Bangladeshi Police Dress Identification using Deep Learning

Mahmuda Khatun
Assistant Professor
Dept. of Computer Science and Engineering
Comilla University
Cumilla, Cumilla-3506
Email: nishucse03@gmail.com

Md. Rakib Mia
B.Sc. Engineering in
Dept. of Computer Science and Engineering
Comilla University
Cumilla, Cumilla-3506
Email: mdrakib.cse11@stud.cou.ac.bd

Ahsan Habib
B.Sc. Engineering in
Dept. of Computer Science and Engineering
Comilla University
Cumilla, Cumilla-3506
Email: ahsanhabib.cou@stud.cou.ac.bd

***Abstract-* The technique of recognizing an item or feature in an image or video is known as image recognition. It is employed in several contexts, including flaw identification, imaging in the medical field, and security monitoring. Most vision-based AI systems and applications today are focused on object recognition. Object recognition is crucial for clear determination, which is helpful in security, transportation, healthcare, and military use cases object detection in retail. Due to the large range of dress designs and the intricate nature of their commonalities, it is difficult to precisely categorize police uniforms. In Bangladesh, uniforms for the police and security forces come in a variety of hues. This makes it difficult to improve performance using a simple classification system. However, using this strategy will ensure that you may conduct study in this exciting field. Consequently, here is the technique for identifying Bangladeshi police uniforms from a picture was developed. In order to achieve these goals, we first constructed the CNN model for predicting Bangla digits since CNN is the most intelligent object prediction model. After a thorough examination, the suggested system attained a better accuracy on our own customized dataset.**

Keywords: Recognition, Identification, Machine Learning, Uniform Identification, CNN.

I. INTRODUCTION

The capacity to identify a previously encountered object as being familiar is known as object recognition. The length of time a research participant appears to spend focusing on the object can be used to gauge how familiar the thing is. The task is known as social recognition when the object is a research participant or another inanimate object. Primate subjects are frequently utilized in object recognition research, particularly in the delayed nonmatching to sample task. The monkey is rewarded for choosing one of two objects that was not selected on the previous trial in this activity, which consists of trials that are separated by delays of many seconds. The fact that primates inherently seek to investigate new things is used to their advantage in this task. Due to this methodology, some researchers have created specific algorithms for tackling particular sub-tasks, such as tracking, estimating relative depth, and detecting independent motion by a moving observer (Ballard and Brown, 1992; Huang and Aloimonos, 1991; Fermuller and Aloimonos, 1992; Sharma and Aloimonos, 1991). Because of how important tracking is, many scholars have been interested in it (Fermuller and Aloimonos, 1992; Coombs and Brown, 1991; the first seven papers in Blake and Yuille, 1992). There are several causes for this interest. In order to gather as much data as possible about a moving item, the observer can 3D-track it and keep it in great focus at the center of the image.

We've already shown how an active strategy can aid in reconstructing the observed scene's three-dimensional structure. The construction of a functional structure from motion module is still challenging even though the geometrical structure might be retrieved through an analysis of image sequences under controlled observer motion conditions. This module would provide us with solutions for the item detection and navigation issues, but not every time we use our visual system will result in a fully reconstructed scene. Many activities that involve visual input only call for the extraction of a small amount of highly particular data. This indicates that not all visual information, nor should every component or object in a scene be extracted or thoroughly examined.

Fig 1.1: Unauthorized entries

Here in figure no 1.1. It shows an unauthorized entry in the office which can be detectable by using this model based on uniform. It will be very much helpful for security purpose in our home/office security system.

Effective object identification systems are necessary for activities in the real world, such as school uniform recognition, medical camp to identify doctors, and conference office security based on allowed individuals. Because of this, it's crucial to recognize uniforms, which have mostly been researched in earlier decades of object recognition/identification study. Numerous studies have been conducted on this subject, including ones for face recognition, but none of them have focused on the image-based identification of Bangladeshi police uniforms.

Today the majority of vision-based AI systems and software are built on object recognition. Clear determination, which is useful in security, transportation, healthcare, and military use cases, depends heavily on object detection. detection of objects in retail.

In conclusion, it is challenging to accurately classify police uniforms due to the wide variety of styles of dress and the complicated nature of their similarities. There are numerous colors for both security and police uniforms in Bangladesh. Because of this, improving performance with a straightforward categorization algorithm is challenging. However, this approach will guarantee the ability to research this fascinating area.

As a result, developing a method for identify Bangladeshi police uniforms from an image. In order to meet these objectives, we created our customized own dataset and then developed the CNN model for predicting Bangla digits, as CNN is the smartest object prediction model.

II. MOTIVATION

There has been a lot of research already done on recognizing a specific object. Only object recognition is one of the main visual issues of interest; nevertheless, other tasks rely on visual information as well. The most crucial is navigation, which is the act of navigating through space with the use of visual sensors. In addition, there are several challenges or sub problems that a perceiving agent that is intelligent, autonomous, and capable of learning must routinely solve that only require the extraction of a small amount of task-specific information from images. Obstacle recognition and avoidance, independent motion detection, tracking of a moving item, interception, hand-eye coordination, and others are a few of these issues. Attempting to address each of these issues independently can result in considerably more cost-effective solutions and the creation of useful working modules that can be activated as needed (Aloimonos, 1990). Object recognition concerns the identification of an object as a specific entity (i.e., semantic recognition) or the ability to tell that one has seen the object before (i.e., episodic recognition). Interest in object recognition is at least partly caused by the development of a new theory of human object recognition by Biederman (1987).

However, there is currently ongoing study on Bangladeshi Police uniform identification, which is quite significant and has numerous academic and commercial interests if it were to be evolved into a widespread practice. This study focuses on the identification of the BD Police uniform, which opens the door to future study in the same area for other uniforms used for other purposes, such as the BD Army, school clothes, hospital aprons, security uniforms, etc. However, dealing with the security function to identify permitted entry is the key issue and duty for it. The opportunity to investigate this fascinating subject is demonstrated by the dearth of similar work in the Bangladeshi uniform.

There are many machine learning algorithms for object recognition, including Deep Learning, CNN, RNN, Faster-RCNN, KNN, etc. Analyzing the works of other researchers, it is found that CNN gives the best performance for recognizing digits and characters. That's why CNN is used to develop the model. The actual purpose of this project is to recognize Bangladeshi Police Dress recognition with the highest accuracy.

III. AIMS AND OBJECTIVES

The research aims to build a Bangladeshi Police Dress Recognition model to predict someone is wearing police dress or not. The primary objectives of this research can be stated as follows:

- Detection of police dress and with-out police dress images by using various object detection algorithms
- Create our own dataset and prepare them to feed in a CNN (Convolution Neural Network) model.
- Development of a CNN object recognition model to predict the BD police dress.

IV. CONTRIVUTION

Our contribution to this study comprises the following:

- Create a Police dress and with-out police dress dataset
- Examine the capabilities of various object detection methods

Highly accurate face shield and non-face shield image detection

V. LIMITATIONS OF THE SYSTEM

Our approach is based on research on utilizing the CNN model to identify Bangladeshi Police Dress as accurately as possible. The researcher can utilize it for a number of things, including security, approved identification, unpredictable entries, etc., but this is only the first stage. Additionally, this research cannot distinguish between various uniforms, which might be the subject of future study. But because it has been trained to recognize them, this system can detect rotated, shifted, and zoomed photos. Without any augmentation, our system obtained 99.82 percent training accuracy, 97.71 percent validation accuracy, and 97.60 percent test accuracy. The system was unable to distinguish between training accuracy that was higher than validation accuracy or test accuracy.

VI. LITERATURE REVIEW

Pathak and A. R., Pandey at al. [1] This research clarifies the function of convolutional neural network-based deep learning algorithms for object detection. A range includes deep learning frameworks and object tracking services is also presented. This research assesses deep instructional strategies for tearing object identification systems. Sharma & Mishra at al. [2] Thus according their assessment, Google Cap and ResNet50 are more adept at feature extraction over Alex Net. Furthermore, they will investigate potential explanations for the significant variability in performance of trained CNNs across broad categories of objects. Nedeljkovic et al. [3] In order to summarize a SPOT imagery using a fuzzy logic clustering, a risen about the spectrometry for certain vegetation types is utilized. The basic concept was to do the categorization phase first employing supervised assessment, and then using fuzzy logic. Results of two approaches, respectively centered on pixel-by-pixel approach, were contrasted, and some encouraging conclusion observations emerged. Zagoruyko and Lerer et al. [4] These innovations have the potential to allow content to transit across their ecosystem via many pathways, including through aspects from other network levels and various object views. Their enhanced predictor is alluded to as a "Multi - path" subnet. Li, Z., Peng et al. [5] Throughout this article, authors suggest DetNet, a manufacturer backhaul infrastructure expressly tailored for object identification. On the MSCOCO metric centered on their DetNet (4.8G FLOPs) cornerstone, state-of-the-art results have been obtained for both feature extraction & helps individuals. Mukhometzianov et al. [6] In this experiment, they assessed the CapsNet method's efficacy against that of three well-known predictors (Fisherfaces, LeNet, and ResNet). On training set comprising a variety of occasions and classes, featuring pictures of people, road signs, and common items, they measured the prediction performance. Bochkovskiy et al. [7] They incorporate several of the most outcome measures, including WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, DropBlock denoising, and CIoU loss, to provide cutting-edge outcomes. Zhang and Wang et al. [8] They find the abovementioned practical method of executing SWA in feature extraction after significant experimentation. Al-Doski and Mansorl et al. [9]

This essay examines the elements of the approaches for categorizing images as well as the associated processes. Ming and Zhou et al. [10] They provide a dynamic anchor learning (DAL) approach that makes use of the defining mirroring extent to thoroughly assess the hooks' capability for navigation and performs a quicker label allocation procedure. Lu, D., & Weng et al. [11] The techniques, issues, and future of picture categorization are examined in this essay. Jiang and Luo et al. [12] To anticipate the IoU across each detectable grid cell and the matching surface, they suggest IoU-Net filtering in their article. Duan, K., Bai et al. [13] This research offers a cost-effective method for efficiently investigating the sensory cues inside each clipped patch. Erhan, D., Szegedy et al. [14] They provide a neural network model approach spotting that takes cues from sensory gating that anticipates a collection of class-neutral embeddings together with a numerical result for each rectangle that reflects the chance that it will comprise any entity. Guo, C., Fan et al. [15] In order to overcome these issues, they initially analyze the showcase hierarchy in FPN's design flaws before implementing recursive neural structure called Aug FPN. Hall, D., Dayoub et al. [16] In effort to be deployed on robots and immersed AI systems in the actual world, their study intends to promote the creation of innovative feature extraction techniques that give intercepts with precisely predicted geographic and labelling uncertainty. Foody et al. [17] It may be possible to guarantee that judgments of categorization precision are reasonable and lessen unjust criticism of thematic maps produced by satellite imagery by having a clearer understanding of the issues confronted in precision evaluation. He, T., Zhang et al. [18] In this research, they will look at a number of these improvements and use an erasure analysis to experimentally assess how they affect the precision prototype. Xu, H et al. [19] To facilitate squishy invitationals choosing and mixing, they create a "filtering infrastructure" within the state's dorsally subsystem. Unlike extra labor, the Deep Region lets structure is taught from beginning to end. Vani, A. K., Raajan et al. [20] First, utilizing Apache Cascading and the French horn expression began to look, the suggested study effort detects individuals and morphological traits such as the eyes, mouth, and nose. Benbrahim et al. [21] The 10,015 thermoscopic pictures in the HAM10000 sample are used to evaluate this design. The categorization of the survey's findings demonstrates that their actress's correctness was attained. Jiang et al. [22] It is crucial to obtaining streamlined farmland because it gives an accurate and trustworthy reasoning foundation for placement and quantifiable agrochemical treatment. Li, A., Li, Y et al. [23] With the use of convolutions, maxpooling, and compact slabs, together with the Residual blocks and the exporting sigmoid activation curve, they created a supervised learning method for the supplied sample of 209 RGB images. Boncolmo et al. [24] In this research, an authentic method for human activity recognition from photographs is presented. The technique is therefore a trustworthy tool for figuring out a recipient's masculinity. The Consecutive Dcnn analysis was conducted to investigate optimum ranges in order to accurately identify the existence of hoods whilst still.

VII. RESEARCH SUMMARY

Creating a dataset and categorizing them based on whether or not they are wearing police attire are the two main components of this research topic. The entire procedure is not a novel idea in other object classification research. However, they are all used to identify if an object is a car, a person, an animal, a tree, a structure, or something else. However, the distinguishing feature of our study is its ability to recognize the BD police uniform using data from our own dataset.

After that, binary images are created from the collected images. The image is then processed to remove noise and strengthen it. After scaling and padding for digit identification, each frame is delivered to the pre-trained CNN model for classification. We first prepared the dataset for our CNN model before training it. We began by rectifying some of the dataset's incorrectly labeled images, after which we deleted some of the incorrect images, such as blank ones. We already included inverted images with edge thickening, inverted forefront and backgrounds, and noise reduction using the median filter in our collection. This system has been trained to detect inputs that have been rotated, moved, expanded, overlapped, and obscured with the best output.

VIII. SCOPE OF THE PROBLEM

There is currently research being done on the identification of Bangladeshi Police uniforms, which is extremely significant and has a wide range of academic and commercial interests if it were to become a common practice. The identification of the BD Police uniform is the primary subject of this study, which pave the way for additional research in the same field on other uniforms worn for different purposes, such as the BD Army, school clothes, hospital aprons, security uniforms, etc. However, the main problem and responsibility for it is to deal with the security function to recognize authorized entrance. The paucity of comparable work in the Bangladeshi uniform demonstrates the opportunity to research this fascinating issue.

IX. CHALLANGES

Due to the large range of dress styles and the intricate nature of their commonalities, it is difficult to precisely categorize police uniforms. In Bangladesh, uniforms for the police and security forces come in a variety of uniforms. This makes it difficult to improve performance using a simple categorization system. However, using this strategy will ensure that you may conduct study in this exciting field. Therefore, it is impossible to tell one uniform from another in this study; however, this could be the subject of future research. The system was unable to differentiate between test accuracy and training accuracy, which was greater than validation accuracy.

X. METHODOLOGY

We will go over the full study technique in this section. The use of all strategies is covered in the methodology section. Here, a thorough description of how to use the model is provided, along with a quick summary of each element of the methodology. This section begins with an explanation of the data collected before going into detail about the entire experiment. The data was preprocessed as necessary, including image resizing, grayscale conversion, and enhancement. The proposed CNN model's operational methodology was provided.

Each component of the strategy is skimmed over in this chapter. Each section is followed during the study project's execution. Greater operational efficiency and grandeur are conferred by a clearer methodology description. Understanding the entire project is made easier by the use of scientific formulas and a visual appearance of the model that includes their descriptions. There is a need for additional research in this field as well as for the research field to be expanded and for clearer technique descriptions.

A. System Workflow

The workflow diagram for police dress detection

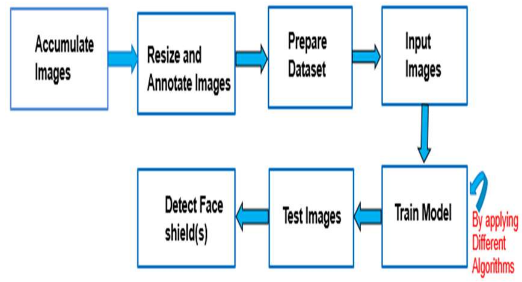

Fig 3.1 : Block diagram of proposed system

used in this suggested system is shown in Fig. 3.1. For input, we gather pictures of people in police

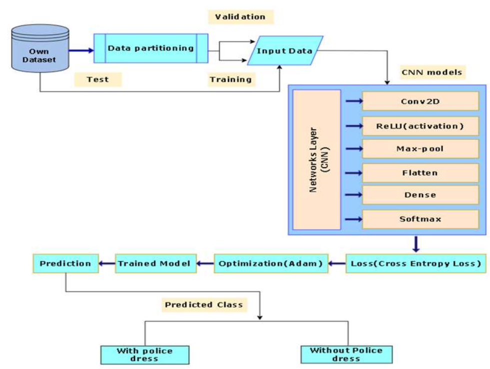

Figure 3.2 : Flow diagram of model implementing Police Dress recognition

uniforms and others without it from a variety of sources. The photos are then resized to produce better results. The resized photos are then annotated using the bounding-box toolkit. Then, using those images, we create our dataset by dividing it into train, val, and test folders. With the CNN model, we plan to train our model. The identification of police attire is then carried out using a set of test photographs, and lastly, we will receive images of wearers and non-wearers of police attire with high prediction accuracy.

B. Data Collection

For this study, various categories of images have been gathered. Most of these were taken with the OnePlus 6T, mi9 lite, POCO X2 and iPhone 11 Pro, while some of them are from Google. In our collection, there are 5000 photos that are both with and without police gear.

C. Data Preprocessing

After gathering all the photographs, the fact that they are of various sizes presents a new issue. Therefore, we downsized all of the photos to 256 by 256 for improved performance. Using the bounding-box toolkit, we then annotate the photos in the train and validation folders. Fig. 3.1 is annotated as follows: Workflow Diagram for Police Dress Detection is kept in a separate text file, and we prepare those files according to the specifications of the various algorithms. We additionally enhanced all of the images by flipping, blurring, shearing, and altering the brightness because our dataset is short.

D. Model Development

Neural networks are commonly used in the recognition of image analysis problems. The most promising tool for doing this is CNN, a deep learning approach that draws inspiration from the neuron connection pattern in animal visual cortex. It is mostly used for jobs like object recognition, picture identification, and other related ones. To create the CNN model for categorizing data, TensorFlow was utilized as the backend and the KERAS Python library as the front end. A Sequential Model, which consists of a layer stack that runs in a straight line, is used as a classifier. CNN neurons have weights and biases that can be taught. Among the algorithms, CNN requires the least amount of preprocessing. While a CNN's input is a multi-channeled picture, a neural network's input is always a vector. A CNN is made up of an input layer, hidden layers, and an output layer. The convolutional layer, the Rectified layer unit (ReLU), which serves as the activation function, pooling layers, normalized layers, and totally linked layers make up the hidden layer. CNN's neurons have weights and biases that can be taught.

Feature extractors and classifiers are the two main parts of a CNN's overall design, which is shown in Figure 3.3. The output of each layer (input layer) in the feature extraction unit is sent to the intermediate layer below it, which is thought of as the preceding layer's output layer, and the immediately next layer receives the current output as input. The categorization sections

then predicts the results based on the input data. Convolution and pooling layers are fundamental components of CNN design. To extract features from the input photos, each node in the convolution layer performs a convolution operation on the input nodes. The max-pooling layer tabulates the feature

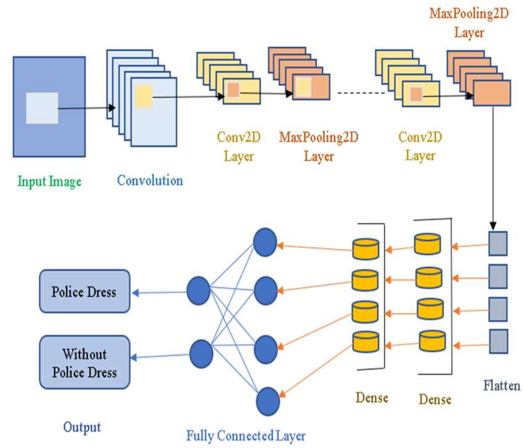

Fig 3.3: Workflow diagram of CNN model

by averaging or maximizing operations on input nodes. In terms of Neural Network output, we have a vector like this. Where Pc is the probability of the class, so if the image doesn't contain anything object, then Pc = 0, otherwise it will be Pc = 1. Then the bounding box so, Bx and By is the coordinate of the center which is indicated an yellow circle here. And Bw, BH is image width and height of this yellow box. C1 is is the class of the shield.

Now here is the image of no object, that's why Pc = 0 and others value doesn't matter.

Fig 3.5: Without Object

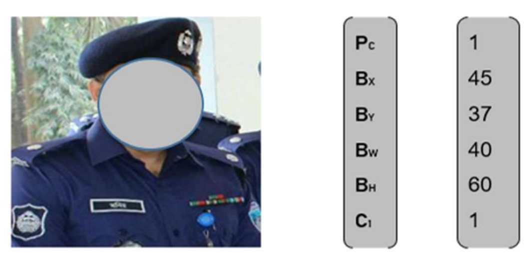

Fig 3.4: Object

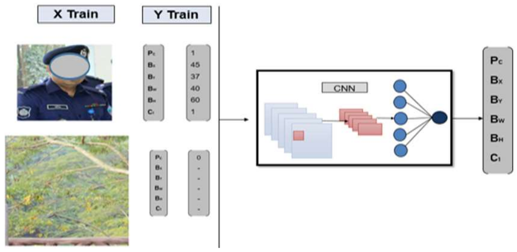

Fig 3.6: Training phase diagram

Then train the neural network in that way

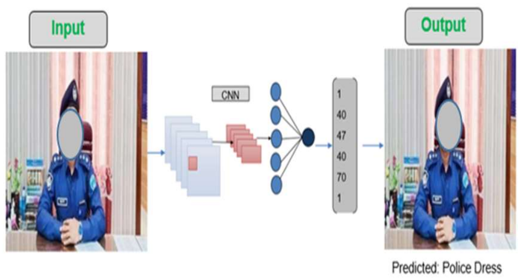

Fig 3.7: Prediction by CNN Model

Now finally If we input a new image, it will find out the particular vector and that vector will say to us is that shield or non-shield image and also make a bounding box when seems it is a shield.

XI. EXPERIMENTAL RESULT AND DISCUSSION

A description is implemented in data science as a programmer, software component, or other computer system using computational thinking. It is the undertaking that converts the conceptual framework into a useful system. The development, setup, and implementation system are all part of the maintenance sequence. Making sure a new system is functioning properly and reliably is crucial to its establishment as a viable one. This proposed system is the end product of the deployment, whose precision we can gauge. We can take the necessary actions for more study depending on the performance.

The material used in the analysis was divided into training, validation, and testing using a split ratio of around 70%, 10%, and 20%. Precision for our research was 99.82%, validation accuracy was 97.71%, and test accuracy was 97.60% is shown in Fig 4.2.1 and the margin of train-data to test-data to validation-data is shown graphically in Fig 4.2.2.

200 epochs were utilized to teach the intended system. The learning was over when 200 epochs were completed. Compared to validation accuracy and test accuracy, training accuracy is improved.

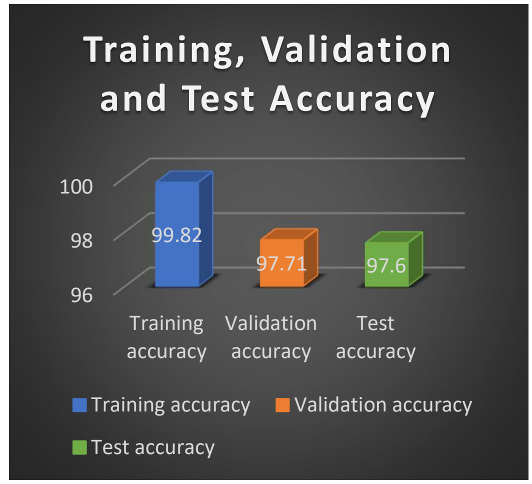

Fig 4.2.1: Training, Validation and Test Accuracy

Since our model was trained to perceive them, our method can detect rotated, shifted, zoomed, superimposed, and veiled

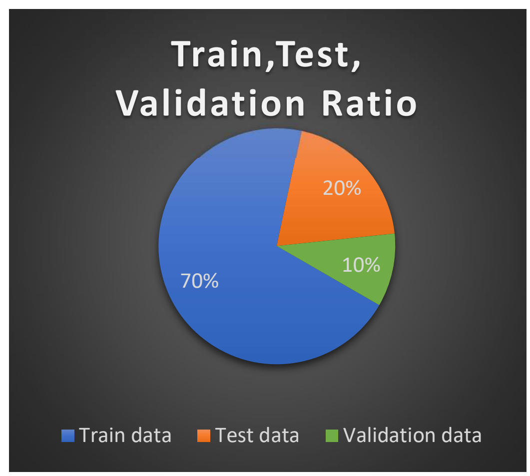

Fig 4.2.2 : Dataset splitting ratio

pictures. Without using enhancement, we were able to attain test accuracy of 97.60%, validation accuracy of 97.71%, and training accuracy of 99.82%. Even though the model's training accuracy was greater than its validation and test accuracy, it was unable to distinguish improved real-time inputs.In Table 4.1 showes the architecture components of the model and the outputs shapes and param of those models.After training, when the engine has been unable to recognize and certify, we obtain our findings. We utilized the validation dataset, which is a set of data patterns produced by our training phase. In order to measure model competence while adjusting parameters of the model, we use it. While we employed 200 epochs,

Table 4.2 shows the training loss, training accuracy, validation loss, and validation accuracy for selected epochs. Brackets, Figure 4.2.3 shows the accuracy of the training and validation of our model, whereas Figure 4.2.4 shows the accuracy decrease throughout training and validation.

Layer (type)	Output Shape	Param #
sequential (Sequential)	(32, 256, 256, 3)	0
conv2d (Conv2D)	(32, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(32, 127, 127, 32)	0
conv2d_1 (Conv2D)	(32, 125, 125, 64)	18496
max_pooling2d_1 (MaxPooling 2D)	(32, 62, 62, 64)	0
conv2d_2 (Conv2D)	(32, 60, 60, 64)	36928
max_pooling2d_2 (MaxPooling 2D)	(32, 30, 30, 64)	0
conv2d_3 (Conv2D)	(32, 28, 28, 64)	36928
max_pooling2d_3 (MaxPooling 2D)	(32, 14, 14, 64)	0
conv2d_4 (Conv2D)	(32, 12, 12, 64)	36928
max_pooling2d_4 (MaxPooling 2D)	(32, 6, 6, 64)	0
conv2d_5 (Conv2D)	(32, 4, 4, 64)	36928
max_pooling2d_5 (MaxPooling 2D)	(32, 2, 2, 64)	0
flatten (Flatten)	(32, 256)	0
dense (Dense)	(32, 64)	16448
dense_1 (Dense)	(32, 2)	130

Table 4.1: Model Layers, Output shape and Param

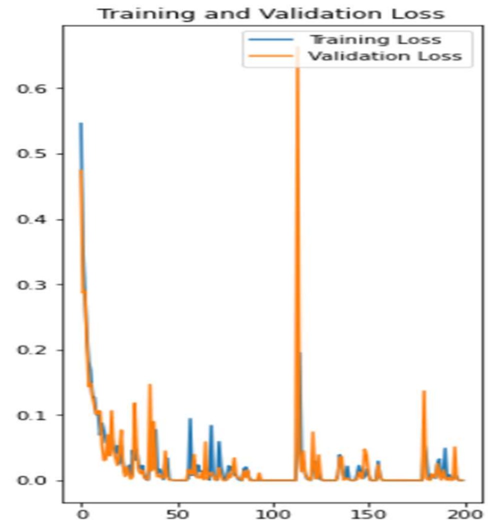

Fig 4.2.4: Training and validation

Here is the summery of the model layers and outputs:

Total params: 183,682

Trainable params: 183,682

Non-trainable params: 0

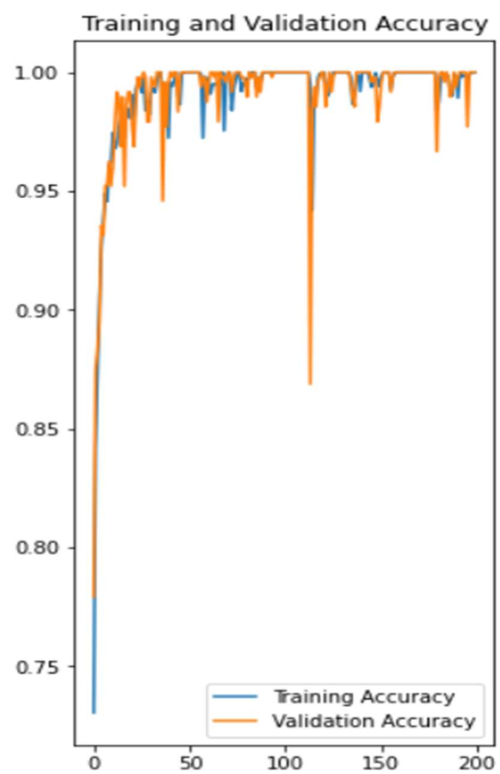

Fig 4.2.3 : Training and validation accuracy

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.6747	0.6044	0.6198	0.6042
50	0.0049	0.9987	0.1264	0.9635
100	0.0379	0.9867	0.0114	1.0000
150	3.4594	1.0000	0.0608	0.9844
200	1.2607	1.0000	0.0095	0.9948

I. System Simulation and Output

The random variable inputs in simulation are typically not precisely understood, although the model is frequently. Inputs are precisely known in machine learning, but the model is unknown before training. The variations in production are relatively subtle. Both provide an output, but there are various sources of uncertainty. The evaluation of data produced by a simulation is known as output analysis. Its goal is to compare the performance of two or more different system designs or to anticipate how well a system will function. We loaded the trained model that was previously stored using Google Colab as the backend. We utilized the CNN technique with the Keras API and Python as the editor and programming language. Here, we used Google Drive to store our own customized dataset. The dataset is then generated using code for further processing. The model effectively recognizes and predicts the Bangladeshi police uniform. As shown in Fig. 4.3.1, our proposed system can accurately predict the class of an input image.

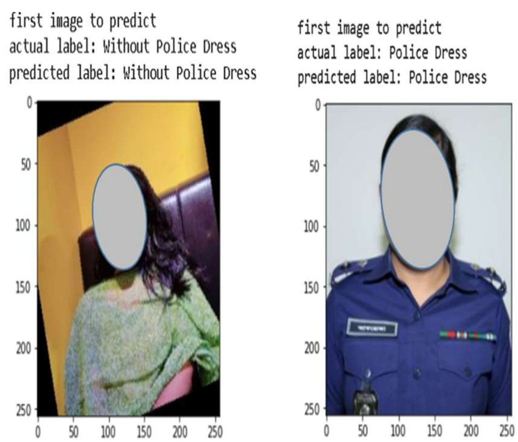

Fig 4.3.1: First prediction

Fig 4.3.2: Initial Test

When predicting the likelihood of a specific outcome, machine learning refers to prediction as the result of an algorithm that has been trained on past data and applied to current data. Here, in Fig 4.3.1 shown the prediction of image of the testing dataset. Actually, the left side image was in the class of without police dress and this model predict it as a level of without police dress. On the other hand, the right-side image was in the class of with police dress and this model predict it as a level of with police dress. So, we can see that this model predicts the input image properly. In the initial test here, it will successfully identify which image are from dataset folder of Police Dress and which one is from the folder of Without Police Dress.

Fig 4.3.3: Final predicted output

XII. CONCLUSSION

The first-time identification of Bangladeshi police attire serves as the focal point of our little effort. In this study, a CNN model with the Keras API was developed and built using data from the Bangladeshi police uniform sample. Thise is very important and, should it spread, has a wide variety of academic and economic interests. The main focus of this study is the identification of the BD Police uniform, which opens the door for future studies in the same field on other uniforms worn for other purposes, such as those worn by the BD Army, in schools, in hospitals, as security uniforms, etc. However, dealing with the security function to identify permitted admission is the primary issue and duty for it. The lack of comparable studies in the Bangladeshi uniform highlights the chance to study this intriguing topic. In future we will try to add other dresses to identify also in realtime. We will also try to make this system to make alarm when it will identify an unouthorizd entries. We will make our dataset more enreach to ensure the better accurecy and also we will shere this dataset as future researcher can use this dataset.

XIII. REFERENCES

[1] Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Application of deep learning for object detection. Procedia computer science, 132, 1706-1717.

[2] Sharma, N., Jain, V., & Mishra, A. (2018). An analysis of convolutional neural networks for image classification. Procedia computer science, 132, 377-384.

[3] Nedeljkovic, I. (2004). Image classification based on fuzzy logic. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34(30), 3-7.

[4] Zagoruyko, S., Lerer, A., Lin, T. Y., Pinheiro, P. O., Gross, S., Chintala, S., & Dollár, P. (2016). A multipath network for object detection. arXiv preprint arXiv:1604.02135.

[5] Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., & Sun, J. (2018). Detnet: A backbone network for object detection. arXiv preprint arXiv:1804.06215.

[6] Mukhometzianov, R., & Carrillo, J. (2018). CapsNet comparative performance evaluation for image classification. arXiv preprint arXiv:1805.11195.

[7] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[8] Zhang, H., Wang, Y., Dayoub, F., & Sünderhauf, N. (2020). Swa object detection. arXiv preprint arXiv:2012.12645.

[9] Al-Doski, J., Mansorl, S. B., & Shafri, H. Z. M. (2013). Image classification in remote sensing. Department of Civil Engineering, Faculty of Engineering, University Putra, Malaysia, 3(10).

[10] Ming, Q., Zhou, Z., Miao, L., Zhang, H., & Li, L. (2021, May). Dynamic anchor learning for arbitrary-oriented object detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 3, pp. 2355-2363).

[11] Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing, 28(5), 823-870.

[12] Jiang, B., Luo, R., Mao, J., Xiao, T., & Jiang, Y. (2018). Acquisition of localization confidence for accurate object detection. In Proceedings of the European conference on computer vision (ECCV) (pp. 784-799).

[13] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6569-6578).

[14] Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014). Scalable object detection using deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2147-2154).

[15] Guo, C., Fan, B., Zhang, Q., Xiang, S., & Pan, C. (2020). Augfpn: Improving multi-scale feature learning for object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12595-12604).

[16] Hall, D., Dayoub, F., Skinner, J., Zhang, H., Miller, D., Corke, P., ... & Sünderhauf, N. (2020). Probabilistic object detection: Definition and evaluation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1031-1040).

[17] Foody, G. M. (2008). Harshness in image classification accuracy assessment. International Journal of Remote Sensing, 29(11), 3137-3158.

[18] He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (2019). Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 558-567).

[19] Xu, H., Lv, X., Wang, X., Ren, Z., Bodla, N., & Chellappa, R. (2018). Deep regionlets for object detection. In Proceedings of the European conference on computer vision (ECCV) (pp. 798-814).

[20] Vani, A. K., Raajan, R. N., Winmalar, D. H., & Sudharsan, R. (2020, March). Using the keras model for accurate and rapid gender identification through detection of facial features. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) (pp. 572-574). IEEE.

[21] Benbrahim, H., Hachimi, H., & Amine, A. (2020). Deep convolutional neural network with tensorflow and keras to classify skin cancer images. Scalable Computing: Practice and Experience, 21(3), 379-390.

[22] Jiang, Z. (2019, December). A novel crop weed recognition method based on transfer learning from VGG16 implemented by keras. In IOP Conference Series: Materials Science and Engineering (Vol. 677, No. 3, p. 032073). IOP Publishing.

[23] Li, A., Li, Y. X., & Li, X. H. (2017). Tensor flow and keras-based convolutional neural network in CAT image recognition. In Proc. 2nd Int. Conf. Comput. Modeling, Simulation Appl. Math.(CMSAM) (p. 5).

[24] Boncolmo, S. D., Calaquian, E. V., & Caya, M. V. C. (2021, November). Gender Identification Using Keras Model Through Detection of Face. In 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (pp. 1-6). IEEE.