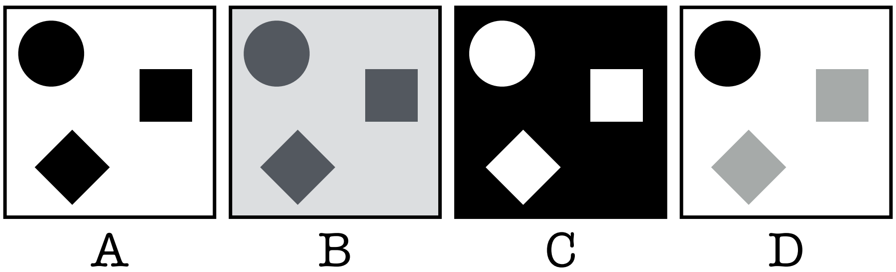
Lab 2: Bonus Questions

- There are two sets of bonus questions for this lab, each worth 3 points. In total, you can gain up to 18 points. Obtaining 12 out of the 18 points will improve your grade by one if you fulfill all requirements to pass the course.
- You have to work on the questions on your own and have to hand in your solutions individually.
- Your answers will be graded on a continuous scale, i.e., we try to reward partially correct answers.
- Please hand in your solutions as a PDF through Canvas.

1 SIFT, 3 points

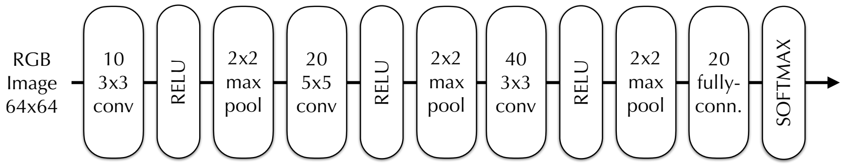
(a) Which of the following patches will result in the same SIFT descriptor? Justify your answer and make sure to explain why the other patches are different.



(b) Given a patch centered around a keypoint, SIFT computes a descriptor from that patch that can be used for image matching.
Describe how SIFT ensures that its descriptors are invariant to changes in scale and rotations (in the image plane), and to uniform additive and multiplicative changes in brightness.

(c) A similarity transformation between two images is defined by a rotation by an angle θ, a change in scale by a factor s, and a 2D translation $\mathbf{t} \in \mathbb{R}^2$.
Explain how a similarity transformation can be computed from a single match between a SIFT feature in one image and a SIFT feature in a second image.

2 Deep Learning, 3 points

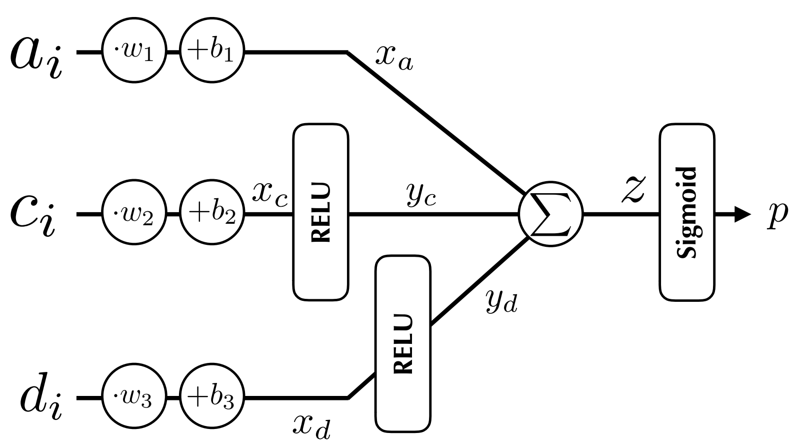


(a) Consider the convolutional neural network shown above and assume that the convolutional layers use padding. You can assume that no bias terms are used.
Assume that you replace the fully connected layer in the network by a fully convolutional layer to be able to apply the network on RGB input images of arbitrary dimensions. You do the replacement such that applying the new network without padding on a 64×64 RGB image results in the same output as applying the original network.
How many trainable parameters does the new fully convolutional layer have? Explain your answer!

(b) The receptive field of a convolutional layer is an area of size $k \times l$ in the original input image that affects the output of the layer. For example, applying a 3×3 convolutional layer (the number of filters has no impact on the receptive field, only the spatial dimensions of the filters) on the input image results in a receptive field of size 3×3 as a 3×3 region is the input to the filters in the layers. Similarly, the receptive field of a 9×9 layer is 9×9. Deeper layers in a CNN have larger receptive fields as their input pixels themselves correspond to larger regions in the input image.
What is the receptive field of the last fully convolutional layer in the network shown above? Justify your answer!

(c) Consider the neural network shown below that performs binary classification on a 3-tuple (a_i, c_i, d_i) of three scalar input values $a_i, c_i, d_i \in \mathbb{R}$. The RELU and Sigmoid functions are given as $\text{RELU}(x) = \max(0, x)$ and $\text{Sigmoid}(z) = \frac{e^z}{1+e^z}$. The derivative of the RELU function is given as

$$\frac{\partial \text{RELU}(x)}{\partial x} = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{otherwise} \end{cases} .$$



Given a positive example $(a_i, c_i, d_i) = (4, 0, -4)$, we want to maximize the probability p. Our loss function is thus given as the negative log-likelihood loss $L_i = -\ln(p)$. Compute the derivatives

$$\frac{\partial L_i}{\partial w_1}, \quad \frac{\partial L_i}{\partial b_1}, \quad \frac{\partial L_i}{\partial w_2}, \quad \frac{\partial L_i}{\partial b_2} \quad \frac{\partial L_i}{\partial w_3}, \quad \frac{\partial L_i}{\partial b_3}$$

through the backpropagation algorithm. To this end, first perform a forward pass to compute values for x_a, x_c, y_c, x_d, y_d, z, and p. Next, use the chain rule to derive formulas for the derivatives in the backward pass. Compute the actual values for the derivatives using your equations and the values for x_a, x_c, y_c, x_d, y_d, z, and p computed during the forward pass.
The current values for w_1, b_1, w_2, b_2, w_3, and b_3 are

$$w_1 = -2, \qquad b_1 = 0, \qquad w_2 = 4, \qquad b_2 = 8, \qquad w_3 = 10, \qquad b_3 = 20 \ .$$