ISSN No:-2456-2165

Convolutional Neural Networks for Image Classification

Jasmin Praful Bharadiya
Doctor of Philosophy Information Technology,
University of the Cumberlands, USA

Abstract:- Deep learning has recently been applied to scene labelling, object tracking, pose estimation, text detection and recognition, visual saliency detection, and image categorization. Deep learning typically uses models like Auto Encoder, Sparse Coding, Restricted Boltzmann Machine, Deep Belief Networks, and Convolutional Neural Networks. Convolutional neural networks have exhibited good performance in picture categorization when compared to other types of models. A straightforward Convolutional neural network for image categorization was built in this paper. The image classification was finished by this straightforward Convolutional neural network. On the foundation of the Convolutional neural network, we also examined several learning rate setting techniques and different optimisation algorithms for determining the ideal parameters that have the greatest influence on image categorization.

Keywords:- *Convolutional neural network, Deep Learning, Transfer Learning, ImageNet, Image classification; learning rate, parametric solution.*

I. INTRODUCTION

Image classification in computer vision is important for our education, jobs, and daily life. Images are classified using a procedure that includes image preprocessing, image segmentation, key feature extraction, and matching identification. With the aid of the most modern image classification techniques, we are now able to acquire image data more quickly than ever before and put it to use in a number of fields, including face recognition, traffic identification, security, and medical equipment. In order to address the shortcomings of the conventional approach of feature selection, feature extraction and classifier have been merged into a learning framework with the emergence of deep learning. The goal of deep learning is to identify several layers of representation with the expectation that high-level characteristics will capture the data's more ethereal semantics. Using Convolutional architectures in image classification is a crucial component of deep learning. The anatomy of the mammalian visual system serves as inspiration for convolutional neural network. Hubel and Wiesel suggested a visual structure model based on the cat visual brain in 1962. For the first time, the idea of a receptive field has been put out. In 1980, Fukushima presented the first hierarchical framework Neocognition would utilise to analyse pictures. In order to achieve network translation invariance, Neocognitionutilised the local connection between neurons.

There are several deep learning architectures available. Convolutional neural networks, the most effective and practical deep neural network for this sort of data, were utilised to create the model reported in this research, a classifier system. As a result, CNNs that have been trained on huge datasets of pictures for recognition tasks may be used to their advantage by applying these learning representations to tasks that need less training data.

Since 2006, a variety of techniques have been created to get around the challenges involved in training deep neural networks. Krizhevsky suggests a traditional CNN architecture Alexnet and demonstrates a considerable advancement over earlier approaches to the picture classification job. Numerous initiatives to boost Alexnet's performance have been recommended in light of its success. VGGNet, GoogleNet, and ZFNet are suggested.

- **Hierarchical Feature Extraction**: CNNs excel at learning hierarchical representations of images. They consist of multiple layers, including convolutional layers and pooling layers, that progressively extract features at different levels of abstraction. This hierarchical approach allows CNNs to capture intricate patterns and structures in images, leading to more accurate classification.

II. KEY REASONS FOR THE SIGNIFICANCE OF CNN

A. Translation Invariance:

CNNs are designed to be translation invariant, meaning they can recognize patterns regardless of their location in an image. This is achieved through the use of convolutional layers that apply filters to an image, detecting features regardless of their position. This property enables CNNs to classify images regardless of their orientation or position, making them more robust and accurate in real-world scenarios.

B. Data Efficiency:

CNNs require fewer training examples than traditional machine learning algorithms. They can learn from a small number of examples due to their ability to capture relevant features and generalize to unseen data. This property makes CNNs ideal for scenarios where large amounts of labeled data are not available.

C. Transfer Learning:

CNNs are capable of transfer learning, meaning they can learn from one task and transfer that knowledge to another related task. This is achieved through the use of pre-trained models, which are trained on large datasets, and can be fine-tuned for specific image classification tasks. Transfer learning reduces the amount of training data required and can lead to significant improvements in classification performance.

D. Scalability:

CNNs are scalable, meaning they can be used for image classification tasks with varying levels of complexity. This scalability is due to their ability to add or remove layers, adjust the number of filters in each layer, and change the size of the filters used in convolutional layers. This flexibility makes CNNs suitable for a wide range of applications, from simple image classification to more complex tasks such as object detection and segmentation.

To achieve this goal, the objectives below have been specified:

- Design a flexible gene encoding scheme of the architecture, which does not constrain the maximal length of the building blocks in CNNs. With this gene encoding scheme, the evolved architecture is expected to benefit CNNs to achieve good performance in solving different tasks at hand.
- Investigate the connection weight encoding strategy, which is capable of representing tremendous numbers of the connection weights in an efficient way. With this encoding approach, the weight connection initialization problem in CNNs is expected to be effectively optimized by the proposed GA.
- Develop associated selection (including the environmental selection), crossover, and mutation operators that can cope with the designed gene encoding strategies of both architectures and connection weights.
- Propose an effective fitness measure of the individuals representing different CNNs, which does not require intensive computational resources.
- Investigate whether the new approach significantly outperform the existing methods in both classification accuracy and number of weights.

III. METHODOLOGY OF EVALUATION

Our research's major goal is to comprehend how effectively networks operate with both static and real-time video streams. Transfer learning on networks using picture datasets is the initial stage in the next process. The next stage is to execute transfer learning on networks with picture datasets. This is followed by testing the next phase. The prediction rate of the same item on still photos and live video streams is then examined.

The various accuracy rates are noticed, recorded, and shown in the tables provided in subsequent sections. The third crucial factor for judging the performance was to see if there were any differences in prediction accuracy between the CNNs used in the study. Videos are utilised as testing datasets, not as a training dataset, it must be highlighted. As a result, we are searching for the best picture classifier where the object is the primary attribute for scene category categorization.

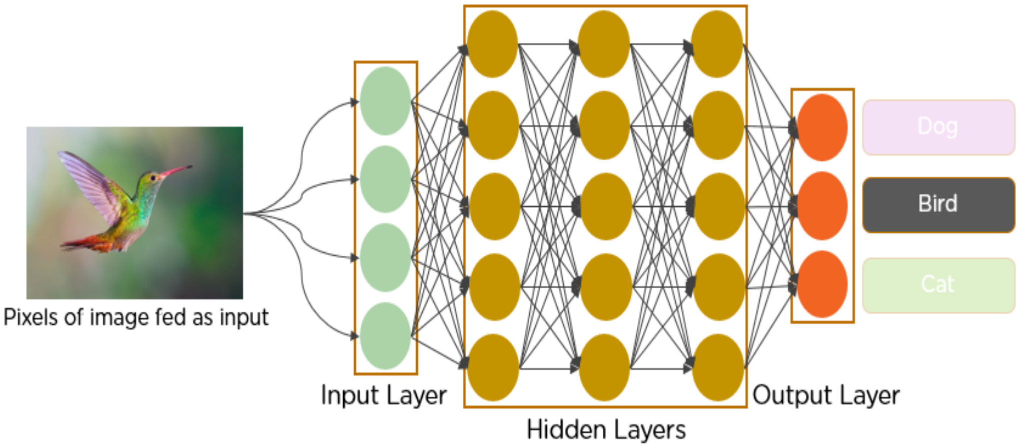

Fig. 1: Primary attribute for scene category categorization

Different layers of the convolutional neural network used are:

- **Input Layer:** The first layer of each CNN used is 'input layer' which takes images, resize them for passing onto further layers for feature extraction.
- **Convolution Layer**: The next few layers are 'Convolution layers' which act as filters for images, hence finding out features from images and also used for calculating the match feature points during testing.
- **Pooling Layer**: The extracted feature sets are then passed to 'pooling layer'. This layer takes large images and shrink them down while preserving the most important information in them. It keeps the maximum value from each window, it preserves the best fits of each feature within the window.
- **Rectified Linear Unit Layer**: The next 'Rectified Linear Unit' or ReLU layer swaps every negative number of the pooling layer with 0. This helps the CNN stay mathematically stable by keeping learned values from getting stuck near 0 or blowing up toward infinity.
- **Fully Connected Layer**: The final layer is the fully connected layers which takes the high-level filtered images and translate them into categories with labels.
- **Basic CNN components:** Convolutional layer, pooling layer, and fully-connected layer are the three major types of convolutional neural network layers.

ISSN No:-2456-2165

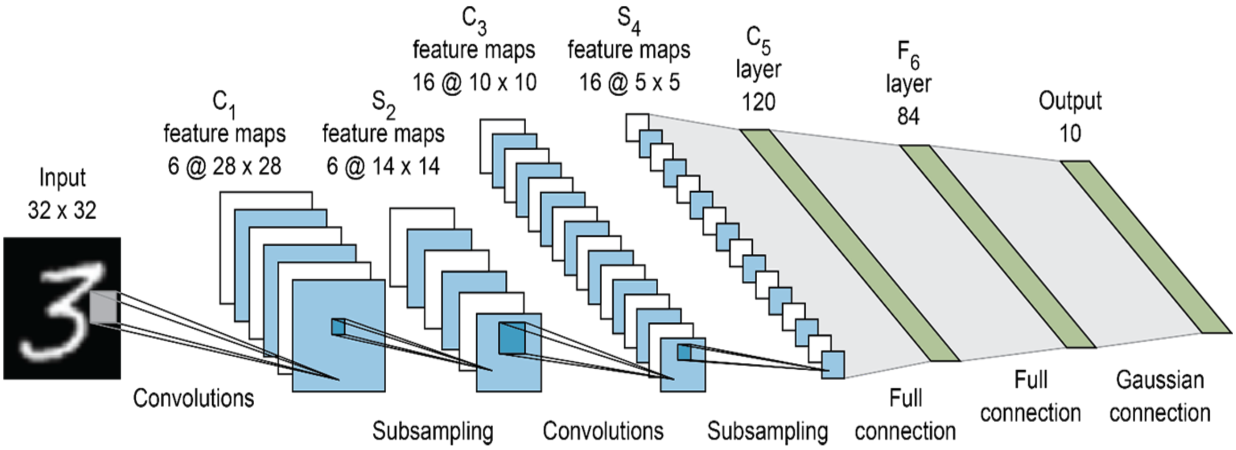

Fig. 2: Basic CNN components

The steps of proposed method are as follows:

- **Creating training and testing dataset:** The super classes images used for training is resized [224,244] pixels for AlexNet and [227,227] pixels GoogLeNet and ResNet50, and the dataset is divided into two categories i.e. training and validation data sets.
- **Modifying CNNs network:** Replace the last three layers of the network with fully connected layer, a softmax layer, and a classification output layer. Set the final fully connected layer to have the same size as the number of classes in the training data set. Increase the learning rate factors of the fully connected layer to train network faster.
- **Train the network:** Set the training options, including learning rate, mini-batch size, and validation data according to GPU specification of the system. Train the network using the training data.
- **Test the accuracy of the network**: Classify the validation images using the fine-tuned network, and calculate the classification accuracy. Similarly testing the fine tune network on real time video feeds for accurate results.

IV. MODELS

There are several intelligent pre-trained CNN; these CNN can transmit learning. Therefore, at its input layer, it just needs the training and testing datasets. The core layers and methods employed in the networks' architecture vary. The Inception Modules in GoogleNet execute convolutions of varying sizes and combine the filters for the following layer. AlexNet, on the other hand, utilises the output of the preceding layer as its input rather than filter concatenation. Both networks have undergone independent testing and make use of the Caffe Deep Learning framework's implementation.

However, as we go further away, neural network training gets challenging and accuracy begins to saturate before declining. Residual Learning makes an effort to address both of these issues. A deep convolutional neural network often has many layers that are layered and trained for the given purpose. At the conclusion of its layers, the network learns a number of low-, mid-, and high-level characteristics. In residual learning, the network tries to learn some residual rather than certain characteristics. Residual is just the feature learnt from the layer's input that is subtracted. ResNet does this through a shortcut connection that connects some (n+x) of the layer's input straight to another layer. The comparison is made among three existing neural networks i.e. the AlexNets, Google Nets and ResNet50. The training of existing networks and the creation of new networks for additional comparison are then followed by the transfer learning ideas. The new models have the same number of layers as the original models, but their performance differs greatly from that of the old networks. The tables in the next section provide the varied accuracy rates that were calculated on the identical photos.

V. CONVOLUTIONAL NEURAL NETWORKS (CNNS) FOR IMAGE CLASSIFICATION OF ADVANCEMENTS

- **Attention Mechanisms:** Recent advancements in CNN architectures have introduced attention mechanisms, which enable the network to focus on specific regions or features in an image that are most relevant for classification. You can explore different attention mechanisms, such as self-attention or spatial attention, and their impact on improving the accuracy and interpretability of CNN models.
- **Transformer-Based Architectures**: The success of the Transformer model in natural language processing has led to its adaptation for image classification tasks. Transformer-based architectures, such as Vision Transformers (ViTs), replace convolutional layers with self-attention mechanisms, enabling the model to capture global dependencies in images. You can investigate the performance and scalability of these architectures compared to traditional CNNs.
- **Meta-Learning and Few-Shot Learning**: Meta-learning approaches aim to enhance the ability of CNNs to learn from a few labeled examples by leveraging prior knowledge learned from similar tasks or datasets. Few-shot learning techniques, such as meta-learning, metric learning, or generative modeling, enable CNNs to generalize to new classes with limited training data. You can explore the advancements in meta-learning and few-shot learning for image classification and compare their performance with traditional CNN models.

- **AutoML and Neural Architecture Search:** Automated Machine Learning (AutoML) techniques, specifically Neural Architecture Search (NAS), have gained attention for automatically discovering optimal CNN architectures for image classification. NAS algorithms leverage reinforcement learning, evolutionary algorithms, or gradient-based optimization to search for architectures with improved performance. You can discuss the progress in AutoML and NAS and evaluate their effectiveness in discovering superior CNN architectures.
- **Explain ability and Interpretability:** As CNNs become more complex, understanding the decision-making process of these models becomes crucial. Future advancements in CNNs for image classification should focus on improving interpretability and explainability. You can explore methods like attention visualization, saliency maps, or class activation maps that provide insights into which regions of an image contribute most to the classification decision.
- **Robustness and Adversarial Defense:** CNNs are susceptible to adversarial attacks, where subtle perturbations to input images can lead to misclassification. Future advancements in CNN architectures should address the robustness and security concerns by incorporating defenses against adversarial attacks. You can discuss different defense mechanisms and compare their effectiveness in improving the robustness of CNN models.

VI. FURTHER DISCUSSIONS

We will go into more detail about the proposed EvoCNN method's fitness evaluation, weights-related parameters, and architectures' encoding strategies in this paragraph. The experimental findings are also reviewed, which may offer helpful information about the potential uses of the suggested EvoCNN approach. Mutation operators serve as the exploration search, or the global search, whereas crossover operators serve as the exploitation search, or the local search. Since local and global searches should compliment one another, only properly developing both of them might significantly boost performance. The commonly employed methods for CNN weight optimisation are based on the gradient data. The gradient-based optimizers' sensitivity to the beginning positions of the parameters that need to be optimised is well known. The gradient-based methods are prone to becoming stuck in local minima without a suitable starting point. It seems impossible to identify a better starting point for the connection weights using GAs due to the vast amount of characteristics. As we have seen, a sizable number of factors cannot be successfully optimised or efficiently stored into the chromosomes. An indirect encoding strategy is used in the proposed EvoCNN technique, which simply encodes the means and standard derivations of the weights in each layer. The final classification accuracy is frequently taken into account by methods now in use to find CNN architectures together with an individual's fitness. The training method normally involves several additional epochs, which takes a long time to get a final classification accuracy.

VII. SUMMARY

This article provides an overview of Convolutional Neural Networks (CNNs) for image classification. It begins by highlighting the importance of image classification in various domains and the liitations of traditional feature selection approaches. Deep learning, particularly CNNs, is introduced as a solution to address these limitations.

The article explains that CNNs excel at learning hierarchical representations of images by utilizing convolutional layers and pooling layers to extract features at different levels of abstraction. CNNs offer translation invariance, allowing them to recognize patterns regardless of their location in an image. They are also data-efficient, requiring fewer training examples due to their ability to capture relevant features and generalize to unseen data.

Transfer learning is emphasized as a key capability of CNNs, enabling them to leverage pre-trained models trained on large datasets and fine-tune them for specific image classification tasks. This reduces the amount of training data required and improves classification performance.

Scalability is another advantage of CNNs, as they can be adjusted by adding or removing layers, changing the number of filters, and modifying the size of filters used in convolutional layers. This flexibility makes CNNs suitable for various image classification tasks, from simple classification to complex tasks like object detection and segmentation.

The article outlines the methodology for evaluating CNN performance, which involves training networks on static and real-time video streams, performing transfer learning, and testing accuracy. It mentions different types of CNN layers, including input layers, convolution layers, pooling layers, rectified linear unit (ReLU) layers, and fully connected layers.

Several models and architectures are discussed, such as AlexNet, GoogLeNet, and ResNet50. The article compares their performance and introduces advancements in CNNs, including attention mechanisms, transformer-based architectures, meta-learning, AutoML, and neural architecture search. It also emphasizes the need for explainability and interpretability in CNNs, as well as robustness against adversarial attacks.

VIII. CONCLUSION

In order to autonomously evolve the architectures and weights of CNNs for image classification challenges, a novel evolutionary technique is being developed in this study. By putting forth a new representation for weight initialization strategy, a new encoding scheme for variable-length chromosomes, a new genetic operator for chromosomes with different lengths, a slacked binary tournament selection for choosing promising individuals, and an effective fitness evaluation method to speed up evolution, this goal has been successfully attained. Understanding deep learning is important, and it is useful since training time is limited. Future study will improve our system by incorporating evolutionary algorithms to address the classification feature extraction challenge and reduce the number of parameters required for this operation.

REFERENCES

[1.] Ahonen, T., Hadid, A., and Pietikinen, "M. Face description with local binary patterns: Application to face recognition." Pattern Analysis and Machine Intelligence, 2037–2041. 2016.

[2.] Bharadiya , J. P., Tzenios, N. T., & Reddy , M. (2023). Forecasting of Crop Yield using Remote Sensing Data, Agrarian Factors and Machine Learning Approaches. Journal of Engineering Research and Reports, 24(12), 29–44. https://doi.org/10.9734/jerr/2023/v24i12858

[3.] Bharadiya, J. (2023). Artificial Intelligence in Transportation Systems A Critical Review. American Journal of Computing and Engineering, 6(1), 34 - 45. https://doi.org/10.47672/ajce.1487

[4.] Bharadiya, J. . (2023). A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. European Journal of Technology, 7(1), 58 - 66. https://doi.org/10.47672/ejt.1473

[5.] Bharadiya, J. . (2023). Machine Learning in Cybersecurity: Techniques and Challenges. European Journal of Technology, 7(2), 1 - 14.

[6.] Bharadiya, J. . (2023). The Impact of Artificial Intelligence on Business Processes. European Journal of Technology, 7(2), 15 - 25. https://doi.org/10.47672/ejt.1488

[7.] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "ImageNet: A large-scale hierarchical image database." In CVPR, 2009.

[8.] Hang Chang, Cheng Zhong, Ju Han, Jian-Hua Mao, "Unsupervised Transfer Learning via Multi-Scale Convolutional Sparse Coding for Biomedical Application." IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 janvier 2017.

[9.] Howard, A. , "Some improvements on deep convolutional neural network based image classi¿cation." ICLR, 2014.

[10.] Nallamothu, P. T., &Bharadiya, J. P. (2023). Artificial Intelligence in Orthopedics: A Concise Review. Asian Journal of Orthopaedic Research, 6(1), 17–27. Retrieved from https://journalajorr.com/index.php/AJORR/article/view/164

[11.] Redmon J, and Angelova A, "Real-time grasp detection using convolutional neural networks", IEEE International Conference on Robotics and Automation, pp. 1316–1322, 2015.

[12.] Van de Sande, K. E. A., Gevers, T., and Snoek, C. G. M, "Evaluating color descriptors for object and scene recognition", IEEE Transactions on Pattern Analysisand Machine Intelligence." 1582– 1596. 2010.

[13.] Zhou, X., Yu, K., Zhang, T., & Huang, T. "Image classi¿cation using super-vector coding of local image descriptors." In ECCV,2010.