Department of Computer Science
University of Abuja
Abuja, FCT, Nigeria
E-mail: alhassan77077@gmail.com
Phone: +2348036331160

ABSTRACT

Face recognition is a very challenging research area in computer vision and pattern recognition due to variations in facial expressions, poses and illumination. Several emerging applications, from law enforcement to commercial tasks, demand the industry to develop efficient and automated face recognition systems. Although, many researchers have worked on the problem of face recognition for many years still several challenges need to be solved. Difference in illumination of the scene, changes in pose, orientation and expression are examples of some of the issues to be dealt carefully. Using machine learning principles and modern computer vision algorithms, it is possible to accurately detect objects of interest in an image frame, and to classify them with above human-level accuracy. To detect and categorize these images, state of the art methods leverage convolutional neural networks (CNNs). These net- works utilize a deep architecture consisting of convolutional layers, pooling layers, fully connected layers, and non-linear activation functions. This form of architecture transforms the input data into a 'deep' low-dimensional feature representation, which facilitates the classification task completed by the fully connected layers. This study consists of thorough reviews of different works that have been done in previous works on Facial Recognition using CNN.

Keywords: Artificial intelligence; CNN; Facial Recognition; Image Processing; Machine Learning

23rd iSTEAMS Conference Proceedings Reference Format
Siyaka, H.O., Owolabi, O. & Hashim, B.I. (2020): Convolutional Neural Network Algorithm and its application in Facial Recognition: A Review. Proceedings of the 23rd iSTEAMS Conference, American University of Nigeria, Yola. April, 2020. Pp 129-136 www.isteams.net/yola2020

1. INTRODUCTION

Face recognition is a very challenging research area in computer vision and pattern recognition due to variations in facial expressions, poses and illumination. Several emerging applications, from law enforcement to commercial tasks, demand the industry to develop efficient and automated face recognition systems. Although, many researchers have worked on the problem of face recognition for many years, there are several challenges that are yet to be solved. Difference in illumination of the scene, changes in pose, orientation and expression are examples of some of the issues to be dealt carefully. Using machine learning principles and modern computer vision algorithms, it is possible to accurately detect objects of interest in an image frame, and to classify them with above human-level accuracy. To detect and categorize these images, state of the art methods leverage convolutional neural networks (CNNs). These net- works utilize a deep architecture consisting of convolutional layers, pooling layers, fully connected layers, and non-linear activation functions.

This form of architecture transforms the input data into a 'deep' low-dimensional feature representation, which facilitates the classification task completed by the fully connected layers. This study consists of thorough reviews of different works that have been done in previous works on Facial Recognition using CNN. In the next sections, discussions will be made on Facial Recognition, CNN and then different works related to the application of CNN in Facial Recognition will be reviewed.

1.1 Facial Recognition

Face recognition refers to an automated or semi-automated process of matching facial images (Ajelolu, Orimogunje, Ewetola & Oladipupo, 2013). The image of the face is capture using a camera and then analysed in order to obtain a biometric ''signature''; different algorithms can be used for this and manufacturers have adopted various proprietary solutions. Face recognition is one of the most popular problems in the field of image analysis and understanding. Identifying a person from an unknown face is usually done by comparing the unknown face with known faces from a face database. The interest of researchers and engineers in face recognition problem has grown rapidly in the recent years since there is a wide range of commercial and law enforcement applications on face recognition (Zhao, Chellappa, Rosenfeld, and Phillips, 2003; Chellappa, Sirophey, Wilson, and Barnes, 1995).

The increasing need for surveillance-related applications, especially due to drug traffic and terrorist activities, has a great impact on the growth of interest in the field of face recognition.
Some of the application areas of face recognition include: personnel identification of credit cards, driver's license, passport checks, entrance control, computer access control, criminal investigations, crowd surveillance, witness face reconstruction, and ATM machines. The studies and findings of psychophysicist and neuroscientists help the engineers who are designing and /or implementing algorithms or systems for machine recognition of faces. Face recognition consists of three steps: face detection, feature extraction and face recognition. Face detection is a process where an object is detected and identified as a human face. Feature extraction involves extracting feature of the detected face, this features include (but are not limited to) hue features, skin texture and edges.

2. CONVOLUTIONARY NEURAL NETWORKS (CNN)

Convolution neural network is one of the deep learning algorithms used in the computer vision applications and object classification accuracy (such as object detection and recognition). CNN is particularly deployed for real-time processing, and parallel or heavy computations (Benjdira et al., 2019). These networks utilize a deep architecture consisting of convolutional layers, pooling layers, fully connected layers, and non-linear activation functions. This kind of structure converts the input data into a deep low-dimensional feature depiction for the purpose of facilitating the task of classification by the fully connected layers (Cameron et al., 2019).

Aside being effective feature depiction approach, CNNs have been extensively applied in the field of computer vision. Following the success story of AlexNet, several novel CNNs have evolved such as ResNet, VGG, DenseNet, etc. In fact, these have offered high performance in all computer vision tasks. The main module of CNN model is the convolution layer that is used extracts features from high-dimensional structural data in more efficient by a collection of convolution kernels (Zhu, Xu, Xu, & Chen, 2018). Whenever considering multi-channel inputs (such as color images), the convolution kernels combines the various channels by adding up the convolution results to produce one single output channel per kernel.
The foremost and notable CNN-based object detector is the Regional CNN (R-CNN), which was discovered to be more efficient when compared to conventional algorithms used in recognition or detection with a 30% improvement for the PASCAL VOC 2012 image dataset (Girshick, Donahue, Darrell, & Malik, 2014). After two years of R-CNN, several other associated algorithms continue to enhance object recognition or detection including Faster R-CNN (Ren, He, Girshick, & Sun, 2017), which offered a speedup of 250 times more over the existing R-CNN. In the case of the Faster R-CNN, a two phase algorithm is involved because of the two distinct

CNNs utilized for the detection of object. The first phase is responsible for finding region proposals in order to identify potential regions of position of an object. And, the second phase is used to classify the objects as well as refine the proposals. Another study by Ramachandran et al. (2012) highlighted the problems of object detection with CNNs based approaches; in particular the pre-CNN methods showed that the appearance of objects differs from a plethora of factors to include: lighting, orientation, the size of the objects, and occlusion (Ramachandran et al., 2012; Cameron et al., 2019).

CNN emulates the features of biological networks (Devi, Ravi, Vaishnavi, & Punitha, 2016). Analogous to AI-based approaches, the typical workflow of the CNN takes an input, operates on the input by means of activation functions, and the output is produced. It is a top choice for anything image classification, natural language processing, and image recognition, and image analysis. The name CNN was coined from the convolutional movement performed on the images during one of the above mentioned operations (Kudva, Keerthana , & Shyamala , Automation of detection of cervical cancer using convolutional neural networks , 2018). Its basic architecture consists of an input layer, hidden layer(s), and an output layer. On a broader scope, the layers in a CNN consist of multiple convolutional layers, pooling layers, and the fully connected layers. A brief description of the functions of each layer is given as follows (Bora, Lipi & Anup, 2016):

- **The Convolutional Layers:** As the name implies, this layer convolves the input and passes its outcome to the next layer.
- **The Pooling Layers:** The pooling layer minimizes the dimensions of the data (reduces the size of the image) by merging the results of a cluster of neurons from one layer into a single unit in the next layer. This operation can be done in two ways; max pooling or average pooling. Max pooling uses the maximum value obtained from the neuron clusters, while average pooling uses the average value obtained from the neuron clusters.
- **The Fully Connected Layers:** This layer connects every unit in a layer to every unit in another layer. Its function is similar to that of the Multi-Layer Perceptron (MLP) neural network.

Other important functionalities in a CNN include:

- **The Receptive Field:** In general, each neuron in the neural network receives input from some points of location in the previous layer. This point of input location is called the receptive field. The neurons in a fully connected layer receive input from all the neurons in the previous layer. In a convolutional layer, inputs are received from a portion of the previous layer.
- **The Weight:** This represents the influence or the strength of the connection between nodes or neurons in the network.
- **Bias:** The bias is an input node which produces a constant value 1. It allows the behaviour of the layer to be controlled.
- **Filter:** Filters are used for the detection of features on an image.
- **Stride:** The stride represents the number of pixels moved by the filter. Stride can help in reducing the image size.
- **Padding:** The size of images often shrinks with the application of a convolutional operation. This can lead to information loss. To curtail this issue, the image can be padded with an additional boarder. This can be done by adding a pixel to the edges of the image. This is known as padding.
- **Parameter Sharing:** Parameter sharing is performed to control free parameters. This could be implemented by using the same weight and bias for every neuron.

- Softmax layer/classifier**:** This layer classifies the input passed to it from the previous layer into several categories.
- **Activation Functions:** An activation function is applied to the output obtained after convolving over an image with a filter. This results in the generation of activations. The various type of activation functions used by previous researchers are the sigmoid (Xie, Fuyong , Xiangfei , Hai , & Lin , 2015), the ReLU (Xie, Fuyong , Xiangfei , Hai , & Lin , 2015), and the Tanh function.

2.1 CNN Architecture

According to various studies, the various CNN's that have been used in the classification of cervical images are the VGG-16 (Hyeon , Ho-Jin , Byung & Kap, 2017; Ma , et al., 2019), VGG – 19 (Plissiti, et al., 2018), LeNet, UNet, Googlenet, and the AlexNet Bora, Lipi, Anup, 2016; (Wu, Yan, Liu, Liu, & Yin, 2018). From these review, it can be sum up that the most popular technique used previously is the AlexNet technique. CNN has inherent capacity to perform exceptionally well for image classification and pattern recognition tasks, which is based on a biological inspired approach (Dixit, Mishra, Shukla, & Tiwari, 2019). The basic architecture of CNN is presented in Figure 1.

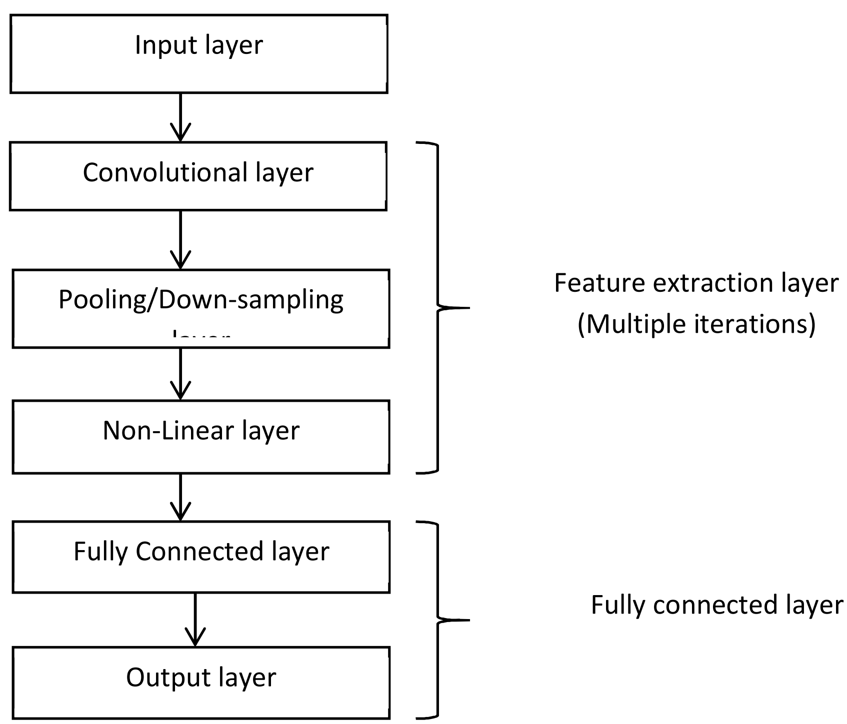

Figure 1: The basic architecture of CNN.

From Figure 1, the typical CNN operations commence at the data input layer by supplying image data to be processed by the convolutional layer. The convolutional layer creates the features map using the filter/kernels whose outcome is the raw features of the image. The feature map dimensionality is reduced with the down-sampling task of the pooling layer. The processed features from the ReLU component are passed to the fully connected layer for the purpose of completing the data classification process.

2.2 The Parameters and Hyper parameters

The hyper parameters in the AlexNet model can be deployed with various trials and errors for the purpose of attaining better results accuracy. The hyper parameters and parameters include: amount of stride, padding, and quantity of filters, filters sizes, and pooling. The parameters include weight, and bias. In general, these parameters and hyper parameters can be implemented on the distinct CNN layers in order to enhance the model effectiveness.

3. APPLICATION OF CNN IN FACIAL RECOGNITION

CNN is particularly deployed for real-time processing, and parallel or heavy computations (Benjdira et al., 2019). These networks utilize a deep architecture consisting of convolutional layers, pooling layers, fully connected layers, and non-linear activation functions. This kind of structure converts the input data into a deep low-dimensional feature depiction for the purpose of facilitating the task of classification by the fully connected layers (Cameron et al., 2019). The main module of CNN model is the convolution layer that is used extracts features from high-dimensional structural data in more efficient by a collection of convolution kernels (Zhu, Xu, Xu, & Chen, 2018). Whenever considering multi-channel inputs (such as color images), the convolution kernels combines the various channels by adding up the convolution results to produce one single output channel per kernel.

Farren (2017) shows retail product classification approach on grocery shelves and makes use of a novel dataset. It uses a greedy algorithm to enhance the network performance via converting the architecture. The parameters include filter size, number of filters of each convolution layer, the number of convolution layers, length and number of fully connected layer. The architecture use a batch normalize layer after first two convolutional max pool layers. The activation function was Relu with a dropout rate of 50 % on each Relu layer to avoid over fitting. The network was trained using of softmax loss function. Each iteration makes use of back propagation to calculate the gradient with respect to the loss at each layer. It then uses Adam algorithm to adapt the learning rate and change in layer weight overtime. The experimental effects show that there is grate increase in the model accuracy. However, when the layer was dropped, it is clear that the filter size did not make much of the relative difference in enhancing the model performance.

Vo, Tran, & Le (2017) proposes a model for classifying images into two groups YES or NO. Yes means for the advertisement being displayed openly, "clear" means that the person can view the content of the advertisement. No refers to not displayed or not clear. The technique entails the usage of Convolutional Neural Network with two parameters (number of layers, number of filters in conv' layer). The proposed model involves four levels that are input, capturing, classification, and output. The input is URL of a website which is saved as screenshot. These images are then resized (32x32 pixels) for processing via CNN. The output is one in every of two conclusions YES or NO. The results showed an accuracy of (85.74%) which reflects feasibility of the proposed model.

The foremost and notable CNN-based object detector is the Regional CNN (R-CNN), which was discovered to be more efficient when compared to conventional algorithms used in recognition or detection with a 30% improvement for the PASCAL VOC 2012 image dataset (Girshick, Donahue, Darrell, & Malik, 2014). After two years of R-CNN, several other associated algorithms continue to enhance object recognition or detection including Faster R-CNN (Ren, He, Girshick, & Sun, 2017), which offered a speedup of 250 times more over the existing R-CNN. In the case of the Faster R-CNN, a two phase algorithm is involved because of the two distinct

CNNs utilized for the detection of object. The first phase is responsible for finding region proposals in order to identify potential regions of position of an object. And, the second phase is used to classify the objects as well as refine the proposals. Another study by Ramachandran et al. (2012) highlighted the problems of object detection with CNNs based approaches; in particular the pre-CNN methods showed that the appearance of objects differs from a plethora of factors to include: lighting, orientation, the size of the objects, and occlusion (Ramachandran et al., 2012; Cameron et al., 2019).

The quest to improve the face quality of video streams and processing speed of face recognition system motivated the work of Qi, Liu and Shuckers (2018). The approach used frame selection of key-frames extraction engine and graphic processing unit (GPU) acceleration, which is used to extract key-frames of high quality faces correctly and speedily. The outcomes improve the face recognition accuracy of the new CNN procedure. Sharma, Jain, & Mishra, (2018) examines the performance and predicting accuracy of three specific convolutional neural networks with datasets (CIFAR10 and CIFAR 100) on ten different class of objects. The result analysis confirmed that GoogLeNet and ResNet50 accomplished better in precision compared to AlexNet. It also found out that transfer learning achieved better than the use of trained network alongside greater accuracy. It is accepted that neural networks is taken into consideration as a new and satisfactory emerging techniques for fixing categorization issues.

Deep learning applies multiple processing layers to learn representations of data with multiple levels of feature extraction. This emerging technique has reshaped the research landscape of face recognition (FR) since 2014, launched by the breakthroughs of Deep face method. Since then, deep FR technique, which leverages hierarchical architecture to stitch together pixels into invariant face representation, has dramatically improved the state-of-the-art performance and fostered successful real-world applications (Wang and Deng, 2019). To improve on the performance of recognition and authentication system of CNN, the layering of the convolutional and sampling layers collapsed into a single layer was proposed by Wang and Li (2018). The concept relied on the optimizing the parameters of face data through pre-training with fully connected layer and softmax classification layers. However, there is need for large-scale images during training phase.

Recently, convolutional neural networks have made great achievements in resolving different image processing problems for FR applications. Yu et al. (2018) proposed a novel method called biometric quality assessment (BQA) for face images, investigating its applicability in FR applications. They used a light CNN with the max-feature-map units to make the BQA method more robust to noisy labels. Their studies have been explored further through experiments on the YouTube, FLW, and CASIA databases. The results of their experiments show a high degree of effectiveness of their proposed BQA method. Nam et al. (2018) proposed a CNN model named PSI-CNN for face recognition. The PSI-CNN model extracts untrained features from the image, then fuses these features with original feature maps. The results of the experiments are shown in terms of matching accuracy, with the model outperforming the model derived from the VGG-Face model. Also, PSI-CNN was able to maintain stable performance when tested on low-resolution images acquired from CCTV cameras. In case of change in image resolution and quality, PSI-CNN is robust.

Prasad et al. (2019) studied deep learning-based face representation for different face recognition challenges, such as misalignment, lower and upper face occlusions, illuminations, and different angles of head poses. They used two approaches—VGG-Face and lightened CNN. The AR face database used to evaluate the approaches' results of the study showed that deep learning approaches provide a good result in terms of recognizing faces and pre-processing. Khan et al. (2019) proposed a system for face recognition using portable smart glasses based on CNN. The detection process was performed using Haar-like features. The method archived detection rate at 98% using 3099 features. They used transfer learning from AlexNet for trained CNN model. The experiments of the study were conducted using 2500 images in a class. The results of the study showed that the accuracy of the system proposed was 98.5%. Qin et al. (2019) proposed a recognition algorithm based on deep CNNs. The algorithm contained face detection, face alignment, and feature extraction. The deep CNNs VGG16 was used to extract facial features. The experiments used the images of five angles (left, right, front, overlook, and look up). The experiment results showed that the algorithm achieved well on recognizing faces for cases of various poses in an indoor environment.

4. CONCLUSION

In this study, varieties of literature by different authors have been reviewed in order to expose the extent to which CNN has been used to tackle different issues in Facial Recognition. From the works reviewed it can be concluded that CNN has been and will continue to be a very important deep learning algorithm used in Facial Recognition and other biometric problems.

REFERENCES

1. Ajelolu, A., Orimogunje, O. K., Ewetola, M., and Oladipupo, E. T. (2013) BIOMETRIC SYSTEM. Term paper on Biometric, 5-7. Oyo State: Unpublished.
2. Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., & Ouni, K. (2019). Car Detection using Unmanned Aerial Vehicles : Comparison between Faster R-CNN and YOLOv3. 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), 1–6. IEEE Bora, Lipi & Anup, 2016
3. Cameron, J. A. D., Savoie, P., Kaye, M. E., & Scheme, E. J. (2019). Design considerations for the processing system of a CNN-based automated surveillance system. 136, 105–114. https://doi.org/10.1016/j.eswa.2019.06.037
4. Chellappa, R., Sirohey, S., Wilson, C. & Barnes, C. (1995). Human and Machine Recognition of Faces: A Survey. Proceedings of IEEE, 4, pp. 705-740.
5. Devi, M. A., Ravi, S., Vaishnavi, J., Punitha, S. (2016). Classification of cervical cancer using artificial neural networks. Procedia Computer Science, 89, 465-72.
6. Dixit, U., Mishra, A., Shukla, A., & Tiwari, R. (2019). Texture classification using convolutional neural network optimized with whale optimization algorithm. SN Applied Sciences, (May). https://doi.org/10.1007/s42452-019-0678-y
7. Farren, D. (2017). Classifying food items by image using Convolutional Neural Networks. http://cs231n.stanford.edu/reports/2017/pdfs/607.pdf
8. Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014). Rich Feature hierarchies for accurate object detection and semantic segmentation. In proceeding of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
9. Hyeon J, Choi HJ, Lee KN, Lee BD (2017). Automating papanicolaou test using deep convolutional activation feature. In Proceeding. 18th International Conference IEEE MDM, pp. 382-85.
10. Khan, S.; Hammad, J. M.; Ehtasham, A.; Shah, S. A. A. and Ali, S. U. (2019). Facial Recognition using Convolutional Neural Networks and Implementation on Smart Glasses. 2019 International Conference on Information Science and Communication Technology (ICISCT)
11. Kudva, Keerthana, & Shyamala, (2018). Automation of detection of cervical cancer using convolutional neural networks, 2018
12. Ma, B., Xia, Y. (2018). "Autonomous Deep Learning. A Genetic DCNN Designer for Image Classification".

13. Nam, G.P.; Choi, H.; Cho, J. (2018) PSI-CNN: A Pyramid-Based Scale-Invariant CNN Architecture for Face Recognition Robust to Various Image Resolutions. Application Science, 8, 1561
14. Plissiti, M. E., Dimitrakopoulos, P., Sfikas, G., Nikou, C., Krikoni, O., & Charchanti, A. (2018). SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 3144-3148). IEEE.
15. Prasad, P.S.; Pathak, R.; Gunjan, V.K.; Rao, H.V.R. (2019) Deep Learning Based Representation for Face Recognition; Springer: Berlin, Germany; pp. 419–424.
16. Qi, X., Liu, C. and Schuckers, S. (2018). Boosting Face in Video Recognition via CNN based Key Frame Extraction. 2018 IEEE 4th Int Conference on Identity, Security, and Behavior Analysis (ISBA)
17. Qin, C.; Lu, X.; Zhang, P.; Xie, H.; Zeng, W. (2019) Identity Recognition Based on Face Image. Journal of Phys. Conference. Series 1302, 032049.
18. Ramachandran, U., Hong, K., Iftode, L., Jain, R., Kumar, R., Rothermel, K. & Sivaku- mar, R. (2012). Large-scale situation awareness with camera networks and multimodal sensing. Proceedings of the IEEE, 100 (4), 878–892.
19. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans on Pattern Analysis and Machine Intel, 39 (6), 1137–1149.
20. Sharma, N., Jain, V. & Mishra, A. (2018). An Analysis of Convolutional Neural Networks for Image Classification. Procedia Computer Science, 132, 377-384.
21. Vo, A. T., Tran, H. S. & Le, T. H. (2017, October). Advertisement image classification using convolutional neural network. In 2017 9th International Conference on Knowledge and Systems Engineering (KSE) (pp. 197-202). IEEE.
22. Wang, J. and Li, Z. (2018). Research on Face Recognition Based on CNN. IOP Conference, 2nd International Symposium on Resource Exploration and Environmental Science. Series: Earth and Environmental Science 170 (2018) 032110
23. Wang, M. and Deng, W. (2019) Deep Face Recognition: A Survey. School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China.
24. Wu, M., Yan, C., Liu, H., Liu, Q. & Yin, Y. (2018). Automatic Classification of Cervical Cancer from Cytological Images by Using Convolutional Neural Network
25. Xie, Y., Fuyong, X., Xiangfei, K., Hai, S., & Lin, Y. (2015). Beyond classification: structured regression for robust cell detection using convolutional neural network. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 358-365. Springer, Cham.
26. Yu, J.; Sun, K.; Gao, F.; Zhu, S. (2018) Face biometric quality assessment via light CNN. Pattern Recognit. Lett. 107, 25–32.
27. Zhao, W.; Chellappa, R.; Phillips, P. J. & Rosenfeld, A. (2003) Face recognition: A literature survey. ACM Computing Surveys (CSUR), 35(4):399–458.
28. Zhu, X., Xu, Y., Xu, H., & Chen, C. (2019). Quaternion Convolutional Neural Networks. ECCV 2018, 1, 1–17.

Towards a Model for Assessing the Sustainability of Rural Community Networks

Auwal, A.T., Longe, O.B., Jean-Paul, C. & Bukhari, B.,
School of IT & Computing
Information Systems Programme
American University of Nigeria
Yola, Adamawa State, Nigeria