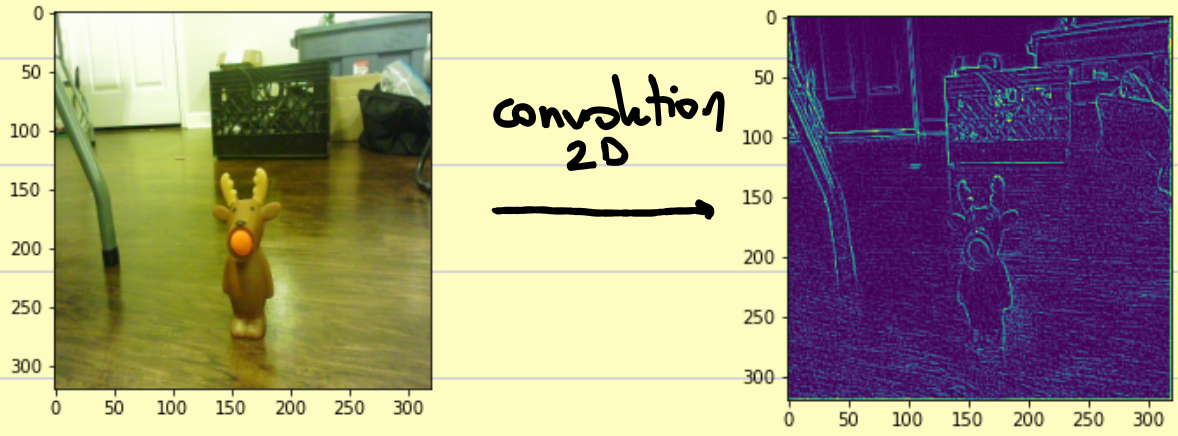
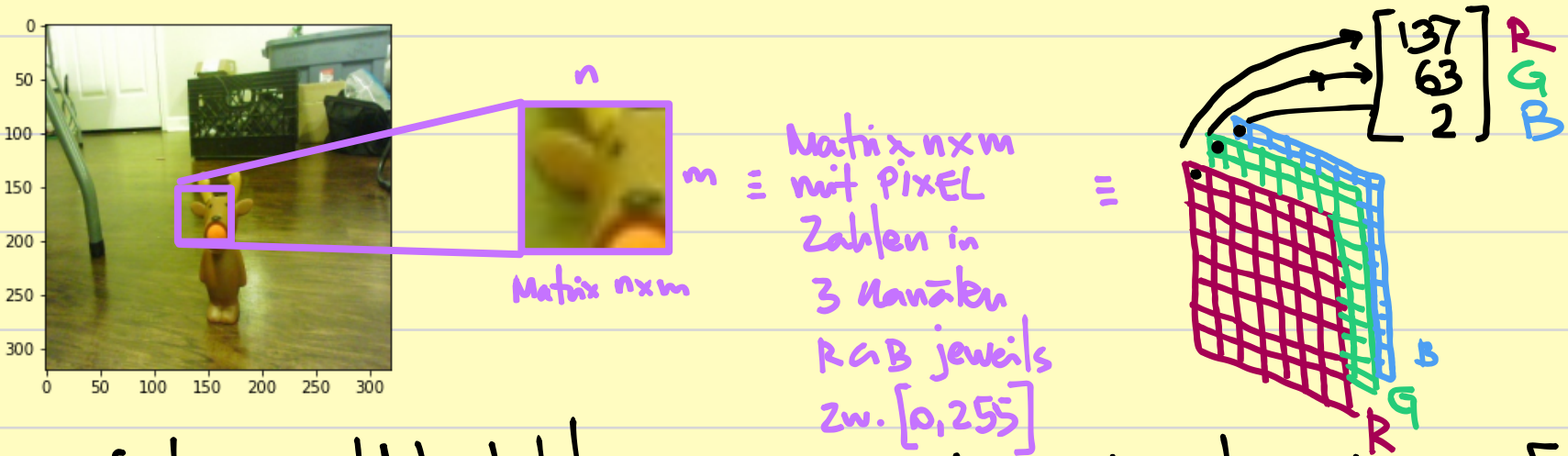
Deep learning by hand (Teil 2)

CONVOLUTION by hand

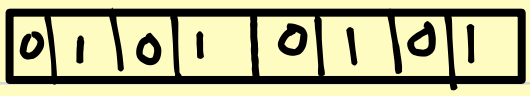


Background: Convolution is a technique used in image processing. Convolution is the process of summing each element of the input image with its local neighbours, weighted by the kernel (*)

(*) A kernel is in this case a matrix that shapes the weights of convolution.



Ein farbiges Bild besteht aus 3 Layer (R,G,B) mit Werten zw [0,255]



1 byte · 8 bits $\{0,1\}$ $2^8 \equiv 256$

The output of the convolution process will depend on the following:

CONV. LAYER OUTPUT SIZE

Input Volume size = $W_1 \times H_1 \times D_1$

with four hyperparameters:

1) $K \equiv$ number of filters
2) $F \equiv$ filter spatial extent
3) $S \equiv$ Stride
4) $P \equiv$ amount of zero padding

Output Volume size = $W_2 \times H_2 \times D_2$

$H_2 = W_2 = (W_1 - F + 2P)/S + 1$

$D_2 = K$

INPUT FILTER or KERNEL

Kernels can come in all shapes and sizes. These kernels make up filters, which are a parameter in convolutional layer. There are kernels for attention, for blurring, for edge recognition (this example), ...

W_0 $[:,:,0]$

1	0	-1
0	0	0
-1	0	1

R

This is an Edge recognition kernel because only the edges (corners) of the kernel are non-zero.

W_1 $[:,:,1]$

0	-1	0
-1	4	-1
0	-1	0

G

This is an Attention kernel because the middle is highlighted (x4).

B

$W_2[:,:,2]$

-1	-1	-1
-1	8	-1
-1	-1	-1

This is a **stronger** attention kernel than W_1. (x8)

Input Volume (Input image)

R[:,:,0]

1	1	0	2	0
2	0	3	3	0
0	2	2	3	1
1	1	2	1	0
0	1	2	2	0

G[:,:,1]

2	0	1	1	3
1	0	1	3	0
1	2	3	3	3
3	0	0	0	2
1	0	2	3	3

B[:,:,2]

0	0	1	2	0
2	3	2	0	3
3	1	0	2	0
1	0	1	2	2
2	2	3	0	1

Zero-Padding

Usually zero-padding is added to input images. In this example we will add a zero-padding P=1, which returns the following:

R[:,:,0]

0	0	0	0	0	0	0
0	1	1	0	2	0	0
0	2	0	3	3	0	0
0	0	2	2	3	1	0
0	1	1	2	1	0	0
0	0	1	2	2	0	0
0	0	0	0	0	0	0

G[:,:,1]

0	0	0	0	0	0	0
0	2	0	1	1	3	0
0	1	0	1	3	0	0
0	1	2	3	3	3	0
0	3	0	0	0	2	0
0	1	0	2	3	3	0
0	0	0	0	0	0	0

B

(idem)

7x7

Zero Padding P=1

Stride

Stride means how many pixels we jump after each convolution. In this case $S=2$.

Convolution

$R[:,:,0]$ $P=1$ $S=2$

7x7

0	0	0	0	0	0	0
0	1	1	0	2	0	0
0	2	0	3	3	0	0
0	0	2	2	3	1	0
0	1	1	2	1	0	0
0	0	1	2	2	0	0
0	0	0	0	0	0	0

$w_0[:,:,0]$

1	0	-1
0	0	0
-1	0	1

$0+7-5+1$ bias

$o[:,:,0]$

3

first convolution : $[0.1+0.0+0.-1]+[0.0+1.0+1.0]+$

$+[0.-1+2.0+0.1]=0$

$w_0[:,:,1]$

0	-1	0
-1	4	-1
0	-1	0

: $[0.0+0.-1+0.0]+$

$+[0.-1]+[2.4]+[0.-1]+$

$+[0.0+1.-1+0.0]=7$

$w_0[:,:,2]$

-1	-1	-1
-1	8	-1
-1	-1	-1

: $[0.-1+0.-1+0.-1]+$

$+[0.-1+0.8+0.-1]+$

$+[0.-1+2.-1+3.-1]=-5$

2. Convolution Stride = S = 2

3+2+1 **+1** bias

7×7 → S=2

0	0	0	0	0	0	0
0	1	1	0	2	0	0
0	2	0	3	3	0	0
0	0	2	2	3	1	0
0	1	1	2	1	0	0
0	0	1	2	2	0	0
0	0	0	0	0	0	0

$W_0[:,:,0]$

1	0	-1
0	0	0
-1	0	1

$O[:,:,0]$

3	7

$W_0[:,:,1]$

0	-1	0
-1	4	-1
0	-1	0

$W_0[:,:,2]$

-1	-1	-1
-1	8	-1
-1	-1	-1

: $[0.1+0.0+0.-1]+[1.0+0.0+2.0]+[0.-1+3.0+3.1]=3$

: $[0.0+0.-1+0.0]+[0.-1+1.4+1.-1]+[0.0+1.-1+3.0]=2$

: $[0.-1+0.-1+0.-1]+[0.-1+1.8+2.-1]+[3.-1+2.-1+0.-1]=1$

Ergebnis der Convolution: bias = immer = 1

$$\begin{bmatrix} 3 & 7 & 4 \\ 17 & -7 & 1 \\ 14 & 25 & 13 \end{bmatrix}$$