Image classification architectures comparison using AlexNet

MD. Samiul Hauque Chowdhury (18-36072-1), MD. Sabbir Hasan (17-33918-1), Mehedi Hasan Joy (18-38413-2),Shihab Hossain(18-38131-2), Fahad, Abdullah Al (18-37904-2)

Department of Computer Sciences, American International University-Bangladesh

Abstract

Training a deep convolutional neural network from scratch requires massive amount of data and significant computational power. However, to collect a large amount of data in medical field is costly and difficult, but this can be solved by some clever tricks such as mirroring, rotating and fine tuning pre-trained neural networks. AlexNet is the first large scale convolutional neural network architecture that does well on ImageNet classification. The Alexnet has eight layers with learnable parameters. The model consists of five layers with a combination of max pooling followed by 3 fully connected layers and they use Relu activation in each of these layers except the output layer. In this paper, we compare the AlexNet image classification algorithm with another existing image classification algorithm like VGG, DenseNet, etc. We have tried to find the advantages and disadvantages of AlexNet compared to other available algorithms. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. The network maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across training cases of the log-probability of the correct label under the prediction distribution. Alexnet has revolutionized the field of machine learning. AlexNet had 60 million parameters, a major issue in terms of overfitting.

Keywords: Deep learning, Computer Vision, Image Classifications, AlexNet, CNN

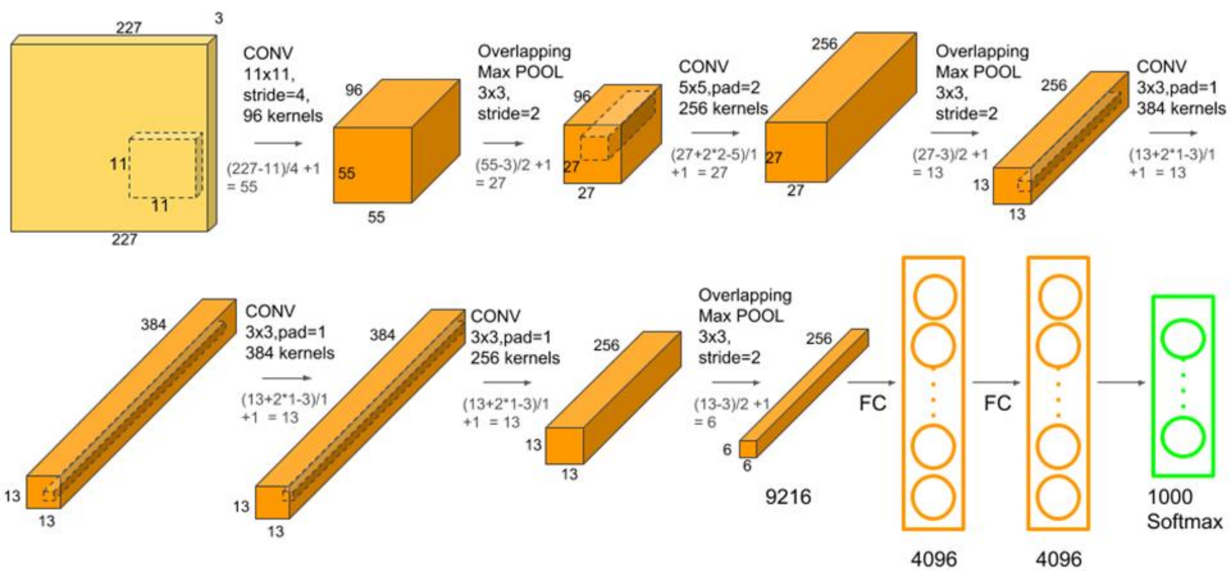
Introduction

AlexNet design is a conv layer followed by pooling layer, standardisation, conv-pool-norm, so a number of a lot of conv layers, a pooling layer, so many absolutely connected layers later on truly appearance terribly kind of like the LeNet network. There are a unit simply a lot of layers in total. There are a unit 5 of those conv layers, and 2 absolutely connected layers before the ultimate absolutely connected layer getting to the output categories. AlexNet was trained on ImageNet, with inputs at a size 227 x 227 x three pictures. If we glance at this 1st layer that may be a conv layer for the AlexNet, it’s eleven x eleven filters, ninety six of those applied at stride four. I had fifty five x fifty five x ninety six within the output and 35K parameters during this 1st layer. The second layer may be a pooling layer and during this case, we've three filters 3 x 3 applied at stride two. The output volume of the pooling layer is twenty seven x twenty seven x ninety six with and zero parameter to be told. The pooling layer doesn't learn something as a result of the parameters area unit the weights that attempting to be told. Convolutional layers have we tend toights that we tend to learn however pooling all we do is have a rule, we glance at the pooling region, and that we take the gamma hydroxybutyrate. Thus there are not any parameters that area unit learned. There are a unit eleven x eleven filters at the start, then 5 by 5 and a few 3 by 3 filters. In the end, we've some of absolutely connected layers of size 4096 and eventually, the last layer, is FC8 getting to the softmax, that goes to the a thousand ImageNet categories. This design is that the 1st use of the ReLu non-linearity.[1]

Hyperparameter:

In this paper this architecture is the first use of the ReLU non-linearity. AlexNet uses a layer of normalization also. In data augmentation, ALexNet used flipping, jittering, cropping, colour normalization and these things. Other parameters are Dropout with 0.5, SGD + Momentum with 0.9, initial learning rate 1e-2 and again reduced by 10 when validation accuracy become flat. The regularization used in this network is L2 with a weight decay of 5e-4. It was trained on GTX580 GPU which contains 3GB of memory.

ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories. The images were collected from the web and labeled by human labelers using Amazon’s Mechanical Turk crowd-sourcing tool. Starting in 2010, as part of the Pascal Visual Object Challenge, an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has been held. ILSVRC uses a subset of ImageNet with roughly 1000 images in each of 1000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. ImageNet consists of variable-resolution images. Therefore, the images have been down-sampled to a fixed resolution of 256×256. Given a rectangular image, the image is rescaled and cropped out the central 256×256 patch from the resulting image. [2]



The architecture depicted in Figure the AlexNet contains eight layers with weights; the first five are convolutional and the remaining three are fully connected. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels. The network maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across training cases of the log-probability of the correct label under the prediction distribution. The kernels of the second, fourth, and fifth convolutional layers are connected only to those kernel maps in the previous layer which reside on the same GPU. The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully-connected layers are connected to all neurons in the previous layer.

Image segmentation in general is defined as a process of partitioning an image into homogenous groups such that each region is homogenous but the union of no two adjacent regions is homogenous. Image segmentation is performed to separate the different regions with special significance in the image [20]. These regions do not intersect each other. Blob detection helps to obtain Regions of Interest for further processing. It is applied for the presence of same type of objects in multiples. Segment the objects of interest from the complex background. The pixels are separated into two classes C0 and C1 (background and object), using a threshold at level K. After then class means ($\mu 0$, $\mu 1$) class variances ($\sigma 0$, $\sigma 1$) are calculated. Then a threshold K is searched, that maximizes.

Literature Review

According to this paper, a modified deep CNN model based on a pre-trained AlexNet for image forgery detection and localization is proposed. The proposed work shows that the proposed model is deemed to be one of the best models to detect tampered images. Not just it can get execution obviously superior to different models, however it is likewise unequivocally strong to the most realized picture preparing. Obviously, the proposed model can adapt to an assortment of tasks with a solid learning ability. Comprehensive test results introduced that the proposed model is stunning in getting controls and accomplishes great generalizability to concealed information and dark altering types. The experimental results likewise show that the improved AlexNet model proposed for identifying and finding the manufactured regions score an impact that is superior to the current models on the datasets previously mentioned. The improved AlexNet is demonstrated to have the ability to become familiar with the frameworks of the manufactured regions and in this manner the capacity to recognize the altered and non-altered regions. Indeed, even with promising outcomes, it's anything but an absolute necessity to remember that no model can settle all falsification assaults altering without anyone else. The model actually needs further exploration to distinguish little produced regions and districts under huge varieties. [4]

After reading this paper we observed that this paper has looked at the presentation of two profound learning models for vehicle location and order in metropolitan video successions. Albeit the AlexNet model is utilized for include extraction in an impromptu arrangement of models situated to metropolitan situations, the pre-prepared Faster R-CNN model accomplishes better outcomes in right discoveries as per F1-score measure. Comment that the Faster R-CNN model doesn't utilize any unique ascribes for vehicle discovery though GMM foundation deduction utilized in AlexNet model. Indeed, as the union rates (MR) result shows, GMM foundation deduction actually has issues with fixed vehicles and blocked situations. In Faster R-CNN, the RPN part results could be improved giving some metropolitan setting data as limitation size of the locales. For future work, we expect to improve the aftereffects of the RPN part of the Faster R-CNN model advancing it with traffic setting data, and improve the grouping segment with highlight extraction utilizing a Deep Architecture as AlexNet, ZF or VGG, with a more extensive arrangement of metropolitan street client. [5]

According to this paper, they used Alexnet to detect disease in rice leaf. According to this paper they detected the three prevalence rice leaf disease using Alexnet neural network and their result provides improved accuracy. In their paper Alexnet was adopted more than CNNs for image tasks and they got an accuracy of 99.42%. [6]

. Discussion

Our findings indicate that with only supervised learning, a massive, deep convolutional neural network might generate record-breaking results on a highly complex dataset. It's important to remember that dropping a single convolutional layer decreases the chance of our network. For example, eliminating any of the middle layers reduces the network's top-1 performance by around 2%. Here did not use any unsupervised pre-training in our tests to keep things simple, even though we expect it to help, especially if we get enough computing capabilities to massively increase the size of the network without a corresponding increase in the amount of labeled data. Our results have improved as we have grown our network and taught it longer, but we still have a long way to go before we can match the human visual system's infero-temporal approach. Eventually want to utilize very big and deep convolutional nets on video sequences, where the temporal structure provides very useful information that is missing or less visible in static images. There are 60 million parameters in our neural network architecture. ILSVRC's 1000 classes impose a 10-bit limit on the mapping from image to label for each training example, however this is insufficient to learn so many parameters without significant overfitting. Application of Alexnet model is limited in image classification because of the large convolution kernel and stride in the first constitutional layer leading to over rapid decline of feature maps resolution and excessive compression of spatial information. This paper is proposed an improved alexnet model according to the design principle of constitutional neural networks (CNNs).Recognition problem is found on image registration, which is a classic problem of computer vision with several applications across like defense, remote sensing, medicine etc. A Feature based image registration methods traditionally used hand-crafted features are matched using a threshold either on distances or radio of distances computed between some of the feature descriptors. Talking about the inefficiency, The AlexNet technology is mostly use in security camera systems, to monitor the scene, while the recognition of human activity is targeted to measure. The anexnet architecture constitutional neural network method is used. Although fortunately the result of obtains a success rate of 100% for detection of human activity , but still a best success rate of 96% for recognizing human activity .

Conclusion

In this paper, we investigate the effect of the concatenation of the AlexNet image classification performance. The reports have done using Convolutional Neural Networks classification algorithms such as AlexNet, VGG, DenseNet, etc. AlexNet is an incredibly powerful model capable of achieving high accuracies on very challenging datasets. However, removing any of the convolutional layers will drastically degrade AlexNet's performance. AlexNet is a leading architecture for any object-detection task and may have huge applications in the computer vision sector of artificial intelligence problems. In the future, AlexNet may be adopted more than CNNs for image tasks. AlexNet can also be credited with bringing deep learning to adjacent fields such as natural language processing and medical image analysis.

References

[1] AlexNet - ImageNet Classification with Convolutional Neural Networks. (2021). Retrieved 1 July 2021, from https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/

[2] Architecture comparison of AlexNet, VGGNet, ResNet, Inception, DenseNet. (2021). Retrieved 1 July 2021, from https://towardsdatascience.com/architecture-comparison-of-alexnet-vggnet-resnet-inception-densenet-beb8b116866d?fbclid=IwAR04NbVP8UsBS9nhsZ8cTB4TjFhx295BfGEoCW6K8OUUVBvsByso3geP1IQ

[3] Iswarya, P., & Maheswari, D. (2019). Paddy Leaf Disease Identification and Classification System A Review. International Journal Of Computer Sciences And Engineering, 7(5), 976-979. doi: 10.26438/ijcse/v7i5.976979

[4] Samir, S.; Emary, E.; El-Sayed, K.; Onsi, H. Optimization of a Pre-Trained AlexNet Model for Detecting and Localizing Image Forgeries. Information 2020, 11, 275. https://doi.org/10.3390/info11050275

[5] Espinosa J.E., Velastin S.A., Branch J.W. (2017) Vehicle Detection Using Alex Net and Faster R-CNN Deep Learning Models: A Comparative Study. In: Badioze Zaman H. et al. (eds) Advances in Visual Informatics. IVIC 2017. Lecture Notes in Computer Science, vol 10645. Springer, Cham. https://doi.org/10.1007/978-3-319-70010-6_1

[6] Matin, Md.M.H, Khatun, Moazzam, Md.G. and Uddin, “An Efficient Disease Detection Technique of Rice Leaf Using AlexNet”, Journal of Computer and Communications, 8, 49-57, 2020, doi: https://doi.org/10.4236/jcc.2020.812005

[7] R.M. Bell and Y. Koren. Lessons from the netflix prize challenge. ACM SIGKDD Explorations Newsletter,

9(2):75–79, 2007.

[8] A. Berg, J. Deng, and L. Fei-Fei. Large scale visual recognition challenge 2010. www.imagenet.

org/challenges. 2010.

[9] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[10] D. Cire¸san, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification.

Arxiv preprint arXiv:1202.2745, 2012.

[11] D.C. Cire¸san, U. Meier, J. Masci, L.M. Gambardella, and J. Schmidhuber. High-performance neural

networks for visual object classification. Arxiv preprint arXiv:1102.0183, 2011.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical

Image Database. In CVPR09, 2009.

[13] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun. What is the best multi-stage architecture for

object recognition? In International Conference on Computer Vision, pages 2146–2153. IEEE, 2009.

[14] A. Krizhevsky. Learning multiple layers of features from tiny images. Master's thesis, Department of

Computer Science, University of Toronto, 2009.

[15] A. Krizhevsky. Convolutional deep belief networks on cifar-10. Unpublished manuscript, 2010.

[16].Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Van Esesn, B.C., Awwal, A.A.S. and Asari, V.K., 2018. The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.

[17]Kavitha, K. and Rao, B.T., 2019. Evaluation of distance measures for feature based image registration using alexnet. arXiv preprint arXiv:1907.12921.

[18]Prastika, Kerly. "Application of individual activity recognition in the room using CNN Alexnet method." In IOP Conference Series: Materials Science and Engineering, vol. 1007, no. 1, p. 012162. IOP Publishing, 2020

Name	ID	PART
MD. Sabbir Hasan	17-33918-1	Abstract & Conclusions
Mehedi Hasan Joy	18-38413-2	Introduction
MD. Samiul Hauque Chowdhury	18-36072-1	Literature Review
Shihab Hossain	18-38131-2	Discussion

Fahad, Abdullah Al	18-37904-2	Discussion