Conducting an Analysis of the Effectiveness of CNNs in Handling Image Recognition Tasks

Dr. Jitendra Singh Chauhan[1], Ms. Anisha Sheikh[2], Mr. Ankur Rao[3]
Head, Professor, Department of Computer Application, Geetanjali Institute of Technical Studies, Udaipur, Rajasthan[1]
Assistant Professor, Department of Computer Application, Geetanjali Institute of Technical Studies, Udaipur, Rajasthan[2,3]

hodmca@gits.ac.in [1], anisha.sheikh@gits.ac.in [2], ankur.rao@gits.ac.in [3]

Abstract -

The purpose of this study is to investigate the capacity of well-known convolutional neural networks (CNNs) to recognise objects in real time. These CNNs include AlexNet, GoogLeNet, and ResNet50. During the inquiry, the accuracy of each network is painstakingly evaluated by making use of benchmark datasets such as ImageNet, CIFAR10, and CIFAR100. The evaluation is carried out only on video data in order to correctly imitate real-time settings. This is an important point to emphasise. According to the results, GoogLeNet and ResNet50 demonstrate greater performance in comparison to AlexNet. Furthermore, significant variances in performance were identified across a variety of object categories according to the findings. Object detection inside CNN architectures is examined in the subsequent discussion, which sheds insight on the complexities of the subject matter. Potential causes that contribute to these performance discrepancies are investigated.

Keywords: **ResNet, GoogLeNet, AlexNet, CNN, Object Detection, Image Recognition, CIFAR10.**

I. INTRODUCTION

Modern digital technology has made complex search algorithms and programmes possible by the abundance of images and videos on the internet. Through more sophisticated and comprehensive overviews of search results, these technologies hope to enhance user experiences by delving into the semantic analysis [1] of visual content. Globally, a great number of studies have reported major advances in scene classification, item recognition, and picture labelling [2, 3]. Further advances in methods that tackle problems with scene categorization and object identification have been made possible by these discoveries. The ground-breaking performance of artificial neural networks—more especially, CNN [4, 5], [6]—in the domains of scene categorization and object recognition has attracted a lot of interest. Finding the network architecture most appropriate for the above described tasks is thus the main goal of this study. These methods start at the critical stage of feature extraction. Accurately capturing the variations across the many item categories requires the extraction of a condensed collection of characteristics from the raw pixel values of a photo. An essential component of this procedure is played by many conventional feature extraction techniques. Among them are Content-Based Image Retrieval (CBIR) [11], Local Binary Patterns (LBP) [10], Histogram of Oriented Gradients (HOG) [8], and Scale-Invariant Feature Transform (SIFT) [7]. These techniques allow one to get a wealth of information about a thing or scene from a few amount of image pixel values. The characteristics are categorised immediately after they are extracted based on the objects seen in the image. It is common procedure to utilise classifiers like Support Vector Machines (SVM), Logistic Regression, Random Forests, Decision Trees, and others for this purpose. These classifiers raise the effectiveness and precision of the research even further.

The field of image analysis has been revolutionised by the use of CNNs, which have shown exceptional performance in tasks including identification, segmentation, detection, and retrieval. One of the numerous applications for its versatility is the classification of objects, as well as the identification of facial expressions and gestures, and the description of scenes. CNNs are able to efficiently and reliably demonstrate outstanding detection rates across a variety of datasets. CNNs, for instance, perform very well, as seen by their 99.77% detection rate on the MNIST dataset and their 97.47% detection rate on the NORB dataset. A significant number of factors have contributed to this achievement, including the availability of enormous labelled datasets like as ImageNet and CIFAR, as well as improvements in deep learning techniques. CNNs are able to significantly enhance their accuracy and generalisation capabilities by fine-tuning their classification performance across millions of photos, which is made possible by these datasets. There is a correlation between the variety and how well they perform on different tasks and datasets. Understanding the nuances of the approaches used to extract features from each CNN model is thus necessary in order to achieve the goal of optimising the performance of each CNN model in certain applications and domains.

The purpose of this research is to investigate the accuracy of CNNs in scene classification based on elements that have been identified. During training, datasets such as CIFAR-100, CIFAR-10, and ImageNet are used, whilst testing encompasses a wide variety of video genres. Through an in-depth analysis of the changes in performance between training photos and real-time video inputs, this study contributes to a better understanding of CNN's representation and learning processes. The findings highlight the importance of object-based image representation for the purpose of enhancing high-level visual identification, particularly in conditions that are complex and include a large number of components. This is because CNNs are so effective at extracting low-level information, which makes them an indispensable tool for improving scene representation. The substantial training that they have received on massive datasets that include millions of photographs has provided them with a solid foundation for scene classification. The study advocates item detection as a feature of scene representation by using the layered architectures of CNNs. The purpose of these features is to evaluate the accuracy of detection over a wide range of complicated real-world scenarios. After doing a study of several prior research, the work is structured in such a way that the problem statement and the proposed solution are detailed in great detail. A comprehensive understanding of the chosen networks and datasets contributes to the enrichment of the study. In-depth investigation of the findings from many datasets demonstrates that CNN's performance varies from one dataset to another. The conclusions of the study provide a concise summary of the results of the work, as well as suggestions for the routes that additional inquiry should take. The findings of this work contribute to our understanding of the performance of CNNs in object-based representation as well as the challenges associated with scene classification based on real video.

1. Related Work

Convolutional neural networks, often known as CNNs, are now being used as powerful tools in a broad variety of domains and applications. A few of the first applications of CNN architecture included the recognition of handwritten numbers [17]. When working on projects like the ImageNet Challenge, convolutional neural networks are widely used, and they are usually employed in conjunction with different sketch dataset combinations [19]. Furthermore, a few studies have tried to establish a comparison between the detection capabilities of human participants and trained networks using image datasets.

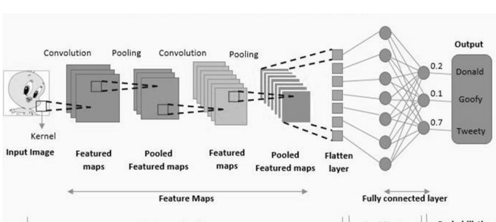

Fig1 Internal Layers of CNNs

results of these comparative research have produced several interesting developments. It was observed, for instance, that a trained CNN provides a little lower accuracy rate of 64%, but a human observer reaches a 73.1% accuracy rate. This was discovered using the dataset. These findings highlight the subtle differences that exist between machine-based recognition and human perception, as well as the remarkable capabilities that CNNs possess when it comes to applications that involve doing picture analysis.

When CNN was applied to the same dataset, their astounding accuracy rate of 74.9% was higher than the performance of humans [21]. Error rates are significantly improved as a result of the use of stroke order in the procedures that are utilised. In the present study [20], the primary emphasis is on gaining a more in-depth understanding of the behaviour of Deep Neural Networks in a variety of settings. The results of these experiments demonstrate how even little adjustments to photos may have a big influence on the results of grouping. In addition, the research displays photographs that are unrecognisable to human eyes but are effectively categorised by trained networks with an extraordinary degree of accuracy [20]. In line with the development of feature detectors and descriptors, there has been a substantial advancement in the algorithms and methodologies used for object and scene categorization. It is interesting to note that the promotion of commonality among object detectors, texture filters, and filter banks is often emphasised throughout the process. Context classifiers developed by Hoeim and present descriptors developed by Felzenszwalb are often used by researchers since there is a wealth of information available on item identification and scene categorization [3].

Work that is being done in the multimedia sector, where various "semantic concepts" are utilised for semantic indexing, image and video annotations [22], is comparable to the work that is being done in the field of fundamental picture interpretation, which involves the construction of several object detectors. In studies that are relevant to our issue, each semantic concept is often taught via the use of either photos or video frames. This makes it challenging to apply and comprehend the concepts when they are presented in pictures that have crowded surroundings. In order to identify and classify individual things, the majority of the previously developed algorithms relied on feature sets that were established by human labour. In the field of scene categorization, however, more recent approaches focus on analysing the relationships that exist between the various elements [3]. The use of a number of different scene categorization techniques, the utility of the object bank has been assessed. Histogram of Oriented Gradient (HOG), GIST, filter banks, and Bag of Features (BoF) coupled with word vocabulary are all examples of low-level feature extraction approaches that have been researched for the purpose of identifying and classifying objects [4]. All of these initiatives, when considered together, contribute to the growing area of image analysis and increase our understanding of the capabilities of neural networks in performing tasks involving visual recognition.

Methodology of Evaluation

Our primary objective is to investigate the performance of the neural network on both live video streams and static photos. The first thing that we are doing is applying transfer learning to neural networks that are utilising image datasets. After that, we compare and contrast the accuracy of the predictions made by static photos and real-time video sources for the identical things. The various accuracy rates are meticulously documented and shown in the tables that are included in the following sections. One of the most important aspects of performance assessment is determining whether or not the CNN architectures that we have employed in our study have different levels of accuracy in their predictions. Due to the fact that we are interested in locating the most effective image classifier, which makes use of the object as the primary feature for scene classification, movies are used as testing datasets rather than training datasets. In this study, we make use of a convolutional neural network that is composed of the five layers.

The 'input layer' is the first stage of any CNN, and its primary function is to reduce the size of images before passing them on to subsequent layers for feature extraction. The images are then filtered by the Convolution Layers in order to extract features and in order to compute matching feature points during testing. After that, the feature sets that have been extracted are sent to the "pooling layer," which optimises the value of each window in order to preserve the best possible fit for each feature. Additionally, the pooling layer simplifies large photos while preserving the essential information. In order to maintain mathematical stability, the 'Rectified Linear Unit' (ReLU) layer then replaces zero for negative values that were obtained from the pooling layer. This prevents the learned values from either increasing to infinity or remaining at zero. In the end, the Fully Connected Layer is responsible for converting high-level filtered images into categories that include labels that correspond to those categories.

Developing a model of a neural network involves going through a number of important steps. In the beginning, the training and testing datasets are produced by resizing superclass photos to [224, 244] pixels for AlexNet and [227, 227] pixels for GoogleLeNet and ResNet50. These photos are classified as belonging to two distinct categories: training datasets and validation datasets to be more specific. In order to modify the architecture of the CNN, the final three layers are then substituted with a fully connected layer, a softmax layer, and a classification output layer. The size of the fully connected layer is then adjusted to correspond with the number of classes that are present in the training dataset. Increasing the learning rate parameters of the fully connected layer is another way to speed up the process of training the network. The next step is to train the network, which involves tweaking the training parameters to take into account the GPU specifications of the system. These parameters include the learning rate, the mini-batch size, and the validation data. Following the completion of training, the network is correctly assessed by virtue of the classification of validation pictures and the evaluation of the classification accuracy. In addition, the enhanced network is tested on real-time video feeds in order to validate its resilience and effectiveness in applications that are used in the actual, physical world.

Models

There are a great number of powerful pre-trained CNNs that are accessible in the deep learning field. All that is required to adjust training and testing datasets at the input layer is their ability to transfer learn. The internal designs of these networks are considerably distinct from one another, and they use a variety of diverse strategies. For example, GoogleNet implements Inception Modules, which are responsible for performing convolutions of varying sizes and concatenating the results. These modules are used to feed data into subsequent layers. AlexNet, on the other hand, does not use filter concatenation and instead transfers the output from one layer directly into the next layer. Caffe, a well-known framework for deep learning, is used in the construction of both networks. ResNet, also known as Residual Network, is a significant advancement in the field since it simplifies the process of training extremely deep neural networks, which are crucial for enhancing visual recognition tasks.

Deep learning has been characterised by the inclination to construct deeper architectures in order to tackle increasingly challenging issues and improve the accuracy of recognition and classification over the course of time. With all of its benefits, deeper networks sometimes lead to more complicated training processes, which eventually result in a plateau in performance development, if not a worsening of the situation. Residual learning is a technique that addresses these issues. It employs skip connections to enable layers to learn residual functions with regard to the layer inputs. As a result, it enables the training of deeper networks without compromising performance. Deep convolutional neural networks, which often consist of several layers, are able to acquire the ability to identify a variety of properties at low, medium, and high complexity levels. Because of its hierarchical learning structure, deep learning models are particularly successful for a wide range of tasks. Additionally, it continues to push the boundaries of what computers are able to identify and comprehend.

When it comes to learning individual features, learning residuals, which is the difference between the input of a layer and the feature that has to be learnt, takes priority over learning specific features simultaneously. This strategy is shown by ResNet, which makes use of shortcut connections to link the input of a layer (n) to the output of a layer (n+x) quickly. As a result, the network is able to skip one or more layers. The use of shortcut connections makes it simpler to train very deep networks. This is accomplished by addressing the problem of vanishing gradients and avoiding the accuracy drop that is often seen in deeper networks that do not have such procedures. In order to better understand ResNet's efficacy and robustness in handling deeper architecture, comparisons have been made with other well-known neural networks such as AlexNet and GoogleNet. The fact that these comparisons demonstrate how modifications to the network architecture may have a significant impact on performance is shown by the use of the fundamental concepts of transfer learning in order to adapt these networks to new tasks. New models have been built as a consequence of this, and although they have a comparable number of layers to the original networks, they have radically different performance characteristics nevertheless. The ensuing outputs and performance metrics of these networks are provided in the form of tables that highlight the various accuracy rates that may be achieved with the same set of photographs across many models. This comparative study not only proves the usefulness of residual learning in ResNet, but it also exhibits the flexibility and potential of pre-trained models via transfer learning. This research serves as an essential benchmark for future developments in neural network design and implementation.

Test Datasets

The CIFAR-100 picture collection is comprised of a wide variety of generic object photographs that are organised into a variety of super-classes, with each super-class section including numerous subclass categories. CIFAR-100 is comprised of 100 separate picture classes, each of which has 600 pictures. So, to be more specific, there are a total of 60,000 photographs that are completely unique. Within each class, there are a total of 500 training photographs and 100 testing images that are selected and divided further. In a crucial step, the 100 classes are merged into twenty superclasses, and each image in the dataset is assigned a "fine" label that indicates the class to which it belongs, as well as a "coarse" label that indicates the superclass of the "fine" label. Mattresses, bicycles, buses, chairs, sofas, motorcycles, streetcars, tables, trains, and wardrobes are only some of the items that fall under the training and testing categories. Other items include tables, trains, and wardrobes.

Certain broad categories from each superclass have been selected for the purpose of training the networks. In particular, the superclasses of automobiles and home goods have been included in this selection. A table that is included with this article provides a compilation of these categories for your convenience. In addition to the CIFAR-100 dataset, the ImageNet dataset is also a significant dataset that is used in combination with it. This dataset also splits super-classes of photos into subclasses that are more precise. Through the process of organising data into meaningful concepts in line with the WordNet hierarchy, ImageNet makes it feasible to handle object identification and classification in a complete manner. Using a big and varied collection of photos, this organised dataset arrangement makes it possible to investigate more challenging picture identification issues. It enhances the capability to train models that are more complex and accurate. A "synonym set" or "synset" on WordNet is a collection of words that collectively describe a certain concept. This provides a complex technique of semantic understanding thanks to the collection of terms. Approximately one hundred thousand of these synsets are included in the dataset; each synset is representative of a different concept. There is consistent and high-quality tagging across the collection as a result of the fact that individuals have diligently tagged each and every photo in the collection. During the course of this effort, we also addressed the issue of enhancing ImageNet's categorization by reorganising labels that were less descriptive into groups that were more pertinent and that better suited the superclasses that were already in place. Images that were once categorised as "table" were, for instance, transferred to the more generally applicable "furniture" category. Because of this reclassification, it is now feasible to create superclasses that provide a framework that is both more relevant and evocative for arranging and analysing the photos.

Our final dataset consisted of the photos that were collected by the CIFAR-10 project. There are 60,000 photographs included in the CIFAR-10 dataset, much as its larger sibling, the CIFAR-100 dataset. However, the CIFAR-10 dataset is separated into ten classes, which makes it simpler to use but does not diminish its effectiveness for specific picture recognition applications. Each CIFAR-10 class is comprised of a total of 6,000 32x32 colour photos, including 50,000 training images and 10,000 testing images. Every single one of the five training batches and the one test batch that comprise the overall dataset has a total of 10,000 photographs among them. Random photographs are selected from each category in order to ensure that the evaluation of the model's performance across a variety of visual representations is conducted in a manner that is both fair and unbiased. Methods and applications for machine learning are being developed as a result of the effective training and testing of image recognition models that is made feasible by this organised approach to the organising of datasets.

Results

In order to accurately evaluate the effectiveness of CNNs, it is necessary to conduct comprehensive testing on established benchmark datasets such as CIFAR-100 and CIFAR-10. The practical efficacy and utility of these networks are critically shown via real-time testing. For example, AlexNet achieves a mere 13% accuracy in correctly identifying items in a given scene. In contrast, GoogleNet and ResNet50 have much better classification accuracy rates of 72.63% and 63.42%, respectively. The discrepancy in performance across the networks highlights the differences in picture processing and classification techniques.

Empirical evidence demonstrates that CNNs may exhibit confusion when distinguishing between closely related categories. ResNet50 often misidentifies canines and cervids as equines. This kind of error indicates challenges in distinguishing visually indistinguishable categories, which is a common issue in many deep learning models. It highlights the need to enhance the network's ability to detect subtle changes in features. Despite these challenges, GoogleNet consistently outperforms the other networks that were assessed, achieving the highest level of detection accuracy. This suggests that the architecture of the model allows it to outperform other models in accurately detecting and categorising images, perhaps because to features such as Inception modules that effectively handle visual cues at different sizes. These evaluations are crucial since they not only demonstrate the current capabilities of neural networks but also guide future advancements in network design and training methods.

Evaluation

Both CNNs generate a probability distribution over all of the potential input classes, which allows their classification decisions to be rationalised mathematically. For the purpose of conducting an analysis of these facts, two distinct methods are used. The first method involves recording the frequency of each class within these top ranks for each and every image that falls into each of the target categories. This is done after the network's predictions are evaluated based on their probability rankings. Taking this technique makes it possible to do a qualitative study of the consistency of the results across all of the categories, in addition to assessing whether or not a substantial probability is assigned to the correct conclusion. When it comes to the capacity of the network to make accurate predictions, the top ten probabilities for each category should basically remain relatively unchanged, indicating that the network is reliable and consistent.

In the second method, a more comprehensive statistical analysis is performed, during which the precise placement of the appropriate class within the overall probability distribution is scrutinised in great detail. According to this method, the results that are produced by the classifier are evaluated in order to determine the frequency with which the appropriate class appears at a higher rank. One indicator of improved performance is a decrease in the standard deviation, which also suggests that there is consistency across samples that belong to the same category. The documentation of the best-case and worst-case possibilities within each category is also made simpler by this technique, which may be extremely beneficial in appreciating and resolving the reasons behind particular findings. This method is helpful since it makes it easier to record the alternatives. The acquisition of significant new information on the performance of a variety of networks is made possible by the broad use of these approaches to a large number of datasets. Based on the results of our experimental inquiry, we provide the following average performance for the CIFAR-100 dataset: AlexNet has an average of 46.90%, whereas ResNet50 has a record of 63.91%, and GoogleNet has a record of 71.66%. Based on the CIFAR-10 dataset, AlexNet achieves a comparable performance of 42.15%, GoogleNet achieves 76.63%, and ResNet50 achieves the highest scoring of 81.62%. These data, which also demonstrate the many advantages and disadvantages of each network, demonstrate the importance of selecting the appropriate network architecture in accordance with the specific requirements and challenges of the task at hand. The optimisation of neural network installations is made easier by this advanced analysis, which is used in a wide variety of practical applications.

Conclusion

This study conducted an in-depth investigation on the accuracy of prediction offered by three CNNs: ResNet50, GoogleNet, and AlexNet. The CNNs were tested on the CIFAR-10 and CIFAR-100 datasets, each of which had ten classes. An evaluation of the consistency and accuracy of these networks' performance on object recognition tasks was the major purpose of this study. Complex conditions have been demonstrated to provide neural networks with considerable obstacles and varying accuracy rates. This is especially true with respect to furniture like as beds, couches, and chairs, which might have varied forms and layouts in different photos. A primary finding that emerged from the research was that transfer learning contributes to an improvement in the performance of these networks. Methods of transfer learning that were used in networks demonstrated much greater rates of accuracy. This improvement was largest in deeper layer networks, which suggests that there is a link between depth and the capacity to successfully extract and generalise from complicated characteristics in the training data. In light of this new finding, it is clear that deeper network topologies have the potential to handle photo recognition jobs that are becoming more difficult. By demonstrating the growing adaptability and efficiency of CNNs, this study highlights the huge potential that CNNs have in the actual categorization of items. Despite the fact that CNNs have very low hardware requirements, they are becoming more useful for a wide variety of applications across a variety of platforms as a result of ongoing technology developments and algorithmic improvements. The availability of this accessibility is essential in order to deploy sophisticated photo recognition algorithms in situations where the power of the computer may be restricted but high accuracy is still required. When everything is taken into consideration, the study highlights the expanding potential of CNNs in a variety of real-world settings where accurate and speedy picture identification is vital.

REFERENCES

[1] Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., &Babu, R. V. (2016) “A taxonomy of deep convolutional neural nets for computer vision.”

[2] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., &Torralba, A. (2014) “Object detectors emerge in deep scene cnns.”

[3] Wang, Y., & Wu, Y. “Scene Classification with Deep Convolutional Neural Networks.”

[4] Lowe, D. G. (2004) "Distinctive image features from scale-invariant keypoints." *International journal of computer vision***60(2).**

[5] Dalal, N., &Triggs, B. (2005, June) "Histograms of oriented gradients for human detection."*In Computer Vision and Pattern Recognition, 2005. CVPR 2005.*

[6] Yang, J., Jiang, Y. G., Hauptmann, A. G.,& Ngo, C. W. (2007, September)"Evaluating bag-of-visual-words representations in scene classification." in Proceedings of *the international workshop on Workshop on multimedia information retrieval*.

[7] Cheung, Y. M., & Deng, J. (2014, October) "Ultra local binary pattern for image texture analysis." in Security*Pattern Analysis, and Cybernetics (SPAC), 2014 International Conference*.

[8] Khan, S. M. H., Hussain, A., &Alshaikhli, I. F. T. (2012, November) "Comparative study on content-based image retrieval (CBIR)." in Advanced Computer Science Applications and Technologies (ACSAT), 2012 *International Conference*.

[9] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997) "Face recognition: A convolutional neural-network approach."*IEEE transactions on neural networks, 8(1)***:**98-113.

[10] K. U. Singh, A. Kumar, G. Kumar, T. Singh, T. Choudhury and K. Kotecha, "Optimizing Handwritten Numeral Recognition for English and Devanagari Using MNIST and CPAR Data," 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS), Istanbul, Turkiye, 2023, pp. 1-6, doi: 10.1109/ISAS60782.2023.10391793.

[11] Bobić, V., Tadić, P., &Kvaščev, G. (2016, November) "Hand gesture recognition using neural network based techniques." in Neural Networks and Applications (NEUREL), *2016 13th Symposium on (pp. 1-4). IEEE.*

[12] Krizhevsky, A., & Hinton, G. (2009)"Learning multiple layers of features from tiny images."

[13] LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., ... &Vapnik, V. (1995) "Learning algorithms for classification: A comparison on handwritten digit recognition."*Neural networks: the statistical mechanics perspective*(pp 261-276).

[14] LeCun, Y., Bottou, L., Bengio, Y., &Haffner, P. (1998) "Gradient-based learning applied to document recognition." proceedings of the *IEEE* **86(11)**: 2278-2324.

[15] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., &Salakhutdinov, R. (2014) "Dropout: a simple way to prevent neural networks from overfitting."*Journal of machine learning research* **15(1)**: 1929-1958.

[16] Eitz, M., Hays, J., & Alexa, M. (2012) "How do humans sketch objects?"*ACM Trans. Graph., 31(4)*.

[17] Ballester, P., & deAraújo, R. M. (2016, February)"On the Performance of GoogLeNet and AlexNet Applied to Sketches." in *AAAI*.

[18] Yang, Y., &Hospedales, T. M. (2015)"Deep neural networks for sketch recognition".

[19] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., &Fei-Fei, L. (2014)"Large-scale video classification with convolutional neural networks." in Proceedings of the IEEE Conference on *Computer Vision and Pattern Recognition*.

[20] Ciresan, Dan, Ueli Meier, and JurgenSchmidhuber, (2012) ""Multi-column deep neural networks for image classification."*2012 IEEE Conference on Computer Vision and Pattern Recognition.*

[21] Ciresan, Dan, Ueli Meier, Jonathan Masci, Luca M. Gambardella, and JurgenSchmidhuber. (2011) "Flexible, High Performance Convolutional Neural Networks for Image Classification." *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence-Volume Two:* 1237–1242.

[22] Lawrence, Steve, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. (1997) "Face Recognition: A Convolutional Neural Network Approach." *IEEE Transactions on Neural Networks, Volume 8; Issue 1*.