University for the Creative Arts

BERLIN SCHOOL OF BUSINESS & INNOVATION

Essay / Assignment Title: Computer Vision with a Special Focus on Different Neural Network Architectures for Image Segmentation

Program title: Computer Vision and Artificial Intelligence

Name: Abijith Mullancherry Asokan

Year: 2023

CONTENTS

Contents

Statement of compliance with academic ethics and the avoidance of plagiarism

I honestly declare that this dissertation is entirely my own work and none of its part has been copied from printed or electronic sources, translated from foreign sources and reproduced from essays of other researchers or students. Wherever I have been based on ideas or other people texts I clearly declare it through the good use of references following academic ethics.

(In the case that is proved that part of the essay does not constitute an original work, but a copy of an already published essay or from another source, the student will be expelled permanently from the postgraduate program).

Name and Surname (Capital letters): ABIJITH MULLANCHERRY ASOKAN

Date: 30/09/2023

INTRODUCTION

A neural network is an artificial intelligence strategy for teaching computers to analyze data in a manner inspired by the human brain. Deep learning is a form of machine learning technique that employs linked nodes or neurons in a layered structure that resembles the human brain. It generates an adaptive system that computers may utilize to learn from their mistakes and constantly improve. Thus, artificial neural networks aim to handle complex issues with increased accuracy, such as summarizing papers or identifying faces.

Importance of Neural Networks:

Neural networks can aid computers in making intelligent judgments with minimal human intervention. This is because they can learn and model nonlinear and complicated connections between input and output data. They can, for example, perform the following duties.

Uses of Neural Networks:

Computer Vision:

The capacity of computers to extract information and insights from photos and movies is known as computer vision. Computers can differentiate and recognize pictures in the same way that humans can.

Speech Recognition:

Despite differences in speech patterns, pitch, tone, language, and accent, neural networks can interpret human voice. Speech recognition is used by virtual assistants such as Amazon Alexa and automatic transcription.

Natural Language Processing:

The capacity to process natural, human-created material is referred to as natural language processing (NLP). Neural networks assist computers in extracting meaning and insights from text data and texts.

Recommendation engines:

To produce individualized suggestions, neural networks can track user activities. They may also evaluate all user activity to uncover new items or services that may be of interest to a particular user.

Architecture of Neural Networks:

A neural network architecture has numerous components. Each neural network shares a few components:

Input - Data that is fed into the model for learning and training purposes is referred to as input.

Weight - Weight organizes variables based on their value and influence.

Transfer function - A transfer function is when all the input variables are summarized and integrated into a single output variable.

The activation function's duty is to determine whether a certain neuron should be stimulated. This judgment is made depending on whether the neuron's input will be useful in the prediction process.

Bias changes the value provided by the activation function.

Convolutional Neural Networks

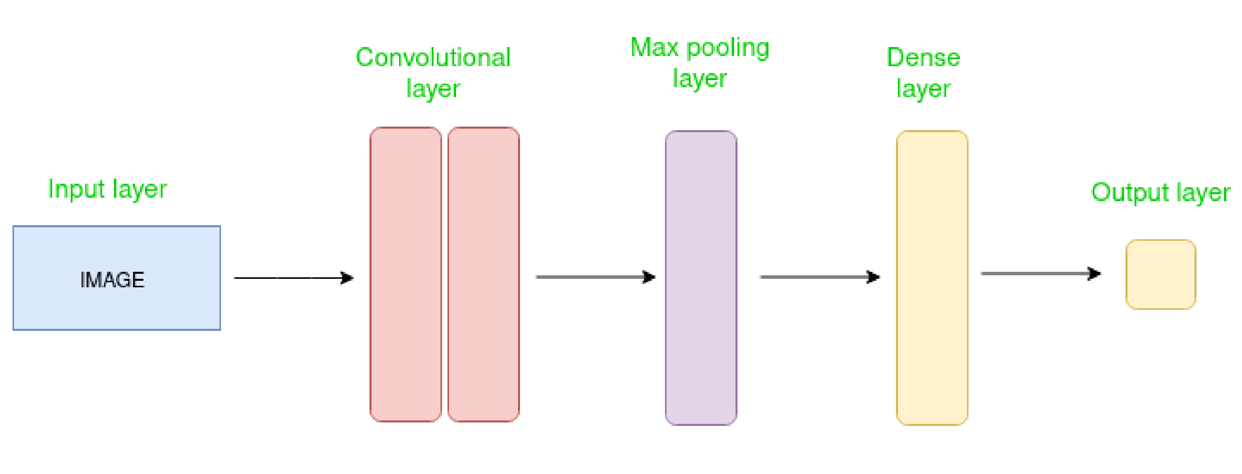
Convolutional neural networks are distinct from other types of neural networks in that they perform better with image, speech, or audio signal inputs. They have three different sorts of layers:

- Layer of convolution
- Layer of aggregation
- FC (fully-connected) layer

The convolutional network starts with the convolutional layer. While convolutional layers can be followed by other convolutional layers or pooling layers, the last layer is the fully-connected layer. With each layer, the CNN grows more complex, identifying more parts of the image. Earlier layers concentrate on fundamental elements like colors and borders. As the visual data is processed by the CNN layers, it learns to discern larger components or characteristics of the item, finally recognizing the target object.

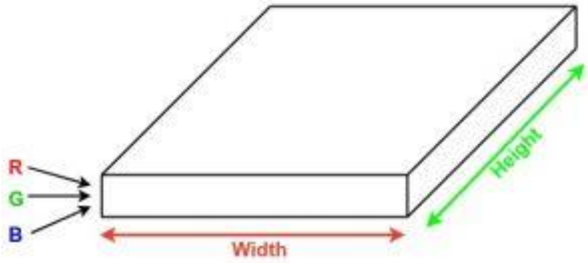
Architecture of CNN:

The input layer, Convolutional layer, Pooling layer, and fully linked layers make up a Convolutional Neural Network.

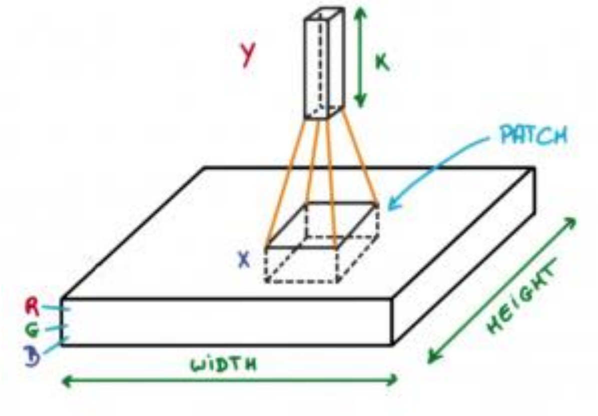


The Convolutional layer extract features from the input picture using filters, the Pooling layer reduces computation by down sampling the image, and the fully connected layer provides the final prediction. Backpropagation and gradient descent are used by the network to learn the best filters.

Convolutional Neural Networks, often known as covnets, are neural networks that share parameters. Assume you have a picture. It may be represented as a cuboid with length, width (picture size), and height.



Consider taking a tiny section of this image and running a little neural network, known as a filter or kernel, on it with, say, K outputs and representing them vertically. Slide the neural network across the whole image, and we will obtain a new image with varying widths, heights, and depths. Instead of simply the R, G, and B channels, we now have more channels but with a smaller width and height. This method is known as convolution. If the patch size is the same as the picture size, the neural network is a normal neural network. We have less weights as a result of this tiny patch.



Mathematics involved in the convolution process:

- Convolution layers are made up of a collection of learnable filters (or kernels) with modest widths and heights and the same depth as the input volume (3 if the input layer is image input).

- During the forward pass, we slide each filter over the whole input volume in steps of stride and compute the dot product of the kernel weights and patch from input volume.

- We will obtain a 2-D output for each filter as we slide them, and we will stack them together to produce an output volume with a depth equal to the number of filters. The network will learn all of the filters.

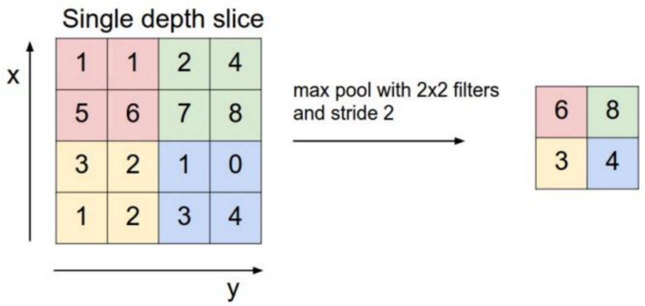
Layers used to build ConvNets:

- **Input Layers:** It is the layer where we provide input to our model. In most cases, the input to CNN will be a picture or a sequence of images. This layer contains the image's raw input, which has 32 width, 32 height, and 3 depth.

- **Convolutional Layers:** This is the layer responsible for extracting the feature from the input dataset. It applies to the input pictures a collection of learnable filters known as kernels. The filters/kernels are typically smaller matrices of 22, 33, or 55 dimensions. It moves across the input image data, calculating the dot product between the kernel weight and the matching input picture patch. This layer's output is known as ad feature maps.

- **Activation Layers:** Activation layers introduce nonlinearity to the network by adding an activation function to the output of the preceding layer. It will apply an element-wise activation function to the convolution layer's output.

- **Pooling Layers:** This layer is added into the covnets on a regular basis, and its major job is to lower the size of the volume, which speeds up computation, saves memory, and avoids overfitting. Pooling layers are classified into two types: maximum pooling and average pooling.



- **Flattening:** Following the convolution and pooling layers, the resultant feature maps are flattened into a one-dimensional vector and sent into a totally connected layer for classification or regression.

- **Fully Connected Layers:** It accepts the preceding layer's input and computes the final classification or regression job.

- **Output Layers:** The output of the fully connected layers is then input into a logistic function for classification tasks, such as sigmoid or softmax, which translates each class's output into the probability score of each class.

Advantages of CNN:

- **Efficient Image Processing:** One of CNNs' main benefits is their capacity to analyze pictures quickly. This is since they employ a process known as convolution, which involves applying a filter to a picture in order to extract characteristics relevant to the job at hand. CNNs can minimize the quantity of information that must be processed in this way, making them quicker and more efficient than other types of algorithms.

- **High Accuracy Rates:** The capacity of CNNs to reach high accuracy rates is another advantage. This is because by studying vast datasets, they may learn to detect complicated patterns in photos. This means they can be trained to recognize specific items or traits with great accuracy, making them excellent for tasks such as facial recognition or object identification.

- **Robust to Noise:** CNNs are also resistant to noise, which means they can identify patterns in pictures that have been deformed or degraded. This is because they utilize numerous layers of filters to extract information from pictures, making them more noise resistant than other types of algorithms.

- **Transfer Learning:** CNNs may also be taught on one job and then utilized to do another activity with little or no extra training. This is because the features retrieved by CNNs are frequently general enough to be employed for a wide range of jobs, making them a useful tool for a wide range of applications.

- **Automated Feature Extraction:** Finally, CNNs can learn to detect patterns in pictures without the need for manual feature engineering since they automate the feature extraction process. This makes them suitable for jobs when the important features are unknown in advance, because the CNN can learn to recognize the necessary features through training.

Disadvantages of CNN:

- **High Computational Requirements:** The high computing needs of CNNs are one of its key drawbacks. This is because CNNs often contain a high number of layers and parameters, which need a significant amount of computing power and memory to train and execute. As a result, they may be impractical for usage in some applications with restricted resources.

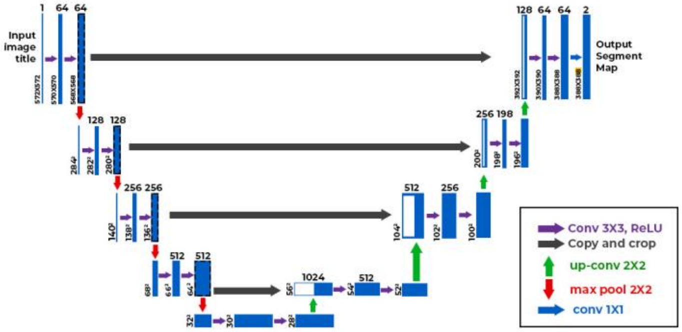
- **Difficulty with small Datasets:** Large datasets are also required for CNNs to obtain high accuracy rates. This is due to the fact that kids learn to detect patterns in pictures by evaluating numerous examples of such patterns. When the dataset is too small, the CNN may overfit, which means it becomes overly specialized to the training dataset and performs badly on fresh data.

- **Vulnerability to Adversarial attacks:** CNNs are also subject to adversarial assaults, which include purposely tampering with the input data in order to trick the CNN into making wrong conclusions. This can be a severe issue in applications such as driverless cars, where safety is paramount.

- **Limited ability to Generalize:** Finally, CNNs can only generalize to new contexts to a limited extent. As a result, they may perform badly on photos that differ much from those in the training dataset. This can be an issue in situations where the CNN must deal with a wide range of pictures.

U-Net Image Segmentation

The publication U-Net: Convolutional Networks for Biomedical Image Segmentation introduces U-Net. The model design is straightforward, consisting of an encoder (for downsampling) and a decoder (for upsampling) with skip connections. As seen in the below figure, it resembles the letter U, thus the name U-Net.



U-Net's design is unique in that it has both a contracting and an expanded route. The contracting path contains encoder layers that capture contextual information and reduce the spatial resolution of the input, whereas the expansive path contains decoder layers that decode the encoded data and generate a segmentation map using information from the contracting path via skip connections.

In U-Net, the contracting path oversees detecting the important characteristics in the input picture. Convolutional processes performed by the encoder layers diminish the spatial resolution of the feature maps while increasing their depth, collecting increasingly abstract representations of the input. This path of contraction is comparable to the feedforward layers in other convolutional neural networks. The expanding route, on the other hand, focuses on decoding the encoded data and identifying the features while keeping the spatial resolution of the input. The expanding path's decoder layers upsample the feature maps while also conducting convolutional operations. The skip connections from the contracting path assist to retain the spatial information lost in the contracting path, allowing the decoder layers to detect the features more correctly.

The diagram depicts how the U-Net network turns a grayscale input picture of size 572*572*1 into a binary segmented output map of size 388*388*2. We can see that the output size is smaller than the input size since there is no padding. However, by using padding, we can keep the input size constant. During the contracting route, the input image's height and breadth are gradually lowered while the number of channels is raised. With more channels, the network may record higher-level characteristics as it continues along the path. A last convolution operation is done at the bottleneck to yield a 30*30*1024 shaped feature map. The expanding route then turns the feature map from the bottleneck into a picture the same size as the original input. This is accomplished with the help of upsampling layers, which raise the spatial resolution of the feature map while decreasing the number of channels. The skip connections from the contraction path are utilized to assist the decoder layers in locating and refining picture features. Finally, each pixel in the output picture correlates to a label in the input image that belongs to a certain object or class. The output map in this example is a binary segmentation map, with each pixel representing a foreground or background area.

Advantages of U-Net:

- Fitted for segmentation: it computes a pixel-wise output (without the convolutions' validity margins). Because we are tackling segmentation jobs here, it should function without change.

- It has a straightforward structure. It's a reiteration of fundamental building blocks: convolutions, ReLu, and max pooling for the downsampling/encoding path, and upsampling, convolutions, and ReLu for the upsampling/decoding path. As a result, it should be simple to implement.

- It has a decent performance. It won several benchmarks when it was first launched, and it still provides for respectable ranks in segmentation problems.

- It works with very minimal training data.

Disadvantages of U-Net:

- **A huge number of parameters**: Because of the skip connections and additional layers in the growing path, UNet has a significant number of parameters. When working with tiny datasets, this might make the model more prone to overfitting.

- **High computational cost**: Because of the skip connections, UNet necessitates additional computations, making it more computationally expensive than comparable systems.

- **Sensitive to initialization**: Because the skip connections might exaggerate any flaws in the initial weights, UNet can be sensitive to model parameter setup. This can make training UNet more complex when compared to other topologies

CNN Vs U-Net

CNN and U-Net differ from each other in every aspect. The below points can be used to convey the differences between them.

- **Architecture:** CNN uses a combination of Faster CNN And FCN whereas U-Net consists of encoders and decoders with skip connections.
- **Types of Images:** CNN is better suited for general purpose images such as natural images, medical images, and satellite images, while U-Net was originally designed for biomedical images. U-Net can be adapted for other images with fine-grained details.
- **Training Process:** CNN requires more data to train compared to U-Net, but can be trained using standard backpropagation techniques. U-Net, on the other hand uses skip connections to transfer low-level features from the encoder to the decoder. This makes the training more challenging, but less data is required when compared to CNN.
- **Performance:** While both models can provide good results in various segmentation benchmarks, CNN provides better results for complex shapes and multiple instances, whereas U-Net provides better results for fine-grained details.

CONCLUDING REMARKS

It is a smart technique to use both CNN and U-Net for picture segmentation. CNNs offer flexibility and scalability for a wide range of computer vision applications, whereas U-Net is designed exclusively for segmentation tasks, providing advantages in situations when fine-grained localization is required. The exact needs of your segmentation task, processing resources, and accessible data should all influence your decision. We have learned and explored how the architecture works for both these datasets. When working with different type of datasets, it is better to work with different networks and find the best suited model by calculating scores. Both these datasets have particular use cases and can yield better results when used with the type of datasets for which each are designed for. There are still several subsets of both these models which can are altered for different designs with small changes in models, which should be taken into consideration while dealing with special use cases.

BIBLIOGRAPHY

- AWS (n.d.). *What is a Neural Network? AI and ML Guide - AWS*. [online] Amazon Web Services, Inc. Available at: https://aws.amazon.com/what-is/neural-network/.
- h2o.ai. (n.d.). *What are Neural Network Architectures? | H2O.ai*. [online] Available at: https://h2o.ai/wiki/neural-network-architectures/#:~:text=The%20architecture%20of%20neural%20networks.
- IBM (2023). *What are Convolutional Neural Networks? | IBM*. [online] www.ibm.com. Available at: https://www.ibm.com/topics/convolutional-neural-networks.
- Convolutional Neural Network (CNN (2019). *Convolutional Neural Network (CNN) | TensorFlow Core*. [online] TensorFlow. Available at: https://www.tensorflow.org/tutorials/images/cnn.
- GeeksforGeeks. (2017). *Introduction to Convolution Neural Network*. [online] Available at: https://www.geeksforgeeks.org/introduction-convolution-neural-network/.
- PyImageSearch. (2022). *U-Net Image Segmentation in Keras*. [online] Available at: https://pyimagesearch.com/2022/02/21/u-net-image-segmentation-in-keras/.
- GeeksforGeeks. (2023). *U-Net Architecture Explained*. [online] Available at: https://www.geeksforgeeks.org/u-net-architecture-explained/.
- AspiringYouths. (n.d.). *Advantages and Disadvantages of Convolutional Neural Network (CNN)*. [online] Available at: https://aspiringyouths.com/advantages-disadvantages/convolutional-neural-network-cnn/.
- VisoByte. (2023). *Mask R-CNN vs U-Net: A Comparison of Two Popular Image Segmentation Methods*. [online] Available at: https://www.visobyte.com/2023/04/maskrcnn-vs-unet-comparison-of-two-image-segmentation-methods.html [Accessed 13 Sep. 2023].
- d’Angelo, Emmanuel. “Why U-Net?” Computers Don’t See (Yet), 5 Mar. 2018, www.computersdontsee.net/post/why-u-net/.