Periodica Polytechnica Electrical Engineering and Computer Science, 66(4), pp. 370–379, 2022

Improving Street View Image Classification Using Pre-trained CNN Model Extracted Features

Meriem Djouadi[1,2*], Mohamed-Khireddine Kholladi[1,3]

[1] Department of Computer Science, Faculty of Exact Sciences, Echahid Hamma Lakhdar University, P. O. B. 789, 39000 El Oued, Algeria
[2] Laboratory of Operator Theory and EDP: Foundations and Application, Echahid Hamma Lakhdar University, P. O. B. 789, 39000 El Oued, Algeria
[3] MISC Laboratory, Abdelhamid Mehri - Constantine 2 University, Ali Mendjeli Campus, P. O. B. 67, 25000 Constantine, Algeria
[*] Corresponding author, e-mail: djouadi-meriem@univ-eloued.dz

Received: 02 February 2022, Accepted: 29 August 2022, Published online: 26 September 2022

Abstract

This paper presents a new approach for the challenging problem of image geo-localization using Convolutional Neural Networks (CNNs). This latter has become the state-of-the-art technique in computer vision and machine learning, particularly in location recognition of images taken in urban environments where the recognition accuracy is very impressive. We cast this task as a classification problem. First, we extract features from images by using pre-trained CNN model AlexNet as a feature extraction tool; where the output of the fully connected layer is considered as the feature representation. Then, the features extracted from the fully connected layer can be used for the classification process by feeding them into the Support Vector Machine (SVM) classifier. We evaluated the proposed approach on a data set of Google Street View images (GSV); the experimental results show that our approach can improve the classification by achieving a good accuracy rate which is 94.19%.

Keywords

image geo-localization, location recognition, pre-trained CNN, Support Vector Machine (SVM), Google Street View images (GSV), classification

1 Introduction

Image geo-localization is a relevant learning and computer vision issue that aims to determine the location shown in a picture primarily based solely on its visual data. Despite the enormous recent advances in computer vision, geo-localization is still a difficult problem. This problem is very challenging under active research, over the past few years, particularly with the increase in the number of images and datasets that are publicly available on the Internet, which provide an opportunity for research in predicting the geographic location of images. Due to the possibility of various applications in landmark recognition [1–3], urban reconstruction [4], place recognition [5–7], visual navigation [8, 9], building recognition [10, 11], and robot vision [12, 13], attention from the research community has been drawn to the location estimation of images [14–21] over the past decade.

Generally, image geo-localization is simplified by limiting the problem to city-scale, for example, recognition of landmarks and cities [22–24] or natural areas such as mountains or deserts [25–27]. Just a few approaches treat the task at a global-scale [28–31].

Images and locations have a close relationship [32], therefore many of these applications require establishing a bidirectional relationship between the visual content and the location; a task that demands careful content analysis. The information about the location at which an image was taken is an important attribute of an image by itself. To handle this task, there are two main approaches based on retrieval and classification. The former retrieves just one relevant image by matching the query and database of geotagged images. The correspondences found must suffice to estimate the camera pose of the query image. The latter tackles the task as a classification problem by extracting features from the input images, and then feeding them into a classifier to predict the location of images. The efficiency of classification-based approaches depends mainly on the power of feature extraction methods.

Cite this article as: Djouadi, M., Kholladi, M.-K. "Improving Street View Image Classification Using Pre-trained CNN Model Extracted Features", Periodica Polytechnica Electrical Engineering and Computer Science, 66(4), pp. 370–379, 2022. https://doi.org/10.3311/PPee.19961

Hence, we aim to enhance the performance of classification from Google Street View images (GSV) which often include varied visual information to indicate their locations. To handle this problem effectively, it is required to learn powerful features.

Lately, the Convolutional Neural Network (CNN) has demonstrated its ability to extract powerful features in many fields, especially in computer vision where Deep Learning outperformed previous state-of-the-art approaches. This can be used in different applications like place recognition [33–35] and location estimation from images [36–38]. Hence, the outcomes of these features are very good and play an important role in improving the accuracy of classification. Therefore, we affirm that image geo-localization can greatly benefit from the power of features extracted by CNN since the performance could be significantly enhanced.

In this paper, we address the problem of improving location recognition from images taken in urban environments by considering it as a classification task. For that reason, we use the pre-trained CNN model AlexNet as a feature extraction tool, and then feed the features extracted from the fully connected layer into the Support Vector Machine (SVM) classifier. To the best of our knowledge, Street View image classification using extracted features from Convolutional Neural Networks Models in image geo-localization problem has not yet been studied. Experimental results demonstrated the efficiency of our proposed approach.

The remainder of the paper is organized as follows. Section 2 reviews related work on image geo-localization. Initial knowledge about Convolutional Neural Networks (CNNs) and the proposed approach are introduced in Section 3. Experimental results are presented and discussed in Section 4. Finally, Section 5 provides a conclusion of the paper.

2 Related work

Related work on image geo-localization can be approximately divided into two groups:

1. approaches based on image retrieval, and
2. approaches based on classification.

The most popular method of image geo-localization is based on image retrieval. Hays and Efros [28, 29] and Vo et al. [30] perform image retrieval using global image descriptors in a geo-tagged images database. It is possible to add different visual features to the image retrieval phase. NetVLAD [24] is directly trained for place recognition task in an end-to-end manner based on ranking loss using Google Street View data. NetVLAD, inspired by the "Vector of Locally Aggregated Descriptors" image representation is widely used to retrieve images. Although global features have the advantage of retrieving various natural scene images based on surroundings information, local image features provide greater accuracy in the retrieval of structured objects like buildings and are therefore used more frequently [1, 7, 20, 22, 39, 40]. Noh et al. [41] propose DELF (DEep Local Feature); an attentive local feature descriptor for image retrieval.

Otherwise, the image geo-localization based on classification formulates the issue as a classification task. Vo et al. in [30] and Weyand et al. in [31] to predict the geo-location of an input image, a classifier is trained. While the geo-location is depicted in a continuous space, approaches based on a classification quantify the map of the whole earth into a set of geo-classes equivalent to partitioned zones. Based on their GPS tags, the training images are labelled into the corresponding geo-classes. At testing time, the geo-class centre with the highest score is considered as the predicted location of an input image. Compared to methods based on image retrieval, this method is lightweight in terms of memory and time complexity. The accuracy of prediction depends on generated geo-class set. Because each image belonging to the same geo-class has a corresponding predicted geo-location, it is preferable to obtain accurate predictions through more fine-grained partitioning. Increasing the number of geo-classes, however, is not always easy as it increases the number of parameters linearly and makes the network susceptible to overfitting training data.

In [14, 20, 42–44], approaches based on pose estimation match query images against 3D models of a region, and use 2D-3D feature correspondences to determine 6-DOF query poses. Sattler et al. [43] first implement image retrieval rather than directly matching against a 3D model to obtain less accurate locations and then estimate poses using retrieved images. Based on a convolutional neural network, PoseNet [45, 46] handles pose estimation as an issue of regression, where the accuracy is enhanced by adding an intermediate LSTM layer to minimize dimensionality [47].

Another related area of research is landmark recognition [40, 48, 49]. The images are grouped by their geo-locations and visual resemblance to create a database of widespread landmarks. The database works as an index of the image retrieval system [50, 51] or as training data of the landmark classifier [52–54]. Cross-view geo-location recognition uses satellite or aerial imagery to determine the location of query [48, 49, 55, 56].

Deep learning is a newly developed approach aimed at showing excellent results and demonstrating that it is capable of deriving powerful features in different tasks of computer vision. It extracts high-level features computed from a whole image component where detailed features based on deep learning like Convolutional Neural Networks can be learned directly from images because it is more effective than traditional handcrafted features. CNNs are the main examples of the potential of neural networks to learn feature representations and patterns from images to perform complicated tasks like object recognition and image classification.

Recently, deep learning methods provide an increase in the performance of location recognition and thus open the direction to new computer vision applications which depend on rich representations of image content. Therefore, the researchers have successfully applied different CNN architectures to estimate the location of images [36, 37, 49, 55, 57] and recognize places and locations [2, 31, 34, 56, 58, 59]. Müller-Budack et al. [36] address the problem of geo-location estimation by using a multi-partitioning approach to incorporate hierarchical experience at different spatial resolutions, furthermore extracting and getting information about various types of environments for example: indoor, natural and urban. They consider image geo-localization as a classification task by dividing the globe into geographical cells with an equal number of images like the work of PlaNet [31]. The authors in [37] show the effectiveness of Delaunay triangles which are a kind of mesh for geo-location in comparatively low volume scenarios. The work contributes to image geo-location estimation by proposing a new global meshing approach, defining an assortment of training methods to avoid large data limitations in the training of these models, and illustrating how to use integrating additional information to improve the performance of a geo-location inference model. In [49], the authors use deep convolutional neural networks to address the problem of cross-view image geo-localization, which entails matching ground-level queries against aerial imagery to locate them. In addition, they propose a network architecture that combines features extracted from aerial images at different spatial scales. Lin et al. [55] provide an approach Where-CNN which inspired by the success of deep learning in face verification to improve the localization of ground-level query images by matching it to a reference database of aerial images. In this work, the image geo-localization is framed as an identity verification task. When it comes to image geo-location, image retrieval methods are commonly used. But in [31], the authors treat the problem as classification by dividing the earth into thousands of multi-scale geographic cells and then training a deep network by using millions of geo-tagged images. The model called PlaNet proved that it outperforms previous methods and reaches superhuman levels of precision in some cases. Tian et al. [56] handle the problem of cross-view image geo-localization to estimate the GPS position of a query street view image by finding its convenient images in a reference database of geo-tagged bird's eye view images, or vice versa. They introduce a novel structure for cross-view image geo-localization that benefits from deep convolutional neural networks' great success in image classification and object detection. The Faster R-CNN is first used to recognize buildings in the query and reference images. Then, using a Siamese network trained on both positive and negative matching image pairs to retrieve the k nearest neighbors from the reference buildings for each building in the query image. Based on dominant sets, the authors develop an efficient multiple nearest neighbors matching method to find the proper NN for each query building. In [58], the authors utilize the convolutional neural network model AlexNet to classify the Street View images. When they use max pooling as a sampling model, adjust the number of samples as 128 for one iteration and cropped at random the same picture out as other parts for the input model to train attains good accuracy. The paper [59] treats the task of defining the geographical location of a photo using a deep neural network based on the ResNet architecture and presents four different strategies of integrating low-level cardinality information. This model achieves an obvious accuracy.

3 Methodology

The overview of our proposed approach is shown in Fig. 1. In this part, we will first go over the classic architecture of Convolutional Neural Networks (CNNs) and the pre-trained CNN architecture, AlexNet, which is assessed in this paper. Results of the suggested approach for image geo-localization are then thoroughly evaluated and extracted.

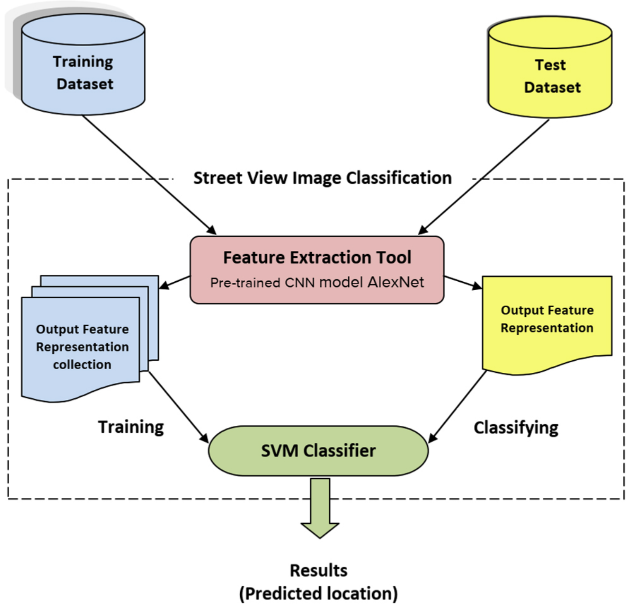

Fig. 1 Flowchart of our proposed approach for Street View image classification

3.1 Typical Convolutional Neural Networks architecture

CNNs are among the most effective learning algorithms for image content understanding and have demonstrated outstanding performance, especially in image classification. They are biologically inspired; wherein each CNN there are four main blocks:

- Convolutional Layer is a group of learnable filters, which help to take spatial features. The output from each filter is an activation map of two-dimensional; they are then piled to get the final result. Convolutional layers utilize the same set of weights across all filters, minimizing the overall amount of parameters needed.
- Non-Linearity (ReLU) is required for any type of neural network to be effective. Non-linearity is required for any type of neural network to be effective. Several CNN uses Sigmoid Non-Linearity but it has some disadvantages. Non-linearity is required for any type of neural network to be effective. ReLU is the abbreviation for Rectified Linear Unit, which is a more powerful non-linear operation. Almost all the latest architectures of CNN, such as AlexNet, used ReLU.
- Pooling Layers are put between convolutional layers. They perform a pooling operation to reduce the size of each feature map and the number of parameters by max or average pooling functions. Also the computations in the network to control the overfitting.
- Fully Connected Layers help to perform tasks of classification. They are following the convolution and pooling layers, which are added to complete the CNN architecture.

3.2 AlexNet architecture

AlexNet is the name of a convolutional neural network architecture created by Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton [60]. It is a very deep convolutional neural network where there are additional filters per layer, also is provided with stacked convolutional layers. As shown in Fig. 2 [61], it contains 5 convolutional layers, 3 max-pooling layers, and 3 fully connected layers. To help to minimize training time, AlexNet used ReLU and to avoid the problem of overfitting to the training data, AlexNet employs dropout layers.

The major outcome of the original study [60] was that the depth of the model was important for its high performance, which was computationally expensive but made possible because of using Graphics Processing Units (GPUs) during training.

On September 30, 2012, AlexNet participated in the ImageNet Large Scale Visual Recognition Challenge and became the pioneer of deep CNN that won the competition with 84.6% accuracy.

3.3 Street View image classification using pre-trained CNN model extracted features

In this paper, we present a method for implementing a supervised learning model for Street View image classification by using CNN extracted features. Our method is comprised of two phases:

1. extraction of features from input images by using CNN model, which is considered the main task of this study, and
2. classification of the input image based on features extracted by utilizing SVM Classifier for determining the location of Google Street View image.

3.3.1 Features extraction tool

We used pre-trained CNN architecture, called AlexNet, as a features extractor; where the output of the fully connected layer is considered as the feature representation. We eliminate the last fully connected layer and extract features from the second fully connected layer to get a feature vector of size 4096 as shown in Fig. 3 that can be utilized directly to predict the location of images.

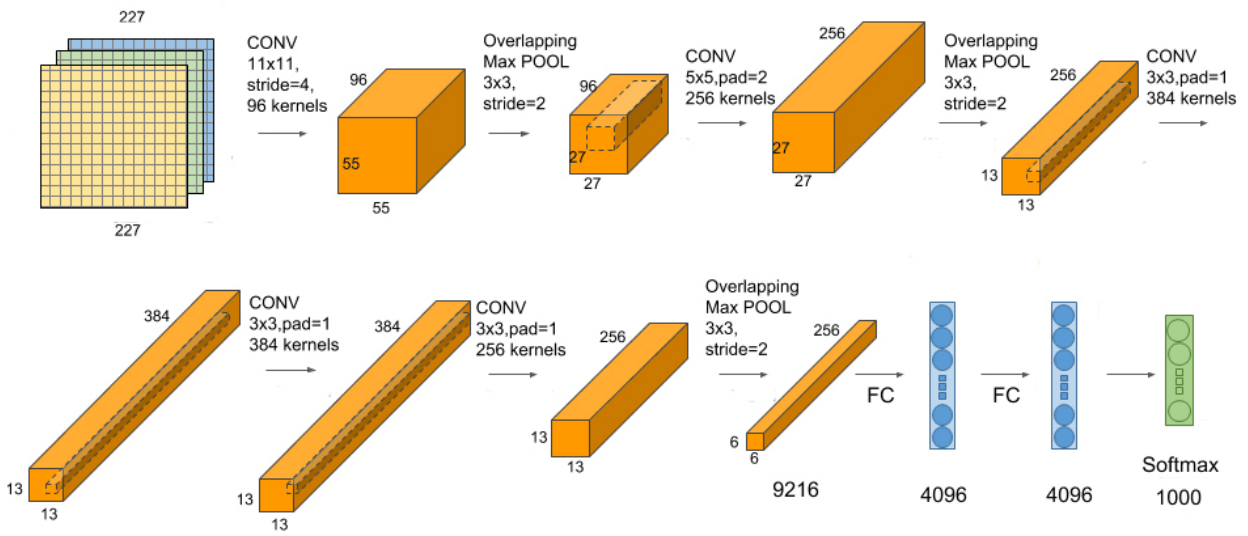

Fig. 2 Architecture of AlexNet (designed by M. Djouadi, adapted from [61])

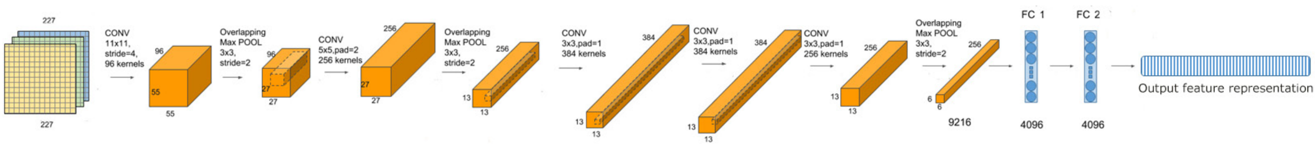

Fig. 3 Feature vector extracted from second Fully Connected layer (FC2), the dimension of this vector is 4096 (designed by M. Djouadi, adapted from [61])

3.3.2 Google Street View image classification

We used a supervised learning approach based on deep feature descriptors generated in the previous step to identify the location of input images by feeding them into a classifier. To accomplish this, we apply an efficient classifier known as Support Vector Machine (SVM) because of its capacity to cope with the hyper-dimensional input vector.

4 Experiments and results

In Section 4, we evaluate the performance of the proposed approach for image geo-localization from Google Street View (GSV) images taken in urban environments that usually include a variety of visual information to indicate their positions. The dataset, comprehensive experiments and experimental results, are described in detail.

4.1 Evaluation dataset

We evaluate the proposed algorithm using the Google Street View dataset [40] contains 62,058 high-quality images. The images depict the city centre as well as the surrounding areas of Pittsburgh, PA; Orlando, FL and Manhattan, NY. Each location mark's 360-degree view is divided into four side views and one top view image. There is one more image for each placemark that displays some overlay signs such as the address, street names, and so on. Fig. 4 illustrates an example of street view images from six placemarks of the dataset.

4.2 Experimental setup

All experiments were conducted on HP ProLiant DL380 G7 Server running under Windows Server 2012 R2 Essentials, with 6-cores Intel Xeon CPU E5649 at 2.53 GHz and 32 GB RAM.

AlexNet, the baseline CNN model which was used in this study is available in Matlab/Simulink. It provides a feature vector of 4096 dimensions from the second fully connected layer (FC2). The input images are re-sized to 227 × 227 pixels, and then the extracted feature vector was fed into a linear SVM classifier.

We randomly selected 90% of the dataset as training samples and 10% as testing samples in our experiments. We performed 60 different runs, and then it was averaged to obtain significant results. To evaluate the classification performance of each run; we use the accuracy (%) as a metric for evaluating classification models. It is calculated as the number of correct predictions divided by the total number of predictions. It can be also calculated in terms of positives and negatives as shown in Eq. (1):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

To get accurate inferences from the numerical results, we conducted a statistical study by using the software

Fig. 4 Sample of street view images belonging to six placemarks of the dataset

program GraphPad Prism [62] on our outcomes. To do that, we firstly used the Kolmogorov-Smirnov test with the Dallal-Wilkinson-Lilliefors *p-value* to determine if the results set is normally distributed (follow the Gaussian distribution). Then, we used the D'Agostino-Pearson test because it is a flexible and powerful normality test that is highly recommended. To apply this test, it is generally required that the sample size be greater than or equal to 20. Also, it is recommended when the sample contains repeated values. In all statistical tests, differences were considered statistically significant at $p < 0.05$.

Table 1 Results of Street View image classification on different datasets and different CNN models

Method	Dataset	Accuracy (%)
AlexNet [58]	Street View shop images	93.60
ResNet [59]	Google Street View dataset of the USA	71.87
NIN-Imagenet [57]	Google Street View images of Pittsburgh, PA; Orlando, FL and Manhattan, NY	86
Our method	Google Street View images of Pittsburgh, PA; Orlando, FL and Manhattan, NY	94.19

4.3 Results and discussion

As shown in Table 1, Google Street View image classification using pre-trained CNN AlexNet model as a features extractor by eliminating the last fully connected layer as well as extracting features from the second fully connected layer and Support Vector Machine (SVM) classifier yielded good results of 94.19% accuracy rate. Similarly to our model, Wang et al. [58] utilize deep learning algorithm to classify Street View images based on CNN AlexNet as the basis of Neural Network model [60] which reached an accuracy of 93.60%. Note that the dataset used is Street View shop images. The study performed by Peddada and Hong [57], the authors were used the NIN-Imagenet model, retrained the final convolution layer and added a fully connected layer. They applied their method on Google Street View images dataset [40] like that we used which found an accuracy of 86%, while in [59], Suresh et al. trained a deep neural network based on the Residual Network (ResNet) architecture, they used an open-source Google Street View images dataset of the United States (50States10K dataset) that created by them, which differs from what we used in our study but they have approximately the same characteristics because they both are Street View images. This model has a 71.87% accuracy rate.

Accuracy distribution analysis revealed that our numerical results obtained by the Kolmogorov-Smirnov test accepted the null hypotheses, and thus, the results were normally distributed as shown in Fig. 5 [63]. Results obtained by the D'Agostino-Pearson test confirmed that by Kolmogorov-Smirnov which indicated that there is a non-significant difference from the normal distribution (follow the Gaussian distribution).

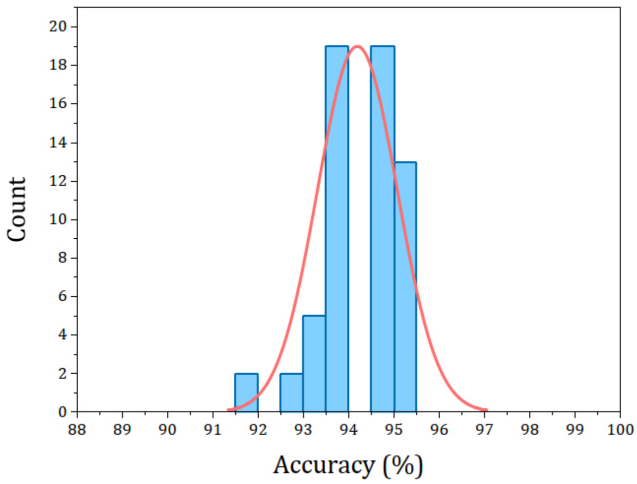

Fig. 5 Accuracy distribution analysis was made by M. Djouadi using Origin(Pro) [63]

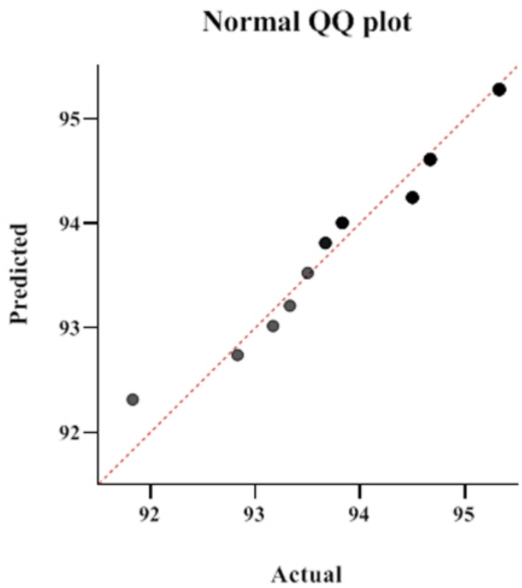

Fig. 6 QQ-normal plot (normality and lognormality tests were performed by M. Djouadi using GraphPad Prism [62])

Fig. 6 [62] illustrates the points in the QQ-normal plot lie on a straight diagonal line, thus the data is normally distributed. There are just minor deviations from the straight line. This indicates that the distribution is normal.

5 Conclusions

In this paper, we proposed a novel approach for the problem of image geo-localization using pre-trained CNN model extracted features and SVM classifier to improve the performance of Street View image classification. The aim of determining the location from images was attained by using a pre-trained CNN model as a feature extractor and SVM classifier. To assess the effectiveness of the proposed method, we used the Google Street View dataset which contains high-quality images from downtown and surroundings of Pittsburgh, PA; Orlando, FL and Manhattan, NY. We conducted statistical validation to prove the efficiency of our proposed approach. The experimental results show that the powerful features derived from the second fully connected layer of the pre-trained CNN model AlexNet can significantly ameliorate the classification accuracy.

For future works, we would like to utilize other pre-trained CNN models and integrate them to enhance the accuracy of Street View image classification. As well as, we want to evaluate the performance of our proposed method on different available datasets.

References

[1] Chen, D. M., Baatz, G., Köser, K., Tsai, S. S., Vedantham, R., Pylvänäinen, T., …, Grzeszczuk, R. "City-Scale Landmark Identification on Mobile Devices", In: CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 737–744. ISBN 978-1-4577-0394-2 https://doi.org/10.1109/CVPR.2011.5995610

[2] Crandall, D. J., Li, Y., Lee, S., Huttenlocher, D. P. "Recognizing Landmarks in Large-Scale Social Image Collections", In: Zamir, A. R., Hakeem, A., Van Gool, L., Shah, M., Szeliski, R. (eds.) Large-Scale Visual Geo-Localization, 2016, pp. 121–144. ISBN 978-3-319-25779-2 https://doi.org/10.1007/978-3-319-25781-5_7

[3] Chen, T., Wu, K., Yap, K.-H., Li, Z., Tsai, F. S. "A Survey on Mobile Landmark Recognition for Information Retrieval", In: 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, 2009, pp. 625–630. ISBN 978-1-4244-4153-2 https://doi.org/10.1109/MDM.2009.107

[4] Cheng, L., Yuan, Y., Xia, N., Chen, S., Chen, Y., Yang, K., Ma, L., Li, M. "Crowd-sourced pictures geo-localization method based on street view images and 3D reconstruction", ISPRS Journal of Photogrammetry and Remote Sensing, 141, pp. 72–85, 2018. https://doi.org/10.1016/j.isprsjprs.2018.04.006

[5] Mishkin, D., Perdoch, M., Matas, J. "Place Recognition with WxBS Retrieval", 2015. [online] Available at: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.710.8936 [Accessed: 30 April 2021]

[6] Gong, M., Sun, L., Yang, S., Yang, Y. "Find where you are: a new try in place recognition", The Visual Computer, 29(11), pp. 1211–1220, 2013. https://doi.org/10.1007/s00371-013-0784-6

[7] Knopp, J., Sivic, J., Pajdla, T. "Avoiding confusing features in place recognition", In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision – ECCV 2010, Springer, 2010, pp. 748–761. ISBN 978-3-642-15548-2 https://doi.org/10.1007/978-3-642-15549-9_54

[8] Feng, G., Ma, L., Tan, X., Xue, H., Guan, K. "Visual Location Recognition Based on Coarse-to-Fine Image Retrieval and Epipolar Geometry Constraint for Urban Environment", International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(11), pp. 25–36, 2016. https://doi.org/10.14257/ijsip.2016.9.11.03

[9] Liu, P., Yang, P., Wang, C., Huang, K., Tan, T. "A Semi-Supervised Method for Surveillance-Based Visual Location Recognition", IEEE Transactions on Cybernetics, 47(11), pp. 3719–3732, 2017. https://doi.org/10.1109/TCYB.2016.2578639

[10] Lee, K., Lee, S., Jung, W. J., Kim, K. T. "Fast and Accurate Visual Place Recognition Using Street-View Images", 39(1), pp. 97–107, 2017. https://doi.org/10.4218/etrij.17.0116.0034

[11] Li, J., Huang, W., Shao, L., Allinson, N. "Building recognition in urban environments: A survey of state-of-the-art and future challenges", Information Sciences, 277, pp. 406–420, 2014. https://doi.org/10.1016/j.ins.2014.02.112

[12] Karasfi, B., Hong, T. S., Jalalian, A., Nakhaeinia, D. "Speedup Robust Features Based Unsupervised Place Recognition for Assistive Mobile Robot", In: 2011 International Conference on Pattern Analysis and Intelligence Robotics, Kuala Lumpur, Malaysia, 2011, pp. 97–102. ISBN 978-1-61284-407-7 https://doi.org/10.1109/ICPAIR.2011.5976919

[13] Yang, S., Mou, W., Wang, H. "Place Recognition using Multiple Feature Types", Advances in Robotics & Automation, S2(1), 008, 2015. https://doi.org/10.4172/2168-9695.S2-008

[14] Li, Y., Snavely, N., Huttenlocher, D. P., Fua, P. "Worldwide Pose Estimation Using 3D Point Clouds", In: Zamir, A., Hakeem, A., Van Gool, L., Shah, M., Szeliski, R. (eds.) Large-Scale Visual Geo-Localization, Springer, 2016, pp. 147–163. ISBN 978-3-319-25779-2 https://doi.org/10.1007/978-3-319-25781-5_8

[15] Li, X., Larson, M., Hanjalic, A. "Geo-Distinctive Visual Element Matching for Location Estimation of Images", IEEE Transactions on Multimedia, 20(5), pp. 1179–1194, 2018. https://doi.org/10.1109/TMM.2017.2763323

[16] Huang, J., Lo, S.-L. "Location Estimation of Urban Images Based on Geographical Neighborhoods", Journal of Physics: Conference Series, 1004, 012019, 2018. https://doi.org/10.1088/1742-6596/1004/1/012019

[17] Chen, C.-Y., Grauman, K. "Clues from the beaten path: Location estimation with bursty sequences of tourist photos", In: CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 1569–1576. ISBN 978-1-4577-0394-2 https://doi.org/10.1109/CVPR.2011.5995412

[18] Sattler, T., Havlena, M., Schindler, K., Pollefeys, M. "Large-scale location recognition and the geometric burstiness problem", In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 1582–1590. ISBN 978-1-4673-8852-8 https://doi.org/10.1109/CVPR.2016.175

[19] Mousavian, A., Košecká, J., Lien, J.-M. "Semantically guided location recognition for outdoors scenes", In: 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 2015, pp. 4882–4889. ISBN 978-1-4799-6923-4 https://doi.org/10.1109/ICRA.2015.7139877

[20] Cao, S., Snavely, N. "Graph-Based Discriminative Learning for Location Recognition", International Journal of Computer Vision, 112(2), pp. 239–254, 2015. https://doi.org/10.1007/s11263-014-0774-9

[21] Qian, X., Wang, H., Zhao, Y., Hou, X., Hong, R., Wang, M., Tang, Y. Y. "Image location inference by multisaliency enhancement", IEEE Transactions on Multimedia, 19(4), pp. 813–821, 2017. https://doi.org/10.1109/TMM.2016.2638207

[22] Schindler, G., Brown, M., Szeliski, R. "City-scale location recognition", In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 2007, pp. 1–7. ISBN 1-4244-1179-3 https://doi.org/10.1109/CVPR.2007.383150

[23] Zheng, Y.-T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., Brucher, F., Chua, T.-S., Neven, H. "Tour the world: Building a web-scale landmark recognition engine", In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 1085–1092. ISBN 978-1-4244-3992-8 https://doi.org/10.1109/CVPR.2009.5206749

[24] Arandjelović, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J. "NetVLAD: CNN architecture for weakly supervised place recognition", In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 5297–5307. ISBN 978-1-4673-8852-8 https://doi.org/10.1109/CVPR.2016.572

[25] Baatz, G., Saurer, O., Köser, K., Pollefeys, M. "Large scale visual geo-localization of images in mountainous terrain", In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) Computer Vision – ECCV 2012, Springer, 2012, pp. 517–530. ISBN 978-3-642-33708-6 https://doi.org/10.1007/978-3-642-33709-3_37

[26] Saurer, O., Baatz, G., Köser, K., Ladický, L., Pollefeys, M. "Image Based Geo-localization in the Alps", International Journal of Computer Vision, 116(3), pp. 213–225, 2016. https://doi.org/10.1007/s11263-015-0830-0

[27] Tzeng, E., Zhai, A., Clements, M., Townshend, R., Zakhor, A., "User-driven geolocation of untagged desert imagery using digital elevation models", In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 2013, pp. 237–244. ISBN 978-0-7695-4990-3 https://doi.org/10.1109/CVPRW.2013.42

[28] Hays, J., Efros, A. A. "IM2GPS: Estimating geographic information from a single image", In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008, pp. 1–8. ISBN 978-1-4244-2242-5 https://doi.org/10.1109/CVPR.2008.4587784

[29] Hays, J., Efros, A. A. "Large-Scale Image Geolocalization", In: Choi, J., Friedland, G. (eds.) Multimodal Location Estimation of Videos and Images. Springer, 2015, pp. 41–62. ISBN 978-3-319-09860-9 https://doi.org/10.1007/978-3-319-09861-6_3

[30] Vo, N., Jacobs, N., Hays, J. "Revisiting IM2GPS in the Deep Learning Era", In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2640–2649. ISBN 978-1-5386-1033-6 https://doi.org/10.1109/ICCV.2017.286

[31] Weyand, T., Kostrikov, I., Philbin, J. "PlaNet - photo geolocation with convolutional neural networks", In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016, Springer, 2016, pp. 37–55. ISBN 978-3-319-46483-1
https://doi.org/10.1007/978-3-319-46484-8_3

[32] Zamir, A. R., Hakeem, A., Van Gool, L., Shah, M., Szeliski, R. "Introduction to Large-Scale Visual Geo-localization", In: Large-Scale Visual Geo-Localization, Springer, 2016, pp. 1–18. ISBN 978-3-319-25779-2
https://doi.org/10.1007/978-3-319-25781-5_1

[33] Chen, Z., Lam, O., Jacobson, A., Milford, M. "Convolutional Neural Network-based Place Recognition", [preprint] arXiv e-prints, arXiv:1411.1509 [cs.CV], 06 November 2014.
https://doi.org/10.48550/arXiv.1411.1509

[34] Chen, Z., Jacobson, A., Sünderhauf, N., Upcroft, B., Liu, L., Shen, C., Reid, I., Milford, M. "Deep Learning Features at Scale for Visual Place Recognition", In: 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, Singapore, 2017, pp. 3223–3230. ISBN 978-1-5090-4634-8
https://doi.org/10.1109/ICRA.2017.7989366

[35] Sünderhauf, N., Shirazi, S., Jacobson, A., Dayoub, F., Pepperell, E., Upcroft, B., Milford, M. "Place Recognition with ConvNet Landmarks: Viewpoint-Robust, Condition-Robust, Training-Free", [pdf] In: Kavraki, L. E., Hsu, D., Buchli, J. (eds.) Proceedings of Robotics: Science and Systems, Rome, Italy, 2015, 22. ISBN 978-0-9923747-1-6
https://doi.org/10.15607/RSS.2015.XI.022

[36] Müller-Budack, E., Pustu-Iren, K., Ewerth, R. "Geolocation Estimation of Photos Using a Hierarchical Model and Scene Classification", In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018, Springer, 2018, pp. 575–592. ISBN 978-3-030-01257-1
https://doi.org/10.1007/978-3-030-01258-8_35

[37] Johns, J. M., Rounds, J., Henry, M. J. "Multi-modal Geolocation Estimation Using Deep Neural Networks", [preprint] arXiv e-prints, arXiv:1712.09458 [cs.CV], 26 December 2017.
https://doi.org/10.48550/arXiv.1712.09458

[38] Iscen, A., Tolias, G., Avrithis, Y., Furon, T., Chum, O. "Panorama to panorama matching for location recognition", In: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, Bucharest, Romania, 2017, pp. 392–396. ISBN 9781450347013
https://doi.org/10.1145/3078971.3079033

[39] Kim, H. J., Dunn, E., Frahm, J.-M. "Predicting good features for image geo-localization using per-bundle VLAD", In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1170–1178. ISBN 978-1-4673-8391-2
https://doi.org/10.1109/ICCV.2015.139

[40] Zamir, A. R., Shah, M. "Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs", IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(8), pp. 1546–1558, 2014.
https://doi.org/10.1109/TPAMI.2014.2299799

[41] Noh, H., Araujo, A., Sim, J., Weyand, T., Han, B. "Large-Scale Image Retrieval with Attentive Deep Local Features", In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 3476–3485. ISBN 978-1-5386-1033-6
https://doi.org/10.1109/ICCV.2017.374

[42] Irschara, A., Zach, C., Frahm, J.-M., Bischof, H. "From structure-from-motion point clouds to fast location recognition", In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 2599–2606, 2009. ISBN 978-1-4244-3992-8
https://doi.org/10.1109/CVPRW.2009.5206587

[43] Sattler, T., Weyand, T., Leibe, B., Kobbelt, L. "Image Retrieval for Image-Based Localization Revisited", In: Bowden, R., Collomosse, J., Mikolajczyk, K. (eds.) Proceedings British Machine Vision Conference, Surrey, UK, 2012, pp. 76.1–76.12. ISBN 1-901725-46-4
https://doi.org/10.5244/C.26.76

[44] Sattler, T., Leibe, B., Kobbelt, L. "Fast Image-Based Localization using Direct 2D-to-3D Matching", In: 2011 International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 667–674. ISBN 978-1-4577-1101-5
https://doi.org/10.1109/ICCV.2011.6126302

[45] Kendall, A., Cipolla, R. "Geometric loss functions for camera pose regression with deep learning", In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6555–6564. ISBN 978-1-5386-0458-8
https://doi.org/10.1109/CVPR.2017.694

[46] Kendall, A., Grimes, M., Cipolla, R. "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization", Santiago, Chile, 2016, pp. 2938-2946. ISBN 978-1-4673-8391-2
https://doi.org/10.1109/ICCV.2015.336

[47] Walch, F., Hazirbas, C., Leal-Taixé, L., Sattler, T., Hilsenbeck, S., Cremers, D. "Image-based localization using LSTMs for structured feature correlation", In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 627–637. ISBN 978-1-5386-1033-6
https://doi.org/10.1109/ICCV.2017.75

[48] Lin, T.-Y., Belongie, S., Hays, J. "Cross-view image geolocalization", In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 891–898. ISBN 978-1-5386-5672-3
https://doi.org/10.1109/CVPR.2013.120

[49] Workman, S., Souvenir, R., Jacobs, N. "Wide-area image geolocalization with aerial reference imagery", In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 3961–3969. ISBN 978-1-4673-8391-2
https://doi.org/10.1109/ICCV.2015.451

[50] Avrithis, Y., Kalantidis, Y., Tolias, G., Spyrou, E. "Retrieving Landmark and Non-Landmark Images from Community Photo Collections", In: Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 2010, pp. 153–162. ISBN 9781605589336
https://doi.org/10.1145/1873951.1873973

[51] Weyand, T., Leibe, B. "Visual landmark recognition from Internet photo collections: A large-scale evaluation", Computer Vision and Image Understanding, 135, pp. 1–15, 2015. https://doi.org/10.1016/j.cviu.2015.02.002

[52] Bergamo, A., Sinha, S. N., Torresani, L. "Leveraging Structure from Motion to Learn Discriminative Codebooks for Scalable Landmark Classification", In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 763–770. ISBN 978-1-5386-5672-3 https://doi.org/10.1109/CVPR.2013.104

[53] Li, Y., Crandall, D. J., Huttenlocher, D. P. "Landmark Classification in Large-scale Image Collections", In: 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 1957–1964, 2009. ISBN 978-1-4244-4420-5 https://doi.org/10.1109/ICCV.2009.5459432

[54] Gronát, P., Obozinski, G., Sivic, J., Pajdla, T. "Learning and calibrating per-location classifiers for visual place recognition", In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 907–914. ISBN 978-1-5386-5672-3 https://doi.org/10.1109/CVPR.2013.122

[55] Lin, T.-Y., Cui, Y., Belongie, S., Hays, J. "Learning Deep Representations for Ground-to-Aerial Geolocalization", In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 5007–5015. ISBN 978-1-4673-6964-0 https://doi.org/10.1109/CVPR.2015.7299135

[56] Tian, Y., Chen, C., Shah, M. "Cross-View Image Matching for Geo-localization in Urban Environments", In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1998–2006. ISBN 978-1-5386-0458-8 https://doi.org/10.1109/CVPR.2017.216

[57] Peddada, A. V, Hong, J. "Geo-Location Estimation with Convolutional Neural Networks", [pdf] CS231n, Stanford University, Stanford, CA, USA, 2015. Available at: http://cs231n.stanford.edu/reports/2015/pdfs/CS231N_Final_Report_amanivp_jamesh93.pdf [Accessed: 28 January 2021]

[58] Wang, Q., Zhou, C., Xu, N. "Street view image classification based on convolutional neural network", In: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 2017, pp. 1439–1443. ISBN 978-1-4673-8980-8 https://doi.org/10.1109/IAEAC.2017.8054251

[59] Suresh, S., Chodosh, N., Abello, M. "DeepGeo: Photo localization with deep neural network", [preprint] arXiv e-prints, arXiv:1810.03077 [cs.CV], 07 October 2018. https://doi.org/10.48550/arXiv.1810.03077

[60] Krizhevsky, A., Sutskever, I., Hinton, G. E. "ImageNet Classification with Deep Convolutional Neural Networks", In: Pereira, F., Burges, C. J., Bottou, L., Weinberg, K. Q. (eds.) Advances in Neural Information Processing Systems 25 (NIPS 2012), Curran Associates, Inc., 2012, pp. 1097–1105. ISBN 9781627480031

[61] Prijono, B. "Student Notes: Convolutional Neural Networks (CNN) Introduction", IndoML.com, 07 March 2018, [online] Available at: https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/ [Accessed: 04 July 2021]

[62] GraphPad Software "GraphPad Prism, version 9.2.0 for Windows", [computer program] Available at: https://www.graphpad.com/updates/prism-920-release-notes [Accessed: 23 September 2021]

[63] OriginLab Corporation "OriginPro 2019b (64 bit), version 9.6.5.169", [computer program] Available at: https://www.originlab.com/index.aspx?go=Products/Origin/2019b&pid=3322 [Accessed: 24 September 2021]