Exercise sheet: 7

Edited on June 18

Task 1: FCN

The following FCN structure is given in a paper:

All CNN layers have a stride of 1, a kernel size of 3x3 and use the *same* padding, ie a padding of 1. The number of filters in the encoder is 20, 40, 60, 80, 100, and in the decoder 80, 60, 40, 5. All pooling layers are normal 2x2 Max pooling layer. The first deconv layer is equivalent to a normal conv-Layer as it has a 1x1 stride „Same "padding used. The next two Deconv layers are used for upsampling (Stride 2x2) to restore the image to bring it to its original size. The first three deconv layers have a kernel size of 3, the last layer only of 1x1. The final output is sent to a softmax to determine the probability distribution. This FCN should be used one page in the background, text, image, headings and marginalia to segment.

(a) Why does the last FCN layer have 5 channels?

Solution: 5 different classes

(b) Does the network have skip connections? Does that make sense?

Solution: This FCN network has no skip connections. Whether that makes sense is can be answered in both directions:

- Yes: There are only large regions in the layout, not individual ones Letters of interest.
- No: To differentiate between marginalia, images and text, you may Sometimes fine resolutions are required.

(c) Enter the pseudo-code for the network.

Solution:

Conv 20, 3x3, padding = same
ReLU

Conv 40, 3x3, padding = same
Pool 2x2
Conv 60, 3x3, padding = same
ReLU
Conv 80, 3x3, padding = same
pool
Conv 100, 3x3, padding = same
ReLU
Deconv 80, 3x3, padding = same (stride = 1x1)
ReLU
Deconv 60, 3x3, 2x2, padding = same
ReLU
Deconv 40, 3x3, 2x2, padding = same
Deconv 5, 1x1
Softmax

(d) Give the input and output dimensions for each layer of the network both the number of trainable weights.

Solution:

Don't forget the bias!

Layer	Input	output	N_{train}
Conv	$h \times w \times 1$	$h \times w \times 20$	1 3 3 20 + 20
ReLU	$h \times w \times 20$	$h \times w \times 20$	0
Conv	$h \times w \times 20$	$h \times w \times 40$	20 3 3 40 + 40
pool	$h \times w \times 40$	$h/2 \times w/2 \times 40$	0
Conv	$h/2 \times w/2 \times 40$	$h \times w \times 60$	40 3 3 60 + 60
ReLU	$h/2 \times w/2 \times 60$	$h/2 \times w/2 \times 60$	0
Conv	$h/2 \times w/2 \times 60$	$h/2 \times w/2 \times 80$	60 3 3 80 + 80
pool	$h/2 \times w/2 \times 80$	$h/4 \times w/4 \times 80$	0
Conv	$h/4 \times w/4 \times 80$	$h/4 \times w/4 \times 100$	$80 \times 3 \times 3 \times 100 + 100$
ReLU	$h/4 \times w/4 \times 100$	$h/4 \times w/4 \times 100$	0
Deconv	$h/4 \times w/4 \times 100$	$h/4 \times w/4 \times 80$	100 x 3 x 3 x 80 + 80

ReLU	$h/4 \times w/4 \times 80$	$h/4 \times w/4 \times 80$	0
Deconv	$h/4 \times w/4 \times 80$	$h/2 \times w/2 \times 60$	80 3 3 60 + 60
ReLU	$h/2 \times w/2 \times 60$	$h/2 \times w/2 \times 60$	0
Deconv	$h/2 \times w/2 \times 60$	$h \times w \times 40$	60 3 3 40 + 40
Deconv	$h \times w \times 40$	$h \times w \times 5$	40 x 1 x 1 x 5 + 5
Softmax	$h \times w \times 5$	$h \times w \times 5$	0

(e) Implement the network structure and train the network on the given data.

Task 2: FCN

In another paper the following FCN structure is given:

H x W x L (output, label probability for each pixel)

H x W x 1 (Input image as grayscale)

H x W x 20

H x W x 20, ReLU

H x W x 60

H x W x 30

Skip connection

H / 2 x W / 2 x 30

H x W x 30, ReLU

H / 2 x W / 2 x 40, ReLU

H / 2 x W / 2 x 80

H / 2 x W / 2 x 40

Skip connection

H / 4 x W / 4 x 40

H / 2 x W / 2 x 40, ReLU

H / 4 x W / 4 x 60, ReLU

H / 4 x W / 4 x 120

H / 4 x W / 4 x 60

Skip connection

H / 8 x W / 8 x 60

H / 4 x W / 4 x 60, ReLU

Conv 5x5

H / 8 x W / 8 x 80, ReLU

H / 8 x W / 8 x 80, ReLU

Max pool 2x2

Deconv 2x2

Concatenate

Conv 1x1

Encoder

decoder

The final output (logits) is sent to a softmax in order to determine the probability to determine the distribution of labor. This FCN is intended to be used on a page Recognize staff lines.

Staff line detection

Input

Target

(a) How many channels does the last layer have?

Solution: Either 2 channels (foreground / background), then with Softmax Activation function and cross-entropy loss, or 1 channel, then with signal moid and logistic regression loss (logistic regression)

(b) Does the network have skip connections? Does that make sense?

Solution: yes! Is useful or necessary, as the staff lines are only very narrow le / are thin objects.

(c) Enter the pseudo-code for the network.

Solution:

All conv / deconv layers with padding = same

L0 = conv 20, 5x5
L1 = ReLU
L2 = conv 30, 5x5
L3 = pool
L4 = conv 40, 5x5
L5 = ReLU
L6 = Conv 40, 5x5
L7 = pool
L8 = Conv 60, 5x5
L9 = ReLU

L10 = Conv 60, 5x5
L11 = pool
L12 = Conv 80, 5x5
L13 = ReLU
L14 = Deconv 60, 5x5, 2x2
L15 = ReLU
L16 = Concat L15, L10
L17 = Deconv 40, 5x5, 2x2
L18 = ReLU
L19 = Concat L18, L6
L20 = Deconv 30, 5x5, 2x2
L21 = ReLU
L22 = Concat L21, L2
L23 = Conv 20, 5x5
L24 = conv 2, 1x1
L25 = Softmax

(d) Give the input and output dimensions for each layer of the network both the number of trainable weights.

Solution: Analogous to issue 1, Concat of course also has no parameters, but be careful that the dimension of the output of the concat layer corresponds to the sum me of the channels (this is indicated in the figure). This in turn is then the dimension for the filters.

(e) Implement the network structure and train the network on the given data.