Application of Convolution Neural Network for Digital Image Processing

Venkata Naga Satya Surendra Chimakurthi

Solutions Architect, CDBDX-Platforms-DAM (Digital Asset Management), Cognizant Technology Solutions, Dallas, **USA**

*Corresponding Contact:
Email: chvnssurendra@gmail.com

ABSTRACT

In order to train neural network algorithms for multiple machine learning tasks, like the division of distinct categories of objects, various deep learning approaches employ data. Convolutional neural networks deep learning algorithms are quite strong when it comes to image processing. With the recent development of multi-layer convolutional neural networks for high-level tasks like object recognition, object acquisition, and recent semantic classification, the field has seen great success in this approach. The two-phase approach is frequently employed in semantic segregation. In the second step of becoming a standard global graphical model, communication networks are educated to deliver good local intelligence with a pixel. Convolutional Neural Networks (CNN or ConvNet) are complicated neural server networks in the field of artificial intelligence. Because of their great accuracy, convolutional neural networks (CNNs) are frequently utilized in picture categorization and recognition. In the late 1990s, Yann LeCun, a computer scientist, was based on the human notion of cognition and came up with the idea. When constructing a network, CNN uses a hierarchical model that eventually results in a convolution layer in which all neurons are linked and output is processed. Using an example of an image processing application, this article demonstrates how the CNN architecture is implemented in its entirety. You can utilize this to better comprehend the advantages of this current photography website.

Key words:

Image Processing, CNN, Digital Image Recognition, Machine Learning

INTRODUCTION

In present years, this modernized system has completely changed the perception of computers. Although they have been around since at least the 1990's, when they became famous they soon had the training of these large, complex networks into very large databases. Using modern computerized GPUs and comparable computer systems has dramatically improved the ability to train CNN models, leading to their development in

both academia and industry. A huge database like ImageNet has made it possible for us to train thousands of neural networks in a relatively short period of time. One of the oldest challenges in computer perception is the problem of picture segregation, in which we categorize an image into one of several focus classes. The proliferation of imaging applications has exacerbated an age-old issue: image segregation. The snapshot of the image is the most significant aspect of the past in automatic driving. Facebook and Google's photo-sharing apps also benefit from modern social media. Images employ picture categorization to improve and tailor the experience of their products for their customers. Many of the common issues in computer perception, such as internal phase variability, closure, flexibility, scale variability, visual variability, and brightness, can be found in the problem of image classification (Chimakurthi, 2017b). Effective image classification techniques may translate into methods that will improve other important computer recognition functions, such as detection, location processing, and segmentation (Wang et al., 2016).

REVIEW OF RELATED LITERATURE

Many high-level operations, such as image categorization, object acquisition, and more recently semantic separation, have proven that convolutional neural networks with numerous layers are useful. Semantic division, the two stages in the stage process, are commonly used. Using this method, networks of communication are taught to generate stunning local features that are cleverly depicted using pixels in step two of the standard world model. Researchers have found all the features of a section in a photograph and, for each example, mark their pixels (Chimakurthi, 2019a). They called this work the Simultaneous Detection and Segmentation (SDS). Unlike the discovery of the old binding box, the SDS needs to be separated not just the box. Unlike the classic semantic division, it requires the conditions of each object (Crounse & Chua, 1995). We describe a new design for the SDS based on recent work using convolutional neural networks to categorize independent regional suggestions (R-CNN). This is followed by an up and down revision of the figure-ground projections that they've already made. A 7-point (16 percent relative) rise in the SDS, a 5-point (10 percent relative) rise in semantic categorization, and state-of-the-art object acquisition performance are all shown in this study. (Liu et al., 1996).

An overall level of resistance is discovered, indicating how likely it is that an image window will include an object of any type. An object having a well-defined border in space (such as livestock or fencing) may be clearly differentiated from an amorphous background (such as grass or the road). According to Bayesia's framework, a few symbols are used to quantify the appearances of items, such as whether or not there is a closed border. Indicator of the closed boundary element is included in this. Using the tough PASCAL VOC 07 database, researchers found that this new indicator outperformed all previous indicators in terms of both the maximum rate and the combined resistance rate (Chimakurthi, 2019). Additionally, we contrast it to the drivers of interesting sites, the HOG detector, and three recent algorithms aimed at separating the default items (Williams, 2018). Finally, they offer two different means of dissent. A tiny proportion of windows were randomly selected for the first test, and an algorithm was devised to determine whether or not they should be used as test locations (Williams & Scott, 2019). This reduces the number of windows checked by the class-specific model, as demonstrated by experiments. When it comes to the second application, they use carelessness as an associated point above a certain model class, leading to some nonexistent good (Chua & Roska, 1993).

The ImageNet challenge is used to assess the performance of CNN's high-performance image classification models. Since establishing AlexNet in 2012, they've made structural and functional improvements that have allowed them to keep growing. For AlexNet, researchers launched GoogLeNet in 2014, which considerably reduced the amount of parameters. Experts also introduced VGGNet in 2014, which improved its performance due to the network's depth (Williams, 2020). ResNet, a system based on "jump connection" and batch normalisation, was launched by another team in 2015. (Wang et al., 2016).

In recent years, there has been a lot of attention paid to the issue of image captions. Using a combination of pre-trained CNN (VGGNet) as a feature detector, the Markov Random Field (MRF) alignment model, and RNN, the developers have been able to provide text descriptions during the past year. A short-term memory model (LSTM) presented by researchers integrates these feature vectors and vectors that embed the name in each sentence directly with the pre-trained VGGNet as a feature reader. Simply by adding the image vector to the first phase of LSTM, which they discovered to improve the results, it was developed in this manner Due to convolutional neural networks' powerful features, great progress has been made in the study of semantic segregation (CNNs) (Chimakurthi, 2019b). Semantic classification techniques based on the most up-to-date research use status information gleaned from the hidden image using CNN feature extraction. Restrictions on picture performance are imposed, and the quality of extracted features may be negatively affected. As an alternative to working on the raw image, they must use thousands of networks in a single image, which is time-consuming (Potluri et al., 2011).

WHAT CNN ACTUALLY IS?

A neural network consists of neurons that communicate with each other by exchanging signals. In order for a well-trained network to respond effectively when given with a visual image or pattern, connections have numeric weights that are modified during training. Layers and layers of neurons receive a feature from a network.
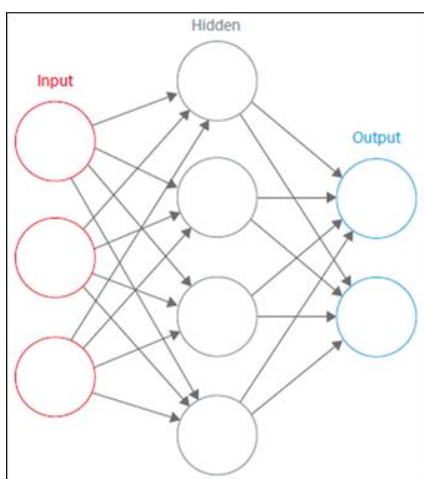

Figure 1: Model of Neural Network

Neurons in each layer respond to distinct combinations of inputs from the previous layer. Layers are created as depicted in Figure 1, with the first layer receiving a collection of ancient

patterns, the second layer obtaining patterns of patterns, and so on. For pattern recognition, common CNNs employ layers ranging from five to twenty-five.

Biological neural systems activate neural networks. The neuron, which is connected to synapses, is the brain's basic calculating unit. Comparing the biological neuron to a simple mathematical model is shown in Figure 2.
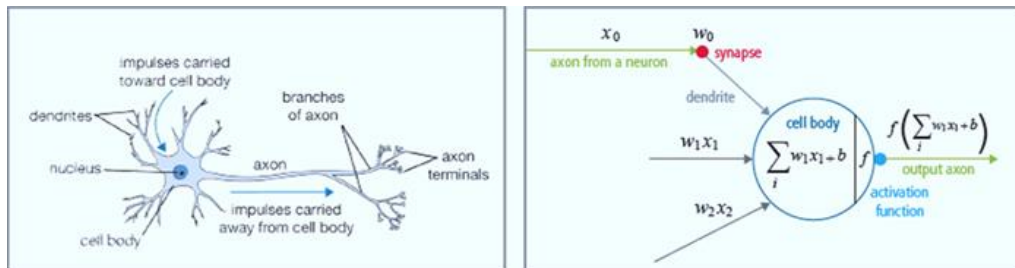
Figure 2: mathematical model and description of a biological brain network

HOW IS CNN VARYING FROM OTHER METHODS USED IN RADIOMICS?

Recent radiomics research uses techniques to extract hand-made features, such as texture analysis, followed by standard mechanical study phases, such as random forests and vector support machines. There are several differences to note between such methods and CNN. First, CNN does not require the release of a handmade feature. Second, CNN structures do not require the separation of tumors or organs by human specialists. Third, CNN is very hungry for data because of the millions of study parameters you can measure, and, as a result, it is very computer-assisted, leading to the need for graphical processing units (GPUs) for model training (Xie et al., 2018).

CNN IN IMAGE PROCESSING

Image and pattern identification, voice recognition, natural language processing, and video processing are only some of the applications of CNNs. Many factors contribute to the rise of convolutional neural networks. Feature extractors in classic pattern recognition models are created by hand. At CNN, the strength of the convolutional layer used to identify the element and the convolution layers used to partition are determined during training.

When it comes to pattern detection and image processing, the best CDRs may be accomplished utilizing CNNs. The MNIST handwriting website yielded a 99.77 percent CDR for CNN, whereas the NORB database yielded a 97.47 percent CDR and a CD for 3D objects [6]. Other information algorithms pale in comparison to CNNs in terms of performance; CNNs outperform humans in tasks like classifying items into fine characters like a specific breed of dog or bird. (Hussain et al., 2018). As shown in Figure 3, the four phases of the common vision pipeline method include image processing, detecting regions of interest (ROI) containing prospective objects, object recognition and decision making. Processing in advance frequently relies on input specifics, particularly the camera system, and is typically performed on a solid cable unit outside of lower vision. When it comes to decision-making at the end of the pipeline, the decisions are frequently based on real items rather than sophisticated computer problems or memory issues. When it comes to object acquisition and recognition, CNN now has a greater impact (Li et al., 2014).
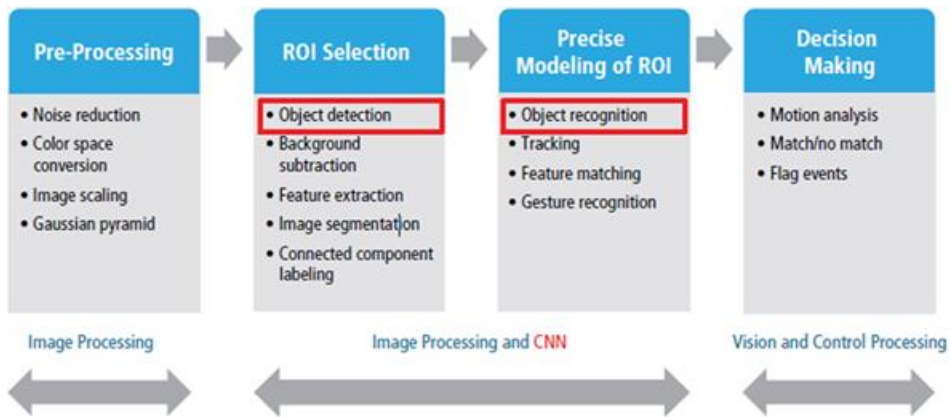
Figure 3: Vision algorithm pipeline

THREE LAYERS OF CNN

Image and video recognition is a specialty of Convolutional Neural Networks. For tasks like image recognition, object detection, and classification, CNN is frequently utilized.

Figure 4 depicts the three different types of layers found in Convolutional Neural Networks:

- Convolutional Layout: Each input neural neuron is linked to the next hidden layer in a shared network. To the hidden layer of neurons in CNN, just a small portion of input layer neurons are connected.
- Blend layer: Blend layer is used to reduce the size of the feature map. There will be multiple layers of unlocking and integration within CNN's hidden layer.
- Fully Cached Layout: Fully Combined Layers form a few storage layers in the network. The entries in the fully integrated layer are the result of the Final Addition or Conversion layer, which is abbreviated and inserted into the fully linked layer (Lee & Kwon, 2017).
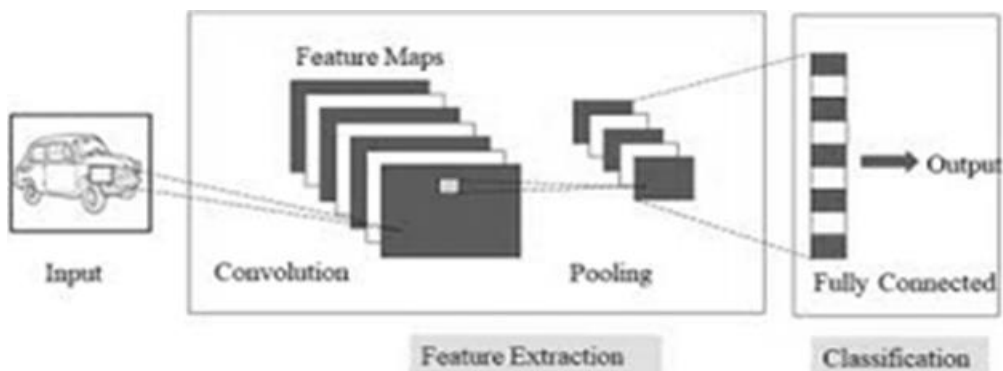

Figure 4: three layers of CNN

BENEFITS OF CNN IN IMAGE PROCESSING

Even though neural networks and other patterns have been there for more than half a century, convolutional neural networks have made considerable strides in recent years. In this section, we'll talk about the advantages of image recognition software like CNN.

Difficulty distortion and changing in the image: Detection with the help of CNN changes position due to the lens, lighting conditions, placement, existence of partial closure, horizontal or vertical shifts, etc. is distorted. CNNs, on the other hand, haven't altered because they all employ the same weight suspension. Theoretically, switching between layers can be made more flexible by employing architectures with completely integrated layers. Nonetheless, the end outcome of this situation is that many units have the same weight patterns regardless of where they are placed. A vast number of test sessions will be needed to examine this weight preparation in order to cover all possible variations.

A few memory requirements: An order of 106 coefficients will be needed to extract features from a 32x32-pixel image and a 1000-encrypted layer in this hypothetical case, requiring a lot of memory. The memory requirements for the convolutional layer are considerably decreased because the same coefficient is applied to multiple locations in space.

Better and easy training: Since the number of parameters in a conventional neural network like CNN might be very big, the training time will rise in tandem. Reduced parameter counts lead to shorter training times for convolutional neural networks (CNN). A common neural network that performs similarly to CNN can be created using full training. There are numerous parameters in a typical neural network that is equivalent to a CNN during training, which can result to more noise than necessary. As a result, the CNN-like performance of a conventional neural network will continue to be exceedingly low (Zhang et al., 2018).

CNN FOR IMAGE CLASSIFICATION AND LOCALIZATION

Aside from reporting the type of object discovered in the image, it is important to include connections to any connecting boxes that may be present. This kind of activity presupposes that there is only one instance of the object in the image.

Attaching the "backing head" over the "split head" to the standard partition network can do this. Convolution and integration, known as "feature mapping," is incorporated in a fully linked network that provides the vector of class opportunities in separation networks. The separating head is a fully linked network that is modified using a phase-loss function (Lc) as categorical cross entropy. The retrospective head, on the other hand, is a fully connected network that captures the feature map and outputs a vector (x, y, w, h) reflecting the linkages remaining over x and y, the width and height of the binding box. As a result, the entire network is tuned using a continuous loss function (Lr) that is a square error.

$L=\alpha LC+(1-\alpha)Lr$

SEMANTIC SEGMENTATION

Separation on the basis of semantics creates even another set of issues. In this case, the goal is to assign a category to every pixel in the image. Creating a separator network for each pixel would be the initial step. The input is a little neighbor right close to each pixel. It is

possible that implementing the picture using convolutions to increase the depth of the feature while keeping the width and length of the image same will improve this approach. It's possible to send a whole feature map of pixels via a fully linked network in order to determine what type of pixels they are. The cost of this is prohibitive, hence it is rarely employed in practice (Chimakurthi, 2018).

There are three different ways to encode a picture using the CNN network. The encoder decreases its width and height but increases its depth (the number of features), while the decoder utilizes updated convolution functions to expand the image and reduce its depth. Going to the other side of the standard convolution with transpose convolution (or up sampling) is what is meant by this technique. The split map is the network's output, while its input is an image. (Han et al., 2018).

CNN IMPLEMENTATION FOR IMAGE DETECTION

Finding an object is a lot like division and local practise in terms of difficulty. Each of us now has to find our own class links and binding boxes in the picture, which is a major change. Additionally, there is no way to predict how many objects will be included or how large they will be. Obviously, this is a big deal and a lot of effort has been put into it. As a first step, we may use the separation and localization networks we discussed before to produce numerous random plants in the input image and for each plant. Because of the damage to the computer, this strategy is unlikely to work well. Use a tool like Selected Search, which employs common computer detection algorithms to locate images in the image that may include objects. The self-identification network known as the "Regional Proposal Network" or R-CNN is made up of these "Regional Suggestions" areas (Figure 5). Resized and networked regions were used to create vector images in the original R-CNN. (Ren et al., 2017).
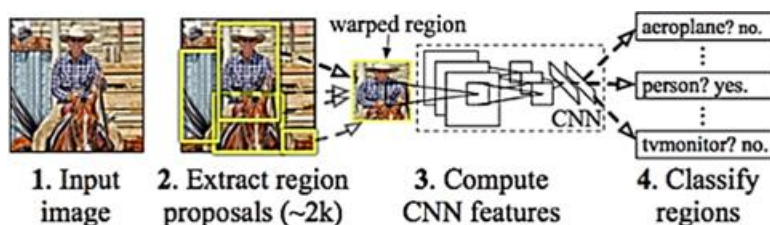

Figure 5: R-CNN: regions with CNN feature

Fast R-CNN was the name given to the next iteration of the R-CNN network. Regional ideas are still sent to Fast R-CNN via an external tool; however, instead of feeding each region's proposal to CNN, the entire image is given to CNN, and the regional suggestions are displayed on the final feature map. In order to create a vector ROI feature, each interested region is given a ROI integration framework and then a completely connected network. Integration of ROI is a common practice in convolutional neural network acquisition. This layer leverages high integration to turn features inside any valid region of interest into a tiny feature map with a defined area of H x W. (where H and W are two parameters). There are now two fully-connected networks, one for the ROI prediction phase, and one for making link box adjustments. Compared to R-CNN, Fast R-CNN is around 25 times quicker. Faster R-CNN (implementation can be obtained) removes the external regional proposal approach and instals a professional component, known as the Regional Proposal Network

(RPN), within the network itself. of R-CNN Faster. You Only Look Once (YOLO), for example, falls into a different type of discovery networks known as Single Shot Detectors (SSD). A grid is used to divide each image into a predetermined number of portions in certain circumstances. YOLO employs a 7 by 7 grid, resulting in the 49 images shown here. A pre-determined plant set with various aspect ratios for each photograph is used to achieve this effect. When given B-binding boxes and classes of C objects, each image generates a $7 * 7 * (5B + C)$-dimensional vector as an output. All of the binding boxes are strong in their confidence and compact in terms of their dimensions (w, h, x, y). It's CNN's YOLO network that is making this shift. The findings of this vector are combined to produce final predictions and binding boxes. One conversion network at YOLO predicts binding boxes and related class possibilities. Even though YOLO is an efficient technique for detecting large objects, the algorithm may miss small ones (Chimakurthi, 2017a).

INSTANCE SEGMENTATION

Although the method of assigning a class name to each of an image's pixels is similar to pattern separation, there are some major differences. Image classification is the first step in making sense of the image. First of all, you don't need to label all of the individual pixels in your shot. While the acquisition of an object is often analogous to this process, we are instead looking for a binary mask that covers each item. Mask R-CNN network information may be found by using the second definition. With an additional CNN in front of its rear head (Fig. 6), the Faster R-CNN is transformed into a binary mask by the Binding Box Links included in the R-CNN (which are reported for each ROI) (Potluri et al., 2011).
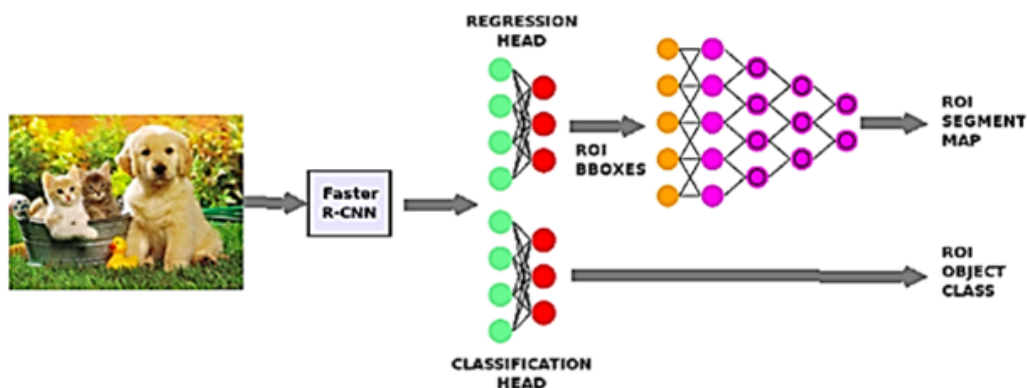

Figure 6: architecture of mask R-CNN

WHAT IS THE FUTURE OF CNN IN IMAGE PROCESSING?

Duplicate neural networks (RNNs) that use short-term memory are among the interesting topics for neural network research (LSTM). Sophisticated time-sensitive tasks, such as speech recognition and handwriting analysis, are covered in these sections. Handwriting, speech, and photography may all be generated by RNNs or auto encoders, and the distribution of these outputs is well-known. Networks that use limited Boltzman (RMBs)/auto encoder tools, such as deep belief networks, can be trained in the most advanced systems since they can train greedily one layer at a time networks (Perez-Munuzuri et al., 1993).

CONCLUSION

On pattern and picture recognition issues, CNN performs exceptionally well, but it does not always outperform humans. CNN's patented algorithms and structures have led to industry-leading results for a number of companies. For the first time ever, researchers have used hierarchical CNNs to recognize images. Complexity reduction has been achieved with 86 aspects of CDR deterioration of less than 2% by a new trade-versus-complexity algorithm designed by innovators. As well as the capabilities required to handle image signal processing, new versions of CNN have all the functionality needed to operate the software effectively. As a result, CNN is currently being employed in a variety of industries and applications for image recognition and processing. In the coming years, this neighborhood will see even more development and increased services.

REFERENCES

Chimakurthi, V. N. S. S. (2017a). Cloud Security - A Semantic Approach in End to End Security Compliance. Engineering International, 5(2), 97-106. https://doi.org/10.18034/ei.v5i2.586

Chimakurthi, V. N. S. S. (2017b). Risks of Multi-Cloud Environment: Micro Services Based Architecture and Potential Challenges. ABC Research Alert, 5(3). https://doi.org/10.18034/abcra.v5i3.590

Chimakurthi, V. N. S. S. (2018). Emerging of Virtual Reality (VR) Technology in Education and Training. Asian Journal of Humanity, Art and Literature, 5(2), 157-166. https://doi.org/10.18034/ajhal.v5i2.606

Chimakurthi, V. N. S. S. (2019). Efficacy of Augmented Reality in Medical Education. *Malaysian Journal of Medical and Biological Research*, 6(2), 135-142. https://doi.org/10.18034/mjmbr.v6i2.609

Chimakurthi, V. N. S. S. (2019a). Implementation of Artificial Intelligence Policy in the Field of Livestock and Dairy Farm. *American Journal of Trade and Policy*, 6(3), 113-118. https://doi.org/10.18034/ajtp.v6i3.591

Chimakurthi, V. N. S. S. (2019b). Application Portfolio Profiling and Appraisal as Part of Enterprise Adoption of Cloud Computing. *Global Disclosure of Economics and Business*, 8(2), 129-142. https://doi.org/10.18034/gdeb.v8i2.610

Chua, L. O., & Roska, T. (1993). The CNN paradigm. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 40*(3), 147-156.

Crounse, K. R., & Chua, L. O. (1995). Methods for image processing and pattern formation in cellular neural networks: A tutorial. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 42*(10), 583-601.

Han, D., Liu, Q., & Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications, 95*, 43-56.

Hussain, M., Bird, J. J., & Faria, D. R. (2018). *A study on cnn transfer learning for image classification.* Paper presented at the UK Workshop on computational Intelligence.

Lee, H., & Kwon, H. (2017). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Transactions on Image Processing, 26*(10), 4843-4855.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014). *Medical image classification with convolutional neural network.* Paper presented at the 2014 13th international conference on control automation robotics & vision (ICARCV).

Liu, M. C., Dubé, L. M., Lancaster, J., & Group, Z. (1996). Acute and chronic effects of a 5-lipoxygenase inhibitor in asthma: a 6-month randomized multicenter trial. *Journal of Allergy and Clinical Immunology, 98*(5), 859-871.

Perez-Munuzuri, V., Perez-Villar, V., & Chua, L. O. (1993). Autowaves for image processing on a two-dimensional CNN array of excitable nonlinear circuits: flat and wrinkled labyrinths. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 40*(3), 174-181.

Potluri, S., Fasih, A., Vutukuru, L. Kishore, A. M. F., & Kyamakya, K. (2011). *CNN based high performance computing for real time image processing on GPU.* Paper presented at the Proceedings of the Joint INDS'11 & ISTET'11.

Ren, X., Guo, H., Li, S., Wang, S., & Li, J. (2017). *A novel image classification method with CNN-XGBoost model.* Paper presented at the International Workshop on Digital Watermarking.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). *Cnn-rnn: A unified framework for multi-label image classification.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Williams, R. T. (2018). Confidence Interventions: Do They Work?. Asian Journal of Humanity, Art and Literature, 5(2), 123-134. https://doi.org/10.18034/ajhal.v5i2.536

Williams, R. T. (2020). A Systematic Review of the Continuous Professional Development for Technology Enhanced Learning Literature. *Engineering International*, *8*(2), 61-72. https://doi.org/10.18034/ei.v8i2.506

Williams, R. T., & Scott, C. D. (2019). The Current State of Outdoor Learning in a U.K Secondary Setting: Exploring the Benefits, Drawbacks and Recommendations. ABC Journal of Advanced Research, 8(2), 109-122. https://doi.org/10.18034/abcjar.v8i2.537

Xie, W., Zhang, C., Zhang, Y., Hu, C., J., H., & Wang, Z. (2018). *An energy-efficient FPGA-based embedded system for CNN application.* Paper presented at the 2018 IEEE International Conference on Electron Devices and Solid State Circuits (EDSSC).

Zhang, M., Li, W., & Du, Q. (2018). Diverse region-based CNN for hyperspectral image classification. *IEEE Transactions on Image Processing, 27*(6), 2623-2634.

--0--