Affix the authentification label over this field.

Surname: ....
First name: ....
Matr.-No.: ....

Institut für Mess- und Regelungstechnik
mit Maschinenlaboratorium
Karlsruher Institut für Technologie (KIT)
Prof. Dr.-Ing. C. Stiller

Exam in "Machine Vision"

Date of exam:	February 20, 2023
Time of exam:	11:00-12:00

Question 1 (5 points)

Provide the 4-by-4 binomial filter.

Question 2 (6 points)

The figures below represent a gray level image and a kernel. Apply the convolution between the kernel and the gray level image. Do not use any padding (leave out boundary pixels).

Note down your calculation steps.

Gray Level Image

20	30	50	80	200
10	20	80	70	110
10	40	50	50	160

Kernel

0	0	0
0	0	1
0	2	4

Question 3 (8 points)

Calculate the line that fits a set of points using the total least squares approach. The set of points is (0,0), (1,0), and (1,1), where the point coordinates are given as (x_1, x_2).

Remark: You can calculate the eigenvalue by zeroing the characteristic polynom:
$det(B - I\lambda) = 0$,
where the determinant of a 2x2 matrix is calculated by:
$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$

Question 4 (6 points)

Below you find six words that describe color, three RGB tuples (min:0, max:1) on the left and three HSV tuples on the right. Connect each RGB tuple with the word that describes its color. Connect each HSV tuple with the word that describes its color. Mind that some words might be connected with several RGB and HSV tuples while other words might not be connected to any.

RGB tuples		*HSV tuples*
(0.3, 1.0, 0.3)	light green	(10°, 1, 0.5)
	orange	
(1.0, 0.6, 0)	pink	(355°, 0.5, 1)
	light blue	
	black	
(0.5, 0, 0)	dark red	(122°, 0.8, 1)

Question 5 (5 points)

Assume that during a camera calibration process the 3d positions of the markers have not been measured. However, the distance of two detected markers is known as $d = 25$ cm. The markers are located in the image at $\vec{p_1}$ and $\vec{p_2}$. The calibration board is assumed to be parallel to the image plane. The intrinsic parameters of the camera are given by matrix $\mathbf{A}$. For simplification, a pinhole camera model is assumed.

Determine the distance to the calibration board.

$$\mathbf{A} = \begin{pmatrix} 800 & 0 & 1000 \\ 0 & 800 & 500 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\vec{p_1} = \begin{pmatrix} 300 \\ 381 \end{pmatrix}, \quad \vec{p_2} = \begin{pmatrix} 350 \\ 381 \end{pmatrix}$$

The points $\vec{p_1}$ and $\vec{p_2}$ are given in $\begin{pmatrix} u \\ v \end{pmatrix}$ format.

Question 6 (2+4 points)

In level set methods for image segmentation the Mumford-Shah-based segmentation approach uses the evolution term

$$\frac{\partial \vec{x}}{\partial t} = \frac{\nabla \phi}{||\nabla \phi||} \cdot \big(- \beta\kappa - \alpha - \lambda_1 (g - \bar{g}_{foreground})^2 + \lambda_2 (g - \bar{g}_{background})^2 \big)$$

where g determines the gray or color value of the image, $\kappa = \nabla \left(\frac{\nabla \phi}{||\nabla \phi||} \right)$ denotes the local curvature of the boundary, and $\alpha \in \mathbb{R}, \beta, \lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$ are tuning parameters.

(a) Assume that we choose $\lambda_1 = \lambda_2 = 1$ and $\alpha = 0$. How does the choice of β influence the result? Compare the cases $\beta = 0$ and $\beta \gg 1$

(b) Assume that we know that the foreground segment should cover approximately $k \in \mathbb{N}$ pixels. Denote with n the present size of the foreground segment in pixels. Extend the Mumford-Shah evolution term to foster foreground segements with the desired size $n \approx k$ over segments with very different sizes $n \not\approx k$. Explain your solution briefly.

Question 7 (5 points)

Prove or disprove whether the following functions can serve as kernel functions in a support vector machine.

(a) $f(\vec{x},\vec{y}) = e^{||\vec{x}||+||\vec{y}||}$ where $||\vec{x}||$ denotes the Euclidean norm of vector $\vec{x}$.

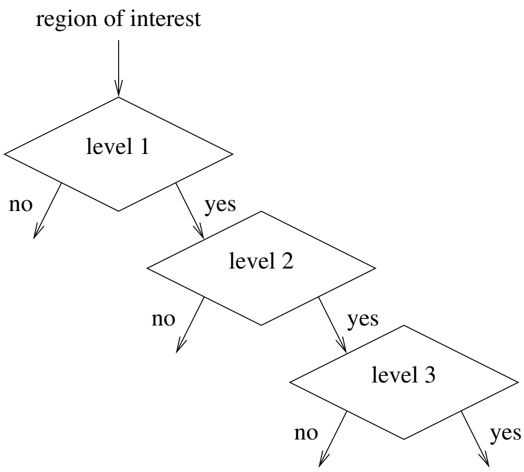
(b) $g(x,y) = x^2y + y$

Question 8 (6 points)

Build a cascade classifier with three levels to detect certain objects in an image. The structure of the cascade is shown in the illustration below. For that purpose you can use the six classifiers described in the table below. Assign to each of the three levels the best classifier from the table. The cascade should be as accurate as possible and it should not require more computational effort than necessary to achieve that accuracy.

For each classifier briefly explain why you have chosen the classifier for a specific level.

classifier	precision	recall	execution time per region of interest
A	82%	55%	0.001 ns
B	49%	98%	8.9 µs
C	53%	98%	6.3 µs
D	92%	87%	1.2 ms
E	99%	28%	0.01 ns
F	34%	100%	0.2 ns



Question 9 (5 points)

Which of the following deep network architectures can be used for which applications? Draw a line between each network architecture and the respective application.

Remark: The given networks should be used without any additional heads or modifications.

network architecture	**application**
Faster R−CNN	image classification
convolutional encoder–decoder network	coloring images, i.e. creating color images from gray value images
ResNet	object detection in images
LSTM network	activity recognition in video streams
YOLO	

Question 10 (8 points)

During the backpropagation of a multi-layer-perception its weights are changed to optimize an error term.

Implement the Python function `change_weights` that implements this process using gradient descent with momentum.

The function should have exactly 5 arguments, the weights of one layer (`weights`), the gradient of the error term with respect to these weights (`gradients`), the previous momentum term (`momentum`), and the hyperparameters epsilon (`epsilon`) and alpha (`alpha`). The function should update the weights and the momentum term and return both.

Provide type hints at your function definition.

Remark: If you use any kind of array or tensor in your solution, use Numpy arrays.

```
import numpy as np
```