Convolutional Neural Network: Analysis and its significance in Image Segmentation

URL:https://medium.com/@sivaranjani.kumar/convolutional-neural-network-analysis-and-its-significance-in-image-segmentation-8b8f9cdb6d0f?sk=e18b1d8535586963850248d72daeaa08

CNN is the most widely used deep learning Neural Network for image processing. It has made some impressive achievements for the past few years in many fields, not just limited to computer vision and natural language processing. We must have heard lot about the applications of CNN in different scenarios, but not from the general perspective. Here, I aim to provide some novel ideas which are proposed recently about CNN and also its significance in image segmentation.

So, where do we start. First, let's talk history for a bit and how it evolved so fast over the past few years and then we dig into analyzing it and talk about applications and prospects of CNN in this fast-growing field. Second, we will continue to explore about its consequences in image segmentation process.

Introduction: Convolutional Neural Network in computer vision has enabled people to accomplish what had been considered impossible in the past few centuries, such as face recognition, autonomous vehicles, self-service supermarket, and intelligent medical treatment. CNN in general is emerged from Artificial Neural Network (ANN).

History: In 1943, a mathematician named McCulloh and pitts proposed the first mathematical model of neurons called as MP (Multi-layer perceptron) model. The concept of CNN was not a new one. It all started from the discovery of Hubel and Wiesel which explained that there are simple and complex neurons in the primary visual cortex and the visual processing always starts with simple structures such as oriented edges. The term '*Convolution*' was first used by LeCun et al, he constructed a convolutional neural network for a handwritten zip code recognition which is the original version of LeNet. In 2012, Alex et al achieved the best classification result using deep CNN in the ImageNet Large Scale Visual Recognition Challenge (LSVRC). Some of the classic CNN-based models are LeNet-5, AlexNet to MobileNet v3 and GhostNet.

Convolutional neural network is a kind of feedforward neural network that is able to extract features from image data with convolution structures. The architecture of CNN is inspired by visual perception. We also have activation functions in CNN which are used to simulate the neurons when electrical signals exceeding a certain threshold is transmitted. Loss functions and optimizers are quite a thing people invented to teach the whole CNN system to learn.

Difference between CNN and ANN:

- Each neuron is no longer connected to all neurons of the previous layer.
- Reduces the number of parameters and speed up convergence.
- A group of connections can share the same weights.
- Down-sampling dimensionality reduction.

In order to build a CNN model, four components are typically needed.

- **Padding:** When setting a convolution kernel with a certain size, we will lose information in the border. Hence, padding is required to enlarge the input size with zeros.

- **Stride:** To control the density of convolving, stride is employed. Larger the stride, lower the density.
- **Feature maps:** Convolution is a pivotal step for feature extraction. The outputs of convolution can be called as feature maps. This map consists of large number of features which is prone to cause overfitting. So, to solve this we will use a technique called pooling.
- **Pooling:** It is used to obviate redundancy. We have two types one is max pooling and average pooling.

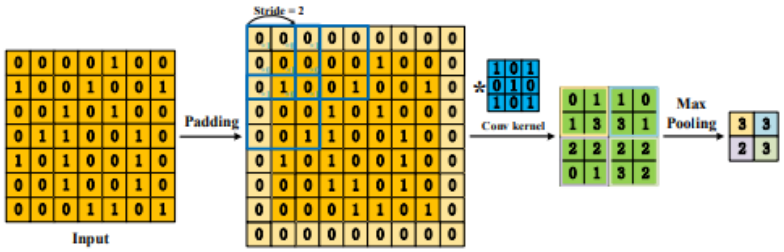
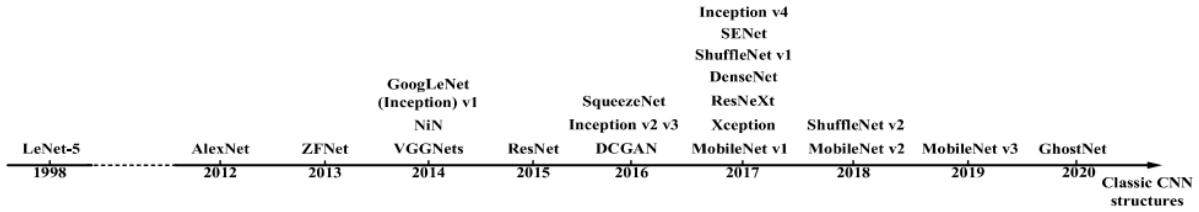


Fig. 1. Procedure of a two-dimensional CNN

There are some pre-trained CNN models that can be used for processing image data, instead of building our own model from scratch which are time consuming. AlexNet is pre trained model which was proposed in 2012, after that researchers have invented a variety of CNN models—deeper, wider, and lighter which gives amazing results. Below is an image for evolution of CNN models and these are called as state-of-the-art models,



CNN is able to harness a massive amount of data to achieve a promising result. Hence, there are lots of applications that come up. It can be used not only in the processing of two-dimensional images but also in one-dimensional and multi-dimensional scenarios. Some of the widely used applications are

- Image classification
- Object detection
- Image segmentation
- Face recognition

Now let's look into one of the most fast-growing application of CNN.

Image Segmentation: As humans, if given a picture of any object we will be able to identify it. For instance, when we want to cross the road at intersection, we typically look left and right and make our decision. It's because our brain is able to analyze a kind of vehicle that comes, or it could be any object in a given scenario. The question is, can machines do that??? The answer was 'no' till few years back. But with the advancements in computer vision, things have changed too fast and too soon. Below is an example of segmented image verses natural image.

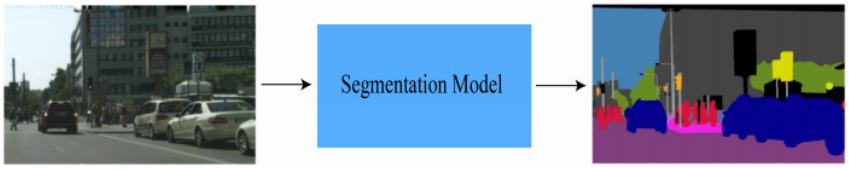
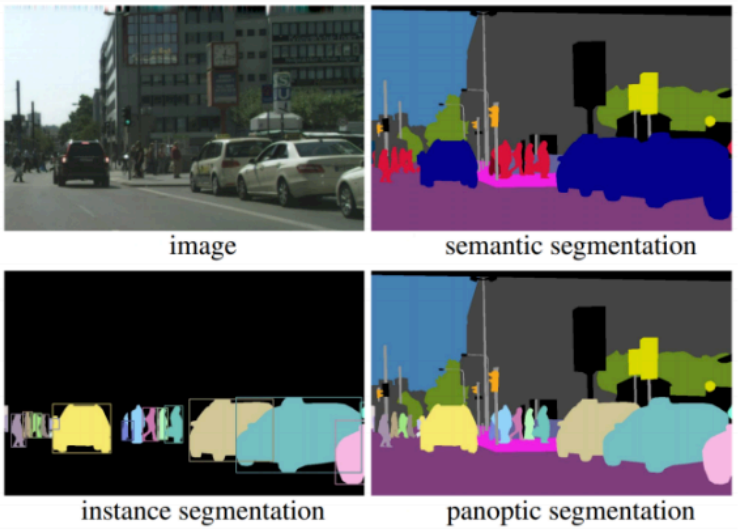


Image segmentation is basically the ability of computer vision model to detect objects and be able to determine its shapes and can also predict the direction the objects that can move. It builds upon the idea of object detection. Unlike classification and object detection, segmenting an image is typically a low-level or pixel level vision task as the spatial information of an image is very important for segmenting different regions semantically. Image segmentation is of two types: semantic segmentation, instance segmentation and the combination of these two can be known as panoptic segmentation.



Ok, so why we need CNN to do image segmentation and why not traditional machine learning algorithm?

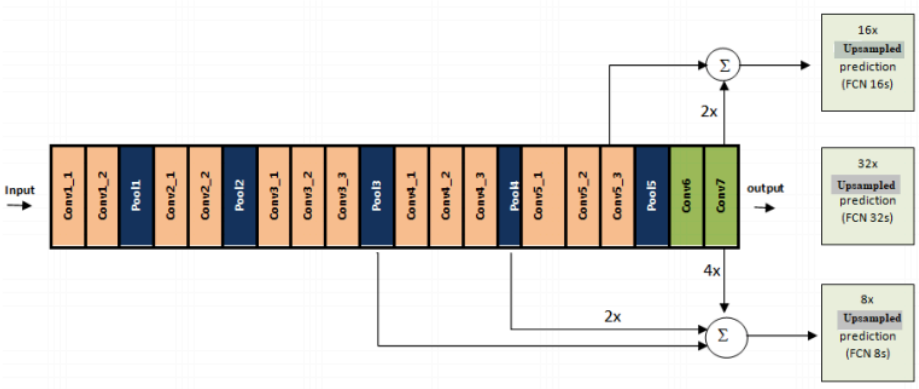
Image segmentation task is not a new field in computer vision. There have been many researches on addressing this task using traditional machine learning algorithms. First of all, feature engineering needs domain expertise and Deep learning does not need traditional handcrafted feature engineering techniques. On the other hand, deep learning has that capability to adapt itself according to the predicted result which cannot be achieved using traditional machine learning algorithm.

There are different ways to extract *segmented images*,

- Multi-scale CNN have been used for scene labeling and achieve state-of-the-art results in the Sift flow.
- R-CNN used selective search algorithm to extract region proposals first and then applied upon each proposal.

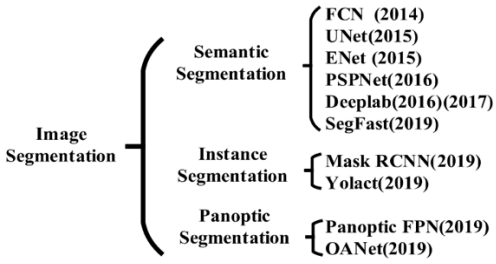
- R-CNN achieved record result over second order pooling which was a leading hand-engineered semantic segmentation system.

Among different CNN based semantic segmentation models, Fully Convolutional Network (FCN) gained the maximum attention and an FCN based semantic segmentation model trend has emerged. The base model which achieved state of the art result is VGG16. FCN has used only local connections which results in ambiguity and so contextual features are used to enhance the model. They used VGG16, GoogleNet and pre-trained data on ILSVRC data as base model and transferred these to dense FCN. This process produces FCN-AlexNet, FCN-VGG16, FCN-GoogleNet and achieved highest mean on validation dataset. Below is the image of architecture of FCN32, FCN-16, FCN-8



Unlike semantic segmentation, *instance segmentation* masks each instance of an object contained in an image independently. In object detection, researchers use the bounding box to detect each object instance of an image with a label for classification. Instance segmentation put this task one step forward and put a segmentation mask for each instance.

In *Panoptic segmentation*, we need to associate all the pixels in the image with a semantic label for classification and also identify the instances of a particular class. The output of a panoptic segmentation model will contain two channels: one for pixel's label (semantic segmentation) and another for predicting each pixel instance (instance segmentation). The image segmentation task completed by CNN is shown below,



On the other hand, fine-tuning of hyper-parameters, data pre-processing methods, choice of the loss function and optimization function, etc. are also play an important role in the success of a model that we choose.

Conclusion:

In this article we have discussed about an overall view of convolutional neural network, its history, analysis, applications and prospects. We have also talked about image segmentation using CNN. The advantages of convolutional neural networks, such as local connection, weight sharing, and down-sampling dimensionality reduction, have been widely deployed in both research and industry projects. Even though convolutions possess many benefits and have been widely used, we reckon that it can be refined further in terms of model size, security, and easy hyperparameters selection.

References:

[1] https://arxiv.org/pdf/2001.04074.pdf , Zewen Li, Wenjie Yang, Shouheng Peng, Fan Liu, Member, IEEE, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects"

[2] https://arxiv.org/pdf/2004.02806.pdf , Farhana Sultanaa, Abu Sufiana, Paramartha Duttab, Department of Computer Science, University of Gour Banga, India. "Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey"

[3] S. Rehman, H. Ajmal, U. Farooq, Q. U. Ain, and A. Hassan, "Convolutional neural network-based image segmentation: a review."

[4] Q. Liu, N. Zhang, W. Yang, S. Wang, Z. Cui, X. Chen, and L. Chen, "A Review of Image Recognition with Deep Convolutional Neural Network."