Deep Learning Applications in the Medical Image Recognition

Song Yukun

High School of Nankai University, Tianjin, China

Email address:
songyukun2002@126.com

To cite this article:
Song Yukun. Deep Learning Applications in the Medical Image Recognition. *American Journal of Computer Science and Technology.* Vol. 2, No. 2, 2019, pp. 22-26. doi: 10.11648/j.ajcst.20190202.11

Abstract: In this essay, the researcher is focusing on the deep learning systems and its major applications in various fields. Song Yukun uses the relu incentive algorithm and the convolution functions to make the program automatically recognize different things or same type of things with different features. Before actually processing the image recognition part, the researcher adds a transforming program which change all kinds of image into one small form. Then, using this modelled image, the program could delicately determine the type of the contents in the image. This technological program is automatic and performs as an essential part of artificial intelligences. The main work it does is imitating the learning process of human brain, which accumulate experiences from thousands of events. It realizes this function by adding different algorithms in the program including the relu incentive algorithm which "teaches" the program particular types of images. After massive input, this technological program could quickly solve current problems with the lack of human labor force doing repetitive but intelligent works like checking particular tumor in the X-ray films. Besides, learning by themselves, the programs could generate results more specific than humans do. This deep-learning principle could be widely utilized since everything in human lives are learning and accumulating experiences. It could change any previous mechanical program into "intelligent" programs which would have an acceleration in their delicacy of determination.

Keywords: Deep Learning, Convolutional Neural Network (CNN), Image Recognition

1. Introduction

During the 1940s, our predecessors invented the first computer in the world, which took up three rooms and was designed to have over ten thousand conductors in order to calculate the projectile of the missile for military use. It was composed of five main parts: the logic components where the computer deal with all the data it holds, chief-controller, input device, output device, as well as the data recoding document. Through the development during these decades, our computers have evolved from the huge machine into the portable laptops, which brings us convenience and facilitates our daily lives.

On the other hand, the software in the computer has also experienced enormous change that people could do things unimaginable in the previous century, for example, individuals could contact face to face with each other using Facebook invented by Zuckburg even if they are actually thousands of miles away from each other. People could also use their phones to read the QR code. And now, human societies are facing an increasing demand of artificial intelligence. As John McCarthy, a professor of computer science at the Stanford university, defines, artificial intelligence is the science of making intelligent machines, especially using computers to understand or act like humans do. This new modern technology is necessary in most of the careers in the current society. Companies need quicker and more comprehensive ways to understand their customers. Face-scanning and fingerprint recognizing systems are employed in the security departments where engineers invented delicate face-scanning locks and safe face-scanning channels for payment as well as departure site in the airport. Virtual fitting rooms are set up in high-tech dress stores where people could change their appearance in the mirror by only touching some icons. Voice detecting and categorizing applications are implemented in the musical applications to make it easier for people to found out the song they hear in their daily lives in a few seconds. Furthermore, there are even some olfactory devices in the

recently invented intelligent phone called Ali-phone for identifying products like the tea. What beyond our imagination is that those widespread different technologies all derive from the same program – deep learning network, and that's far less than its limit.

Deep learning could be also utilized to do something more influential and helpful to the society. There are millions of disabled people and deeply suffering patients who got diseases that expired their ability of moving. The deep-learning program could help them. The intelligent sticks that help those blind people are perfect examples in which the deep learning systems could play important roles to detect and avoid possible dangers like moving cars, stairs, and walls. Deep learning programs could also help detect whether the patient is opening their eyes or not, through which the nurses can let patients make choices by winkling their eyes. With the help of deep learning artificial intelligence, there could be much simpler tasks for people to deal with instead of those complicated and dangerous tasks. But how?

2. Convolutional Neural Network Approach

Starting from the simplest geometrical shapes could be more acceptable for explanation since the actual recognition of the human eyes are too complicated while they actually share the intrinsic essence. Imagine the most common figurative shapes on public bathrooms: male and female. How could human distinguish them during our life and never make mistakes? Dresses stands for women, so all the places take the dress as the most distinguishable sign on their doors for female bathrooms. Dress, in abstract, represents an important factor and tool for deep learning – particular characteristics. People see those shapes of trapezoids, so do computers. Mathematics now play an important role in recognizing the figures. As is widely acknowledged, the figures are combined by pixels and each single pixel have red, green, and blue (RGB) to alternate the color in that pixel. Actually, all colors are stored in computers as numbers from 000000 to FFFFFF, corresponding to the three original colors.
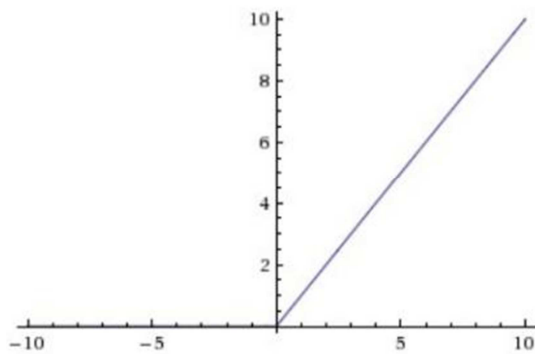

Figure 1. Relu activation function.

Therefore, the first thing is to decode the graph received from the cameras, which could be accomplished by adding

decoding lines in our programs thus changing the initial image inputs into readable four dimensional matrixes, which contains four parameters corresponding to batches, rows, columns, and layers. Then, the second question comes, how could computers "learn" from those data? The researcher uses incentive relu functions for solving this problem. The relu functions are lines that reserve useful data and eliminate the irrelevant lines of numbers.

As the graph shows, the relu function actually excludes all the negative values regarding irrelevant data in our programs. [1-4] This function make it much easier for computers to conclude and induce the internal unique characteristics of the product. Then, based on this relu incentive, the computers could do regularization to the matrix of weight it generated before. This process is mimicking the function of our brain composed of billions of neurons and it is just making certain pat of the "brain" active. [5-10] Apart from that, convolution to achieve deep learning is still needed. First, the computer may randomly generate a matrix for the following convolution. During each round of process, the computer multiplies every number recorded in one pixel with the corresponding element in the random matrix and get a new matrix of sum. Some of the pictures may be too large or small, so the program contains normalizing functions and padding elements in the program to avoid blind area in the procedure of convolution. In addition, the picture could be huge if it is calculated in the number of pixels, for which the program should contain pooling system to combine multiple numbers together. [11-18] The "pooling" function is:

$$b_{x,y}^i = a_{x,y}^i / \left(k + \alpha \sum_{j=\max\left(0,i-\frac{n}{2}\right)}^{min(N-1,i+n/2)} \left(a_{x,y}^j\right)^2 \right)^\beta$$

After the computer get the matrix full of sums, it could use the relu incentive to manage the weight matrix while the results are too small or two huge. Then repeat the whole procedure again and again until the weight matrix could measure one feature of the graph.
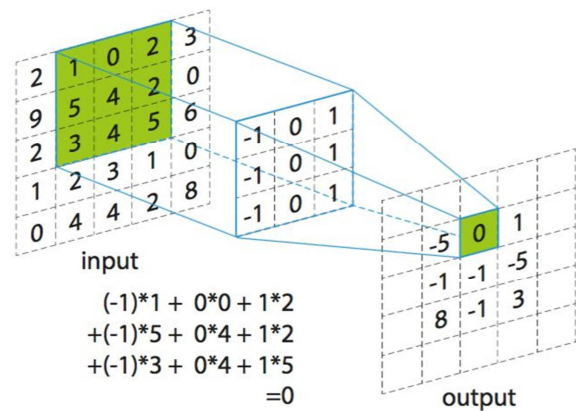

(-1)*1 + 0*0 + 1*2
+(-1)*5 + 0*4 + 1*2
+(-1)*3 + 0*4 + 1*5
=0

Figure 2. Convolution calculations between matrix.

Figure 3 describes the procedure of convolution. In order to cover each single pixel, the program could either make the

stride of multiplying be one, which is very ineffective because the picture is actually huge for the weight matrix to crumble on it, or make the picture have some more pixels so that the length and width could be divided by the weight matrix without remainders, called padding. Normally programmers add 0 or the marginal value of the input image on its margin so that it doesn't affect the accuracy of convolution.

```
data/0_84.jpg      real:Plane => predicted:Plane
data/0_85.jpg      real:Plane => predicted:Plane
data/0_89.jpg      real:Plane => predicted:Plane
data/0_9.jpg       real:Plane => predicted:Plane
data/0_91.jpg      real:Plane => predicted:Plane
data/0_97.jpg      real:Plane => predicted:Plane
data/1_0.jpg       real:Car => predicted:Car
data/1_1.jpg       real:Car => predicted:Car
data/1_101.jpg     real:Car => predicted:Car
data/1_102.jpg     real:Car => predicted:Car
data/1_103.jpg     real:Car => predicted:Car
data/1_12.jpg      real:Car => predicted:Car
data/1_13.jpg      real:Car => predicted:Car
data/1_14.jpg      real:Car => predicted:Car
```

Figure 3. Convolutional neural network to predict the figures.

But the images don't only have one characteristic. It's necessary for program to generate hundreds of weight matrixes and do convolution to each layer of the graph, which is divided by colors. Due to the fact that the characteristics are relatively randomly distributed on the range of number from 00 to FF, when the program generates two random matrixes, it's possible that they get totally different results by convolution only because they are at different position on the number axis. Therefore, the reasonable number of matrixes generated is required in order to cover each characteristic. Besides, for obtaining more possible characteristics, the program could rotate the picture to different angle to mimic any possible visual angle in the reality. First of all, the original two-dimensional array (or matrix) is [[0, 1, 2, 3], [4, 5, 6, 7], [8, 9, 10, 11], [12, 13, 14, 15]]. Then, after reversing the picture, it becomes [[0, 4, 8, 12], [1, 5, 9, 13], [2, 6, 10, 14], [3, 7, 11, 15]], which enables the convolution to cover more characteristics from different angle.

After all, the computer would get a lot of valuable weight matrix after the training using relu incentive function, and now it need to "recognize" the new picture. When the computer get the new picture, it still transforms it into arrays, then, it multiplies the array to each weight matrix to get hundreds of sums. Using the determination function based on the variance of the whole sample of data, the credibility of each sum could be get, thus get the credible interval of each single characteristic. As a result, if the program could generalize the image into some certain characteristic of, for example, the mouse, that the computer "learned" before, it could output a percentage standing for our confidence of regarding the input image as a mouse, or it could also give a percentage standing for the confidence of not regarding the input image as a mouse. The credibility of the confidence could also be learned through the process of deep learning and comparing to the

standard deviation of the whole population's credible interval. Thus, the program could conclude out criteria by which it could more specifically determine whether the image is one of the target images that the computer "learned" with titles before. After being "trained" by sources of different types of pictures, taking planes, cars, and birds for examples, the deep-learning program could precisely predict the upcoming pictures as previous types of images.

3. Applications in Image Recognition

This spectacular technology could be utilized in multiple fields [19-24]. For example, in the car-number detecting system on the road, the cameras could deliver the images to the central controlling computer so that it could utilize programs which are previously generalized by millions of car-number images. Then, the computer could know whether the car has a number plate and whether the car owner blocks the plate. With the support of great database, the program could match the car owner with the illegal behavior towards the car-number plate and finally punish him or her. As Fujiyoshi Hironobu *et al.* write, "Their features are a labeling of the smoothed image using Gaussian filter, template matching using averaged patterns and some artificial neural networks (NNs) based on the Back-Propagation algorithm. First, in the case of Gaussian filtering method, the smoothed image is used so that the separation of car body and number plate is not sufficient in the making labels. Therefore, a body color of the car influences to the detection score. The template matching is given the high score because of the edge emphasizing. It does not depend on a body color. Finally, we show that the NNs are able to detect with much higher score." Based on this real implementation, the program could evolve and become able to directly connect the pedestrians, riders, and drivers who violate the transportation law and automatically warn and punish them on the personal trust network which is under constructing, which could save all the police who have to stand in the middle of the road every day in metropolitan. In the future, maybe the self-driving car implementing this program to distinguish pedestrians and other cars could also get into peoples' lives. Besides, this technology could also be applied in the medical field. The programs could figure out some kinds of hidden figure that human could not notice based on our natural sensing, which may make the programs play unreplaceable important status in diagnosing diseases. For instance, eye is one of the most delicate apparatus in human body and there's hundreds of different kinds of disease that cause trouble to human eyes, which are difficult for doctors to recognize since the structure of eyes are too complex. However, utilizing the deep-learning program, "It has a higher accuracy rate on finding indications of diseases on retinal images than slower conventional diagnoses by doctors. As such, it has the potential to make preventative healthcare available to millions of people, not just in China, but around the world." The extreme high speed of diagnosing could also bring incredible efficiency to the hospitals, which helps save millions of people's eyesight or

even lives. Using the Airdoc machine with intelligent learning systems, "The machine has just taken high-resolution medical-grade images of both your retinas. It instantly sends them to the cloud where it takes 20 to 30 milliseconds." Furthermore, there are millions of patients suffering from serious diseases and must stay in bed, some of whom even could only move their eyes, this deep learning program could also be trained with types of human with eyes opening and closing which may make those patients being able to express their own ideas by choosing the choices on a large screen in front of them with blinking different times.

Based on this program, the nurses and doctors don't need to stuck at the ICU and looking every single movement of the patients while doing more precise detections through the automatic programs. [25-31] As a result, this program could massively bring benefits and convenience to multiple fields.

Codes could be finished based on convolutional neural network to recognize the pictures of human face. At first step, the nurses could train the codes with hundreds of pictures of patients' faces. These pictures can be classified to a few types, therefore each picture is labelled as one specific name. After training the codes with these pictures and the corresponding labels, our convolutional neural network can capture the features of each kind of pictures. The figure below present the learning rate accuracy of the codes. After training many steps, the codes can fit their feature matrix and ready for predictions. In the hospital, these codes can monitor people's face and "know" if this kind of face is painful or not. In the below figure, with more training steps, the mean loss reflects the discrepancy between code predictions and the real data. It decreases with steps, and the code prediction accuracy increases with steps, which means the codes can learn the training pictures.
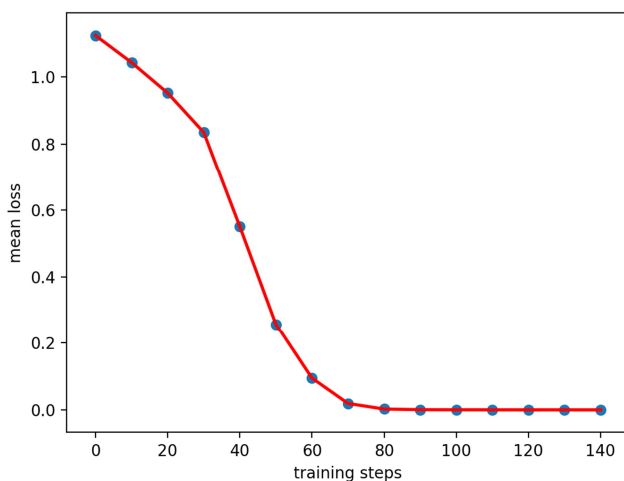

Figure 4. *Loss function calculated with the codes decreases with the training steps, which reflects the improvement of its prediction accuracy.*

4. Conclusion

In conclusion, those inherent written lines in the programs constructed more and more intelligent system with high speed and accuracy of dealing with information which is beyond human's reach and imagination. Based on these profound programs, human civilization could make a great leap and improve in every corner in the society. Artificial intelligence with data regression and distinguishing systems, which is more and more popular and developed in the status quo, is actually changing our lives, both in front of us and under the surface.

References

[1]   Abien Fred M. Agarap, arXiv: 1803.08375.

[2]   Hidenori Ide; Takio Kurita, Proceeding of 2017 International Joint Conference on Neural Networks (IJCNN).

[3]   Huan Zhang, Tsui-Wei Weng, Pin-Yu Chen, Cho-Jui Hsieh, Luca Daniel, Proceeding of Thirty-second Conference on Neural Information Processing Systems.

[4]   Siamak Ravanbakhsh, Barnabas Poczos, Jeff Schneider, Dale Schuurmans, Russell Greiner; Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, PMLR 51: 809-818, 2016.

[5]   Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, Proceedings of the 32 nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37.

[6]   Vidushi Sharma, Sachin Rai, Anurag Dev, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10.

[7]   Juergen Schmidhuber, Neural Networks, Vol 61, pp 85-117, Jan 2015.

[8]   Waseem Rawat, Zenghui Wang, Neural Computation 29, 2352–2449 (2017).

[9]   Saad Albawi; Tareq Abed Mohammed, Saad Al-Zawi, proceeding of 2017 International Conference on Engineering and Technology (ICET).

[10] Wenling Shang, Kihyuk Sohn, Diogo Almeida, Honglak Lee, Proceedings of the 33 rd International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48.

[11] Hayoung Eom, Heeyoul Choi, proceeding of ICASSP 2019.

[12] Jawad Nagi, et al, proceeding of 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA).

[13] Manli Su, Zhanjie Song, et al, Neurocomputing, Volume 224, 8 February 2017, Pages 96-104.

[14] Qi Zhao, Shuchang Lyu, et al, Wireless Communications and Mobile Computing, Volume 2018, Article ID 8196906, 15 pages.

[15] Chen-Yu Lee, Patrick W. Gallagher, Zhuowen Tu, Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS) 2016, Cadiz, Spain. JMLR: W&CP volume 51.

[16] Dingjun Yu, Hanli Wang, Peiqiu Chen, Zhihua Wei, Proceeding of International Conference on Rough Sets and Knowledge Technology.

[17] Jawad Nagi, Frederick Ducatelle, et al, Proceeding of 2011 IEEE International Conference on Signal and Image Processing Applications.

[18] Li Deng, Ossama Abdel-Hamid, Dong Yu, IEEE International Conference on Acoustics, Speech, and Signal Processing, May 2013.

[19] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 22, NO. 10, OCTOBER 2014.

[20] Yoon Kim, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751.

[21] Sameer Khan, Suet-Peng Yong, Proceedings of APSIPA Annual Summit and Conference 2017.

[22] Ruiting Shao, Yang Lei, Jian Fan, Jerry Liu, Proceeding of IS&T International Symposium on Electronic Imaging 2018.

[23] Mathias Niepert, Mohamed Ahmed, Konstantin Kutzkov, Proceedings of the 33 rd International Conference on Machine Learning, New York, NY, USA, 2016.

[24] Eduardo Carabez, Miho Sugi, Isao Nambu, Yasuhiro Wada, Computational Intelligence and Neuroscience, Volume 2017, Article ID 8163949.

[25] Nima Tajbakhsh; Jae Y. Shin, proceeding of IEEE Transactions on Medical Imaging (Volume: 35, Issue: 5, May 2016).

[26] Dinggang Shen, Guorong Wu, Heung-Il Suk, Annual Review of Biomedical Engineering, Vol. 19: 221-248 (Volume publication date June 2017).

[27] Onur Ozdemir, Benjamin Woodward, Andrew A. Berlin, proceeding of Second workshop on Bayesian Deep Learning (NIPS 2017), Long Beach, CA, USA.

[28] Nikou Gunnemann, Jurgen Pfeffer, Proceedings of Machine Learning Research 74: 92–102, 2017.

[29] Holger R. Roth, Le Lu, Jiamin Liu, Jianhua Yao, Ari Seff, Kevin Cherry, Lauren Kim, and Ronald M. Summers, Proceeding of IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 35, NO. 5, MAY 2016.

[30] Edward Choi, Andy Schuetz, Walter F Stewart, Jimeng Sun, Journal of the American Medical Informatics Association, Volume 24, Issue 2, March 2017, Pages 361–370, https://doi.org/10.1093/jamia/ocw112.

[31] Le Lu, Yefeng Zheng, Gustavo Carneiro, Lin Yang, book < *Deep Learning and Convolutional Neural Networks for Medical Image Computing* >.