ANALYSIS OF DIFFERENT DEEP LEARNING APPROACHES BASED ON DEEP NEURAL NETWORKS FOR PERSON RE-IDENTIFICATION

Adnan **RAMAKIĆ**[1], Zlatko **BUNDALO**[2], Dušanka **BUNDALO**[3]

[1] *Technical Faculty, University of Bihać, Bihać, Bosnia and Herzegovina*, [2] *Faculty of Electrical Engineering, University of Banja Luka, Banja Luka, Bosnia and Herzegovina*, [3] *Faculty of Philosophy, University of Banja Luka, Banja Luka, Bosnia and Herzegovina*

adnan.ramakic@unbi.ba, zlatbun2007@gmail.com, dusbun@gmail.com

Keywords: ***Deep Neural Network (DNN), Person Identification, Person Re-Identification, Convolutional Neural Network (CNN)***

Abstract*: In this work, different deep learning approaches based on deep neural networks for person re-identification were analyzed. Both identification and re-identification of people are frequently required in various fields of human life. Some of the most common applications are in various security systems where it is necessary to identify and track a particular person. In the case of person identification, the identity of a particular person needs to be established. In the case of re-identification, the main task is to match the identity of a particular person across different, non-overlapping cameras or even with the same camera at different times. In this work, three different deep neural networks were used for the purpose of person re-identification. Two of them were user-defined, while one of them is a pre-trained neural network adapted to work with a specific dataset. Two neural networks used were Convolutional Neural Networks (CNN). For the defined experiment, it was used own dataset with 13 subjects in gait.*

1. INTRODUCTION

Person identification and re-identification are important tasks in many aspects of human life. It is often necessary to determine the identity of a particular person, that is, to identify a particular person. This is a challenging task for which many methods have been developed in the last decades. Most of these methods are based on certain physiological or behavioral characteristics of the human body. The methods based on the mentioned characteristics are called biometric methods. Biometric methods are usually divided into two

groups: physiological and behavioral biometric methods. Accordingly, there are methods based on a person's fingerprint or palm print, iris or retina (eye elements), face, gait, voice, or signature that are used in various applications.

The methods listed above are implemented in different ways and use different features based on the above characteristics. In general, an identification system can be divided into two parts. The first part is used to create a database (or in this paper denoted as *dataset*) in which images are usually captured for each person, e.g., using an RGB camera (Red, Green, Blue) or an RGB-D device (Red, Green, Blue - Depth). The images captured are stored in the database and depend on the method used. For example, if the implemented method is based on a person's face (face recognition), images containing people's faces are captured. On the other hand, if the method is based on gait (gait recognition), images with a person walking upright are usually captured and used. In the further course of the process, the aforementioned images can be subjected to different types of processing, depending on the method implemented. Features can also be extracted from the images and used, but this also depends on the method used. Accordingly, the extracted features may be also contained in the database. The second part of the identification system is the identification part, where the new image (or extracted features) of a particular person is matched with the images or features stored in the database. The above described can be roughly represented as in *figure* 1.

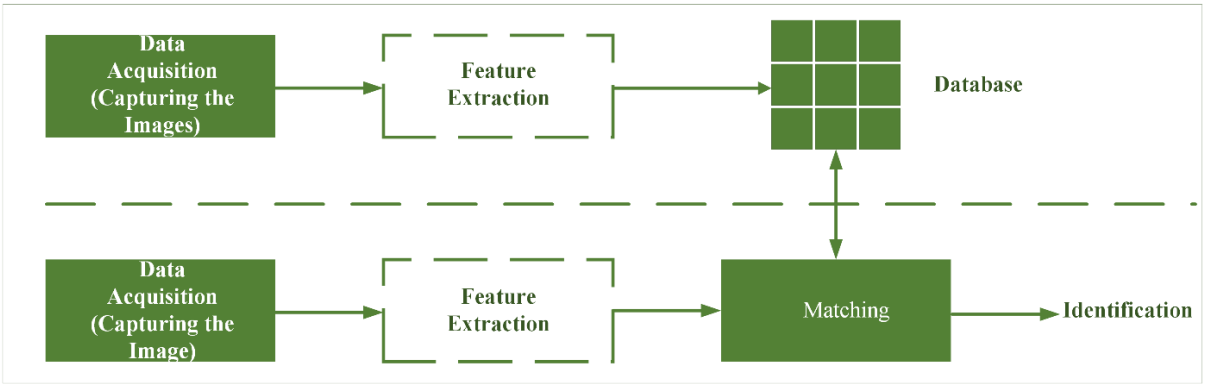

Fig. 1. An Example of Identification System with Defined Steps

While person identification involves establishing the identity of a particular person, re-identification involves matching the identity of a particular person across different, non-overlapping cameras or even with the same camera at different time frames.

Nowadays, various methods for identification and re-identification are implemented using machine learning in such a way that a model is created, trained with the data, and the created model is then used for identification or re-identification tasks. Machine learning approaches typically use classifiers such as k-Nearest Neighbors (kNN) [1], Support Vector Machines (SVM) [2], or Linear Discriminant (LD) [3]. In addition to machine learning, deep learning approaches are also used, usually using a deep neural networks (DNN).

In this work, different approaches based on deep learning have been investigated. In this context, different deep neural networks were created and analyzed. An experiment was conducted with deep neural networks using a custom dataset. Accordingly, the experiment, the experimental setup and the obtained results have been described in the following chapters.

2. THE DATASET AND EXPERIMENTAL SETUP

In this work, a custom dataset was used for the defined experiment. The dataset used contains 13 people, during a walk (in gait), recorded with different camera positions. A stereo camera was used to create the dataset and multiple video footages were available for each person. The dataset was recorded in nice weather. The video footages have high resolution. Accordingly, the extracted images also have a high resolution. A drawback of the dataset can be during extraction of silhouettes, because some people wear clothes with similar colors as background. The size of the dataset (video footages in *.avi*) is about 1,5 GB. For each of the 13 people, different images were extracted from the recorded video footages and used in the experiment.

This was implemented so that in each video containing a particular person, the person was detected and tracked in each frame. To detect upright people, *vision.PeopleDetector* in Matlab [4] [5] may be used. In this context, a bounding box was formed around the person and this part of the scene was extracted and saved as an image in RGB format. This is shown in *figure* 2. The resolution of the extracted images containing only a person (green rectangular part in *figure* 2) was 185 x 375.

Fig. 2. Detected Person in One Video Frame

This procedure was performed for each of the 13 people. In other words, in the dataset there are 13 folders (*Person1, Person2 ... Person13*) containing the images for each person. The mentioned procedure is illustrated in *figure* 3. The aforementioned extracted images are suitable for re-identification applications because the images are in RGB format and the people in all the captured images are wearing the same clothes. More specifically, said images can be used for short-term re-identification applications. For identification applications and long-term re-identification applications [6] [7] [8] [9] [10] [11], some longer-term features should be defined and used.

Various representations of silhouettes of people are often used as longer-term features, and many of the methods presented are based on them. An example of such a method is the well-known gait recognition method called Gait Energy Image (GEI) [12]. GEI is defined as an image containing silhouettes of a person over a gait cycle that are normalized, aligned, and temporally averaged. Some examples of silhouette images and GEI images from the Casia Dataset B [13] [14] [15] are shown in *figure 4* and *figure 5*.

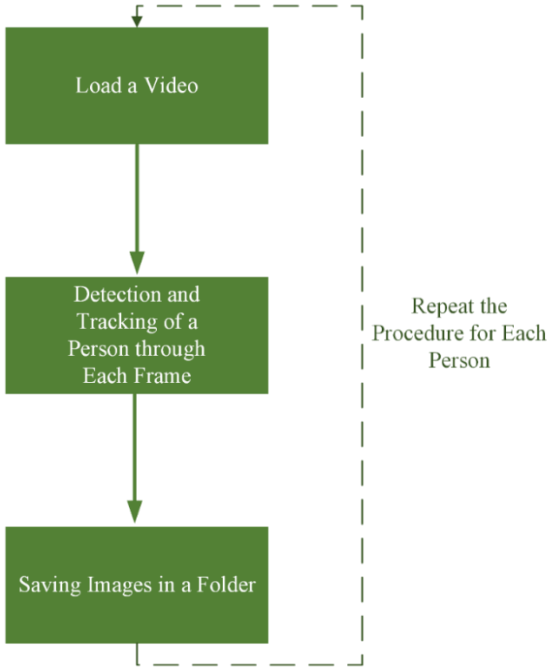

Fig. 3. The Procedure for Dataset Creation

Once the dataset was created, three different deep neural networks were created and used for the experiment. The main idea was to create two different neural networks, with the first neural network having a feature input layer as the first layer. The features from the images would first be extracted and stored in a table and then used with the neural network created. In the case of the second neural network, the first layer is intended to be an image input layer. In this case, only images should be loaded to be used with the created neural network without prior feature extraction. In the third case, a pre-trained neural network was defined for use.

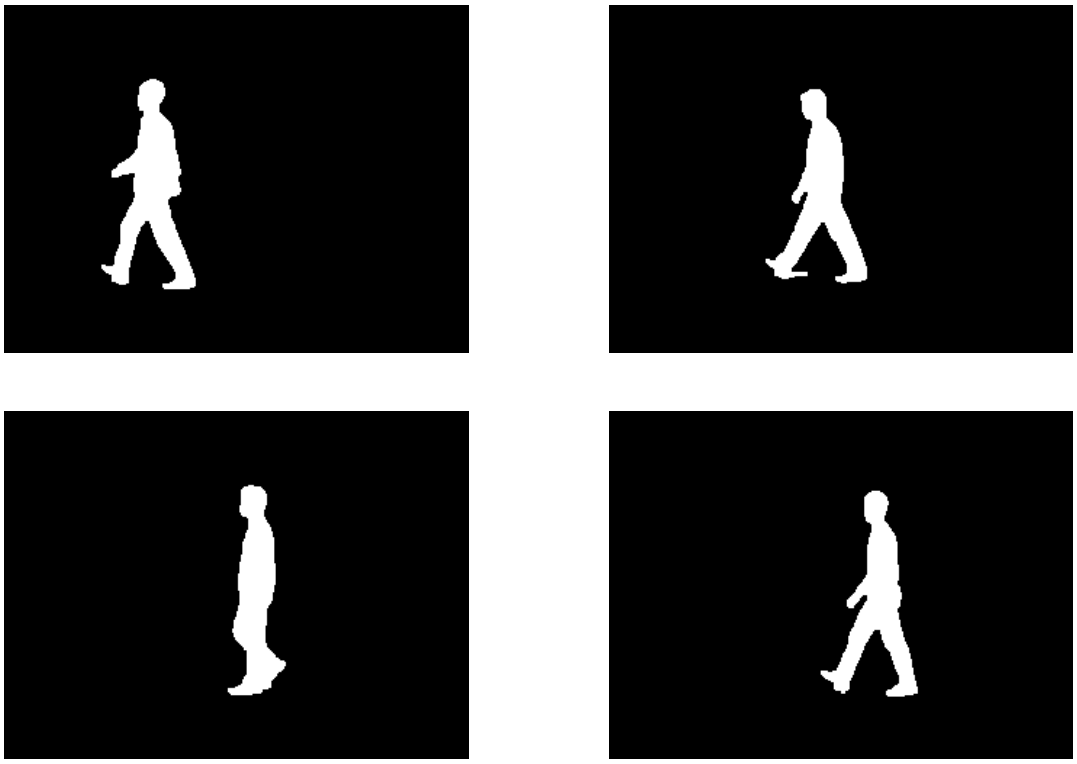

Fig. 4. Examples of Silhouette Images from Casia Dataset B [13] [14] [15]

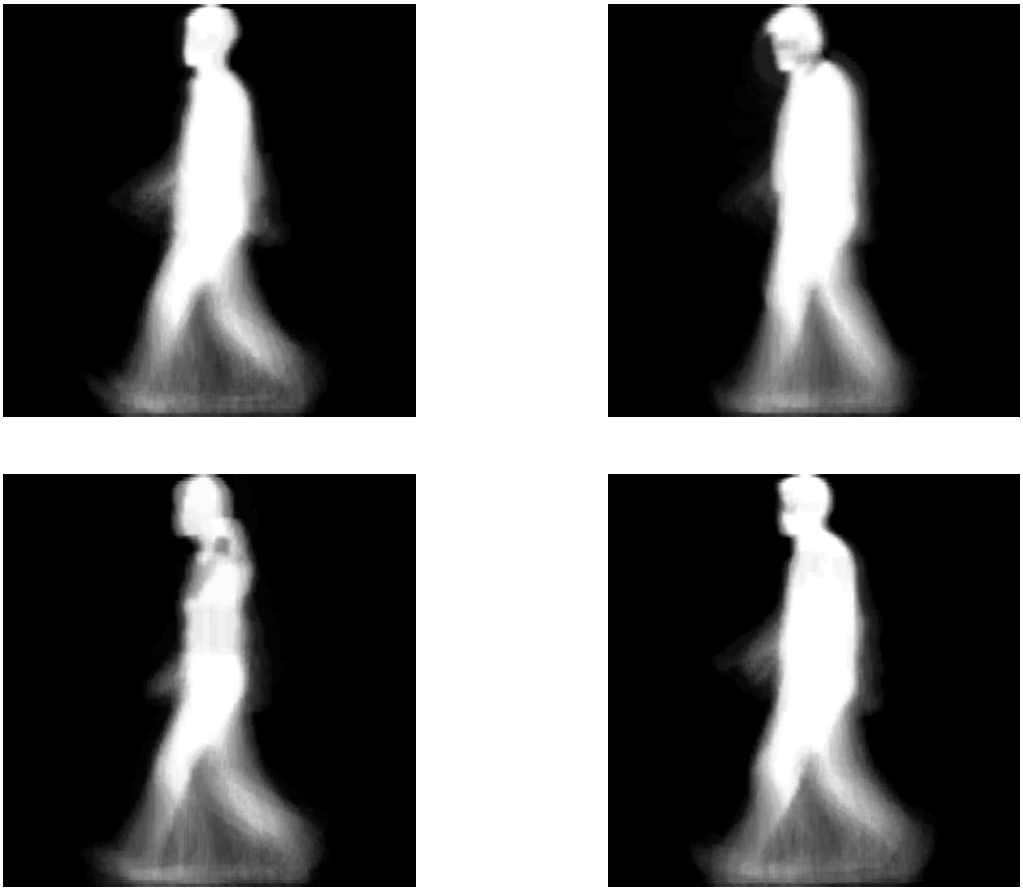

Fig. 5. Examples of GEI Images from Casia Dataset B [13] [14] [15]

Matlab was used for the mentioned experiment and for the creation of the dataset. For the creation of the dataset, a separate program was created for this purpose. It should be noted that Python, TensorFlow and Keras have also been analyzed, tested and used for the same purpose.

The first neural network (hereinafter marked with DNNf) uses extracted features from the images of the dataset. This was done using a *bag of visual words* (*bagOfFeatures* in Matlab, with defined parameters *VocabularySize - 500* and *PointSelection* as *Detector*) [16] [17], where the visual vocabulary was created by default from Speeded-Up Robust Features (SURF) [18]. The mentioned features were stored in a table. DNNf consists of seven layers, the first layer being the feature input layer (*featureInputLayer*). Different numbers and types of layers were tested, but with the mentioned seven layers and defined parameters, satisfactory results were obtained.

The seven defined layers are:

1. *featureInputLayer*
2. *fullyConnectedLayer*
3. *batchNormalizationLayer*
4. *reluLayer*
5. *fullyConnectedLayer*
6. *softmaxLayer*
7. *classificationLayer.*

The extracted features stored in the table were divided into a training and a testing part, with 70 percent used for training and 30 percent for testing. Other training options for the DNNf include 30 epochs, a learning rate of 0,001 and the Adaptive Moment Estimation Optimizer (Adam) [19] was used. The best results were obtained with the above settings.

The second neural network (hereinafter marked with DNNi) is a Convolutional Neural Network (CNN). The DNNi uses the images without prior feature extraction. The images were only loaded as the first layer is the image input layer (*imageInputLayer*). Also, in this case, different numbers and types of layers were analyzed and tested. With defined eight layers and defined parameters, satisfactory results were obtained.

DNNi consists of following eight layers:

1. *imageInputLayer*
2. *convolution2dLayer*
3. *batchNormalizationLayer*
4. *reluLayer*
5. *maxPooling2dLayer*
6. *fullyConnectedLayer*
7. *softmaxLayer*
8. *classificationLayer.*

The images used were also split in the ratio of 70 percent for training and 30 percent

for testing. Other options defined for DNNi are 30 epochs, a learning rate of 0,001 and Stochastic Gradient Descent with Momentum (SGDM) [20] was used. Also, the best results were obtained with the above defined settings.

In addition to the deep neural networks created and described above (DNNf and DNNi), a pre-trained neural network was also used. The pre-trained neural network used is *GoogLeNet* [21] [22], a convolutional neural network. *GoogLeNet* [21] [22] was used and adopted to work with the dataset described above. This was done in such a way that two layers were replaced and adapted to the dataset. The layers mentioned are *fullyConnectedLayer* and *classificationLayer*. In the *fullyConnectedLayer*, the *OutputSize* parameter was set to *13*, which corresponds to the number of subjects in the dataset. The images used were split in the ratio of 70 percent for training and 30 percent for testing. Other training options for the *GoogLeNet* include 30 epochs, a learning rate of 0,001 and the SGDM was used, as in case DNNi.

3. RESULTS AND DISCUSSION

With the settings defined above and the neural networks described, the following results were obtained using the dataset described. In the case of DNNf, the accuracy was 90,8%. When DNNi was used, the accuracy was 91,7% which is higher compared to DNNf. In the case of *GoogLeNet*, pre-trained neural network, the accuracy was 99,4%. The results presented above are shown in table 1 and *figure 6*.

Table 1. The Obtained Results with Defined Settings and Used Dataset

Deep Neural Network Used	**Accuracy**
DNNf	90,8%
DNNi	91,7%
GoogLeNet	99,4%

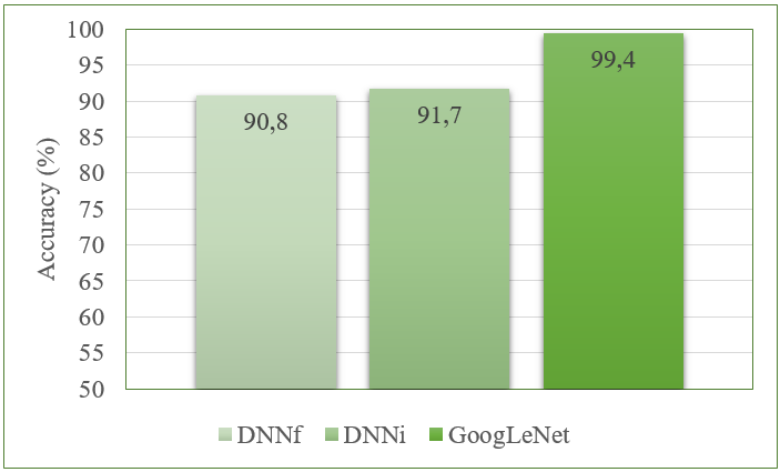

Fig. 6. The Obtained Results for the Deep Neural Networks Used

As can be seen from table 1 and *figure* 6, the pre-trained deep neural network *GoogLeNet* achieved the best overall result. This was to be expected, since *GoogLeNet* is a more complex neural network that has been pre-trained and validated on a large number of different images. With a relatively simple adaptation to use a custom dataset, the aforementioned deep neural network can easily be used for this type of application.

The created deep neural network, called DNNi, had the second best results and slightly better results compared to another created deep neural network (DNNf) that uses extracted features. On the other hand, DNNf has a much shorter training time. Moreover, the results of DNNf and DNNi can be improved by additional optimizations and adding some extra layers.

It should be noted that it is easier to work with deep neural networks such as DNNi and *GoogLeNet* compared to DNNf. The two deep neural networks mentioned, DNNi and *GoogLeNet,* have an image input layer as the first layer. This means that no explicit feature extraction is required in this case. For use with DNNi and *GoogLeNet*, only images containing people in gait should be loaded. In the case of DNNf, explicit feature extraction is required because the first layer is a feature input layer.

4. CONCLUSION

In this work, different deep learning approaches were analyzed. Person identification and re-identification applications are important in many areas of human life. In person identification, the identity of a particular person needs to be established. In person re-identification the main task is to match the identity of a particular person across different, non-overlapping cameras or with the same camera at different times. For example, in different security systems, some kind of identification or re-identification is often required.

Various methods have been developed for the aforementioned identification and re-identification applications. The mentioned methods are usually based on various physiological or behavioral characteristics of a person. Nowadays, identification and re-identification methods are usually implemented using various machine learning and deep learning approaches.

In this work, three different approaches based on deep neural networks were analyzed. For this purpose, two deep neural networks were created, while the third deep neural network used is pre-trained and adapted for use with a specific dataset. The first deep neural network created (DNNf) has a feature input layer as its first layer and uses extracted features from the images of the dataset. The second deep neural network (DNNi) is a convolutional neural network (CNN) and has as its first layer an image input layer into which only images to be used with said deep neural network are loaded. The third deep neural network used is the pre-trained neural network *GoogLeNet*.

The experiment with the defined deep neural networks was performed and the results

were presented. For this purpose, a custom dataset containing 13 people in gait was used. The best overall result had the pre-trained deep neural network *GoogLeNet*.

In future research, it is planned to analyze and use a larger dataset containing a larger number of people in gait. In addition, it is also interesting to study different points of view and conditions where people wearing similar clothing. Accordingly, other deep neural network architectures will also be analyzed and studied.

REFERENCES

[1] L. E. Peterson, *K-Nearest Neighbor*, Scholarpedia, 4(2), 1883, 2009.

[2] S. R. Gunn, *Support Vector Machines for Classification and Regression*, ISIS Technical Report, 14(1), 5-16, 1998.

[3] R. A. Fisher, *The Use of Multiple Measurements in Taxonomic Problems*, Annals of Eugenics 7(2), 179-188, 1936.

[4] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 886-893, IEEE, 2005.

[5] Official Web Page of Mathworks, *vision.PeopleDetector (Documentation)*, Link:*https://www.mathworks.com/help/vision/ref/vision.peopledetector-system-object.html* [Accessed 15/5/2023]

[6] K. Lenac, D. Sušanj, A. Ramakić and D. Pinčić, *Extending Appearance Based Gait Recognition with Depth Data*, Applied Sciences, 9(24), 5529, MDPI, 2019.

[7] A. Ramakić and Z. Bundalo, *Gait Recognition as an Approach for People Identification*, In: International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies, 717-726, Springer, 2023.

[8] A. Ramakić, Z. Bundalo and D. Bundalo, *An Example of Solution for Data Preparation Required for Some Purposes of People Identification or Re-Identification*, Journal of Circuits, Systems and Computers, *https://doi.org/10.1142/S0218126623501645*, World Scientific, 2022.

[9] A. Ramakić, Z. Bundalo and D. Bundalo, *A Method for Human Gait Recognition from Video Streams Using Silhouette, Height and Step Length*, Journal of Circuits, Systems and Computers, 29(7), 2050101, World Scientific, 2020.

[10] A. Ramakić, Z. Bundalo and Ž. Vidović, *Feature Extraction for Person Gait Recognition Applications*, Facta Universitatis, Series: Electronics and Energetics, 34(4), 557-567, 2021.

[11] A. Ramakić, D. Sušanj, K. Lenac and Z. Budalo, *Depth-based Real-time Gait Recognition*, Journal of Circuits, Systems and Computers, 29(16), 2050266, World Scientific, 2020.

[12] J. Han and B. Bhanu, *Individual Recognition Using Gait Energy Image*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2), 316-322, IEEE, 2005.

[13] S. Yu, D. Tan, and T. Tan, *A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition*, In: 18th International Conference on Pattern Recognition (ICPR), 441-444, IEEE, 2006.

[14] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, *Robust View Transformation Model for Gait Recognition*, In: 18th International Conference on Image Processing, 2073-2076, IEEE, 2011.

[15] Official Web Page of the Institute of Automation, Chinese Academy of Sciences, Link: *http://www.cbsr.ia.ac.cn/english/Gait\%20Databases.asp* [Accessed 15/5/2023]

[16] G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray, *Visual Categorization with Bags of Keypoints*, In: Workshop on Statistical Learning in Computer Vision (ECCV), 1-2, 2004.

[17] D. Nister and H. Stewenius, *Scalable Recognition with a Vocabulary Tree*, In: Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2161-2168, IEEE, 2006.

[18] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, *Speeded-Up Robust Features (SURF)*, Computer Vision and Image Understanding, 110(3), 346-359, Elsevier, 2008.

[19] D.P Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv preprint arXiv: 1412.6980, 2014.

[20] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT PRESS, 2012.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich, *Going Deeper With Convolutions*, IEEE Conference on Computer Vision and Pattern Recognition, 1-9, IEEE, 2015.

[22] Official Web Page of Mathworks, GoogLeNet, Link:*https://www.mathworks.com/help/deeplearning/ref/googlenet.html* [Accessed 15/5/2023]