

Original Article

Intelligent Visual Place Recognition using Sparrow Search Algorithm with Deep Transfer Learning Model

S. Senthamizhselvi¹, A. Saravanan²

^{1,2}Department of Computer & Information Science, Annamalai University, Annamalai Nagar, India.

¹Corresponding Author : selvikarthi2002@gmail.com

Received: 12 November 2022

Revised: 30 March 2023

Accepted: 13 April 2023

Published: 25 April 2023

Abstract - The application of Deep Learnings (DLs) is thriving in the domain of Visual Place Recognitions (VPRs) that serves an indispensable role in visual Simultaneous Localization and Mapping (vSLAM) applications. The usage of Convolutional Neural Networks (CNNs) attains superior performance compared to handcrafted feature descriptors. However, still, VPR is a difficult task because of the two major issues they are perceptual variability and perceptual aliasing. This study develops an Intelligent Visual Place Recognition using Sparrow Search Algorithm with Deep Transfer Learning (IVPR-SSADTL) model. The presented IVPR-SSADTL technique recognizes the visual places effectively and accurately. It involves a three-phase process: feature extraction, hyperparameter tuning, and place recognition. At the initial phase, the IVPR-SSADTL technique employs the MixNet model as a feature extractor with the sparrow search algorithm (SSA) as a hyperparameter optimizer. Next, in the later phase, the IVPR-SSADTL technique applies Manhattan distance-based similarity measurement to recognize the places promptly. To exhibit the higher performance of the IVPR-SSADTL system, an extensive range of simulations were performed. A wide range of comparison studies stated the improved achievement of the IVPR-SSADTL algorithm over other models.

Keywords - Visual places recognition, Transfer learning, Deep learning, MixNet model, Feature extraction, Sparrow search algorithm.

1. Introduction

Autonomous systems functioning in dynamic, unstructured, and challenging atmospheres need excessive localization capabilities demonstrating robustness to accumulative odometry faults [1]. A general way to raise such robustness is by using a place recognition engine which is a mechanism that employs the recognition of revisited scenes for recovering the position of robots in localization failure scenarios or rectifying the estimated odometry [2]. A place detection engine depends on visual sensing methods, usually classified as Visual Place Recognition (VPR). VPR techniques should display environmental differences to achieve augmented robustness, invariance over lighting, and viewpoint [3]. Especially, long trajectory situations, with extreme appearance variations because of different year seasons (winter and summer) or day periods (night and day), have fostered VPR into one of the difficult errands in robotic vision [5]. Visual cues related to VPR were usually not uniformly dispersed across an image; thus, concentrating on significant areas, as opposed to confusion or irrelevant regions, was key to enhancing the VPR performance [6]. For instance, while detecting a street scene, utilizing features derived from time-varying objects, like moving cars or pedestrians, as opposed to those derived from static

structures, like road signs or buildings, could present misleading data in place detection [8].

Inspired by such achievements, several researchers were conducted to investigate the Convolutional Neural Networks (CNNs) efficiency of features implemented to the VPR issue [9]. The authors gained attributes from various layers of a specific CNN method and compared them in contrast to several existing sequence-related VPR methods. They display that the CNNs middle layer outpaces any other algorithm or layer, even in the event of single image matching [10]. Long-term VPR imposes a vital challenge in robot navigation due to the reason that a single place undergoes substantial appearance variations because of illumination and seasonal or weather changes. In recent times, the success of deep learning (DL) in computer vision (CV) has activated a range of inspections into how to generate a feature representation from CNN that is robust to such variations [11]. To further enhance CNN's performance, the CV community put forth efforts not just towards constructing more complicated and deeper network structures but even towards a better insight into how their Daedalian framework performs in distinct stimuli and occasions [12].



This study develops an Intelligent Visual Place Recognition using Sparrow Search Algorithm with Deep Transfer Learning (IVPR-SSADTL) model. The presented IVPR-SSADTL technique recognizes the visual places effectively and accurately. It involves a three-phase process: feature extraction, hyperparameter tuning, and place recognition. At the initial phase, the IVPR-SSADTL technique employs the MixNet model as a feature extractor with the sparrow search algorithm (SSA) as a hyperparameter optimizer. Next, in the later phase, the IVPR-SSADTL technique applies Manhattan distance-based similarity measurement to recognize the places promptly. To exhibit the higher performance of the IVPR-SSADTL system, an extensive range of simulations were performed.

2. Literature Review

In [13], a novel deep distance learning structure for VPR was introduced. With in-depth research of several restrictions of the distancing connection from the VPR issues, the multiconstraint loss function has been predicted to optimize the distancing restriction connections from the Euclidean space. A novel infrastructure assists different types of CNN, such as VGGNet, AlexNet, and other user-defined networking, to derive further individual aspects. In [14], deep metric learning that combined optimization extracting feature and similarity metric was employed to train end-to-end networking in detail for place identification tasks to handle the occurrence altering over time. A self-adaptively improved similarity metric was planned to strengthen the discrimination capability and compute the similarity betwixt descriptors of image pairs that can be extracted in a CNN.

Park et al. [15] project a lightweight CNN technique for VPR. The presented system specifically targets the embedding method. To reduce the computational difficulty of the networks, the author's proposal is an FCN structure with some filters and layers. The presented network directly learns a vector space, whereas their distance is equivalent for placing similarity with metric learning. Mao et al. [17] introduce a new technique for building a multi-scale feature pyramid and project 2 techniques for utilizing the pyramid for augmenting the place recognition ability. A primary method fuses the pyramid for obtaining novel mapping features that is attention to either local or semi-global presence. In the secondary system, learn an attention method in the feature pyramid for weighting the spatial grid on novel mapping features. Both techniques integrate the multiscale features from the pyramid to suppress the confusing local feature but address the problems from 2 distinct approaches.

Zhu et al. [19] introduce a novel technique dependent upon CNN, with place image as to the pre-training network system for obtaining automatically learned image descriptor, and with several functions of binarization, pooling, and fusion for optimizing them, afterward, the resemblance outcome of location recognitions are projected with

Hamming distance of place sequences. Pei et al. [20] examine a rapid and dependable technique utilizing BoW of structural lines demonstrated by Extended Line Band Descriptors, including Pose estimate ability termed ELBDP. Additionally, the robust system that only utilizes one set of vanishing points and a single structural line for estimating the comparative posing betwixt imaging sets was presented.

3. The Proposed Model

In this research, an automated IVPR-SSADTL technique has been developed to recognise visual places. The presented IVPR-SSADTL technique automatically recognizes the visual places effectively and accurately. It involves a three-phase process: feature extraction, hyperparameter tuning, and place recognition.

3.1. Feature Extraction: MixNet Model

At the beginning stage, the IVPR-SSADTL technique employs the MixNet model as a feature extractor. A CNN architecture built by the classical complex function is hard to operate for mobile terminals as a consequence of excessive parameters and complex computation [22]. A set of lightweight convolutional operators was introduced to guarantee the model accuracy and enhance the efficiency of the mobile terminal. In particular, the depthwise separable convolution layer was more commonly used, and they split the convolutions into pointwise and depthwise convolutions. At first, it convolves a single channel consecutively using 3 convolution kernels. Next, it applies the feature maps with a kernel size of 1×1 convolution layer. Consider that N $D_k \times D_k$ feature view and 1 convolution sliding steps were exploited to convolve factor maps with $D_F \times D_F \times M$ dimension, along with the output feature maps $D_F \times D_F \times N$ dimension, and it is given in the following expression:

$$D_k \times D_k \times M \times N \quad (1)$$

The parameter included in the depthwise convolution function can be given as follows:

$$D_k \times D_k \times M + 1 \times 1 \times M \times N \quad (2)$$

The calculation included in the convolutional function is shown below:

$$D_k \times D_k \times M \times N \times D_F \times D_F \quad (3)$$

The calculation included in the depthwise convolution function is shown in the following expression:

$$D_k \times D_k \times M \times D_F \times D_F \times M \times N \times D_F \times D_F \quad (4)$$

The ratio of two functions is determined by:

$$\frac{D_k \times D_k \times M \times D_F \times D_F \times M \times N \times D_F \times D_F}{D_k \times D_k \times M \times N \times D_F \times D_F} \quad (5)$$

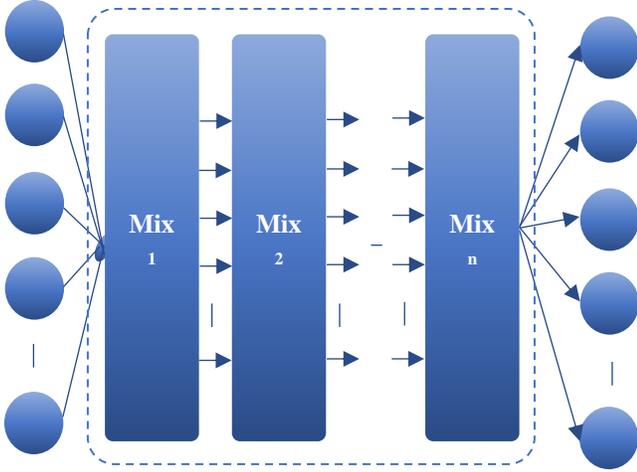


Fig. 1 Structure of MixNet system

The depthwise separable convolution function uses a similar kernel size of 3×3 convolution layer in the computation model. However, a network with a huge complexity kernel of 5×5 or 7×7 determines that a huge complexity kernel enhances the accuracy and efficacy of the algorithm. On the other hand, the study demonstrates that it is an uncommon case that the bigger the complexity kernels are, the better since larger complexity kernels minimize the model precision. Now, the MDConv divides input channels with M size into C ensembles and later convolutes every ensemble with dissimilar kernel magnitudes. Fig. 1 illustrates the architecture of the MixNet method.

3.2. Hyperparameter Tuning: SSA Technique

In this study, the SSA is used as a hyperparameter optimizer. SSA is a novel nature-inspired technique based on the behaviors of sparrows [23]. Many animals search for food and avoid predators with the swarm intelligence in nature. They are classified into 2 classes based on their suitability and are defined by the unique posture of every sparrow. The individual who has the best ft will belong to the producer. The residual sparrows are onlookers. In the entire populace, distinct individuals have dissimilar eating behaviors.

Furthermore, some sparrows are answerable to avoid predators during the foraging processes amongst the population. They decide to fly closer or farther to another sparrow to face this danger. The sparrow colony finds more low-risk food by continually updating its location. SSA is recommended by mimicking the sparrow's anti-hunting group and search behavior. This approach has faster performance, fewer parameters, and a robust search capability. The major phases of SSA are discussed in the following: Step 1, initialize and create the solution. Here, determine the ratio of sparrows in intensive care (PV), the magnitude of the populace, the maximal amount of reproduces, and the Producer Ratio (PD). The first location of the sparrow population has been demonstrated as follows. They are randomly generated:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,d} \\ x_{2,1} & x_{2,2} & x_{2,d} \\ \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & x_{n,d} \end{bmatrix} \quad (6)$$

In Eq. (6), n denotes the count of sparrows, and d indicates the dimension of chosen parameters. Every person's fitness for the subsequent process is defined below. Every row value in FX signifies every individual's ft.

$$F_X = \begin{bmatrix} f[x_{1,1} & x_{1,2} & x_{1,d}] \\ f[x_{2,1} & x_{2,2} & x_{2,d}] \\ \vdots & \vdots & \vdots \\ f[x_{n,1} & x_{n,2} & x_{n,d}] \end{bmatrix} \quad (7)$$

Step2: In this phase, producers with the highest suitability value have prioritized over those who produce cuisine. Since the producer is responsible for searching for cuisine and directing the movements of the whole population, the producer could find cuisine in a wider range than the explorer location. Based on steps (1) and (2), in all the iterations, the manufacturer updates the status using the following equation:

$$X_{ij}^{t+1} = \begin{cases} X_{i,i}^t \times \exp\left(\frac{-i}{\alpha \times iter_{max}}\right) & \text{if } R_2 < ST \\ X_{ij}^t + Q \times L & \text{if } R_2 \geq ST \end{cases} \quad (8)$$

Let, $iter_{max}$ be the constant with the maximum iteration amount. t denotes the existing iteration, and $j = 1, 2, \dots, d$, X_{ij}^t characterize the resulting value of j -th sparrows at t iteration. α indicates a random integer between zero and one. R_2 (Alert value) is a value within $[0,1]$, and Safe Threshold (ST) was a number ranging from 0.5 to 1.0. Q denotes a uniformly distributed arbitrary integer. L signifies $a1 \times d$ matrix where every component is 1. Once it is $R_2 \geq ST$, few sparrows have found the hunter, and each sparrow needs to fly rapidly to another safe area. If $R_2 < ST$, then no hunting individual is around, and the manufacturing individual arrives in the wide searching phase.

For explorers, Rules 4 and 5 should be followed. From the abovementioned, some explorer keeps a tab on the most manufacturer. They leave their present position to contend for food once they learn that producers have recognized delicious food. Once they win, they could eat immediately; or else Rule 5 will be applied.

Updating location for the explorer can be determined based on the following equation. x_p does either manufacturer occupies the optimum location. X_{worst} signifies the worst location. A implies a $1 \times d$ matrixes that are arbitrarily allocated 1 or-1 to all the elements inside $A^+ = A^T(AA^T)^{-1}$. When $i > \frac{n}{2}$ then it specifies that i -th probes with worst ft values are highly possible to be hungry.

$$X_{i,t}^{t+1} = \begin{cases} Q \times \exp\left(\frac{X_{worst}^t - X_{ij}^t}{i^2}\right) & \text{if } > \frac{n}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \times A^+ \times L & \text{otherwise} \end{cases} \quad (9)$$

Step 3 Afterward, position updating of the entire population, many sparrows are chosen as scouts (exploration) accountable for warning and identification. Usually, they make up 10 to 20% of the overall population. Position updating can be determined based on Rule 6 as follows:

$$X_{ij}^{t+1} = \begin{cases} X_{best}^t + \beta \times |X^t - X_{best}^t| & \text{if } f_i > f_g \\ X_{ij}^t + K \times \left(\frac{|x_{ij}^t f - x_{worst}^t|}{(f_i - f_w) + \varepsilon}\right) & \text{if } f_i = f_g \end{cases} \quad (10)$$

Now, X_{best} indicates the present global optimum location. ε denotes a smaller constant for the avoidance of zero-division error. β denotes a control variable for step size and uniformly distributed random integer within [0, 1]. f_g and f_w denotes the present best and worst fitness value, correspondingly. K refers to a random integer ranging from 1 to -1. f_i indicates the present value of the sparrow. $f_i = f_g$ shows that sparrows among the populace are conscious of risk and must be approached by the rest. When $f_i > f_g$ the individual is at the edge of the group. X_{best} shows the central position of the population and was safer around it. K signifies the direction in which the individual moves and the step size controlling aspect.

Step 4: every individual's present location is associated with the latter recurrence. The updating process can be completed when the novel location is more efficient than the preceding one and save the better location. The existence of a few sparrows might enhance the final two stages afterwards.

Step 5 when the number of recurrences is lesser than the maximum number, go to step2. Or else the process ends, and a better solution is attained. Fig. 2 demonstrates the steps involved in SSA.

3.3. Similarity Measurement: Manhattan Distance

In the last phase, the IVPR-SSADTL technique applied the Manhattan distance based on resemblance measurement to recognize the places promptly. To provide QI, the factor vectors will be compared and extracted with all factor vectors in the IFD [24]. For Manhattan distance measuring and every factor vector, the adjacent N candidate images were supplemented to the histogram of locations. Once the total querying vectors are considered, the output histogram is leveraged in deriving a list of N high-ranking candidate images. The Manhattan distancing can be defined as the count of absolute variances among 2 vectors. In 2D space, the Manhattan distancing is defined as:

$$d = |p_1 - q_1| + |p_2 - q_2| \quad (11)$$

In order to n-dimensional space, the Manhattan distancing for 2 data points p_i and q_i is denoted in Eq. (11):

Algorithm 1: Pseudocode of SSA

```

Input:
G: Maximal amount of iterations
PD: The count of producers
SD: The agent counts (sparrows) warned of the danger
R2: the alarming value
n: The count of agents
Generate a populace of n agents and define the crucial parameter.
Output: Xbest, fg.
While (t < G)
Ranking of the fitness value to define the present worst and finest individuals.
R2 = random(1)
For i in (1, PD)
Upgrade the agent location based on Eq. (8);
Out of For
For i in ((PD + 1), n)
Upgrade the location of the agent based on Eq. (9);
Out of For
For T in (1, SD)
Upgrade the location of the agent based on Eq. (10);
Out of For
Acquire the new location;
if the novel location is more efficient than the previous one, then upgrade;
t = t + 1
End While
Return Xbest, fg.
    
```



Fig. 2 Steps involved in SSA

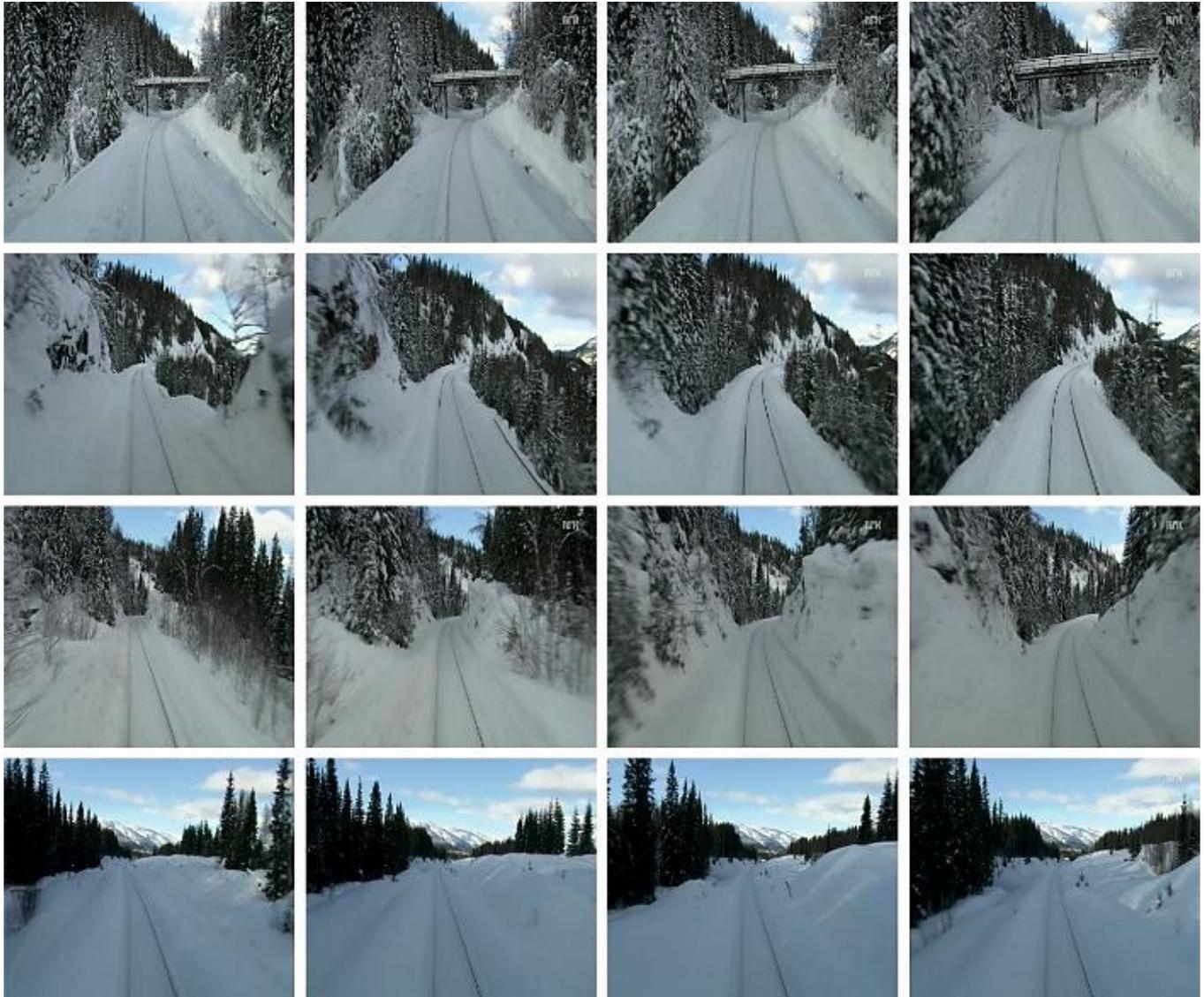


Fig. 3 Sample Image

4. Results and Discussion

The experimental validation of the IVPR-SSADTL algorithm is investigated under distinct performance measures. Fig. 3 depicts a few trial imageries.

Table 1. Precision-recall investigation of IVPR-SSADTL system with current approaches under four datasets

Recall (%)	Precision (%)				
	IVPR-SSADTL	CoHOG	AMOS Net	Hybrid Net	Dense VLAD
Nordland Dataset					
0	100.00	100.00	100.00	100.00	100.00
10	98.33	98.25	96.55	96.92	95.00
20	98.33	98.25	96.55	93.85	95.00
30	98.33	98.25	94.83	90.77	93.33
40	96.67	96.25	87.93	90.77	91.67
50	96.67	96.49	77.59	84.62	81.67

60	91.67	94.74	63.79	75.38	66.67
70	70.00	69.47	48.28	60.00	53.33
80	46.67	33.68	32.76	40.00	36.67
90	23.33	22.11	15.52	20.00	18.33
100	0.00	0.00	0.00	0.00	0.00
SPEDTest Dataset					
0	100.00	100.00	100.00	100.00	100.00
10	98.28	98.25	98.28	96.92	96.67
20	98.28	96.49	98.28	92.31	93.33
30	96.49	94.83	94.83	89.23	91.67
40	94.74	93.10	93.10	87.69	90.00
50	94.74	82.76	82.76	81.54	80.00
60	89.47	72.41	72.41	72.31	65.00
70	82.46	53.45	53.45	55.38	50.00
80	61.40	36.21	36.21	36.92	35.00
90	31.58	17.24	17.24	18.46	18.33
100	0.00	0.00	0.00	0.00	0.00

Synthia Night to Fall Dataset					
0	98.33	100.00	100.00	100.00	100.00
10	98.33	98.25	98.28	96.55	96.67
20	98.33	98.25	98.28	96.55	93.33
30	98.33	98.25	94.83	94.83	91.67
40	98.33	96.49	93.10	87.93	88.33
50	98.33	94.74	82.76	81.03	78.33
60	98.33	92.98	72.41	70.69	63.33
70	96.67	85.96	53.45	53.45	48.33
80	96.67	70.18	36.21	36.21	31.67
90	91.67	36.84	17.24	17.24	15.00
100	0.00	0.00	0.00	0.00	100.00
Living Room Dataset					
0	100.00	100.00	100.00	100.00	100.00
10	98.33	98.25	98.25	98.46	98.33
20	98.33	98.25	96.49	98.46	98.33
30	98.33	98.25	96.49	95.38	96.67
40	98.33	96.49	91.23	90.77	90.00
50	98.33	94.74	89.47	81.54	65.00
60	96.67	92.98	78.95	66.15	50.00
70	85.96	80.00	59.65	50.77	33.33
80	70.18	53.33	40.35	32.31	16.67
90	36.84	26.67	21.05	16.92	8.00
100	0.00	0.00	0.00	0.00	0.00

Table 1 and Fig. 4 report an overall precision-recall examination of the IVPR-SSADTL model with compared methods [25].

The experimental outcomes inferred that the IVPR-SSADTL technique had depicted improved results.

For instance, with $reca_l$ of 10%, the IVPR-SSADTL approach has given a higher $prec_n$ of 98.33%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD models have attained lower $prec_n$ of 98.25%, 96.55%, 96.92%, and 95% respectively.

Eventually, with $reca_l$ of 40%, the IVPR-SSADTL technique has offered a higher $prec_n$ of 96.67%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques have acquired lower $prec_n$ of 96.25%, 87.93%, 90.77%, and 91.67% respectively.

Meanwhile, with $reca_l$ of 90%, the IVPR-SSADTL technique has presented a higher $prec_n$ of 23.33%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques have attained lower $prec_n$ of 22.11%, 15.52%, 20%, and 18.33% correspondingly.

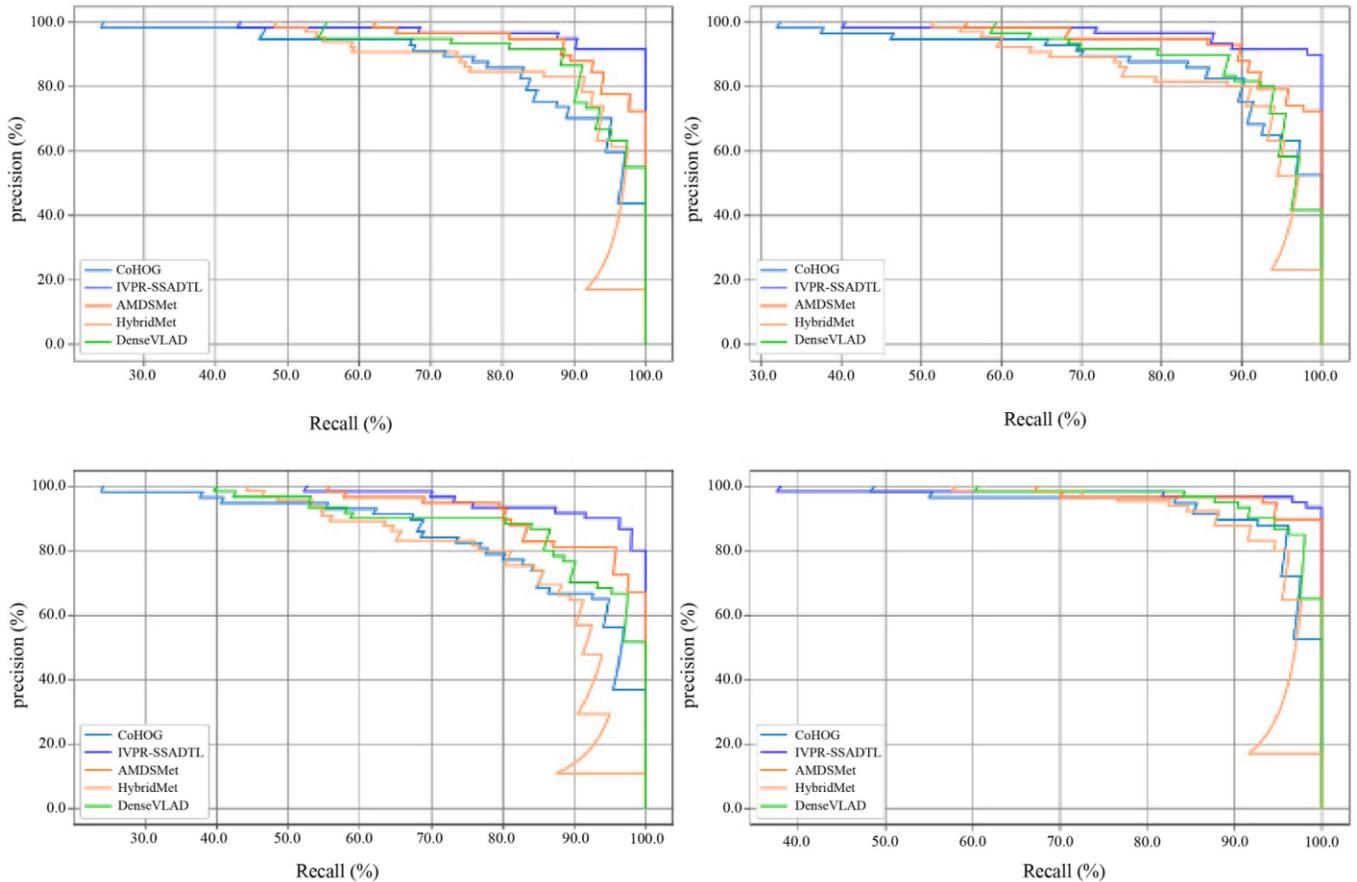


Fig. 4 Precision-recall analysis of IVPR-SSADTL system (a) Nordland Dataset, (b) SPEDTest Dataset, (c) Synthia Night to Fall Dataset, and (d) Living Room Dataset

Table 2. Recall rate analysis of IVPR-SSADTL system with existing approaches under four datasets

N (steps)	Recall Rate(%)				
	IVPR-SSADTL	CoHOG	AMOS Net	Hybrid Net	DenseVLAD
Nordland Dataset					
0	0.00	0.00	0.00	61.54	0.00
2	1.67	43.86	72.41	63.08	55.00
4	91.67	70.18	77.59	78.46	63.33
6	91.67	75.44	84.48	84.62	66.67
8	95.00	78.95	87.93	87.69	73.33
10	95.00	85.96	89.66	89.23	75.00
12	96.67	87.72	94.83	92.31	75.00
14	96.67	91.23	94.83	93.54	86.67
16	98.33	92.98	96.55	95.38	91.67
18	98.33	98.25	98.28	98.46	93.33
20	100.00	100.00	100.00	100.00	95.00
SPEDTest Dataset					
0	0.00	0.00	0.00	0.00	0.00
2	1.67	52.63	72.41	23.08	41.67
4	90.00	64.91	74.14	63.08	58.33
6	91.67	68.42	79.31	80.00	71.67
8	91.67	82.46	84.48	81.54	81.67
10	93.33	87.72	87.93	86.15	83.33
12	96.67	89.47	93.10	89.23	90.00
14	96.67	92.98	94.83	90.77	91.67
16	98.33	94.74	98.28	95.38	95.00

18	100.00	98.25	98.28	98.46	96.67
20	100.00	100.00	100.00	100.00	100.00
Synthia Night to Fall Dataset					
0	0.00	0.00	0.00	0.00	0.00
2	80.00	56.14	67.24	29.23	51.67
4	86.67	66.67	72.41	56.92	68.33
6	90.00	73.68	81.03	69.23	76.67
8	91.67	77.19	82.76	75.38	80.00
10	93.33	80.70	89.66	81.54	88.33
12	95.00	84.21	93.10	86.15	90.00
14	96.67	89.47	94.83	90.77	91.67
16	98.33	92.98	96.55	93.85	95.00
18	100.00	96.49	98.28	96.92	98.33
20	100.00	100.00	100.00	100.00	100.00
Living Room Dataset					
0	0.00	0.00	0.00	0.00	0.00
2	93.33	52.63	1.72	16.92	65.00
4	93.33	71.93	89.66	64.62	85.00
6	95.00	87.72	89.66	80.00	86.67
8	95.00	89.47	94.83	83.08	90.00
10	96.67	91.23	94.83	87.69	93.33
12	96.67	94.74	96.55	93.85	95.00
14	98.33	96.49	96.55	95.38	96.67
16	98.33	98.25	98.28	96.92	98.23
18	100.00	98.36	98.28	98.46	98.23
20	100.00	100.00	100.00	100.00	100.00

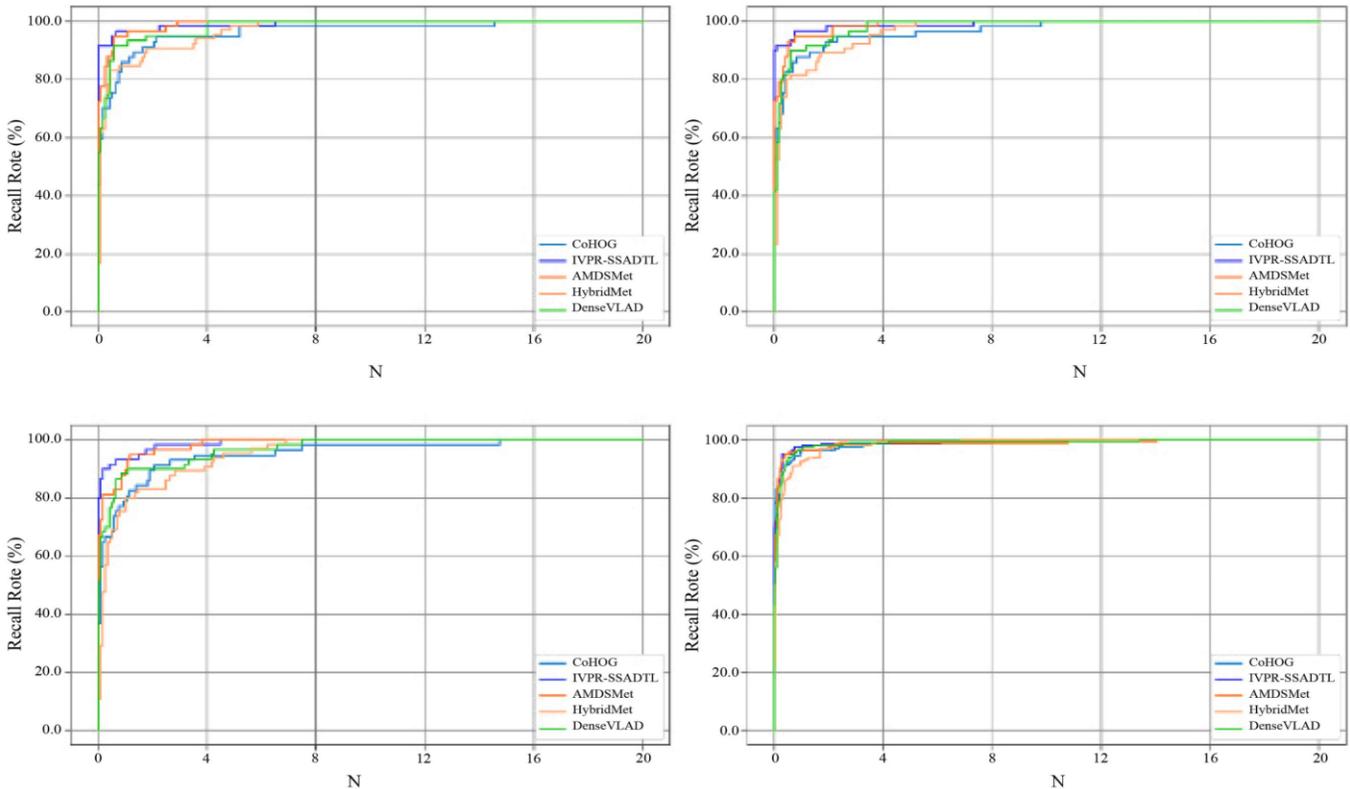


Fig. 5 Recall rate analysis of IVPR-SSADTL system (a) Nordland Dataset, (b) SPEDTest Dataset, (c) Synthia Night to Fall Dataset, and (d) Living Room Dataset

Table 2 and Fig. 5 provide a comprehensive RR of the IVPR-SSADTL approach with recent models. The results signified that the IVPR-SSADTL model has resulted in increased values of RR. With the Norland dataset and 4 steps, the IVPR-SSADTL model has gained an improved RR value of 91.67% while the CoHOG, AMOSNet, HybridNet, and DenseVLAD models have obtained decreased RR values of 70.18%, 77.59%, 78.46%, and 63.33% respectively. Concurrently, with the Norland dataset and 16 steps, the IVPR-SSADTL technique has achieved an improved RR value of 98.33%. At the same time, the CoHOG, AMOSNet, HybridNet, and DenseVLAD models have gained decreased RR values of 92.98%, 96.55%, 95.38%, and 91.67%, correspondingly. Simultaneously, with the SPEDTest dataset and 4 steps, the IVPR-SSADTL techniques have gained an improved RR value of 90% while the CoHOG, AMOSNet, HybridNet, and DenseVLAD technique has gained decreased RR values of 64.91%, 74.41%, 23.08%, and 41.67% correspondingly. Along with that, with the SPEDTest dataset and 10 steps, the IVPR-SSADTL technique has gained an improved RR value of 93.33%. In contrast, the CoHOG, AMOSNet, HybridNet, and DenseVLAD methods have gained decreased RR values of 87.72%, 87.93%, 86.15%, and 83.33%, respectively. Finally, with the SPEDTest dataset and 16 steps, the IVPR-SSADTL model has obtained an enhanced RR value of 98.33% while the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques have gained decreased RR values of 94.74%, 98.28%, 95.38%, and 95% correspondingly.

Fig. 6 exhibits a comparison study of the IVPR-SSADTL model with existing models on SPEDTest and Nordland datasets. On the SPEDTest dataset, the IVPR-SSADTL model has reached increased AUC_{score} of 94.48%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD models have resulted in reduced AUC_{score} of 47.90%, 91.41%, 90.29%, and 84.83% respectively. Besides, on the Nordland dataset, the IVPR-SSADTL method has reached increased AUC_{score} of 76.42%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques have resulted in reduced AUC_{score} of 8.46%, 29.99%, 17.37%, and 12.96% correspondingly.

Fig. 7 exhibits relative research of the IVPR-SSADTL approach with prevailing models on Living Room and Synthia datasets. On the Living Room dataset, the IVPR-SSADTL approach has reached increased AUC_{score} of 99.46%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques have resulted in reduced AUC_{score} of 85.27%, 99.90%, 96.97%, and 99.24% respectively.

Similarly, on the Synthia dataset, the IVPR-SSADTL method has reached increased AUC_{score} of 99.19%, while the CoHOG, AMOSNet, HybridNet, and DenseVLAD techniques have resulted in reduced AUC_{score} of 79.40%, 88.93%, 91.35%, and 98.78% correspondingly.

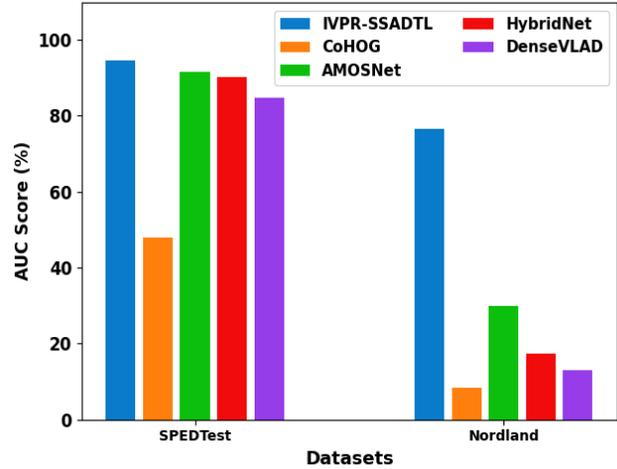


Fig. 6 AUC_{score} analysis of the IVPR-SSADTL system under SPEDTest and Nordland datasets

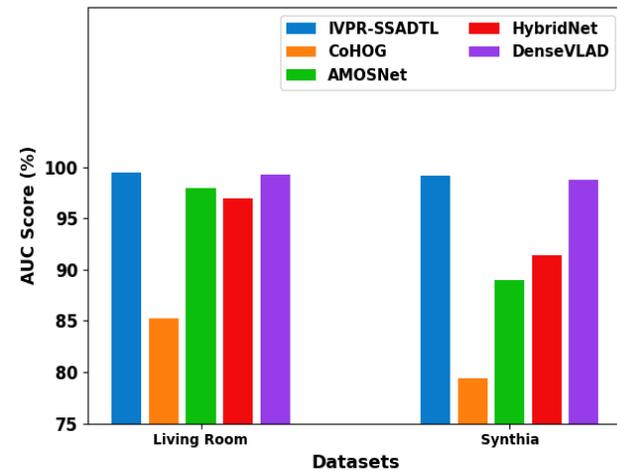


Fig. 7 AUC_{score} analysis of IVPR-SSADTL system under Living Room and Synthia datasets

5. Conclusion

In this study, an automated IVPR-SSADTL technique is developed to recognise visual places. The presented IVPR-SSADTL technique automatically recognizes the visual places effectively and accurately. It involves a three-phase process: feature extraction, hyperparameter tuning, and place recognition. At the beginning stage, the IVPR-SSADTL technique employs the MixNet model as a feature extractor with the SSA as a hyperparameter optimizer. At the same time, the IVPR-SSADTL technique applied the Manhattan distance-based resemblance measurement to identify the locations promptly. A wide range of simulations was performed to exhibit the enhanced performance of the IVPR-SSADTL algorithm. A large-scale comparison study stated the improved accomplishment of the IVPR-SSADTL algorithm over other models. As a part of the future scope, the outcomes of the IVPR-SSADTL technique can be extended to hybrid DL models.

References

- [1] Xiwu Zhang, Lei Wang, and Yan Su, "Visual Place Recognition: A Survey from Deep Learning Perspective," *Pattern Recognition*, vol. 113, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Sourav Garg et al., "Semantic–Geometric Visual Place Recognition: A New Perspective for Reconciling Opposing Views," *The International Journal of Robotics Research*, vol. 41, no. 6, pp. 573-598, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Alex Junho Lee, and Ayoung Kim, "EventVLAD: Visual Place Recognition with Reconstructed Edges from Event Cameras," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2247-2252, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] M. Sakthivadivu, and P. Suresh Babu, "Analytical and Empirical Survival Study on Natural Image Compression and Classification using Machine Learning Techniques," *International Journal of Computer Trends and Technology*, vol. 70, no. 8, pp. 21-29, 2022 [[CrossRef](#)] [[Publisher Link](#)]
- [5] Mubariz Zaffar et al., "Cohog: A light-Weight, Compute-Efficient, and Training-Free Visual Place Recognition Technique for Changing Environments," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp.1835-1842, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Stephen Hausler et al., "Patch-netVLAD: Multi-scale Fusion of Locally-Global Descriptors for Place Recognition," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14141-14152, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] A. Zulaiha Parveen, and T.Senthil Kumar, "The Deep Learning Methodology for Improved Breast Cancer Diagnosis in MRI," *International Journal of Computer and Organization Trends*, vol. 11, no. 3, pp. 11-14, 2021. [[CrossRef](#)] [[Publisher Link](#)]
- [8] Luis G. Camara, and Libor Preucil, "Visual Place Recognition by Spatial Matching Of High-Level CNN Features," *Robotics and Autonomous Systems*, vol. 133, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Guanyu Huang et al., "SMCN: Simplified Mini-Column Network for Visual Place Recognition," *2nd International Conference on Computer Vision and Data Mining*, vol. 2024, no. 1, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Nurul Nadiah Abd Rahman, Marina Yusoff, and Murizah Kassim, "HOG and Haralick Feature Extractions with Machine Learning Methods for Covid-19 X-Ray Image Classification," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 8-17, 2022. [[CrossRef](#)] [[Publisher Link](#)]
- [11] Anu Mathews, N. Sejal, and K. R. Venugopal, "Text Based and Image Based Recommender Systems: Fundamental Concepts, Comprehensive Review and Future Directions," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 124-143, 2022. [[CrossRef](#)] [[Publisher Link](#)]
- [12] S. Shashikala, and G. K. Ravikumar, "Deep Studying Signature for Obstruction obscure in Copy Move Image Forgeries," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 262-270, 2022. [[CrossRef](#)] [[Publisher Link](#)]
- [13] Liang Chen, Sheng Jin, and Zhoujun Xia, "Towards a Robust Visual Place Recognition in Large-Scale vSLAM Scenarios Based on a Deep Distance Learning," *Sensors*, vol. 21, no. 1, pp. 310, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Chenyang Zhao, Runwei Ding, and Hong Liu Key, "End-to-End Visual Place Recognition Based on Deep Metric Learning and Self-Adaptively Enhanced Similarity Metric," *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 275-279, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Chanjong Park et al., "Light-Weight Visual Place Recognition Using Convolutional Neural Network for Mobile Robots," *IEEE International Conference on Consumer Electronics (ICCE)*, pp. 1-4, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] M. C. Shanker, and M. Vadivel, "Hybrid Transfer Learning of Mammogram Images for Screening of Micro-Calcifications," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 8, pp. 40-47, 2022. [[CrossRef](#)] [[Publisher Link](#)]
- [17] Jun Mao et al., "Learning to Fuse Multiscale Features for Visual Place Recognition," *IEEE Access*, vol. 7, pp. 5723-5735, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Stanislav Yamashkin et al., "Classification of Metageosystems by Ensembles of Machine Learning Models" *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 258-268, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Jianliang Zhu et al., "Visual Place Recognition in Long-Term and Large-Scale Environment Based on CNN Feature," *IEEE Intelligent Vehicles Symposium*, pp. 1679-1685, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Ling Pei et al., "IVPR: An Instant Visual Place Recognition Approach Based on Structural Lines in Manhattan World," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 4173-4187, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] K. Kavitha, and Suneetha Chittineni, "Efficient Sentimental Analysis using Hybrid Deep Transfer Learning Neural Network," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 155-165, 2022. [[CrossRef](#)] [[Publisher Link](#)]
- [22] Yanxin Wang et al., "Optimizing GIS Partial Discharge Pattern Recognition in the Ubiquitous Power Internet of things Context: A MixNet Deep Learning Model," *International Journal of Electrical Power & Energy Systems*, vol. 125, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [23] Yanlong Zhu, and Nasser Yousefi, "Optimal Parameter Identification of PEMFC Stacks Using Adaptive Sparrow Search Algorithm," *International journal of hydrogen energy*, vol. 46, no. 14, pp.9541-9552, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Sasikumar, P., and S. Sathiamoorthy, "Optimal Densely Connected Networks with Pyramid Spatial Matching Scheme for Visual Place Recognition," *Pervasive Computing and Social Networking*, vol. 475, pp. 123-137, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Mubariz Zaffar et al., "VPR-bench: An Open-Source Visual Place Recognition Evaluation Framework with Quantifiable Viewpoint and Appearance Change," *International Journal of Computer Vision*, vol. 129, no. 7, pp. 2136-2174, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]