Using Deep Learning in Image Classification

Saven Hasan Rasool[1], Assist. Prof. Dr. Laith R.Flaih[2]

[1]Department of Computer Science and Information Technology, Collage of Science, Salahaddin University-Erbil, Erbil, Iraq
[2]Department of Computer Science, College of Science, Cihan University-ERBIL

***Abstract*— In our daily life the computer and technology took almost all aspects of life through the use of smartphones, smartwatch, laptops, internet, and so many different tools. This makes people around the globe utilize services and tools more than before which lead to introduce hundreds of algorithms to enhance such services. The deep learning tools in AI become an urgent need to simulate a human behavior by the machine as both consists of neurons, and a weighted input to that neurons. The trip from a single neuron to AlexNet and then to EfficientNet will predict a bright future of this area in AI especially with dealing with images. Computer vision considered as mid-way that collect both AI and image processing to gather to perform a classification of images through AI to increase the understandability of such images and give the ability for machine for better decision making upon these images without human interfere, this article provides a survey study about a number of most recent papers about techniques AlexNet, ResNet, GoogLeNet, VGG, MobileNet, DenseNet, LetNet, Gabor CNN, Siamese CNN, PPF CNN and 3D GAN, and summarize there efficiency and weak points according to the papers so to we can provide a good guide for this area researchers.**

***Index Terms*— deep learning, image classification, AI, Computer vision, image processing**

I. INTRODUCTION

Nowadays the world is rapidly directing toward digitization and the amount of data is accumulated day after another, this huge amount of data gaining, collecting, and extracting such big data presents a wide variety of obstacles, among processing and interpreting them including data volume, variety of sources, quality of data, security challenges, integration, and validation, these mentioned challenges make companies, universities, and enterprises try to find a clear way to deal with it. Such accumulative amount of data was the key factor to what is known as "data Science".The spread of cheap cameras, smartphones camera besides this exhaustive amount of CCTV and survey lines make images a very common data type especially since they are considered as one of the most important parts of data science and data mining as one image can express thousands of words and people forwarded to document everything through the image and pictures. Dealing with images is important in many areas such as medical image processing, fine arts, digital forensics, digital multimedia systems, and so on, the first medical image processing research paper was introduced in 1972 when Sir Godfrey Hounsfield and Dr. Allan Cormack, independently developed the technology and they awarded a Nobel Prize. But still with the current explosion of data traditional image processing won't work. This motivated me to intervene in artificial intelligence algorithms. After the positive increase of AI algorithm's. It seen that these algorithms are applied efficiently on the field of image processing such as classification of images, image restoration, feature extraction, etc. The classification of images means grouping a limited amount of neighborhood pixels to the same label which is used in present day application's like classifying human lung to either has cancer or not. (W Rahane, et al 2018) implements machine learning techniques for such classification. Nowadays a sufficient amount of dataset are available for researchers to train their model using both machine learning and deep learning algorithms. AlexNet was the first sufficient deep learning model introduced in 2012, the model introduced by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, they use the deep convolutional neural network to overcome the limitations of old methods of image processing and computer vision. The error rate where significantly reduced. Now there are about 10 different techniques in deep learning have an ability to work on different images and can be trained by different international or local datasets.

II. IMAGE CLASSIFICATION TECHNIQUES

Image classification is merely answer the question; what is the image? Or either the image belongs to a specific class or not? Image classification consider as a cornerstone of computer vision. A number of algorithms and techniques had been introduced in this term which mainly classified in deep learning and machine learning.Both two methods are used to classify the images, machine learning means learn from data using one of the ML algorithms like SVM, Naïve Bayes probability, etc. and in the other hand the deep learning algorithms are complex structure algorithm used for unstructured data like text and image. Below are the most used algorithms in image processing and image classification.

Below are some brief description about some common techniques.

1- Supervised learning is the most common classification approach where you train your model according to a dataset to prepare the future behavior of the model.

2- Unsupervised learning is come up with the absence of dataset, so the model designer must depend on statistical common feature within the image like distance and color.

DOI:http://doi.org/10.24086/cocos2024/paper.1082

3- Convolutional Neural Networks (CNNs) it is mostly used for classification tasks like image and sound classification besides computer vision, this network consists of three layers Convolutional layer which can accept a great amount of data like image high, width, colors image features, etc., Pooling layer and Fully-connected (FC) layer, figure 1 illustrate the architecture of CNN.

4- Transfer Learning this algorithm consider that if a knowledge is obtained in a specific topic, it will work well for similar topics as well, for example a network trained to classify cars, may work for other vehicles successfully.

5- Recurrent Neural Networks (RNNs) this network mainly used for streaming like immediate language translation, a feedback process is similar to other networks excluding sum of errors.

Attention Mechanisms a model that improve the deep learning using a careful selection of inputs.

Figure 1: A Simple Convolutional Neural Network Architecture

III. Deep Learning techniques developed for Image Classification

Deep learning is a technique in a machine learning and both are techniques in Artificial intelligent. It's known as deep because it consists of single hidden layer or more, a network like LeNet-5 holds seven layers, these layers are the alternative method of human interference to determine the class in machine learning, Figure 2 illustrate the deep learning in image classification. (Yann LeCun et al. At 1989) assumed that one layer of neural network will not be sufficient, this leads them toward a multilayer neural network, at 1990 they proposed LeNet-5 neural network for handwriting recognition. The network used convolution to extract the image features with hyperbolic tan activation function. And a fully connected perceptron at the final layer to provide the decision. The LeNet-5 suffer from a high amount of computationally which make researchers stay on tradition algorithms like SVM. Until the dramatically grew of GPU (Sanskruti Patel, 2020). The LeNet-5 is still a good choice but some time it causes of overfitting, this lead the researchers to introduce benchmark architecture. The first benchmark architecture was AlexNet, where introduced at 2012, it was an image classification deep convolutional neural network technique, were the convolutional layers used to learn image features. The AlexNet consists of eight layers, five convolutions for feature extraction and three fully connected layers for classification purpose, in spite of its ability for classification, still it is not deep enough like VGGNet and ResNet (Md Zahangir, et al 2018).The visual geometry group (VGGNet) is developed at 2014 by Karen Simonyan et al. to increase the depth of neural network, with its two version VGG-16 (16 convolutional layer) and VGG-19 (19 convolutional layer), in order to maximize the efficiency of the convolutional neural network, the VGG-6 has an ability to accept about 138 million parameters, and 128, 256 or 512 filters in its output layer, it is observed that it had 92.7% accuracy on ImageNet dataset, which is an international dataset contain about 14 million images labeled with 1000 different labels. The major problem in this method is the exhaustive amount of calculation and considered as very time consuming method (Karen 2015). Residual Network (ResNet) also proposed at 2015, it is another convolutional neural network generations mainly used by computer vision through creating hundreds or thousands of convolutional layers. The reason behind adding such amount of layers is to minimize the error rates, but a problem of vanishing/exploding gradient was introduced, which means approaching the weights gradually either to 0 or to a very large number, which leads to increasing the error rate by the increase of layer numbers. It is seen that 56-layers CCNs error rate is more than 20 layers network. From the above problem ResNet introduce a residual blocks. Which is simply skip some connections and layers, whenever a specific layer seen to be vanished or exploded, it will be skipped (Kaiming He et al, 2015).

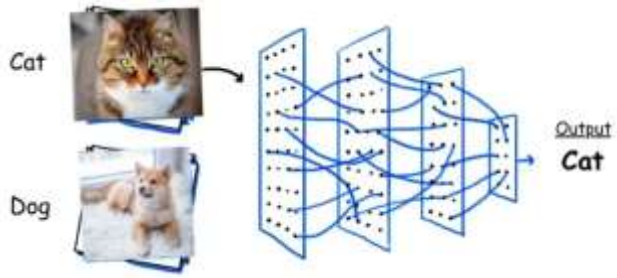

Figure 2 Image classification in Deep Learning.

Andrew G. Howard et al. at 2017 proposed a MobileNets network which is a lighter version with an ability to work on mobile devices and other embedded system. It depends on a sequence of depth to provide a light weight neural networks. This system used widely in mobile applications such as finger print and face detection devices (Andrew G et al, 2017). EfficientNet is also an efficient technique which is released at 2019 uses compound coefficient to scale depth, broadness and resolution of the network, where the coefficient changes with the change of image size or available time for performing computation (Mingxing Tan et al, 2020). Vision Transformers(ViT) is another deep learning technique used in image recognition. It breaks the image down to subgroups and process them independently, then aggregate the result to perform the classification, it is proposed by (Alexey Dosovitskiy et al. at 2021). NAS (Neural Architecture Search)

DOI:http://doi.org/10.24086/cocos2024/paper.1082

was also proposed in the same year and still it is a subject of more than 1000 papers. This model automatically build the architecture for a specific task (Colin White et al, 2023). Another technique is called Federated learning where the neural network broken into sub networks and train theme independently. This makes terminal device trained locally instead of sending the whole data to server computer for training, hence only the raw data will transfer to the server side this causes of dramatically increase of the efficiency and privacy as well, this technique was introduction in 2016 but it's still a worm subject in deep learning field (Mammen 2021).

IV. Deep learning Research Implemented on dataset

After machine learning and deep learning introduction, a sufficient amount of researches and studies had been introduced to provide a smooth comparison between these techniques, actually the classification according to dataset is needed. Some models work well on some dataset objects and fail for others, this review paper generally classified the researches to either homogenies dataset or heterogeneous dataset.

V. Homogenies dataset

Here some researches explored to that used the image of homogenies objects in dataset like dog and cats for example, both objects have some similar features which increase the challenge of this type classification. (Jason Wang et al. in 2017) worked on effective data augmentation using deep learning. They employed a Generative Adversarial Network (GAN) for various styles of image generation. Two datasets, MNIST and tiny-magnet-200, were utilized. The tiny-magnet-200 dataset, incorporated into the proposed system, consists of 100,000 training images, 10,000 testing images, and 10,000 validation images, all in the dimensions of 64x64x3 (for X, Y, and color channels). This dataset comprises 200 different classes, each with 500 sample images. The second dataset, MNIST, comprises 60,000 handwriting images for training, and 10,000 testing images in the size of 28x28x1 (X, Y, and grayscale), the images classified into 10 different classes. To achieve better results, multiple copies of images were generated with various zooms, flips, crops, shifts, etc. For each input image, six different styles were generated using GAN. Ten experiments were conducted on the data for 40 epochs, the learning rate was 0.0001 and Adam optimization was used. After the augmentation of images, they were fed into a network with only three convolutional layers. The researchers used the VGG16. The first experiment focused on dogs and cats, followed by two experiments on goldfish and dogs. According to the researchers, these experiments went more smoothly due to the orange color of goldfish. They also observed that neural augmentation performed better than not augmented images data. In the dog experiment, the accuracy was 91.5%, and for cats and goldfish, it was 85.5% and 85.5%, respectively. The researchers concluded that both neural augmentation and GAN did not enhance classification as traditional augmentation did. (Qing Lv, Suzhen et al. in 2022) worked on the flowers dataset provided by Oxford University, which contains variant information about shapes, light changes, and proportions of the flower. The dataset comprises 17 different types of flowers, each with 80 images. The dataset was divided into three parts: train, test, and validate. Three classification models were applied, LetNet, AlexNet, and VGGNet. After 50 iterations, the accuracy were 28%, 17%, and 48% for the three networks, respectively. When the number of iterations doubled, the network performance improved: LetNet reached 45%, AlexNet reached 39%, and VGG reached 72%. After converging, VGG achieved the highest accuracy (75%), followed by LetNet (58%) and AlexNet (41%). (Gupta et al, 2023) used three insect dataset NBAIR, Xie2 and Xie1with 40 class, 40 class and 24 class of insects consequently to evaluate the performance of some deep learning techniques, the images resolution were (227 X 227 X 3). 80% of dataset used for training and 20% for testing. After implementing in python using Anaconda IDE, the observed Accuracy and training time was as follow for each technique. (AlexNet 98.95% finished in 1.75 hours), (ResNet-50 was %99.1 finished in21.42 hours), (GoogLeNet 98.97% finished in 4.72 hours), (VGG16 was %99.49 finished in 129.27 hours), (Inceptionv3 %99.46 finished in 23.11), (MobileNetv2 %99.03 finished in 8.6 hours), (DenseNet201 %99.71 finished in 83.19 hours), the researchers concluded that DenseNet201 had the best performance. the other techniques like (Alexnet, ResNet, andVGG, Inceptionv3 and GoogLeNet) on the given dataset were in same accuracy, they perform well only if a big dataset available for training, and they assume also that the Mobilenetv2 was suitable for real time applications or for robot. They also didn't recommend the AlexNet due to its large memory consumption in its work.

VI. Heterogeneous dataset

The research covered in this section is where the researcher took two different objects available in dataset like flowers and dogs. (Harshali Mane et al. in 2018) used the ImageNet dataset, consisting of 1,281,167 training images, 50,000 validation images, and 100,000 test images. The researchers selected four different objects in the dataset, Sea Anemone, Barometer, Stethoscope, and Radio Interferometer. They applied a five layer AlexNet for classification, in the first layer 96 (11 x 11) filters used. The model included fully connected convolutional neural networks, and in the normalization layer, two norms were present ReLU, and htan, the ReLU trained much faster than tanh. The researchers concluded that the model could be classify without any problem. (Shutao Li et al. at 2019) tried a number of machine learning and deep learning techniques on hyper spectral images, the researchers used GRSS which built by IEEE, this dataset has 2.5 million different images distributed over 15 class (Healthy grass, tree, soil, etc.) and Houston dataset with 9 classes within 3.7 million images, the image resolutions were (4349 X 1905). The deep learning techniques used by this paper was (3D CNN, Gabor CNN, Siamese CNN, PPF CNN, 3D GAN and deep feature fusion network CNN) with a SVM and JSR machine learning techniques, the training and testing were vary from a class to class for example (healthy grass 198 training and 1053 testing),

DOI:http://doi.org/10.24086/cocos2024/paper.1082

(tree 188 training and 1056 testing), etc. The result of SVM, EMP, EPF, 3D GAN and JSR were not satisfied the researchers, the accuracy in some classes were 99% or 100% while in some other classes stay between 43% and 51%. The CNN techniques were between (81.3 for PPF CNN and 84.56 for DFFN) the experiment result show that CNN frame works were displayed better result than others. While the 3D GAN was 0.799% which was better than other tradition method. (Krishna et al, 2019) proposed a network with five convolutional layer and three layers fully connected, over three days of training using ImageNet Fall 2011 dataset, which contain approximately 14 million images. The transformation of learning performed on the network through the train of AlexNet, ResNet and VGG16, using augmentation to increase the amount of input images. In the data augmentation face, the researchers used 6 dataset. flowers102, dogs, caltech101, etc. Nine different pre trained model created by other researchers used for learning transformation. The accuracy was as follow (GoogleNet 66%, ResNet 52%, Alexnet 13%, combine (Alex, VGG16) was vary between 72% to 95% according to dataset, Inceptiov3 on face %96.3, CNN deep 96%, Genetic DCNN vary between 66% to 99% according to dataset, Inception-v3 on dog dataset 96%, combine (CNN,AdaBoost) 88.4%) the researchers concluded that the Inceptiov3 that proposed by Hussain, M.et.al. had the best effects of learning.

VII. Discussion and Comparison

The image classification just like any other area started from scratch where the image features had been extracted and manipulated, the grew of AI affect this area of image processing too, where many techniques of machine learning had been implemented like SVM, and K nearest, in 2012 with introduction of AlexNet, the efficiency of deep learning had been noticed in image classification which leads to many other invention and development in this area, which embark the number of algorithms significantly. Hence implication on real dataset and compression between these techniques become an urgent need in this important part of computer vision. (Jason Wang et al. in 2017) used a dataset to classify the images between of dog and cat and dog and goldfish using VGG, they proof that the augmentation of image using different styles are very important to increase the amount of training images. They also concluded that classification of cat-dog was less accurate that the classification of dog-goldfish, the VGG works well if there is a noticed difference in image features whereas (Harshali Mane et al. in 2018) worked on the old version of AlexNet but with changing the activation function from ReLU to hyperbolic TAN, and they gain a good result, so however a plenty number of algorithm introduced but still we can improve the old one to provide further enhancement instead of inventing the new complex algorithms. With this incremental numbers of algorithms and techniques, the comparison between these techniques were raised rapidly, (Shutao Li et al. at 2019) applied the same dataset on 8 different techniques (3D CNN, Gabor CNN, Siamese CNN, PPF CNN, three dime national GAN and deep fusion of feature network CNN, SVM, JSR) they conclude that Convolutional Neural networks accuracy were much higher than other techniques. While (Krishna et al, 2019) worked on other techniques (AlexNet, ResNet, VGG16, CNN, AdaBoost and Inceptiov3) they concluded that AdaBoost and Inceptiov3 had the best performance unlike AlexNet which had the worst performance. (Qing Lv, Suzhen et al. in 2022) worked on more challengeable dataset with wide variety of shapes, light changes, and proportions of 1360 flowers, using three different techniques LetNet, AlexNet, and VGGNet over 50 iteration of learning, they concluded that the VGG had the best performance while the AlexNet had the worst. The (Gupta et al, 2023) used another three challenging insect dataset and on seven different techniques (AlexNet, ResNet-50, GoogLeNet, VGG16, Inceptionv3, MobileNetv2 and DenseNet201), considering the calculation time too, the researchers concluded that DenseNet201 had the best performance, the (Alexnet, ResNet, and VGG, Inceptionv3 GoogLeNet) were equal in accuracy. The also concluded that Mobilenetv2 is suitable for real time applications and robot and whereas the AlexNet term consumption was in high level.

VIII. Conclusions

After exploration of various techniques and discussions some comparison studies, this survey study concludes that deep learning techniques outperform traditional feature extraction methods and machine learning techniques. The preprocessing step is identified as a critical factor for the whole process success, ensuring that the data fed into the network is in sufficient amount and of the required quality. With the introduction of the OpenCV library and pre-built functions in programs like Python, the parallel testing of variant techniques becomes more feasible, some researchers incorporating up to eight methods in a single program to provide a wide view of overall techniques. A comprehensive understanding of these diverse techniques on various datasets is now available. The study also concludes that older techniques, such as AlexNet, can be improved and effectively implemented in image classification. However, certain algorithms may perform well on specific datasets and fail on others, emphasizing the necessity for identical experiments in each domain to identify the best fit for a given dataset.

DOI:http://doi.org/10.24086/cocos2024/paper.1082

IX. References

- Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam(17) , MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 17 Apr 2017, Cornell University, available at: https://arxiv.org/abs/1704.04861v1
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit and Neil Houlsby, (2021) An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 3 Jun 2021, Cornell University. Available at: https://arxiv.org/abs/2010.11929v2
- Colin White, Mahmoud Safari, Rhea Sukthanker, Binxin Ru, Thomas Elsken, Arber Zela, Debadeepta Dey, Frank Hutter, (2023) Neural Architecture Search: Insights from 1000 Papers, 25 Jan 2023, Cornell University, available at: https://arxiv.org/abs/2301.08727
- Harshali Mane and Venu Gopala Rao Matcha, (2018) Image classification using Deep learning, International Journal of Engineering & Technology · March 2018, DOI: 10.14419/ijet.v7i2.7.10892
- Jason Wang and Luis Perez, The Effectiveness of Data Augmentation in Image Classification using Deep Learning, [Submitted on 13 Dec 2017], Cornell University
- Karen Simonyan and Andrew Zisserman, (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition 10 Apr 2015, Cornell University, available at: https://arxiv.org/abs/1409.1556
- Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun,(2015) Deep Residual Learning for Image Recognition, [Submitted on 10 Dec 2015], Cornell University, available at: https://arxiv.org/abs/1512.03385
- Md Zahangir Alom, Tarek M. Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, (2018) Brian C Van Esesn, Abdul A S. Awwal, Vijayan K. Asari, The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches, 12 Sep 2018 available at: https://arxiv.org/abs/1803.01164
- Mingxing Tan and Quoc V. Le,(2020) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 11 September 2020, available at: https://arxiv.org/pdf/1905.11946.pdf
- Priyanka Mary Mammen, (2021) Federated learning, Federated Learning: Opportunities and Challenges, 14 Jan 2021], Cornell University, available at: https://arxiv.org/abs/2101.05428
- Qing Lv,Suzhen Zhang and Yuechun Wang, (2022) Deep Learning Model of Image Classification Using Machine Learning, Volume 2022 Article ID 3351256 https://doi.org/10.1155/2022/3351256, available at: https://www.hindawi.com/journals/am/2022/3351256/
- Sanskruti Patel, (2020)A Comprehensive Analysis of Convolutional Neural Network Models, International Journal of Advanced Science and Technology, Vol. 29, No.4, (2020), pp.771-777
- Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi and Jón Atli Benediktsson, (2019) Deep Learning for Hyperspectral Image Classification: An Overview, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 57, NO. 9, SEPTEMBER 2019
- Sajja Tulasi Krishna and Hemantha kumar Kalluri,(2019) Deep Learning and Transfer Learning Approaches for Image Classification, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.
- Vinita Abhishek Gupta, M.V. Padmavati, Ravi R. Saxena, Pawan Kumar Patnaik and Raunak Kumar Tamrakar,(2023) A study on image processing techniques and deep learning techniques for insect identification insect identification, Karbala International Journal of Modern Science: Vol. 9 : Iss. 2, Article 16. 2023, Available at: https://doi.org/10.33640/2405-609X.3289
- W Rahane, H Dalvi, Y Magar, A Kalane, S Jondhale, Lung cancer detection using image processing and machine learning healthcare, 2018 International Conference on Current Trends towards Converging …, 2018•ieeexplore.ieee.org

DOI:http://doi.org/10.24086/cocos2024/paper.1082