Convolutional Neural Networks for Image Classification

Laura Anger
Institute for Media and Imaging Technology
TH Köln – University of Applied Sciences Cologne
laura.anger@smail.th-koeln.de

ABSTRACT

CONTENT OF ABSTRACT Seitenzahlen fehlen!!!

Keywords

Convolutional Neural Networks ; Image Classification

1. INTRODUCTION

Image classification is important in many different areas. For Example: automatically Traffic sign recognition, facerecognition, MOTIVATION: 1) Bildklassifizierung ist in sehr vielen Bereichen wichtig: pornographische Inhalte filtern, google suche, Straßenschilderkennung, Gesichtserkennung, medizinische Erkennung von Zellen.

3) Bildklassifizierung kann unter umständen besser als durch Menschen geschehen ([1] im Abstract)

2) Schwierigkeit bei realen Bildern: difficult challenges due to real-world variabilities such as viewpoint variations, lighting condi- tions (saturations, low-contrast), motion-blur, occlusions, sun glare, physical damage, colors fading, graffiti, stickers and an input resolution as low as 15x15 ([3] in Introduction)

4) Analogie von NN zu menschlichem Körper (Viele Eingänge aber nur einen Ausgang; Synapsen verbinden Ausgang mit den Eingängen anderer Neuronen; Zellkern berechnet aus dem Eingangsignal das Ausgangsgnal)

GLIEDERUNG DES PAPERS: - Einführung in die Thematik und Benennung der Problematik von normalen nn - Was ist besser an cnn und kurze Funktionsweise - Cnn paper müssen sortiert dargestellt werden

ÜBERSICHT ENTWICKLUNG NEURONALE NETZE: - Neocognitron von Kunihiko Fukushima (1980): hierarchisches mehrschichtiges künstliches neuronales Netz, welches zum Beispiel bei der Erkennung handschriftlicher Zeichen und bei anderen Mustererkennungs-Aufgaben zum Einsatz kommt ([2])

2. GRUNDLAGEN NEURONALE NETZE

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Motivation

A regular neural network or multi-layer perceptron (MLP) is made up of input/out layers and a series of hidden layers.

There are three main problems if we use MLP for image recognition tasks:

MLPs do not scale well The generalization performance of the MLP will be eclipsed by its excessive free parameters, i.e., weights. For example, each image from the ImageNet has a size of 256x256x3. That means each neuron in the hidden layer will have 256x256x3=196608 connections with each pixel in the input image. The total number of weights would scale up quickly if we want to add more neurons or hidden layers. The enormous number of weights produced by full connectivity would quickly lead to overfitting.

MLPs ignore pixel correlation It is an important property of images that nearby pixels are more strongly correlated than distant pixels. This property is not taken advantage of by MLP due to the full connectivity. This property suggests local connectivity is preferred.

MLPs are not robust to image transformations It is expected that the recognition algorithm should be robust to transformations applied to the input image. For example, for handwritten digit recognition, a particular digit should be assigned the same value regardless of its position within the image or of its size.

For MLPs, any subtle change in scale or position from the input layer would produce significant changes in following layers. It is therefore advantageous to incorporate into the network some invariance to common changes that could occur in images, e.g., translation, scale, etc.

Accordingly image recognition tasks require a new type of neural network: cnn.

Unsupervised learning methods applied to patches of natural images tend to produce localized filters that resemble off- center-on-surround filters, orientation-sensitive bar detectors, Gabor filters. These findings as well as exper- imental studies of the visual cortex justify the use of such filters in the so-called standard model for object recognition, whose filters are fixed, in contrast to those of Convolutional Neural Networks (CNNs), whose weights (filters) are learned in a supervised way through back-propagation (BP).([1] Muss noch feiner zitiert werden)

- Multilayer-Perceptron

2.1 Schichten

- Eingangsschicht: verteilt nur - 1. Verdeckte Schicht: erzeugt Linien - 2. Verdeckte Schicht: erzeugt Fläche (Polygone ohne Löcher) - Ausgangsschicht: erzeugt beliebige Flächen (konvex, konkav, mit Löchern, beliebig)

2.2 CNN

Convolutional Neural Networks

To utilize the prior knowledge on image recognition, CNNs incorporate the following concepts into the design:

Local connectivity Local connectivity is a solution to the over-parameterization problem. The advantage of using local features and the derived high order features has been demonstrated in classical work in visual recognition. This knowledge can be easily built into the network, by forcing the neurons to receive only local information.

Figure 2. A neural network with local connectivity [5]

This notion is illustrated in Fig. 2 where layer m-1 can be considered as input images, and layer m is the hidden layer. It can be seen from the figure that each neuron in the hidden layer connects to only 3 adjacent input pixels in the image. Local connectivity also takes advantage of the pixel correlation from input image, as each neuron in the hidden layer only cares about nearby pixels in a neighborhood.

Weight sharing One problem of image recognition tasks is that images that contain the same semantic object could have the object at various locations on the image. As a result, classical work in visual recognition detects local features at various location on the input image. Weight sharing is a solution that stimulates the approach of applying local feature detectors at different positions of the image. It is also a solution to improve network robustness against image transformations.

Figure 3. A neural network with weight sharing [5]

As depicted in Figure 3, the three neurons in hidden layer m share the same weight. Weights of the same colour are constrained to be identical. Note that the hidden layer m could contain multiple planes of neurons that share the same weights. These planes are referred to as "feature maps". Neurons in every feature map are constrained to have identical weights, which is equivalent to performing the same operation in different parts of the image.

Since each neuron is only connected to part of an image, and keeps the information in the corresponding feature map. This behavior is equivalent to a convolution with a small size kernel, followed by a squashing function [2]. This is also how CNN was named.

Sub sampling As an object can appear at various locations on the input image, the precise location of a local feature is not important to the classification. That means the network can afford to lose the exact position information. However an approximate position information should be retained so that the next layer can possibly detect higher- order feature. As a result the feature maps need not be as the same size of the input image. Instead, a feature map can be built by sub sampling the input image. Sub sampling provides a form of invariance to distortions and translations.

Piecing these concepts together results in LeNet, the first CNN, which is illustrated in Fig 4 [5]. Although Fig 4 does not precisely reflect the architecture presented in [2], it can still be used as an meaningful illustration.

2.3 Unterschiede zwischen max-pooling layer und sub-sampling layer

max-pooling layer: [1]

3. REFERENCES

[1] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber. A committee of neural networks for traffic sign classification. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1918–1921, July 2011.

[2] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.

[3] P. Sermanet and Y. LeCun. Traffic sign recognition with multi-scale convolutional networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2809–2813, July 2011.