# Complex Convolutional Neural Networks

Shivam Misra
March 11, 2020

***Abstract*—Convololtional Neural Networks (CNNs) are commonly used image classification networks that use machine learning to recognize components of images and then learn to combine these components to recognize structures that are much larger. This paper first goes into what a CNN is and some of the problems CNNs face in the scope of complex-number based CNNs and their benefits into image classification.**

## I. Introduction

Convolutiona; Neural Networks are advanced techniques for image classification in the field of computer science. Rather than traditional CNNs for computer vision tasks, a complex number-based CNN can capture natural images, where the complex field is simply a generalization over the real model. Traditional CNNs are prone to overfitting, and can be hard to regulate. A complex-based CNN can be a better solution for problems in classification where phase structure detection is important. With a potential of having more input parameters than traditional convolutional networks, complex networks have the variability in sensitivity to produce better data. Complex number representations have the potential for easier optimization, allowing for images to be generalized better than before.

Images are read by computers in representations of pixels, which is expressed by matrices, as height by width by depth (X by Y by Z). These images have red, blue, and green channels (RGB). Convolutional layers utilise learnable filters , which help identify patterns and features in the input image, expressed in the form of a matrix with a smaller dimensin, but same depth. Filters differentiate between each other based on their different features, which are respectibely convolved on the input file, after which a set of activation maps are outputted, passing along to the next later of of the CNN.

Complex convolution is relatively new in the field of research and implementation, so use cases are still being developed. In this paper, we will discuss areas that have a benefit when looking at deep learning techniques. Some of these areas include exploiting complex synthetic aperture radar, a form of radar that is used to create two-dimensional images or three-dimensional reconstructions of objects and even speech and audio detection, which can be enhanced from the sensitivity complex neural networks can bring.

Finally, we will look at the benefits complex networks can have over real networks, in terms of accuracy and efficiency.

## II. Background

### A. *What are Neural Networks?*

Neural networks are based on a biological motivation. More specifically, the human brain is the basis for networks. Artificial Neural Networks (ANNs) are simulations of biological networks, which have units of neurons. These neurons work similarly to how neurons function in biology - taking in an input, passing the value through a set function, and then outputting a final value after the function. These input and output values depend on what type of network, what type of dataset, and what type of application the network is being used for. At the end of each neuron, more output neurions follow, which take the output and push them through their own function, continuing this process until the final layer is reached, where values can be interpreted as ”0” or ”1”s.

We can showcase an network that classifies letters of our English alphabet from a picture. If we have a 32x32 image, we need to have enough input neurons for the size of the picture - in this case, we need 1024 input neurons. Multiple layers of these neurons would follow, and there would have to be 8 output neurons at the end, since each one represents a byte of an ASCII character, to be able to represent the alphabet.

Networks work by implementing graph-like structures, which are initialized with randomly assigned weight values at edge of the graph. This means that each connection from neuron to neuron can be looked at by how weak or strong the signal between the two is.

Weight values are representations of how much attention neurons pay to the certain input values. Alterations of these weights result in a different output values. Knowing that, now, input that we know the expected output of can be fed into the network, upon which we compared the expected and actual output. If we feed in a picture of the letter ’A’ (ASCII character # 65). Doing this, we know the output neurons should correspond to the proper binary representation: 01000001. We check each output neuron. If our output does not match with the value we know to be true, we can use the preceding weight valyes to create a calculation of a numerical representation of how wrong the input values initally were. This training allows the weights for each neuron to be adjusted to make subsequent values correct on each iteration. Each new alteration means that we can work backward to find the value of error for each neuron back to the initial input. This training process is called backpropagation, and we keep changing inputs and matching with outputs that we know the real value of, making the network more accurate each time.

The above is a very simple example of an artificial neural network. In detail, these networks are composed of alternating layers of two types, affine and activation function. In an affine layer, each neuron’s value is a weighted sum of the previous layer’s neurons. In an activation function layer, each neuron’s value is set to be a non-linear function of exactly one neuron from the previous layer. Figure 1 shows a depiction of an ANN.

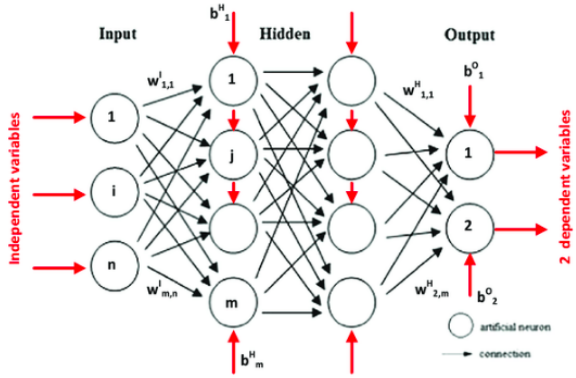


Fig. 1. An ANN with weights is shown, along with the depiction of hidden affine layers, with corresponding connections which are values that are ouputs of corresponding activation functions [1]

### B. (Complex) Convolutional Neural Networks

In CNNs, the neurons in each layer are organized as a three dimensional array rather than as a vector. Guberman explains this really well - ”The first two dimensions are called spatial, and the third is a division channels. The CNN model follows three principles characteristic of natural images - locality, sharing and pooling. The locality property, is the fact that pixels depend only on their neighbors, rather than on far away pixels. Sharing is the restriction that different pixels should undergo the same processing. Pooling is used to induce invariance to small translations, which is a characteristic of natural images. A pooling layer does so by splitting each input channel into patches, and replacing each patch with a single representative value in the output layer. Essentially, what this means is that in CNNs, there are identical clusters of neurons that all have the same activation functions. This is done to make sure each part of the image is treated the same, since image recognition is essentially recognition of smaller sub-images. Typical choices the maximal or average value, in max and average pooling, respectively” [2]. The activation function, the rectified linear unit (ReLU) function is usually given by:

$$\begin{cases} x & x \geq 0 \\ 0 & otherwise \end{cases}$$

Each image will be processed with a input of real values $x$, until an output of 0 is reached or the end of the network is reached.

A complex number (as we know), $z = a + ib$ has a real component, $a$, and an imaginary component $b$. We represent the real part and the imaginary part of a complex number as logically distinct real valued entities and simulate complex arithmetic using real-valued arithmetic internally. Then, four-dimensional tensors are made to link input and output feature maps of convolution layers, with complex weight values. This concept is particularly well-explained in a conference paper that was presented at the International Conference on Learning Representations [3], where the math derivations are explained with theorems and derivations. When comparing real and complex inputs, the difference is in the number of potential inputs a complex representation can give, having twice the amount of inputs. Networks, however, cannot take in a complex input just by separating it into two real inputs - transformations must be made to the data to make sure the outputs correspond to the type of input (complex) we are giving the network.

Nitzan Guberman, in his thesis paper, goes into detail about how a complex network is transformed, defining complex functions that take the place of functions that would usually end up in the pooling and preservation layers of a complex neural network. Using the Cauchy - Reimann equations, he introduces concepts for transformations, and proceeds to use advanced topics like the Wirtinger derivative operators to derive the needed fitting functions for his network. One example is the modified ReLU point-wise function, which now is represented as:

$$\begin{cases} Z & Z \in A \\ 0 & otherwise \end{cases}$$

for some $A \in \mathbb{C}$

## III. Applications and Implications

### A. Architecture Comparison

In comparison, architecture can also be analyzed. One example of this is done through the work of Guberman, who used Complex CNN’s to model cell identification. fluorescence microscopy images were taken and then put into networks to identify cells, using the corresponding formats for the ReLU functions. Figure 2 shows the architecture of a real-network, while Figure 3 shows the architecture of a complex network (transformed). Keeping the same architecture, twice as many inputs have to be interpreted in the real network, as we consider a complex number $Z$ to be the primary input parameter.

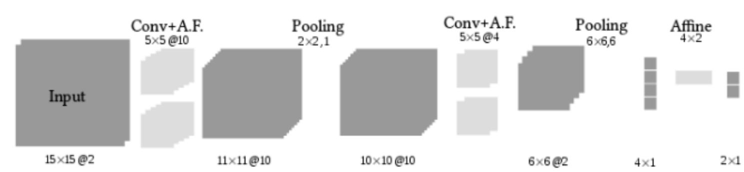


Fig. 2. The real network architecture equivalent to the complex one in 9. There are twice as many channels and kernels. To obtain an output of two classes, rather then four the projection layer from the previous image was replaced by an affine layer[2]

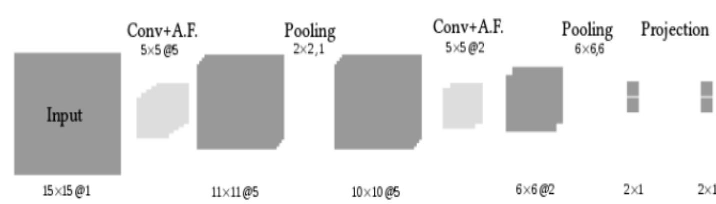


Fig. 3. The complex network architecture, with two convolutions, activation function and pooling.To obtain real valued labels, a projection layer is added. [2]

In his experiment, Guberman notes that the accuracies (shown in the table below) are comparable, with the real model performing slightly better. The training loss of the real network is much lower then its test loss, while those of the complex network are comparable. The real model suffers from overfitting.

| | Train loss | Train Accuracy | Test loss | Test Accuracy |
|---|---|---|---|---|
| Complex network | 0.056 | 97.4% | 0.0690 | 97.3% |
| Real network | 0.007 | 99.8 % | 0.1450 | 97.5% |

Fig. 4. Note that the real model's test loss is significantly higher than its training loss. [2]

What is overfitting? This deals with the model a neural network follows. Neural networks aim to achieve a model that will perform optimally on the data used to train it as well as new data. We want the model to learn from pre-determined examples and use them to generalize from the given examples for new examples in the future. Some methods used can be like a train/test split or k-fold valudations help us give an idea of the ability of the model to generalize new data. If there is less learning than expected, the model will be prone to perform poorly and undefit the problem. Likewise, too much learning will mean the training dataset will overfit and perform badly on new data. In both these cases, the model has not been generalized properly.

Going back to the data presented, the complex network does not present overfitting, as the training and test loss of the complex network remain close, and lie between the real network's train and test loss. These results suggest that the complex model serves as a regularization. The accuracies, however, do not present the same pattern. The real network's test accuracy does not decrease as the loss rises, and is higher than that of the complex network. Guberman suggests that more data is required to see if this phenomenon occurs for other network architectures.

### B. Other Examples

Some of the uses cases researchers have looked into include audio and speech detection. Because audio and speech varies so much over a variety of regions and dialects, even a singular language and pose hurdles for networks to identify. Traditional speech recognition relies on a model that takes up far more space and has no ability to learn over time based on a user. This comes into play when audio files are analyzed, especially those of which of music - recognizing notes from different instruments and recognizing what song or rhythm is being played is a herculean task. Researchers at Dartmouth [5] had binaural recordings synthesized by convolving source audio with head-related impulse responses (HRIRs), which are recordings of wide-band sounds that have been directed at a human head and measured inside the ears using in-ear microphones. The sounds are propagated at regularly spaced angles with respect to the human head.

Another study at the University of Michigan looked at an approach for fully exploiting complex synthetic aperture radar (SAR) data using a convolutional neural network (CNN). A similar normalization technique as the other papers in terms of transformation was also used here, and the benefits of the multitude of inputs was also mentioned. The results detailed that phase detection in images was made significantly easier, due to the increased sensitivity having complex inputs brought. A very detailed image of how images are processed is shown on the next page.

## IV. Discussion

Looking at the research on SAR data, the results of this research suggest that due to the detection and accuracy properties of complex CNNs, it is worthwhile looking into a less-traditional point of view when it comes to neural networks. Image visualization of SAR data is pretty hard to do, and if this is an example of where reconstruction is possible, use cases can be made everywhere from landscape reconstruction to smaller image formats for 3D image data. The music detection research showed complex neural networks, although harder to train, allow for a better recognition of otherwise hidden audio cues in speech and music transcriptions. It showed that complex networks indeed have a place in specific uses, from image identification to identifying music and speech as well. The higher accuracies for these deep complex networks put them above real networks for the tested use cases (music classification, etc). The paper describes how different models are created to help in each of these use cases in terms of architecture.

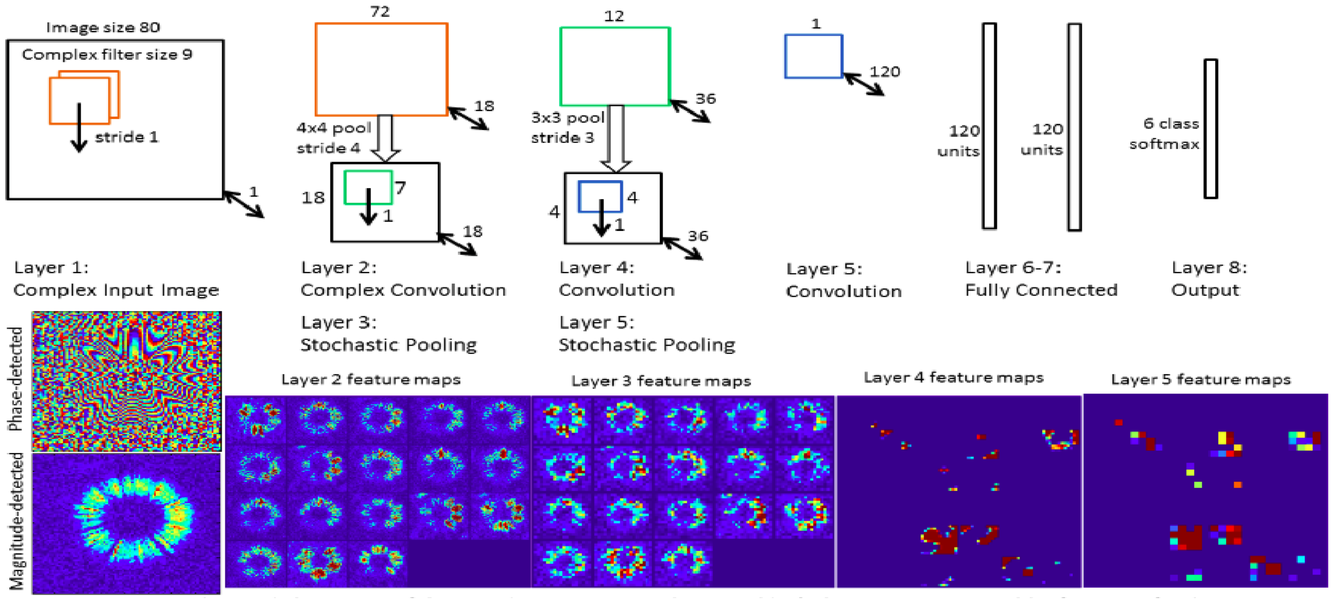


Fig. 5. This image is a very detailed layer-by-layer representation of how a network deals with images, and shows the properties of phase-detection that a complex network enables to be analyzed. [4]

## REFERENCES


1 Image: Guimarães, Daniela Rosa, Maria Thereza. (2018). Experimental measurements and simulation of the fouling phenomena of natural proteins. International Journal of Heat and Mass Transfer. 129. 1085. 10.1016/j.ijheatmasstransfer.2018.10.057.

2 Guberman, Nitzan (2016). On Complex Valued Convolutional Neural Networks. CoRR. abs/1602.09046.

3 Trabelski,Chiheb; Bilaniuk, Olexa; Serdyk, Dimitry, et al. (2017) Deep Complex Networks. CoRR. abs/1705.09792

4 Michael Wilmanski, Chris Kreucher, Alfred Hero, ”COMPLEX INPUT CONVOLUTIONAL NEURAL NETWORKS FOR WIDE ANGLE SAR ATR”, IEEE SigPort, 2016. [Online]. Available: http://sigport.org/1351. Accessed: Feb. 23, 2020.

5 Sarroff, Andy M. (2018) Complex Neural Networks for Audio. Dartmouth College. Thesis