AN EFFICIENT METHOD FOR ANIMAL CLASSIFICATION USING DEEP LEARNING

Komarasamy G[1], Mapakshi Manish, Devraj Dhar, Vadde Dheemanth, Manash Bhattacharjee[2]
[1]*Associate Professor*, [2]*UG students*,
Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jain University, Ramanagara District, Karnataka 562112, India

Abstract: Many existing methods such as CNN are used to show promising results for human recognition cannot properly represent the diversity of animal classes with complex intra-class variability and inter-class similarity. There are several kinds of approaches for solving this problem with each one having its advantages and disadvantages. Proposed approach constructs complex features which represents and discriminates sample images better but creating such a feature is complicated and it is problem dependent. Instead of using complex representation, the information is used for decision making.

This paper, shows that proposed method can be applied to other classification and recognition problems. It is believed that performance rate can be increased when the information of two different types of classification system are consolidated. For this purpose, at classifier systems based on features extracted from CNN is constructed.

Keywords: Animal classes, CNN, Decision making.

I. INTRODUCTION

Animal recognition and classification is an important area which has not been discussed rapidly. Animal classification which relies on the problem of distinguishing images of different animal species is an easy task for humans, but evidence suggests that even in simple cases like cats and dogs, it is difficult to distinguish them automatically. Animals have flexible structure that could self-mask and usually they appear in complex scene. Also, as all objects, they may appear under different illumination conditions, viewpoints and scales. There are attempts to apply recognition methods on images of animals but the specific problem of animal categorization has recently attracted limited interest.

Many existing methods showing promising results for human face recognition cannot properly represent the diversity of animal classes with complex intraclass variability and interclass similarity. There are several kinds of approaches for solving this problem with each one having its advantages and disadvantages. The first approach constructs complex features which represents and discriminates sample images better but creating such a feature is complicated and it is problem dependent. The second approach combines the extracted features from different methods and concatenates them to build a more powerful feature vector. Increasing the size of feature space causes increased problem computation cost. Instead of using complex representation, the information is consolidated from different classifiers and a decision is made according to it. This method is known as score-level fusion.

Observing wild animals in their natural environments is an important task in ecology. The fast growth of human population and the endless pursuit of economic development are making over exploitation of natural resources, causing rapid and novel and substantial changes to Earth's ecosystems. An increasing area of land surface has been transformed by human action, altering wildlife population, habitat and behavior. More seriously, many wild species on Earth have been driven to extinction and many species are introduced into new areas where they can disturb both natural and human systems. Monitoring wild animals, therefore, is essential as it provides researchers evidences to help make conservation and management decisions to maintain diverse, balanced and sustainable ecosystems in the face of those changes, Various modern technologies have been developed for wild animal monitoring, including radio tracking, wireless sensor network tracking, satellite and global positioning system (GPS) tracking, and monitoring by motion sensitive camera traps. Motion-triggered remote cameras or "camera traps" are an increasingly popular tool for wildlife monitoring, due to their novel features equipped, wider commercial availability, and the ease of deployment and operation. For instance, a typical covert camera model is capable of not only capturing high definition images in both day and night, but also collecting information of time, temperature and moon phase integrated in image data. In addition, flexible camera settings allow tracking animals secretly and continuously.

Once being fully charged, a camera can snap thousands of consecutive images, providing a large volume of data. These specifications make camera traps a powerful tool for ecologists as they can document every aspect of wildlife. Visual data, if can be captured is a rich source of information that provide scientists evidences to answer ecology related scientific questions such as: what are the spatial distributions of rare animals, which species are being threatened and need protection such as bandicoot, which cohort of pest species such as fox and rabbit, need to be controlled; they are examples of key questions to understand animal population, ecological relationships and population dynamics. To this end, a recently widely used approach by ecologists is to set up several camera traps in the wild to collect image data of wild animals in their natural habitat.

The overwhelming amounts of data from camera traps highlight the need for image processing automation. From data analysis and machine learning point of views, there are some immediate techniques to make wildlife identification automated classifier with manual object bounding on hand-crafted features, convolutional neural network (CNN) model with automatic object detection, or fine-tuning CNN models inheriting model weights pretrained on a very large scale dataset such as the ImageNet. These approaches addressed the problem of automating wildlife and demonstrated promisingly empirical results. However, two primary challenges which holds back the feasibility of an automated wildlife monitoring application in practice, are still remaining. The first obstacle is that, to obtain applicable image classification accuracy, an enormous amount of manual preprocessing is still required to input images for detecting and bounding animal objects. The second limitation is poor performance obtained by wildlife monitoring system, in spite of complete automation it requires much more improvements for practical application.

In [7] the authors Tilo Burghardt, Janko Calic presented an algorithm for detecting and tracking animal faces from the videos which use Haar-features and adaboost classifiers, this was tested on lion faces. In [9], authors Alexander Loos and Andrea Ernst presented and evaluated a unified image-based automated face identification of captive and free living chimpanzees in uncontrolled free environments. In [10], authors Dalila Cherifi, Fateh, Yacini and Nait-Ali have experimented on increasing the accuracy of the face recognition. This work is done on analysis and performance based on global methods like PCA, FLD and local methods like SIFT and LBP. In [12], authors Zan Gao, Hai Zhen Xuan and Hua Zhang have worked on human face recognition using several angles of facial images. Their paper proves that having the images from several facial angles and using those images for identification, recognition and even in feature extraction produces more accuracy compared to using image consisting of just one angle. In [11], authors Bhalke D.G and Archana P have worked on multimodal biometrics which produces enhanced accuracy and least error rates like FAR and FRR. Based on the present situation regarding the quality of security and accuracy of biometrics, this can become a revolution in e-commerce industry which suffers a lot of security vulnerability these days in warehouse and as well as in the logistics department.

2. METHOD SPECIFICATION

A system architecture diagram is used to show the relationship between different components. Usually they are created for systems which include both hardware and software and these are represented using diagram to show the interaction between them.

A novel method is proposed for animal face classification based on one of the popular convolutional neural network (CNN) features. We are using CNN, a method which can automatically extract features, learn and classify them. CNNs relatively does little pre-processing compared to other image classification algorithms. This means that the network learns the filters which were hand-engineered in previous traditional algorithms. This independence from prior knowledge and human effort in feature design is a major advantage in image classification.

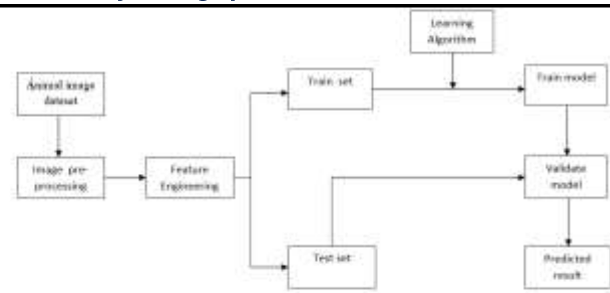

Fig: 2.1 – Model Architecture

As shown in the above Fig 2.1, the following figure demonstrates the architecture of this model.

3. ENCRYPTION METHOD

This model is built by using Convolutional neural network. Convolutional neural networks (CNN) are a special architecture of artificial neural networks, proposed by Yann LeCun in 1988. CNN uses some of the features of the visual cortex. Now that the pre-processing is done, neural network can be implemented. It will produce 3 convolution layers with 2 x 2 max-pooling.

Max-pooling is a technique used to reduce the dimensions of an image by taking the maximum pixel value of a grid. This also helps reduce over fitting and makes the model more generic. Next, 2 fully connected layers is added. Since the input of fully connected layers should be two dimensional and the output of convolution layer should be four dimensional. So, a flattening layer is needed between them. At the very end of the fully connected layers is a layer called softmax layer.

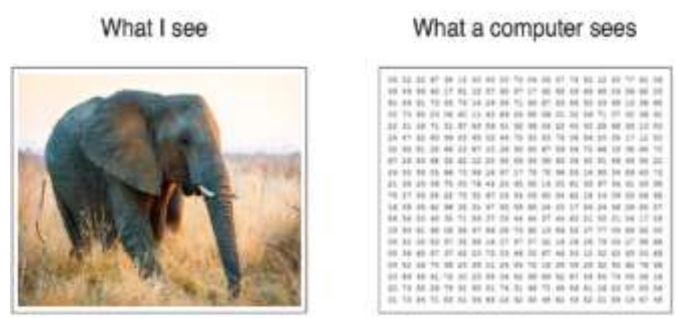

Fig: 3.1 – What A Computer Sees
Source: medium.com

Instead of an image, the computer sees an array of pixels. For example, if image size is 300x300, In this case, the size of the array would be 300x300x3. Where 300 is width, next 300 is height and 3 is RGB values. The computer is assigned a value from 0 to 255 to each of these numbers. This value describes the intensity of the pixel at each point.
To solve this problem, the computer looks for the characteristics of the base level. In human understanding such characteristics are seen as, for example the trunk or large ears. For the computer, these characteristics are seen as, boundaries or curvatures. And then through the groups of convolutional layers, computer constructs more abstract concepts.
In the above fig 3.1, the two images show us what a human eye and computer see when they are presented with an image.

4. DATA FLOW DIAGRAM

Data flow diagram (DFD) is a graphical representation of the flow of data through an information system, modeling its process aspects describes the first stage process of this project. An animal dataset is passed as an input and the system will perform the preprocess and extract the important features. Fig 4.1 describes first level of overall process.

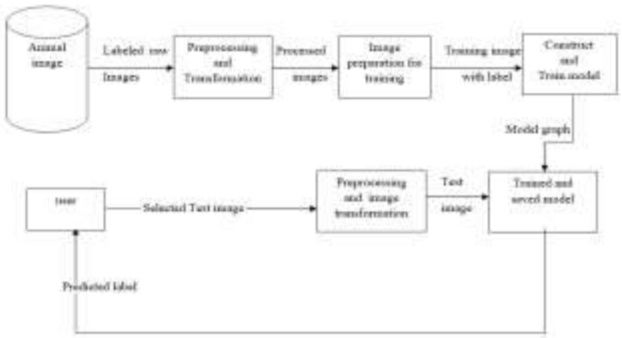

Fig: 4.1 – Overall Process

The dataset is passed and the system will do preprocessing and extract features and transforms them into content readable by the system. A training model is then constructed by using the training images with labels. A model graph is plotted and the model is saved by the system. Now for verification, the user evaluates the trained model with the test model to check the percentage of accuracy. This can be repeated with minor changes as per the requirements until the expected accuracy is achieved.

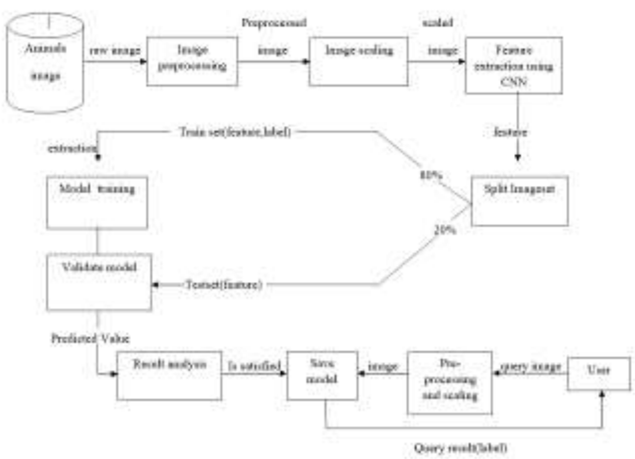

Fig: 4.2 – Stepwise Process

The above fig: 4.2 describes the final stage of the process. extracted features from level 1 are being passed and trained data as a input the system will classify the given animal is matched or not using CNN model.Fig:4.2 describes the internal working of model

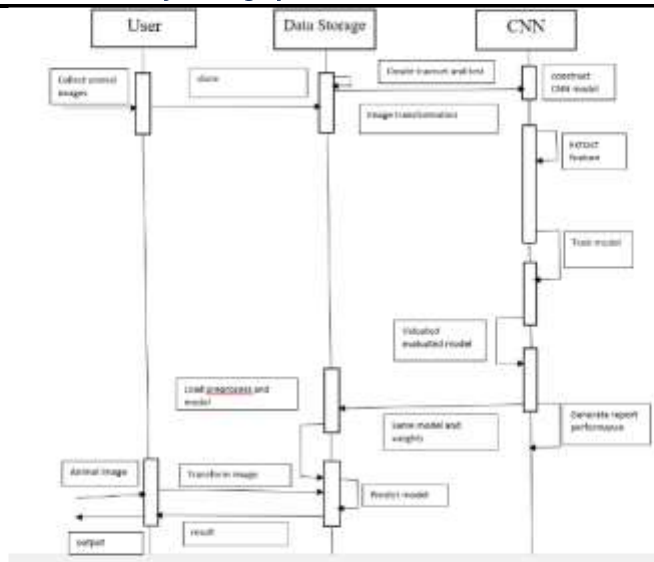

Fig: 4.3 – Sequence Diagram

The above figure 4.3 is the sequence diagram of the process. A sequence diagram simply represents interaction between objects in a sequential order, which is the order in which these interactions take place. The terms, event diagrams or event scenarios can be used for referring a sequence diagram. Sequence diagrams describe how and what order the objects in a system function.

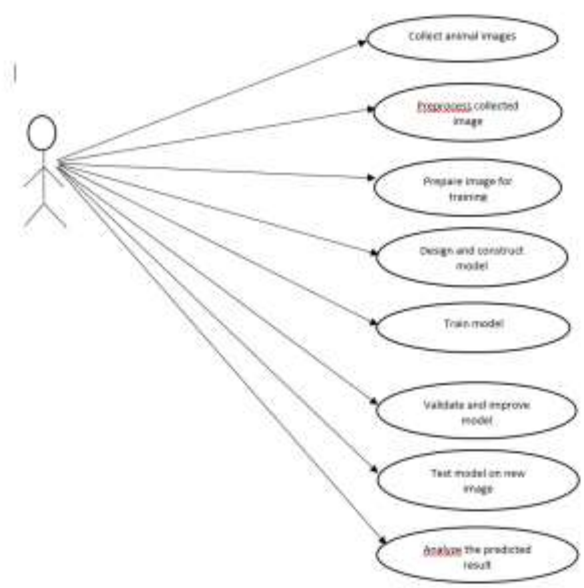

Fig: 4.4 – Use Case Diagram

The above figure 4.4 is the use-case diagram of the process. It will cover the details explanation of methodology that is being used to make this project complete and working well. Many findings from this field mainly generated into journal for others to take advantages and improve as upcoming studies. The method is used to achieve the objective of the project that will accomplish a perfect result. The below figure 4.5, is the SDLC life cycle of the process. In order to evaluate this project, the methodology based on System Development Life Cycle (SDLC), generally three major steps as shown in the below figure 4.6, which is planning, implementing and analysis.

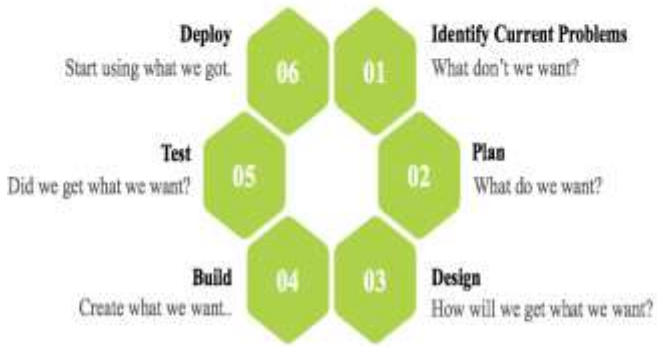

Fig: 4.5 – System Development Life Cycle

Source: workshop.readthedocs.io

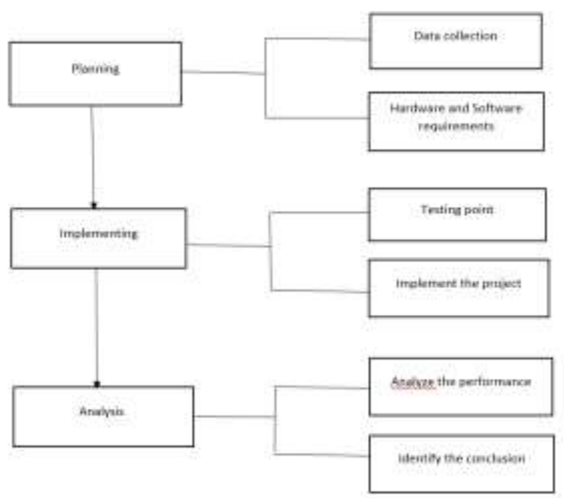

Fig: 4.6 – Major Steps Followed

Machine learning needs two things to work: data and models. When acquiring the data, It must be made sure that there are enough features (aspect of data that can help for a prediction, like the surface of the house to predict its price) populated to train learning model correctly. In general, the more the data, the more the ability of making enough rows.

The primary data collected from the online sources remains in the raw form of statements, digits and qualitative terms. The raw data contains error, omissions and inconsistencies. It requires corrections after careful scrutinizing the completed questionnaires. The following steps are involved in processing the primary data. A huge volume of raw data is collected through field survey needs to be grouped for similar details of individual responses.

Data Preprocessing is a technique that is used to convert raw data into a clean data set. In other words, whenever the data is gathered from different sources it is congregated in a raw format which is not convenient for the analysis.

Therefore, certain steps are executed to convert the data into a small and clean data set. This technique is performed before the execution of Iterative Analysis. These set of steps is known as Data Preprocessing. It includes –

- Data Cleaning
- Data Integration
- Data Transformation
- Data Reduction

Data Preprocessing is necessary because of the presence of unformatted real-world data. Mostly real-world data is composed of –

- **Inaccurate data (missing data) -** There are many reasons for missing data such as data is not continuously collected, a mistake in data entry, technical problems with biometrics and much more.
- **The presence of noisy data (erroneous data and outliers) -** The reasons for the existence of noisy data could be a technological problem of gadget that gathers data, a human mistake during data entry and much more.

5. CONCLUSION

In this work, a business intelligent model has been developed, to classify different animals, based on a specific business structure deal with Animal classification using a suitable deep learning technique. This model was evaluated by a scientific approach to measure accuracy. Convolutional Neural Network (CNN) is used to build the model.

6. EXPERIMENT AND RESULTS

Several experiments have been carried out in order to show the performance of the proposed method and the other state-of-the-art methods. In the following subsections, the dataset description and experimental setup and results are presented.

6.1 Dataset

The CIFAR-10 dataset is used in this project. It is an online google dataset which is readily available.CIFAR-10 dataset consists of 32X32 color images with 10 classes. Each class consists of 6000 images. The dataset consists of even few low resolution images in it, which helps in making this experiment more efficient in classifying low resolution images and other images with lesser lighting and low quality as well.

6.2 Results

In all of the following experiments, the CIFAR-10 dataset is taken into a training set and test set is taken as a collection of images of the random animals with 32X32size resolution.in jpg format.

6.3 Model construction

The model is built by using Convolutional neural network. Convolutional neural networks (CNN) are a special architecture of artificial neural networks, proposed by Yann LeCun in 1988. CNN uses some features of the visual cortex. Now that the pre-processing phase is done, neural network can be implemented now. It is going to have 3 convolution layers with 2 x 2 max-pooling.

Max-pooling is a technique used to reduce the dimensions of an image by taking the maximum pixel value of a grid. This also helps reduce over fitting and makes the model more generic. Next, add 2 fully connected layers. Since the input of fully connected layers should be two dimensional, and the output of convolution layer is four dimensional, a flattening layer is required between them. At the very end of the fully connected layers is a softmax layer.

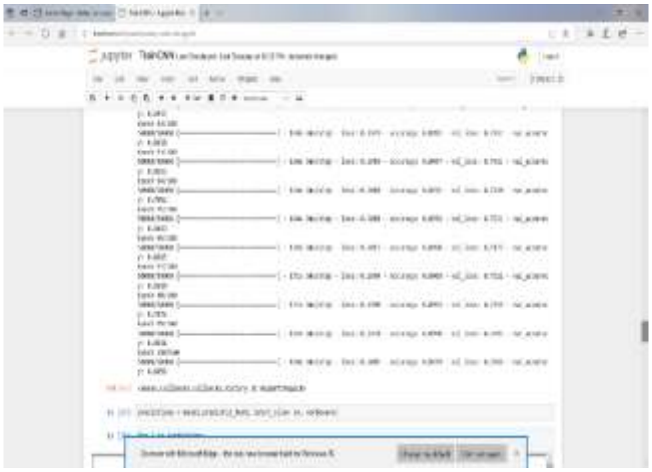

Fig: 6.1 – Execution of Code

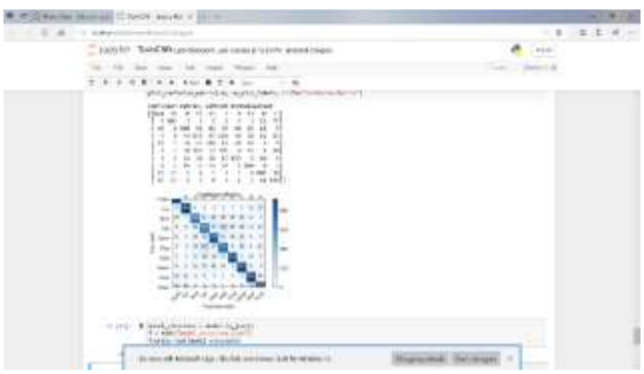

Fig: 6.2 – Constructing, Testing and Evaluating Experiment

6.5 Model testing and evaluation

Once the model has been trained, **model testing** can be carried out**.** During this phase a test set of data is loaded. This data set has never been seen by the model and therefore its true accuracy will be verified. Finally, the saved model is ready to be used in the real world. The name of this phase is called **model evaluation**. This means that the model can be used to evaluate new data. In the below figure 6.3, a tiny code has been used to show that the testing is done appropriately using the images in the desktop.

Fig 6.3- conclusion and result

6.6 Conclusion

This system comes under deep learning which is advanced technique at present. CNN is more suitable for image processing especially in image classification. It concludes the experimental result what which is obtained from developed system is comparatively more accurate than many procedures followed before. The above figure 6.1 shows us the accuracy rate. The above figure 6.2 shows the the convolution matrix where the image evaluation is done.

6.7 Future work

Image recognition and prediction is a wide concept which is very deep. Deep learning concepts can be used in this subject to get efficient results. Concept like Convolutional neural networks(CNN) can be used, which is effective in obtaining results for image identification and classification.

7. REFERENCES

[1] W. Tang, J. Ren, and Y. Zhang, "Enabling trusted and privacy-preserving healthcare services in social media health networks," IEEE Transactions on Multimedia, vol. 21, no. 3, pp. 579–590, 2019.

[2] X. Du and H. H. Chen, "Security in wireless sensor networks," IEEE Wireless Communications, vol. 15, no. 4, pp. 60–66, 2008.

[3] M. A. Sahi, H. Abbas, K. Saleem, X. Yang, A. Derhab, M. A. Orgun, W. Iqbal, I. Rashid, and A. Yaseen, "Privacy preservation in e-healthcare environments: State of the art and future directions," IEEE Access, vol. 6, pp. 464–478, 2018.

[4] H. K. Maji, M. Prabhakaran, and M. Rosulek, "Attribute-based signatures," in Proc. of Cryptographers' Track at the RSA conference, 2011, pp. 376– 392.

[5] X. Chen, J. Li, X. Huang, J. Li, Y. Xiang, and D. S. Wong, "Secure outsourced attribute-based signatures," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 12, pp. 3285–3294, 2014.

[6] Waziha Kabir, OmairAhmad.M, M.N.S Swamy, "A multi-biometric system based on feature and score level fusion", IEEE Transactions on Information and Forensics and Security, vol.10, no.8, pp.964-971, 2018.

[7] Tito Burghardt, Janko Calic, “Real time face detection and tracking of animals”, IEEE Seminar on Neural Network applications in Electrical Engineering, vol.12, no.6, pp.1228-1235, 2006.
[8] UmitCacar, MurvetKirci, “Scorenet : Deep cascade score-level fusion for unconstructed ear recognition.”, IET Transaction on Biometrics, vol.2, no.3, pp.215-224, 2019.
[9] Alexander Loos, Andrea Ernst, “Detection and identification of chimpanzee faces in the wild.”, IEEETransaction on Unified Automatic Image-based Face Detection, vol.7, no.4, pp.847-852, 2012.
[10] Dalila Cherifi, Fateh Cherfauoui, “Fusion of face recognition methods at score level.”, BioSMARTTransaction on Global and Local Methods of Possible Fusions, vol.5, no.2, pp.289-297.
[11] Balke.D.G, Archana Patil, “Fusion of fingerprint, palmprint and iris for person identification.”, ICACDOTTransaction on Authentication Process and Errors, vol.3, no.6, pp.406-415.
[12] Zan Gao, Hai Zen Xuan, Hua Zhang, “Adaptive fusion and category-level dictionary model for Multiview human action recognition.”, IEEETransaction on Multiple Resource from Various Domains in Analysing Human Actions, vol.5, no.3, pp.562-576.
[13] Carmen Bisogni, Michelle Nappi, “Multibiometric score level fusion using training and optimization.”, BioSMART Presentation on Optimization and Training to Generate Total Score of Multibiometric System, vol.6, no.4, pp.674-682.