Programming with neural networks: Exercise sheet 10 SS 2020

University of Würzburg - Chair for Computer Science VI Prof. Dr. F. Doll, A. Hekalo, M. Krug

Exercise sheet: 10

Edited on July 16

Task 1: Classification

Consider the following network for classifying capital letters in images of size 256x256 (grayscale), whereby the conv-layers have valid padding and the activation function ReLU is. You can use several different letters appear in a picture. However, each type of letter can only be used once appear in a picture.

The numbers above the layers indicate the number of feature maps and the Filter size or the number of nodes.

(a) Enter the pseudo-code and the output dimension of all layers for the im- implementation of the network.

(b) Enter a suitable loss function (with the corresponding formula for a training example) for the training.

(c) Which forms of regularization could be built in and which operational Are data augmentation useful here? Why why not?

(d) How must the network and the loss function be adjusted so that one can only classified exactly one letter per picture?

Solution:

(a) (input 256x256x1)

Conv2D 6x5x5 252x252x6

Page 7

Conv2D 6, 5x5	252x252x6
ReLU	252x252x6
MaxPool2D 2x2	126x126x6
Conv2D 16, 5x5	122x122x16
ReLU	61x61x16
MaxPool2D 2x2	61x61x16
Flatten	59,536

FC 120	120
ReLU	120
FC 84	84
ReLU	84
FC 26	26
Sigmoid	26

(b) Binary Cross Entropy (negative log likelihood). Formula (for a training example):

$$\sum_{c=1}^{26} - y_c \log P(Y_c = 1 \mid X) - (1 - y_c) \log P(Y_c = 0 \mid X)$$

(c)
- Dropout according to the FC layers (e.g. rate = 0.5)
- Dropout to conv layers at a low rate (e.g. 0.1)
- Regularization terms for Weigth Decay
- Noise on input / weights / output
- Early stopping
- (Batch Normalization)
- Data augmentation

For data augmentation it would be useful:

- (Small) rotations
- Brightness shift
- Zoom / Crop (small)
- (Small) translations (in both directions)
- Conditional reflections for symmetrical letters (eg A, I, H Etc.)

Not suitable:

- Large rotations and reflections in which the label changes becomes
- Scaling, because we need a fixed resolution
- large cropping / translations in which parts of the letters be removed

(d) For only one letter per picture: Sigmoid in the last layer to soft- Use max and categorical instead of binary cross entropy.

Task 2: sequence classification

In the following, the network from exercise 1 is intended to identify the sequence of letters

be expanded. For this we assume, in simplified form, that only the letters from A to J can occur. Sequences of 0 to 5 letters should be recognized can be.

(a) Let the following output of the network $y_{t\ be\ given}$ $_k$. Decode the sequence using the CTC greedy decoder.

y_{k}^{t}	0	1	2	3	4th
-	0.9	0.1	0.2	0.3	0.15
A.	0	0.8	0.1	0	0
B.	0	0	0.7	0.4	0
C.	0	0	0	0	0
D.	0	0	0	0	0.1
E.	0	0	0	0.3	0.15
F.	0	0	0	0	0.1
G	0.1	0.1	0	0	0.15
H	0	0	0	0	0.15
I.	0	0	0	0	0.2
J	0	0	0	0	0

Solution: Argmax at every point results in **-ABBI** and thus becomes **ABI**

(b) The ground truth for the output is actually AI. Calculate the forward and backward variables α and β . Use it to calculate the total probability possibility of AI.

Solution:

α	0	1	2	3	4th
-	0.9	0.09	0.018	0.0054	0.00081
A.	0	0.72	0.081	0	0
-	0	0	0.144	0.0675	0.010125

I.	0	0	0	0	0.0135
-	0	0	0	0	0

β	0	1	2	3	4th
-	0.0135	0.0006	0	0	0
A.	0	0.0144	0.006	0	0
-	0.00108	0.0012	0.012	0.06	0
I.	0	0	0	0	0.2
-	0.00081	0.0009	0.009	0.045	0.15

P.	0.0135	0.0135	0.0135	0.0135	0.0135

Task 3: Image segmentation and object recognition

(a) How do image segmentation and object recognition differ, ie how is classified, what are the expenses etc.?

Solution: When segmenting the image, each pixel of the image is usually individually classified, ie the output is a mask of the same dimension as the picture. With some architectures, the mask also has a slightly smaller one Resolution and needs to be scaled up. The classes are mostly exclusive sive, ie the regions do not overlap. For each pixel there is a Class vector assigned, the length of which corresponds to the number of classes.

When detecting objects, bounding boxes are calculated, the regions of the image and classify this region. The classes are never exclusive, ie the boxes can overlap. It can also multiple instances of the same class occur. The exact number of Boxing depends on the architecture or on the Convolutions, but 4 coordinates and *n* classes are assigned to each box .

(b) Given the well-known FCN U-Net (Figure 1) with the variation that have the conv-layer padding = same. The Up-Convolutions are supposed to be the have the same filter size as the normal Convolutions. Enter the pseudo docode (without dimensions) for this architecture. The activation function can be understood here as an attribute of the conv-layer. The output should be 10 different, exclusive classes may be possible.

Figure 1: U-Net

Solution:

All convs / deconvs with padding = same

Parameter Deconv: number of filters, filter size, strides

L0 = Conv 64, 3x3, ReLU
L1 = Conv 64, 3x3, ReLU

L2 = max pool, 2x2
L3 = Conv 128, 3x3, ReLU
L4 = Conv 128, 3x3, ReLU
L5 = max pool, 2x2
L6 = Conv 256, 3x3, ReLU
L7 = Conv 256, 3x3, ReLU
L8 = Max Pool, 2x2
L9 = Conv 512, 3x3, ReLU
L10 = Conv 512, 3x3, ReLU
L11 = Max Pool, 2x2
L12 = Conv 1024, 3x3, ReLU
L13 = Conv 1024, 3x3, ReLU
L14 = Deconv 512, 3x3, 2x2
L15 = Concat L14, L10
L16 = Conv 512, 3x3, ReLU
L17 = Conv 512, 3x3, ReLU
L18 = Deconv 256, 3x3, 2x2

L20 = Conv 256, 3x3, ReLU
L21 = Conv 256, 3x3, ReLU
L22 = Deconv 128, 3x3, 2x2
L23 = Concat L22, L4
L24 = Conv 128, 3x3, ReLU
L25 = Conv 128, 3x3, ReLU
L26 = Deconv 64, 3x3, 2x2
L27 = Concat L26, L1
L28 = Conv 64, 3x3, ReLU
L29 = Conv 64, 3x3, ReLU
L30 = conv 10, 1x1
L31 = Softmax

(c) List the (main) approaches to classification in object recognition and name an example architecture for each.

Solution: Main approaches:

- Two stage detectors: First determination of many regions of interest ("Objectness Score"), then classification of these RoIs (and additionally Adjustment of the coordinates). An example of this is Faster R-CNN.
- One / single stage / shot detectors: regression and classification in one step. Examples are YOLO, Single Shot Detector and anchor free methods.
- Anchor-free methods can be viewed as a separate approach, too if they can also be defined as a single stage method.

Examples are CornerNet and CenterNet.

(d) What types of segmentation are there? What makes them stand out?

Solution: There are three different types:

- Semantic Segmentation: Classifies the entire image into regions, but only by label, i.e. person, building, soil, animal, tree, etc. and does not differentiate between different people, cars, Animals etc.
- Instance Segmentation: Classifies only countable objects in the image, such as

Areas such as buildings, streets, etc. This is done between individual People are differentiated, each with its own label.

- Panoptic segmentation: Combines semantic and instance segments tation, that is, both uncountable regions and countable objects are classified and the countable objects are different from one another. distinguishable.

Task 4: Medical application

Consider the following application in medicine: You want to diagnose Doctors provided support by looking at the (entire) lung volume from X-ray scans

or, as a first step, determine the patient's lung area.

Describe which type of neural network you could use and which Kind of data that medical professionals would need to provide for training.

Solution: segmentation task, e.g. B. U-Net. For training, the lung gene wing be marked, z. B. as a binary mask or as a polygon (then convert). One possibility would also be to consider the lungs as separate classes to be able to carry out independent calculations.