Task 1

a) Adding padding to the image and flipping the kernel :

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 2 & 3 & 1 & 0 \\ 0 & 3 & 9 & 1 & 1 & 4 & 0 \\ 0 & 4 & 5 & 0 & 7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \quad , \quad \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

We then apply hadamard product between the kernel matrix and each of the matrix below, obtained by sliding the kernel with stride $= 1$:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 3 & 9 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 2 & 1 & 2 \\ 3 & 9 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \\ 9 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 4 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 2 & 1 \\ 0 & 3 & 9 \\ 0 & 4 & 5 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 2 \\ 3 & 9 & 1 \\ 4 & 5 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 \\ 9 & 1 & 1 \\ 5 & 0 & 7 \end{bmatrix} \quad \begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 4 \\ 0 & 7 & 0 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & 0 \\ 1 & 4 & 0 \\ 7 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 3 & 9 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 3 & 9 & 1 \\ 4 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 9 & 1 & 1 \\ 5 & 0 & 7 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 4 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 4 & 0 \\ 7 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We obtain :

$$-\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 9 \end{bmatrix} \quad -\begin{bmatrix} 0 & 0 & 0 \\ -4 & 0 & 4 \\ -3 & 0 & 1 \end{bmatrix} \quad -\begin{bmatrix} 0 & 0 & 0 \\ -2 & 0 & 6 \\ -9 & 0 & 1 \end{bmatrix} \quad -\begin{bmatrix} 0 & 0 & 0 \\ -4 & 0 & 2 \\ -1 & 0 & 4 \end{bmatrix} \quad -\begin{bmatrix} 0 & 0 & 0 \\ -6 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

$$-\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 18 \\ 0 & 0 & 5 \end{bmatrix} \quad -\begin{bmatrix} -2 & 0 & 2 \\ -6 & 0 & 2 \\ -4 & 0 & 0 \end{bmatrix} \quad -\begin{bmatrix} -1 & 0 & 3 \\ -18 & 0 & 2 \\ -5 & 0 & 7 \end{bmatrix} \quad -\begin{bmatrix} -2 & 0 & 1 \\ -2 & 0 & 8 \\ 0 & 0 & 0 \end{bmatrix} \quad -\begin{bmatrix} -3 & 0 & 0 \\ -2 & 0 & 0 \\ -7 & 0 & 0 \end{bmatrix}$$

$$-\begin{bmatrix} 0 & 0 & 9 \\ 0 & 0 & 10 \\ 0 & 0 & 0 \end{bmatrix} \quad -\begin{bmatrix} -3 & 0 & 1 \\ -8 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad -\begin{bmatrix} -9 & 0 & 1 \\ -10 & 0 & 14 \\ 0 & 0 & 0 \end{bmatrix} \quad -\begin{bmatrix} -1 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad -\begin{bmatrix} -1 & 0 & 0 \\ -14 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

the result is finally the 3×5 matrix where the ij^{th} component is the sum of the elements of the ij^{th} above :

$$-\begin{bmatrix} 11 & -2 & -4 & 1 & -7 \\ 24 & -8 & -12 & 5 & -12 \\ 19 & -10 & -4 & 3 & -15 \end{bmatrix}$$

b) The Convolutional layer (i) is the one reducing translational sensibility since a kernel for the convolution will detect a feature regardless of its position in the image

c) The output height H_O and width W_O are computed as follow, where H_I and W_I are the image height and width, S the stride, F the kernel (receptive **F**ield) side size and P the padding size :

$$H_O = \frac{H_I - F + 2P}{S} + 1 = H_I - 6 + 2P$$
$$W_O = \frac{W_I - F + 2P}{S} + 1 = W_I - 6 + 2P$$

Because we want $H_O = H_I$ and $W_O = W_I$, we need $2P = 6 \Rightarrow P = 3$

d) Using same equation as above, we have

$$F = H_I + 2P - S(H_O - 1) = 512 - 508 - 1 = 3$$

Kernel has dimensions 3×3

e) Again using same equation,

$$H_O = \frac{H_I - F + 2P}{S} + 1 = \frac{508 - 2}{2} + 1 = 254$$
$$W_O = \frac{W_I - F + 2P}{S} + 1 = \frac{508 - 2}{2} + 1 = 254$$

Pooled feature maps have dimensions 254×254

f) Similarly, feature maps of the second layer have dimension 252×252.

g) Each Conv layer filter has one weight kernel and one bias per input depth. Since the kernels have size 5×5, each conv layer has $D_I \cdot 26 \cdot D_O$ parameters. The Pooling layers and Flatten layer do not introduce parameters.
Each FC layers has a bias for each of its neuron and one weight per input unit for each of all its neurons, so each FC layer has $(I + 1) \cdot O$ parameters, where I is the number of input units of the layer, and O is the number of output units.
Each pooling layer divides width and height of input by 2 each, leaving depth untouched :

$$32 \times 32 \times 32 \underset{Pool1, Conv2}{\Rightarrow} 16 \times 16 \times 64 \underset{Pool2, Conv3}{\Rightarrow} 8 \times 8 \times 128 \underset{Pool3}{\Rightarrow} 4 \times 4 \times 128 \underset{Flatten}{\Rightarrow} 2048 \times 1$$

ConvLayer 1 has a 3-channel input and produce 32 feature maps :
$3 \cdot 26 \cdot 32 = 2\,496$ parameters.
ConvLayer 2 has a 32-channel input and produce 64 feature maps :
$32 \cdot 26 \cdot 64 = 53\,248$ parameters.
ConvLayer 3 has a 64-channel input and produce 128 feature maps :
$64 \cdot 26 \cdot 128 = 212\,992$ parameters.
FClayer1 has 2048 inputs and 64 output : $(2048 + 1) \cdot 64 = 131\,136$ parameters.
FClayer2 has 64 inputs and 10 output : $(64 + 1) \cdot 10 = 650$ parameters.
Taking the sum, we find that the model has $400\,522$ parameters.

Task 2