The Relationship Between Deep Learning and Brain Function

Sukhjinder Nahal, Jamal Wilson, Abel Renteria, and Nusseir Moath
Seidenberg School of CSIS, Pace University, Pleasantville, NY 10570

***Abstract*— Humans have been interested in understanding how the brain functions for many centuries. More recently the focus has been on the human visual system as it works coherently with the brain to interpret what we observe. This paper compares the human visual system with deep learning technologies such as Convolutional Neural Networks. A better insight into how the brain learns and recognizes objects can be used to improve current deep learning technologies. This paper discusses and compares how the brain and deep learning receive, process and interpret visual data. As an emerging technology, deep learning has the potential to affect military, medical, law enforcement. It can be of help in situations requiring objects to be identified accurately such as facial recognition, tumor identification, autonomous vehicles, unsupervised learning, and ultimately artificial intelligence.**

***Index Terms*— Brain function, Convolutional Neural network, Deep learning, Pattern Recognition, Visual Cortex, Visual System**

I. Introduction

This session briefly describes how the Human Visual System functions and how it relates to deep learning technologies. Convolutional neural networks are a deep learning technology that come close to the functionality of the human visual system. The technology can be used in various real world applications where object identification or pattern recognition can be of benefit.

Deep learning is a relatively new technology that is still in its infancy. There are many improvements to be made before it reaches or exceed the capabilities of the human brain. It is currently restricted due to hardware limitations as it requires a great deal of computing power. This limitation has been partially alleviated with the use of graphics processing units (GPU). The use of GPUs allows for parallel processing which allows for more computations in comparison to a central processing unit (CPU). Another factor that has made deep learning technology viable is the availability of large amounts of data and the decreasing cost of storage [4]. The brain has evolved over millions of years and is efficient and effective for the actions it performs. Deep learning has to evolve much like the brain to reach the same type of effectiveness and efficiency as the brain.

There are many similarities and differences between deep learning and how the brain functions. The brain uses the eyes as input for visual data. Deep learning on the other hand uses a data set of images as an input. Both the brain and CNNs use neurons to extract information about the object before passing it onto other layers in the system. However, they differ in the way the learn and recognize objects. CNNs are only able to name an object it recognizes whereas a human can provide much more information about the object.

A better understanding of brain functions in processing and learning information can help improve deep learning technology. Current deep learning technology is based on the present knowledge and understanding of how the brain functions. Algorithms have been developed based on theories on how the brain functions since the function of the brain is not fully understood. Deep learning technology will be more efficient and effective if an algorithm can be developed that functions much like the brain.

II. Human Visual System

The human visual system is a complex system consists of many components. These components all work together to allow you to view your surroundings.

The transparent cornea, and the lens all play a role in focusing light as it passes from our environment through the pupil before it is projected onto the region of the eye known as the retina. The data is then sent to the optic nerve which transmits the signal to the brain for processing [10].

The dilating and constricting of the iris which resides between the cornea and lens regulates the amount of light that enters the eye. The light then proceeds to hit the lens which bends the light and projects onto the retina. The retina contains photoreceptors which are cells that are sensitive to light. These photoreceptors have the task of converting the incoming light into nerve impulses that the brain can interpret. Rods and cones are the two types of photoreceptors (light-sensitive cells) which are present in most vertebrates and sit adjacent to the pigment epithelium. Both rods and cones are elongated cells that rely on light sensitive molecules in their outer regions to capture light [6].

The primary function of the pigment epithelium is to maintain the rod and cone photoreceptors. Rods are light sensitive and can only pick up a grayscale image and thus they are mainly for your peripheral vision. Cones are the opposite as they can detect fine detail and color. Humans are considered trichromatic in that they have three different types of cones which are sensitive to red, green, and blue light wavelengths. These cones fire off nerve impulses depending on the wave length of the light that comes in through the pupil and allow you to perceive color. The color of the light is dependent upon the wavelength of the light. Shortwave length light is perceived as blue whereas long wavelength light is perceived as red [7].

The optic nerve is responsible for transferring the nerve impulses from the retina to the visual cortex. Before the nerve impulses reach the visual cortex, it goes through the optic chiasm and then to lateral geniculate nucleus (LGN). The optic chiasm is where the information from the left visual field is sent to the right side of the brain and the right visual field is sent to the left side. The LGN consists of 6 layers which alternates inputs from both eyes [9].

The visual cortex is comprised of 6 different layers as shown in figure 1 V1, V2, V3, V3a, V4, and V5. V1 & V2 are responsible for sorting and sending out the information to the other layers for further processing. These layers process color, form, motion, and size. The other layers are accountable for identification of objects and storage of visual memories.

V1 is the primary visual cortex and is dedicated to processing information about static and moving images. This layer is further divided into 6 sublayers. Layer 4 is where the input from the LGN is received [27]. Neurons in this layer can distinguish color, direction, and orientation. These cells are effective when they are stimulated by bars or edges in certain placements [18]. As shown in figure 1 V1 receives the input and then V2 sends the signals to the other various parts of the brain for further processing. The other sections extract motion, color, and form to complete the image we observe.

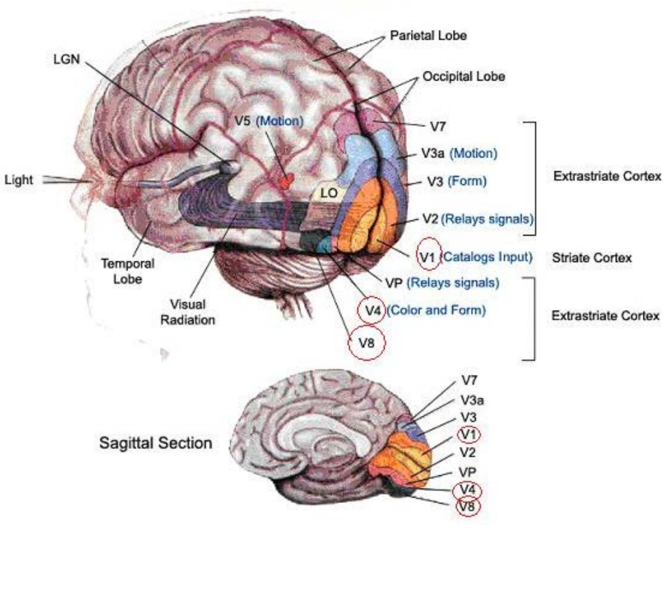

FIGURE 1: VISUAL CORTEX [2]

III. DEEP LEARNING

Deep Learning is a branch of the broader field known as machine learning which is created by using algorithms to mimic the structure and function of the human brain. Deep learning is concerned with creating artificial neural networks which are composed of many layers stacked on top of one another. These layers are inspired by the layers of the human brain and the neuron interaction between that passes information among these layers. In an artificial neural network these multiple layers enable the neurons to store information and pass them up the layers to achieve a more abstract version of the original data. This abstract version of the data helps to make sense of the original data in a way that a machine can be trained to recognize this data.

The technology can be applied to emulate the human visual system. It does so by imitating how neurons function in the neocortex using an artificial neural network.

A Convolutional neural network (CNN) is a type of deep learning technology. It functions similarly to how the human brain processes visual information. CNNs process information much like neurons pass information from one neuron to another. Convolutional Neural Networks (CNNs), first introduced by Yann LeCun in 1989 exhibit exceptional performance in classification tasks especially for object classification [14].

CNNs are comprised of multiple layers of receptive fields which are small neuron collections that process pieces of the inputted image. The layers consist of convolutional layer, RELU (Rectified linear unit) layer, Pooling layer, and a fully connected layer). These layers build upon each other like building blocks to provide you a result. The output of one layer becomes the input to the next layer.

This can be compared to how the brain processes visual information. The raw information is processed by the first layer in the CNN much like it is processed by the primary visual cortex. Then the information is sent to the other layers in the CNN to extract features just like the primary visual cortex.

The convolutional, RELU, and pooling layer are repeated multiple times. When the process is repeated three or more times it is considered deep (i.e. Deep Learning). Then it is followed by fully connected layer to support classification.

The application of CNNs are currently being employed to research and advance: Medical imaging for example MRI based brain tumor imaging.

CNNs have also attempted to caption images, conduct facial analysis, recognize emotions, detect sarcasm, differentiate between types of percussions instruments, classify events in microblogs and detect and track social networking trends, advance speech recognition, detect twitter spam, recognize, detect and segment objects.

As shown in figure 2 CNNs can accurately detect the license plate number from a car. It does so by separating the image of the plate from the car [26].

FIGURE 2: LICENSE PLATE IDENTIFICATION [26]

a) Convolutional Layer

This is the main component in the CNN process and it performs the bulk of the heavy computation. The input of this layer consists of a set of learnable filters. The inputted image is split into smaller receptive fields for processing. For example, if an image is 38x38x3 its receptive field can be 5x5x3. The dimensions of the image are length, width and depth which is the color channels red, green, and blue values. The color values are separated just like how there are different cones in the eye to distinguish color. As these filters are applied it detects certain visual features and creates a 2D activation map.

Figure 3 demonstrates how convolution occurs on an image. In this example, the CNN is trying to determine if the image is an X. CNNs evaluate multiple areas for matches in every possible combination. In the figure the ones indicate a strong match and the negative ones indicate a weak match. The math behind convolution involves multiplying the numbers from the section being evaluated against the features that were learned. Then you sum up the values and divide by the total number of pixels in the image. A resulting value of 1 indicates a match which could be 1X1 or -1x-1. A resulting value of -1 indicates that there is not a match. In the figure, all the values result in a one and thus indicate that it is an exact match. This process is repeated for every possible combination and results in a feature map that contains the values that are generated from this process [20].

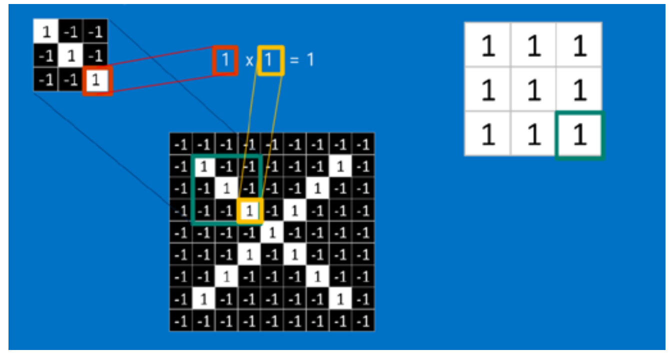

FIGURE 3: CONVOLUTION PROCESS [20]

b) Pooling

The pooling layer is responsible for down sampling the inputted image. It is usually placed in between convolutional layers to reduce the amount of computation. It reduces computation by decreasing the amount of input since the size of the image is decreased. It would take a massive amount of time and computation to process an image that is for example 7 megapixels which would create 7 million data points.

As shown in Figure 4, the pooling process takes the feature map values generated from the convolutional layer and shrinks them down while retaining the important aspects. It does the shrinking by taking the maximum value from sections of the image. In the figure, it takes the maximum value from a grid of 2x2 pixels which effectively reduces the size to a quarter of the original [20].

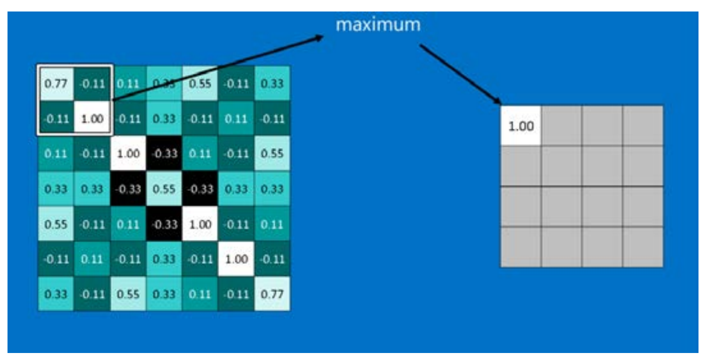

FIGURE 4: POOLING PROCESS [20]

c) ReLU (Rectified Linear Unit)

An issue with CNNs is the computational cost. Even with the advancements of computational power and the use of multiple Graphical Processing Units (GPUS), CNNs face the challenge of matching or exceeding human performance. The computational cost issue was addressed by Krizhevsky et al.

Krizhevsky showed that in practice, CNNs with Rectified Linear Unit Activations (ReLU) neural activations trained faster than other activation functions [13]. The approach simplified turned all negative values into 0.

The ReLU layer acts as a normalization process. Normalization is referred to as regularization in some texts. “A step to keep the math from blowing up by tweaking the values a bit.” According to Bandon Rohrer. The approach simplified turns all negative values into 0 [20]. In the ReLU layer a stack of analyzed images becomes a stack of images with no negative values.

At the 2016 CVPR, the premier annual computer vision conference Ben Haeffele and René Vidal of Johns Hopkins University presented some of the recent improvements in the performance of CNNs. Cited was improved regularization among others. Various techniques are used in the ReLU layer, including weight decay and dropout. The dropout method is cited as performing better than weight decay [23].

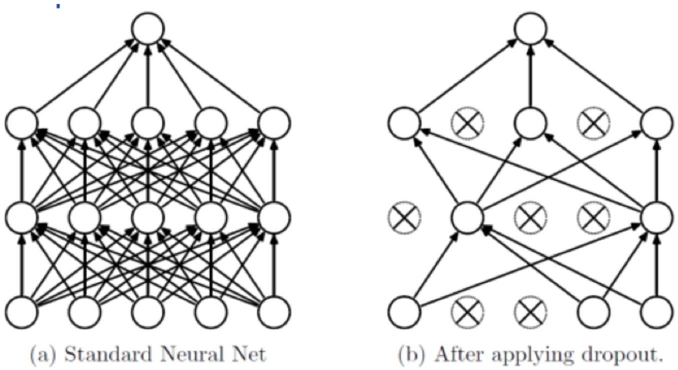

FIGURE 5: EXAMPLE OF DROPOUT IN THE RELU LAYER [23]

Dropout is a normalization function where each element of a layer’s output is kept with probability p. When the output fails to meet the probability criteria the output is set to 0.
When applied to the outputs of the fully connected layer the equation is written as such: r = m * a (W v)

In this equation, r is regularization, m is mask being applied, * the elements product, a is an activation function, W is the weight of the mask, v represents input from the fully connected layer. Other variations apply dropout to the input and the equation is represented as such: r = a (m * W v) [25].

d) Fully Connected Layer

After a few repetitions of Convolution and pooling a fully connected layer is created to connect the neurons between the current layer and the previous layer.

Figure 6 Shows the different layers of the CNN and how they are used to come up with a result. In this case, the input is an image of a car. After many repetitions the image is identified as a car with a high probability [11].

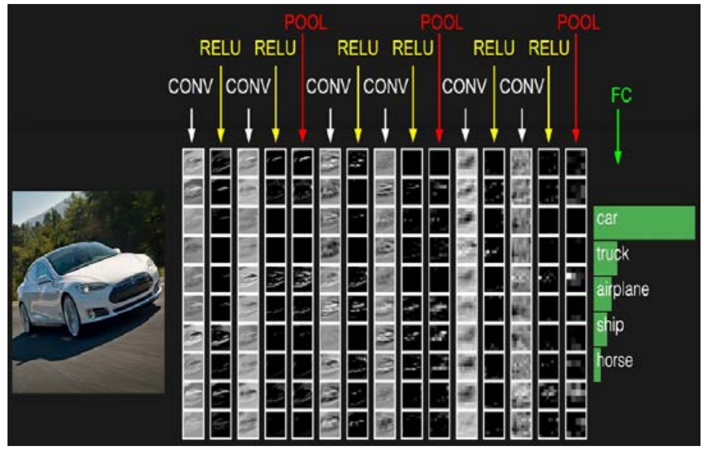

FIGURE 6: CNN LAYERS [11]

The final results are based on weights which are given by the neurons in the network. Each neuron in the neural network assess the image and gives it a weight based on attributes in a process called backpropagation. In figure 6 the neurons would look for things such as wheels, headlights, and windshield to determine if it is a car. According to the weights a probability vector is created which is a highly-educated guess. In figure 6 the first result is a car which shows the highest probability. The next result is a truck which is likely since they share some characteristics. The rest of the results have lower probabilities and are not likely to be the correct identification of the image. The result of the CNN is highly dependent on the quality of the dataset that is inputted. If the CNN gets a dataset that does not accurately represent a car it will make mistakes and will not be as effective at recognizing the object.

IV. BACKPROPAGATION

Backpropagation is a process that generates weights for the inputted image based on the features it has learned. Every image that is processed receives a weight and the error or cost function in this vote can show you how good your learned features are. The error is calculated by subtracting the right answer which would be a 1 by the generated weight. Then gradient descent is used to minimize the error in the weights. This process adjusts the features and weights up and down to see how the error changes. The amount of adjustment is dependent on the size of the error, so a large error will require a large adjustment and a small error will require a small adjustment [20].

Figure 7 shows the computation behind the back-propagation process. The backpropagation process is “nothing more than a practical application of the chain rule for derivatives” [15]. A weighted sum of error derivatives is calculated at each hidden layer.

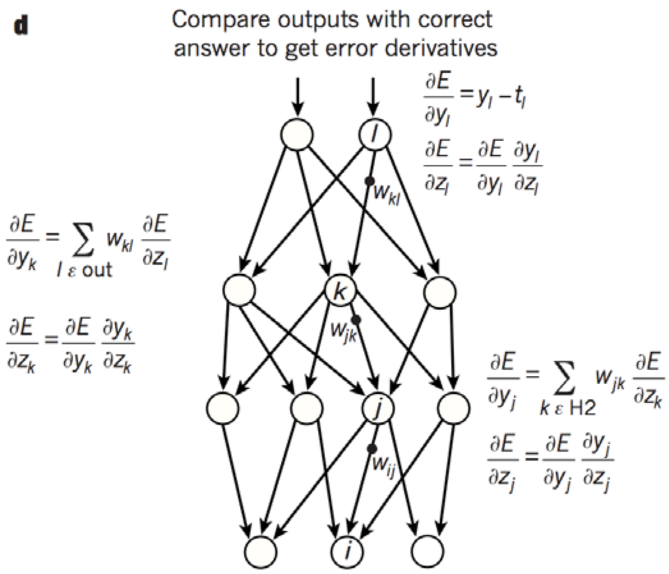

FIGURE 7: BACKPROPAGATION PROCESS [15]

V. COST FUNCTION

Cost functions give an indication of how well a neural network performed given a training sample and its expected output. Deep learning in part is based on efficiently optimizing these cost functions as shown in the backpropagation process.

In the publication "Towards an integration of deep learning and neuroscience" they have 3 hypotheses about the brain and cost functions.

The first hypothesis states that the brain optimizes cost functions. They claim that the brain has mechanisms for credit assignment while learning and adjusts the properties of each neuron to tune the process. This process of assigning cost is a vital component to the learning process and is similar to how CNN functions [16]. CNNs use backpropagation to reduce errors by tuning the weights like neurons in the brain.

The second hypothesis they proposed is that cost functions are diverse across areas and change over development. Neurons are located throughout the brain and thus may optimize things differently. In addition, the cost functions may change over time as the brain develops [16]. Deep learning is currently not able to optimize the learning process over time.

The third hypothesis they proposed is that specialized systems allow efficient solution of key computational problems. Different parts of the brain seem to specialize in certain functions. This is indicated by how the information flows to different regions of the brain as show in figure 10. This is similar to how computer programs have specialized functions to efficiently solve a problem.

They also argue that the brain solves problems differently depending on the problem. The brain might use genetic pre-specification circuits, local optimization or a host of proposed circuit structures that would allow it to perform backpropagation.

VI. MEMORY

Memory or storage is a vital part in the learning process of both the human brain and in deep learning. Memory in deep learning networks must store input data, weight parameters, and other computed data [8]. Deep learning uses GPU DRAM (dynamic random access memory) and external DRAM as storage. The brain on the other hand uses neurons and synapses throughout the brain as memory.

Three types of memory are need for learning. The first type of memory needed is a long-term memory (content addressable memory). It is used to store information for a long period of time, which can be accessed later. The second type of memory is short-term memory (working memory). It is used to store information for a short period that can be overwritten rapidly. The last type of memory needed is an implicit memory which is unconscious memory. This type of memory is learned using gradient descent on errors [16].

a) Content Addressable Memory

Content addressable memory allow us to recognize a pattern that we have seen before. Including such memory in deep learning "allows deep networks to learn to solve problems that previously were out of reach" [16] Current deep learning technology is able to extract features from

b) Working Memory

Working memory is a short-term memory that can be overwritten quickly. It is a vital for human like intelligence such as reasoning as it provides short term storage for processes [16].

c) Implicit Memory

Implicit memory is a type of long term memory that allows you to use "past experiences to remember things without thinking about them" [28]. For example, activities such as walking, riding a bike, and driving a car are things you do without thinking.

Tagging the importance of memory can be of beneficial as it can remove unnecessary data from the training set. The brain contains a salience network as indicated by the black region in figure 8. It "is a collection of regions of the brain that select which stimuli are deserving of our attention." [24] The brain receives various sensory inputs and therefore must prioritize a given sense depending on the situation. The salient network is considered to play an integral part in preparing the brain for an event [22]. For example if you feel pain in some part of your body, your attention is completely directed to the area of concern.

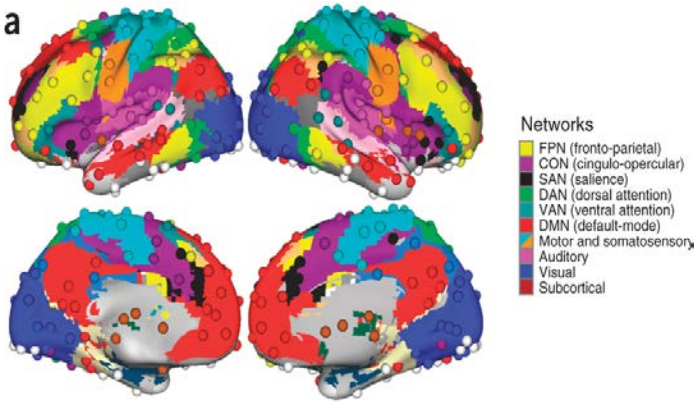

FIGURE 8: FUNCTIONAL NETWORKS OF THE BRAIN [3]

VII. SUPERVISED VS UNSUPERVISED TRAINING

Deep learning can learn in two different methods, a supervised method and the other is unsupervised. Supervised training requires a large amount of data that is labeled for training purposes. Unsupervised training on the other hand does not require labeled data. It can sort and classify the data it is given without human intervention.

CNNs are trained using the supervised learning method. Thus, require a large dataset that is labeled to create feature maps. Which are used to identify an object that you input. The brain's architecture of neurons is comparable to convolutional neural networks but they differ in the way they learn [5].

Section A of figure 9 demonstrates how the supervised learning occurs using labeled data [16]. The brain can also learn using the supervised method but it is able to use it in conjunction with unsupervised learning. Section B of figure 9 demonstrates how the brain uses supervised and unsupervised learning. Section C of figure 9 shows how information enters the brain from the sensory inputs and the outputs that are generated as a result. The brain has different sections that specialize in a given function. This is similar to the way a specialized algorithm can be optimized for a specific function thus increasing its efficiency [16]. The section below that shows how a deep learning network functions in comparison. The deep network takes the inputs and computes a cost function which it passes on to the other layers in the network for further processing.

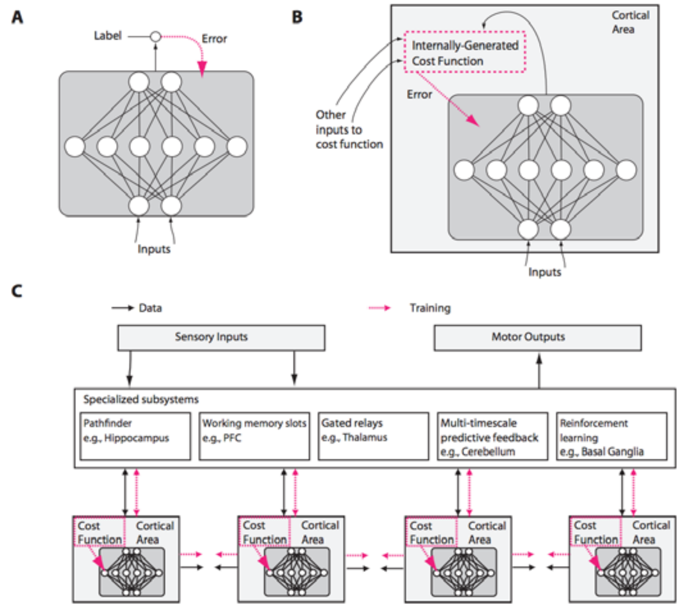

Figure 9: DIFFERENCES BETWEEN CONVENTIONAL AND BRAIN-LIKE NEURAL NETWORK DESIGNS [16]

Differences between CNN and human visual system is that CNNs are good at labeling things. The human visual system is not a labeling machine but recognizes surrounding objects and allows you to make sense of it. For example, if you see a flower outside you might not know the exact name but you know other details about it. You may know how it smells or the season it grows in but may not be able to identify the name.

The world around us gives us a limited amount of information that we can use for supervised learning. However, there is a large amount of data available for unsupervised learning since it doesn't have to be manually labeled. The caveat with unsupervised learning is that it may not learn things as a human would and thus may not make sense to us [16].

Humans can infer information about an object based on how it looks without knowing the exact name. Currently Deep learning is not able to logically think about objects as humans can. The technology is far away from human level performance since it is only capable of labeling objects based on user inputted data [5].

VIII. IMAGENET

ImageNet is a project that contains a large set of visual data which has been hand classified. It is used to train and test deep learning technology.

ImageNet is a widely-used dataset because it contains more than fourteen million annotated images organized by the semantic hierarchy of WordNet. [21]. As of the publication of this article ImageNet contains 14,197,122 images with 21841 synsets indexed [21]. A "synonym set" or "synset" is a concept represented in the image and labeled with a word, multiple words or word phrases.

In the publication “Deep Learning: mathematics and neuroscience” Tomaso Poggio states “Deep networks trained with ImageNet seem to mimic not only the recognition performance but also the tuning properties of neurons in cortical areas of the visual cortex of monkeys.” [19] This shows that deep learning technology can perform like the brain if it is given a proper labeled dataset. Tomaso Poggio also states that networks that are trained using ImageNet perform well on other datasets with very little additional training.

ILSVRC (ImageNet Large Scale Visual Recognition Challenge) is an annual held contest. In this challenge participants train their algorithm using images from a dataset and then automatically label a set of test images. These predicted labels are then sent to a server that evaluates the correctness of the labels [21].

The challenge consists of three different tasks [21].

1) **Image classification** in which the algorithm identifies all the objects in an image.

2) **Single object localization** in which the algorithm identifies one instance of each object category and their location in the image.

3) **Object detection** in which the algorithm identifies multiple instances of an object and their locations.

GoogLeNet is a deep learning algorithm that was developed by google. It was entered into the ILSVRC in 2014 and placed 1st in the image classification and object identification tasks. The error percentage for the image classification task was 6.66 percent. The average precision percentage for object detection was 43.93 percent. The algorithm placed second for single object localization with an error percentage of 26.44 percent.

In the publication “ImageNet Large Scale Visual Recognition Challenge” they conducted an experiment between 2 human annotators and GoogLeNet [21]. Both annotators A1 and A2 trained using a set of images and annotated a set of test images. The first annotator (A1) trained using a set of 500 images and then annotated 1500 images. The second annotator (A2) trained using a set of 100 images and then annotated 258 images. The GoogLeNet algorithm was trained using a much larger dataset of 100,000 images. The results of the experiment are shown in the figure 10. The first annotator was able to achieve a lower classification error than GoogLeNet even though they trained using a smaller set of training images. This shows that the brain can be trained using a much smaller set of images and is able to outperform deep learning technology. The second annotator achieved a much bigger error percentage which can be attributed to smaller set of training data.

Relative Confusion	**A1**	**A2**
Human succeeds, GoogLeNet succeeds	1352	219
Human succeeds, GoogLeNet fails	72	8
Human fails, GoogLeNet succeeds	46	24
Human fails, GoogLeNet fails	30	7
Total number of images	1500	258
Estimated GoogLeNet classification error	6.8%	5.8%
Estimated human classification error	5.1%	12.0%

FIGURE 10: CLASSIFICATION ERROR HUMAN VS DEEP LEARNING

IX. SCALABILITY OF DEEP LEARNING

Deep learning can come up with better results when it is given larger sets of data as an input. In comparison, older learning algorithms don’t scale as well. As shown in figure 11 the performance of deep learning increases as the amount of data increases [1]. We have access to large datasets now due to the decreasing cost of storage. This is one of the reasons that deep learning has become a feasible technology.

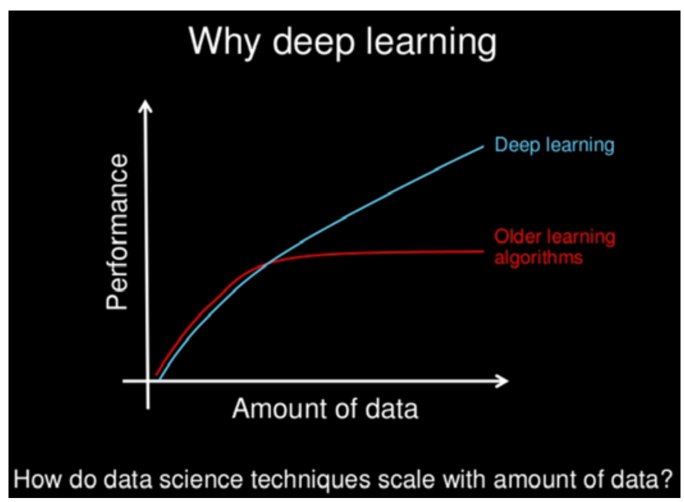

FIGURE 11: SCALABILITY OF DEEP LEARNING [17]

X. LNP MODEL (LINEAR-NONLINEAR-POISSON CASCADE MODEL)

Linear-nonlinear-Poisson cascade (LNP) is a popular model that mimics the functional behavior of neurons. This model is similar to how neurons in deep learning function. The LNP model consists of three steps. The first is estimating linear filters. The next stage consists of transforming the output of the filters an instantaneous firing rate. The last stage then generates a poisson spike from the nonlinear transformations.

XI. Conclusion and Future work

The human visual system is a complex system that allows you to view the world. Light is inputted from the eyes which is then converted into impulses that the neurons send to the brain for processing. This functionality is similar to how CNNs function. They both have some sort of input and use a network of neurons to evaluate and recognize the image. They differ in the way the learn. CNNs learn by using a set of labeled images and extract features from the images which it uses to recognize objects.

Currently deep learning technology is bottlenecked because of hardware and the quality of data used for training. Deep learning as of now is employing GPUs for computation since it is much faster than CPU computation. The accuracy of the deep learning technology is dependent on the inputted data set.

As advances in the field of deep learning are made the visual learning systems will improve and come close to the ability of the human visual system. Currently CNNs learn using the supervised method, where data is labeled manually and put through the algorithm. Also, deep learning is not able to infer information about an object based on how it looks as humans capable of. In the future, there may be a deep learning technology that may learn without human involvement and be able to infer information. This would remove the manual process of finding, sorting, and labeling images which in turn will save time.

A better understanding of how the brain learns will provide valuable information and help extend the capabilities of current deep learning technology. Our current understanding of how the brain fully functions is limited. Thus, current deep learning technologies are based on assumptions and theories as to how the brain functions in attempt to emulate it. A better insight will lead to better systems that will provide users with useful information in a quick and efficient manner.

The application of visual deep learning in the real world are infinite. Various fields can benefit from such a technology. For example, it can be used in autonomous car systems. In this application, it can be used to detect red lights, pedestrians, and vehicles. Then using that data, it can drive the car to a destination without hitting obstacles and abide by the laws. Another application is that it can be used to identify people or license plates in security footage which can greatly help law enforcement. In addition, it can be used to help the blind navigate the world with the use of some type of wearable technology and haptic feedback [12]. It can also be used in the medical field to detect cancer or other illnesses and detect it before it becomes deadly.

XII. References

[1] Brownlee, Jason. "What is Deep Learning?" Machine Learning Mastery. August 16, 2016. Accessed March 2017

[2] Claffey, Mike. "Notes: Vision." 2013. Accessed April 03, 2017.

[3] Cole, Michael W., Jeremy R. Reynolds, Jonathan D. Power, Grega Repovs, Alan Anticevic, and Todd S. Braver. "Multi-task connectivity reveals flexible hubs for adaptive task control." Nature Neuroscience 16, no. 9 (2013): 1348-355.

[4] Copeland, Micheal. "What’s the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning?" July 29, 2016. accessed March 2017

[5] Dettmers, Tim. "The Brain vs Deep Learning Part I: Computational Complexity — Or Why the Singularity Is Nowhere Near." July 27, 2015.

[6] Dowling, John E., and Joseph L. Dowling. Vision: How It Works and What Can Go Wrong. MIT Press, 2016. accessed March 2017

[7] Dowling, John E., and Joseph L. Dowling. "Capturing Light—The Photoreceptors: Retinitis Pigmentosa and Age-Related Macular Degeneration." In Vision: How It Works and What Can Go Wrong, 45-72. MIT Press, 2016. accessed March 2017

[8] Hanlon, Jamie. "Why is so much memory needed for deep neural networks?" January 31, 2017. Accessed April 10, 2017.

[9] Heeger, David. "Perception Lecture Notes: LGN and V1." accessed March 2017

[10] Ihaka, Ross. "Human Vision." accessed March 2017

[11] Karpathy, Andrej. "CS231n Convolutional Neural Networks for Visual Recognition." accessed March 2017

[12] Kobielus, James. "Algorithms for eyes: How deep learning can help the blind." InfoWorld.com, June 22, 2015. Computer Database. accessed March 2017

[13] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E.. 2012. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, P. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.). 1106–1114. accessed March 2017

[14] LeCun, Yann, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4):541{551, 1989. accessed March 2017

[15] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." May 28, 2015. Accessed April 3, 2017.

[16] Marblestone, Adam Henry, Greg Wayne, and Konrad P. Kording. "Towards an integration of deep learning and neuroscience." 2016. doi:10.1101/058545. accessed March 2017

[17] Ng, Andrew. "Extract Data Conference." SlideShare. Accessed March 2017

[18] O'Reilly, Randall C., Yuko Munakata, Michael J. Frank, and Thomas E. Hazy. "Perception and Attention." In Computational Cognitive Neuroscience. Wiki Book. accessed March 2017

[19] Poggio, Tomaso. "Deep Learning: mathematics and neuroscience." The Center for Brains, Minds and Machines (CBMM). April 26, 2016. Accessed April 14, 2017.

[20] Rohrer, Brandon. "How do Convolutional Neural Networks work?" August 18, 2016.

[21] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision 115, no. 3 (2015): 211-52.

[22] Singer, Emily. "Inside a Brain Circuit, the Will to Press On." Inside a Brain Circuit, the Will to Press On. December 5, 2013. Accessed April 14, 2017.

[23] Srivastava, et al, "Dropout: a simple way to prevent neural networks from overfitting." JMLR, 2014.

[24] Uddin, Lucina Q. Salience network of the human brain. Amsterdam: Elsevier/Academic Press, 2017.

[25.26] Wan et al, “Regularization of neural networks using dropconnect.” ICML, 2013

[26] Zang D, Chai Z, Zhang J, Zhang D, Cheng J; Vehicle license plate recognition using visual attention model and deep learning. J. Electron. Imaging. Accessed March 2017

[27] Zeki S, JD Watson, CJ Lueck, KJ Friston, C Kennard, RS Frackowiak. “A direct demonstration of functional specialization in human visual cortex”. Journal of Neuroscience 1 March 1991, 11 (3) 641-649, accessed March 2017

[28] Zimmermann, Kim Ann. "Implicit Memory: Definition and Examples." LiveScience. February 12, 2014. Accessed April 14, 2017.