Useful Applications Based On Deep Learning Technology

Deep Learning (DL) is a new research direction in the field of Machine Learning (ML). DL was imported into this field to make machine learning achieve closer its original goal -- Artificial intelligence (AI). The basic idea of deep learning is to learn the inherent laws and representation levels of the sample data. The information obtained during these learning processes is very helpful for the interpretation of data such as text, images and sounds. Its ultimate goal is to enable machines to analyze and learn like humans, such as recognize the sample data. Deep learning is a complex machine learning algorithm that achieves far more results in speech and image recognition than previous related technologies. Deep learning has achieved many results in search technology, data mining, machine learning, machine translation, natural language processing, multimedia learning, speech, recommendation and personalization technologies, and other related fields. Deep learning technology solves many complex pattern recognition problems by enables machines to imitate human activities such as audiovisual and thinking. This approach has made great progress in artificial intelligence-related technologies [4].

Deep learning is a general term for a class of pattern analysis methods. As far as specific research content is concerned, three types of methods are mainly involved which are convolutional neural network, autoencoder and deep belief networks.

Convolutional Neural Network (CNN) is a neural network system based on convolution operations. It contains eight learned layers — five convolutional and three fully-connected. The convolutional neural network mimics the visual perception mechanism of organisms in order to perform supervised and unsupervised learning. The convolutional kernel parameter sharing in the hidden layers and the sparseness of the connection between the layers allow the convolutional neural network to minimize the amount of calculation to learn grid-like topology features, such as pixels and audio. As one of the most important representative algorithms of deep learning, convolutional neural networks have the ability to learn from appearance, that is, they can extract high-level characteristic from the input information it received. Specifically, the convolutional layer and the pooling layer in the convolutional neural network can respond to the transition invariance of the input characteristic. It can identify similar features located in different positions in the same space. The ability to extract transition invariant is one of the reasons why convolutional neural networks are used in computer vision problems. The transfer of transition-invariant features inside convolutional neural networks has a general law. In the image processing problem, the feature map at the front of the convolutional neural network usually extracts the most representative highest-frequency and lowest-frequency features; then the pooled feature map will show the aliasing artifacts of the input image [1]. When the signal enters in deeper hidden layer, its more general and complete features will be extracted. The hidden layer features of convolutional neural networks are visualizable by deconvolution and un-pooling the layer. In a successful convolutional neural network, the feature map that passed to the fully connected layer should contain the same features in the learning target, for example, the complete image of each category under the image classification.

Autoencoder (AE) is a type of artificial neural networks used in semi-supervised and unsupervised learning. Its function is to form representation learning by using the input information as the learning target. The autoencoder includes two parts: encoder and decoder. According to the learning paradigm, autoencoders can be divided into undercomplete

autoencoders, regularized autoencoders, and variational autoencoders. The undercomplete autoencoders and regularized autoencoders belongs to discriminative models, but variational autoencoders belongs to generative models. Depending on the type of construction, the autoencoder can be defined as feedforward or recursive neural network [2]. Autoencoders have the function of representing learning algorithms in a general sense which are applied to dimensionality reduction and anomaly detection. Autoencoders with convolutional layer construction can be applied to computer vision problems, including image denoising, neural style transfer, and etc.

Deep belief networks (DBN) is a probabilistic generation model which in contrast to the traditional discriminative model of neural networks. The generation model is to establish a joint distribution between data observations and labels. DBN can be used to identify features, classify data, and also can use it to generate data. The DBN algorithm is a very practical learning algorithm which applied to a wide range of applications and has strong scalability. It can be applied to the fields of machine learning, handwriting recognition, speech recognition, and image processing. DBN is composed of multilayer neurons. These neurons are divided into dominant neurons and recessive neurons. Dominant neurons are used to accept input, and recessive neurons are used to extract features. Therefore, the recessive neurons are also called feature detectors. The connections between the top two layers are undirected and form associative memory, but lower layers have direct connections from up and down. The bottom layer represents data vectors, and each neuron represents one dimension of the data vector [3].

There are many common structures among these three methods. For example, they are all formed multi-layer processing, in which gradually transforming the initial low-level feature representation into high-level feature representation, and all using simple models to complete complex classification and other learning tasks. Therefore, deep learning can be further defined as feature learning and representation learning process. In recent years, researchers have gradually combined these types of methods, such as unsupervised pre-training of convolutional neural networks based on supervised learning and self-encoded neural networks. Compared with traditional learning methods, deep learning methods preset more model parameters. In this case, model training in deep learning process is more difficult. According to the general law of statistical learning, the more model parameters, the larger the amount of data required participate in data training. During the early of 20^{th} century, the amount of data available for analysis was not enough due to the limited computing power of computers and the limitations of related technologies. Deep learning did not show excellent recognition performance in pattern analysis. Since 2006, Dr. Hinton proposed the CD-K algorithm to quickly calculate the weights and deviations of restricted Boltzmann machine (RBM) networks, RBM has become a powerful tool for increasing the depth of neural networks, leading to the widespread use of DBN which has been used by companies such as Microsoft in speech recognition and other deep networks. At the same time, sparse encoding is also used in deep learning because it can automatically extract features from data. Different from traditional shallow learning, deep learning is designed to establish an appropriate number of neuron computing nodes and multi-level computing

hierarchies, select the appropriate input and output layers, learn and tune the network to establish a functional relationship from input to output, although it cannot be one-hundred percent but it can approximate the actual relationship as much as possible. Using the successfully trained network models, deep learning can then achieve its automation requirements for completing the complex transactions.

The applications of deep learning are very wide, including speech recognition, face recognition, self-driving car, artificial intelligent, and computer vision.

Speech recognition technology, also known as Automatic Speech recognition (ASR), it is aims to convert vocabulary content in human speech into the computer-readable inputs, such as keys, binary codes, and character sequences. Speech recognition working like the auditory system of a machine, which allows the machine to convert speech signals into corresponding text and commands through the machine recognition and understanding. The earliest research on speech recognition technology started in the 1950s. In 1952, Bell Labs developed a recognition system for 10 isolated numbers. Since the 1960s, Reddy and other professions at Carnegie Mellon University in the United States have carried out research on continuous speech recognition, but related research area was developed slowly in this period. Pierce J of Bell Labs even stated speech recognition development is impossible in an open letter in 1969. In the late 1980s, the artificial neural network (ANN), the predecessor of deep neural network (DNN), has also become a new direction of speech recognition research. But this kind of shallow study of the neural network does not perform efficiently on speech recognition tasks, and its performance not even as good as the GMM-HMM model. In 2006, Dr Hinton proposed to use a restricted Boltzmann machine to initialize the nodes of a neural network, that is, a deep belief network which solves the problem that it is easy to fall into a local optimum solution during the training of deep neural networks. In 2011, the deep neural network-based modeling method officially replaced GMM-HMM as the main and most important speech recognition modeling method [5].

The principle of speech recognition is actually not complicated. In order to extract features more effectively, it is often necessary to perform preprocessing tasks such as filtering and framing the collected sound signal and extracting the signal to be analyzed from the original signal. After that, feature extraction works convert sound signals from the time domain to the frequency domain to provide appropriate feature vectors for the acoustic model; the acoustic model then calculates the score of each feature vector on the acoustic features according to the acoustic characteristics; The language model is based on the linguistic-related in theory to calculate the probability that this sound signal corresponds to any possible phrase sequence. Finally, the phrase sequence is decoded according to an existing dictionary to obtain the last possible textual representation. As the premise and basis of speech recognition, the preprocessing stage of speech signals is very important. In the final template matching, the characteristic parameters of the input speech signal are compared with the characteristic parameters in the

template library. Therefore, only when the characteristic parameters that can characterize the essential characteristics of the speech signal had obtained in the pre-processing stage, then these characteristic parameters can be matched for speech recognition in a high recognition rate. The pre-process stage generally implemented by using a bandpass filter to set the upper and lower ring frequencies for filtering, and then quantizing the original discrete signal to make this stage. After the signal pre-processing stage is completed, the key feature extraction operation is performed. Currently, the most commonly used feature parameters in mainstream research institutions are Linear Prediction Cepstrum Coefficient (LPCC) and Mel Frequency Cepstrum Coefficient (MFCC). These two characteristic parameters operate on the speech signal in the cepstrum domain. LPCC taking the vocalization model as the starting point and uses Local Procedure Call (LPC) technology to obtained cepstrum coefficient. MFCC simulates the auditory model and uses the output of the speech through the filter bank model as acoustic features, and then uses the discrete Fourier transform (DFT) to obtain the value. After figured out the feature parameters, now it is time to study about human language model. The language model is mainly used to describe the habit of the expression of human language and emphasizing the internal relationship between words and the arrangement of words. In the process of speech recognition decoding, the intra-word transfer refers to the vocal dictionary and the inter-word transfer refers to the language model. A good language model can not only make the decoding process being more efficiency, but also improve the recognition correction rate. Language models have been widely used in the fields of speech recognition, machine translation, and emotion recognition. After that, it is about decoding. The decoder is the core component of the recognition stage. It decodes the speech by its trained model which generates a recognition lattice for subsequent components to perform the next step. The core algorithm of the decoder is the dynamic programming algorithm called Viterbi. Due to the huge decoding space, we usually use a token passing method that limits the search width in practical applications.

In general, the process of speech recognition can be divided into few steps. First of all, analyze and process voice signals in order to remove redundant information, and then extract key and feature information that affects speech recognition and the meaning of the language. Use the smallest unit to identify words according to their respective grammars in different languages in order. Taking the context relation as an auxiliary recognition condition to help with the recognition. According to the semantic analysis, divide the key information into paragraphs, take out the identified words and connect them again by adjust the sentence composition according to the meaning of the sentence. Finally, making appropriate corrections to the statements that being processed currently.

Face recognition is a popular field of computer technology research, including technologies such as face tracking detection, image magnification automatic adjustment, night infrared detection, exposure intensity automatic adjustment. This technology is based on the facial features of the person. First of all, the input face image and video stream will be

determined. If there is a face, then the position and size of each facial organs are further given. Based on this information, the identity characteristics contained in each face are extracted and compared with known faces to identify the identity of each face. The generalized face recognition actually includes a series of related technologies for building a face recognition system, including face image acquisition, face positioning, face recognition preprocessing, identity confirmation, and identity search. There are some main face recognition methods. Geometric method calculates geometric relationship between eyes, nose, mouth, etc. (such as the distance between each other). This algorithm advantages in fast recognition speed and require less memory but have a lower recognition correction rate as its disadvantage part. Eigenface method is a face recognition method based on KL transform. KL transform is an optimal orthogonal transform for image compression. A high-dimensional image space is transformed into a new set of orthogonal bases after KL transformation. The important orthogonal bases are retained, and these bases can be expanded into a low-dimensional linear space. If it is assumed that the projection of the human face in these low-dimensional linear spaces is separable, then these projections can be used as the feature vectors for recognition, and this is the basic idea of the eigenface method. But the disadvantage is this method require lots of training samples and it cannot be one-hundred percent accurate. The elastic map matching method defines a distance in a two-dimensional space that has a certain invariance to ordinary face deformation and uses attribute topology maps to represent faces. Each vertex of the topological graph contains a feature vector, which is used to record information about the face near the vertex position. This method combines grayscale characteristics and geometric factors and allows image elastic deformation during comparison. It has achieved good results in overcoming the impact of facial expression changes on recognition, and no longer requires multiple samples training for a single person. Line Hausdorff Distance (LHD) method is based on the line segment map extracted from the gray image of the face. It defines the distance between two-line segment sets. What is different is that LHD does not establish a one-to-one correspondence between different line segment sets, so it is more able to adapt to small changes between line segment graphs. Experimental results show that LHD performs very well under different lighting conditions and different postures when doing the face recognition, but it does not recognize well under a huge facial expression change. Support Vector Machine (SVM) method is a new hot spot in the field of statistical pattern recognition. It tries to make the learning machine achieve a compromise in between empirical risk and generalization ability, thereby improving the performance of the learning machine. The support vector machine mainly focusing on solves two-classification problem. Its basic idea is to try to transform a low-dimensional linearly inseparable problem into a high-dimensional linearly separable problem. The usual experimental shows that SVM has a good recognition correction rate, but it requires large number of training samples (300 per class), which is often unrealistic in practical applications. Moreover, the SVM takes a long time to train and the method is complicated to implement. Compare to other biometric face recognition technologies, face recognition advantages in it does not required contact, the user does not need

to be in direct contact with the device and face image information can be proactively obtained. It can be happened concurrency, that is, multiple faces can be sorted, judged, and identified in the same time. But face recognition also has its opposite site, for example, it is sensitive to the surrounding light environment, in which may affect the accuracy of identification. Obstacles such as hair and ornaments on the human face may affect identify accuracy.

Self-driving cars are a type of smart car which rely on computer-based intelligent pilots in the car to achieve the purpose of self-driving. According to the latest Thomson Reuters report on intellectual property and technology, from 2010 to 2015, there were more than 22,000 invention patents related to car driverless technology [7]. Since the 1970s, the United States, Britain, Germany and other developed countries have begun to conduct research on driverless cars. The world's most advanced self-driving cars have been tested for nearly half a million kilometers, of which the last 80,000 kilometers were completed without any human safety intervention. Like many other things, self-driving car also has a process of progressive development of technology. The first phase is the assisted driving phase. Assisted driving functions such as lane keeping and adaptive cruise belong to this stage of technology, but the driver is still the main operator. The second phase is the semi-autonomous driving. At this stage, computer-controlled self-driving car can already complete the process of going to the destination. It can be used as a backup system, but due to factors of country laws and regulations, it still cannot be the subject of the entire driving process. The third phase is the fully self-driving stage. By that factors such as technology, cost, and regulatory deregulation are no longer affect it, which means computer-controlled systems already take role as the main driving subject, and drivers can also take over the operating system at any time, but due to technical and regulatory restrictions, most of the current self-driving vehicles are in the second stage.

Google can be regarded as the earliest Internet company which doing the cross-border research and development the self-driving vehicles. At the same time, relying on its unique map and big data computing resources, Google has a leading advantage in this field. Google ’s self-driving vehicles use a 64-bit 3D lidar provided by Velodyne to map the surrounding environment into a 3D map, and then combined the map with Google ’s high-precision maps, using computers and the cloud network performs big data processing, and finally accomplish the automatic driving function. Early Toyota Prius prototypes were equipped with video cameras, lidar, position sensors, and ranging radars. The video camera is used to judge traffic lights and any moving objects; the laser radar is used to form a 3D map of the real road environment; the ranging radar is used to detect obstacles around the vehicle. Once an object approaches, the vehicle will automatically slow down. Position sensors at the rear wheels are used to detect and estimate the vehicle's lateral position offset to determine the vehicle's position on the map. After years of testing, Google launched its own self-driving prototype. This prototype is also equipped with many radars and sensors, as well as lidar towering on the roof. Google ’s self-driving car

has eliminated the steering wheel, and the car is completely controlled by an on-board computer. It is the closest car prototype to the self-driving concept.

Technically speaking, before 2014, target detection usually used a more traditional method. First, think of a way to generate some candidate boxes, and then extract the characteristics of each box, such as Histogram of Oriented Gridients (HOG). Finally, use a classifier to confirm whether this box is the target object or not. There are also many ways to generate candidate frames, such as sliding preselected frames with different sizes in the picture, or like the Selective Search algorithm that generated frames based on the texture and other characteristics of the picture itself. However, since 2013, with the development of deep learning-related technologies, new models have appeared continuously which achieve end-to-end training and network detection, and this technology significantly improved the efficiency compared to traditional methods. Region-CNN (R-CNN) is an earlier model proposed to solve the object detection using deep learning. The idea is to first use the selective search algorithm to extract a certain number of candidate regions, then use CNN to extract features for each candidate region, and then add proposed features followed by regression and Support Vector Machine(SVM) classification to predict the position and category of the target object. The disadvantage of R-CNN is obvious. The entire process is divided into several steps and cannot be completely trained. In addition, because each candidate frame feature is calculated independently, the entire process includes lots of redundant calculations. Fast-RCNN is an improved version based on this, which mainly solves the problem of redundant calculation when extracting features. First convolve the entire image, extract features to get a layer of feature map, and then extract features of each candidate frame directly on this feature map. However, the size of each candidate frame is different, and the features must be fixed in length for classification and regression. In order to solve this problem, the Roi-Pooling layer is proposed in Fast-RCNN, which can extract features of fixed dimensions for regions of different sizes, so that subsequent classification and regression can operate normally. This model reduces a large number of redundant calculations and improves the speed of the entire model. Faster-RCNN is our last step about to solve this question perfect. The problem with Fast-RCNN is that the extraction of candidate regions is still using the selective search algorithm, which disrupts the continuity of the entire model. Faster-RCNN proposed a Region Proposal Network (RPN) structure in order to improve this. RPN can extract many candidate frames of different sizes and shapes at each position of the feature map, which is also called anchor. Each anchor is followed by a binary classification to determine whether the anchor is the background or not, and then a regression is used to fine-tune the position. Specific categories and locations will be further adjusted at the end of the network. At this point, the entire target detection process can be trained end-to-end. Segmentation is the second critical component in developing self-driving car. Semantic segmentation is more commonly used in self-driving technology. For example, road surface division, crosswalk division, and so on. The early and classic model of semantic segmentation is Fully Convolutional Networks (FCN). FCN

has several classic improvements. The first is to replace the fully connected layer with a full convolution layer. The second is a smaller resolution feature map after convolution. After sampling from the upper layer, the result of the original resolution is obtained. Finally, FCN uses a cross-layer connection that can better combine the high-level semantic features with the bottom-level location features, and making the segmentation result more accurate [8].

Artificial intelligence is an inevitable trend in the development of intelligent robots. Among them, deep learning occupies a pivotal position in artificial intelligence development field. It completely changes the traditional robot's image and speech recognition technology and solves the basic problem of robot positioning and navigation in a better way. It has achieved making the most powerful robot vision and hearing technology. In practical applications, the robot needs to extract text information when detecting the position of the text in the library. At this time, it often encounters the issue about things like text adhesion. It is necessary to use a picture of the text area with incomplete adhesion to train the neural network, so that not only the text position can be obtained. And it also could avoid missing position detection issues. In the process of object recognition and large-scale natural scene recognition, convolutional neural networks and super-pixels can be combined with DBM, which is like what we had seen in the speech reignition. The CNN is used to preprocess large-scale scene images to obtain volumes. After accumulating features, the results are used as the visual layer input of a DBM for feature extraction. After that, the scene is classified using the Soft-max classifier. Super-pixels are formed by preprocessed image with simple linear iterative clustering algorithm and then aggregate pixels that has similar distance and color. In this way, it can make the image outline clearer, and can also handle complex scene graph. The whole process is similar to the method we used in the speech recognition and face recognition. In indoor scenes, it is necessary to realize the association between indoor 3D maps and semantic information. Use decentralized modular technology to enable the robot to perform scene object recognition and map reconstruction at the same time, so it will be able to realize its indoor identification function. Based on the 3D scene construction technology RGB-D information and combine with the depth-first algorithm to perform 3D reconstruction of the environment map. Introduce an object recognition system based on a convolutional deep learning model to realize the identification and classification of indoor environmental items。 Finally, in order to solve unsynchronized system problem, create a method to increase synchronization identification is necessary. Using all these methods can solve the problem about rebuilding reliable environmental maps for the robots indoor [9].

Computer vision benefits by deep learning technology in many parts, such as image classification. The image classification task is to assigning labels to images. For example, a photo can be classified as a day or night shot. In addition, in the field of transportation, image classification can be used to detect whether a car is in a parking space or not, that is, it can determine whenever the parking space is occupied. Secondly, Image reconstruction tasks is to reconstructing missing or damaged parts of an image. This task can be considered as a photo

filter and transformation without objective evaluation. Although it is indeed possible to ensure that the visible attributes of the image are closely matched. But asking the computer to recreate details without reference is obviously unreasonable. Therefore, image reconstruction systems have some limitations, which largely depend on how many original images are available for learning. Finally, an important goal of computer vision is to be able to identify events that occur over a certain of time. One example is object tracking, where the goal is to track a specific object in an image or video. Object tracking is important for almost all computer vision systems that contain multiple images. For example, in football training, time tracking position information of each player can be obtained through target tracking, and scientific training can be carried out by studying its physical and tactical characteristics [11].

In addition to these four mainstream deep learning applications, deep learning has expanded many other amazing applications, for example, an app called Face2Face, this system can use face capture to let your face replace another person's face in the video in real time. In simple terms, you can transplant your facial expressions to a certain person in the video in real time. The same principle can also be used for 3D reconstruction of scenes in video and use for movie special effects [10]. Deep learning networks have been able to help animators a lot in estimating motion. We can perform real-time estimation today. Dr. Zhe Cao in Cornell University had taught a neural network to estimate changes in the position of the human skeleton so that it will estimate the next move you might take through this neural network. Use computers to automatically sort photos is common in our real life. For example, Facebook can tag your friends in shared photos, and Google can tag your photos for more efficient searches. There is a deep learning application can describe various elements in photos. Dr.Andrej Karpathy trained a deep learning system that can recognize elements in different regions in a photo and describe this deep learning system photo in one sentence which users can generate a new photo from none. The Oxford Geometry Group developed an application using deep learning technology. It can read the text in the video. By entering the text, the application can directly show the BBC news videos that contain these texts in the image. Google Sunroof can estimated how much solar energy can be collected on the roof of the panel by creating a 3D model of the roof of your home, and based on the aerial map of Google Earth, use deep learning to distinguish the roof from the surrounding number. Then, based on the sun's trajectory and weather conditions, the result will be easy to calculate. Scientists from Oxford University and Deep-mind have jointly completed the project called Lip-Net, and its lip-reading accuracy reached 93%, which far exceeding the average level of human lip readers' 52% accuracy.

In conclusion, deep learning proposes a method for computers to automatically learn pattern features and integrates this feature learning into the process of building different models, thereby reducing incompleteness that caused by human artificial design. At present, some machine learning applications that has deep learning technology as the core have achieved better recognition and classification performance which beyond any existing algorithms in some

specific conditions. Furthermore, the neural network has a strong fitting ability and can do very complicated non-linear mapping. There are many parameters in a deep neural network, so it has a strong representation ability. Most modern deep neural networks can be used to extract the abstract features like images, speech, and text, and the extracted features have stronger generalization performance than human set specifics. It can be applied to multiple fields. No matter learning shallow or deep semantic information, it can always provide features for the tasks. From the earlier Alex-net to the later vgg16, vgg19, google-net, and res-net, they all focus on reducing the number of parameters, adjust richer features, and speed up the training speed. In this case, neural networks have great flexibility in which can adapt to multiple different tasks.

There are still some disadvantages of deep learning. In application scenarios that can only provide a limited amount of data, deep learning algorithms cannot make unbiased estimates of the law of the data. In order to achieve good accuracy, big data support is needed. Due to the complexity of the graph model in deep learning, the time complexity of the algorithm has increased dramatically. In order to ensure the real-time nature of the algorithm, higher parallel programming skills and better hardware support are required. Therefore, only some scientific research institutions that have relatively strong economic background can use deep learning as the main tool to develop cutting-edge applications. Furthermore, neural networks are easy to fall into a local optimum solution, which leads to the poor generalization ability of the model, and making the data distribution become over-fitting, and not perform good in predicting the unknown data. The amount of neural network parameters is huge, so it is very slow to train, and it is very inconvenient for storage. Therefore, based on this problem, the network structure is constantly changed, the large network is trained first, and then migrating to a small network structure, which greatly reduces the parameter amount, but also the accuracy will be significantly reduced. In the case of very deep networks, the gradient will disappear, so the command such as rectified linear units [-relu], and [-tanh] should be prepared to prevent the disappear of gradient.

Overall, the reconstruction potential of deep learning has been proven in cognitive problems, and it has even become a potential algorithm basis in cloud to work as the decision support. However, if these reconstruction algorithms have a very different from the real environment, then the actual application is likely to bring huge risks, and especially considering the application of deep learning in the fields of face recognition, which means someone can steal your identity by capture your facial characteristic. Although it is impossible to prevent deep learning from continuously integrating into our lives, we can further improve its transparency, that is, understand how these algorithms make their own judgments. We should examine the identification process of specific algorithms in deep learning applications (for example, from source information to end-to-end graphic transformations, statistical models, and even metadata, etc.), and then learn how to take a better action in different specific situations.

Reference

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017. [Accessed: 28-Oct-2019].

[2] A. Dertat, "Applied Deep Learning - Part 3: Autoencoders," Medium, 08-Oct-2017. [Online]. Available: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798. [Accessed: 28-Oct-2019].

[3] "Deep Belief Networks," Encyclopedia of Machine Learning and Data Mining, pp. 338–338, 2017. [Accessed: 28-Oct-2019].

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature News, 27-May-2015. [Online]. Available: https://www.nature.com/articles/nature14539. [Accessed: 30-Oct-2019].

[5] Graves, A.-rahman Mohamed, and G. Hinton, Speech recognition with deep recurrent neural networks - IEEE Conference Publication. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6638947. [Accessed: 05-Nov-2019].

[6] Bruce and A. Young, "Understanding face recognition," British Journal of Psychology, vol. 77, no. 3, pp. 305–327, 2014. [Accessed: 05-Nov-2019].

[7] M. Daily and S. Medasani, Self-Driving Cars - IEEE Journals & Magazine. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8220479. [Accessed: 05-Nov-2019].

[8] Jonathan, Shelhamer, Evan, Darrell, and Trevor, "Fully Convolutional Networks for Semantic Segmentation," Page Redirection, 01-Jan-1970. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html. [Accessed: 10-Nov-2019].

[9] Samek, Wojciech, Wiegand, Thomas, and Klaus-Robert, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models," arXiv.org, 28-Aug-2017. [Online]. Available: https://arxiv.org/abs/1708.08296. [Accessed: 01-Dec-2019].

[10] "Detecting Both Machine and Human Created Fake Face Images In the Wild," ACM Digital Library. [Online]. Available: https://dl.acm.org/citation.cfm?id=3267367. [Accessed: 01-Dec-2019].

[11] "A Modern Approach," Computer Vision. [Online]. Available: https://dl.acm.org/citation.cfm?id=580035. [Accessed: 10-Dec-2019].