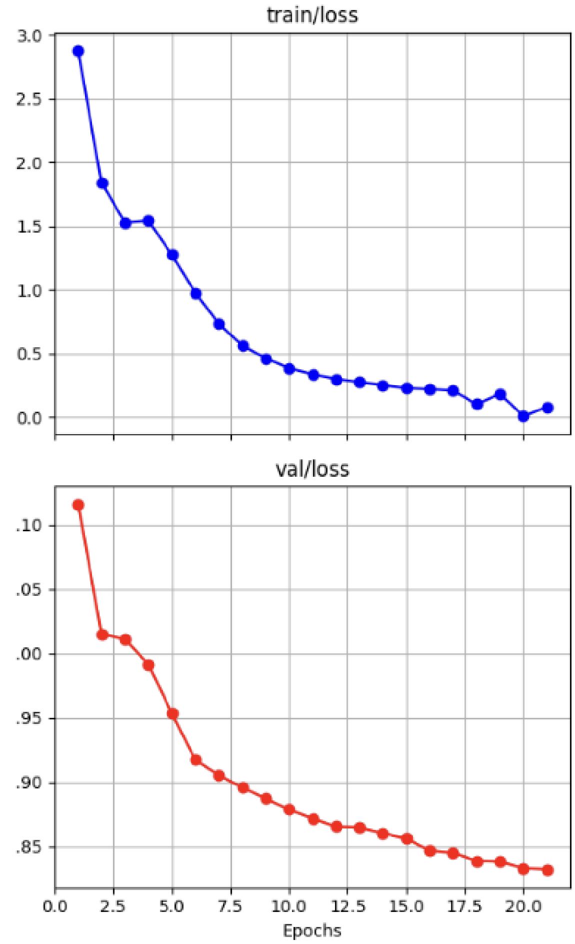


Fig. 7. Training and Validation Loss.

nearly 0 by epoc 22. At the end of the training duration, the model seemingly has very minimal errors and according to the validation loss graph, should be able to generalize any new data given to it well for accurate predictions.

IV. Discussion

Despite the training and validation loss graphs approaching 0, this may lead to overfitting which directly affects the accuracy of GeoBot. The training loss being so low may be indicative of the model memorizing the training data along with any noisy data and outliers within it. This would cause the model to directly compare any unseen data to the training data rather than looking for general patterns and relations thus, affecting its ability to accurately label unseen images. This would also explain why the accuracy_top1 is lower than the accuracy_top5 as it is not extracting enough patterns and relations of each country to distinguish them from each other. The single guess it is given may include data that is found in other countries which reduces the confidence of identifying them correctly.

While overfitting may be reduced through a larger and more diverse training set, this task would be difficult given the challenges already mentioned as well as being limited to

Google Street View as the primary source of obtaining data. Since we only had access to countries through Google Street View, it was difficult to eliminate any biases due to common features that all or large groups of countries share. Some of these features being

- Geography and Climate: Countries that are geographically grouped together may share similar climates, landscapes and vegetation
- Urban and Rural Areas: Urban areas have different features than rural areas so the model may associate other features in the environment specifically to one or the other leading to bias
- Demographics: Countries sharing similar population density, ethnicity, and/or age distributions will also introduce bias

If we were not limited to Google Street Views and could be more selective of the images chosen, we could make the training set larger and use images that highlight other features such as language, landmarks, and culture; things that are more specific to a particular country. These improvements would make it easier for the model to find patterns and relations associated with each country, allowing it to make more confident predictions thus, increasing the overall accuracy.

With a higher accuracy rate that has more emphasis on the patterns and relations found within a country to identify it, GeoBot would be able to do more than simply label countries. If it has the information to classify individual countries, that same information could be utilized to instead find and search for locations given a set of characteristics; a functionality that can be of use in real estate where agents can easily look for locations that satisfy a client's needs. Furthermore, it would be able to accurately describe the physical state of a location of a country; providing insight and possibly assessing the infrastructure and architectural characteristics of cities within that country. This information can then be used to improve transportation systems, optimize the usage of urban spaces, assign areas to be residential, commercial, or industrial, and much more.

While we accomplished the task of developing a bot that can correctly classify images to country names, improving the bot would expand its functionality; allowing it to be used as a tool to assist and/or improve fields other than image classification.

V. Conclusion

In this research endeavor, we embarked on a journey to harness the power of artificial intelligence and machine vision for the challenging task of country prediction from images. We introduced GeoBot, a sophisticated machine-learning model, as our vehicle to navigate this complex terrain. Through meticulous training on a diverse dataset comprising 50,000 images from Google Street View, GeoBot learned to decipher the intricate cues embedded in geographical landscapes, ranging from landmarks to road signs and natural elements. This research encompassed the critical phases of dataset curation, model training, and fine-tuning to optimize precision and computational efficiency.

Our methodology was anchored in the adoption of state-of-the-art tools, prominently featuring the YOLOv8 architecture for image classification. This choice ensured that GeoBot was well-equipped to handle the nuanced and multifaceted visual data encountered in our task. The transition from a smaller model, YOLOv8n-cls, to the more robust YOLOv8x-cls, with its significantly increased parameters, proved pivotal in enhancing the accuracy of country recognition. An extended training duration of 21 epochs further honed GeoBot's abilities, resulting in a top-1 accuracy of 65%.

The implications of our study extend beyond the confines of image classification and country prediction. GeoBot represents a significant milestone in the convergence of machine learning and computer vision, showcasing the immense potential of artificial intelligence in real-world applications. Our findings shed light on the challenges and possibilities within the realm of image classification, especially concerning diverse geographical landscapes.

Moreover, this research contributes to the broader discourse on machine vision, demonstrating its relevance in diverse fields, from augmented reality to robotics and self-driving vehicles. As we advance towards a future where machines can interpret and interact with visual information from the world, our work serves as a testament to the ever-evolving landscape of artificial intelligence and its capacity to transform our understanding of the world around us.

In conclusion, the journey undertaken in this study reaffirms the profound synergy between machine learning and machine vision. GeoBot's ability to accurately predict countries from images paves the way for innovative applications in tourism, content organization, and beyond. As we continue to explore the frontiers of AI, GeoBot stands as a testament to the remarkable strides we are making in utilizing machine vision to decipher the intricacies of our world.

References

[1] D. Wilson, X. Zhang, W. Sultani, and S. Wshah, "Image and object geo-localization," International Journal of Computer Vision, 2023. Accessed: Nov. 5, 2023. doi:10.1007/s11263-023-01942-3

[2] D. M. Forsyth and J. Ponce, "Classifying Images," in *Computer Vision A Modern Approach*, 2nd ed. Upper Saddle River, New Jersey: Pearson Education, Inc., 2012, ch. 16, pp.482

[3] A. R. Zamir, A. Hakeem, L. V. Gool, M. Shah, and R. Szeliski, *Large-Scale Visual Geo-Localization*, Cham, Switzerland: Springer International Publishing, 2015,

[4] F. Fontanella, F. Colace, M. Molinara, A. Scotto Di Freca, and F. Stanco, "Pattern recognition and artificial intelligence techniques for Cultural Heritage," Pattern Recognition Letters, vol. 138, pp. 23–29, 2020. doi:10.1016/j.patrec.2020.06.018

[5] S. Suresh, M. Abello, and N. Chodosh, "DeepGeo: Photo Localization with Deep Neural Network," arxiv. doi: https://doi.org/10.48550/arXiv.1810.03077, Oct. 2018.

[6] Pouke, M.: Using GPS data to control an agent in a realistic 3D environment, pp. 87–92. IEEE, September 2013. https://doi.org/10.1109/NGMAST.2013.24

[7] E. Müller-Budack, K. Pustu-Iren, and R. Ewerth, "Geolocation estimation of photos using a hierarchical model and scene classification," Computer Vision – ECCV 2018, vol. 11216, pp. 575–592, Oct. 2018. doi:10.1007/978-3-030-01258-8_35

[8] G. Jocher and A. Chaurasia, "Home," Ultralytics YOLOv8 Docs, https://docs.ultralytics.com/ (accessed Dec. 6, 2023).