The Computer Vision Training Course MIPT

Comparison of SIFT and CNN methods

The report is prepared by Aleksandr Arkhipov

Abstract

This is the report based on the paper “SIFT Meets CNN: A Decade Survey of Instance Retrieval” by Liang Zheng, Yi Yang, and Qi Tian. The information that is given in this article is very important because Convolutional Neural Network is the basic algorithm in computer vision. This paper describes an image retrieval problem and modern algorithms that solve this problem. I describe each algorithm separately and compare them below.

Introduction

In this paper, CNN and SIFT methods are compared. But there is a hybrid method that contains something from CNN and SIFT. CNN is the latest algorithm of computer vision. It is the most effective algorithm for the vast majority of tasks. But there are several tasks that solved more efficiently using the algorithm SIFT.

Convolutional Neural Network

This is the Deep Learning algorithm that is based on convolutions. Convolution is componentwise multiplication kernel tensor by each area of the picture tensor. Componentwise multiplication is also named as the Hadamard product. This operation can be modified with using different kernels, using different stride etc. After convolution artificial neural network architecture can contain a pooling layer. It is a layer that compresses the feature map by writing the average or maximum value of several cells instead of these cells. This technique helps to optimize the computational expenses of the neural network. It is very useful because the convolutional neural network is very computationally-expensive. Besides this, there are nonlinear layers that allow the neural network to learn. Even though neural networks are computationally expensive, there is a high accuracy of this method.

Scale-invariant feature transform

The SIFT algorithm is published by David Lowe in 1999. Matching features across different images in a common problem in computer vision. When all images are similar in nature

(same scale, orientation, etc) simple corner detectors can work. But when you have images of different scales and rotations, you need to use the Scale Invariant Feature Transform. SIFT isn't just scale-invariant. You can change the following, and still get good results: Scale, Rotation, Illumination, Viewpoint. SIFT is quite an involved algorithm. It has a lot going on and can become confusing, So I have split up the entire algorithm into multiple parts. Here's an outline of what happens in SIFT: Constructing a scale space, LoG Approximation(The Laplacian of Gaussian is great for finding interesting points (or key points) in an image. But it's computationally expensive. So we cheat and approximate it using the representation created earlier), finding key points, get rid of bad key points, assigning an orientation to the keypoints, Generate SIFT features.

Comparison

SIFT and CNN are innovative algorithms. There are a lot of differences between them. Nowadays computing power is growing, but mobile devices haven’t enough power for modern CNN. It is an actual problem for Huawei. SIFT is needed less processing power, memory for storing features for each image. SIFT has a simpler design and fewer parameters to set compared to CNN. CNN is more relevant for classification and categorization tasks, it has very good generalization abilities. CNN has very good generalization.

Nowadays CNN is standard in computer vision in the vast majority of problems.

Conclusion

There are a lot of useful subtleties in the paper “SIFT Meets CNN: A Decade Survey of Instance Retrieval”. I figured out the intricacies of the principles of operation of CNN и SIFT. I read a lot of additional information to make out in this paper. In my opinion, these algorithms are not mutually exclusive. The choice of algorithms should depend on the task.