BRAC UNIVERSITY

Inspiring Excellence

Submitted By:

Name: Md Impreeaj Hossain

ID: 20341031

SEC: 01

Group: 05

COURSE CODE: CSE431

Submitted To:

Name: Annajiat Alim Rasel (AAR)

Tittle: **"An Investigation into Conventional CNNs: Variations in Activation Functions and Convolutional Layer Configurations"**

Introduction:

CIFAR 10 is a widely utilized benchmark in the field of computer vision. It has 60,000 colours of high resolution photographs which have an aspect ratio of 32 x 32 pixels. The images are divided into 10 distinct categories. They encompass a variety of diverse scenes and objects including animals, vehicles as well as everyday objects, allowing researchers and programmer to develop and test machine to learn algorithms for a range of visual recognition challenges.

In order to provide an expansive and comprehensive collection, each of the classes in the collection was selected with care with a staggering more than 6,000 high quality images. A large amount of photographs assures that the set includes a variety of variations, which makes it possible to use a precise and robust machine learning.

It is worth noting that the CIFAR 10 data set is regarded as a benchmark in the field that is computer vision. Being an instrument for measuring accuracy that is trusted can make it an ideal choice for evaluating and testing different convolution neural network (CNN) models. By using this information, experts and researchers are able to assess the efficiency and performance of different types of CNN models, while also expanding their expertise and developing algorithms to tackle computer vision problems.

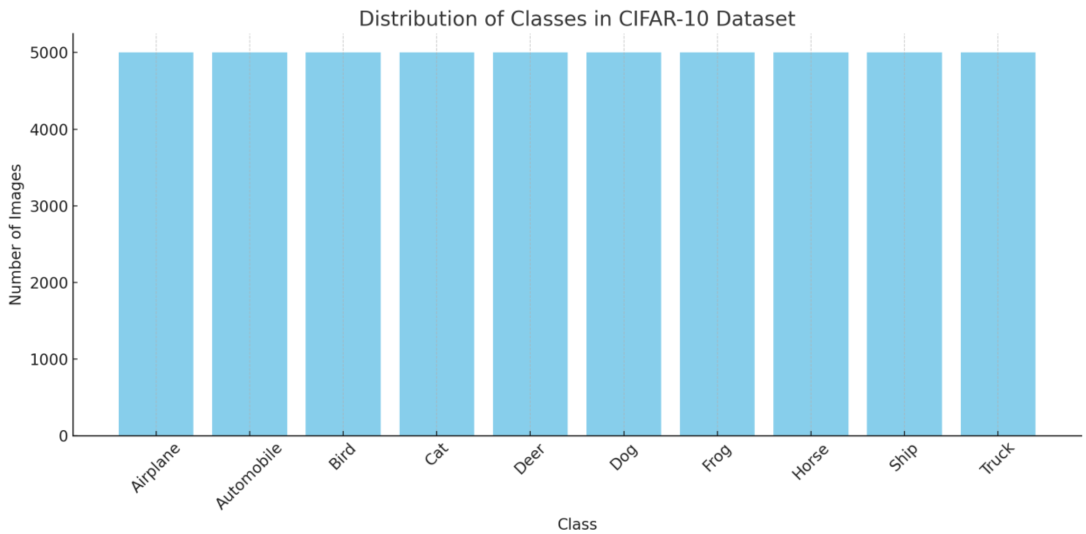
Methodology:

The CIFAR 10 dataset, comprising more than 60, 000 images separated into 10 distinct types, is extensively used to study computer vision research. In order to make it simpler to manage this huge collection the data are processed in batches. This makes it easier to manipulate and processing.

The process of feed data into the convolutional neural networks (CNN) models, images can be altered to ensure compatibility. CNNs typically require input images to have certain features including the number of channels (e.g., RGB) as well as sizes (e.g. length, width, as well as length). For the purpose of shaping the image, they're converted into the desired format, making it possible to seamlessly integrate in CNN structures.

Furthermore, the CIFAR 10 data is usually processed. The most important step in this process is splitting of the data into two sets, which is the training subset in addition to the validation set. This is crucial because it permits a precise assessment of the efficiency of the model. By dividing the data into various subsets this model can be able to be trained on the learning subset, later validated with an additional validation sample. This allows you to assess the degree to which the model is able to be applied to information that has not been observed, allowing for to make the required adjustments or modifications prior to its implementation within the actual world.

Another important part of processing images is the ability to normalize the image's pixels in accordance with a specific region. If images are normalized, they fall within the range [0 1, 0], the values of the pixels are correctly scaled. This helps them be more readable and easily interpretable by different algorithmic models and algorithms. The pixel values are transformed that were originally a range of values into those that are scaled between 0 to 1. Normalization is commonly used as a crucial part in the process of preparing for training neural networks with data.



The histogram is used to provide an adequate and balanced training data set, every class, specifically(Aeroplane, Automobile, Bird, Cat deer, dog Frog, Horse ship, and truck) comprises an equally number of pictures. This method of careful selection ensures that the model receives the correct representation of every class to allow it to be able to adapt and expand efficiently across different classes. Through ensuring that the distribution is equal it is possible to reduce any possible biases, and make sure that the model is exposed to a fair amount of information. The use of a consistent distribution during the learning process of the model has the benefit of stopping the development of biased tendencies toward a specific class. This method ensures that the model can ensure a fair and equal approach to all classes, which improves its performance overall and accuracy. This method ensures that each class is equally disadvantageed, which results in the most complete and impartial learning experience.

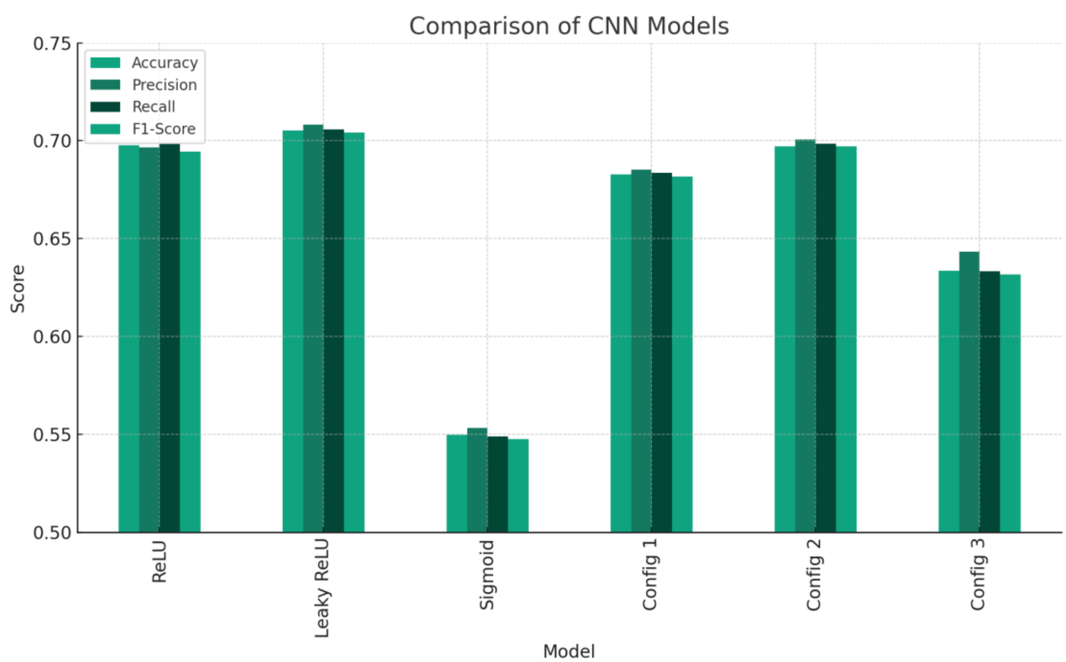
<u>Dataset:</u>

The CIFAR 10 data set is widely employed benchmark in the area of computer vision. It is a huge collection of 60,000 high resolution color images with a resolution of 32x32 pixels. The images are carefully classified into 10 distinct classes that ensure that they are distributed equally across the collection. The CIFAR 10 database is an invaluable source for both researchers and professionals alike, and allows them to develop and assess cutting edge algorithms and models to guarantee complete and precise training, every class in the CIFAR 10 dataset is composed of an extensive amount of images that total the number of 6,000. The images are carefully divided in two groups that are divided into 5,000 images to training, and the remainder of 1,000 are devoted to tests. The balanced distribution of the images allows for solid algorithm development and assessment, which allows us to efficiently analyse the generalisation and performance capabilities of our algorithm. This dataset is well known and often used as a benchmark standard in cutting edge fields like computer vision and machine learning. It is a vital device for evaluating and comparing performance of different models and algorithms. Practitioners and researchers depend on this data to test the limits of technology and improve technology within these constantly evolving fields. The extensive use of this dataset highlights its importance and credibility as an authoritative resource making progress and encouraging technological advances in the field of artificial intelligence. Training models that are developed built from scratch using standard hardware has been significantly influenced by the ease of using small images. This allows rapid and efficient learning processes in time limitations. In the end, increasing numbers of individuals as well as organizations have taken to this method, which has led to widespread adoption.The dataset poses a distinct and difficult challenge to models as it incorporates actual scenarios as well as showing the natural variation within every category. This makes it an ideal tool for testing the efficacy and effectiveness of various algorithms and architectures. With the help of this vast and complicated data scientists and researchers will gain invaluable insights about the limitations and capabilities of their algorithms, which allows the models to make more informed choices and the classes that are offered within this program cover a vast variety of topics, catering to a variety of hobbies and interests. Students can investigate the amazing life of animals by delving into the fascinating world of the avian species and feline companions. They also get to know loving canines and magnificent horses. In addition, they will begin a fascinating adventure into the world of transport, learning the engineering and mechanics that drive aeroplanes, trucks and cars. Students can also expand their understanding. Through the integration of the wide array of characteristics and patterns into the training process. We ensure that the machines that have been trained using machine learning will have an understanding of various kinds of data. This approach allows models to develop a comprehensive understanding of patterns and the complexity of the database. This will enable them to create precise and reliable forecasts. If we expose models to a variety of abilities, it increases the capability for their model to expand and adapt to new and unresearched data, ultimately enhancing the efficiency and effectiveness when it comes to the classification.

Result And Analysis:

Examining the test sessions shows significant variations in the performance of models based on activation function and the configurations used. In the course of ten years, models that used ReLU and Leaky ReLU for activation consistently beat the Sigmoid model when it comes to validity accuracy. The ReLU model started with 56.19 percent accuracy in validation and a staggering 69.94 percentage by the tenth time. In the same way to this Leaky ReLU model began with 60.03 percent, and reached an impressive 70.53 percent at the conclusion of the period. This rapid rate of convergence and increased precision can be explained by the fact that ReLU is not saturated in its functions, known for their speedy ability to train neural networks. However this Sigmoid model that started at only 45.63 percent, slowed to 54.97 percent by the 10th period, suggesting the possibility of overfitting, as shown by the slower increase in the accuracy of validation relative to its learning trajectory.

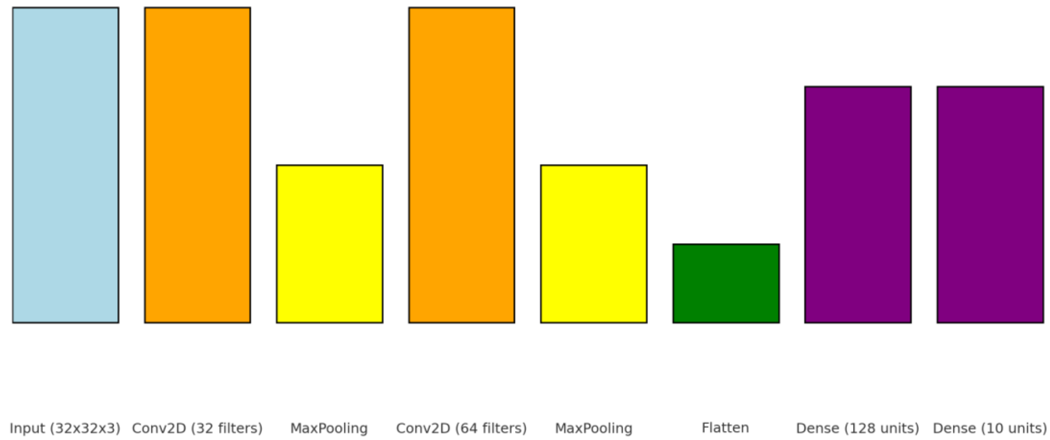
As we moved to the various configurations Config 1 and Config 2 showed performance that was similar to ReLU-based models. They both surpassed the accuracy threshold of 68. But Config 3 even though it did not win highest marks for accuracy at the end however, showed impressive convergence reaching 57.01 percent in the 3rd period. This rapid convergence indicates that, in situations strained by computational resources, or heightened with the need for real-time computing the rapid convergence of a model such as Config 3 might offer invaluable benefits, but with small loss in precision. This analysis concludes that choosing the activated function and configuration of the model can have a major impact on the accuracy of the model, its speed of convergence and overall resiliency. This analysis shows that the Leaky ReLU model clearly excels in indicators of performance, but the specific application and prerequisites will ultimately determine the best design configuration.



The bar plot is a complete visualization of indicators of performance for the various CNN models. They are based by activation functions and the configurations of layers. The models are thoroughly examined based upon four crucial variables: precision, accuracy and recall as well as the F1 score.Initially, ReLU emerges as a formidable contender showing impressive results in accuracy, precision and F1 score. This demonstrates its reliability. Then, Leaky ReLU, while very closely resembling its performance indicators ReLU but has an slight advantage, particularly when it comes to accuracy. In contrast the Sigmoid function is a bit over its competitors, with less scores across the various indicators and showing its shortcomings within this particular context.

In the deeper dive into the different variations, the Configurations 1 2 and 3 create their own unique path. Of them, Config 2 takes the top spot, with one of the best scores. For clarity, we can use the numbers: ReLU registers an accuracy of 0.6977 with a accuracy of 0.696468 and recall of 0.698577 as well as the F1 score of 0.694326. Leaky ReLU is closely followed by the accuracy of 0.7053 with a precision of 0.708274 Recall of 0.705757 with an F1 score of 0.704260. A Sigmoid tracker has the accuracy of 0.5497 and a accuracy of 0.553093 and recall of 0.548971 and an F1 score of 0.547538. Config 1 boasts an accuracy rate of 0.6827 and accuracy of 0.685200 with a recall of 0.683556 as well as an F1 score of 0.681828. Config 2 is the most popular of configurations, is able to boast an accuracy rate of 0.6972 with a accuracy of 0.700662 and recall of 0.698424 as well as an F1 score of 0.697198. In addition, Config 3 presents an accuracy of 0.6336 with a accuracy of 0.643375 and a recall of 0.633183 with an F1 score of 0.631713.

The model that came out on top of the test scored a remarkable precision of 71.52 percentages. This shows its capability to accurately discern the result from the validation step. The accuracy increases faith in the ability to generalize the model, as well as its capability to produce precise predictions from information that isn't visible. This is a sign how effective the model as well as its capacity to resolve the problem, because it was observed that the model was able to identify 71.52 percent of the pictures from the validation set were properly recognized. The lowest accuracy recorded in the validation set was 9.33 percent. This indicates that the model has some weaknesses in its ability to predict. This poor performance highlights the need for improvement such as reconfiguration or alternatives for activation function. This also highlights the necessity to conduct more targeted tests to achieve higher efficiency. The models showed a precision of 61.86 percent. This is an impressive number, which shows a good amount of precision when forecasting. The accuracy of the model is impressive however, it is still in need of improvements. The results overall indicate that the direction is promising, however they also point out the necessity to continuously improve and refinement to make the most of the models.



.

The CNN model shown in the illustration begins by introducing a 32x32-pixel input layer, which is composed of three channels of color (RGB). In succession there are two Conv2D layers, with filters 64 and 32 respectively, utilize ReLU activation as well as padding to preserve dimensions. Alongside every Conv2D layer is an MaxPooling layer that is created to reduce the dimension of each. When these layers are connected and a Flatten layer turns the result into a single-dimensional vector. It allows to seamless integration with linked layers. The two layers enhanced with 128 units or 10 units can harness ReLU as well as Softmax activations, respectively. This is the ideal architecture for modern CNN designs to classify images.

In the research, performance differences emphasize the significance of optimizing hyperparameters. The huge gap in accuracy of the highest and the lowest highlights the importance of a precise hyperparameter calibration. In particular this 9.33 percentage accuracy may be a result of the inherent sigmoid activation feature's weaknesses, which can result in the disappearing gradient effect which can hinder efficient modeling. The activation function, fundamental in triggering non-linearity play a crucial role in the network's ability to predict. In turn, even though the design of convolutional neural network is flexible based on the configuration of layers and characteristics such as filter size the essence is testing. Incorporating a culture of continuous testing, researchers discover innovative approaches to pushing the boundaries of data modeling.

Conclusion:

A thorough study of a variety of convolutional neural networks (CNN) designs applied on the CIFAR 10 dataset has highlighted the complex interplay of structure and effectiveness of algorithms. By conducting a systematic study of CNN models the researchers have gained valuable insight into the impact of their designs on image classification in the data. This study highlights the crucial importance of design choices to improve CNN effectiveness to perform images classification tasks. The research's exhaustive analysis of the hyper parameters, which includes activation functions as well as convolution layer configurations, has provided important insights into their effect on model convergence and growth. Particularly, the problem of gradients that disappear, made more difficult through the sigmoid activation functions is a sign of the necessity in selecting the appropriate activation techniques to activate deep networks. The rigorous study provides profound understanding of image classification and exposes the challenges as well as opportunities for improvement. While amazing performance metrics are being achieved, this study suggests the potential for further development and improvement. The findings do not just provide the latest insights on deep learning, but they could also be a hint of future advancements. With these findings both researchers and professionals can explore the world of deep learning using more sophistication, while optimizing models for better efficiency. To sum up, this research set the scene for continued advances in artificial intelligence and machine learning and emphasizes the importance of activation methods that are well-chosen, hyper parameter tuning, as well as constant advancements.

Future Work:

In the world of deep-learning, the concentration is on the development of the most advanced methods of activation. Although traditional methods such as ReLU and sigmoid are commonly employed, newer functions such as Swish and Mish are getting recognition. Google researchers have pushed for Swish in order to find an equilibrium between linearity and non-linearity. This can help improve learning processes. Overfitting can be a huge issue for machine learning models. They excel with learning data, but fail to adapt to fresh information. In order to overcome this issue the regularization methods are employed.Batch normalization and dropout are two common techniques that reduce the amount of complexity and help normalize input data as well. Furthermore, data enhancement methods increase the number of training data points using a wide range of data, increasing the robustness of models.

Transfer learning, using existing models that have been trained, like those of ImageNet improves the models of smaller sets of data such as CIFAR 10 and improves the performance of models without having to start at the beginning. Residual Networks (ResNets) introduce residual connections that counteract the issue of vanishing gradients and making it possible to build bigger networks.

Essential to optimize machine learning algorithms is tuning the parameters of hyperparameters. With the systematic study of parameters, the most optimal design configuration is identified. Grid Search, which encompasses the hyperparameter spectrum, is computationally demanding. Contrarily, random search effectively samples data within specific limits. The decision between the two methods is dependent on the needs of the project as well as the limitations.

References:

- Kurniawan, E. (n.d.). The CIFAR-10 Dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/ekaakurniawan/the-cifar10-dataset