https://doi.org/10.31891/2307-5732-2025-355-41
УДК 004.42:004.932

LAVINSKYI HLIB
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
e-mail: lavinskygleb@ukr.net
SKURAT OKSANA
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
https://orcid.org/0000-0001-7633-9121
e-mail: shkurat@pzks.fpm.kpi.ua

ALGORITHMIC AND SOFTWARE IMAGE RECOGNITION METHOD FOR TRACKING OBJECT GEOLOCATION

This article introduces a software method for image recognition designed to track the geolocations of objects captured in digital imagery. The proposed approach combines advanced computer vision techniques with geospatial data processing to enable automatic identification, classification, and spatial mapping of objects of interest within photographic content. At the core of the method lies a convolutional neural network architecture, which ensures high recognition accuracy by extracting and analyzing visual features from images. A distinctive feature of the system is the integration of an environmental context classification module, which segments the surrounding scene into semantic subclasses such as urban, natural, or industrial environments. This segmentation provides the model with additional spatial and contextual cues, significantly improving the reliability and precision of geolocation predictions, particularly in complex or ambiguous visual settings. The method is tailored for applications in environmental monitoring, reconnaissance, digital cartography, archival image analysis, and autonomous navigation systems. Experimental results demonstrate the effectiveness of the proposed approach compared to traditional algorithms, highlighting its potential to automate geolocation processes for diverse visual datasets. The implementation utilizes convolutional neural networks ResNet50 and EfficientNetB0, both pre-trained on large image datasets, which ensures high generalization capability. Experiments conducted on the Im2GPS3k dataset demonstrated a country-level classification accuracy of 57.3% and also 74.6% when the correct country was one of the five suggested by the network. The findings support the method's scalability and future development, including the integration of multimodal data sources and more sophisticated deep learning models for enhanced accuracy and generalization across varying geographic and visual conditions.

Keywords: Convolutional Neural Network, Image-Based Geolocation, ResNet50, EfficientNetB0, Image Recognition.

ЛАВІНСЬКИЙ ГЛІБ, ШКУРАТ ОКСАНА
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

АЛГОРИТМІЧНО-ПРОГРАМНИЙ МЕТОД РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ ДЛЯ ВІДСТЕЖЕННЯ ГЕОПОЗИЦІЙ ОБ'ЄКТІВ

У статті представлено метод глибинного навчання для розпізнавання зображень та визначення геолокації об'єктів на цифрових зображеннях. Розроблений метод ґрунтується на інтеграції моделей глибинного навчання архітектури EfficientNetB0 та ResNet50 для початкового розпізнавання типу сцени та подальшого визначення геолокації. Запропонований метод було протестовано на наборах даних ImageNet та Im2GPS3k. Проведені експерименти демонструють підвищення точності визначення геолокації. Для набору даних Im2GPS3k точність відстеження геолокації становила 57,3% та 74,6%, коли вірна країна була одна з п'яти запропонованих мережею. Запропонований метод є перспективним напрямом для розроблення фреймворку машинного навчання для відстеження геолокації на основі зображення.

Ключові слова: згорткова нейронна мережа, геолокація на основі зображень, ResNet50, EfficientNetB0, розпізнавання зображень.

Стаття надійшла до редакції / Received 15.05.2025
Прийнята до друку / Accepted 06.06.2025

Introduction

Accurate image-based geolocation remains a complex and open research challenge, particularly in dynamic and diverse environments such as urban areas, rural landscapes, and natural scenes. Existing image recognition systems often rely on generic deep learning models that do not take into account the semantic context of the scene. As a result, they suffer from degraded performance when encountering domain shifts or ambiguous visual cues.

A critical limitation in current geolocation approaches is the lack of adaptability to the type of environment depicted in the image. For instance, features that are informative in an urban context - such as buildings, road signs, or street layouts – may be irrelevant or entirely absent in natural landscapes, which instead require focus on terrain, vegetation, or water bodies. Applying the same model uniformly across all scene types leads to suboptimal performance and reduced accuracy.

The geolocation of an image, particularly in heterogeneous and visually complex environments, remains a fundamentally ill-posed problem. Traditional convolutional architectures struggle to generalize across scene types due to the inherent variability in spatial texture, layout regularity, and semantic content. In response to these challenges, proposed a algorithmic and software method for geolocation recognition that explicitly leverages scene-dependent

priors by introducing a conditional routing mechanism. The proposed method decomposes the overall inference task into two sequential stages: high-level semantic scene recognition, and specialized geolocation inference via scene-specific recognition branches.

Research objective

The objective of this research is to develop an efficient deep learning method for tracking image-based geolocation. Applying the EfficientNetB0 model, the proposed method initially classifies an image scene type. Based on the classified image, the models of the ResNet50 architecture recognize the geolocation of the image object.

Analysis of the latest research and publications

The task of image-based geolocation has attracted considerable attention in recent years, driven by advances in computer vision, availability of large-scale geotagged datasets, and the growing demand for visual localization in applications such as autonomous navigation, augmented reality, and digital forensics. Researchers have developed a wide range of algorithmic solutions, most of which can be broadly classified into two categories: global image-based classification approaches and retrieval-based or matching approaches. However, relatively few works have focused on the explicit use of hierarchical or adaptive models that account for the semantic nature of the scene.

One of the foundational works in global classification for image geolocation is PlaNet [1], which frames the problem as a classification task over a discretized map of the Earth. Using a convolutional neural network (CNN) trained on millions of geotagged photos, PlaNet predicts the geographic cell that most likely corresponds to the input image. This model demonstrated that geolocation could be tackled with end-to-end learning, bypassing the need for hand-crafted features or retrieval-based pipelines. However, PlaNet also exposed key limitations of global models most notably, a drop in performance when dealing with visually ambiguous scenes or when the model is exposed to underrepresented geographies in the training data. Moreover, since the architecture does not distinguish between different types of environments, it often relies on generic features, which may be insufficient in challenging or visually homogeneous regions.

To address the coarse granularity of models like PlaNet, CPlaNet [2] introduced the idea of combinatorial partitioning, where the Earth's surface is divided into overlapping regions. A coarse-to-fine strategy is used, combining predictions across multiple partitions to improve spatial resolution. This method improves accuracy but still relies on a unified CNN backbone that is applied uniformly across all image types, without explicit consideration for the semantic content of the scene. This again highlights a gap in adaptability, where the same model may be used for urban images rich in architectural features as well as for natural landscapes with minimal structural elements.

Several studies have started to recognize the importance of contextual information in improving geolocation accuracy. For example, GeoGuess [3] presents a pipeline that begins with high-level scene classification identifying whether an image depicts a beach, desert, forest, or urban area - and then uses this information to refine the geolocation estimate. Their results demonstrate a clear improvement over baseline methods, suggesting that the integration of semantic priors can serve as a powerful guide for downstream tasks. However, this system remains largely static; that is, once a scene class is detected, the model applies predetermined strategies without further adaptive tuning or model reconfiguration.

On a more dynamic front, the work in [4] explores modular neural networks that can adapt their internal processing based on environmental cues. These architectures consist of multiple specialized modules (filters, encoders, decoders) that are selectively activated depending on the detected domain. While this approach has shown promising results in domain adaptation for classification and segmentation tasks, its application to geolocation is not well studied. Furthermore, such dynamic systems often entail higher computational costs and more complex training procedures, which can be challenging in real-time or resource-constrained applications.

The importance of robust scene classification as a standalone task has also been emphasized in the literature. For example, the Places365 dataset and classification network [5] is specifically designed to recognize a wide variety of environments, such as forests, highways, markets, and stadiums. The dataset contains over 10 million labeled images across 365 categories, making it one of the most comprehensive resources for scene understanding. The pre-trained models derived from this dataset have been shown to generalize well to diverse domains and can be effectively used as the first step in a hierarchical geolocation pipeline [6].

Despite these advancements, few existing systems attempt to bridge the gap between scene semantics and adaptive geolocation strategies in a fully integrated manner. Most models remain either fixed-function or only loosely modular, without exploiting the full potential of hierarchical reasoning. There is currently no dominant framework that incorporates high-level scene understanding as a control mechanism for switching between specialized geolocation models, or for dynamically adjusting model parameters during inference.

Thus, the state-of-art technologies shows a growing recognition of the importance of environmental context in geolocation, but a lack of unified methodologies that bring together scene classification, adaptive model selection, and hierarchical decision-making into a single cohesive system. This represents an opportunity for novel research focused on software frameworks capable of leveraging scene semantics to intelligently guide the geolocation process for higher accuracy and robustness across diverse terrains.

Proposed method

The proposed method of image recognition for tracking image-based geolocation is based on two types of architectures of convolutional neural networks (CNN). The architecture of the proposed method is shown in Fig. 1.

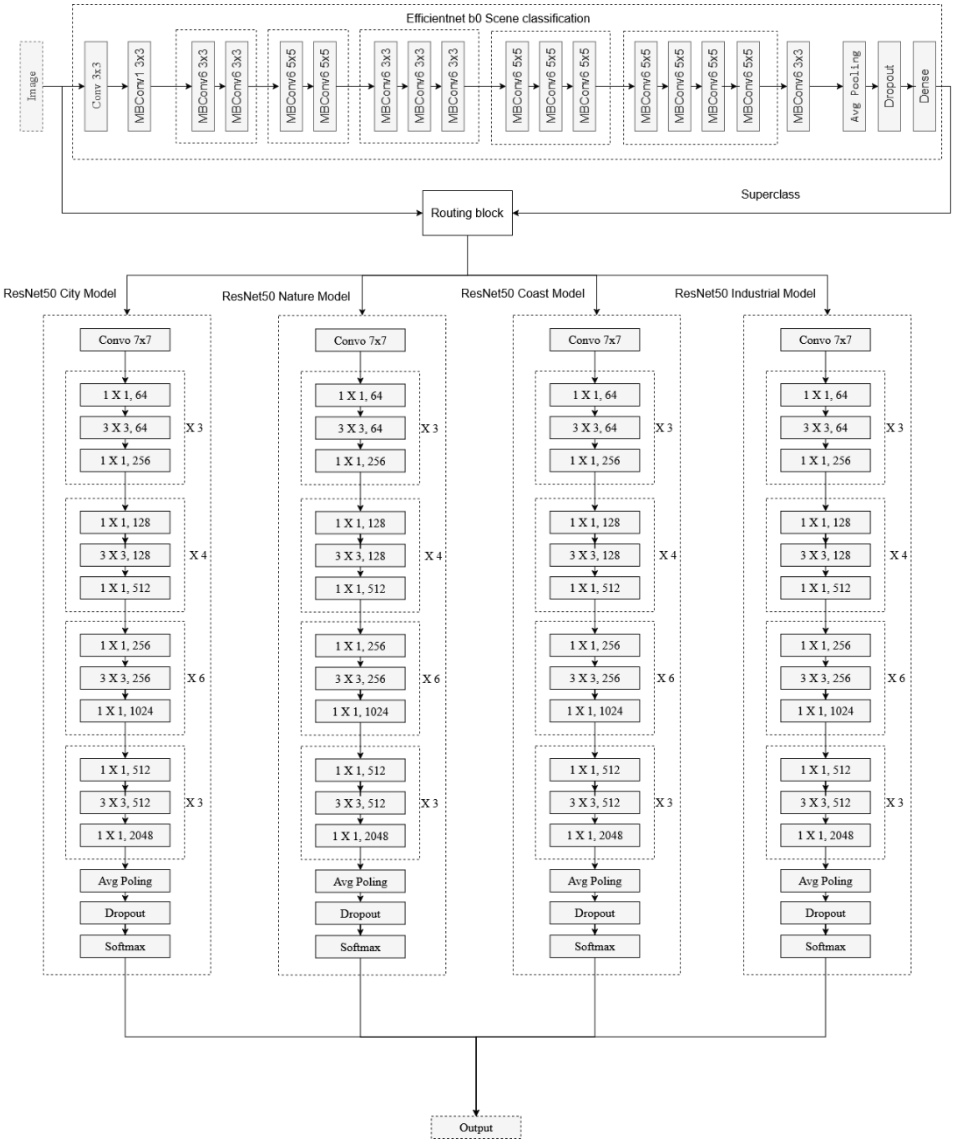

Fig.1. Architecture of the proposed method of image recognition for tracking image-based geolocation based on deep learning models of EfficientNet-B0 and ResNet50 architectures

The EfficientNet-B0 architecture neural network model is used for initial image recognition according to four categories (city, industry, coast, nature). The EfficientNet-B0 model is trained on the ImageNet [10] dataset. Training was performed using the Adam optimizer with an initial learning rate of 0.001. The objective function was categorical cross-entropy, and the model was trained for up to 50 epochs with a batch size of 32 (see Fig.2).

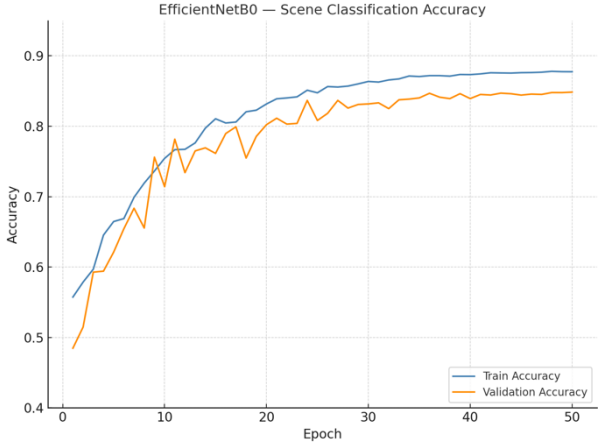

Fig.2. The results of training and validation of the proposed EfficientNet-B0 model

To facilitate stable convergence and effective transfer learning, a two-phase training strategy was adopted. During the first 10 epochs, all layers of the EfficientNet-B0 backbone were frozen, and training was limited to the newly added classification head. In the second phase, the top 30% of the backbone layers were unfrozen to enable

fine-tuning on the target dataset. Table 1 shows the result of applying the proposed EfficientNet-B0 model to image recognition for the ImageNet dataset.

Table 1

Evaluation of the proposed EfficientNet-B0 model for the ImageNet dataset.

Scene Type	Accuracy
City	85.1%
Nature	92.5%
Coast	91.6%
Industrial	84.5%

Based on the initial recognition results, four neural network models of the ResNet50 architecture are used to recognize countries of the world. The ResNet50 models are trained on the ImageNet dataset independently. The ResNet50 models employ convolution layers, the global average pooling layer, the dropout layer with the rate of 0.5, and the fully connected output layer with softmax activation function (1) for probability result of recognition.

$$softmax\ (x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (1)$$

Training was conducted using the Adam optimizer with a fixed learning rate of 0.001. The loss function used was categorical cross-entropy, and training proceeded for up to 40 epochs with a batch size of 32 (see Fig.3).

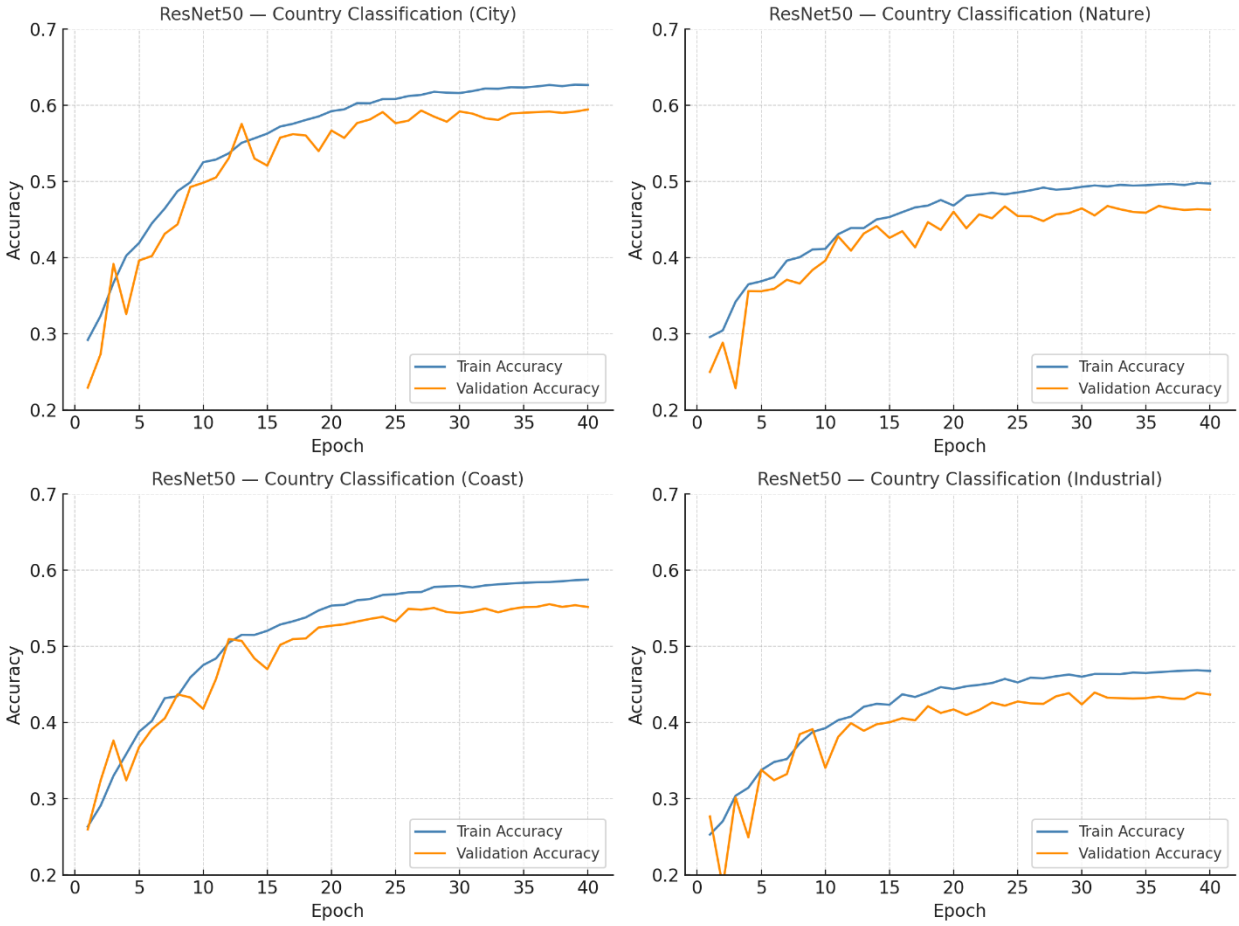

Fig.3. The results of training and validation of the proposed ResNet50 models

A two-phase training scheme was applied. In the initial 10 epochs, all layers of the ResNet-50 backbone were frozen to constrain learning to the added classification head. During the subsequent phase, the upper layers of the backbone were unfrozen to facilitate fine-tuning and enhance adaptability to the target dataset.

To improve generalization and robustness to variations in image content, standard data augmentation techniques were applied during training. These included random horizontal flipping, random rotation up to 15 degrees, random zooming up to 20%, and randomly selecting an area that covers at least 60% of the image, with an aspect ratio falling within the range $3/4 \leq \text{Ratio} \leq 16/9$. All images were normalized using the mean and standard deviation of the ImageNet dataset. Augmentations were applied dynamically during training using TensorFlow's data pipeline.

Table 2 shows the result of applying the proposed algorithmic and software image recognition method for tracking object geolocation for the Im2GPS3k dataset [11]. The Im2GPS3k dataset contains 3000 geotagged images

from a diverse range of countries and environments and serves as a standard tested for evaluating geolocation models on real-world scenes with varied visual complexity.

Table 2

Evaluation of the proposed algorithmic and software image recognition method for tracking object geolocation for the Im2GPS3k dataset.

Scene Type	Accuracy	Accuracy (the correct country was one of the five suggested)
City	63.5%	81.1%
Nature	50.3%	72.8%
Coast	59.6%	77.0%
Industrial	47.7%	68.3%

The results of image-based geolocation recognition by the proposed deep learning method based on the EfficientNet-B0 and ResNet50 models are shown in Fig.4.

Fig.4. The results of image-based geolocation recognition by the proposed deep learning method based on the EfficientNet-B0 and ResNet50 models for the Im2GPS3k dataset

The output of the model is a predicted country class, drawn from a set of countries that were present in the training data. This setup reflects a closed-set classification problem, where the model is not required to predict unseen countries, but rather to generalize across image styles and scene types. Each image is first processed by the scene classification module, which determines the environment type (city, nature, coast, industrial), and then routed to the corresponding geolocation head trained to predict the country. The proposed method achieved an overall country classification accuracy of 57.3% on the Im2GPS3k dataset. When considering the top 5 predicted countries, the model correctly identified the ground truth country in 74.6% of cases. A breakdown by scene type showed that the model performed best on urban scenes, reaching an accuracy of 63.8%. For natural landscapes, the accuracy was 52.1%, while coastal areas yielded 56.4%, and rural scenes were classified with an accuracy of 49.7%. For comparison, in existing deep learning methods based on the Inception v3 model, the recognition accuracy is significantly lower and is 28.4% and 48%, respectively.

Conclusion

The paper introduces the algorithmic and software image recognition method for tracking object geolocation. The proposed image-based geolocation recognition method employs a neural network model of the EfficientNet-B0 architecture for initial image recognition according to four categories (city, industry, coast, nature), and based on the results of the initial recognition, four neural network models of the ResNet50 architecture are employed to recognize the countries of the world.

The main point of the proposed image recognition method is module architecture with adaptive computation pathways, where the model dynamically selects the most appropriate geolocation subnetwork based on contextual understanding of the scene. This not only improves recognition accuracy by exploiting domain-specific spatial structures but also offers enhanced model interpretability, training efficiency, and flexibility for future extensibility.

Practical experiments have shown the significant increase in recognition accuracy of countries, different places of which ones are visualized on an image using the proposed algorithmic and software image recognition method for tracking object geopositions. The proposed image recognition method demonstrated a country recognition accuracy of 57.3% and also 74.6% when the correct country was one of the five suggested by the network. The highest recognition accuracy is achieved in urban scenes. Feature extraction from distinct architectural and infrastructural images is more accurate than natural and rural scenes, which are characterized by visual uniformity.

The developed module system based on the proposed image recognition method for tracking object geolocation is a scalability system that allows future improvement, applying region-based initial classification and integration of multimodal data.

References

1. Weyand, T., Kostrikov, I., & Philbin, J. (2016). PlaNet: Photo geolocation with convolutional neural networks. In *European Conference on Computer Vision (ECCV 2016)* (pp. 37–55). https://doi.org/10.1007/978-3-319-46487-9_40
2. Vo, N., & Jacobs, N. (2018). CPlaNet: Enhancing image geolocalization by combinatorial partitioning of maps. In *European Conference on Computer Vision (ECCV 2018)* (pp. 715–731). https://doi.org/10.1007/978-3-030-01234-2_43
3. Liu, X., Zou, D., & Tan, P. (2021). GeoGuess: A scene classification approach to image geolocalization. *arXiv preprint*. arXiv:2104.07702. https://arxiv.org/abs/2104.07702
4. Zamir, A. R., Sax, A., Shen, W., Guibas, L., Malik, J., & Savarese, S. (2018). Dynamic modular networks for visual domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8776–8785). https://doi.org/10.1109/CVPR.2018.00915
5. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009
6. Hofmann, K., Scherrer, Y., & Hollenstein, N. (2022). GeoAdapt: Leveraging intermediate tasks to improve geolocation prediction. *arXiv preprint*. arXiv:2203.08565. https://arxiv.org/abs/2203.08565
7. EfficientNet: Rethinking model scaling for convolutional neural networks [Electronic resource]. (2025). Retrieved July 17, 2025, from https://keras.io/api/applications/efficientnet/
8. Places [Electronic resource]. (2025). Retrieved July 17, 2025, from http://places2.csail.mit.edu/
9. ResNet50 TensorFlow [Electronic resource]. (2025). Retrieved July 17, 2025, from https://www.tensorflow.org/api_docs/python/tf/keras/applications/ResNet50
10. Im2GPS [Electronic resource]. (2025). Retrieved July 17, 2025, from https://paperswithcode.com/dataset/im2gps
11. ImageNet [Electronic resource]. (2025). Retrieved July 17, 2025, from https://www.image-net.org/

Література

1. Weyand, T., Kostrikov, I., & Philbin, J. (2016). PlaNet: Photo geolocation with convolutional neural networks. In *European Conference on Computer Vision (ECCV 2016)* (pp. 37–55). https://doi.org/10.1007/978-3-319-46487-9_40
2. Vo, N., & Jacobs, N. (2018). CPlaNet: Enhancing image geolocalization by combinatorial partitioning of maps. In *European Conference on Computer Vision (ECCV 2018)* (pp. 715–731). https://doi.org/10.1007/978-3-030-01234-2_43
3. Liu, X., Zou, D., & Tan, P. (2021). GeoGuess: A scene classification approach to image geolocalization. *arXiv preprint*. arXiv:2104.07702. https://arxiv.org/abs/2104.07702
4. Zamir, A. R., Sax, A., Shen, W., Guibas, L., Malik, J., & Savarese, S. (2018). Dynamic modular networks for visual domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8776–8785). https://doi.org/10.1109/CVPR.2018.00915
5. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009
6. Hofmann, K., Scherrer, Y., & Hollenstein, N. (2022). GeoAdapt: Leveraging intermediate tasks to improve geolocation prediction. *arXiv preprint*. arXiv:2203.08565. https://arxiv.org/abs/2203.08565
7. EfficientNet: Rethinking model scaling for convolutional neural networks [Electronic resource]. (2025). Retrieved July 17, 2025, from https://keras.io/api/applications/efficientnet/
8. Places [Electronic resource]. (2025). Retrieved July 17, 2025, from http://places2.csail.mit.edu/
9. ResNet50 TensorFlow [Electronic resource]. (2025). Retrieved July 17, 2025, from https://www.tensorflow.org/api_docs/python/tf/keras/applications/ResNet50
10. Im2GPS [Electronic resource]. (2025). Retrieved July 17, 2025, from https://paperswithcode.com/dataset/im2gps
11. ImageNet [Electronic resource]. (2025). Retrieved July 17, 2025, from https://www.image-net.org/