ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse931022021.pdf
https://doi.org/10.30534/ijatcse/2021/931022021

Criminal Detection: Study of Activation Functions and Optimizers

Chaithra Sudheer[1], Nithya M S[2], Nandana Vijayan[3], Sreelekshmi S R[4],Vishnuraj N R[5], Libin M Joseph[6]
[1] PG Student, Department of Computer Applications -Saintgits College of Engineering Pathamuttom Kottayam Kerala, India, mailtomechaithra@gmail.com
[2] PG Student, Department of Computer Applications - Saintgits College of Engineering Pathamuttom Kottayam Kerala, India, nithyamohananpillai1997@gmail.com
[3] PG Student, Department of Computer Applications - Saintgits College of Engineering Pathamuttom Kottayam Kerala, India, nvnair2@gmail.com
[4] PG Student, Department of Computer Applications - Saintgits College of Engineering Pathamuttom Kottayam Kerala, India, sreelekshmisr01@gmail.com
[5] PG Student, Department of Computer Applications - Saintgits College of Engineering Pathamuttom Kottayam Kerala, India, vishnurfor@gmail.com
[6] Assistant Professor, Department of Computer Applications - Saintgits College of Engineering Pathamuttom Kottayam Kerala, India, libin.mj@saintgits.org

ABSTRACT

Face is the primary means of recognizing a person, transmitting information, communicating with others, and inferring people's feelings, among others. Our faces will reveal more than we think. A facial image may show personal characteristics such as ethnicity, gender, age, fitness, emotion, psychology, and occupation. In addition to the recent specialisation of deep learning models, the exponential output and memory space growth of computer machines has greatly increased the role of images in recognising semantic patterns. Facial photographs can reveal those personality features in the same way as a textual message on social media reveals the author's individual characteristics. We investigate a new degree of image comprehension by using deep learning to infer a criminal proclivity from facial images. A convolutional neural network (CNN) deep learning model is used to differentiate between criminal and non-criminal facial images. Using tenfold cross-validation on a set of 5500 face pictures, the model's confusion matrix, training, and test accuracies are registered. In learning to achieve the highest test accuracy, CNN was more reliable than the SNN, which was 8% better than the SNN's test accuracy. Finally, CNN's dissection and visualization of convolutional layers showed that CNN distinguished the two sets of images based on the shape of the face, eyebrows, top of the eye, pupils, nostrils, and lips. In this project we focus on Activation functions and optimizers. Activation functions are of two types Saturated and Non-Saturated. Here we use non saturated activation functions like ReLU, SELU and SOFTMAX. When we combine ReLU and SOFTMAX, we get 99.3 percentages as test accuracy. By combining SELU and SOFTMAX we get 99.6 as test accuracy. Therefore, SELU and SOFTMAX combination give the better accuracy.

Key words: Activation Functions, Convolutional Neural Networks, Deep Learning,

1. INTRODUCTION

In addition to the recent specialisation of deep learning models, the exponential output and memory space growth of computer machines has greatly increased the role of images in recognising semantic patterns. Deep learning models are used to distinguish between criminal and non-criminal facial images. Deep learning is an artificial intelligence function that simulates the operations of the human brain in data processing for object detection, speech recognition, language translation, and decision-making. Convolutional neural networks (CNNs or ConvNets) are a form of deep neural network used in visual imagery research. Here we focus on Activation functions and optimizers for better accuracy.

Machine learning (ML) is the empirical analysis of algorithms and mathematical models used by computer systems to perform a particular task, relying instead on patterns and inference, without using explicit instructions. It's classified as an artificial intelligence sub-discipline. Machine learning algorithms construct a mathematical model from sample data, referred to as "training data," in order to make predictions or decisions without having to be explicitly programmed. Biological neural networks are responsible for the organisation of the human brain, and the term Artificial Neutral Network is derived from them. Artificial neural networks' neurons are often connected in different layers of the networks, similar to how neurons in the human brain are linked. Nodes are the names given to these neurons.

A CNN (Convolutional Neural Network) is a deep learning algorithm that recognises computer vision features in images. It's a multi-layer neural network that can process visual inputs and perform tasks such as image recognition, segmentation, and object detection, which may be useful in self-driving vehicles. CNNs can also be used in healthcare for deep learning applications including medical imaging. A CNN consists of several types of layers:

- Convolutional layer - Create a feature map by using a filter to predict the class probabilities for each function by scanning the entire image, a few pixels at a time.
- Pooling layer (down sampling) -Reduces the amount of data generated by the convolutional layer for each function while retaining the most relevant data (the process of the convolutional and pooling layers usually repeats several times).

- Fully connected input layer - "Flattens" previous layer outputs into a single vector that can be used as an input for the next layer.
- Fully connected layer - Apply weights to the output of the function analysis to predict an effective mark.
- Fully connected output layer - Generates the final probabilities for the image to determine a class.

1.1 Activation Functions

The performance of a neural network is determined by activation functions, which are mathematical equations. Depending on whether the input of each neuron is relevant to the model's prediction, each neuron in the network has a role that decides whether or not it should be triggered ("fired"). Activation functions are often used to normalise a neuron's output to a range between 1 and 0 or -1 and 1. Numeric data points, known as inputs, are fed into the neurons in the input layer of a neural network. Each neuron has a weight, which is multiplied by the input number to decide the neuron's output, which is then passed to the next layer. The activation function is a mathematical "gate" that sits between the input and output of the current neuron, which is then sent to the next layer. A phase function, for example, may turn the neuron output on and off based on a rule or threshold. It may also be a mapping of input to output signals that the neural network requires to function.

Sigmoid Function

The sigmoid function is a logistic function, which means that anything you put in would result in a number between 0 and 1. To put it another way, each neuron, node, and activation you join will have a scale of 0 to 1.

$$sigmoid(a) = \sigma = \frac{1}{1+e^{-a}}$$

The Sigmoid function has a "S"-shaped curve or sigmoid curve characteristic, is continuous, distinguishable, has a non-zero derivative everywhere and takes real value as input and gives output between 0 and 1.

Hyperbolic Tangent (Tanh) Function

The hyperbolic tangent (tanh) function is also S-shaped, continuous and distinguishable, but its output value differs from -1 to +1, making each output layer more or less centered around 0.

$$\tanh a = \frac{\sinh a}{\cosh a} = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}} = \frac{e^{2a}-1}{e^{2a}+1}$$

Since it has gradient 1 near the origin, the tanh function is stronger than the sigmoid function. The tanh function, like the sigmoid activation function, saturates at -1 or +1 when the input becomes broad (negative or positive) with the derivative extremely close to zero.

ReLU Function

The rectified linear activation function, also known as the short ReLU, is a piecewise linear function that will directly output the input if it is positive and zero otherwise. For ReLU, this is the equation

$$ReLU(a) = \max(0, a)$$

The ReLU equation tells us this:

- Set input to 0 if the input an is less than 0.
- Set input equal to input if the input an is greater than 0.

Scaled Exponential Linear Unit (SELU) Function

SELU is a version of the scaled ELU activation function. Two fixed parameters are used, namely α and λ, and their value is extracted from the inputs. The proposed values for standardized inputs, however (mean of 0 and standard deviation of 1) α=1.6733, λ=1.0507.

$$selu(a) = \lambda \begin{cases} a \; if \; a > 0 \\ \alpha e^{a} - \alpha \; if \; a < 0 \end{cases}$$

The key advantage of using SELU is that it provides self-normalization (that is, through output from SELU activation, the average of 0 and standard deviation of 1 will be maintained) and this solves the issue of gradients disappearing or exploding.

Softmax Function

The Softmax function computes the probability distribution for 'n' different events. In general, this function can calculate the probability of each target class for all possible target classes. The calculated probabilities will be useful later in deciding the target class for the given inputs.

$$Softmax(a_i) = \frac{\text{Exp}(a_i)\, i = 0,1,2,\ldots,k}{\sum_{j=0}^{k} Exp(a_i)}$$

The performance probability list is the main benefit of using Softmax. The number is 0 to 1, and the sum of all the possibilities equals one. The range is 0 to 1, and the total number of possibilities is one. When the softmax function is used for the multi-classification model and the target class is given a high probability, the probabilities of each class are returned.

2.2 Optimizers

Optimizers are algorithms or methods that change the features of a neural network, such as weights and learning rate, to minimise losses.

RMSProp (Root Mean Square Propagation)

By using a moving average of the square gradient, RMSProp attempts to overcome the radically declining learning rates of Adagrad. To normalize the gradient, it utilizes the magnitude of the recent gradient descents. The learning rate is automatically changed in RMSProp, and each parameter chooses a different learning rate. RMSProp divides the learning rate by the squared gradient exponential decay average.

$$\theta_{t+1} = \theta_t \quad \frac{\eta}{\sqrt{(1-\gamma)g^2{}_{t-1} + \gamma g_t + \varepsilon}} \cdot g_t$$

γ is the decay term that takes value from 0 to 1. g_t is moving average of squared gradients.

Adaptive Moment Estimation (Adam)

Another technique slows down Adagrad's rapidly decreasing learning rates by measuring the person's adaptive learning for each parameter based on estimates of the first and second gradient moments. Adam can be used in conjunction with Adagrad to create sparse gradients in online and non-stationary environments, and RMSprop, which works well. Instead of a simple average as in Adagrad, to scale the learning rate, Adam uses an exponential moving average of the gradients. It keeps the average of previous gradients decaying exponentially. Adam is computationally powerful and needs very little memory. The Adam optimizer is one of the most widely used algorithms for gradient descent optimization. Adam algorithm first updates the exponential moving averages of the gradient (m_t) and the squared gradient (v_t) these are the first and second moment figures. Hyper-parameters β1, β2 ∈ [0, 1) control the exponential decay rates of these moving averages as shown below

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$$

m_t and v_t are estimates of first and second moment respectively.

Moving averages are set to zero at the start, resulting in moment estimates that are skewed toward zero, particularly in the early time steps. This initialization bias is easily overcome, resulting in bias-corrected estimates.

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \qquad \hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

$\hat{m}_t$ and $\hat{v}_t$ are bias corrected estimates for the first and second moments

Finally, we make the required changes to the parameter as shown below,

$$\theta_{t+1} = \theta_t \quad \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

2. LITERATURE REVIEW

Criminal tendency detection from facial images and the gender bias effect

The exponential growth in computer efficiency and memory space, coupled with the recent specialisation of deep learning models, has significantly increased the role of images in the identification of semantic patterns, according to this paper [1]. Facial photos can show those personality features in the same way as a textual post on social media exposes individual attributes in order to suit the blogger. This research is the first step toward inferring personality characteristics from facial images. With this end goal in mind, we investigate a new level of image comprehension by using deep learning to infer criminal tendencies from facial images. Researchers used two deep learning models to distinguish between criminal and non-criminal facial images: a normal feed forward neural network (SNN) and a convolutional neural network (CNN). Using tenfold cross-validation on a collection of 10,000 images of the face, the uncertainty matrix, training and test accuracies for both models were reported. In learning to achieve the highest test accuracy, CNN was more reliable than the SNN, which was 8% better than the SNN's test accuracy. Then, to test the classifier's possible gender bias, we only used male facial images. There were no major variations in classification accuracies or learning accuracy, implying that the classifier was not biased towards women. Finally, CNN's dissection and visualisation of convolutional layers showed that CNN distinguished the two sets of images based on the shape of the face, eyebrows, top of the eye, pupils, nostrils, and lips. It takes care to identify people in any way, but to predict whether a person is a criminal demands even more caution and diligence and must be looked at with suspicion [1]. This technology's danger lies in its imperfection, because misclassifying people can have severe consequences. It would be too naive to argue that the 97 percent accuracy of the test, obtained by CNN in this job, can easily be generalized to face shots from any other source. This is due to a variety of factors, including the small size of our dataset and the fact that various sources come from both criminal and non-criminal images. As a result, the conditions in which the photographs were taken are not identical, raising the question of whether the deep classifier has captured this disparity in peripheral conditions in order to discriminate arbitrarily between the two classes. Both face shots, both criminal and non-criminal, will be taken with the same camera and under the same conditions in the ideal dataset, i.e. Illumination, perspective, distance, history, resolution, mustache, hat, glasses, and make-up.

Multi-Faces Recognition Process Using Haar Cascades and Eigen face Methods

The input image, a face, is detected using a Haar Cascades technique, which is then used to extract features from the eyes and mouth using Sobel edge detection. [2]This project proposes methods for a faster and more efficient face recognition process. A hybrid process of Haar Cascades and Eigen face methods was used to carry out the proposed face recognition process, which can detect multiple faces (55 faces) in a single detection process. With an accuracy of 91.67 percent, this improved approach to face recognition was able to recognise multiple faces [2]. Face recognition has been used as an authentication process in a variety of fields, especially in computer security-related activities, such as home security, building access security, criminal identification, and user identification on small mobile devices. Face recognition is also important in the fields of biometrics and computer vision. A face recognition system's objective is to provide a marginal rate of misclassification. For authentication, biometric technology is used and it can evaluate human behaviour. Each biometric device has its own set of benefits and drawbacks, so careful consideration is needed when selecting one for use in a given application. Within the defined goal, the facial recognition process has been successfully optimised.

Effectiveness of Self Normalizing Neural Networks for Text Classification

The activation function suggested in SNN is the Scaled Exponential Linear Units (SELU) (Self Normalizing Neural Networks) [3]. FNN and CNN use recited linear units (ReLU) as activation in general. ReLU activation clips the negative values to 0 and hence the user's dying problem with ReLU. The activation function should include both positive and negative control mean values, high variance reduction saturation regions and higher than one slope to increase variance if its value is too small. SELU activation is then enforced to maintain the aforementioned properties.

3. METHODOLOGY

This system is actually a research methodology which focuses on variations in model accuracy rate based on activation functions and Optimizers. Criminal Detection" models recognize the face of suspected person in a photo and predict whether that the person is criminal or non-criminal. By using Deep learning model called Convolutional Neural Network (CNN), we discriminate the face as criminal or non-criminal. CNN model Generated with activation function combination Selu, Softmax and Optimizer RMSprop gives an accuracy rate of 99.9 % which contradicts the model in the reference paper with activation functions ReLU, Softmax and optimizer Adam which offers 97% accuracy. Even though the Tanh and Sigmoid ,and, Tanh and Softmax combinations generates accuracy more than 98%,still they may lead to saturation and cause vanishing gradient problem. Hence these combinations are recommended in any of the research papers.

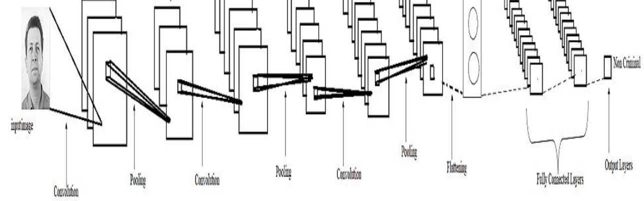

Figure 1: Architecture of the Proposed System

4. RESULTS

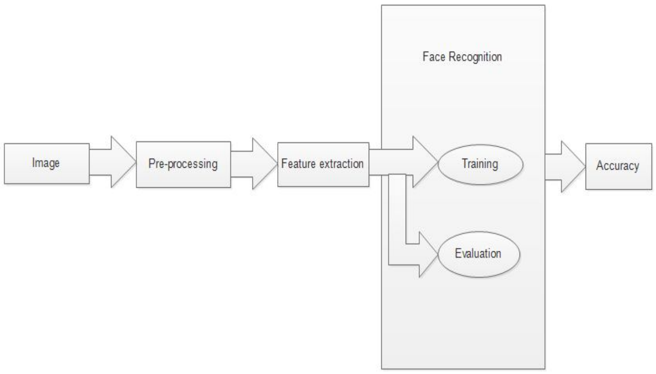

Figure 2: Flow chart of Face Detection

Models	Accuracy
CNN(activation=Tanh, Softmax optimizer=adam)	98.25
CNN(activation=Relu, Sigmoid optimizer=adam)	50
CNN(activation=Selu, Sigmoid optimizer=adam)	50
CNN(activation=Tanh, Sigmoid optimizer=adam)	92.17
CNN(activation=Sigmoid, Relu optimizer=RMS prop)	50
CNN(activation=Selu, Softmax optimizer=adam)	99.13
CNN(activation=Relu, Softmax optimizer=adam)	95.82
CNN(activation=Relu, Softmax optimizer=RMS prop)	97.57
CNN(activation=Selu, Softmax optimizer=RMS prop)	99.61
CNN(activation=Selu, Sigmoid optimizer=RMS prop)	50.5
CNN(activation=Tanh, Sigmoid optimizer=RMS prop)	99.22
CNN(activation=Tanh, Softmax optimizer=RMS prop)	98.45

Figure 3: Result summary model wise

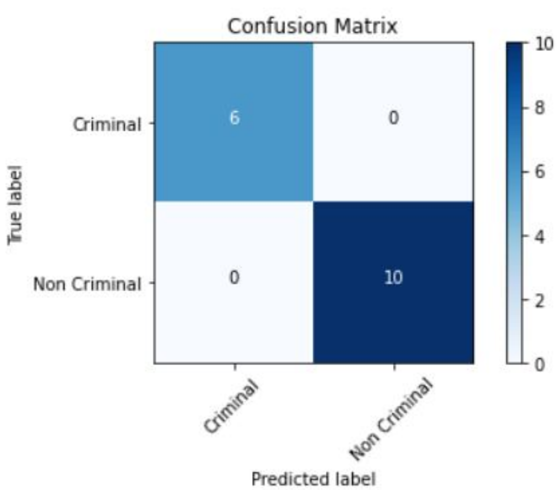

Figure 4: Confusion Matrix of CNN Model with Selu, Softmax, RMSProp Combination

```
img_pred=load_img("img_n.jpg",target_size=(128,128))
plt.imshow(img_pred)
img_pred=img_to_array(img_pred)
img_pred = np.expand_dims(img_pred, axis=0)
rslt=modelCNN.predict(img_pred)
print(rslt)
if rslt[0][0]==[1]:
    prediction="Criminal"
else:
    prediction="Not Criminal"
print("Prediction=",prediction)
```

```
[[0. 1.]]
Prediction= Not Criminal
```

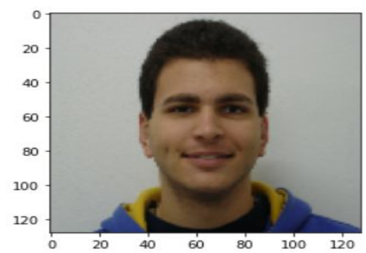

Figure 5: Predicted Image

5. CONCLUSION

Accuracy rate of model generated with activation function combination Selu and Softmax with optimizer RMSprop is much higher than any other model. This project also enhances the efficiency of operations handle by police departments. It can also help us to minimize the terrorist activities and may help to nullify various terrorist attacks happen in the cities. The ability of storing suspect images that were arrested for at least one crime incident and the efficiency in identifying possible criminals for a crime incident are the characteristics that makes Criminal Detection System perfect than other systems.

REFERENCES

1. Teddy Mantoro, Media A Avu **Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods** Faculty of Engineering and Technology Sampoerna University, Jakarta, Indonesia.
2. Mahdi Hashemi and Margret Hill **Criminal tendency detection from facial images and the gender bias effect**
3. Avinash Madasu and Vijjini Avnesh Rao **Effectiveness of Self Normalizing Neural Networks for Test Classification**, Samsung R & D Institute