Rohit Basu
December 12 ‘20
CS390
Dr. Ravikumar

CONVOLUTIONAL NEURAL NETWORKS AND MODERN APPLICATIONS

A CNN(Convolutional Neural Network) is a class of deep neural networks used particularly to analyze visual imagery. In this paper we will be taking an indepth look at two fields that CNN’s are helping break major ground in the modern day: CT Scans and Population Diversity. We will also cover a brief overview of the methodologies used to validate CNN’s and as well as some data to see truly how effective these neural networks are at classification and identification.

The CNN starts off by ingesting images in the form of three separate strata of color stacked on top of each other. We identify the colors on the image through a rectangular box whose width and height are determined by the pixels from those dimensions. The first layer of the CNN is the CONVOLUTIONAL LAYER which does the most computational work and is the foundation for this algorithmic process.The convolutional layer applies a series of filters or kernels. The filters are dependent on the RGB value of the image to decide what depth the filter should be at. Ultimately this process allows us to summarize the image input and take the key pixels needed to accurately analyze the image; which in turn significantly reduces the computational complexity. For example if we had a convolution with a 3D filter the output would be a summative 2D matrix. The next layer is called the ACTIVATION LAYER which implements the Rectified Linear Unit. This helps us increase non-linearity in the CNN due to the lack of linearity when it comes to Images made of different objects. The third layer, the POOLING LAYER, involves the extraction of features as shown in Figure 1. Lastly , the

FULLY CONNECTED LAYER is responsible for flattening the image which involves transforming the matrix into a single column which after being fed to the neural network for processing will give us activation function(soft max or sigmoid) to classify the output.

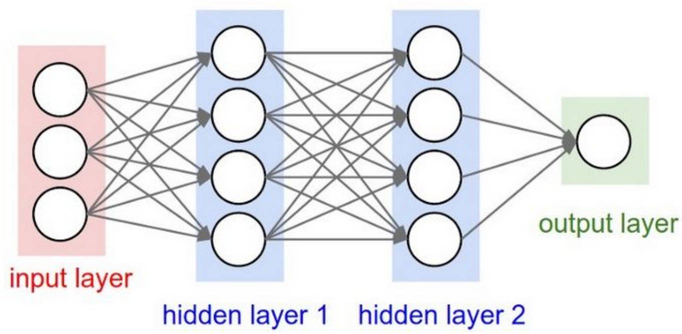

Figure 1.

Some interesting questions that motivate the use of CNN's in the field of ecology are , "What are spatial distributions of rare animals", "Which species are being threatened and need protection such as brandicoot?" and "Which cohort of pest species such as Red Fox and Rabbit need to be controlled?". A key obstacle, as stated by Dr. Gill, when approaching some of these questions is data annotation and the fact that doing this manually is a very time-intensive and laborious process. For example the Snapshot Serengeti project(a public database created to help with the automation of data annotation) took more than two months to annotate a 6 month batch of images by a group of 28,0000 registered volunteers and 40,000 unregistered volunteers. With the use of CNN's we can make this process much more efficient and accurate with the use of a binary classifier and a multiclass classifier. A binary classifier provides automated feedback on raw data and returns the images(taken by a Trap Camera) that actually have animals in them. Solving this problem makes the rest of the process of image classification much faster and robust. This in turn helps us with our second problem, multi class classification, because now

Citizen Scientists can spend less time on trying to figure out if the image has an animal in it and more time on trying to annotate what species class the animal in the image is from.

The use of Trap Cameras is growing rapidly as a method of collecting wildlife data unobtrusively. The cost of production for these Trap Cameras is going down significantly which has led to much more scalability in terms of data collection. It is due to this newfound scalability that the SnapShot Serengeti project gathered 3.2 million images through 225 camera traps across the serengeti national park. The Wild Life spotter project has also shown impressive statistics with a 94.6% accuracy rate in correctly identifying images with animals in them and a 90.4% accuracy rate in correctly identifying the three most common species among the sets of images taken in South Central Australia with a single labeled dataset.

Comparing performance percentages among the three different architectures used in the paper "Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring" Nguyen et. we see similar accuracy distributions across the board. The VGG-16 Architecture showed a 96.6% accuracy , the ResNet 50 architecture came in with a 95.96% and the Lite AlexNet, the simplest architecture with only 5 layers, showed an accuracy level of 92%. A big problem that comes with these types of architecture's is the computational complexity associated with compiling these neural networks. Therefore, Fine Tuning has proven to be a very cost-effective tool as it decreased the amount of time it took to complete the VGG-16 neural network from 40,000 seconds to 65 seconds.

Chest computed tomography (CT) scans are primarily used in detection and classification of pulmonary diseases. Manual review of these scans is a very time and labor intensive job which motivates the application of a Computer Aided Diagnostic(CAD) system. The CAD system's objective is to identify whether the lungs in the scan are normal or infected by interstitial lung disease (ILD). A key problem here again is the lack of readily available annotated data needed to to develop the CAD system. Therefore a process called Transfer Learning which is a machine learning method where a model developed for a task is reused as the starting point for a model on the second task. In layman's terms this method allows you to recycle data repeatedly and use the results each time to build a better model.

An alternative resource saving solution that has been proposed is the use of a clustering algorithm to reduce the workload of manually preparing the data set for the training of the CNN. This process involves the splitting of CT slices into an image patch which then utilizes a k-means cluster algorithm. A k-means clustering algorithm is a method that essentially takes observations belonging to the cluster with the nearest mean, which then serves as a prototype of the cluster. In order to find the achieved average value we calculate the F-score(F_{avg}) using the following formula(check figure two). In the paper “Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset” Xu, M., Qi, S., Yue, Y. *et al.* The team used 23 parameters in the training process of the CNN model but only saw significant impact on classification results from 9 variations of those parameters. Some specified default settings used in this study is a kernel size of 5, local response normalization layer of 3 and a max pooling type were used for the specific parameters used in this model. This combined with the use of validation methods such as cross-shaped verification, volume intersection, a connected component analysis and as well as an a patch expansion allowed the team to create the final data set. Secondly the team also made use of a CNN architecture that consists of one convolutional layer with six kernels, one maximum pooling layer as well as two fully connected layers. Using this method to generate data allowed the team to train a variety of models and evaluate them through an eight fold cross-validation.

$$F_{\text{avg}} == \frac{Precision_{nlp} \times Recall_{nlp}}{Precision_{nlp} + Recall_{nlp}} + \frac{Precision_{lp} \times Recall_{lp}}{Precision_{lp} + Recall_{lp}}$$

Figure 2.

Precision and Recall variables in the formula shown in figure two stand for the positive prediction rate and the sensitivity of the class of non lung parenchyma.

These two methodologies, with a sample size of 121,728 patches, saw an average F-score of 0.9917 in identifying the scans with some variant of lung parenchyma and saw an area of a curve up to 0.9991 for classification of lung parenchyma and non-lung parenchyma. The results are ultimately demonstrating that the proposed k-means clustering algorithm can be used in generating a training dataset for CNN models with accuracy and consistency. The obtained CNN model can quite clearly differentiate between lung parenchyma very accurately and has the potential to locate and analyze lung lesions.

A current roadblock, as mentioned before ,that researchers face is the ability to obtain an applicable image classification accuracy. Due to the enormous amount of manual processing required to input images and combined with poor performance obtained by wildlife monitoring systems that application of a variety of Neural Networks are encouraged to be applied in this domain and not just a CNN.

Methods such as a Linear support vector machine (SVM) provide enormous benefits as well a limited amount of data. In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that are able to efficiently perform a non-linear classification using a kernel trick which essentially maps inputs to a high-dimensional feature space. The motivation for the application of this learning model is that each data point is observed in terms of p-dimensional vector and then the model is determined to whether such points can be plotted on a (p-1) dimensional hyperplane. Essentially if it can be mapped then there is a successful classification attempt for the input.

So in summation of all the points stated clearly we see a growing need for automation in fields beyond just the ones listed in this paper. Computer aided identification and classification systems can save organizations an exorbitant amount of time and money if implemented properly. These models are only becoming increasingly accurate the more non-linear data inputs are fed into the system as seen in many of the studies presented in this paper. Obstacles in acquiring and evaluating data on a widespread basis are consistently being tackled with new algorithms being implemented to synthetically recreate data. Time will only tell how many other fields we will see using CNN's in the future.

Works Cited:

Dr. Gurman Gill: Convolutional Neural Networks

Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset:
https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-018-0619-9#citeas

Animal Recognition and Identification with Deep Convolutional Neural Networks for

Automated Wildlife Monitoring:
https://www.researchgate.net/publication/319392368_Animal_Recognition_and_Identification_with_Deep_Convolutional_Neural_Networks_for_Automated_Wildlife_Monitoring

Transfer Learning:
https://en.wikipedia.org/wiki/Transfer_learning

F-Score:
https://en.wikipedia.org/wiki/F-score

Convolution Neural Network:
https://en.wikipedia.org/wiki/Convolutional_neural_network