

Convolutional Neural Network

By Bastian Berle & Ron Holzapfel

Motivation & Vergleich mit traditionellen NN

Bei Use Cases, wie der Klassifizierung von Objekten auf Bildern, stoßen ANNs an ihre Grenzen. Da diese fully connected (dense) sind, steigt der Trainingsaufwand bei der Bildverarbeitung immens. Die Faltungsschichten im CNN reduzieren hingegen sowohl Inputparameter ($height \cdot width \cdot layer$ - (meist 3 Layer für RGB)), wodurch sowohl die Speicheranforderungen, als auch die Trainingszeiten sinken. Zusätzlich wird die räumliche Struktur berücksichtigt.

Sparse Interactions

CNN

- Kern kleiner als Input
- Weniger Operationen zur Berechnung des Outputs
- Reduziert Speicherbedarf
- Verbesserungen der Effizienz

NN

- Matrixmultiplikation
- Gewicht beschreibt Interaktion zwischen Output und Input
- Jeder Output interagiert mit jeder Input Einheit

Probleme bei der Verwendung von Bildern

Damit CNNs Objekte innerhalb eines jeden Bildes erkennen können, müssen folgende Unterschiede beachtet werden:

- Viewpoint Variation,
- Scale Variation,
- Illumination Conditions,
- Deformation,
- Background Clutter,
- Occlusion, and
- Intra-class Variation.

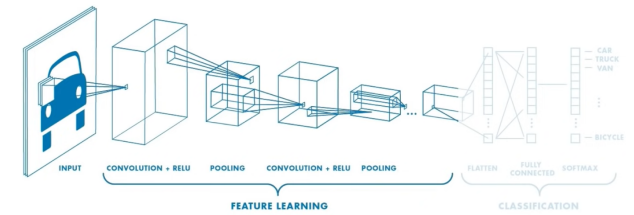
Anwendung

Semantic Segmentation nutzt Fully Convolutional Networks. Dabei besteht das Netzwerk vollständig aus Convolutional Layers. Ein Teil wird zum Down-sampling des Bildes genutzt und der zweite zum abstrahierten Upsampling. Im Downsampling werden die Regionen bzw. Pixel ermittelt, die zueinander gehören. Im Upsampling werden diese durch einheitliche Farben markiert.

Probabilistic Robotics ist ein Gebiet in der Robotik, das sich mit der Wahrnehmung und Steuerung von Robotern unter Unsicherheit befasst. Probabilistische Ansätze werden u.a. zur Zustandsschätzung und Lokalisierung von Robotern in der mobilen Robotik eingesetzt. Beim „Schätzen“ einer Größe aus Sensordaten berechnet der probabilistische Ansatz eine Wahrscheinlichkeitsverteilung darüber, was in der Welt der Fall sein könnte, anstatt nur eine einzige „beste Schätzung“ zu erstellen. Mit weiteren Sensordaten, wie z.B. Google Maps, kann mit den aus probabilistische Ansätze ermittelten Möglichkeiten eine deterministische Auswahl getroffen werden. Ein Roboter, der probabilistische Methoden anwendet, kann sich daher von Fehlern erholen, mit Mehrdeutigkeiten umgehen und Sensordaten auf konsistente Weise integrieren.

Object Detection nutzt ein CNN, um ein oder mehrere Objekte innerhalb eines Bildes zu erkennen und räumlich zu umrahmen. Ein Ansatz zur Bestimmung der Regionen und Typen der Objekte ist R-CNN, wobei das R für Region based steht. Hierbei werden Regionen unterschiedlicher Größe gefunden, die Objekte enthalten. Anschließend werden die Regionen als Bilder extrahiert und auf eine einheitliche Größe verzogen. Die neuen Bilder dienen als Input für ein CNN. Eine Erweiterung ist Faster R-CNN, welche die Notwendigkeit, das CNN für jedes Teilbild einzeln zu durchlaufen, eliminiert und somit schneller ist.

Architektur



Convolution Erlernen und Anwenden von verschiedenen Filtern, um kleinere Feature Maps zu generieren. Die Convolution-Funktion ist durch $\sum_{i=1}^4 \sum_{j=1}^4 w_{ij} x_{i+p,j+q} + b$ gegeben, wobei 4×4 die Größe des Filters ist und p sowie q die Neuronen darstellen. w sind die zu erlernenden Gewichtungen und b der Bias.

Non-linearity Anschließend wird eine nicht-lineare Funktion angewendet, wie die ReLU.

Pooling Das Pooling beschreibt eine Verkleinerung der Feature Map durch Anwendung weiterer Funktionen. Häufig findet das MaxPooling, welches nur den maximalen Wert einer gegebenen Teilmatrix übernimmt, Anwendung.

