

6. Übungsblatt

Aufgabe 24 Convolution

In der Bildverarbeitung werden Graustufenbilder oft als zweiparametrische Funktion $f(x, y)$ dargestellt und als Matrix beschrieben. Dabei sind x und y die Koordinaten eines Bildpunktes und der Funktionswert / Zellwert gibt die Helligkeit an.

Das folgende Bild ist in Graustufen gegeben:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 10 & 10 & 10 & 0 \\ 0 & 10 & 10 & 10 & 10 & 0 \\ 0 & 10 & 10 & 0 & 0 & 0 \\ 0 & 10 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Convolutional Neural Networks sind in der Lage effizient Bilddaten unabhängig von Rotation und Position enthaltener Objekte zu verarbeiten. Hierfür werden Kernel verwendet, welche jeweils mit Ausschnitten der Matrix verrechnet werden. Die auf diese Weise berechneten Convolved Features können zur Erkennung von Objekten verwendet werden.

Der Sobel-Operator dient der Kantenerkennung in Bilddaten. Nimmt man an, dass bei einer Kante die Helligkeit stark wechselt, so kann man Kanten an den Extremwerten der ersten Ableitung der Funktion f erkennen. Gegeben seien die Operatoren S_x und S_y .

$$S_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

- a) Berechnen Sie die Features G_x und G_y durch ausschnittsweise Multiplikation der Operatoren S_x und S_y mit der Matrix A . Verwenden Sie Zero Padding von 2.

$$G_x = S_x * A$$

$$G_y = S_y * A$$

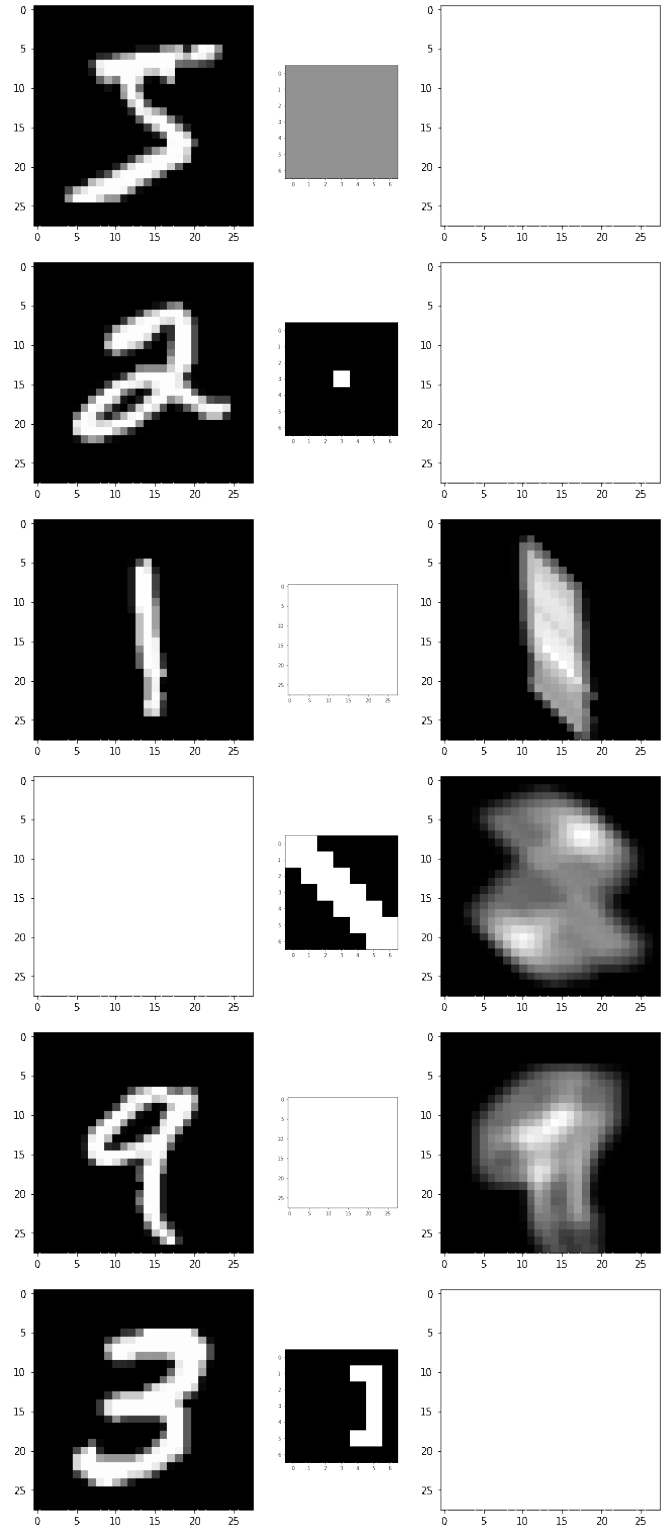
- b) Aus den richtungsabhängigen Matrizen soll jetzt eine richtungsunabhängige Matrix G erstellt werden. Dafür werden **die Einträge** von G_x und G_y **jeweils** quadriert, dann aufsummiert und aus dem Ergebnis die Wurzel gezogen.

$$G = \sqrt{G_x^2 + G_y^2}$$

Berechnen sie die Matrix G mit Hilfe der Teilergebnisse aus Aufgabenteil a) und beschreiben sie das Endergebnis!

Aufgabe 25 Convolution

Skizzieren Sie das fehlende Bild. Hierbei befindet sich in der linken Spalte das Ausgangsbild, in der Mitte der Kern und in der rechten Spalte die resultierende Feature Map.



Aufgabe 26 Receptive Field

Nehmen Sie an, dass ein CNN Bilder mit 32×32 Pixeln als Eingabe erhält. Bestimmen Sie, aus wie vielen Schichten von Konvolution mit einem Stride von 1 und einem 2×2 Kernel ein Netz bestehen muss, damit das ganze Bild von der letzten Schicht einbezogen werden kann.

Was bedeutet dies für das Klassifizieren von hochauflösenden Bildern ?

Aufgabe 27 Stride, Padding, Pooling

Die Ausgabedimension O eines beliebigen Convolution Layers lässt sich mit Hilfe der Formel

$$O = \frac{(W - F + 2P)}{S} + 1$$

berechnen, wobei W die Eingabedimension, F die Filtergröße, P das Zero Padding und S der Stride ist.

- Berechnen Sie die Ausgabedimension O für eine Eingabe der Dimension 32×32 , Filtern der Dimension 6×6 , Zero Padding von 2 und Stride von 2.
Wie ändert sich die Ausgabe, wenn wir die Filter auf 5×5 und den Stride auf 3 setzen ?
- Wie viele Reihen muss man durch Zero Padding an eine Eingabe beliebiger Größe anfügen, damit Ausgabe und Eingabe die selbe Dimension besitzen ?
- Wenden Sie jeweils Max pooling, Average pooling und L2-norm pooling mit einem Filter der Dimension 2×2 und Stride 2 auf die folgende Matrix an.
Warum werden Pooling Layer in CNNs eingesetzt ? Welche Vorteile bieten sie ? Warum wird meist nur Max pooling verwendet ?

$$A = \begin{bmatrix} 1 & 1 & 2 & 0 & 2 & 3 \\ 5 & 6 & 7 & 6 & 6 & 8 \\ 3 & 2 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 4 & 3 & 0 \\ 1 & 4 & 2 & 2 & 5 & 8 \\ 0 & 0 & 1 & 7 & 4 & 1 \end{bmatrix}$$