

MNASreID: Grasshopper Optimization based Neural Architecture Search for Motorcycle re-IDentification

Medipelly Rampavan (✉ rampavan58@student.nitw.ac.in)

National Institute of Technology Warangal

Earnest Paul Ijjina

National Institute of Technology Warangal

Research Article

Keywords: Motorcycle re-IDentification, Neural Architecture Search, Grasshopper Optimization Algorithm, Auto re-IDentification

Posted Date: August 3rd, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3214105/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

MNASreID: Grasshopper Optimization based Neural Architecture Search for Motorcycle re-Identification

Medipelly Rampavan and Earnest Paul Ijjina

Computer Science and Engineering, National Institute of Technology,
Warangal, 506004, Telangana, India.

*Corresponding author(s). E-mail(s): rampavan58@student.nitw.ac.in;
Contributing authors: iep@nitw.ac.in;

Abstract

Object re-Identification (reID) is the process of matching the same object across multiple cameras. An approach for object reID is crucial for identifying objects like people, cars, motorcycles etc., in video surveillance systems. This is an active area of research in both industry and academia due to the ever-growing population and need for smart surveillance, public safety and traffic management. This work proposes an approach for automatically designing a deep convolutional neural network, named MNASreID, for motorcycle re-identification task. Most current reID methods use deep convolutional neural networks as the backbone that are manually designed, which does not have the optimum settings as the network complexity increases. MNASreID uses Neural Architecture Search (NAS) to search for the backbones of the Deep Neural Network (DNN) model. In addition, the designed search space includes parameters of the architecture and its hyperparameters to find the optimum architecture for the reID task. This work uses Grasshopper Optimization Algorithm (GOA) based NAS to find the optimal DNN model. The outcomes of experiments on two motorcycle datasets demonstrate that our method can automatically find the efficient DNN model for motorcycle reID.

Keywords: Motorcycle re-Identification, Neural Architecture Search, Grasshopper Optimization Algorithm, Auto re-Identification

1 Introduction

Object re-identification is the process of detecting and tracking objects across multiple cameras in a surveillance system, which has shown significant advancement in the last decade. The purpose of object re-identification is to match the same object/person across several cameras to ensure continuous tracking and monitoring of the object/person as they move across cameras in the surveillance system. Object re-identification is a challenging task due to variations in illumination, occlusions, changes in appearance, and other characteristics that can make it difficult to recognise and track the same object/person across different cameras or viewpoints. The object/person re-Identification(reID) is an important task in computer vision and has numerous applications in the fields of security and surveillance, traffic monitoring, retail analytics, robotics, and sports analysis.

In recent years, there has been significant progress in methods for reID task, which primarily focussed on person reID [1–3] and four-wheeler vehicle reID [4–7]. The tasks of person re-identification and vehicle re-identification differ significantly due to the difference in object characteristics, camera angle of view, and movement patterns. People are distinguished by physical characteristics such as clothes, hairstyle, and facial features, but vehicles are commonly distinguished using license plate, color, make & model information. Furthermore, people have more height variations than vehicles, which can assist in the re-identification of the most commonly used vehicle, the motorcycle. People also move slowly compared to vehicles. However, there is little research on the object in this context that the reID community ignored: motorcycles. The re-identification of motorcycles presents a unique challenge as they involve both a person and a vehicle, which have different characteristics that need to be considered. For example, motorcycle riders often wear helmets that occlude the rider’s face, and riders may wear diverse clothing that can conceal the physical characteristics of the person, needed for discrimination or recognition. In addition, motorcycles are much smaller than cars, which means that they appear in the smaller regions of the surveillance video. Despite of these challenges, there are approaches in the literature for motorcycle reID. A few studies have explored using traditional methods [8] and deep learning methods [9, 10]. However, further research is needed to develop effective techniques for motorcycle re-identification task, particularly in real-world scenarios, where there may be limited data or environmental factors that could impact the performance of the reID task.

Motorcycles are an essential mode of transportation in many developing nations of the world. They are a convenient and inexpensive means of transportation, especially for short distances and in cities. However, motorcycles can contribute to traffic problems like reckless driving and hit-and-run accidents, and some motorcycle riders engage in criminal activities, including drug trafficking, extortion, and violent crimes. Therefore, recognizing motorcyclists and monitoring their activities across wide areas is vital, which is a difficult task. Nowadays, surveillance cameras are present almost everywhere and can assist traffic monitoring personnel in performing this work efficiently. However, there are various challenges to be handled when designing an effective system, including low-resolution visual information, occlusion by other vehicles, motion blur, and variations in lighting conditions. Furthermore, as most motorcycles only have license

plates on the back, the off-the-shelf license plate recognition systems can be useful for reID only for a specific camera angle of view. As a result, to address these issues, there is a need to create a motorcycle re-identification system based on appearance that differs from traditional reID approaches based on person and vehicle characteristics. Existing methods based on traditional approaches use manually designed deep learning models for re-identification task, which typically requires expert knowledge and optimized considering specific dataset characteristics. This dependence on domain expertise and dataset characteristics can hinder the generalization capabilities of these methods across different datasets. To address these limitations, we propose a Neural Architecture Search (NAS) based approach for automatically designing a Deep Neural Network (DNN) architecture for motorcycle re-identification task.

The key contributions of this work are as follows.

1. A GOA based NAS optimization framework is proposed to automate the design of a deep neural network for motorcycle re-identification task.
2. A simple but powerful search space is designed to include the backbone architecture as well as hyperparameters for GOA based NAS to identify the optimal model.
3. Experiments were conducted to search and train different architectures for reID without pretraining on two motorcycle reID datasets which outperformed the existing state-of-the-art approaches.

The remainder of the paper is structured as follows. The related work for motorcycle re-identification is covered in Section 2. Our proposed approach is described in Section 3. Section 4 presents the experimental results and analysis. Finally, Section 5 summarizes the conclusions and future work.

2 Related work

In this section, we outline some of the existing approaches to re-identification task as well as NAS optimization techniques for various computer vision tasks.

Re-identification:

The deep learning methods have been successfully used for many computer vision tasks, and many deep learning methods have been developed in the literature for person reID and vehicle reID. The state-of-the-art methods typically use a deep neural network to extract features of the person/vehicle visual representation. For object reID, two popular approaches were proposed in the literature, which are based on ranking and classification concepts. In ranking-based methods, three images are used as input: two of them depict the same identity, while the third belongs to a different identity. The model is trained using a loss function such as Triple Loss [11], Quadruplet Loss [12], or Margin Sample Mining Loss (MSML) [13]. The objective of these loss functions is pulling samples with the same ID together, while pushing those with different IDs far apart. On the other hand, classification-related approaches [14, 15] use classification loss with a carefully designed special structure for object reID.

Luo et al. [16] introduced a baseline model for person reID that incorporates a technique called BNNeck. This approach aims to improve performance by combining ID loss and triplet loss in the training process. Huynh et al. [4] presented a model for vehicle reID that introduces the concept of multi-head attention combined with Supervised Contrastive Loss [17]. Chen et al. [18] introduced an end-to-end distance-learning deep network for vehicle reID. This network integrates global features and local features at a more detailed level, aiming to improve the performance of vehicle reID systems. Khorramshahi et al. [19] employed a key point detection approach to segment vehicle images and extract detailed information. They then utilized deep learning techniques to refine these features and effectively combine both coarse and fine features, resulting in more discriminative and efficient representations. Cheng et al. [20] introduced a multi-granularity deep feature fusion method for vehicle reID. Their approach involved designing two distinct branches to extract and fuse global features and local features, ultimately utilizing the fused feature representation for vehicle image representation. Wang et al. [21] introduced an efficient deep convolution method for re-identification tasks. Their approach involved learning deep features guided by important attributes, which collectively contributed to improving the overall re-identification performance. Zheng et al. [22] proposed a method to enhance feature representation by utilizing a multi-head architecture to extract multi-scale information. He et al. [23] investigated the utilization of vision transformers for re-identification tasks. They explored several domain adaptations and proposed a robust baseline called ViT-BoT, which served as the backbone network. To address the specific characteristics of re-identification data, the researchers introduced two modules: side information embedding and jigsaw patch module. Zhang et al. [24] explored the use of transformers for person re-identification in videos. However, they noted a challenge related to training transformers, namely the requirement for a large amount of data. Insufficient data can lead to a higher risk of overfitting.

Yuan et al. [8] have created BPreID, a large-scale dataset that focuses on bike person re-identification, which is collected in a campus environment. An approach based on handcrafted features of the bike and person parts of the image is used for bike-person re-identification. While this approach worked on the campus dataset, the challenges faced by real-time urban monitoring systems can't be handled by this approach. The MoRe dataset, which was first proposed by Figueiredo et al. [9], is the first large-scale dataset that focuses entirely on motorcycles and is captured from urban traffic surveillance cameras in real time. A strong baseline approach is developed using a deep learning model with a combination of triple loss [11, 25], quadruplet loss [12], and MSML [13], which are metric learning losses. They have also used techniques like warmup learning rate and label smoothing to increase the performance of motorcycle re-identification. J. Li et al. [10] proposed the pyramid attention mechanism to enhance the strong baseline introduced by Figueiredo [9] for capturing the essential information of the rider's images. The effectiveness of this approach was evaluated on BPreID and MoRe datasets.

Neural Architecture Search(NAS):

The field of NAS has numerous methods proposed in the literature. Auto-reID [26] differs from expert-designed neural networks [1, 9, 10] as it focuses on automating the search for neural network architectures. Neural network architectures developed by NAS-based approaches have surpassed expert-designed architectures not only in image classification [27, 28] but also in object detection [29, 30] and segmentation [31, 32]. NAS based approaches can be categorized into different groups based on the search strategy used: gradient based [33–35], evolutionary algorithms based [31, 32, 36], and reinforcement learning based [37–39] approaches. In the literature, few NAS approaches have been specifically designed for the reID task. Zhou et al. [40] introduced NAS based attention modules to learn spatial and channel attention feature maps. These feature maps were then combined to enhance the model’s feature representation capabilities without the need for pretraining. To optimize the efficiency of attention placement in the re-identification module, an Attention Search Space (ASS) was proposed. Fu et al. [41] proposed a cross-domain deep network architecture search method that specifically explores Batch Normalization (BN) layers to discover the optimal architecture. Their work emphasizes the importance of learning shared information between different modalities in existing representation learning methods, which primarily aim to improve feature extraction. Chen et al. [42] introduced a NAS approach for person reID task. This method involved searching for an optimal cell structure by making greedy decisions during the search process. Further, they introduced a triplet loss with batch hard mining [11] as the retrieval loss, aimed at enhancing the feature representation capability of the backbone network and improving the overall generalization performance. Unlike these methods, which primarily concentrate on the search for backbone architectures in DNN, our work aims to automate the search process for the optimal backbone architecture of DNN and its hyperparameters for motorcycle reID task.

3 Proposed approach

This section presents the proposed use of NAS with GOA for finding the optimal model for the re-identification of motorcycles.

Object reidentification typically involves two main phases: feature learning and reidentification. During the feature learning phase, images are used to train a classification network to extract feature vectors. In the reidentification phase, test images are split into gallery images of one camera view and probe images of another camera view, and these are then fed into the trained classification network to extract their respective feature vectors. Once these feature vectors have been extracted, a distance measure is used to match objects and identify the same object across different views. In this work, we have modelled this entire process as a NAS framework to design optimal deep neural network structure for re-identification. The flow diagram of the proposed approach is shown in Fig. 1. In this figure, the process begins with the random initialization of grasshoppers. The grasshoppers were then converted to DNN models. These DNN models undergo feature learning with training images. Following that, the rank-1 accuracy is computed by comparing feature vectors of probe and

gallery images. Subsequently, the grasshopper optimization is applied. This entire process is repeated until the maximum number of iterations is reached.

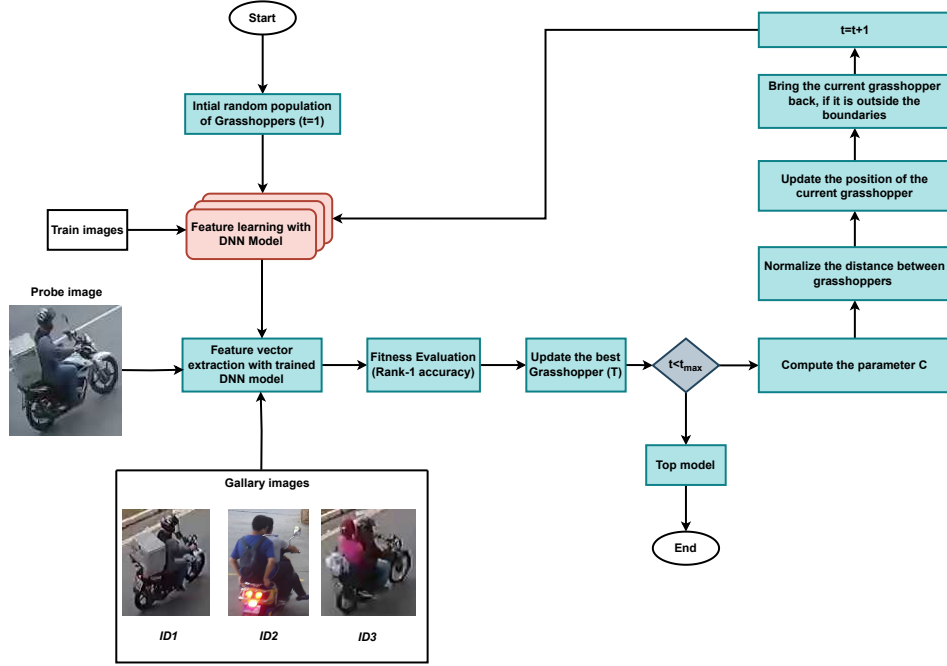


Fig. 1: The flow diagram of the proposed GOA based NAS approach for finding an optimum DNN model

A. Neural Architecture Search(NAS)

The goal of NAS is to automate the process of designing a neural network architecture using an optimization algorithm. There are two steps in using which are: A) search space design and B) search optimization algorithm to explore the search space. The search space defines the set of possible architectures to explore during the search process. This can be done by specifying the type and depth of layers, the optimizers, the activation functions, and other hyperparameters. More versatility can be obtained by adding more parameters to the search space. However, as the size of the search space grows, the cost of finding the best deep neural network architecture increases. For finding the best deep neural network architecture, even with a simple search space, one has to evaluate a significant number of candidates in the search space through trial and error. Hence, an optimization algorithm is needed to search for the neural network architecture. Thus, the search optimization algorithm finds the best architecture within the search space. This process involves selecting architectures that performed well during evaluation for the given objective function and using them to

generate new architectures. We can formulate NAS as an optimization problem, as given in Eq. 1.

$$\max_{x \in S} \mathcal{M}_{r1}(\mathcal{A}(x, w^*(x)), P, G) \quad (1)$$

$$\mathcal{A}(x, w^*(x)) = \min_w \mathcal{L}_{train} \mathcal{A}(x, w) \quad (2)$$

Here, S represents the search space of the DNN architecture denoted by a Directed Acyclic Graph (DAG), $x \in S$ is a specific path in the DAG corresponding to a DNN architecture, w represents the weights of the DNN architecture, $\mathcal{A}(x, w)$ denotes a DNN model with architecture x weight w and \mathcal{L}_{train} is the training loss. Now, Eq. 2 represents the computation of the DNN model \mathcal{A} with optimal weights $w^*(x)$ for the DNN architecture x , that minimizes the training loss \mathcal{L}_{train} . In Eq. 1, \mathcal{M}_{r1} is the rank-1 accuracy, and G, P represents a set of gallery and probe images, respectively. Eq. 1 represents the identification of DNN architecture $x \in S$, such that its rank-1 accuracy, \mathcal{M}_{r1} is maximum, corresponding to the optimum DNN model \mathcal{A} identified by Eq. 2.

B. Search space design

The performance of the re-identification task is dependent on the features extracted by the Convolutional Neural Networks (CNNs) backbone architecture. Due to this reason, we include the parameters related to the backbone in the search space. The search space of the backbone architecture consists of different kinds of blocks. Backbone architecture could be divided into several stages according to the resolution of the output features, where the stage refers to a number of blocks fed by the features with the same resolution. In this work, we proposed search space with four kinds of blocks: Resnet block [43], EfficientnetV2 block [44], Regnet block [45, 46] and Densenet block [47].

In addition to the parameters of the backbone, we have also included parameters associated with training like *Size of input*, *Pooling operation*, *Activation function*, *Optimization algorithm*, *Loss function* and *Distance metric* in our search space. The details of the search space used in this work is described in Table 1.

Table 1: The Parameters in the search space and their values considered in this work

Parameter	Range
Block type	[Regnet, EfficientnetV2, Densenet, Resnet]
Size of input	[320×320, 256×256, 224×224, 128×128]
Pooling operation	[Average, Max]
Activation function	[ReLU, Swish]
Optimization algorithm	[Adam, SGD]
Loss function	[Cross Entropy Loss, Triplet Loss, Quadruplet Loss, Margin Sample Mining Loss]
Distance metric	[Euclidean, Cityblock, Minkowski, Squeclidean]

C. Encoding Scheme

In order to train the DNN model for the given task, it is necessary to encode the search space parameters. The encoding scheme enables the conversion between the grasshopper and its corresponding DNN models.

Block type: We have selected four block types: ResNet (depths: 18, 32, 50, 101, 152), DenseNet (depths: 121, 161, 169, 201), EfficientNetV2 (variations: bo, b1, b2, b3, v2s, v2m, v2l), and RegNet (depths: rgenety320, y160, y080, y040, y016, y008, y006, y004). A variable length encoding is used to encode these blocks. ResNet uses skip connections to feed the output of one layer as the input to a layer further down the network to avoid loss of information between layers. DenseNet employs dense connections to combat the vanishing gradient problem. EfficientNetV2 uses variants of inverted residual block and scaling for better generalization. RegNet is a variant of ResNet with a regulator mechanism. These blocks are considered in this study due to their extensive usage in the literature and the diversity of the block types they employ, including basic blocks, bottleneck blocks, dense blocks, squeeze-and-excitation blocks, and stem blocks. This diversity of block types enables us to explore different DNN architectures by exploring the impact of usage of a block on performance to build high-performance DNN models for reID task.

Size of input: The size of the input image affects the performance of the DNN model that uses it. Hence, we have selected four input sizes: 320×320 , 256×256 , 224×224 , and 128×128 .

Pooling operation: The Max pooling and Average pooling are considered for pooling operation in the search space. The exploration scheme of NAS will identify the optimum pooling operation to be used in the DNN architecture.

Activation function: The activation function in a DNN model introduces non-linearity, enabling the model to learn intricate patterns and representations from the data. The choice of activation function plays a vital role in building a high-performance DNN model. In this work, for the activation function, we consider ReLU and Swish. ReLU is well-known for its simplicity, ability to induce sparsity, and reducing the likelihood of encountering the vanishing gradient problem. On the other hand, Swish is a non-monotonic smooth function. By incorporating ReLU and Swish as activation functions in our experiments, we aim to leverage the strengths of both activation functions to effectively optimize the performance of the DNN model.

Optimization algorithm: The choice of optimization algorithm plays a crucial role in reducing the miss-classification of the DNN model. In this work, we have opted for two standard optimization algorithms: Adam and SGD. These optimization algorithms are widely used and have proven to be effective in training DNN models.

Loss function: We have selected various metric learning losses, such as Cross Entropy Loss, Triplet Loss, Quadruplet Loss, and Margin Sample Mining Loss, for training. Each loss has its own significance: Cross Entropy Loss for classification, Triplet Loss for embedding similar samples close and dissimilar samples far, Quadruplet Loss for robust learning with four samples, and Margin Sample Mining Loss for informative sample selection. The use of appropriate training loss is crucial in optimizing embeddings that, intern improve the performance of the reID task. These diverse loss functions aim to enhance our model’s ability to accurately identify and distinguish individuals in reID

task.

Distance metric: The choice of distance metric is crucial in evaluating the rank-1 accuracy of the reID task. We have selected four distance metrics: Euclidean, Cityblock, Minkowski, and Squeclidean. Each metric offers unique advantages that impact the reID system’s performance. Euclidean calculates straight-line distances, Cityblock is robust to outliers and non-isotropic distributions, Minkowski is a generalization of Euclidean and Cityblock, while Squeclidean emphasizes larger differences between feature vectors.

D. Search Objective

Rank-1 accuracy is a key metric for evaluating the performance of object reidentification. The rank-1 accuracy measures the percentage of correctly identified probe images where the top-ranked match is the correct one. The rank-1 accuracy can be defined as follows:

Let P be the set of probe images, and G be the set of gallery images. For each probe image p in P , the reidentification algorithm returns a ranked list of gallery images g_1, g_2, \dots, g_n in G , where n is the total number of gallery images. The rank-1 accuracy denoted as $r1$ is then defined as:

$$r1 = \frac{1}{m} \sum_{(p \in P)} [f(p, g_1)] \quad (3)$$

where m is the total number of probe images, and $f(p, g_1)$ is an indicator function that equals 1 if the top-ranked gallery image g_1 is the correct match for the probe image p , and 0 otherwise.

Eq. 3 represents the average number of correct top-ranked matches across all probe images, expressed as a percentage of the total number of probe images.

E. Search optimization algorithm (GOA)

GOA is a recent swarm intelligence algorithm inspired by the foraging and swarming behavior of grasshoppers in nature. It is first introduced by Saremi et al. in [48]. It has been shown to be effective in solving various optimization problems, including medical image segmentation [49, 50], image enhancement [51], image fusion and feature selection [52, 53]. Its simplicity, efficiency, and robustness make it a popular optimization technique. From our review of the existing literature, we noticed that there is no existing approach utilizing GOA for the reID task. This motivated us to use GOA as the search strategy in this work for finding an optimum DNN model for motorcycle reID task.

In GOA, the grasshoppers move based on the attractiveness of the food source and the distance to other grasshoppers. The swarming behavior of grasshoppers is mathematically modelled in Eq. 4

$$Y_i = S o_i + G f_i + W a_i \quad (4)$$

Here, Y_i indicates the i^{th} grasshopper? position, So_i is the social interaction, Gf_i denotes the gravity force on the i^{th} grasshopper, and Wa_i is the wind advection. The social interaction So_i is defined as follows:

$$So_i = \sum_{j=1, j \neq i}^N s(d_{ij}) \frac{Y_j - Y_i}{d_{ij}} \quad (5)$$

where N denotes the number of grasshoppers, $d_{ij} = |Y_j - Y_i|$ denotes the Euclidean distance between the i^{th} and the j^{th} grasshopper $s(r)$ represents the social forces that can be computed from the formula in Eq. 6.

$$s(r) = F * e^{-r/L} - e^{-r} \quad (6)$$

where F and L are the attraction intensity and attraction length scale, respectively. The social interaction between grasshoppers can be defined as attraction and repulsion. The gravity force Gf_i can be computed using the formula in Eq. 7.

$$Gf_i = -g * \hat{e}_g \quad (7)$$

Where g denotes the gravitational constant and \hat{e}_g represents a unit vector toward the centre of the earth.

The wind advection Wa_i is given by the formula in Eq. 8.

$$Wa_i = -u * \hat{e}_w \quad (8)$$

where u represents the drift constant and \hat{e}_w is a unit vector in the wind direction.

According to Saremi et al. [48], Eq. 4 is modified to suit the actual conditions of the grasshoppers? movement, using the formula in Eq. 9.

$$Y_i^d = C \left(\sum_{j=1, j \neq i}^N C \frac{ub - lb}{2} s(|Y_j^d - Y_i^d|) \frac{Y_j - Y_i}{d_{ij}} \right) + T_d \quad (9)$$

Where lb , ub is defined as the lower-bound and upper-bound in the d^{th} dimension. T_d is defined as the best value of the d^{th} dimension in the target. The coefficient C is defined in Eq. 10.

$$C = C_{max} - t \frac{C_{max} - C_{min}}{t_{max}} \quad (10)$$

Algorithm 1 Proposed GOA based NAS approach

Input: train_images (I)_{train}, test_images (I)_{test} = (P,G), N .

Output: rank-1 accuracy of the best model (T)

- 1: Generate the initial random population of Grasshoppers Y_i of size N
 - 2: Initialize $C_{max}, C_{min}, t_{max}, F, L, t=1$.
 - 3: **while** $t < t_{max}$ **do**
 - 4: Encode and train the model using given (I)_{train}, search space and encoding scheme
 - 5: Evaluate the fitness (rank-1 accuracy given in Eq. 3) of each trained model, using (I)_{test} corresponding to its grasshopper (Y_i)
 - 6: The grasshopper with the highest fitness is assigned to T
 - 7: **for** $i = 1, 2, \dots, N$ **do**
 - 8: Update the value of C by using the computation given in Eq. 10
 - 9: Normalize the distance between grasshoppers
 - 10: Update the position of the current grasshopper using Eq. 9
 - 11: Bring the current grasshopper back, if it is outside the boundaries
 - 12: Update $t=t+1$
 - 13: **end for**
 - 14: **end while**
-

We first calculate the fitness of the initial random population of Grasshoppers of size N . The position of Grasshoppers is updated based on their fitness values using Eq. 9. This process will continue until the maximum number of iterations is reached. The pseudo-code to this approach is given in Algorithm 1.

To streamline the computationally expensive process of NAS, we adopted a strategy to accelerate the optimization. Firstly, we incorporated pre-termination criteria, which involved monitoring the model’s loss after 20 epochs. If the loss does not decrease significantly within this period, we terminated the training early. Otherwise, we continued training the model for a total of 80 epochs and recorded the results. Additionally, to further optimize the NAS process, we implemented a caching strategy to avoid retraining the same model if it had been generated in previous iterations of the GOA. This approach allowed us to skip redundant training of previously generated model architectures during optimization, thereby speeding up the search for the best architecture.

4 Experimental results and analysis

In this section, we present the experimental evolution of the proposed approach on two motorcycle reID datasets, i.e., MoRe [9] and BPreID [8]. We first discuss the evaluation metrics and implementation Details. The comparison of the performance of the proposed approach against the existing reID also approaches is presented in this section. Finally, an analysis of the results of the proposed approach is presented.

Datasets:

We evaluate the proposed approach on two motorcycle reID datasets, i.e., MoRe and BPreID. The MoRe dataset [9] consists of 14,141 images of 3,827 identities. As per the standard evaluation scheme, 1913 identities are used for training, and 1914 identities are used for testing. The BPreID dataset [8] contains of 18,763 motorcycle images of 940 identities. The details of these datasets is given in Table 2. The Rank-1 accuracy and the mean average precision (mAP) [18] are considered in this work for the evaluation of reID task. The mean average precision (mAP) is widely used

to evaluate the performance of the convolutional networks for reID. We compare the performance of the proposed approach against the existing approaches by using these two evaluation metrics on the two datasets.

Implementation Details:

In this work, We use the TensorFlow framework for implementing the proposed approach. The experimental study in this work is performed on a computer with an Intel Xeon(R) Silver 4110 CPU running at 2.10GHz, 64GB of RAM, and one NVIDIA GeForce RTX 2080 Ti GPU, with CUDA 11.1 in Linux platform. The Rotation, Translation, and Horizontal flipping operations are used for data augmentation. Attraction intensity F is considered as 0.5, and attraction length scale L is considered as 1.5. C_{max} is 1, C_{min} is 0.00004 and maximum number of iterations t_{max} is considered to be 20. The batch size is 24, and each batch is sampled with randomly selected 6 identities and 4 images per identity.

Table 2: Number of Images, Number identities in MoRe and BPreID datasets

Dataset	#Cameras	#IDs	#Images
Motorcycles BPreID	6	940	18,763
MoRe	10	3,827	14,141

Comparative study of existing algorithms:

We evaluate the performance of existing Motorcycle re-identification algorithms against our proposed approach on the MoRe dataset, BPreID dataset and summarize these in Table 3. Since we specifically focus on motorcycles, the number of images our approach considers differs from the rider reID method mentioned in [10]. Therefore, we cannot directly compare our results with theirs for the BPreID dataset. From Table 3, it can be observed that our approach, MNASreID, achieves a significant improvement of +6.83% and +5.76% in performance for r1 and mAP metrics, respectively, compared to the MoRe approach given in [9]. When compared to the recent state-of-the-art rider reID [10], MNASreID achieves an improvement of +1.14% and +1.24% on the r1 and mAP metrics, respectively, for the MoRe dataset. For the BPreID dataset, MNASreID outperforms the MoRe approach in [9] with a substantial margin, achieving a performance improvement of +26.82% and +29.56% on the r1 and mAP metrics, respectively. From Table 3, it can be concluded that our proposed approach surpasses the baseline model as well as the recent model, demonstrating its effectiveness on both the MoRe and BPreID datasets.

Table 3: Comparison of rank1 and mAP of the existing re-identification approaches against the top-1 model obtained from the proposed approach on MoRe dataset and BPreID dataset

Method	Dataset			
	MoRe		BPreID	
	r1 (%)	mAP (%)	r1 (%)	mAP (%)
MoRe [9]	83.41	86.38	16.94	23.08
Rider reID [10]	89.1	90.9	-	-
MNASreID	90.24	92.14	43.76	52.64

Note: The best values are highlighted in bold.

Discussion:

The number of times the NAS search space parameter value is used by the NAS exploration process for identifying the optimal DNN model for the reID task is analyzed. The results of this analysis depicting the frequency of usage of each parameter value by the proposed MNASreID is shown in Figure 2. This reveals the trends in the use of parameter values by the NAS exploration process. Among the block types, ‘Resnet’ and ‘EfficientnetV2’ are the most commonly chosen block types, surpassing other block types. For pooling operation, ‘Max’ pooling is more favoured than ‘Avg’ pooling. The ‘ReLU’ activation function emerged as the dominant choice, outperforming ‘Swish’. The ‘SGD’ optimizer is the preferred option over ‘Adam’ optimizer. Among the distance metrics, the ‘Squeclidean distance’ metric is selected most frequently compared to other distance metrics. As per the size of input, 128×128 pixels is more prevalent compared to other sizes, and for the loss function, ‘Margin Sample Mining Loss’ is the dominant choice. These observations reveal the preference for these particular parameter values during the NAS optimization process.

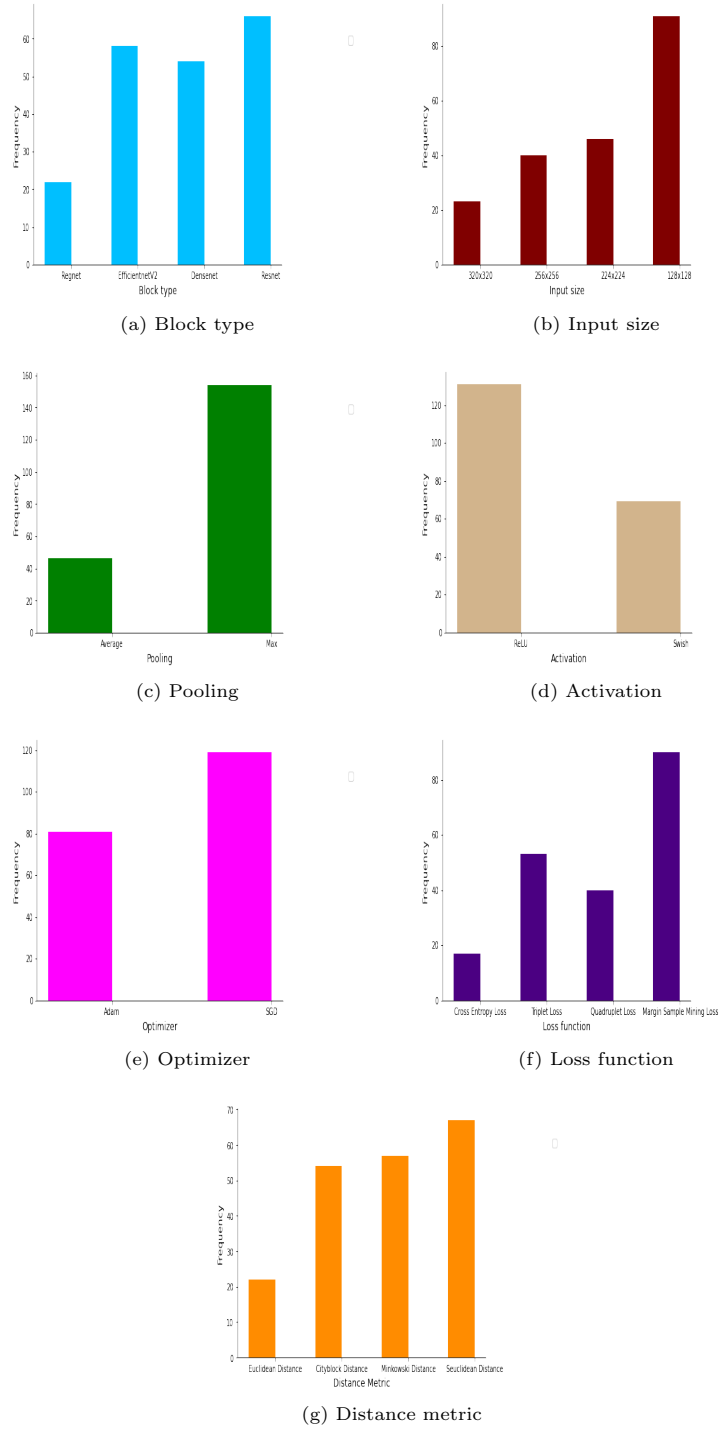


Fig. 2: The frequency of parameter values appearing in the population during 20 iterations of the MNASreID optimization process on the MoRe dataset.

The analysis of the rank-1 accuracy for each iteration is conducted by evaluating the performance of the best grasshoppers. Fig. 3 shows the trajectory details of the MNASreID rank-1 accuracy achieved by the best grasshopper after each iteration. It is observed from the figure that, as the iterations progress, the accuracy steadily increases and eventually reaches a convergence point at subsequent iterations.

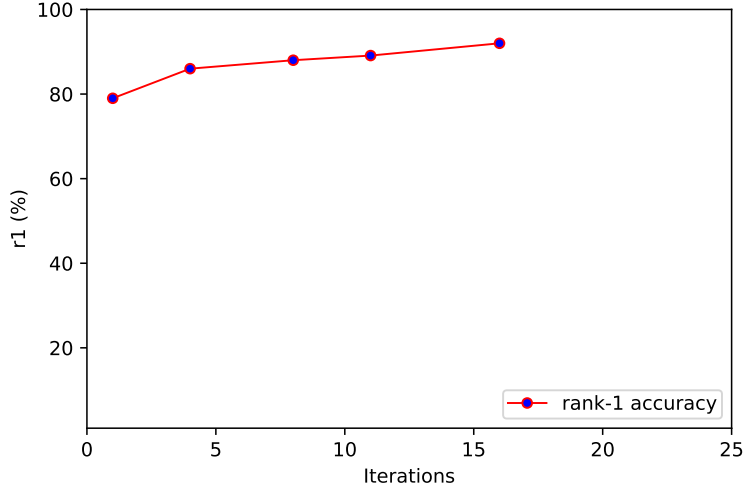


Fig. 3: Iteration wise rank-1 accuracy of the MNASreID approach for the best Grasshopper

5 Conclusion and future work

A new automated Neural Architecture Search (NAS) approach based on the Grasshopper optimization algorithm, called MNASreID, is proposed for the motorcycle re-IDentification (reID) task in this work. This approach efficiently explores both neural network architecture and hyperparameters to find the optimal deep neural network architecture. The Grasshopper optimization algorithm is utilized as the search strategy to discover the optimal model. The performance of MNASreID is compared with the existing approaches for motorcycle re-identification on the MoRe and BPreID datasets. The experimental results suggest that the proposed MNASreID outperforms the existing methods, indicating its effectiveness for the motorcycle re-identification task.

In future, this approach can be extended for other object re-identification tasks. One potential future extension is the incorporation of part-wise learning, which can further improve performance.

References

- [1] Fengliang Qi, Bo Yan, Leilei Cao, and Hongbin Wang. Stronger baseline for person re-identification. *arXiv preprint arXiv:2112.01059*, 2021.
- [2] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook. *in: IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2872–2893, 2022.
- [3] Nianchang Huang, Jianan Liu, Yunqi Miao, Qiang Zhang, and Jungong Han. Deep learning for visible-infrared cross-modality person re-identification: A comprehensive review. *in: Information Fusion*, 91:396–411, 2023.
- [4] Su V. Huynh, Nam H. Nguyen, Ngoc T. Nguyen, Vinh Tq. Nguyen, Chau Huynh, and Chuong Nguyen. A strong baseline for vehicle re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4142–4149, 2021.
- [5] Zhedong Zheng, Tao Ruan, Yunchao Wei, and Yezhou Yang. Vehiclenet: Learning robust feature representation for vehicle re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, volume 2, page 3, 2019.
- [6] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Embedding adversarial learning for vehicle re-identification. *in: IEEE Transactions on Image Processing*, 28(8):3794–3807, 2019.
- [7] Diego M. Jiménez-Bravo, Álvaro Lozano Murciego, André Sales Mendes, Héctor Sánchez San Blás, and Javier Bajo. Multi-object tracking in traffic environments: A systematic literature review. *in: Neurocomputing*, 494:43–55, 2022.
- [8] Yuan Yuan, Jian’an Zhang, and Qi Wang. Bike-person re-identification: A benchmark and a comprehensive evaluation. *in: IEEE Access*, 6:56059–56068, 2018.
- [9] Augusto Figueiredo, Johnata Brayan, Renan Oliveira Reis, Raphael Prates, and William Robson Schwartz. More: A large-scale motorcycle re-identification dataset. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 4033–4042, 2021.
- [10] Jiaze Li and Bin Liu. Rider re-identification based on pyramid attention. In *Proceedings of the Pattern Recognition and Computer Vision (PRCV)*, pages 81–93. Springer, 2022.
- [11] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

- [12] W. Chen, X. Chen, J. Zhang, and K. Huang. Beyond triplet loss: A deep quadruplet network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1320–1329. IEEE Computer Society, 2017.
- [13] Qiqi Xiao, Hao Luo, and Chi Zhang. Margin sample mining loss: A deep learning based method for person re-identification. *arXiv preprint arXiv:1710.00478*, 2017.
- [14] Lin Wu, Chunhua Shen, and Anton van den Hengel. Personnet: Person re-identification with deep convolutional neural networks. *arXiv preprint arXiv:1601.07255*, 2016.
- [15] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 907–915, 2017.
- [16] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1487–1495, 2019.
- [17] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *in: Advances in Neural Information Processing Systems*, volume 33, pages 18661–18673. Curran Associates, Inc., 2020.
- [18] Xu Chen, Haigang Sui, Jian Fang, Wenqing Feng, and Mingting Zhou. Vehicle re-identification using distance-based global and partial multi-regional feature learning. *in: IEEE Transactions on Intelligent Transportation Systems*, 22(2):1276–1286, 2021.
- [19] Pirazh Khorramshahi, Amit Kumar, Neehar Peri, Sai Saketh Rambhatla, Jun-Cheng Chen, and Rama Chellappa. A dual-path model with adaptive attention for vehicle re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6131–6140, 2019.
- [20] Yiting Cheng, Chuanfa Zhang, Kangzheng Gu, Lizhe Qi, Zhongxue Gan, and Wenqiang Zhang. Multi-scale deep feature fusion for vehicle re-identification. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1928–1932, 2020.
- [21] Huibing Wang, Jinjia Peng, Dongyan Chen, Guangqi Jiang, Tongtong Zhao, and Xianping Fu. Attribute-guided feature learning network for vehicle reidentification. *in: IEEE MultiMedia*, 27(4):112–121, 2020.

- [22] Zhedong Zheng, Minyue Jiang, Zhigang Wang, Jian Wang, Zechen Bai, Xuanmeng Zhang, Xin Yu, Xiao Tan, Yi Yang, Shilei Wen, and Errui Ding. Going beyond real data: A robust visual representation for vehicle re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2550–2558, 2020.
- [23] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14993–15002, 2021.
- [24] Tianyu Zhang, Longhui Wei, Lingxi Xie, Zijie Zhuang, Yongfei Zhang, Bo Li, and Qi Tian. Spatiotemporal transformer for video-based person re-identification. *arXiv preprint arXiv:2103.16469*, 2021.
- [25] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1335–1344, 2016.
- [26] Ruijie Quan, Xuanyi Dong, Yu Wu, Linchao Zhu, and Yi Yang. Auto-reid: Searching for a part-aware convnet for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3749–3758, 2019.
- [27] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8697–8710, 2018.
- [28] Zhaohui Yang, Yunhe Wang, Xinghao Chen, Boxin Shi, Chao Xu, Chunjing Xu, Qi Tian, and Chang Xu. Cars: Continuous evolution for efficient neural architecture search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1826–1835, 2020.
- [29] Yukang Chen, Tong Yang, Xiangyu Zhang, GAOFENG MENG, Xinyu Xiao, and Jian Sun. Detnas: Backbone search for object detection. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 32. Curran Associates, Inc., 2019.
- [30] Medipelly Rampavan and Earnest Paul Ijjina. Genetic brake-net: Deep learning based brake light detection for collision avoidance using genetic algorithm. *in: Knowledge-Based Systems*, 264:110338, 2023.
- [31] Jiahong Wei, Guijie Zhu, Zhun Fan, Jinchao Liu, Yibiao Rong, Jiajie Mo, Wenji Li, and Xinjian Chen. Genetic u-net: Automatically designed deep networks for retinal vessel segmentation using a genetic algorithm. *in: IEEE Transactions on*

- Medical Imaging*, 41(2):292–307, 2022.
- [32] Chilukamari Rajesh and Sushil Kumar. An evolutionary block based network for medical image denoising using differential evolution. *in: Applied Soft Computing*, 121:108776, 2022.
- [33] Han Cai, Ligeng Zhu, and Song Han. Proxylessnas: Direct neural architecture search on target task and hardware. *arXiv preprint arXiv:1812.00332*, 2018.
- [34] Bichen Wu, Xiaoliang Dai, Peizhao Zhang, Yanghan Wang, Fei Sun, Yiming Wu, Yuandong Tian, Peter Vajda, Yangqing Jia, and Kurt Keutzer. Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10726–10734, 2019.
- [35] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*, 2018.
- [36] Zichao Guo, Xiangyu Zhang, Haoyuan Mu, Wen Heng, Zechun Liu, Yichen Wei, and Jian Sun. Single path one-shot neural architecture search with uniform sampling. In *Proceedings of the 16th European Conference on Computer Vision (ECCV)*, pages 544–560. Springer, 2020.
- [37] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016.
- [38] Bo Chen, Golnaz Ghiasi, Hanxiao Liu, Tsung-Yi Lin, Dmitry Kalenichenko, Hartwig Adam, and Quoc V. Le. Mnasfpn: Learning latency-aware pyramid architecture for object detection on mobile devices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13604–13613, 2020.
- [39] Gabriel Bender, Hanxiao Liu, Bo Chen, Grace Chu, Shuyang Cheng, Pieter-Jan Kindermans, and Quoc V Le. Can weight sharing outperform random architecture search? an investigation with tunas. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14323–14332, 2020.
- [40] Qinqin Zhou, Bineng Zhong, Xin Liu, and Rongrong Ji. Attention-based neural architecture search for person re-identification. *in: IEEE Transactions on Neural Networks and Learning Systems*, 33(11):6627–6639, 2022.
- [41] Chaoyou Fu, Yibo Hu, Xiang Wu, Hailin Shi, Tao Mei, and Ran He. Cmnas: Cross-modality neural architecture search for visible-infrared person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11803–11812, 2021.

- [42] Shengbo Chen, Kai Jiang, Xianrui Liu, Kangkang Yang, and Zhou Lei. Tgas-reid: Efficient architecture search for person re-identification via greedy decisions with topological order. *in: Applied Intelligence*, 53(7):7343–7354, 2023.
- [43] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [44] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *Proceedings of the International Conference on Machine Learning*, volume 139, pages 10096–10106. PMLR, 2021.
- [45] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10425–10433, 2020.
- [46] Piotr Dollár, Mannat Singh, and Ross Girshick. Fast and accurate model scaling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 924–932, 2021.
- [47] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.
- [48] Shahrzad Saremi, Seyedali Mirjalili, and Andrew Lewis. Grasshopper optimisation algorithm: Theory and application. *in: Advances in Engineering Software*, 105:30–47, 2017.
- [49] Nassrallah Faris Abdukader Al Shalchi and Javad Rahebi. Human retinal optic disc detection with grasshopper optimization algorithm. *in: Multimedia Tools and Applications*, 81(17):24937–24955, 2022.
- [50] Chilukamari Rajesh, Ravichandra Sadam, and Sushil Kumar. An evolutionary u-shaped network for retinal vessel segmentation using binary teaching–learning-based optimization. *in: Biomedical Signal Processing and Control*, 83:104669, 2023.
- [51] Ashish Kumar Bhandari and Kusuma Rahul. A novel local contrast fusion-based fuzzy model for color image multilevel thresholding using grasshopper optimization. *in: Applied Soft Computing*, 81:105515, 2019.
- [52] Phu-Hung Dinh. A novel approach based on grasshopper optimization algorithm for medical image fusion. *in: Expert Systems with Applications*, 171:114576, 2021.
- [53] Haouassi Hichem, Merah Elkamel, Mehdaoui Rafik, Maarouk Toufik Mesaaoud, and Chouhal Ouahiba. A new binary grasshopper optimization algorithm for

feature selection problem. *in: Journal of King Saud University - Computer and Information Sciences*, 34(2):316–328, 2022.