ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 6, June 2023

Semantic Segmentation using Convolutional Neural Networks

[1] Maaz Ansari*, [2] Dr.Archana Choudhary, [3] Dr. Surendra Bhosale

[1] [2] [3] Veermata Jijabai Technological Institute (VJTI), Mumbai, Maharashtra, India
Corresponding Author Email: [1] maansari_m19@ee.vjti.ac.in, [2] apchoudhary@ee.vjti.ac.in, [3] sjbhosale@ee.vjti.ac.in

***Abstract*— The task of partitioning an image into multiple segments is known as image segmentation. This helps in analyzing the given image. This thing we are looking for in computer vision. It has applications such as self-driving cars, delivery drones, object detection, medical image analysis, robotic perception, video surveillance, etc. In this paper, semantic image segmentation is performed using convolutional neural network for training and validating computer vision algorithms and to organize the data contained within images and videos into meaningful categories.**

***Index Terms*— Semantic Segmentation, Convolutional Neural Networks, Computer Vision, Deep Learning, encoder-decoder models, Max Pooling.**

I. INTRODUCTION

Image segmentation is a process in which an image is torn into small groups. These groups reduces the complexity of the image. This makes further analysis of image easier. Segmentation is assigning labels to pixels. Pixels belonging to the same class have a similar label allotted to them. Take a photo as input for object detection. Instead of working on the whole image, the detector can be inputted by a segmentation algorithm. This will reduce inference time.

In this, pixels need to be grouped together on the basis of some characteristics. On the basis of these characteristics, image segmentation can be divided into the following:

1). Semantic Segmentation
2). Instance Segmentation

Semantic Segmentation:

In this, each and every pixel in the picture belongs to one particular class. Pixels belonging to one class have been given one color. To give a definition, semantic segmentation is the work of allotting a label to every pixel in a given picture. This is different from classification. Classification allots one class to the whole picture whereas semantic segmentation divides every pixel of the picture to one of the classes. Self-driving vehicles and portrait mode on google's pixel phone are two examples of semantic segmentation.

Instance Segmentation:

In instance segmentation, different instances of the same class are divided individually. Different labels are given to different instances of the same class.

II. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

The capacity of machine learning to close the space between anthropoid and computer literacy has been increasing drastically. Both office workers and layman concentrate on abundant aspect of the discipline to attain great outcome. The area of machine vision is one of such field. The motive of this area is to make it happen for machines to do variety of things including picture and video recognition, image analysis, media reproduction, recommendation systems, natural language processing, etc. The improvements in computer vision made possible by deep learning have been developed over time, notably through the use of one particular algorithm: the convolutional neural network.

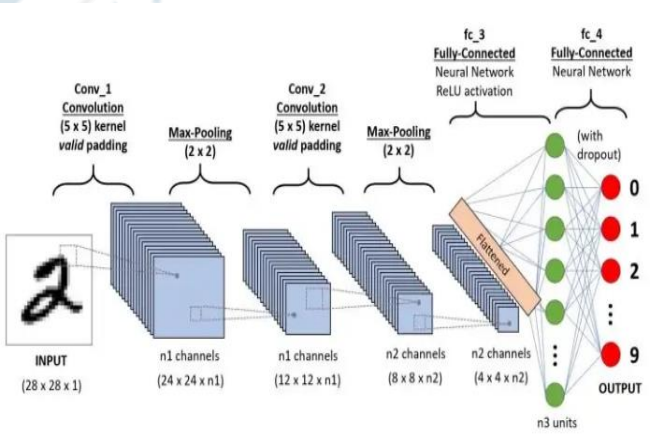

Figure 1. CNN sequence to classify handwritten digits [11].

Convolutional Neural Network (CNN) is a system that can take in an input picture, give distinct things in the picture graspable weights to differentiate between them. CNN requires considerably less initialisation than other categorization procedure. CNNs can acquire a knowledge of these traits, whereas in manual filtering techniques, these features must be hand-engineered. The shape of a CNN was affected by the way the visual cortex is arranged and is alike to the connectivity design of neurons in the anthropoid brain.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 6, June 2023

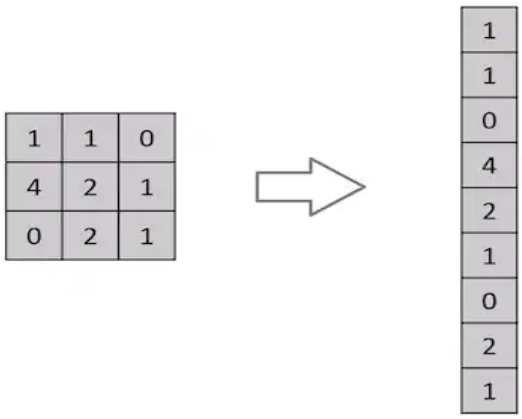

Figure 2. Flattening of a 3×3 image matrix into a 9×1 vector [12].

A matrix of pixel values is all that makes up an image. Therefore, for classification purposes, just flatten the image's 3 by 3 matrix into a 9 by 1 vector and input it to a multi-level perceptron.

III. SEMANTIC IMAGE SEGMENTATION

Semantic segmentation allots a class to each pixel of a picture. We are not segregating instances of the same class, which is a critical distinction to make. Only the categorization of each pixel is important to us. In other words, the segmentation map does not automatically differentiate two things of the same class in the input picture as self-sufficient ones. A dissimilar category of representation called instance segmentation distinguishes between different things belonging to the same class. A few tasks that segmentation models are beneficial for include:

- **Medical image diagnostics**

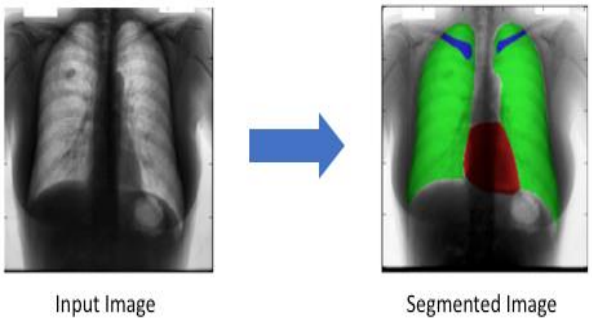

Figure 3. A chest x-ray [13].

Machines can augment analysis performed by radiologists, greatly reducing the time required to run diagnostic tests.

Portraying the job

The objective is to generate a segmentation map from a grayscale image or an RGB colour image in which each pixel includes a class name encoded as an integer. We will generate our goal one-hot encoding by treating the class labels in a manner similar to how we handle normal categorical values, thereby forming an output path for each of the potential categories. By putting a target on top of the observation, we can quickly inspect it. This is referred to as a mask because it highlights the areas of a picture where a particular category is present when we cover a single path of our forecast.

Making an architectural structure

To create a neural network for this job, the simplest way is to just pile several convolutional layers with the identical stuffing to conserve measurements and produce a last segmentation map. By successively transforming feature mappings, this immediately acquire a knowledge of an aligning from the input picture to its appropriate portion. However, maintaining full resolution across the network is highly computationally expensive. Remember that in deep convolutional networks, the initial layers often acquire low-amount idea while the final layers evolve higher-amount and more specialised characteristic alignments. As we go further into the network, we often need to increase the number of feature mappings in order to preserve expressiveness.

Figure 4. Prediction map [13].

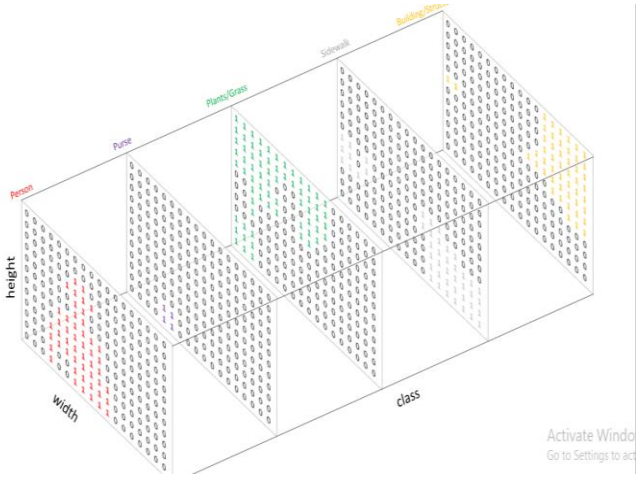

Figure 5. Segmentation map [13]

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 6, June 2023

Figure 6. Regions of an image [13].

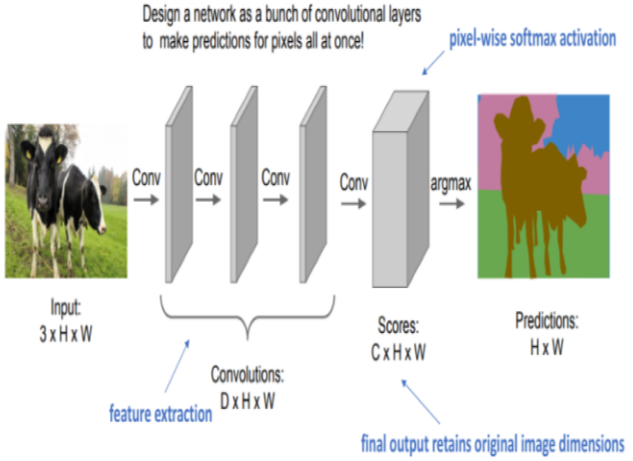

Figure 7. Successive transformation of feature mappings [13].

Because we only care about the content of the image and not its location, this does not necessarily pose a barrier for the task of classifying images. We could reduce the computational load by routinely compressing the spatial resolution of our feature maps using pooling or strided convolutions. However, we intend our prototype to generate a thorough design semantic guess for image segmentation. A common method for creating picture fragmentation prototype is to use an encoder-decoder design. In this method, the dimensional design of the input is downsampled, bottom design characteristic alignments are created and are found to be very effective at classifying images, and then the feature representations are upsampled to create a full-resolution segmentation map.

Approaches to upsampling

We can utilise a variety of techniques to increase the resolution of a feature map. Unpooling action randomize the design by dispersing one merit into a greater design, whereas pooling operations downsample the resolution by summing a local region with a single value, i.e. average or max pooling. Transpose convolutions, however, are by far the most widely used strategy since they enable us to create a learnt upsampling. A transpose convolution effectively accomplishes the reverse of what a regular convolution operation would do, which is to lay hold of the inner product of the merits now in filters sight and construct one merit for the analogous output spot. In order to project those weighted values into the output feature map for transpose convolution, we select one merit from the bottom-design characteristic map and multiply all of the weights in our filter by this merit.

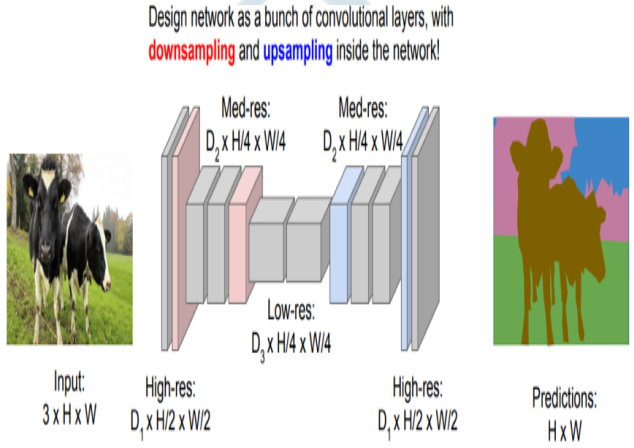

Figure 8. Downsampling and upsampling [13].

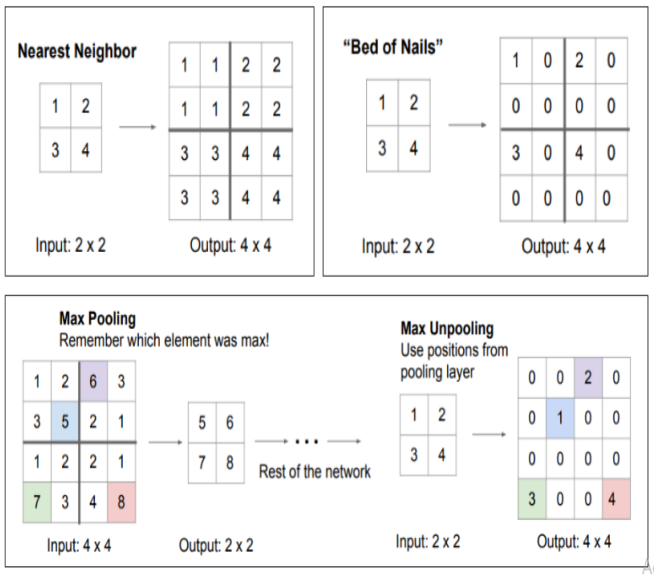

Figure 9. Max pooling and Max unpooling [13].

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 6, June 2023

IV. RESULTS

Figure 10. VJTI Image

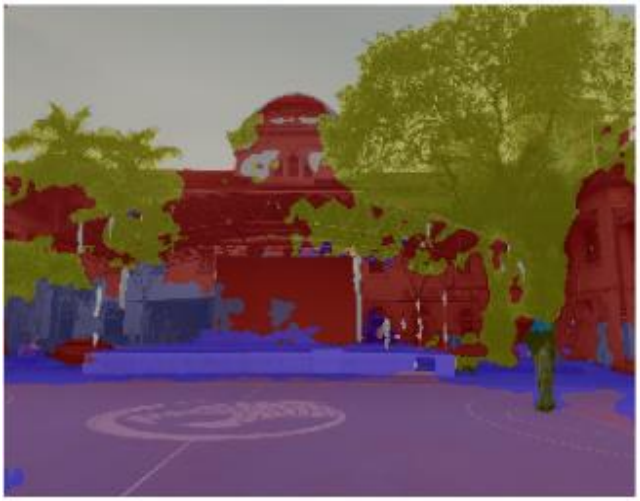

Figure 11. Semantic Segmented Image

In the above figure, VJTI image is shown and output image is semantically segmented.

V. CONCLUSION AND FUTURE SCOPE

The process of grouping together portions of photographs that correspond to the same object class is known as semantic segmentation. The detection of road signs, the detection of cancers, the detection of surgical tools, the segmentation of colon crypts, and the categorization of land use and land cover are only a few applications for this form of segmentation. Non-semantic segmentation, in contrast, just groups pixels based on the general traits of single objects. Because various segmentations may be appropriate, the task of non-semantic segmentation is not well defined. In contrast to semantic segmentation, object detection must distinguish between various instances of the same thing. While having a semantic segmentation is undoubtedly helpful when attempting to identify object instances, there are a few issues to be aware of. For example, adjacent pixels in the same class may belong to separate object instances, and regions that are not related may do so as well. Semantic segmentation has applications in facial recognition, self-driving cars, virtual fitting rooms, medical imaging and diagnostics, etc.

REFERENCES

[1] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[2] D. E. Rumelhart, G. E. Hinton, R. J. Williams et al., "Learning representations by back-propagating errors," Cognitive modeling, vol. 5, no. 3, p. 1, 1988.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[5] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[6] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," arXiv preprint arXiv:1701.07875, 2017.

[7] W. Liu, A. Rabinovich, and A. C. Berg, "Parsenet: Looking wider to see better," arXiv preprint arXiv:1506.04579, 2015.

[8] A. G. Schwing and R. Urtasun, "Fully connected deep structured networks," arXiv preprint arXiv:1503.02351, 2015.

[9] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," arXiv preprint arXiv:1909.11065, 2019

[10] X. Xia and B. Kulis, "W-net: A deep model for fully unsupervised image segmentation," arXiv preprint arXiv:1711.08506, 2017.

[11] Simonyan, K., and Zisserman, A. (2014) "Very deep convolutional networks for large-scale image recognition". arXiv preprint arXiv:1409.1556.

[12] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way

[13] https://www.jeremyjordan.me/semantic-segmentation/