

DEEPMIND CREATION PVT LTD

**CONVOLUTIONAL NEURAL NETWORK THEORY
INTERNSHIP
2021**



Submitted by

**Hem Bahadur Gurung
AI internship**

Submitted to

**Aayush Adhikari
Cofounder**

Submitted Date: June 1th,2021

Table of Contents

Introduction.....	3
CNN architecture.....	3
Conclusion	9
Drawback of Convolutional Neural Network (CNN).....	9
Reference	10

Convolutional Neural Network (CNN) Theory

Introduction

A convolutional neural network (CNN) is class of artificial neural network (ANN) used in computer vision and processing which is specifically designed to process pixel data. Convolutional neural network (ConvNets) were first introduced in the 1980s by Yann LeCun, computer science researcher named as LeNet-5 to recognize hand-written numbers. CNN structure was first proposed by Fukushima in 1988. Alex Net, VGG, Google Net, Dense CNN and Fractal Net are generally considered the most popular architectures because of their performance in object recognition. (Alom, et al., 2019)

CNN is particularly used in image & video recognition, image analysis & classification, segmentation, media recreation, recommendation systems, brain-computer interfaces, natural language processing and financial time series. (Bhandare, et al., 2016)

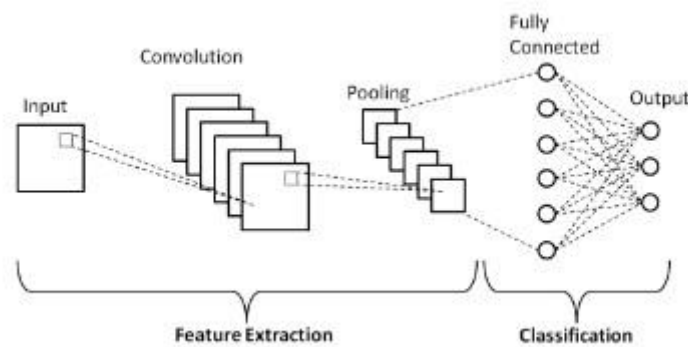


Fig 1: Schematic diagram of a basic CNN architecture (Source (Phung & Rhee, 2019))

CNN architecture

CNN architecture consists two blocks feature extraction and classification. Feature extraction is also called convolution block which contains repetitions of a stack of convolution layers and Pooling layer. Similarly, classification block is also known as fully connected block where simple neural networks are connected to each other. (Yamashita, et al., 2018).The process in which input data are converted into output through these layers is called forward propagation as shown in figure.

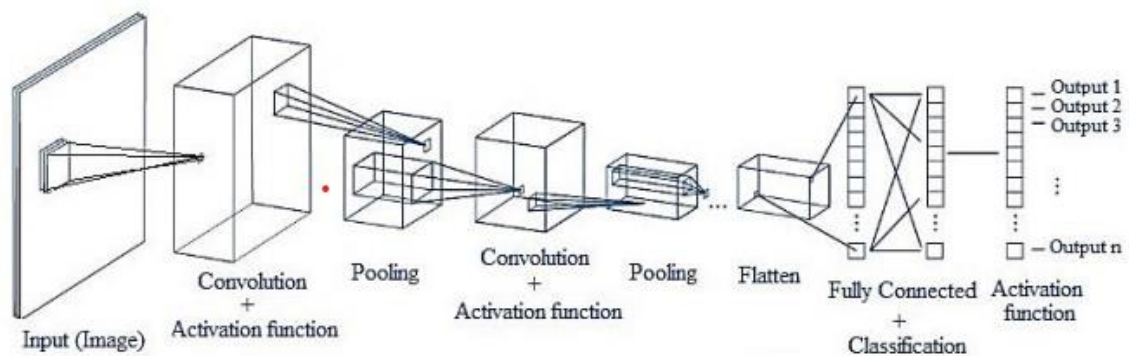


Fig 2: Convolution neural networks (CNN) architecture in image classification

Source: Hacibeyoglu (2018)

A. Features Extraction

i. Convolutional Layer

A convolutional neural network is core part of CNN architecture. responsible for feature extraction performs linear and nonlinear operations such as convolution operation and activation function. The aim of Convolutional layer is to learn feature representations of the input. (Guo, et al., 2017) Convolutional layer contains set of convolutional kernels or filters which gets convolved with the input image (N-dimensional metrics) to generate an output feature map. Filter is set of weights in a matrix applied on an image or a matrix to obtain the required features.

Let's considered a 5 x 5 whose image pixel values are 0, 1 and filter matrix 3 x 3 as shown in below.

Input Feature Map					Convolutional Filter		
3	5	2	8	1	1	0	0
9	7	5	4	3	1	1	0
2	0	6	1	6	0	0	1
6	3	7	9	2			
1	4	9	5	1			

Fig 3: 5x5 Image matrix (feature map) multiplies kernel or filter matrix 3x3(Source: (Developers , 2021)

Then the convolution of 5 x 5 image matrix multiplies with 3 x 3 filter matrix which is called "Feature Map" as output shown in below.

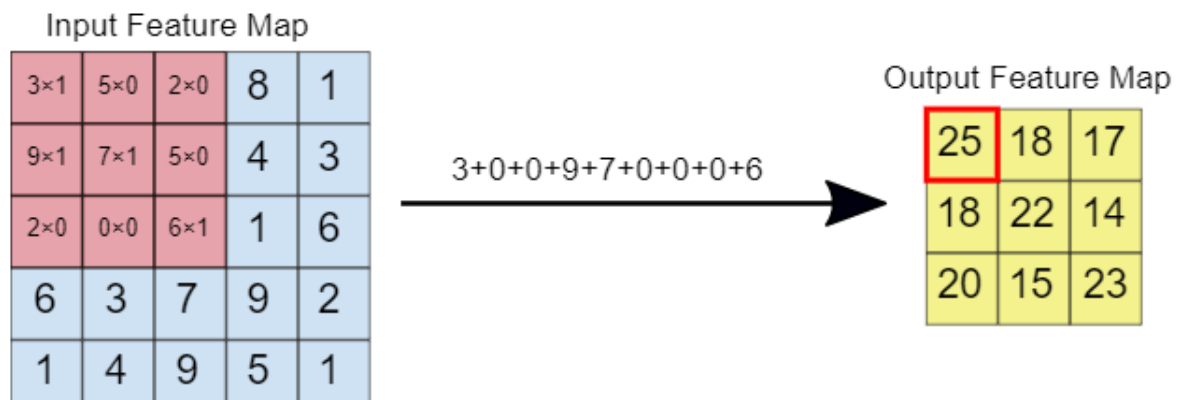


Figure 4: 3 x 3 Output matrix Source: Developers (2021)

ii. Non-Linearity (ReLU)

ReLU has been used after every Convolution operation.

The output is $f(x) = \max(0, x)$.

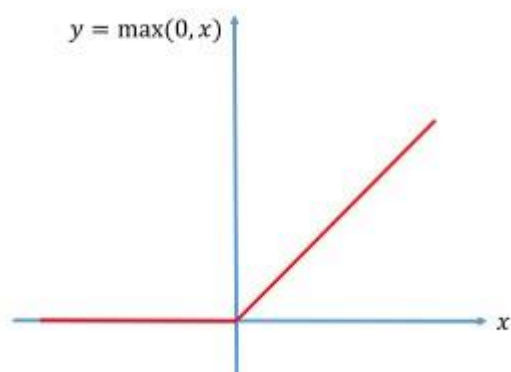


Fig 5: ReLU function on Graph. Source: Wu(2017)

ReLU's purpose is to introduce non-linearity in our ConvNet. Since, the real-world data would want our ConvNet to learn would be non-negative linear values. There are other nonlinear functions such as tanh or sigmoid that can also be used instead of ReLU.

On the following image we apply the Relu operation and replaces all the negative numbers by 0.

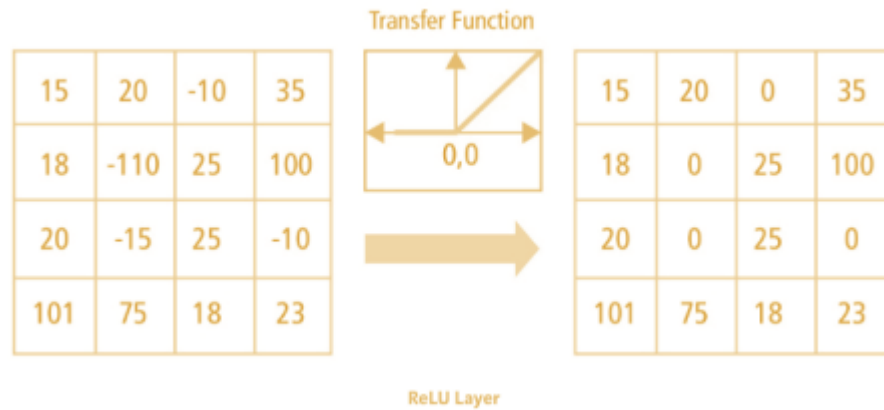


Fig 6: ReLU Operation Source: Hijazi, et al.(2015) & Prabhu (2018)

iii. Pooling Layer:

Pooling layer is second layer of CNN architecture. The pooling operation consists in reducing the size of the images while preserving their important characteristics. It summarizes value overall the value present. The purpose of the pooling layers is to reduce the spatial resolution of the feature maps and thus achieve spatial invariance to input distortions and translation.

There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. Max Pooling also performs as a Noise Suppressant by discards the noisy activations altogether. Meanwhile; Average Pooling returns the average of all the values from the portion of the image as a noise suppressing mechanism covered by the Kernel as shown in figure. Thus, Max Pooling performs a lot better than Average Pooling. There are two important parameters in the pooling layer, one of which is the filter size and the second is the stride parameter. (Saha, 2018)

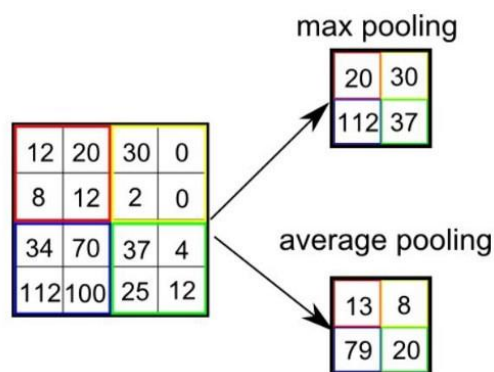


Figure 7: Comparison between Max and average pooling. Source: (Rawat & Wang, 2017)

Striding

We use striding in such case when we do not want to capture all the information. So, we skip some neighboring cells. For example-

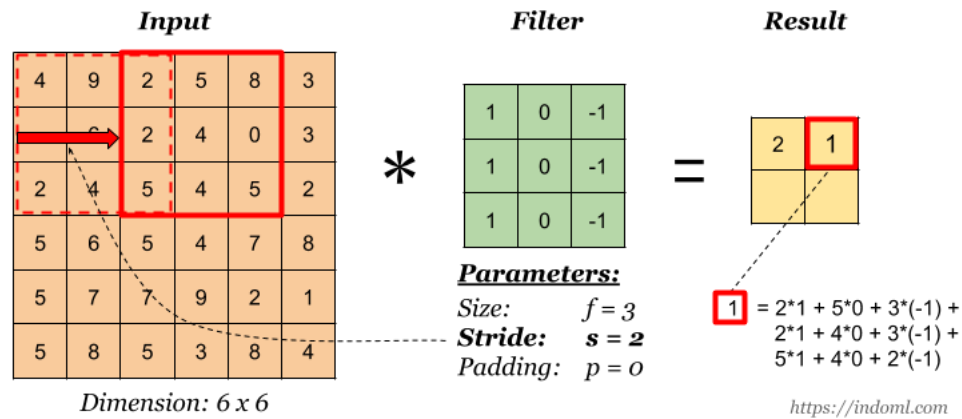


Fig 7: Result of output image after striding

Padding

While applying convolutions, sometime we willnot get output equivalent to input because of loss of data over borders. So that, we append a border of zeros amd recalculate the convolutions covering all the input values as shown in below.

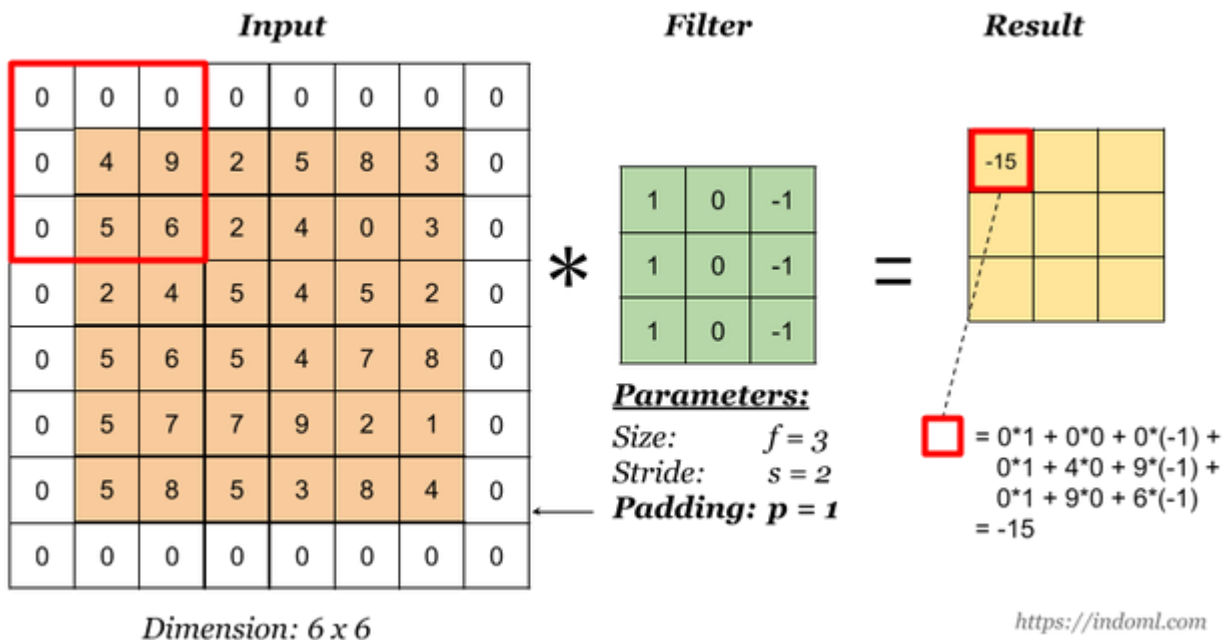


Fig 8: Image matrix after padding and striding

Classification Block

iv. Fully connected Layer

Fully-connected layers are the last part of every CNN architecture where each neuron inside a layer is connected with each neuron from its previous layer also known as dense layers. The last layer of Fully-Connected layers is used as the classifier of output image of the CNN architecture. The final fully connected layer typically has the same number of output nodes as the number of classes. Fully-Connected Layers are feed-forward artificial neural network (ANN) works under the principle of traditional multi-layer perceptron neural network (MLP).

The FC layers take input as feature maps from the final convolutional or pooling layer in the form of a set of metrics and those metrics are flattened to create a vector and this vector is then pass into the FC layer to generate the final output of CNN. (Ghosh, et al., 2020). Flattening is the function that converts the pooled feature map to a single column that is passed to the fully connected layer. Finally, we use an activation function such as SoftMax or sigmoid to classify the outputs.

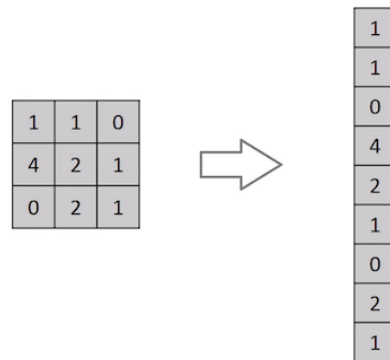


Fig 9: Flattening of a 3x3 image matrix into a 9x1 vector

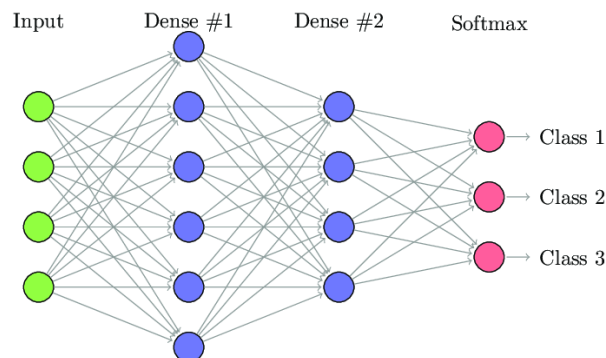


Fig 10: Example of fully-connected neural network (Source: (Pelletier, et al., 2019))

Conclusion

In conclusion, Convolutional Neural Network (CNN) consists convolution layer and fully connected layer. The convolution layer always contains two basic functions, namely convolution and pooling. The convolution operation using multiple filters generate feature map by extracting features from the data set without losing important features. The pooling operation, also called subsampling, is used to reduce the dimensionality of feature maps from the convolution operation. These features are passed to the fully connected layer which consists of activation function and classify the image. The detail diagram of CNN is listed following as shown in figure.

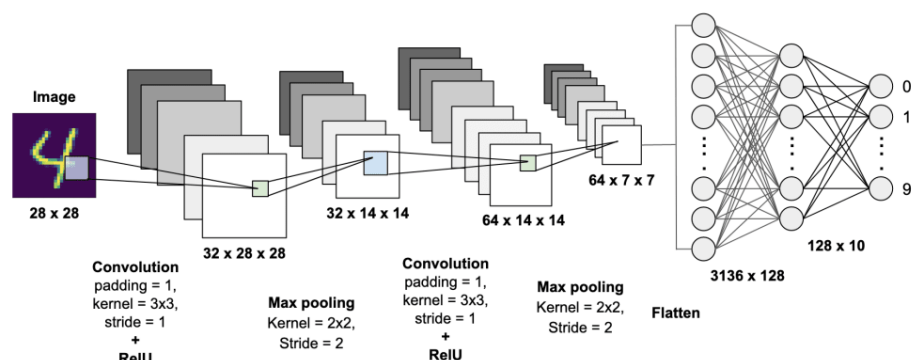


Fig11: Convolutional Neural Networks (CNN) architectures presentation

(Source: <https://dev.to/afrozchakure/cnn-in-a-brief-27gg>)

Drawback of Convolutional Neural Network (CNN)

CNN has many limitations on its applications. It makes prediction and classifies based on certain components are present in that image. One of them is its inability to encode the position and orientation of object. Let's consider our face where have the face oval, two eyes, a nose and a mouth. It's enough for CNN in having presence of these components can be a very strong indicator to consider that there is a face in the image. Orientational and relative spatial relationships between these components were not considered for CNN. (Pechyonkin, 2017)The example is given below.

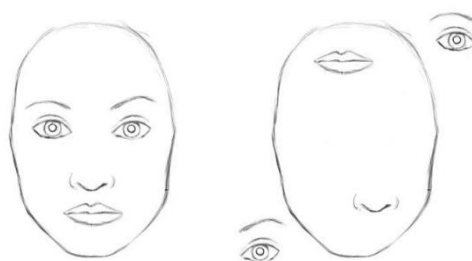


Figure 12: Both pictures are similar to CNN, since they both contain similar components. Source: Pechyonkin(2017)

Reference

- Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network, in: 2017 International Conference on Engineering and Technology (ICET). Presented at the 2017 International Conference on Engineering and Technology (ICET), IEEE, Antalya, pp. 1–6. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A.S., Asari, V.K., 2019. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* 8, 292. <https://doi.org/10.3390/electronics8030292>
- Bhandare, A., Bhide, M., Gokhale, P., Chandavarkar, R., 2016. Applications of Convolutional Neural Networks 7, 10.
- Dertat, A., 2017. Applied Deep Learning - Part 1: Artificial Neural Networks [WWW Document]. Medium. URL <https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6> (accessed 5.30.21).
- Deshpande, A., n.d. A Beginner's Guide To Understanding Convolutional Neural Networks [WWW Document]. URL <https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/> (accessed 5.30.21).
- Drawbacks of Convolutional Neural Networks [WWW Document], n.d. URL <https://www.linkedin.com/pulse/drawbacks-convolutional-neural-networks-sakhawat-h-sumit> (accessed 6.1.21).
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., De, D., 2020. Fundamental Concepts of Convolutional Neural Network. pp. 519–567. https://doi.org/10.1007/978-3-030-32644-9_36
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G., Cai, J., Chen, T., 2017. Recent Advances in Convolutional Neural Networks. arXiv:1512.07108 [cs].
- Guo, T., Dong, J., Li, H., Gao, Y., 2017. Simple convolutional neural network on image classification, in: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(. Presented at the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE, Beijing, China, pp. 721–724. <https://doi.org/10.1109/ICBDA.2017.8078730>
- Hacibeyoglu, M., 2018. Human Gender Prediction on Facial Mobil Images using Convolutional Neural Networks. *International Journal of Intelligent Systems and Applications in Engineering* 3, 203–208. <https://doi.org/10.18201/ijisae.2018644778>
- Hijazi, S., Kumar, R., Rowen, C., n.d. Using Convolutional Neural Networks for Image Recognition 12.
- Hinton, G., Sabour, S., Frosst, N., 2018. MATRIX CAPSULES WITH EM ROUTING 15.
- Indolia, S., Goswami, A.K., Mishra, S.P., Asopa, P., 2018. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science* 132, 679–688. <https://doi.org/10.1016/j.procs.2018.05.069>

Khan, A., Sohail, A., Zahoora, U., Saeed, A., 2020. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. Artificial Intelligence Review 53. <https://doi.org/10.1007/s10462-020-09825-6>

ML Practicum: Image Classification. Google Developers. URL <https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks> (accessed 6.1.21).

O'Shea, K., Nash, R., n.d. An Introduction to Convolutional Neural Networks 12.

Pandya, M.D., Shah, P.D., Jardosh, S., 2019. Medical image diagnosis for disease detection: A deep learning approach, in: U-Healthcare Monitoring Systems. Elsevier, pp. 37–60. <https://doi.org/10.1016/B978-0-12-815370-3.00003-7>

Pechyonkin, M., 2018. Understanding Hinton's Capsule Networks. Part I: Intuition. [WWW Document]. Medium. URL <https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b> (accessed 6.1.21).

Pelletier, C., Webb, G., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. Remote Sensing 11, 523. <https://doi.org/10.3390/rs11050523>

Phung, Rhee, 2019. A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences 9, 4500. <https://doi.org/10.3390/app9214500>

Prabhu, 2019. Understanding of Convolutional Neural Network (CNN) — Deep Learning [WWW Document]. Medium. URL <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148> (accessed 5.30.21).

Rawat, W., Wang, Z., 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation 29, 2352–2449. https://doi.org/10.1162/neco_a_00990

Sci-Hub: removing barriers in the way of science [WWW Document], n.d. URL <https://sci-hub.do/> (accessed 6.1.21).

Shamsaldin, A., Fattah, P., Rashid, T., Al-Salihi, N., 2019. The Study of The Convolutional Neural Networks Applications. UKH Journal of Science and Engineering 3, 31–40. <https://doi.org/10.25079/ukhjse.v3n2y2019.pp31-40>

Student Notes: Convolutional Neural Networks (CNN) Introduction, 2018. . Belajar Pembelajaran Mesin Indonesia. URL <https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/> (accessed 6.1.21).

Wu, J., Introduction to Convolutional Neural Networks 31.

Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629. <https://doi.org/10.1007/s13244-018-0639-9>