VIR exercises

Exercice 1 :

Consider a CNN with following layers :
l1 = Conv2d(in_channels=3, out_channels=n, kernel_size=(7,7), padding=3, stride=2)
l2 = Conv2d(in_channels=n, out_channels=2*n, kernel_size=(3,3), padding=1, stride=1)
l3 = Conv2d(in_channels=2*n, out_channels=4*n, kernel_size=(3,3), padding=1, stride=2)
l4 = Linear(in_features=x, out_features=5)
where x is unknown and n is a parameter. Input is passed sequentially and output of layer3 is flattened. The network is supposed to process 64×64 RGB images.

— What is the receptive field of filters at layer3, that is, what is the size of the part of an input image that they can "see" (disregard padding)

> **Solution:** Receptive field of layer3 and layer 2 combined is 5×5 (not 9×9 due to overlap) that combined with layer 1 gives us 15×15 (7×7 base and 2(stride=2) added per remaining "pixel" of previous receptive field)

— express x int terms of n

> **Solution:** The image is downsampled twice via stride by a factor of 2 that makes $x = 16 \cdot 16 \cdot 4n = 1024n$

— What is the maximal value of n if we the network parameters are needed to be less than 128KiB given the parameters are of type float32 ? (1KiB = 1024B)

> **Solution:** Let's first calculate the number of paramaters for each layer :
> layer1 : $(7 \cdot 7 \cdot 3 + 1)n = 148n$
> layer2 : $(3 \cdot 3 \cdot n + 1)2n = 18n^2 + 2n$
> layer3 : $(3 \cdot 3 \cdot 2n + 1)4n = 72n^2 + 4n$
> layer4 : $1024n \cdot 5 + 5 = 5120n + 5$
> That gives us : $90n^2 + 5274n + 5 \leq \frac{128}{4} \cdot 1024 \longrightarrow 90n^2 + 5274n - 32763 \leq 0$
> after solving the quadratic we get that the answer is 5