A Review of Deep Learning-based Computer Vision Applications in Agriculture

Jayan P Vijayan

M.Tech student, Dept. of ECE.

Govt. College of Engineering,

Idukki, Kerala, India

Abstract- **A deep learning algorithm consists of an artificial neural network [(AI)], which resembles the biological brain structure. It belongs to the field of artificial intelligence, where machines perform tasks that typically require some kind of human intelligence. Mimicking the learning process of humans with their senses, deep learning networks are fed with sensory data, like texts, images, videos, or sounds. There is another branch of artificial intelligence called Computer vision which uses convolutional neural networks (CNNs) to process visual data at the pixel level and deep learning recurrent neural networks (RNNs) to understand how one pixel relates to another. There are several review articles about deep learning and computer vision, which are focused on specific scientific fields or applications, for example, deep learning advances in computer vision or in specific tasks like plant disease detection. This review aims to discuss deep learning-based computer vision applications in agriculture.**

Index Terms— Deep learning, Artificial neural networks, Machine learning, Computer vision.

I. INTRODUCTION

Deep learning is the subset of artificial intelligence which is trying to achieve this by mimicking the learning of a human brain. Imitating the physiological structure of a brain, which consists of billions of neurons and connections between them, a deep learning algorithm consists of an artificial neural network of interconnected neurons. Fig.1 shows the relationship between artificial intelligence and deep learning.
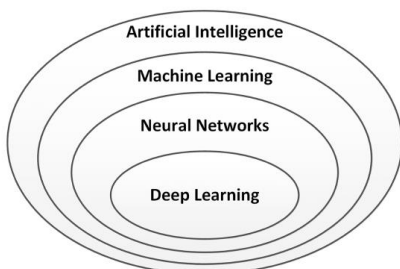

Fig.1 Artificial intelligence and deep learning

II. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks (CNNs) [1] are a special type of artificial neural networks (ANNs), which usually include more than one convolutional layer. CNNs are widely used in image processing and language processing. They have a superior ability of information handling for a large amount of data.

2.1 Different layers

The following are definitions of different layers shown in the above architecture:

2.1.1 Convolutional layer

Convolutional layers are made up of a set of filters called kernels that are applied to an input image. The output of the convolutional layer is a feature map, which is a representation of the input image with the filters applied. Convolutional layers can be stacked to create more complex models, which can learn more intricate features from images.

2.1.2 Pooling layer

Pooling layers are a type of convolutional layer used in deep learning. Pooling layers reduce the spatial size of the input, making it easier to process and requiring less memory. Pooling also helps to reduce the number of parameters and makes training faster. There are two main types of pooling: max pooling and average pooling. Max pooling takes the maximum value from each feature map, while average pooling takes the average value. Pooling layers are typically used after convolutional layers in order to reduce the size of the input before it is fed into a fully connected layer.

2.1.3 Fully connected layer

Fully-connected layers are one of the most basic types of layers in a convolutional neural network (CNN). As the name suggests, each neuron in a fully-connected layer is Fully connected- to every other neuron in the previous layer. Fully connected layers are typically used towards the end of a CNN- when the goal is to take the features learned by the previous layers and use them to make predictions. For example, if we were using a CNN to classify images of animals, the final Fully connected layer might take the features learned by the previous layers and use them to classify an image as containing a plant, leaf, etc.

As the name describes, CNNs are based on convolution shown in fig. 2. At first, in CNN as shown in fig. 3 the image is flipped for both, the rows and columns. Then the kernel slides over the flipped image, each element is multiplied by its corresponding pixel in the flipped image and summed up. The size of the output depends on both, the input image and the kernel.

Special Issue - 2022

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
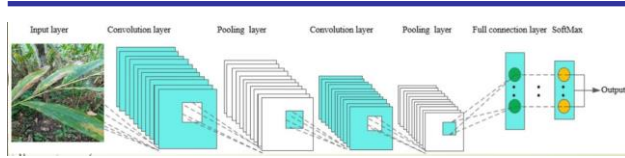**ICCIDT - 2022 Conference Proceedings**

Fig. 2 A CNN applied for image classification

The images are given as input to the network, propagated through several convolutional layers, pooling layers and one fully connected layer. Finally, the possibility of the input image representing required class is displayed.
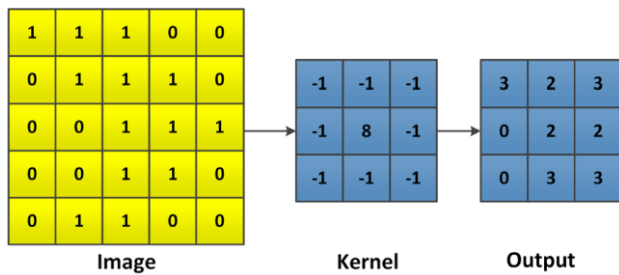

Fig.3 Example of a general 2D convolution.

The mathematical expression of a 2D convolution of an image is:

$$g(i,j) = l \times k(i,j) = \sum_{j=0}^{m-1} \sum_{i=0}^{n-1} l(m,n)k(i-m, j-n),$$

in which k(i, j) is the input image, l is the convolution or filter kernel and g(i, j) is the convolved/filtered image. The equation shows the basic form of a 2D convolution, which can also describe the similar calculation of a 3D convolution as shown in Fig. 11. Like traditional neural networks, a convolutional neural network consists of an input and an output layer, as well as several hidden layers which usually consist of convolutional layers in combination with other layer types, such as pooling layers and fully connected layers. The input images of CNNs are usually in the size of image height $\times$ image width $\times$ image depth. After the convolution, the images are abstracted to so-called feature maps and passed to the next layer. A convolutional layer has the following parameters:

1. The number of input and output channels;
2. Convolutional kernels are defined by their shape.

During the training process of CNNs, the feature maps can become very large, which results in a big amount of computational resources required. One of the methods to solve this problem is the pooling layer. Pooling layers reduce the dimensions of feature maps from convolutional layers and pass the reduced data to the next layer. Different functions for pooling are available, such as max pooling, average pooling and sum pooling, where max pooling is most commonly used. With max pooling or average pooling operations, the 4×4 feature map is transformed to a 2×2 feature. The stride controls how the pooling filter is moved over the input volume.

At the end of a CNN, there are usually one or several fully connected layers, which connect every neuron in the last layer to every neuron in itself. The fully connected layers are used to concatenate the feature maps

into the desired output values. For example the probability of an image showing weed or crop.

2.2 Fully convolutional neural networks

As one of the most common tasks in medical imaging, automatic segmentation is challenging because of the huge difference between different patients (anatomy and pathology). In this field, however, neural networks have shown great advantages to learn image features automatically from the medical images and corresponding ground truths [2] In addition, the development of fully convolutional neural networks (FCNs) further improved the advantages of deep learning in the area of image segmentation and particularly semantic segmentation. Semantic segmentation is to understand what is in the image on a pixel level, which can be also defined as to label each pixel of an image with a corresponding class. Figure 4 shows an example for semantic segmentation.
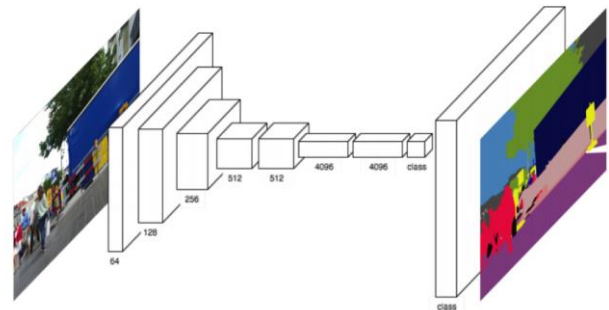

Fig.4 FCN for a semantic segmentation

Fully convolutional neural networks were developed by Long, [3] based on normal convolutional neural networks. In FCNs, the final fully connected layer of CNNs are replaced by convolutional layers, so that the images are not downsized and the output will not be a single label as in CNNs. Instead, a pixel-wise output can be calculated to represent each pixel in the input image. Fig.4 shows a typical setup of an FCN for a semantic segmentation.

Based on the improvement of FCNs, there are currently many deep learning methods in computer visions that are applied for segmentation tasks in medical imaging. Most of these methods are based on FCNs that learn the features of spatial dimensions from the original images.

2.3 Different CNN Architectures

A CNN is made up of several layers that process and transform an input to produce an output. Table 1 shows the comparison among different architectures.

Table 1 Comparison among different CNN Architectures

Model	No. of Layer	Parameters (Million)	Size
AlexNet	8	60	-
VGGNet-16	23	138	528 MB
VGGNet-19	26	143	549 MB
Inception-V1	27	7	-
Inception-V3	42	27	93 MB
ResNet-152	152	50	132 MB
ResNet-101	101	44	171 MB
InceptionResNetV2	572	55	215 MB
MobileNet-V1	28	4.2	16 MB
MobileNet-V2	28	3.37	14 MB
EfficientNet B0	-	5	-

Special Issue - 2022

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCIDT - 2022 Conference Proceedings

III COMPUTER VISION APPLICATIONS

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs and take actions or make recommendations based on that information. Computer vision uses convolutional neural networks (CNNs) to processes visual data at the pixel level and deep learning recurrent neural network (RNNs) to understand how one pixel relates to another.

Computer vision systems use (1) cameras to obtain visual data, (2) machine learning models for processing the images, and (3) conditional logic to automate application-specific use cases. The deployment of artificial intelligence to edge devices, so-called edge intelligence, facilitates the implementation of scalable, efficient, robust, secure, and private implementations of computer vision. Uses for computer vision include:

1. *Biometric access management* - CV plays an important role in both facial and iris recognition.
2. *Industrial robots and self-driving cars* - CV allows robots and autonomous vehicles to avoid collisions and navigate safely.
3. *Digital diagnostics* - CV can be used in tandem with other types of artificial intelligence programming to automate the analysis of X-rays and MRIs.
4. *Augmented reality* - CV allows mixed reality programming to know where a virtual object should be placed.
5. *Plant disease detection* - CV allows accurate estimation of disease severity is essential for food security, disease management, and yield loss prediction.

Computer vision algorithms lead to numerous real-world applications, like automatic face or license plate detection and recognition, or even self-driving cars, just to name a few. However, making the algorithms reliable enough for their specific task remains challenging. A missed person on a group photo for automatic tagging on a social network website may not be very dramatic, but a missed pedestrian by a self-driving car can have fatal consequences.

IV COMPUTER VISION IN AGRICULTURE

The agricultural sector has witnessed a lot of contributions when it comes to artificial intelligence (AI) and computer vision in areas like plant health detection and monitoring, weeding, harvesting, etc [4]. Some of the applications are discussed below.

4.1 Crop Monitoring

The yield and quality of important crops such as rice and wheat determine the stability of food security. Traditionally, crop growth monitoring mainly relies on subjective human judgment and is not timely or accurate. Computer Vision applications allow to continuously and non-destructively monitor plant growth and the response to nutrient requirements.

Compared with manual operations, the real-time monitoring of crop growth by applying computer vision technology can detect the subtle changes in crops due to malnutrition much earlier and can provide a reliable and accurate basis for timely regulation.

In addition, computer vision applications can be used to measure plant growth indicators or determine the growth stage.

4.2 Plantation Monitoring

In intelligent agriculture, image processing with drone images can be used to monitor palm oil plantations remotely. With geospatial orthophotos, it is possible to identify which part of the plantation land is fertile for planted crops. It was also possible to identify areas less fertile in terms of growth and parts of plantation fields that were not growing at all.

4.3 Plant Disease Detection

Automatic and accurate estimation of disease severity is essential for food security, disease management, and yield loss prediction. The deep learning method avoids labor-intensive feature engineering and threshold-based image segmentation. Automatic image-based plant disease severity estimation using Deep CNN applications were developed, for example, to identify apple black rot.

4.4 Insect Detection

Rapid and accurate recognition and counting of flying insects are of great importance, especially for pest control. However, traditional manual identification and counting of flying insects are inefficient and labor-intensive. Vision-based systems allow the counting and recognizing of flying insects (based on You Only Look Once (YOLO) object detection and classification).

4.5 Weed Detection

Weeds are considered to be harmful plants in agronomy because they compete with crops to obtain the water, minerals, and other nutrients in the soil. Spraying pesticides only in the exact locations of weeds greatly reduces the risk of contaminating crops, humans, animals, and water resources [5].

The intelligent detection and removal of weeds are critical to the development of agriculture. A neural network-based computer vision system can be used to identify potato plants and three different weeds for on-site specific spraying.

4.6 Automatic Harvesting

In traditional agriculture, there is a reliance on mechanical operations, with manual harvesting as the mainstay, which results in high costs and low efficiency. However, in recent years, with the continuous application of computer vision technology, high-end intelligent agricultural harvesting machines, such as harvesting machinery and picking robots based on computer vision technology, have emerged in agricultural production, which has been a new step in the automatic harvesting of crops.

Special Issue - 2022

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCIDT - 2022 Conference Proceedings

The main focus of harvesting operations is to ensure product quality during harvesting to maximize the market value. Computer Vision powered applications include picking cucumbers automatically in a greenhouse environment or the automatic identification of cherries in a natural environment.

V CONCLUSION

Agriculture is the scientific process that deals with the cultivation of plants to produce food. To address the real time problems using computers deep learning-based computer vision algorithms are used. Deep learning is the subset of artificial intelligence which is trying to achieve this by mimicking the learning of a human brain. In the case of deep learning Convolutional neural networks (CNNs) are used as a special type of artificial neural networks (ANNs), which usually include more than one convolutional layer. It is a network architecture for deep learning. It learns directly from images. A CNN is made up of several layers that process and transform an input to produce an output. CNNs are widely used in computer vision applications. Various research happened in the field of agriculture computer vision techniques are used for crop monitoring, plantation monitoring, plant disease detection, insert detection, weed detection, automatic harvesting, etc.

VI REFERENCES

[1] Krizhevsky A, Sutskever I, Hinton GE. 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25:1097–1105 DOI 10.1145/3065386.

[2] Hesamian MH, Jia W, He X, Kennedy P. 2019. Deep learning techniques for medical image segmentation: achievements and challenges. Journal of Digital Imaging 32(4):582–596 DOI 10.1007/s10278-019-00227-x.

[3] Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 3431–3440.

[4] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, ``Deep learning for plant identi_cation using vein morphological patterns,'' *Comput. Electron. Agricult.*, vol. 127, pp. 418_424, Sep. 2016.

[5] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, ``A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network,'' *Comput. Electron. Agricult.*, vol. 154, pp. 18_24, Nov. 2018.