IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF CNN BASED IMAGE CLASSIFICATION TECHNIQUE

[1] M. Senthil Kumar, [2] B. Chidambararajan, [3] R.Ranjith, [4] M. Sabarish, [5] J. Samuel Jayaraj

[1] Associate Professor, [2] Director, [345] UG Scholar
[12345] Department of Computer Science and Engineering,
SRM Valliammai Engineering College, Kattankuathur, Chengalpattu District, India.

Abstract: Even-though technologies and organizations evolve, some technological paradigms remain in the developing stage. Image Classification, Image Recognition, and Image Detection are yet some other fields that are been in the research and development stage and have been the most fundamental, popular, and crucial research direction. Image Classification technology based on Machine-Learning (ML) has been broadly used in featured Image Classification, Segmentation, and Recognition, and is a hot trend that developing rapidly in various fields. This paper proposed an ML Model which recognizes the image that the user feeds and predicts the image class to which it belongs. This paper mainly focused on methods and algorithms used to build the model and improve the accuracy. The Model proposed uses the most famous and powerful libraries TensorFlow (TF), Keras, etc to build the Sequential model which predicts and classifies the given image using a Convolutional Neural Network (CNN). CNN is specifically designed for Image Recognition and Processing (It is a type of Artificial Neural Network - ANN).

Key Words - Machine Learning, Deep Learning, Sequential Model, Convolutional Neural Network

I. INTRODUCTION

Image Classification Techniques emerged to bridge the gap between Computer Vision (CV) and Human Vision by providing corresponding information to the Computer. Artificial Intelligence (AI) is the superset, in which Machine Learning (ML) and Deep Learning (AI) reside as a subset. This AI has been a field of research for decades, with both scientists and engineers working intensely to solve the puzzle of Robots and Computers coming up with the ability to see and understand real-world objects. One of the most crucial aspects of this research is training the Computers on how to recognize visual information such as images and videos, that are generated all around us daily. CV is the study of making computers see and understand visual information. Image Classification was created to bridge the gap between CV and Human Vision by giving the required information to the Computer. Because there were hundreds of thousands of various ways to represent an entity and thousands (even millions) of different scenes and objects that only once, finding the most optimized way and accurate mathematical Models to represent all of the possible features of every object and scene, and is a job that will reach the highest point. The concept of ML was introduced in the year 1990 when Computers were told what and how to recognize from the scenes and objects in the images and videos. Instead, of designing various algorithms that will allow Computers to learn how to recognize scenes and objects in images on their own, it's like a child starts to learn and to know his/her environment and things around them through exploration. ML has made this possible and achievable for Computers to learn to recognize practically any object or scene.

II. RELATED WORKS

Over the past 20 years, where the neural networks take up a large part of the world's research and development [7], the Neural Networks were first mentioned in the mid-1960s in the book of Aleksey G. Ivakhnenko and Valentine G. Lapa. The very first model of CNN was called "neocognitron" and was discovered in 1980 by Kunihiko Fukushima [3]. Kunihiko Fukushima proposed many methods and algorithms for both supervised and unsupervised ML, and this neocognitron itself is a multi-layered deep structure [6]. Analysed the Neural Network Architecture (NNA) as an image classification method [1]. The framework is made out of two sets of imitators of human eye as well as sequence of varying auto-encoding. It involves a lot of complex image sets, but the system was successively improving the CIFAR10 models during the research. The open-source database CIFAR10 will be utilized as the training set. Given that medical image identification has lately gained the attention of academics, there are various challenges in studying this subject, including
1) Lack of training Datasets,
2) Differences in scaling images and images with noises.
These drawbacks were considered when developing a network model that includes a full-scale convolutional layer that extracts patterns from various receptive fields using a shared set of convolutional nuclei, allowing scale-invariant patterns to be captured [13].

Researchers and developers have been paying close attention to convolutional networks during the last decade.
ImageNet is one of the most famous AI and computer vision competitions in the world. The prize was won utilizing a convolutional network structure, such as AlexNet [1], VGG [12], GoogleNet [13], and ResNet [8] among the other artificial networks.
These networks perform admirably, with over 90% recognition.

III. METHODOLOGY

3.1 MACHINE LEARNING

Machine learning (ML) resides under the categorical classification of Artificial Intelligence (AI). ML algorithms were employed in the software applications to improve the prediction accuracy without any external efforts. ML algorithms predict the output values based on the given historical data and training made to the ML model. ML is essential because it helps businesses in many ways in which acquiring the information about customer actions in public places like Mall, Theatre, etc., and the business patterns will also be generated to assist in new product development. ML takes many of today's most businesses to success, like Google, Facebook, Uber, etc. For many organizations, ML has become an essential competitive differentiation.

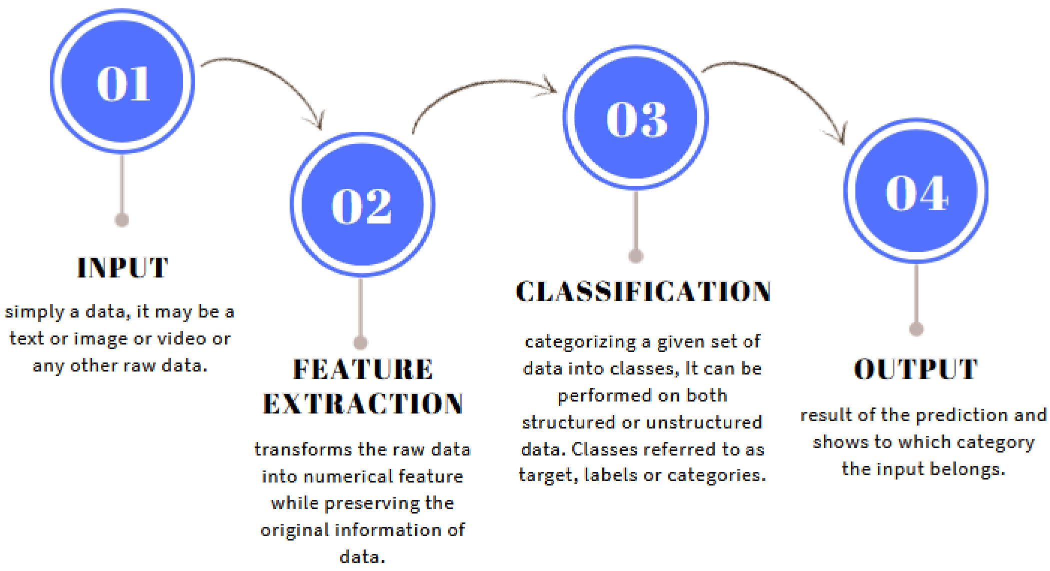

Figure 1 working process of ML.

3.2 DEEP LEARNING

Deep Learning is again a subset of Machine Learning technique that gives Computers, the ability to learn by example like the way human learns. Deep Learning is an essential component of self-driving transport, allowing the vehicles to detect traffic signs, pedestrians, posts, and other objects. It enables voice control over consumer electronics such as Mobiles, Tablets, Televisions, Speakers, etc. Deep Learning has grown and shows a drastic improvement in multidisciplinary fields. It's accomplishing the previously missed accomplishment, which means Deep Learning is developing rapidly with numerous ideas and their uses.

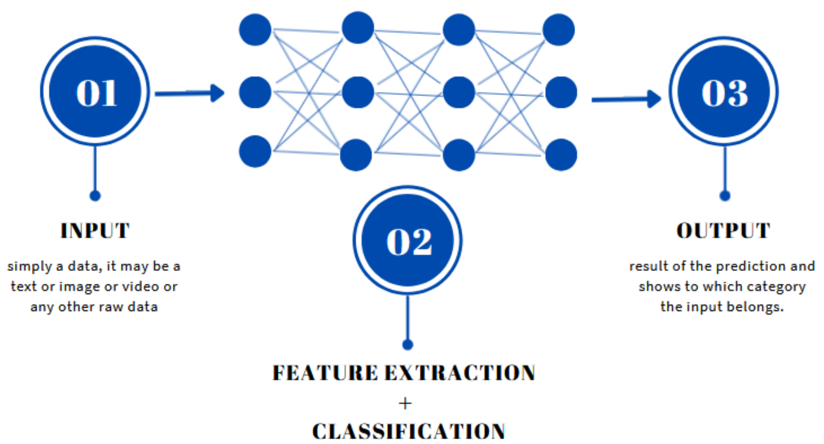

Figure 2 Working process of Deep Learning

3.3 NEURAL NETWORKS

A Neural Network (NN) is a type of network which is made up of highly interconnected processing nodes, just like the neurons in our human brain. Each node is linked to each other nodes from different levels like a mesh. These nodes transfer data via the network in a feed-forward fashion, which means the data only flows in one direction. The Deep Learning network comprises the Input layer, Convolutional Layer, Pooling Layer, Dense Layer, and finally the Output layer, but between these layers, hidden layers are there

too. In a Deep Learning network, nodes in each layer give their output as an input to the next layer. The nodes present in the next layer accept that input data and train on them. The number of layers is directly proportional to the prediction of complex information, which means more layers than greater the ability to find complex information. Each connected node is assigned a "weight/cost", which helps the network to make choices. When a node passes information to other nodes, then it initially determines the weights or value of that information before processing it. The information can be passed to the next layer if the weights reach a specified range of thresholds. The information can't be passed to the next layer if the weights fall below the threshold value. All the nodes of a newly created Neural network hold with the set of random values assigned to the weights and values.
If more training data is supplied to the Input Layer, then most probably outputs will be right.

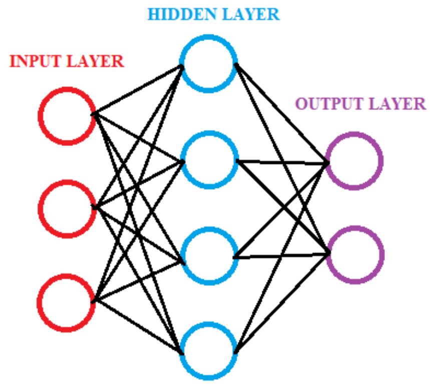

Figure 3 Structure of simple NN

3.4 CONVOLUTIONAL NEURAL NETWORK

A Convolutional Neural Network (CNN or ConvNet) is a deep learning network design or an algorithm that learns from input data without the requirement of human-like feature extraction. CNN's are more particularly useful for object, face, and scene recognition by looking for patterns in images. They're also useful for categorizing non-image data including audio, signal data, text information, etc. The application that uses Image detection, Classification used in self-driving vehicles, facial recognition, etc are just made possible by this CNN.

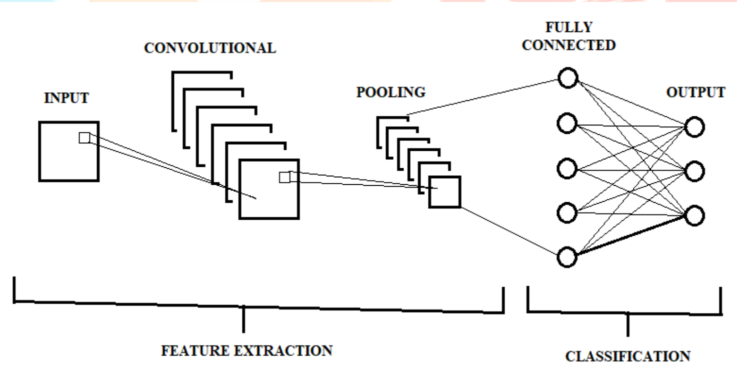

Figure 4 Structure of CNN

IV. IMPLEMENTATION

The CIFAR10 datasets contain 60,000 images of 10 classes. Those images were with a resolution of 32 x 32 pixels, which means those images were kept in a form of an array. Figure 5 shows the structure of the first tested CNN model. From the figure, the model contains 2 CNN layers with the input of image dimensionality (height=32 x width=32x scale=3). After that, the Pooling layer with 2x2 dimension followed by flattening of output. Finally, two dense layers and the final output got as 10 units.

Improving the above model with a further layer of CNN and totally 3 layers of CNN. Figure 6 shows the second tested model. 1st Conv layer gets the input as 32x32x3 and the output as in the form of 32x32x32 which is input to the 2nd Conv layer. The output of the 2nd Conv layer is 16x16x32 which is again input to the 3rd Conv layer. The output of that layer is in the form of 16x16x64. Further flattening inputs occur and results are obtained after two dense layers.

Figure 7 again shows the improved version of the 2nd tested model, here again, 4 Conv layers were included.

The above three models were trained for 20 epochs, which leads to getting the accuracy of 67.070%, 71.080%, and 73.500% respectively. Considering all the above models and their training with the corresponding outcomes, the final model is designed. Figure 8 shows the final model designed.

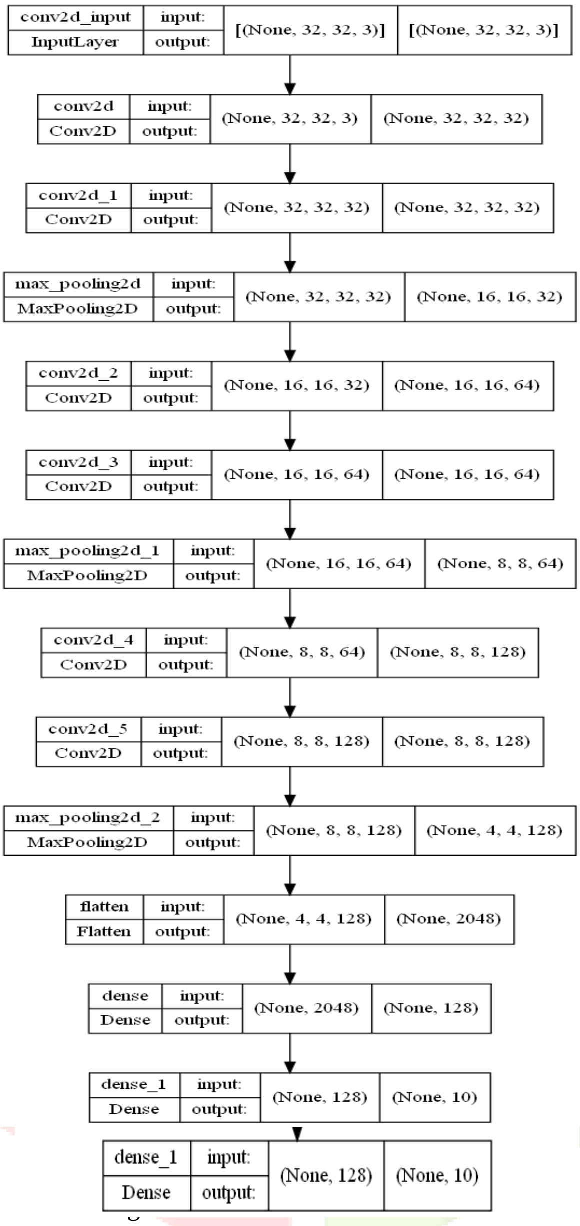

Figure 5 First tested CNN model

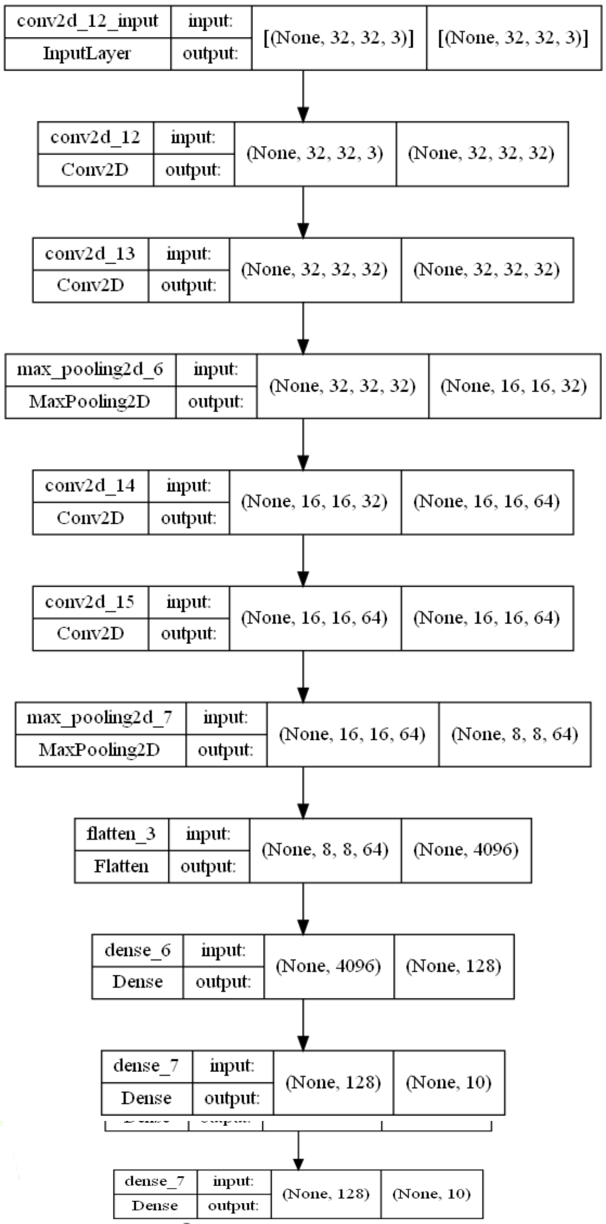

Figure 6 Second tested CNN model

In the final CNN model, the 1st Conv layer gets the input as 32x32x3 and output as 32x32x3 after this Pooling occurs after the 2nd layer and the output from the pooling layer is as 16x16x32. This output is then given as the input to the 3rd Conv layer and 4th Conv layer, then the output after these layers is in the form of 16x16x64. Then again pooling occurs after this process is transformed into 8x8x64 from 16x16x64. Then this output format is flattened to 4096 units. the 1st dense layer gets input as 4096 units and outputs to 128 units. then 2nd dense layer again gets the input of 128 units and outputs 10 units. Every model and its weights are saved periodically into a .h5 file. Finally, compiling the model and evaluated against the tested dataset. It can able to produce (93-95) % of accuracy in the prediction of image classes.

V. DISCUSSION

So, from the above, three models were explored, and based on the analysis of their outcomes, the final model is designed and evaluated. Figure 9 shows the epochs training process. As said above model 1 is capable of producing accuracy rates up to 67%, model 2 produces 71% and model 3 produces 73.5 %. By changing some hyper parameter values like adding dropouts and max-pooling layers, the final model is capable to produce 94% accuracy in predicting the images. Figure 10 shows the results obtained in evaluating model 3.

```
y: 0.6555
Epoch 14/50
1563/1563 [==============================] - 227s 145ms/step - loss: 0.8612 - accuracy: 0.6952 - val_loss: 0.9567 - val_accurac
y: 0.6662
Epoch 15/50
1563/1563 [==============================] - 521s 333ms/step - loss: 0.8200 - accuracy: 0.7086 - val_loss: 0.9402 - val_accurac
y: 0.6727
Epoch 16/50
1563/1563 [==============================] - 231s 148ms/step - loss: 0.7793 - accuracy: 0.7236 - val_loss: 0.9367 - val_accurac
y: 0.6779
Epoch 17/50
1563/1563 [==============================] - 226s 145ms/step - loss: 0.7487 - accuracy: 0.7350 - val_loss: 0.9360 - val_accurac
y: 0.6766
Epoch 18/50
1563/1563 [==============================] - 224s 143ms/step - loss: 0.7114 - accuracy: 0.7488 - val_loss: 0.9122 - val_accurac
y: 0.6858
Epoch 19/50
1563/1563 [==============================] - 229s 147ms/step - loss: 0.6796 - accuracy: 0.7586 - val_loss: 0.9307 - val_accurac
y: 0.6812
Epoch 20/50
1563/1563 [==============================] - 230s 147ms/step - loss: 0.6426 - accuracy: 0.7734 - val_loss: 0.9207 - val_accurac
y: 0.6879
Epoch 21/50
1563/1563 [==============================] - 224s 143ms/step - loss: 0.6067 - accuracy: 0.7856 - val_loss: 0.9227 - val_accurac
y: 0.6894
Epoch 22/50
1563/1563 [==============================] - 222s 142ms/step - loss: 0.5764 - accuracy: 0.7965 - val_loss: 0.9187 - val_accurac
y: 0.6922
Epoch 23/50
1563/1563 [==============================] - 226s 145ms/step - loss: 0.5399 - accuracy: 0.8085 - val_loss: 0.9340 - val_accurac
y: 0.6960
Epoch 24/50
1563/1563 [==============================] - 235s 151ms/step - loss: 0.5179 - accuracy: 0.8148 - val_loss: 0.9391 - val_accurac
y: 0.6931
Epoch 25/50
 722/1563 [============>.................] - ETA: 2:01 - loss: 0.4777 - accuracy: 0.8321
```

Figure 9 Epochs training module.

Figure 10 Results obtained on model 3 evaluation.

VI. RESULTS

Figure 11 it is showing that after rigorous training the accuracy is increasing after every level of iteration. Figure 12 represents the actual label of the image and the predicted label of the image. For instance, in the first part of the image actual image value is sandal, and the predicted label is also sandal. Similarly, the output is successfully predicted by labelling dog, truck, automobile, etc.,

```
In [*]: history = model.fit(train_X,train_Y,validation_data=(test_X,test_Y),epochs=50,batch_size=32)
1563/1563 [==============================] - 158s 101ms/step - loss: 0.1908 - accuracy: 0.9325 - val_loss: 1.1507 - val_accurac
y: 0.7070
Epoch 45/50
1563/1563 [==============================] - 163s 104ms/step - loss: 0.1844 - accuracy: 0.9351 - val_loss: 1.1832 - val_accurac
y: 0.7058
Epoch 46/50
1563/1563 [==============================] - 165s 106ms/step - loss: 0.1778 - accuracy: 0.9375 - val_loss: 1.2271 - val_accurac
y: 0.7066
Epoch 47/50
1563/1563 [==============================] - 161s 103ms/step - loss: 0.1734 - accuracy: 0.9388 - val_loss: 1.1888 - val_accurac
y: 0.7073
Epoch 48/50
1563/1563 [==============================] - 215s 138ms/step - loss: 0.1646 - accuracy: 0.9421 - val_loss: 1.1950 - val_accurac
y: 0.7116
Epoch 49/50
1563/1563 [==============================] - 208s 133ms/step - loss: 0.1623 - accuracy: 0.9439 - val_loss: 1.1983 - val_accurac
y: 0.7116
Epoch 50/50
 135/1563 [=>............................] - ETA: 3:19 - loss: 0.1423 - accuracy: 0.9521
```

Figure 11 Epochs training of the final CNN model

Figure 12 Final Result

VII. CONCLUSION

The random picture testing proved to be successful. The image dataset was CIFAR-10. For classification, a CNN is utilized in conjunction with Keras Sequential API. The Neural Network model is successfully developed and tested on the image dataset. The model can classify the test data successfully. The model can be trained on any data set and be implemented to classify different sets of images of different sizes. In the present model, an accuracy of 93.94% is obtained.

VIII. FUTURE WORKS

This paper has proposed a way of effective image classification using CNN. For which, the open-source dataset CIFAR10 is used as a training set and testing set. Yet the images in the dataset were static and all images are scaled in one range. This CNN model can be developed to recognize multiple images, scaled in various ranges, and at different angles. For which custom datasets of various images need to be trained to the model by changing their scale, skew, and angle positions.

IX. REFERENCES

[1] Alex Krizhevsky, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, 2012, 100-110.

[2] Chauhan K, "image classification with deep learning and comparison between different convolutional neural network structures using Tensorflow and Keras," International journal of advance engineering and research, 2018, 533-538.

[3] Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 1980, 193–202.

[4] Fusion Imaging in Pixel Level Image Processing Technique – A Literature Review, International Journal of Engineering & Technology, ISSN : 2227-524X.

[5] An Amplifying Image Approach: Non-Iterative Multi Coverage Image Fusion, International Conference on Clinical and Medical Image Analysis, 2018.

[6] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016, 770–778.

[7] Matthieu Courbariaux et al, "Training deep neural networks with low precision multiplications", Accepted as a workshop contribution at ICLR 201, ICLR, 2015, 371-410.

[8] Mohd Azlan Abu., "A study on image classification based on Deep Learning and Tensorflow", International journal of engineering research and technology, 2019.

[9] Shaoqing Renetal., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", work was done at Microsoft Research, NIPS, January 2016.

[10] Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. In: International Conference on Learning Representations, 2015.

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going Deeper with Convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[12] Wang X, Li X, Leung V C M. Artificial Intelligence-Based Techniques for Emerging Heterogeneous Network: State of the Arts, Opportunities, and Challenges", IEEE Access, 2017, 1379-1391.

[13] Zhang Y, Kwong S, Wang X, et al. Machine learning-based coding unit depth decisions for flexible complexity allocation in high-efficiency video coding.", IEEE Transactions on Image Processing, 2015, 2225-2238.