

Computer Vision (CV) (English) – JMU Wuerzburg, Master, Spring 2021

There are **4 home assignments** roughly corresponding to the teaching sessions. They are counting for a 10% bonus in the final grade for this course and are meant for familiarizing the students with the written exam which will contain similar tasks.

Each home assignment (HA) has 8 topics that are weighted equally. The solutions to each HA should be sent as a single *pdf* by email at radu.timofte@uni-wuerzburg.de no later than May 26 for HA1, June 10 for HA2, July 1 for HA3, and July 15 for HA4, respectively. Each student shall include with her/his email full name and the details of the study program. The Subject of the email should be of the form:

"CV 2021 HA[#home_assignment] [master_program] [lastname] [firstname]"

where [#home_assignment] should be replaced with the number of the home assignment, the [master_program] with the corresponding short abbreviated program (like XtAI, LuRI) and [lastname] and [firstname] are the student's last name and first name, respectively.

The grades on the home assignments will be provided together with the final course grade.

Plagiarism and misconduct are not tolerated!

The grading of Computer Vision (taught in English) is as follows:

- **The final CV (Eng) grade (maximum 100%)** is composed from 100% written exam.
- Up to **10% extra bonus** will be awarded for those attending the lectures.
- Up to **10% extra bonus** will be awarded for those attending the exercise sessions.
- Up to **10% extra bonus** will be awarded for those sending their solutions to the HAs.
- For **exam** the students are required to prepare the contents covered during the lectures, exercise sessions, home assignments and found on the provided slides and lab materials. Examples of potential exam topics are found in the HAs.

Home assignment IV

(due July 15th, 2021)

25. What is over-fitting in the context of image classification with a CNN? Enumerate five strategies to prevent over-fitting? Justify your answers.

26. Given the following 3x3 convolution kernel:

$$\begin{bmatrix} 1 & -2 & 0 \\ 0 & 4 & -1 \\ 0 & 2 & 0 \end{bmatrix}$$

(i) Compute and report the result of applying the convolution kernel to the grayscale 5x5 image:

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 1 & 0 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(ii) Compute and report the result of applying the kernel in a transposed convolution to the grayscale 3x3 image:

$$\begin{bmatrix} 0 & 1 & 0 \\ 2 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

27. In a CNN architecture for image classification, the sequence of layers is as follows (no biases are added following either the fully-connected or convolutional layers):

```
input rgb image with 32x32 pixels (total values: 32x32x3)
→ 5x5 convolutional layer with stride 1 and 16 output feature maps
→ activation layer→avg-pooling layer with 2x2 spatial extent and stride 2
→ 3x3 convolutional layer with stride 1 and 8 output feature maps
→ activation layer→avg-pooling layer with 2x2 spatial extent and stride 2
→ vectorize the feature maps
→ fully-connected layer with 32 output units→ activation layer
→ fully-connected layer with 16 output units
```

What is the number of parameters of this specific CNN architecture?

(for partial points you should provide intermediate number of parameters per each layer)

What is the receptive field of this specific CNN architecture?

What is the number of classes?

Justify your answers.

28. In a CNN architecture for image classification, the sequence of layers is as follows (no biases are added following either the fully-connected or convolutional layers):

```
input grayscale image with 32x32 pixels
→ 3x3 convolutional layer with stride 1 and 64 output feature maps
→ activation layer→max-pooling layer with 2x2 spatial extent and stride 2
→ 3x3 convolutional layer with stride 1 and 32 output feature maps
→ activation layer→max-pooling layer with 2x2 spatial extent and stride 2
→ 3x3 convolutional layer with stride 1 and 16 output feature maps
→ activation layer→max-pooling layer with 2x2 spatial extent and stride 2
→ vectorize the feature maps
→ fully-connected layer with 16 output units→ activation layer
→ fully-connected layer with 10 output units
```

What is the number of parameters of this specific CNN architecture?

(for partial points you should provide intermediate number of parameters per each layer)

29. How the receptive field can be increased in a CNN architecture? Justify your answer.

30. In a CNN architecture for image classification, the sequence of layers is as follows (no biases are added following either the fully-connected or convolutional layers):

```
input grayscale image with 64x64 pixels
→ 3x3 convolutional layer with stride 1 and 32 output feature maps
→ activation layer→max-pooling layer with 2x2 spatial extent and stride 1
→ 5x5 convolutional layer with stride 1 and 32 output feature maps
→ activation layer→max-pooling layer with 2x2 spatial extent and stride 2
→ 3x3 convolutional layer with stride 1 and 32 output feature maps
→ activation layer→max-pooling layer with 2x2 spatial extent and stride 2
→ vectorize the feature maps
→ fully-connected layer with 128 output units→ activation layer
→ fully-connected layer with 10 output units
```

What is the number of parameters of this specific CNN architecture?

(for partial points you should provide intermediate number of parameters per each layer)

31. Enumerate a couple of applications where translation invariance is desirable and a couple of applications where translation variance is desirable. Justify your answers.

32. How a CNN that is not fully-connected can be trained or designed such that to achieve output results that are invariant to affine translations of the input image? Justify your answer.