Time: 30 minutes Max marks: 10

Instructions:
- Do not plagiarize. Do not assist your classmates in plagiarism.
- Show your full solution for the questions to get full credit.
- Attempt all questions that you can.
- In the unlikely case a question is not clear, discuss it with an invigilating TA. Please ensure that you clearly include any assumptions you make, even after clarification from the invigilator.

1. Fig. 1 shows the convolution operation on an input array using a single kernel (filter) to generate the output feature map. Answer the following questions.

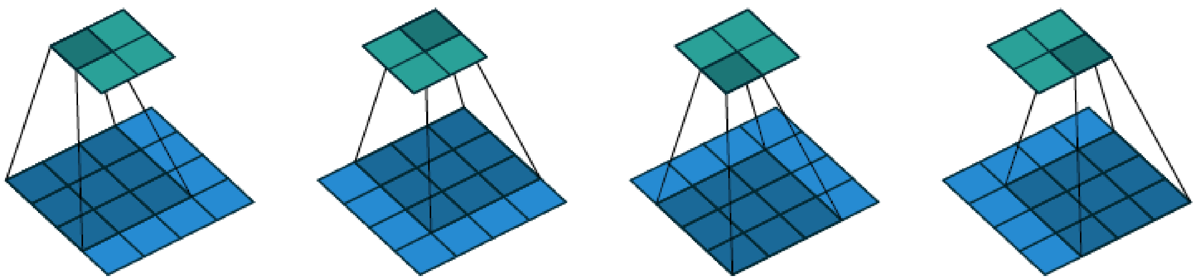

Figure 1: Steps of the convolution operation. The 4×4 patch is the input. The dark-shaded 3×3 patch overlaid on the input is the convolution kernel (filter). The 2×2 grid is the output feature map.

(a) ($1/2$ point) What are the values of stride and padding used in Fig. 1?

(b) ($1/2$ point) What is the total number of *learnable* parameters in this case?

(c) ($1/2$ point) If the input was an RGB image of 6×6, and we used a 3×3 kernel with a stride of 1 with no padding, what would be the spatial dimensions (height $\times$ width) of the output feature map?

(d) ($1/2$ point) The feature map from part (c) above is passed through a max-pooling layer with a 2×2 filter and a stride of 2 and no padding (Usually pooling layers don't use padding). What will be the spatial dimensions of the feature map at the output of this pooling layer?

(e) (1 point) Say you have the convolution layer with 8 such filters (as in part (c)) to process the 6×6 RGB image. What would be the total number of *learnable* parameters for this layer?

(f) (2 points) For Fig. 1, let the weights of the 3×3 kernel be $w_{i,j},\ i, j \in \{0, 1, 2\}$. Write the convolution operation as a matrix multiplication when the input is the 4×4 single-channel image and the output is the 2×2 feature map. You may ignore the bias term for this part, however, for full credit, describe how would the input / output need to be processed in order to implement the convolution operation as a matrix multiplication, e.g., processing may need operations like reshape etc.

Total for Question 1: 5

Solution:

(a) ($1/2$ point) stride $= 1$, padding $= 0$.

Rubric: binary.

(b) (½ point) 9 for the kernel (filter) weights and 1 for the bias = 10.

Rubric: binary.

(c) (½ point) The output feature map size O is given by the formula

$$O = \left\lfloor \frac{(I - K + 2P)}{S} \right\rfloor + 1$$

where I is the input dimension (calculated separately if height and width are not the same, similarly if kernel height and width are not the same; usually the kernel dimensions are the same, the input not often), K is the kernel dimension, P is the amount of padding (in terms of rows / columns, with the multiplication of 2 accounting for top & bottom rows / left & right columns), and S is the stride and $\lfloor x \rfloor$ indicates the *floor* function (largest integer smaller than x). Using this formula, the output dimension of the feature map will be 4×4.

Rubric: binary.

(d) (½ point) We apply the 2×2 max-pooling filter to the 4×4 feature map, which gives us a 2×2 output feature map.

Rubric: binary.

(e) (1 point) Since it is an RGB image, there are 3 input channels, and we have a 3×3 kernel. Therefore we have $3 \times 3 \times 3 = 27$ weights for each kernel and 1 bias term. For eight filters, the total number of learnable parameters is $28 \times 8 = 224$.

Rubric:
+1 if calculation fully correct
+0.5 if bias not considered

(f) (2 points) Let $\mathbf{I}$ be the 4×4 input image shown in Fig. 1. We *vectorize* this image row-wise to obtain the 16-dimensional column vector $\mathbf{x} = [I_{0,0}, I_{0,1}, \ldots, I_{0,3}, I_{1,0}, \ldots, I_{3,3}]^\top$. We then write the convolution operation as a matrix multiplication as shown below.

$\mathbf{C}_{4\times 16} =$

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

Then the 4-dimensional vectorized output $\mathbf{y} = \mathbf{C}\mathbf{x}$ and is then reshaped to obtain a 2×2 matrix that is the feature map.

Rubric:
+1 point for correct matrix
+0.5 points for mentioning vectorizing input
+0.5 points for mentioning reshaping output vector

2. (a) (1 point) Is $\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ a valid rotation matrix? Explain why or why not?

(b) (1 point) If $\mathbf{R}$ above is not a valid rotation matrix, convert it into one by only changing the sign of the non-zero entries. Find the axis and angle of rotation, for the corrected rotation matrix, or for the original if it already was a valid rotation matrix. {*Hint*: This should be possible just by inspection, even if you don't recall the formula.}

Total for Question 2: 2

Solution:

(a) (1 point) Yes, $\mathbf{R}$ is a valid rotation matrix, as it is orthogonal and has a determinant of 1.

(Edit) Comparing it to the standard form of the rotation matrix is also valid.

Rubric:
0.5 points for correctness; 0.5 points for justification

(b) (1 point) The axis of rotation is the third axis, or Z, i.e., $[0, 0, 1]^\top$, and the angle of rotation is $-90°$ or equivalently $270°$. One could also alternatively claim that the axis of rotation is the negative Z axis, i.e., $[0, 0, -1]^\top$ and the angle of rotation is $90°$.

Rubric:
0.5 points for correct axis; 0.5 points for correct angle

3. (a) (1 point) Write the most general form of the intrinsic camera parameter matrix $\mathbf{K}$ and identify the name of each parameter.

(b) (1 point) Given two vectors $\mathbf{n} = [n_1, n_2, n_3]^\top$ and $\mathbf{x} = [x_1, x_2, x_3]^\top$, give the expression for the cross product $\mathbf{n} \times \mathbf{x}$. Write the form of the 3×3 cross-product matrix denoted by $\mathbf{N}$ (also denote by $[\mathbf{n}]_\times$), such that $\mathbf{Nx} = \mathbf{n} \times \mathbf{x}$. {*Hint*: You may use of the determinant-based approach for computing cross-products of 3D vectors.}

(c) (1 point) Show that $\mathbf{Nn} = (\mathbf{Nx})^\top \mathbf{n} = 0$ for any arbitrary non-zero $\mathbf{x}$ and $\mathbf{n}$.

Total for Question 3: 3

Solution:

(a) (1 point)

$$\mathbf{K} = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

where

- f_x and f_y are the **focal length** parameters (measured in #pixels along the x and y axes respectively; recall that a pixel may not be square and can have different horizontal and vertical side lengths.)
- p_x and p_y are the pixel coordinates of the **principal point**, which is the projection of the center of projection on the image plane.
- s is the **skew parameter** that captures the correlation in the pixels horizontal and vertical axes (non-zero for non-rectangular pixels).

Rubric:
+0.5 points for correct matrix; 0 if skew not present
+0.5 points for naming the parameters

(b) (1 point) We can compute $\mathbf{n} \times \mathbf{x}$ as:

$$\begin{aligned}
\mathbf{n} \times \mathbf{x} &= \begin{vmatrix} \widehat{i} & \widehat{j} & \widehat{k} \\ n_1 & n_2 & n_3 \\ x_1 & x_2 & x_3 \end{vmatrix} \\
&= \begin{bmatrix} n_2x_3 - n_3x_2 \\ n_3x_1 - n_1x_3 \\ n_1x_2 - n_2x_1 \end{bmatrix} \\
[\mathbf{n}]_\times &= \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix}
\end{aligned} \tag{1}$$

Rubric:
+1 for correct matrix

(c) (1 point) We can see that for any non-zero $\mathbf{n}$, we necessarily have

$$\begin{aligned}
[\mathbf{n}]_\times \mathbf{n} = \mathbf{N}\mathbf{n} &= \begin{bmatrix} 0 - n_3n_2 + n_2n_3 \\ n_3n_1 + 0 - n_1n_3 \\ -n_2n_1 + n_1n_2 + 0 \end{bmatrix} \\
= [\mathbf{n}]_\times^\top \mathbf{n} = \mathbf{N}^\top \mathbf{n} &= \begin{bmatrix} 0 + n_3n_2 - n_2n_3 \\ -n_3n_1 + 0 + n_1n_3 \\ n_2n_1 - n_1n_2 + 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\end{aligned} \tag{2}$$

$$\left([\mathbf{n}]_\times \mathbf{x}\right)^\top \mathbf{n} = (\mathbf{N}\mathbf{x})^\top \mathbf{n} = \mathbf{x}^\top [\mathbf{n}]_\times^\top \mathbf{n} = \mathbf{x}^\top \mathbf{N}^\top \mathbf{n} = \mathbf{x}^\top \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0$$

Rubric:
+0.5 points for showing $Nn = 0$
+0.5 points for showing $(Nx)^T n = 0$