

Aufbau von Convolutional Layers

Erarbeitet von
Dr. Ludmila Himmelspach

Lernziele	1
Inhalt	2
Einstieg.....	2
Bestandteile der Convolutional Layer.....	2
Bestimmung der Filtermasken in CNNs	4
Take-Home Message	7
Quellen	8
Weiterführendes Material.....	8
Disclaimer	8

Lernziele

- Du kannst die wichtigsten Bestandteile der Convolutional Layers in CNNs erklären
- Du lernst, wie die Verbindungen zwischen den Neuronen einzelner Schichten in CNNs realisiert werden
- Du lernst, welche Parameter innerhalb der Convolutional Layers in einem Trainingsprozess optimiert werden

Inhalt

Einstieg

Wir haben bereits darüber gesprochen, dass ein Convolutional Neural Network aus aufeinanderfolgenden Convolutional Layers besteht. Die Neuronen innerhalb eines Convolutional Layers sind ausschließlich mit Neuronen des darunterliegenden Convolutional Layers verbunden, die innerhalb eines sogenannten Wahrnehmungsfeldes liegen. In diesem Video lernst Du, wie die einzelnen Convolutional Layer aufgebaut sind und wie die Verbindungen zwischen den Neuronen einzelner Schichten in CNNs realisiert werden. Außerdem erfährst Du, welche Parameter innerhalb der Convolutional Layer in einem Trainingsprozess optimiert werden.

Bestandteile der Convolutional Layer

Hubel und Wiesel stellten in ihren Experimenten fest, dass im visuellen Cortex einige Neuronen mit gleichen Wahrnehmungsfeldern auf Linien unterschiedlicher Ausrichtung reagieren.

Quelle [1]

Dieser Aspekt wurde in Convolutional Neural Networks durch Anwendung von Filtern realisiert. Jedes Convolutional Layer innerhalb der CNNs enthält mehrere Filter, die durch ihre Masken, auch *Kernel* genannt, definiert werden.

Quelle [2]

Im ersten Convolutional Layer werden Filtermasken auf die Nachbarpixel innerhalb eines kleinen Rechtecks, häufig Quadrats, im Eingabebild angewendet. Also entspricht dieser Ausschnitt des Eingabebildes dem *lokalen Wahrnehmungsfeld* des *Neurons* im Convolutional Layer. Die Verbindung zwischen dem Neuron und seinem Wahrnehmungsfeld wird durch die Berechnung der Pixelwerte innerhalb des Wahrnehmungsfeldes im Eingabebild und den Gewichten der Filtermatrix realisiert. In Folge dieser Filteroperation entsteht ein Ausgabebild. In Convolutional Neural Networks werden die Ausgabebilder als *Feature Maps* bezeichnet und bilden zusammen mit den Filtermasken Bestandteile eines jeden Convolutional Layers.

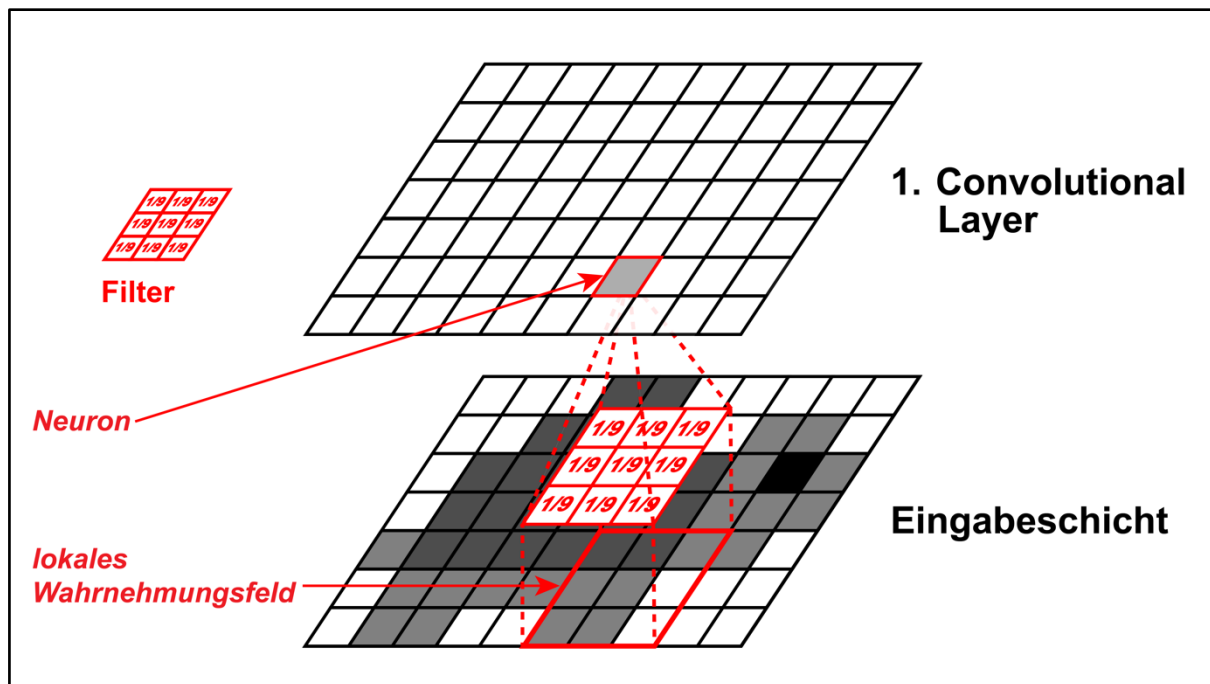


Abbildung 1: Verbindung zwischen einem Neuron und seinem Wahrnehmungsfeld in CNNs.

Da durch die Anwendung eines jeden Filters eine Feature Map entsteht, hängt die Anzahl der Feature Maps von der Filteranzahl ab, die man übrigens bei der Implementierung für jede Schicht selbst festlegen kann. Des Weiteren kannst Du auch die Größe des Wahrnehmungsfeldes bzw. die Größe der Filtermaske selbst bestimmen. Diese gilt aber für alle Filter innerhalb eines Convolutional Layers. In unseren vorigen Beispielen haben wir mit 3 x 3-großen Filtermasken gearbeitet. Man kann aber auch Filter mit 5 x 5-, 7 x 7- oder auch mit größeren Masken anwenden. Dabei muss man beachten, dass mit der Vergrößerung der Filtermasken sich auch die Anzahl der Optimierungsparameter im CNN vergrößert. Später in diesem Video siehst Du auch, warum.

Zuerst aber rufen wir uns nochmal in Erinnerung, wie die Mustererkennung im menschlichen visuellen Cortex stattfindet. Hubel und Wiesel fanden heraus, dass die Neuronen der höheren Hierarchieebenen auf komplexere Muster reagieren, die Kombinationen einfacher Muster sind.

Quelle [3]

Wenn wir annehmen, dass Feature Maps des ersten Convolutional Layers nur einfache Muster enthalten, müssen wir diese miteinander kombinieren, um auf komplexere Muster zu kommen. Das wird erreicht, indem die Wahrnehmungsfelder bzw. die Filtermasken in höheren Convolutional Layers um eine Achse bzw. eine Dimension erweitert werden. Diese Dimension erstreckt sich über alle Feature Maps des darunterliegenden Convolutional Layers.

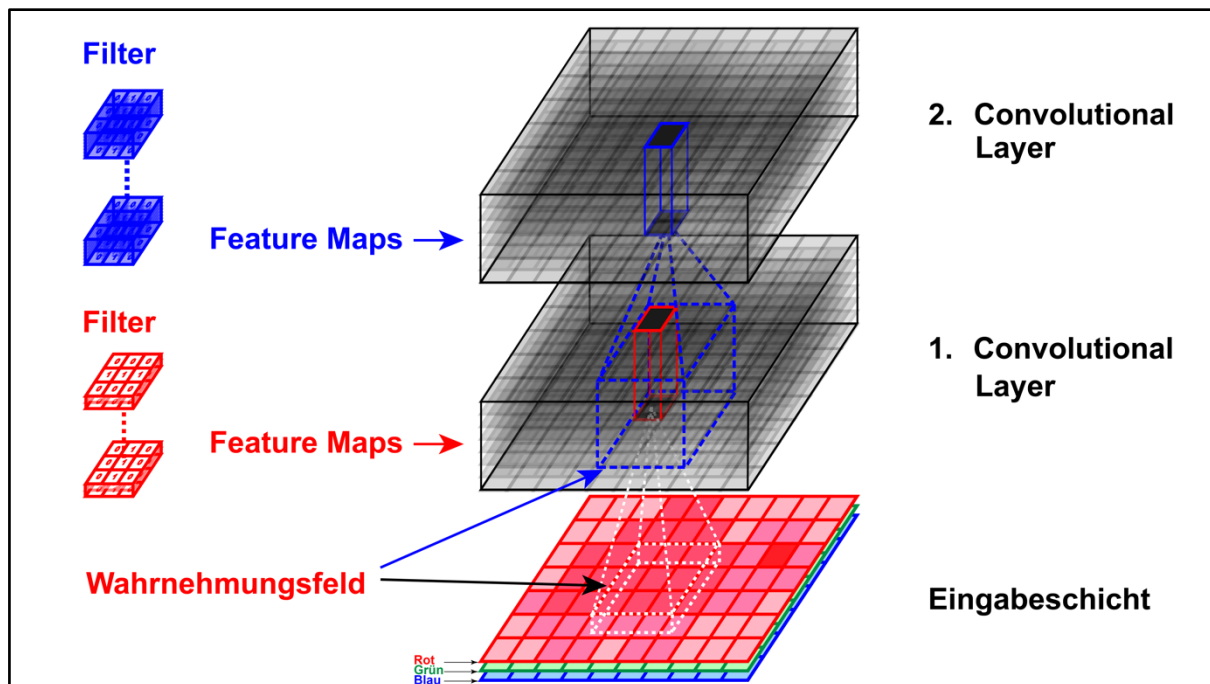


Abbildung 2: Aufbau der Convolutional Layer.

In CNNs, die Farbbilder verarbeiten, sind übrigens Filtermasken des ersten Convolutional Layers schon dreidimensional, weil Farbbilder aus drei Schichten bestehen, die den Farbkanälen Rot, Grün und Blau entsprechen.

Aufgrund der Vielzahl der möglichen Kombinationen der Feature Maps in einem Convolutional Layer, muss die Filteranzahl und somit die Anzahl der Feature Maps im nächsten Convolutional Layer erhöht werden. Um die Arbeitsspeicherkapazitäten nicht zu überschreiten, wird in den meisten CNN-Architekturen die Anzahl der Filter in jedem oder in jedem zweiten Convolutional Layer verdoppelt. Die Berechnung aller möglichen Kombinationen der Feature Maps macht außerdem keinen Sinn, weil nicht jedes mögliche Muster, das aus den Eingabebildern extrahiert wird, zur Lösung der gestellten Aufgabe beiträgt.

Bestimmung der Filtermasken in CNNs

Um die Berechnung der Feature Maps besser verstehen zu können, haben wir in den vorigen Beispielen Filtermasken als bekannt vorausgesetzt. Wenn man die Filtermasken in einem CNN tatsächlich festlegen würde, so würden dem Netz nur die Muster bei der Lösung der Aufgaben zur Verfügung stehen, die durch diese Filter aus den Bildern extrahiert werden. Schauen wir uns das Graustufenbild vom „Haus der Universität“ an.



Abbildung 3: Graustufenbild von "Haus der Universität"
(Originalbild: Mayr, I. Haus der Universität /©HHU).

Wenn wir darauf einen 3 x 3 Sobel-Filter zur Extraktion der horizontalen Kanten

$$F_h = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

anwenden, erhalten wir diese Feature Map:



Abbildung 4: Feature Map nach Anwendung des horizontalen Sobel-Filters auf "Haus der Universität"
(Originalbild: Mayr, I. Haus der Universität /©HHU).

Wenden wir einen 3 x 3 Sobel-Filter zur Extraktion der vertikalen Kanten

$$F_v = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix}$$

auf das gleiche Bild an, sieht die entsprechende Feature Map folgendermaßen aus:



Abbildung 5: Feature Map nach Anwendung des vertikalen Sobel-Filters auf "Haus der Universität"
(Originalbild: Mayr, I. Haus der Universität /©HHU).

Aus den beiden vorher berechneten Feature Maps können wir höchstens diese Feature Map erhalten:



Abbildung 6: Kombination der Feature Maps aus Abbildungen 2 und 3
(Originalbild: Mayr, I. Haus der Universität /©HHU).

Im Vergleich zum Originalbild enthält die zusammengesetzte Feature Map viel weniger Informationen. Anhand dieser Feature Map kann man zwar noch bestimmen, dass im Bild ein Gebäude zu sehen ist. Man kann sogar noch die einzelnen Buchstaben auf der Hausfassade gut erkennen. Aber wenn es um die Verzierungen der Fassade, besonders um die Skulpturen geht, wird es schwieriger. Da braucht man weitere Filter, um die entsprechenden Muster aus dem Originalbild extrahieren zu können. Wie aber werden diese bestimmt?

In Convolutional Neural Networks werden die Gewichte der Filtermasken für jede einzelne Anwendung trainiert. Das heißt, ein CNN muss selbst herausfinden, welche Bildeigenschaften bzw. Muster bei der Lösung des gestellten Problems ausschlaggebend sind. Dementsprechend werden die Gewichte der Filtermasken, die wir hier als Variablen w_{ij} darstellen, in jedem Convolutional Layer anhand der Trainingsbilder und ihrer Klassenzuordnungen in einem Trainingsprozess optimiert.

w_{11}	w_{12}	w_{13}
w_{21}	w_{22}	w_{23}
w_{31}	w_{32}	w_{33}

Abbildung 7: Allgemeine 3x3-Filtermaske.

Wenn unsere Aufgabe darin besteht, Lungenentzündung in Röntgenbildern des Brustkorbs zu bestimmen, bedeutet es für ein Convolutional Neural Network die Gewichte der Filtermasken in jedem Convolutional Layer so zu lernen, dass nach der Verarbeitung eines jeden Trainingsbildes mit diesen Filtern seine Klassenzugehörigkeit richtig bestimmt wird.

Die Gewichte der Filtermasken in CNNs sind vergleichbar mit dem menschlichen visuellen Gedächtnis, das wir alle von Geburt an trainieren und das sich von Mensch zu Mensch oder von Expert*in zu Expert*in unterscheidet, abhängig davon, auf welche Art der visuellen Aufgaben man sich spezialisiert hat.

Neben den Gewichten der Filtermasken wird für jede Feature Map zusätzlich noch der Bias-Term trainiert. Diesen Term kann man sich als eine Art Regler vorstellen, der bestimmt, wie stark eine Feature Map in die Berechnung der Feature Maps höherer Convolutional Layer eingeht. Die Bias-Terme sind in Convolutional Neural Networks nicht unbedingt erforderlich, bieten aber den Netzen mehr Flexibilität und werden deswegen oft verwendet.

Take-Home Message

In diesem Video hast Du gelernt, dass jeder Convolutional Layer innerhalb eines CNNs hauptsächlich aus mehreren Filtern und den entsprechenden Feature Maps besteht. Du hast erfahren, wie die Verbindungen zwischen den Neuronen einzelner Schichten realisiert

werden. Außerdem weißt Du jetzt, welche Parameter innerhalb der Convolutional Layer während eines Trainingsprozesses optimiert werden.

Quellen

- Quelle [1] Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3), 574.
- Quelle [2] Géron, A. (2023). *Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme* (aktualisierte und erweiterte Auflage.). O'Reilly Verlag.
- Quelle [3] Bellebaum, C., Thoma, P., Daum, I. (2011). *Neuropsychologie*. Germany: VS Verlag für Sozialwissenschaften.

Weiterführendes Material

Chollet, F. (2018). *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG.

Disclaimer

Transkript zu dem Video „Bildklassifikation und Bildsegmentierung: Aufbau von Convolutional Layers“, Dr. Ludmila Himmelsbach.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.