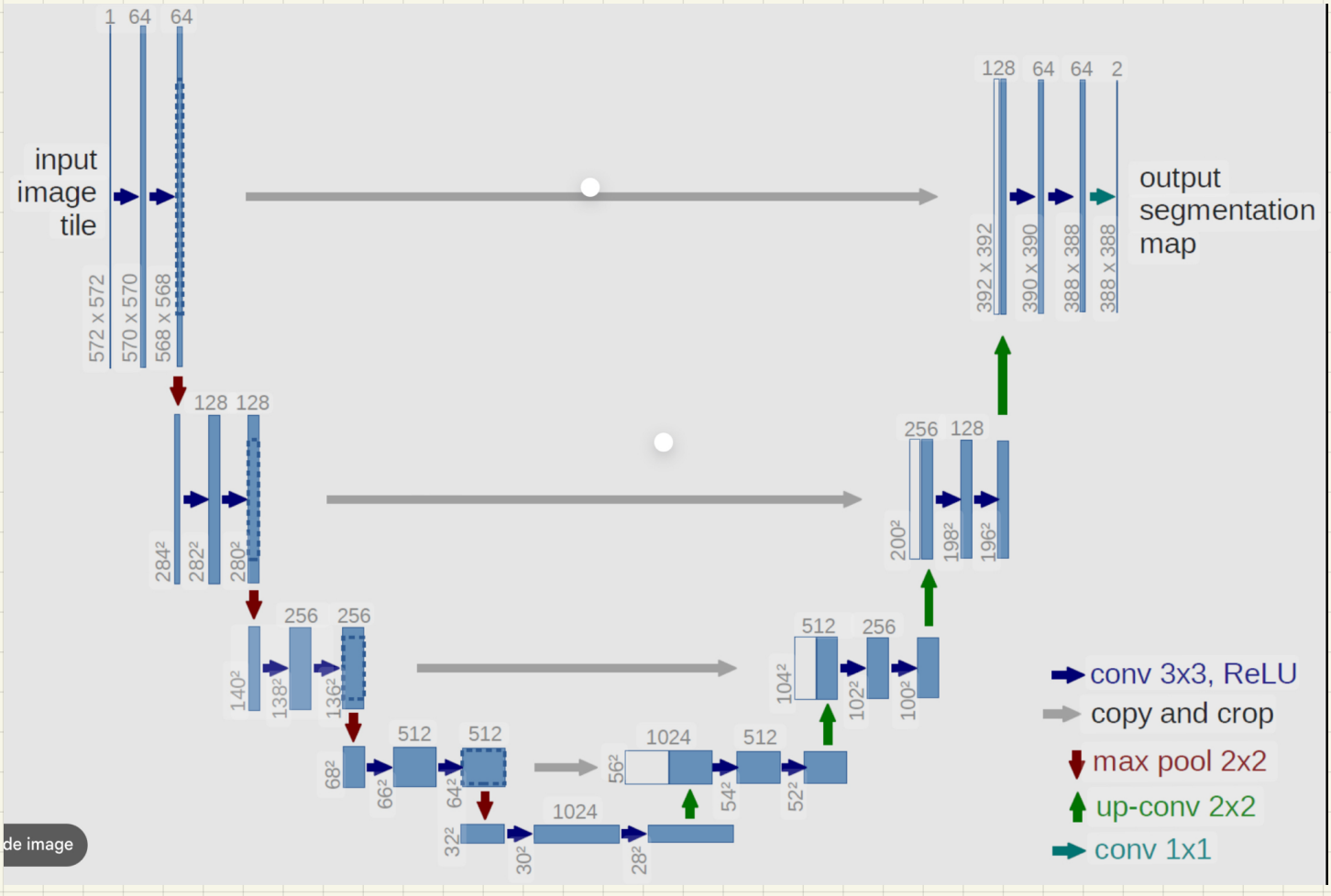
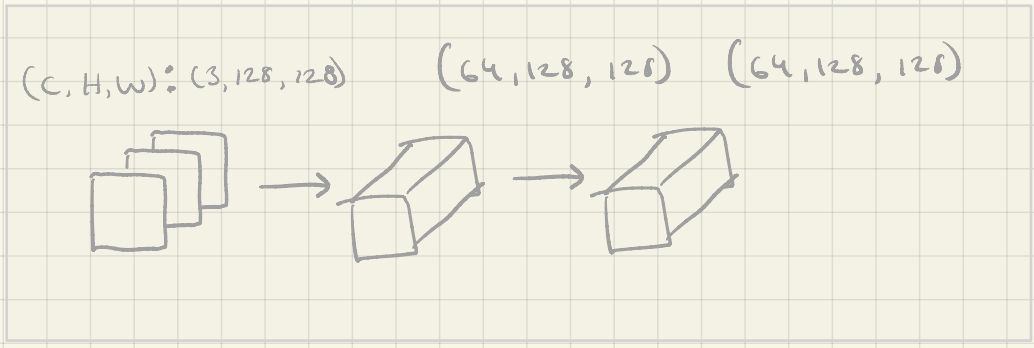

input image tile
output segmentation map
1
64
64
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
56²
1024
512
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

$$H_{out}, W_{out} = \left\lfloor \frac{H_{in} - 2p + k}{s} \right\rfloor + 1$$

1 layer of UNet



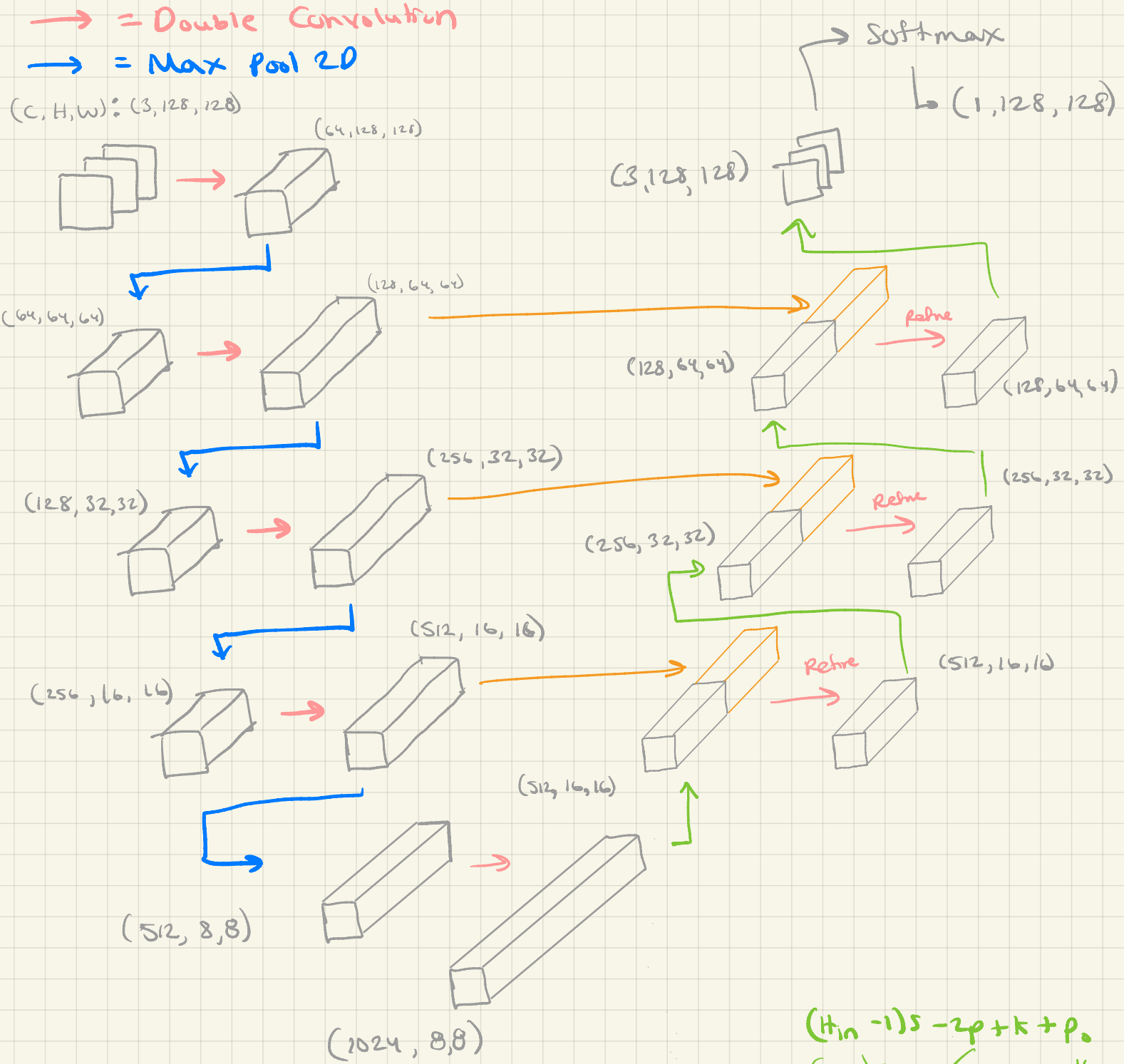
Why Double Conv?

1) increased receptive field
   3x3 → 5x5 without extra param
2) more nonlinearity → more complex feature interactions
3) Feature refinement
4) Stability & proven effectiveness

$$RF_{new} = RF_{prev} + (k-1)$$

→ = Double Convolution

→ = Max Pool 2D



$$(H_{in} - 1)s - 2p + k + p_o$$

$$(8-1)2 - 2(0) + 2 = 16$$

$$L_{CE} = -\sum \log(p_{l(x)}(x)) \quad \text{, where} \quad p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'=1}^{C} \exp(a_{k'}(x))}$$

C classes = [pet, backgound, boundary]

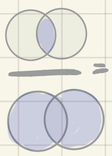
logits $a(x) = (a_1(x) \text{ — } a_C(x))$ for each pixel $x \in \Omega$

ground truth label at each pixel $l(x) \in \{1, \text{—}, C\}$

A is predicted pixels in Pet class
B is ground truth pixels in Pet class

Jaccard similarity

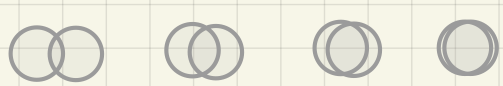
$$\text{IoU} \quad \frac{|A \cap B|}{|A \cup B|} =$$



$$= \frac{TP}{TP + FP + FN} \rightarrow [0, 1]$$

0 → 1

as IoU →



$$\text{Dice} \quad \frac{2|A \cap B|}{|A| + |B|} = \frac{2PR}{P + R} = \frac{2TP}{2TP + FP + FN}$$

precision $\quad p(Y'=1 \mid y=1) \quad$ y=1 / y'=1 $\quad \frac{TP}{TP + FP}$

Recall $\quad p(Y=1 \mid Y'=1) \quad \frac{TP}{TP + FN}$

Overlapping Kernels Sum

$(H_{in}-1)s-2p+k+p_o$

$W_{in} H_{in} = (4, 4)$

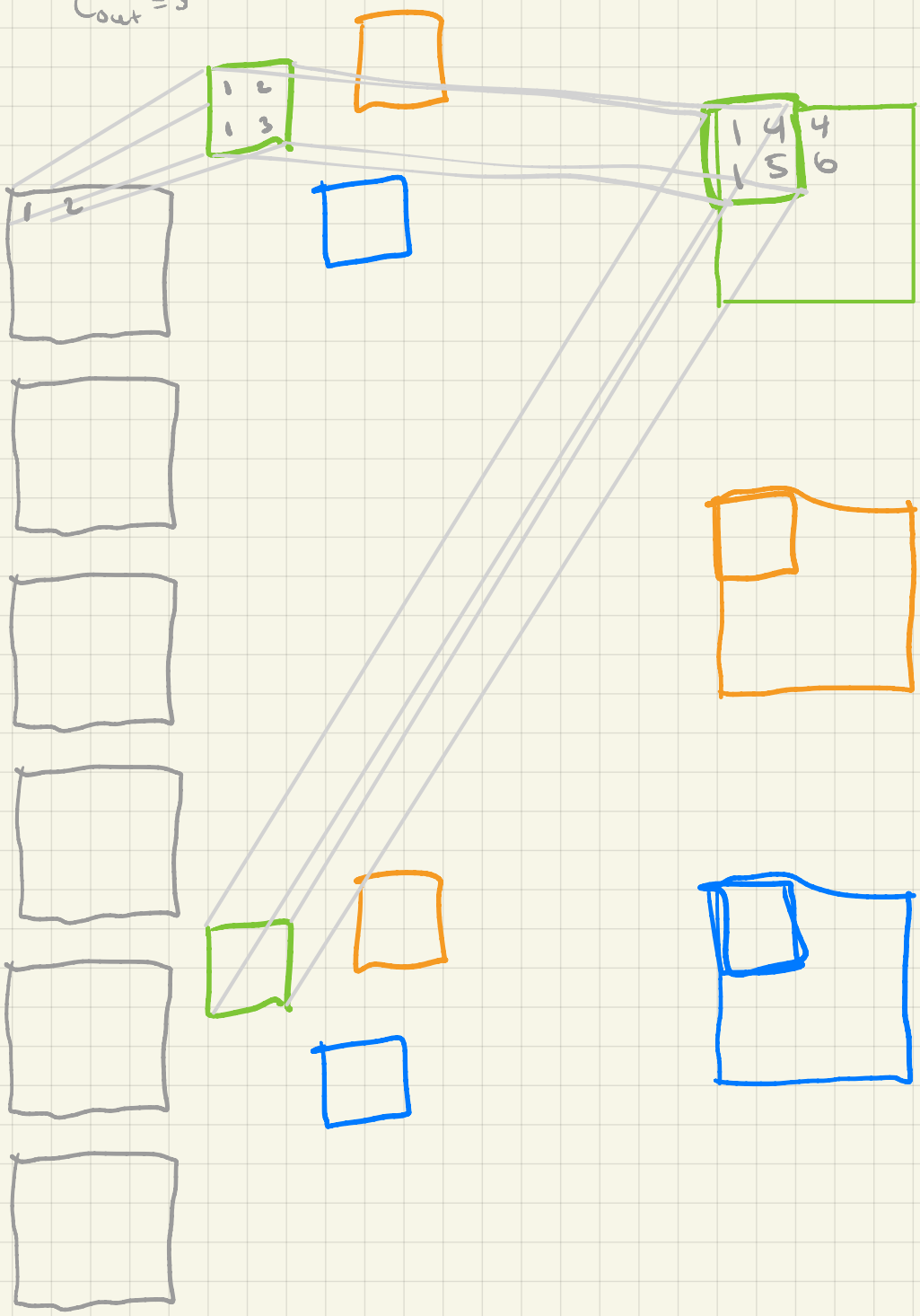
$(4-1)1-2(0)+2=5$

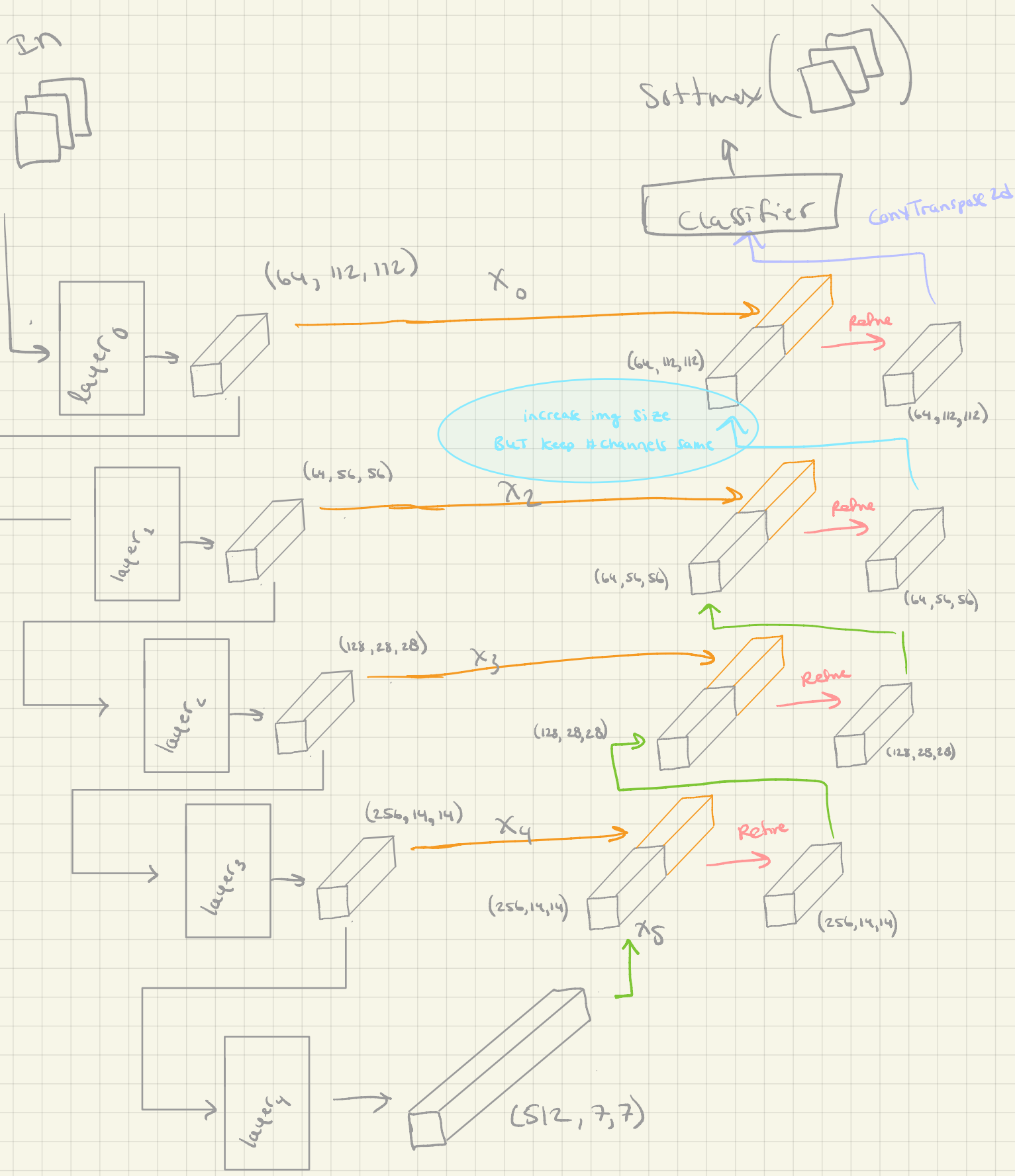
$3 \cdot 1 + 2 =$

$C_{in} = 6$

$C_{out} = 3$

Kernel = 2
stride = 1
padding = 0




Softmax
Classifier
ConvTranspose 2d
layer0
layer1
layer2
layer3
layer4
(64, 112, 112)
x0
Relu
(64, 112, 112)
(64, 112, 112)
increase img size
BUT keep # channels same
(64, 56, 56)
x2
Relu
(64, 56, 56)
(64, 56, 56)
(128, 28, 28)
x3
Relu
(128, 28, 28)
(128, 28, 28)
(256, 14, 14)
x4
Relu
(256, 14, 14)
(256, 14, 14)
x5
(512, 7, 7)