Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-2, February 2026
DOI: 10.53469/jrse.2026.08(02).12

A Robust Visual Place Recognition Model Based on Cuckoo Search Algorithm and Deep Learning

Kumar Thrisha[1], Madhura Vaishali[2]

[1]Research Scholar, Department of Computer and Information Sciences, Annamalai University
thrisha@yahoo.co.in

[2]Assistant Professor / Programmer, Department of Computer and Information Sciences, Annamalai University
vaishali@gmail.com

Abstract: ***Visual place recognition (VPR) utilizing deep learning (DL)is developed in the domain of computer vision (CV) by connecting the power of neural networks (NNs) for automatically recognizing and finding particular land marks or positions within visual scenes. Leveraging DL model, namely convolutional neural networks (CNNs) and recurrent neural networks (RNNs), this technology has achieved important developments in tasks like object localization autonomous navigation, and scene understanding. Consequently, this study presents a Cuckoo Search Algorithm with Deep Learning Driven Robust Visual Place Recognition (CSADL-VPR) method. This introduces CSADL-VPR approach for VPR that leverages Gaussian filtering (GF) based pre-processing, Cuckoo Search optimization algorithm (CSA), MobileNetV2 feature extractor, and Manhattan distance-based similarity measurement. The MobileNetV2 feature extractor could be exploited for extracting compact and distinctive features from input images. For enhancing the performance of the place recognition model, the CSA has been implemented for parameter tuning and feature selection (FS). The Manhattan distance metric is employed for evaluating the similarity among feature vectors extracted from query images and those in the reference datasets. A series of simulated analyses can be executed to ensure the enhanced performance of the CSADL-VPR approach. The experimental outcomes represented the improvements of the CSADL-VPR method on the place recognition task.***

Keywords: Visual place recognition; Deep learning; Cuckoo Search optimization algorithm; MobileNetv2; Similarity measurement

1. Introduction

The main objective of Visual Place Recognition (VPR) has to support a vision-based navigation network and resolves if it recognizes prime websites that are visited [1]. It is said to be very complex problem in computer vision (CV) as well as automation area. In recent days, these areas have rushed in VPR utilization for a lot of applications. However, VPR has increased substantial attention and has been studied widely in automation and CV groups [2], but still numerous open difficulties to concerned with. The difficulty process with VPR is twofold. First, various dominant VPR methods are appearance-related [3]. But, the appearance of related places can extremely differ with several enlightenment circumstances, distance, standpoints, seasons, false place recognition, and background clutter first and occlusion permits interfering to localize approaches, which decrease exactness and then consequences in catastrophic localization collapse for navigation systems [4]. Hence, it could be hard to know an equivalent location appropriately after it goes via appearance modifications. Primary visual descriptors are classified into two types namely: global and local descriptors [5].

Global descriptors produce an individual compact feature vector and then describe the complete image [6], therefore it has advantages in enlightenment and then arrival invariance however it is ineffectual in managing standpoint changes. Whereas the Local descriptors are used to remove the feature nearby interest points that describe images and tell robustness alongside viewpoint differences but endure presence changes [7]. A CoHOG remarkable handcrafted feature related employs on the entropy map for removing RoI and utilized the HOG (Histogram of Oriented Gradients) descriptor to make use of the cooperative local illustrations [8]. Presently, the concentration on region built VPR could be transformed to learning-based methods, mainly Convolutional Neural Networks (CNNs), since its ultimate successes in identification and image recovery [9]. The great way of joining CNN techniques as well as local region descriptors is well-known via the initial results in the VPR. However, CNN-related descriptors are in necessity of additional computing networks namely hardware based acceleration by using a GPU that is inappropriate for resource-deficient devices [10].

This study introduces a Cuckoo Search Algorithm with Deep Learning Driven Robust Visual Place Recognition (CSADL-VPR) method. This presents CSADL-VPR approach a novel approach to visual place recognition that leverages the Gaussian filtering (GF) based pre-processing, MobileNetV2 feature extractor, the Cuckoo Search optimization algorithm (CSA), and Manhattan distance-based similarity measurement. The MobileNetV2 feature extractor is employed to extract discriminative and compact features from input images. To optimize the performance of the place recognition model, the CSA is employed for feature selection and parameter tuning. A series of experimental analyses can be performed to ensure the improved performance of CSADL-VPR system.

2. Related Works

Zaffar et al. [11] developed training free, compute efficient techniques related to HOG descriptor to attain obtainable performance in VPR. The stimulation for this CoHOG relies on convolution scanning and then regions depends on the feature extraction are used by CNN. This approach attained more effectiveness in VPR in changing surroundings. The authors [12] developed a learning-based result, so a CNN to

create image level illustrations that are invariant to conditions like lighting and weather. The authors developed a technique that ensures visual localization via an image level illustration which is planned from sequence of images. The authors proposed Gaussian Process Particle Filter organization that introduces two main improvements which allow localization by using databases that cover huge regions with strengthening the behaviour each time handling with improper initialization lack of the filter. Ultimately, the author for convolution neural organizations presents two novel common objective modules. Initially, the authors proposed CNN-COSFIRE method for the image identifying work. CNN-COSFIRE spreads as well as familiarizes the COSFIRE organization for its presence in CNN design.

Schubert et al. [13] made onto our existing study on graph enhancement for identifying places, where graph has been used for demonstrating further organizational information. To improve the performance of place classification, the following nonlinear least squares optimization (NLSQ) is utilized. This study spoke about the long run-time and the prominent memory utilized to boost finer place identification performance quicker on complex matters. The author projected a novel graph maximization technique that relies on Iterated Conditional Modes (ICM). The authors checked novel cost function for an edge in the graph. Yang et al. [14] projected a novel and effective technique namely Multi scale Sliding Window (MSW) to boost up the place recognition result for landmark generation (LG). Desperate classical method of LG usually depends on classifying objects that size allocations are different so it could not be highly effective in understanding viewpoint and shift invariance.

Xie et al. [15] designed a fusion network that takes the point cloud descriptors and then images to remove the place recognition problem. This method could be reduced equally as farming a compacted fusion system that caught both robust representation of the image and then 3D point cloud by employing the removed process in point cloud global feature combination for boosting classification result, a suitable metric to illuminate the similarity of this fused worldwide feature. Islam et al. [16] designed an Independent Component Analysis (ICA) and AE complexity for recognizing the route over the machine.

3. The Proposed Model

In this study, we have presented a new CSADL-VPR system for automatically and accurately recognizing visual places. The aim of the CSADL-VPR technique is the proper recognition of the visual places using DL method. In the CSADL-VPR approach, a 4 sub-stages are comprised namely Minkowski Distance based visual recognition, GF based noise elimination, CSA based hyper parameter tuning, andMobileNet-v2 feature extractor.

3.1 GF based Pre-processing

The GF is used to remove the noise in the input images.GF has an extensively employed preprocessing method in image and signal processing that performs an important function in feature enhancement and noise reduction [17]. It can depend on the principles of convolution, where a Gaussian kernel, normally a 2D bell-shaped curve, was implemented to the input signal or image. This kernel is represented by a 2 parameters namely mean (μ) and standard deviation (σ). The Gaussian kernel's shape efficiently blurs the image or smooths the signal by decreasing higher-frequency noise when maintaining main structural data. The convolution process measures a weighted average of the pixel values along with a local neighbourhood, where the weights could be evaluated by the Gaussian kernel. The outcome is that pixels nearer to one another in the input image provide most importantly to the output, whereas pixels have a more reduced effect. This feature makes GF an efficient tool for minimizing both non-Gaussian and Gaussian noise, like salt-and-pepper noise. Moreover, GF is used frequently to attain various scales of smoothing, permitting users to fine-tune the trade-off between feature preservation and noise reduction. In General, GF functions as a multipurpose and basic preprocessing stage in diverse applications comprising feature extraction, edge detection, and image denoising.

3.2 Feature Extraction using MobileNetv2 Model

Here, the MobileNetv2 model is implemented for extracting feature vectors. Feature extraction utilizing the MobileNetV2 architecture is a main constituent of CV tasks, especially in conditions where computational efficiency is significant like embedded and mobile devices [18]. MobileNetV2 is a lightweight CNN model that exceeds to extract useful features from images but, decreases computational efficiency. This method is represented by its linear bottlenecks and inverted residual blocks that permit it to attain significant stability among model accuracy and size. Fig. 1 represents the structure of MobileNetv2.

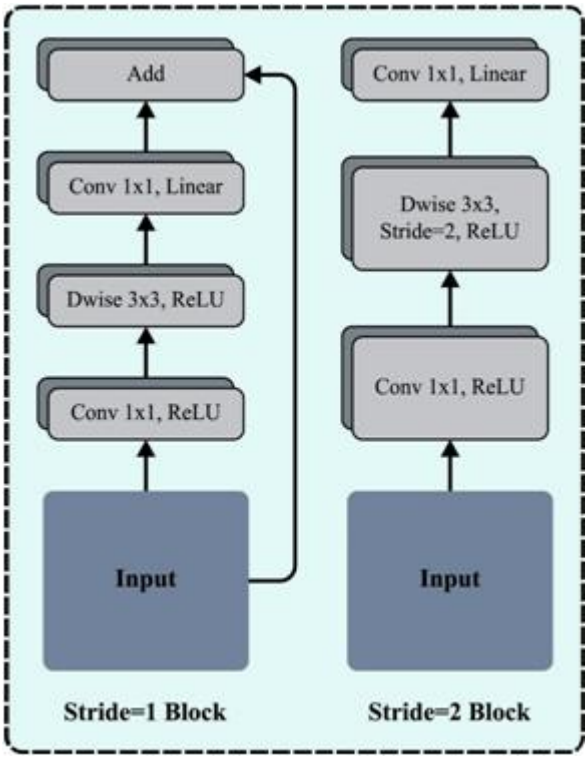

Figure 1: Structure of MobileNetv2

MobileNetV2 uses depthwise separate convolutions that comprise a 2 sequential processes namely depthwise and

pointwise convolutions. Firstly, depthwise convolution implements a separate convolutional filter to all input channels for decreasing the number of computations. Secondly, pointwise convolution integrates these channels linearly to generate the outputs. This technique majorly minimizes the computational cost and number of parameters in comparison with standard convolutional layers. MobileNetV2 also presents skip connections that enable the data flow through various layers, allowing the network to acquire features at multiple levels. The resultant feature maps from MobileNetV2 are applied as input to the following models or layers for different CV processes like as object detection, image classification, or semantic segmentation. Now, a simplified mathematical equation of depthwise separable convolution is given:
Depthwise Separable Convolution:

$$DepthwiseSeparableConv(x) = PointwiseConv\big(DepthwiseConv(x)\big) \quad (1)$$

Where x denotes the input feature map, $PointwiseConv$ signifies pointwise convolution and $DepthwiseConv$ represents depthwise convolution.This effective feature extraction ability of MobileNetV2 creates it a common selection in resource-limited environments but both computational efficiency and accuracy can be essential considerations.

3.3 Hyperparameter Tuning

To vary the hyperparameter values of the MobileNetv2 model optimally, the CSA is used.CSA is based on the egg-laying pattern of cuckoo with flood flight [19]. They commence by laying eggs on their nest of other birds and later wait for the host bird to lay the eggs near its own.

The table to enhance the given habitat should construct the values of the problem variable. The next N_{var} is an array of one N_{var} in the optimization problems:

$$habitat = [X_1, X_2, .., X_{Nvar}] \quad (2)$$

The (f_p) benefit function for habitat is estimated to define fitness, where “benefit” represents the present habitat in the existing formula (3):

$$profit = fb(habitat) = fb(X_1, X_2, \ldots\ldots\ldots, X_{Nvar}) \quad (3)$$

The optimization techniques require a habitat matrix of size $N_{pop} * N_{var}$. All the habitats provide a random chance to acquire one egg.
All the nesting range and cuckoo’s egg production are evaluated. Also, the cuckoos begin putting in the region that is near to the existing best possible area, and the throw radius can be defined as follows:

$$ELR = a * \frac{current\ cuckoo\ eggs\ number}{total\ eggs\ number} * (Var_{hi} - Var_{low}) \quad (4)$$

Each female cuckoo lay their eggs in nest within Eggs Laying Radius (ELR), and destroy a given proportion of the eggs that arenot as cost-effective as others (low-profit function). Chicks in host nest gains size and strength from the additional nutrients obtained.

When it’s time to start the family, cuckoo fly towards a newest position. Cuckoo is known to form a group in different positions before the optimum location of group is selected as a goal and others relocate there.
Cuckoo covers a large region by adjusting both values $(\Lambda, \varphi)_P$. is a random value ranges from $-\frac{\pi}{6}$ to $\frac{\pi}{6}$ and Λ denotes the random values within [0,1].

$$X_{Next\ Habitat} = X_{current\ Habitat} + F(X_{GoalPoin\ 1} - X_{current\ Habitat}) \quad (5)$$

Over 95% of cuckoos finish the CSO. Here, the eggs have a great affinity with the host and can exploit the host’s more robust food supply. Egg mortality is minimized, and potential profit is maximized.

3.4 Place Recognition Process

Lastly, the Minkowski distance is utilizedfor recognizing places through similarity measurement method. Minkowski Distance has a multipurpose distance metric employed for computing the similarity or dissimilarity among a 2 points in a multi-dimensional space, including numerous other distance metrics as particular condition. It's a generalized distance determination that regards the distance among points for their dissimilarity through all dimensions. The mathematical equation for Minkowski Distance can be given below:

$$D(x, y) = \left(\sum_{i=1}^{n} | xi - yi |^p \right)^{\frac{1}{p}} \quad (6)$$

In this equation, 'x' and 'y' denotes a 2 points in the multi-dimensional space, 'p' is a parameter that calculates the order of the Minkowski distance, and 'n' represents the number of dimensions. If 'p' is set to 1, the Minkowski Distance can be equal to the Manhattan Distance that determines the distance at grid lines. If 'p' is set to 2, it is the Euclidean Distance that measures the straight-line distance among points. By changing the values of 'p,' Minkowski Distance is adapted to various applications and cases, which permits it to take different notions of distance and the same in various conditions namely classification, clustering, and recommendation techniques.

4. Results and Discussion

The experimental validation of the CSADL-VPR approach is investigated under distinct performance measures. Fig. 2 represents the sample images.

Figure 2: Sample Images

Table 1 describes the comparative AUC_{score} analysis of CSADL-VPR approach with other existing models under four datasets [20].

Table 1: AUC_{score} analysis of CSADL-VPRapproach with other methods under four datasets

AUC Score (%)					
Datasets	CSADL-VPR	CoHOG	AMOSNet	HybridNet	DenseVLAD
SPEDTest	97.90	47.90	91.40	90.30	84.80
Nordland	82.27	10.46	30.00	17.50	12.95
Living Room	99.78	85.50	98.00	97.00	99.25
Synthia	99.57	79.50	89.00	91.30	98.80

Fig. 3 represents the AUC_{score} analysis of CSADL-VPR approach with existing methods on SPED Test database. The experimental values denote that CSADL-VPR algorithm has resulted in raised values of AUC_{score}. According toAUC_{score}, the CSADL-VPR model has gained an increased value of AUC_{score} of 97.90% whereas the CoHOG, AMOSNet, HybridNet, and Dense VLAD systems have attained minimum values of AUC_{score} of 47.90%, 91.40%, 90.30% and 84.80% correspondingly.

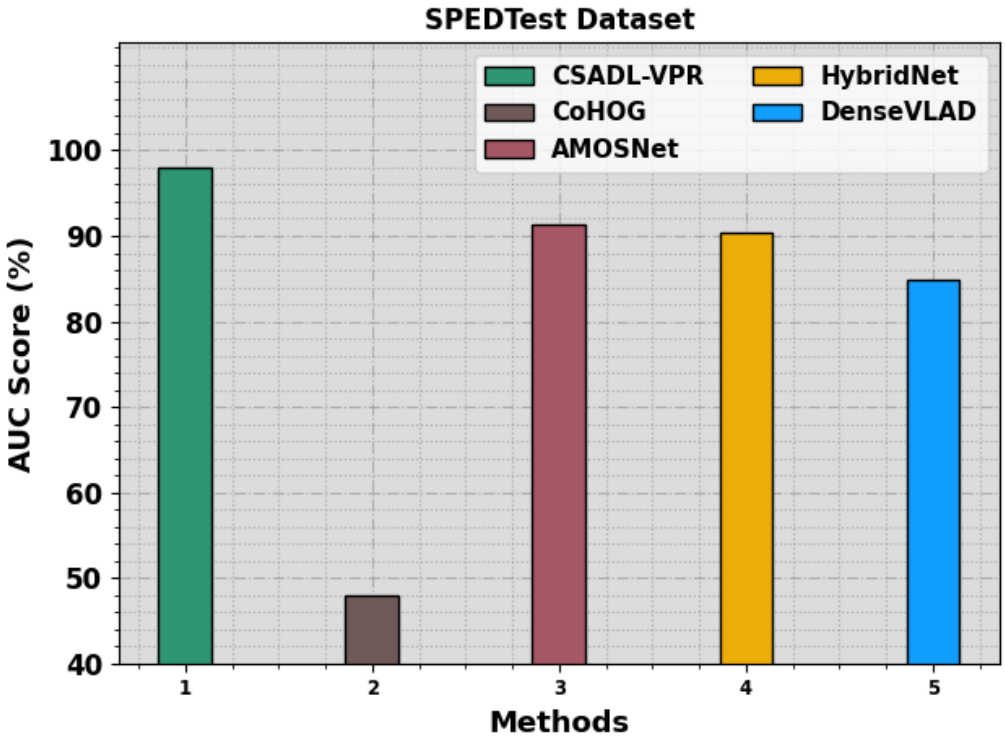

Figure 3: AUC_{score} analysis of CSADL-VPR approach on SPED Test dataset

Fig. 4 shows the AUC_{score} analysis ofCSADL-VPR algorithm with existing models on Nordland dataset. The outcome value simply that CSADL-VPR system leads to raised values of AUC_{score}. According to AUC_{score}, the CSADL-VPR technique has reached an improved value of AUC_{score} of 82.27% while the CoHOG, AMOSNet, HybridNet, and DenseVLAD methodologies obtained a reduced value of AUC_{score} of 10.46%, 30.00%, 17.50% and 12.95% respectively.

Fig. 5 illustrates the AUC_{score} analysis ofCSADL-VPR models with existing techniques on Living Room dataset. The simulation values indicate that CSADL-VPR methods have resulted in raised values of AUC_{score}. Additionally, based on AUC_{score}, the CSADL-VPR approach has achieved an improved value of AUC_{score} of 99.78%but the CoHOG, AMOSNet, HybridNet, and DenseVLAD systems acquired less values of AUC_{score} of 85.50%, 98.00%, 97.00% and 99.25% individually.

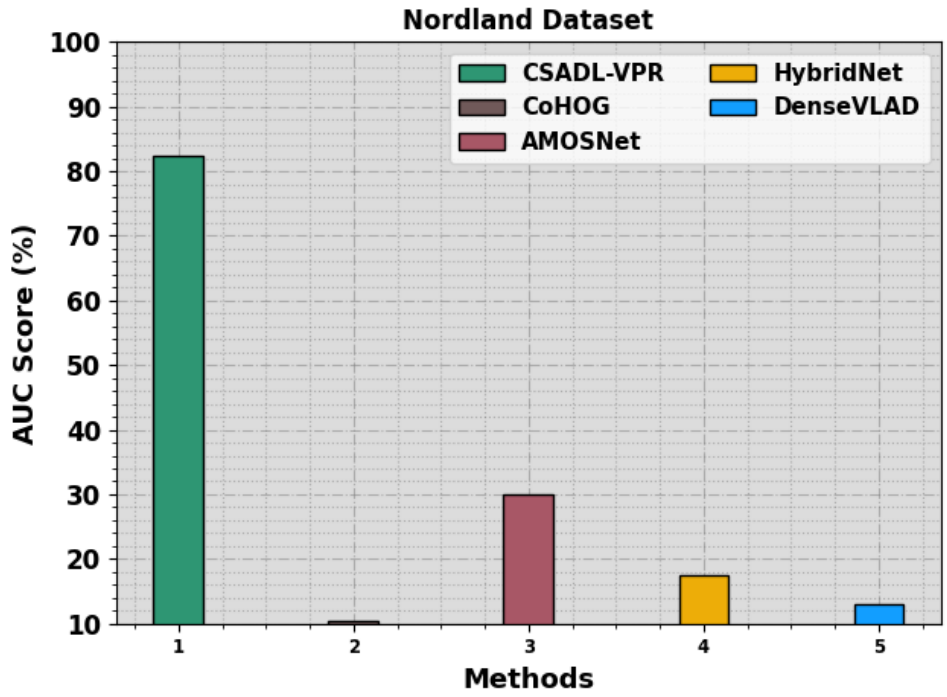

Figure 4: AUC_{score} analysis of CSADL-VPR approach on Nordland dataset

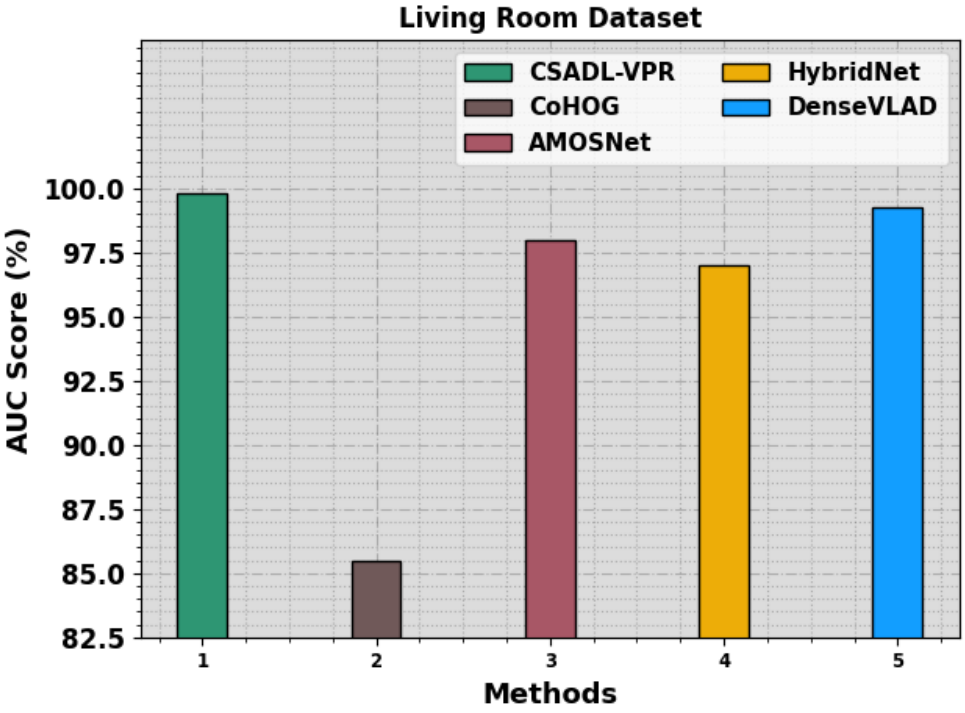

Figure 5: AUC_{score} analysis of CSADL-VPR approach on Living room dataset

Fig. 6 exhibiting the AUC_{score} analysis of CSADL-VPR system with existing techniques on Synthia dataset. The simulation values represent that CSADL-VPR approach leads to improved values of AUC_{score}. Furthermore, based on AUC_{score}, the CSADL-VPR models have reached an increased value of AUC_{score} of 99.57% whereas the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques acquired lower values of AUC_{score} of 79.50%, 89.00%, 91.30% and 98.80% correspondingly.

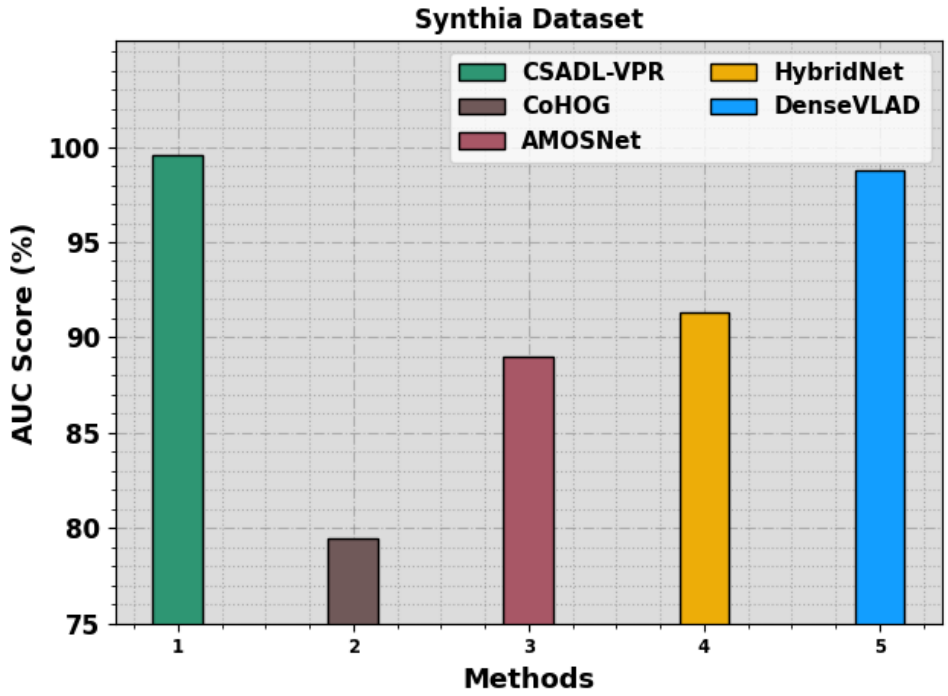

Figure 6: AUC_{score} analysis of CSADL-VPR approach on Synthia dataset

5. Conclusion

In this study, we introduce a CSADL-VPR method system to accurately and automatically recognize visual places. The purpose of the CSADL-VPR technique is the proper recognition of the visual places using DL model. This presents CSADL-VPR approach a novel approach to visual place recognition that leverages the GF based pre-processing, MobileNetV2 feature extractor, CSA, and Manhattan distance-based similarity measurement. To ensure the enhanced performance of the CSADL-VPR approach, a series of experimental analyses can be

performed. The simulation outcomes portrayed the enhancements of the CSADL-VPR system on place recognition process.

References

[1] Arcanjo, B., Ferrarini, B., Milford, M., McDonald-Maier, K.D. and Ehsan, S., 2022. An efficient and scalable collection of fly-inspired voting units for visual place recognition in changing environments. *IEEE Robotics and Automation Letters*, *7*(2), pp.2527-2534.

[2] Khaliq, A., Ehsan, S., Chen, Z., Milford, M. and McDonald-Maier, K., 2019. A holistic visual place recognition approach using lightweight cnns for significant viewpoint and appearance changes. *IEEE transactions on robotics*, *36*(2), pp.561-569.

[3] Maltar, J., Marković, I. and Petrović, I., 2020. Visual place recognition using directed acyclic graph association measures and mutual information-based feature selection. *Robotics and Autonomous Systems*, *132*, p.103598.

[4] Chen, B., Song, X., Shen, H. and Lu, T., 2021. Hierarchical Visual Place Recognition Based on Semantic-Aggregation. *Applied Sciences*, *11*(20), p.9540.

[5] Xu, D., Liu, J., Hyyppä, J., Liang, Y. and Tao, W., 2022. A heterogeneous 3D map-based place recognition solution using virtual LiDAR and a polar grid height coding image descriptor. *ISPRS Journal of Photogrammetry and Remote Sensing*, *183*, pp.1-18.

[6] Zhang, W., Yan, Z., Wang, Q., Wu, X. and Zuo, W., 2020. Learning second-order statistics for place recognition based on robust covariance estimation of CNN features. *Neurocomputing*, *398*, pp.197-208.

[7] Ferrarini, B., Milford, M.J., McDonald-Maier, K.D. and Ehsan, S., 2022. Binary neural networks for memory-efficient and effective visual place recognition in changing environments. *IEEE Transactions on Robotics*, *38*(4), pp.2617-2631.

[8] Camara, L.G. and Přeučil, L., 2020. Visual place recognition by spatial matching of high-level CNN features. *Robotics and Autonomous Systems*, *133*, p.103625.

[9] Tomită, M.A., Zaffar, M., Milford, M.J., McDonald-Maier, K.D. and Ehsan, S., 2021. Convsequential-slam: A sequence-based, training-less visual place recognition technique for changing environments. *IEEE Access*, *9*, pp.118673-118683.

[10] Islam, T., Rabiul Islam, S. and Rahman, M., 2022. Learning Condition-Invariant Scene Representations for Place Recognition across the Seasons Using Auto-Encoder and ICA. *Journal of Electrical and Computer Engineering*, *2022*.

[11] Zaffar, M., Ehsan, S., Milford, M. and McDonald-Maier, K., 2020. Cohog: A light-weight, compute-efficient, and training-free visual place recognition technique for changing environments. *IEEE Robotics and Automation Letters*, *5*(2), pp.1835-1842.

[12] SC, D., PETKOV, N. and ANTEQUERA, M.L., 2019. COMPUTER VISION TECHNIQUES FOR CALIBRATION, LOCALIZATION AND RECOGNITION.

[13] Schubert, S., Neubert, P. and Protzel, P., 2021. Fast and Memory Efficient Graph Optimization via ICM for Visual Place Recognition. In *Robotics: Science and Systems*.

[14] Yang, B., Xu, X., Li, J. and Zhang, H., 2019. Landmark generation in visual place recognition using multi-scale sliding window for robotics. *Applied Sciences*, *9*(15), p.3146.

[15] Xie, S., Pan, C., Peng, Y., Liu, K. and Ying, S., 2020. Large-scale place recognition based on camera-lidar fused descriptor. *Sensors*, *20*(10), p.2870.

[16] Islam, M.T., Hasib, K.M., Rahman, M.M., Tusher, A.N., Alam, M.S. and Islam, M.R., 2022. Convolutional Auto-Encoder and Independent Component Analysis Based Automatic Place Recognition for Moving Robot in Invariant Season Condition. *Human-Centric Intelligent Systems*, pp.1-12.

[17] Nyemeesha, V. and Ismail, B.M., 2021. Implementation of noise and hair removals from dermoscopy images using hybrid Gaussian filter. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *10*, pp.1-10.

[18] Srinivasu, P.N., SivaSai, J.G., Ijaz, M.F., Bhoi, A.K., Kim, W. and Kang, J.J., 2021. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, *21*(8), p.2852.

[19] Basil, N. and Marhoon, H.M., 2023. Towards evaluation of the PID criteria based UAVs observation and tracking head within resizable selection by COA algorithm. *Results in Control and Optimization*, p.100279.

[20] Zaffar, M., Garg, S., Milford, M., Kooij, J., Flynn, D., McDonald-Maier, K. and Ehsan, S., 2021. Vpr-bench: An open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change. *International Journal of Computer Vision*, *129*(7), pp.2136-2174.