RCNN Siblings

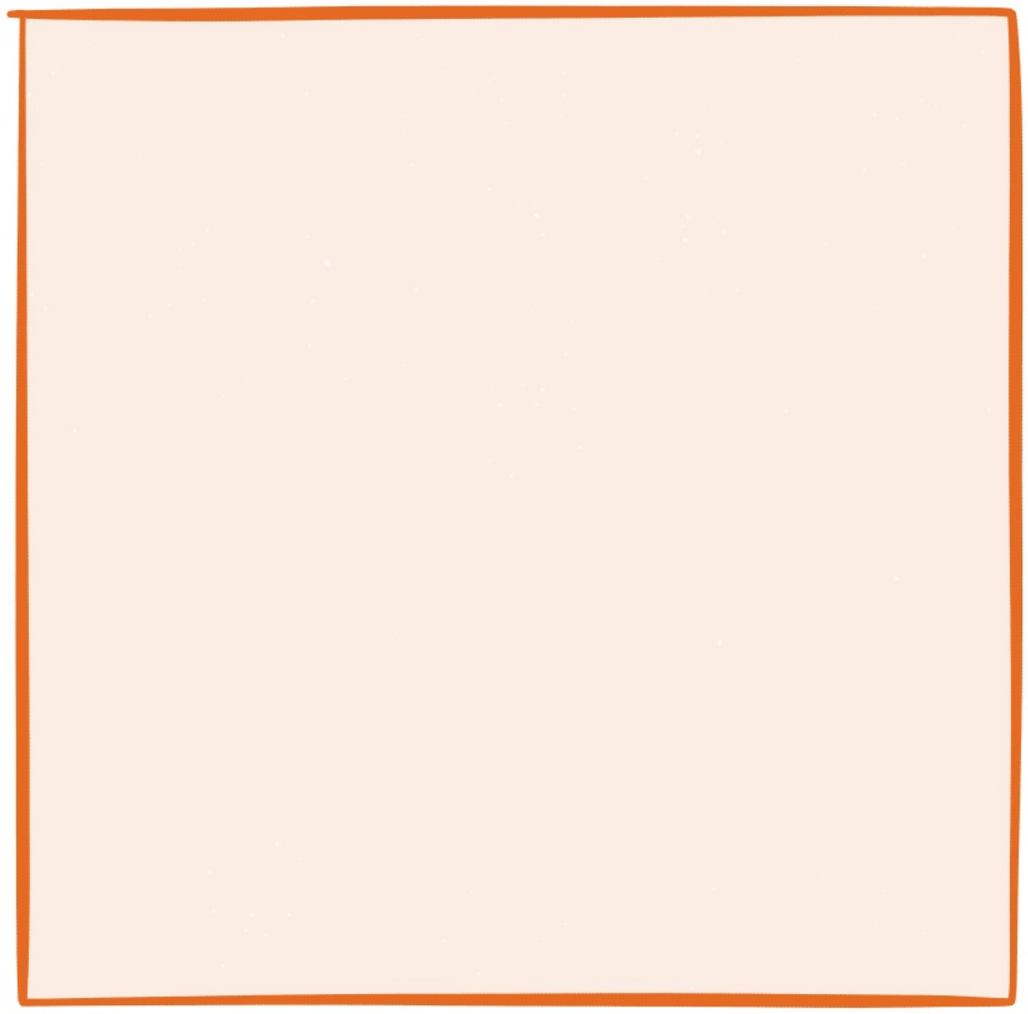
R CNN

SPPnet

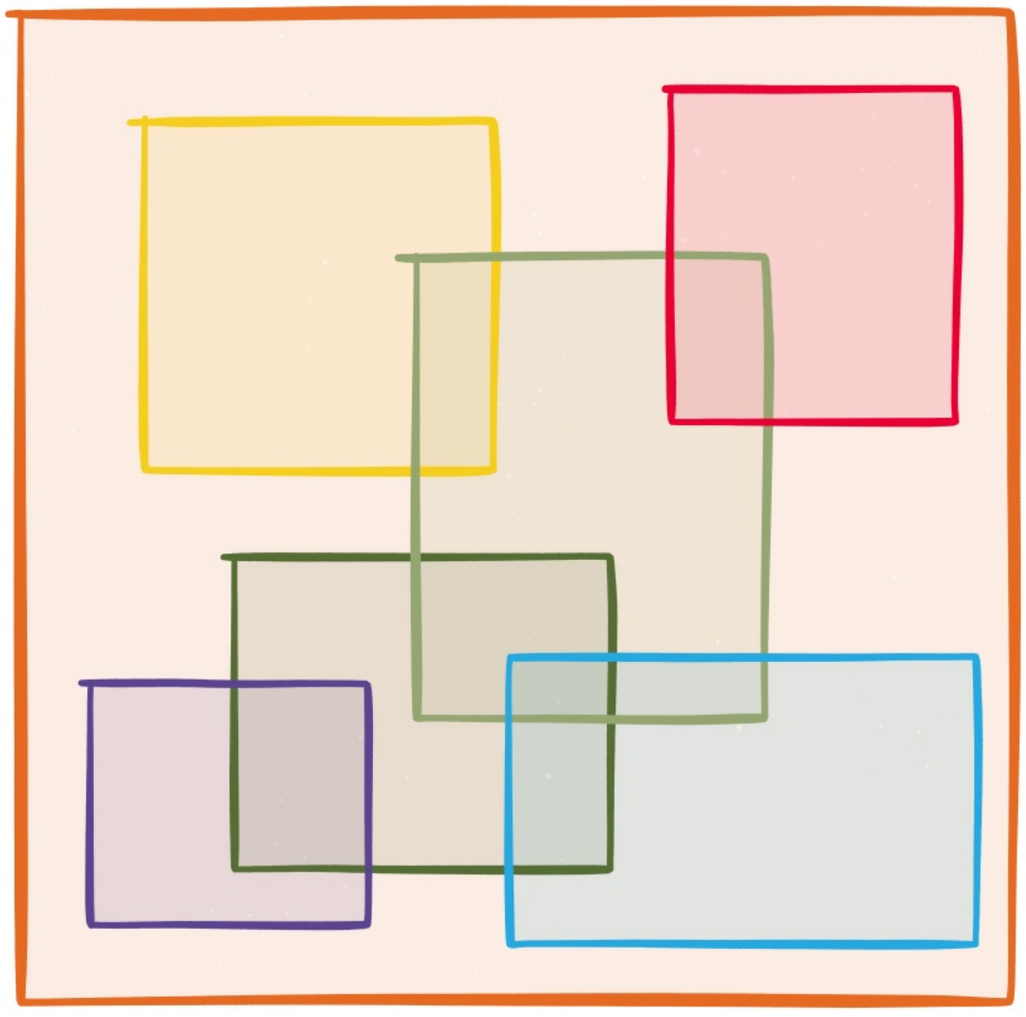
Fast R CNN

Faster R CNN

RCNN Siblings



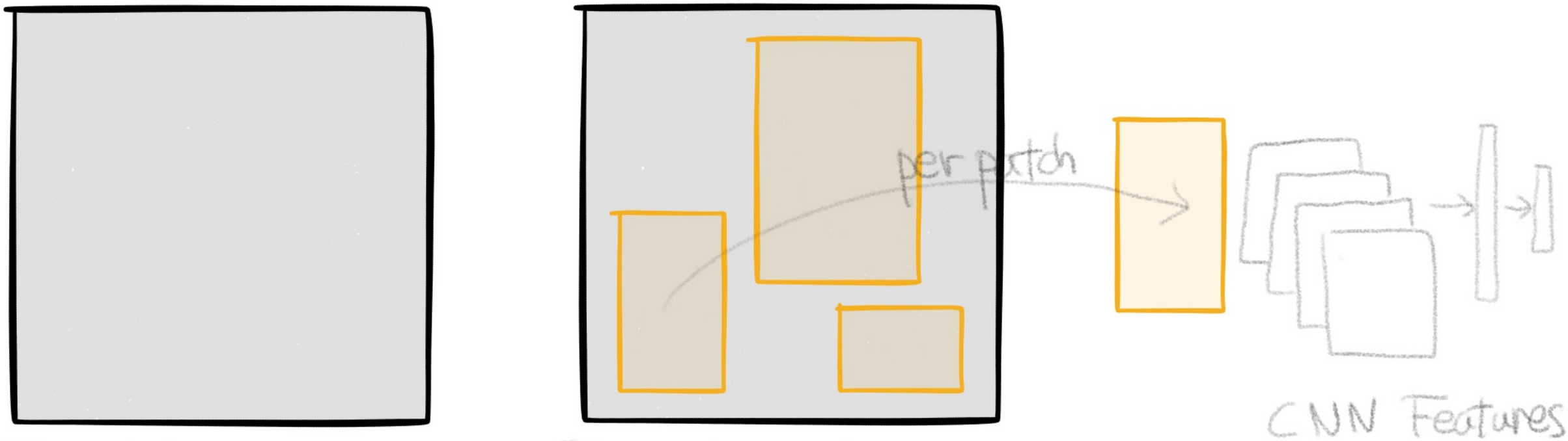
1000



20 + 1

Classification vs. Detection

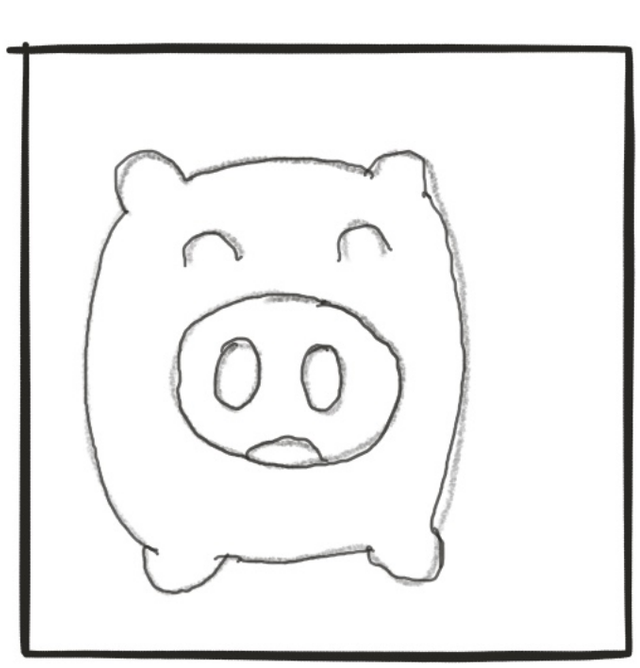
R-CNN Architecture

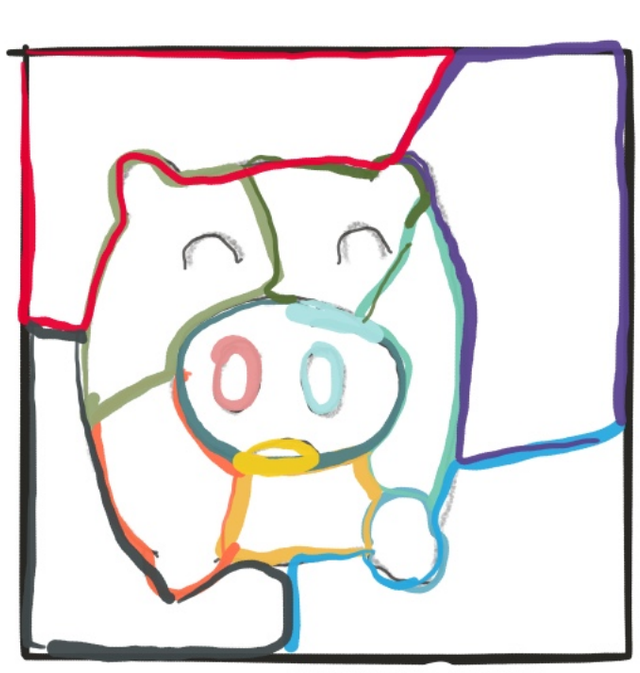


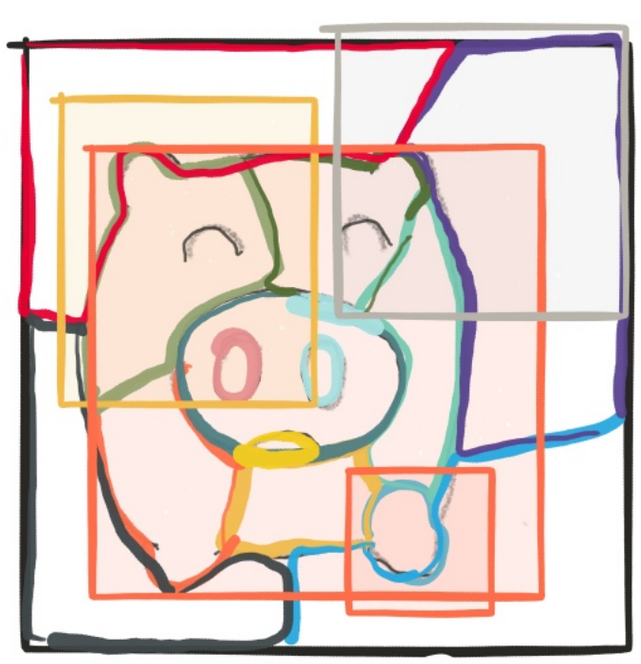
R CNN

Step 1 : Region proposal

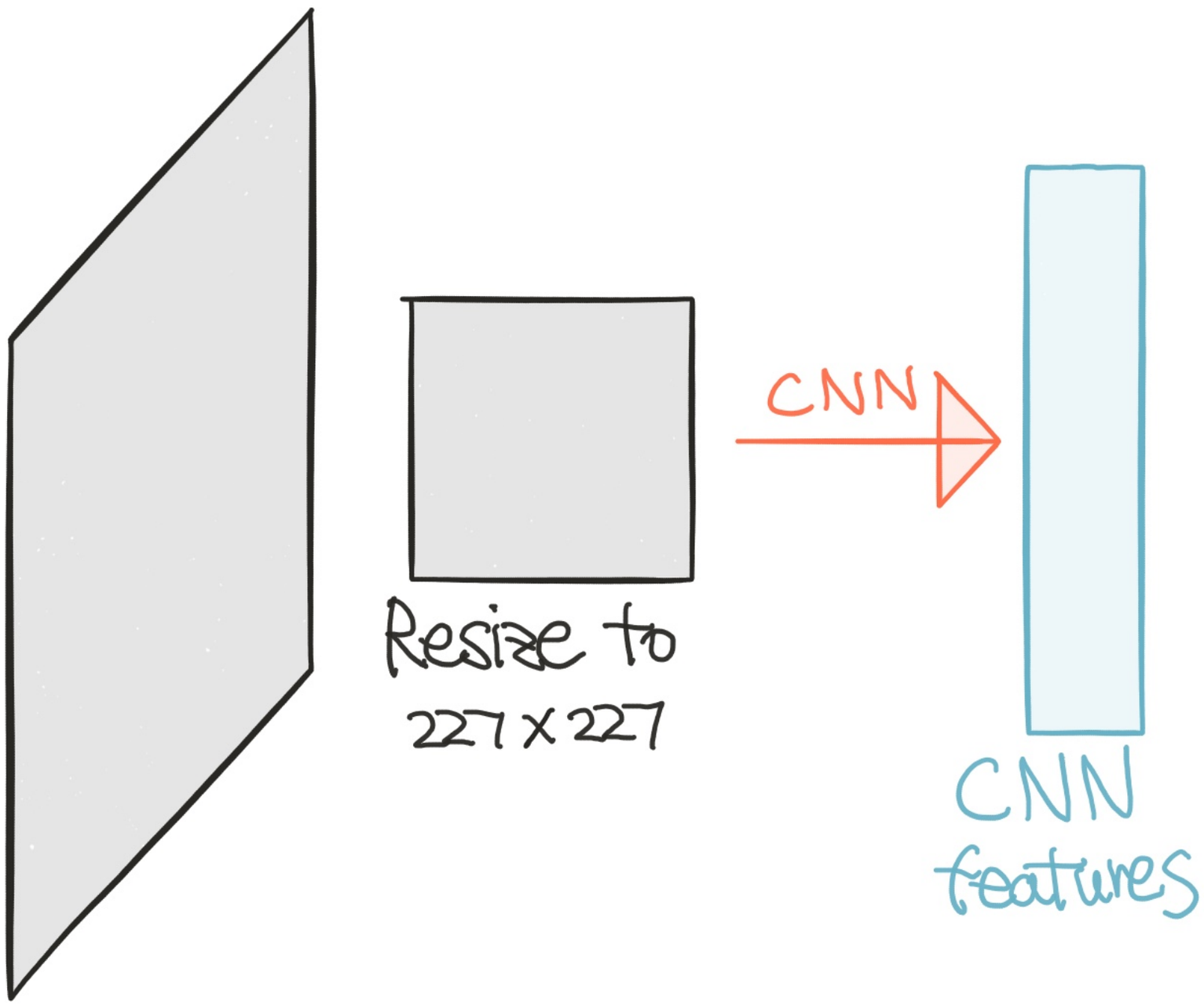
Selective search method has been used



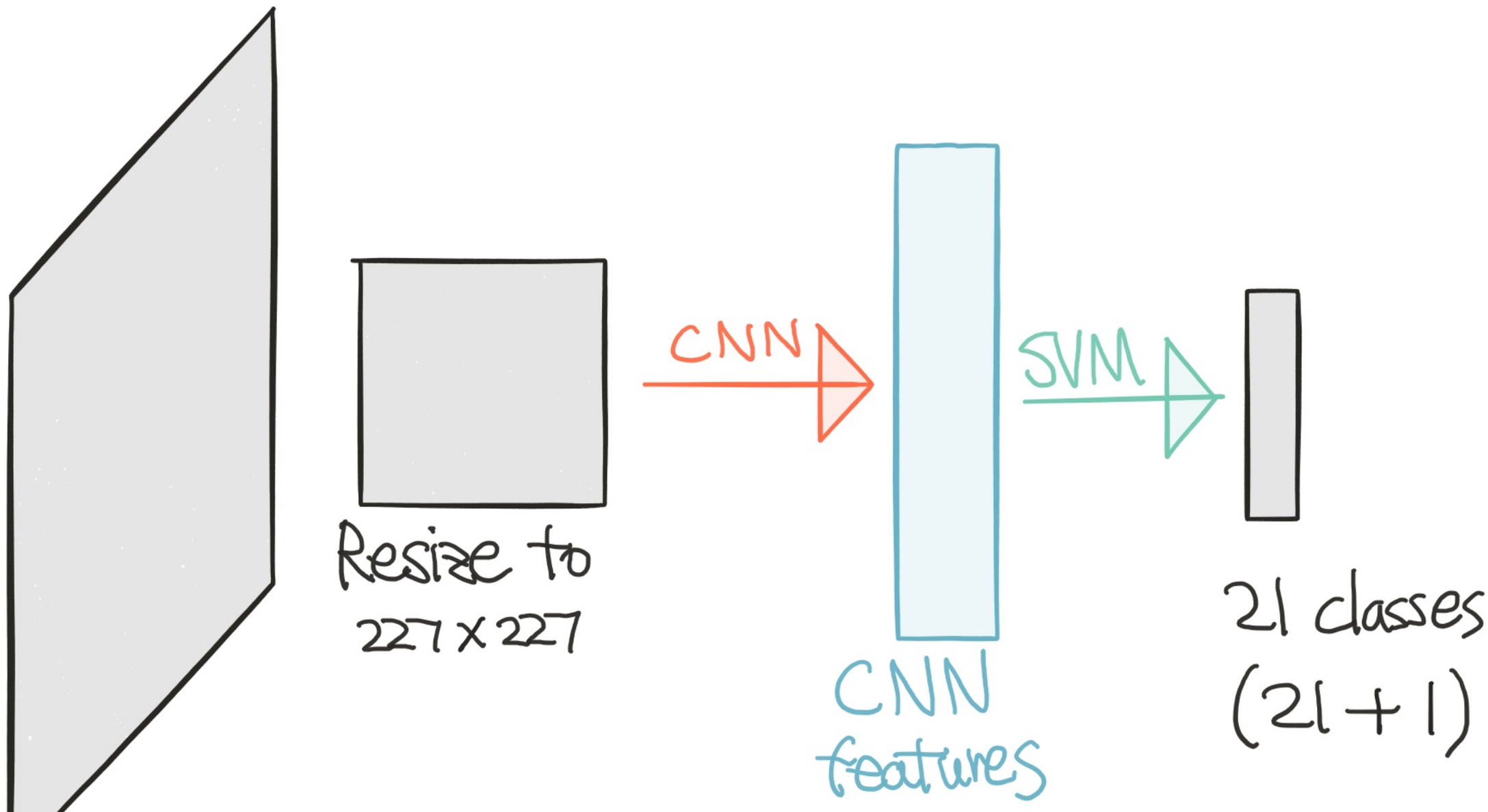




Region Proposal



CNN Feature Computation


Step 3 : Linear SVM
Resize to
227 x 227
CNN
CNN
features
SVM
21 classes
(21 + 1)

Linear SVM

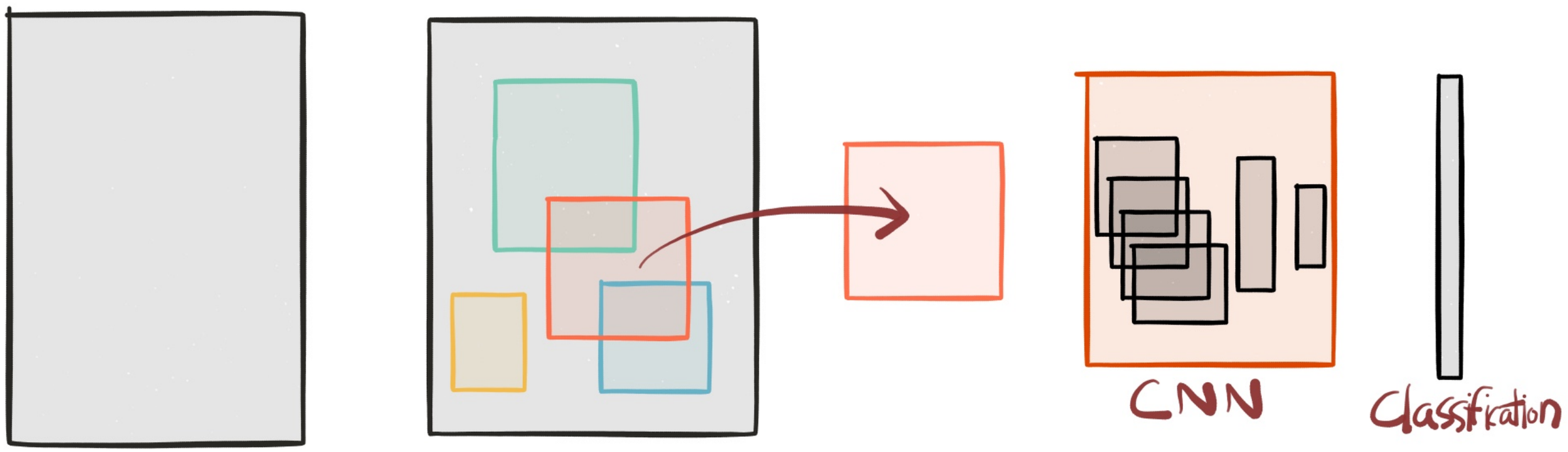
Training Phase

1. Pre traing

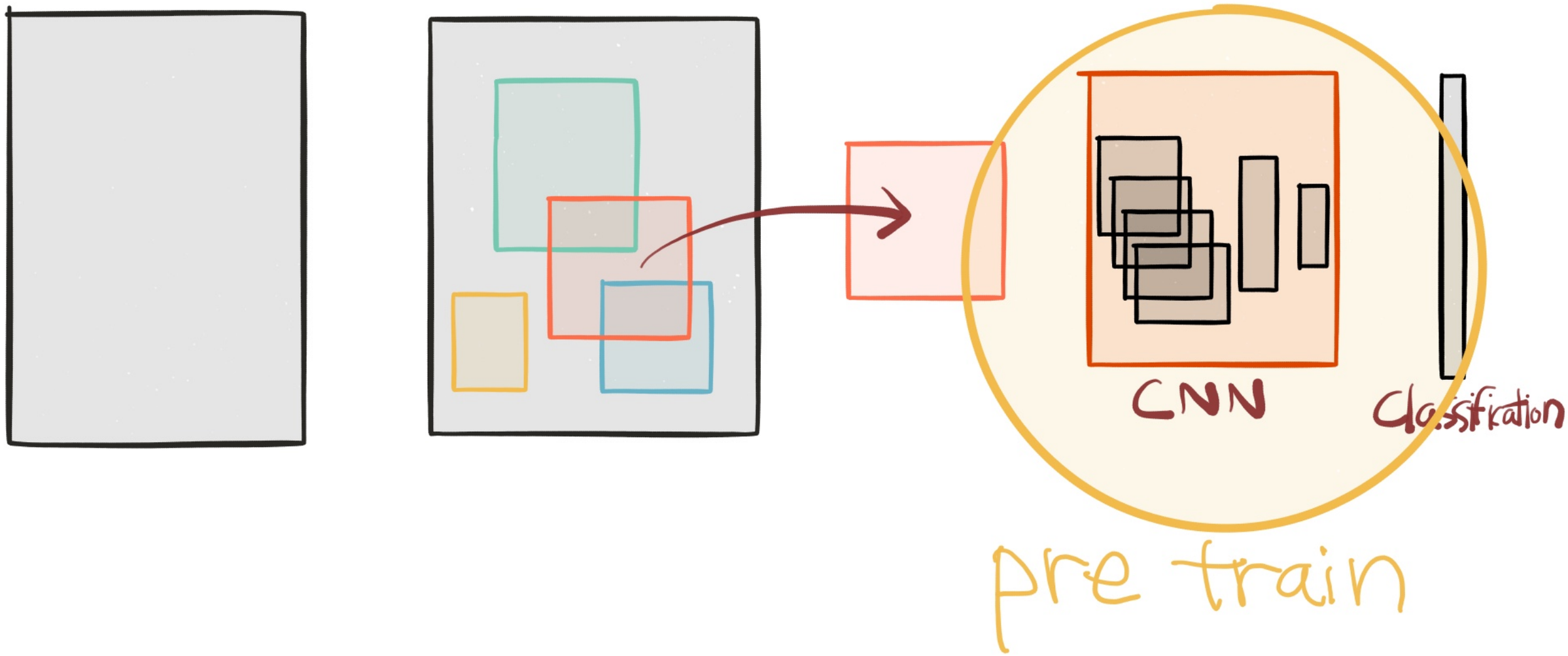
2. Fine tuning

3. Bounding Box regression

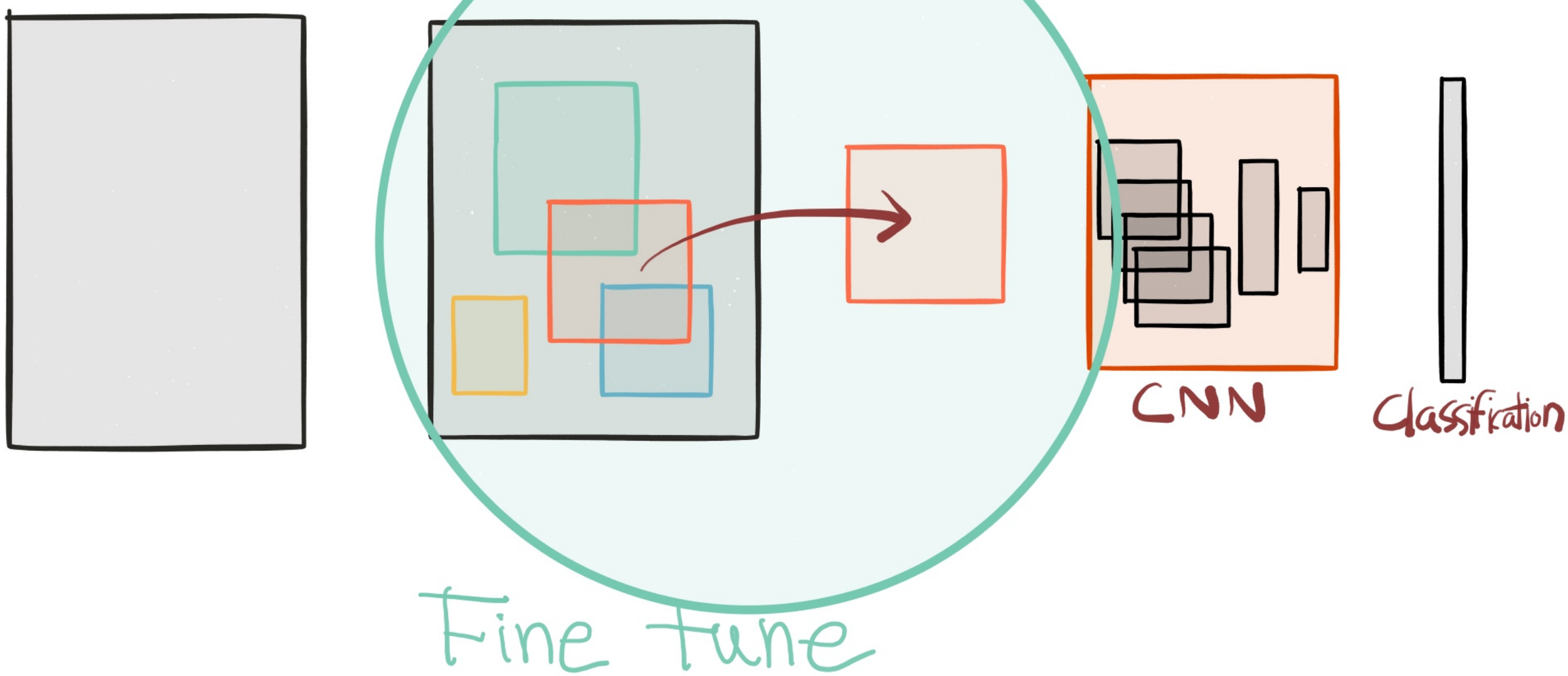
Training RCNN



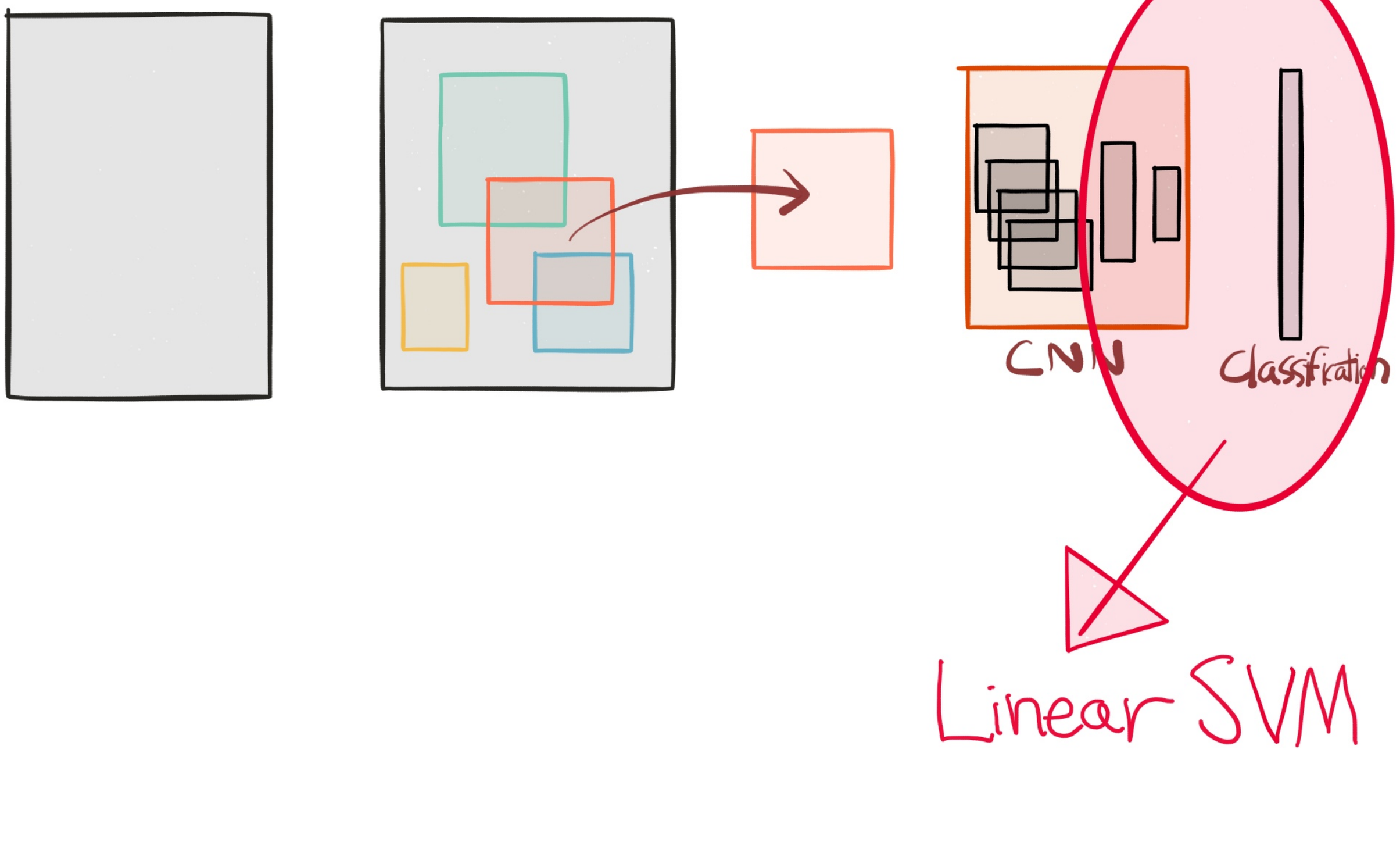
Training RCNN



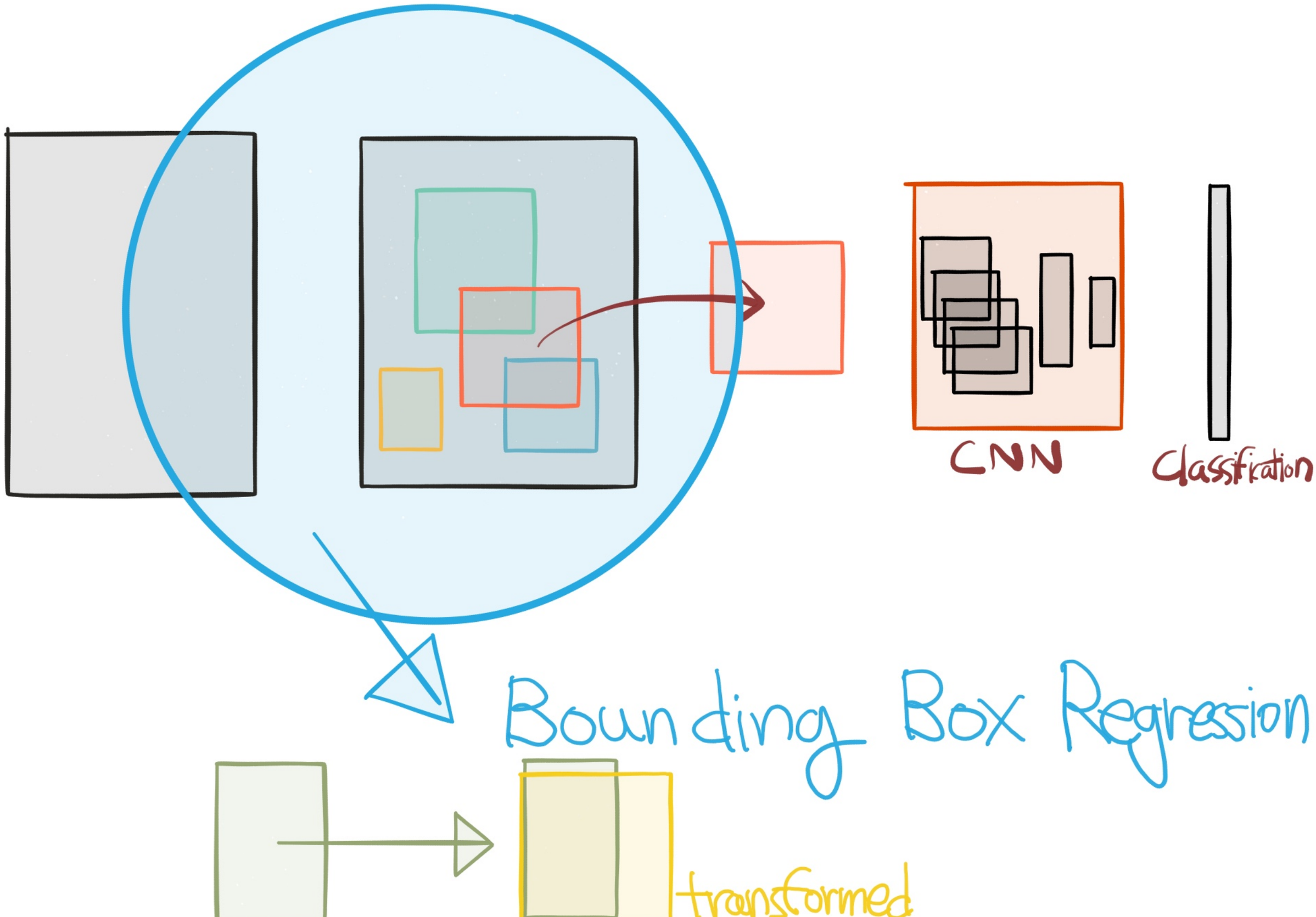
Pretrain



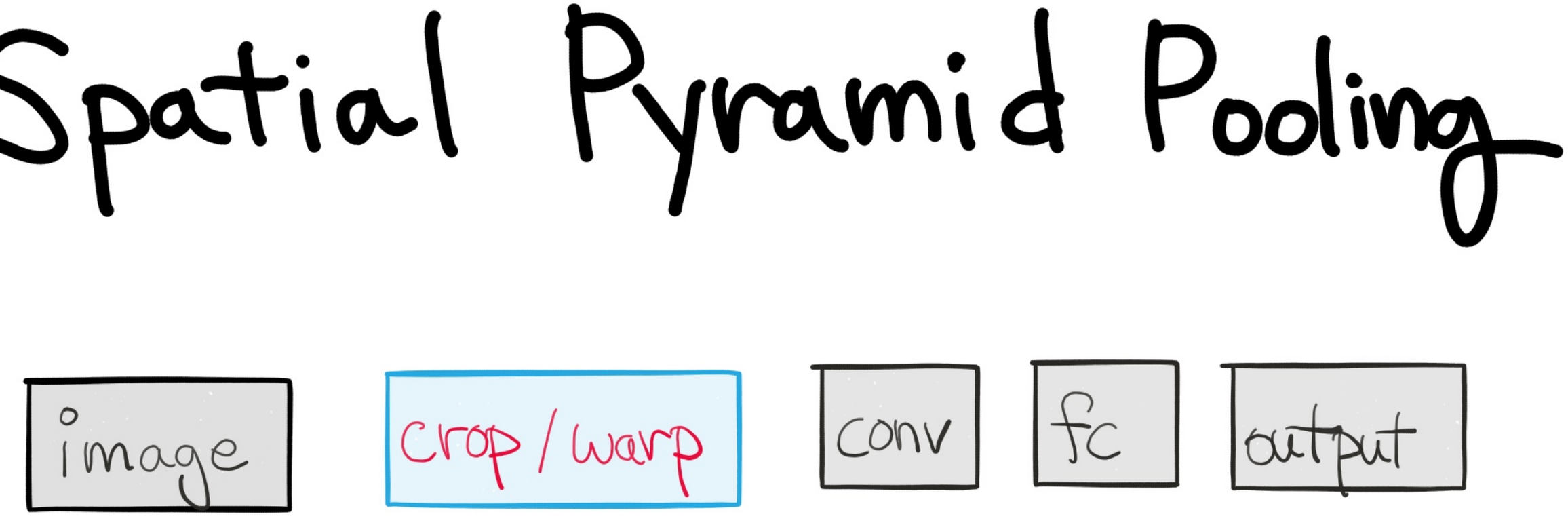
Fine tuning



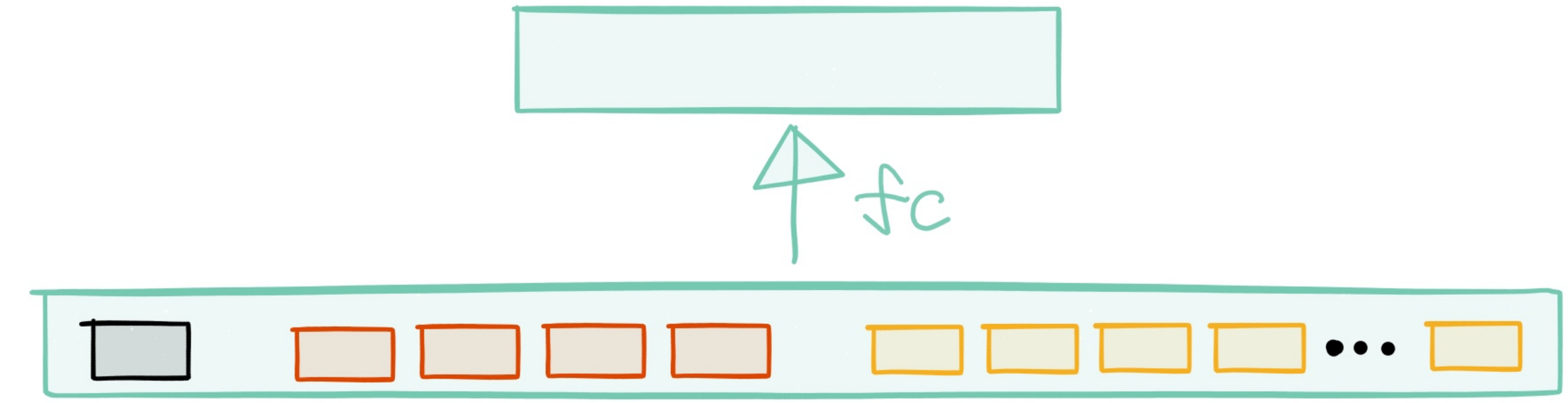
Train SVM



Bounding Box Regression



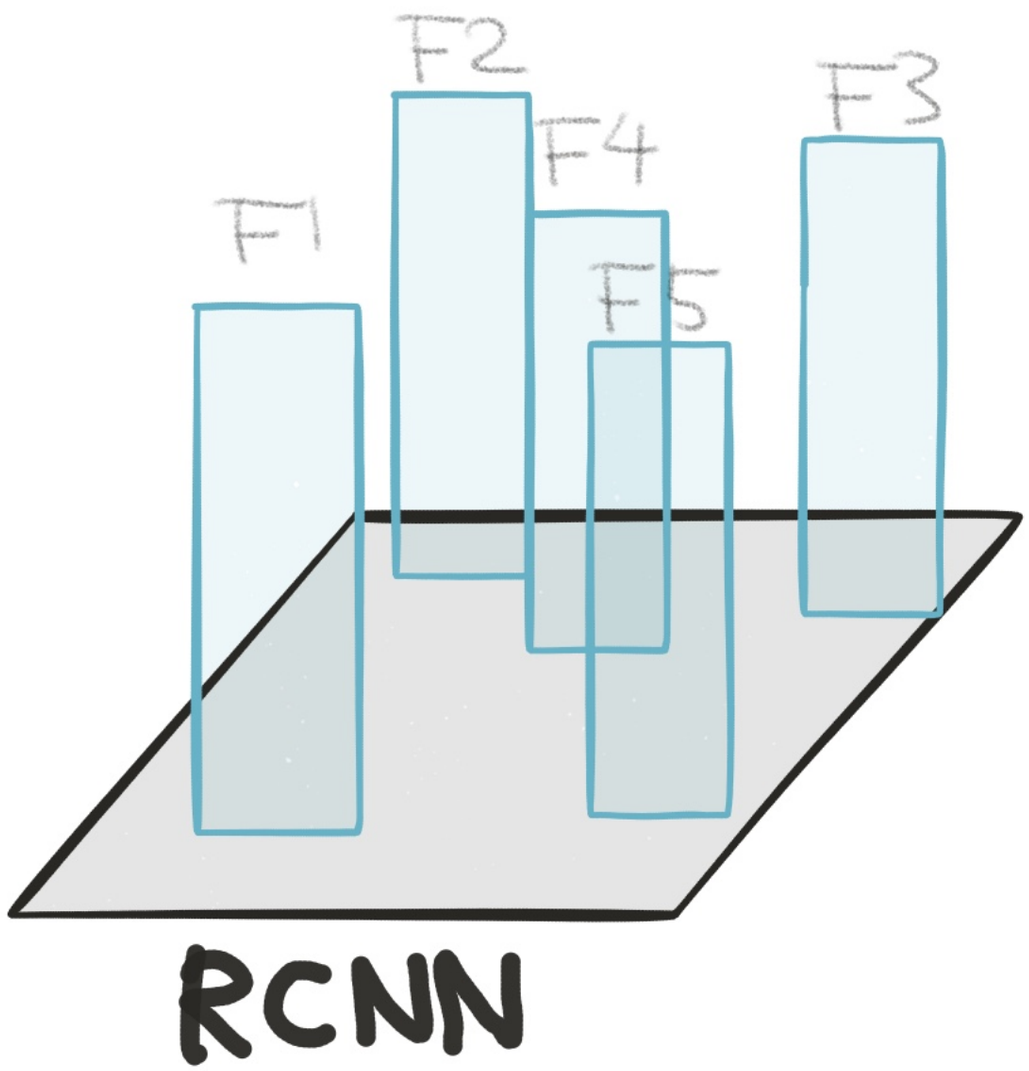
Spatial Pyramid Pooling



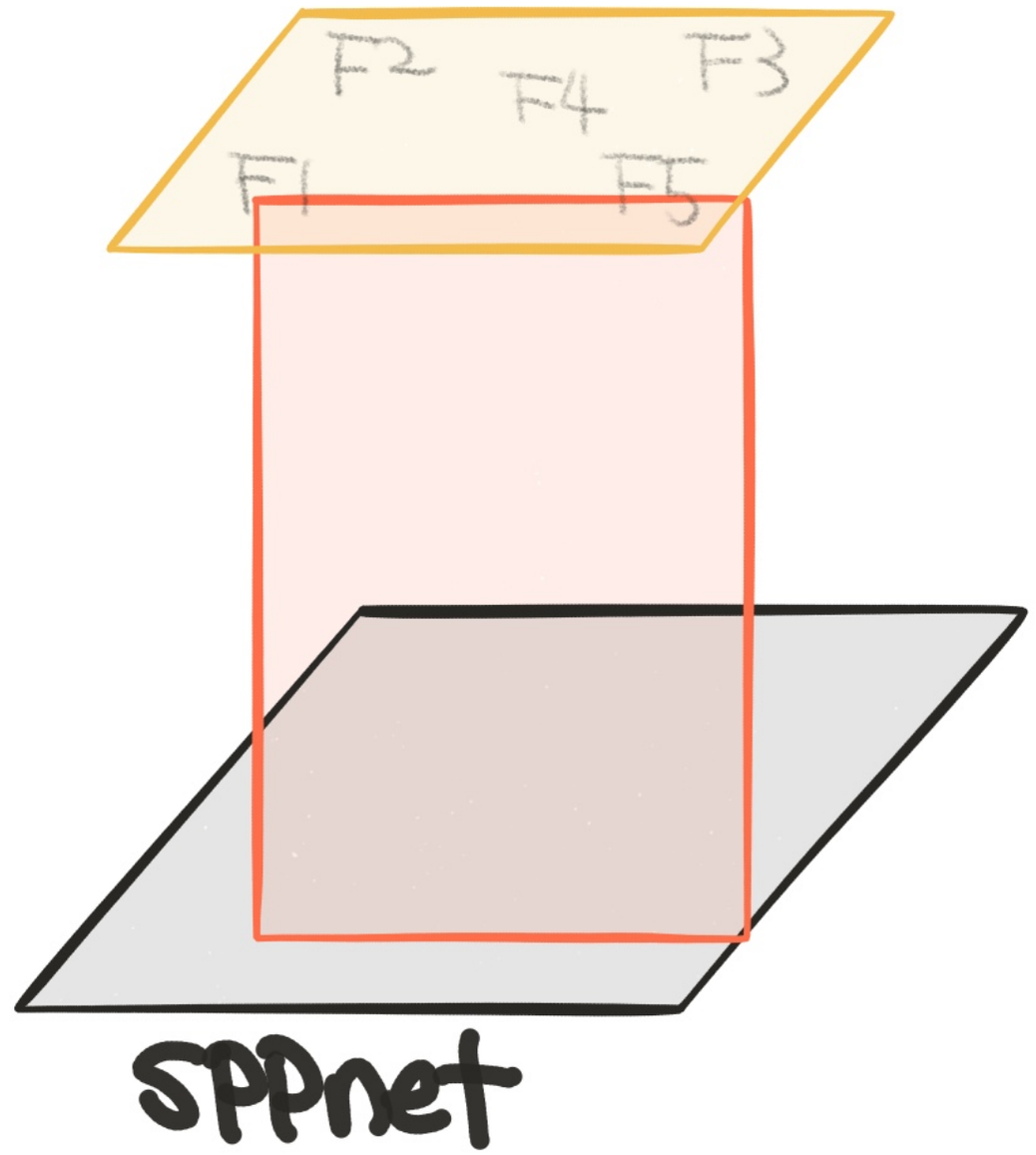
Spatial Pyramid Pooling

Layer

SPPnet vs. RCNN



RCNN

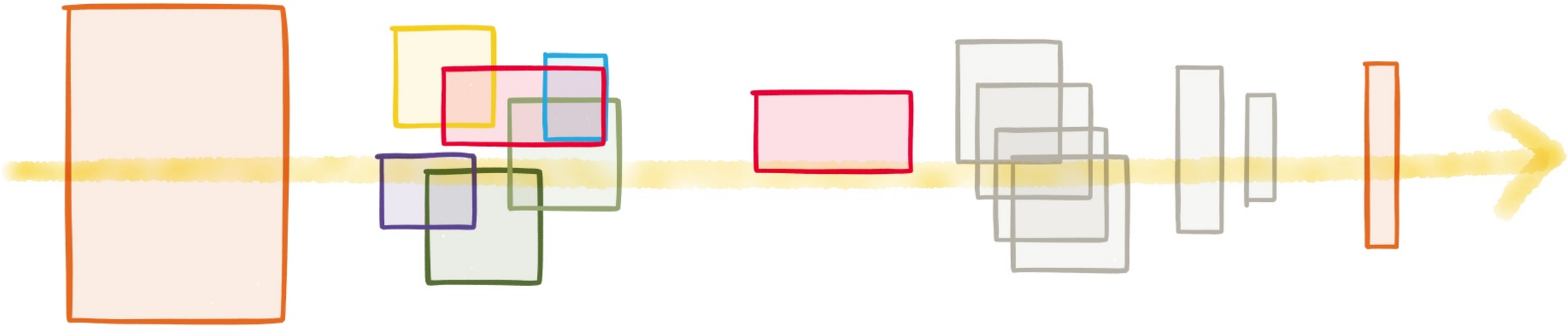


SPPnet

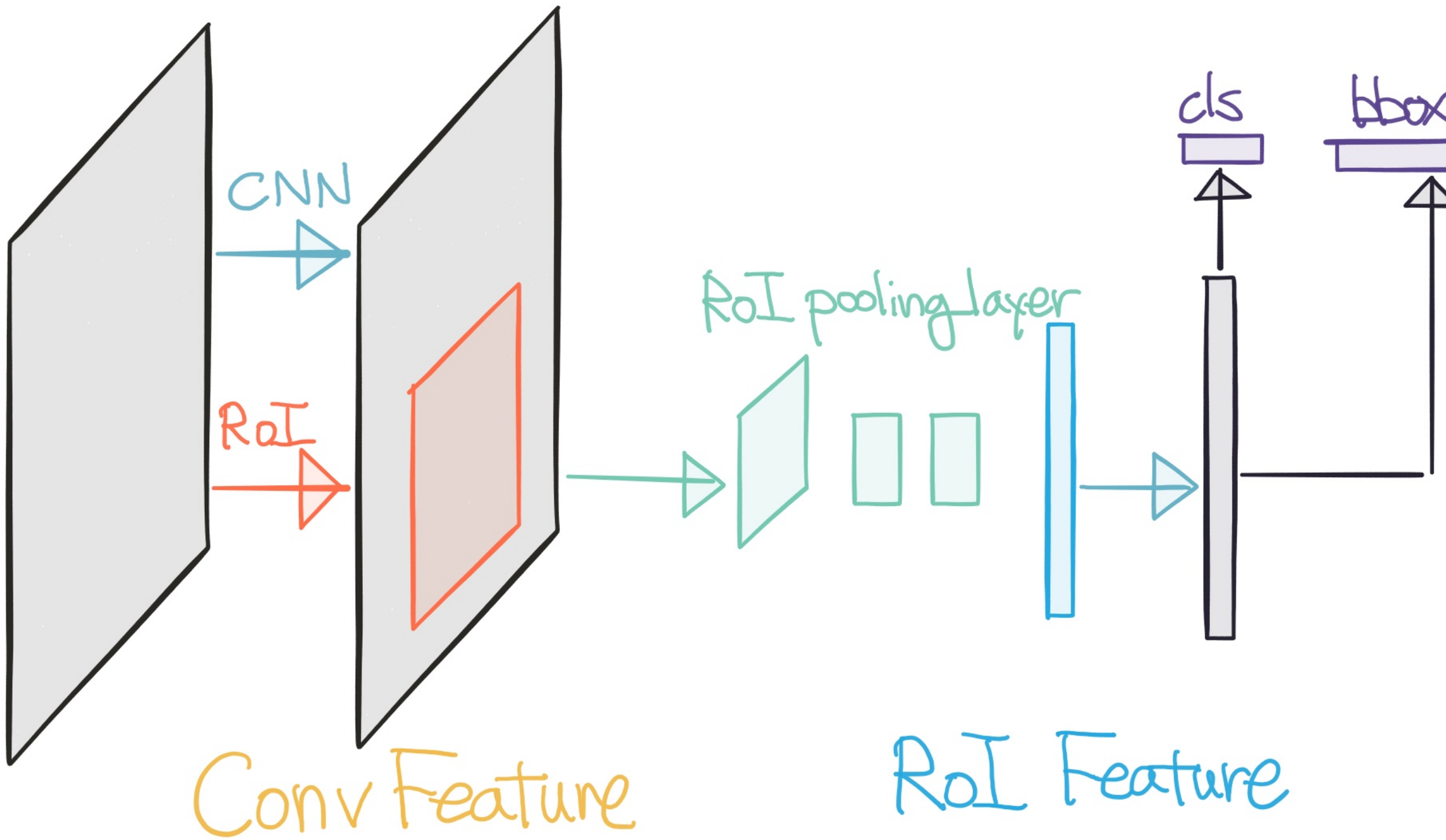
SPPnet vs. RCNN

Fast R-CNN

Drawbacks of R-CNN

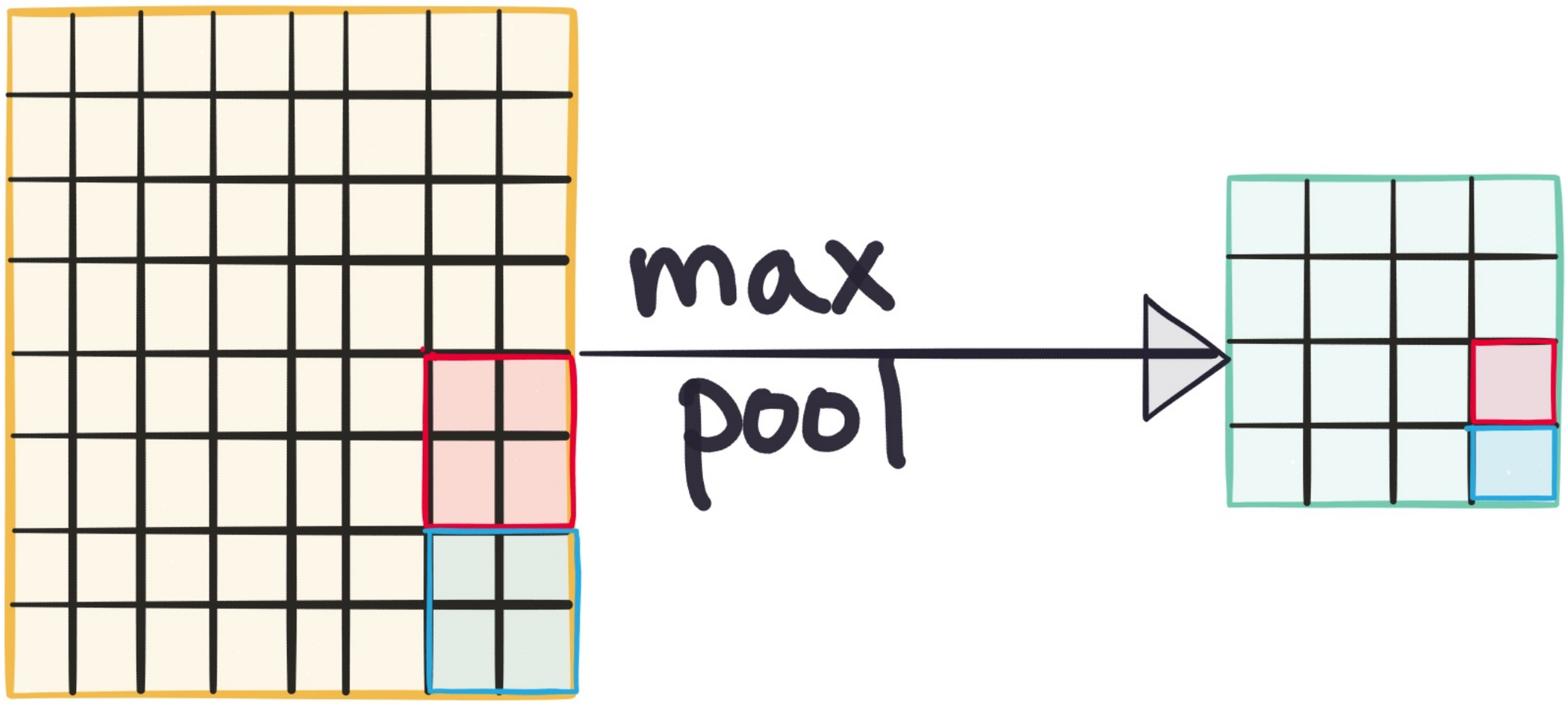


Fast RCNN



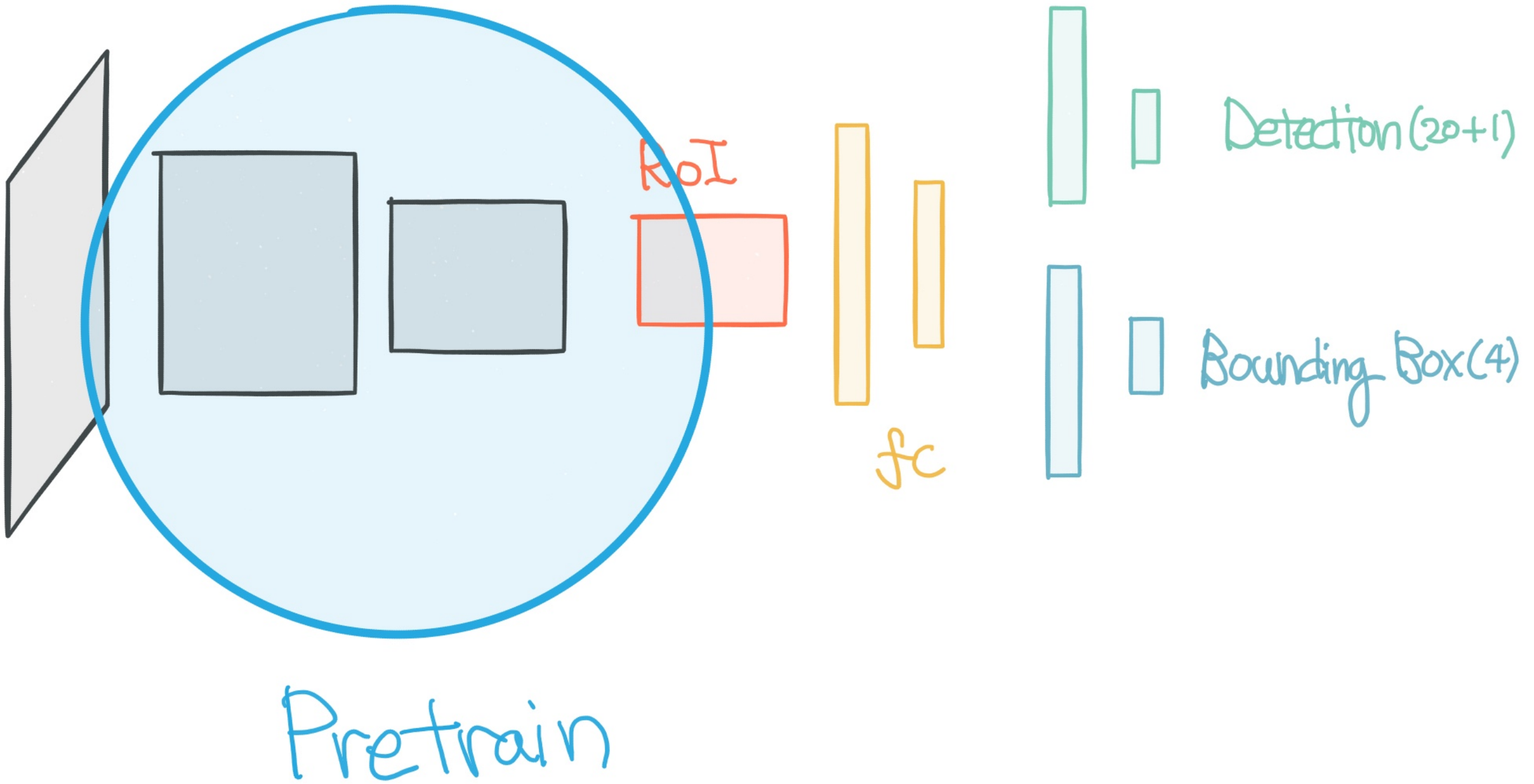
Architecture of Fast RCNN

Fixed size feat. map

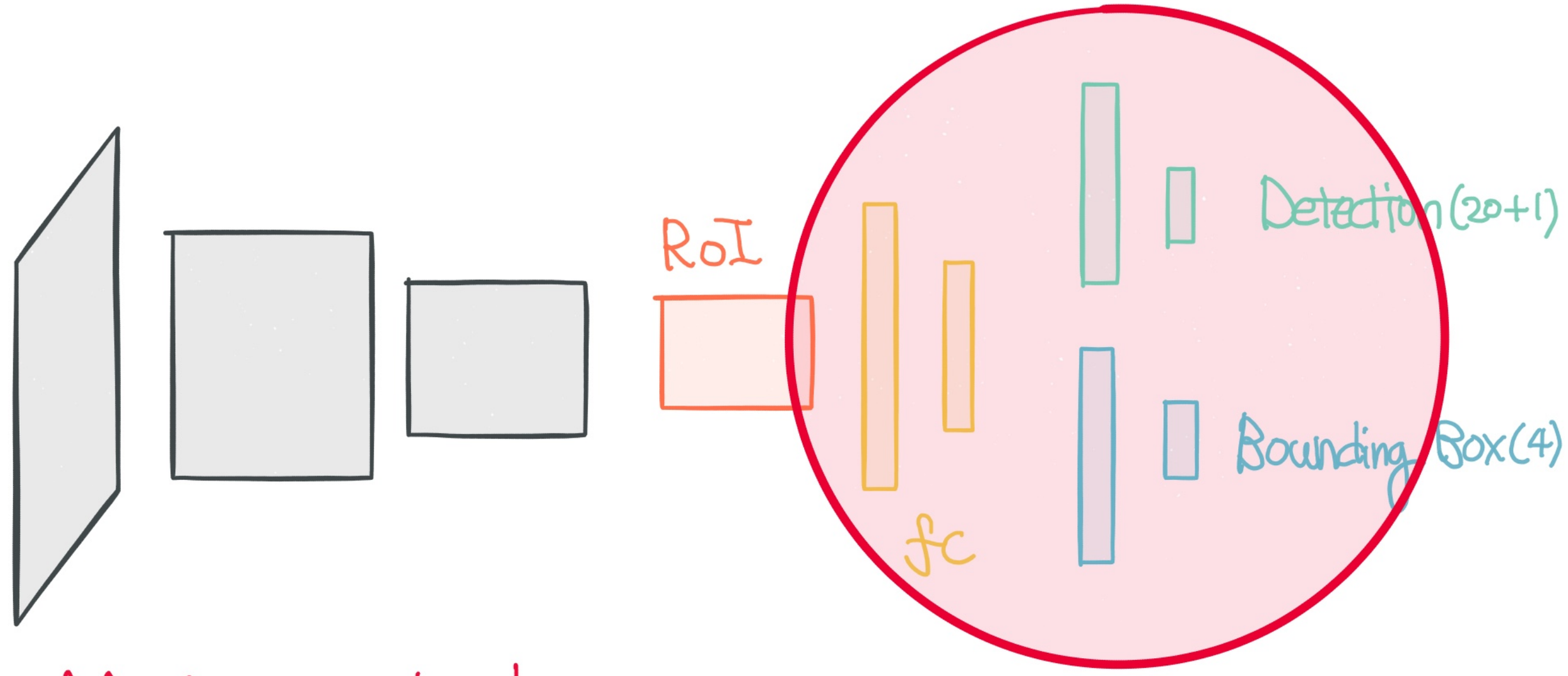


RoI Pooling Layer

Training Fast RCNN



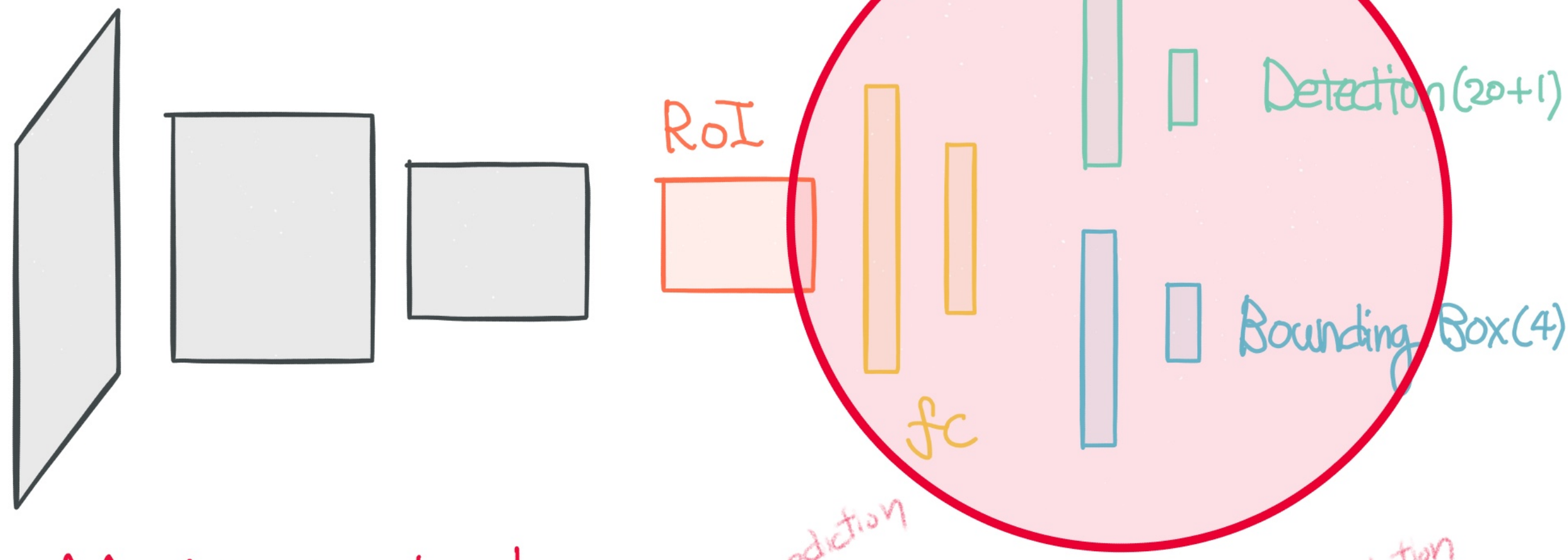
Training Fast RCNN



Multi task loss

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda [u \geq 1] L_{loc}(t^u, v)$$

Training Fast RCNN

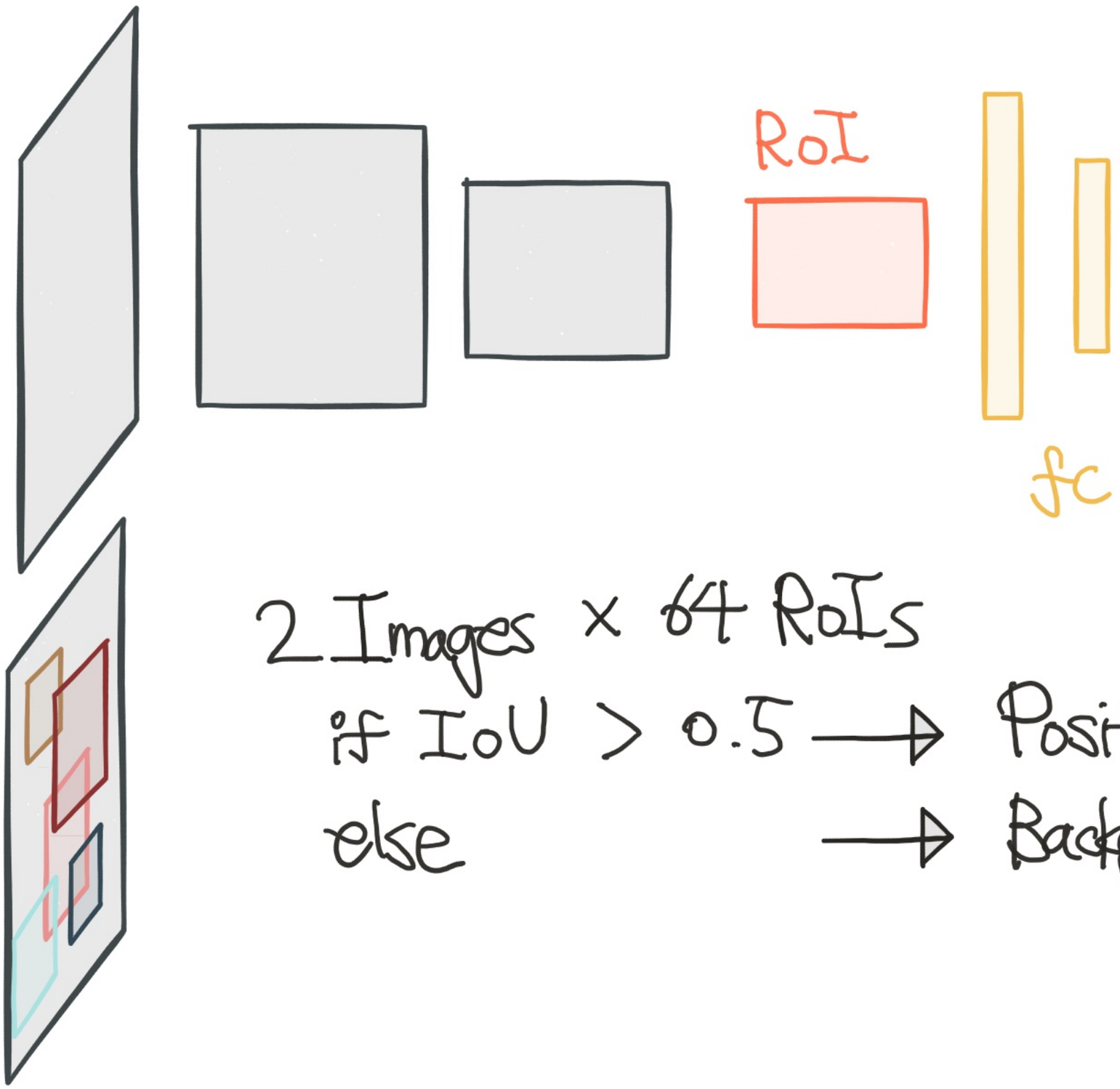


Multi task loss

$$L(p, u, t^u, v) = \underbrace{L_{cls}(p, u)}_{\text{classification}} + \underbrace{\lambda [u \geq 1] L_{loc}(t^u, v)}_{\text{localization}}$$

class prediction
target class
Ignore BG class
BB prediction
target BB

Training Fast RCNN

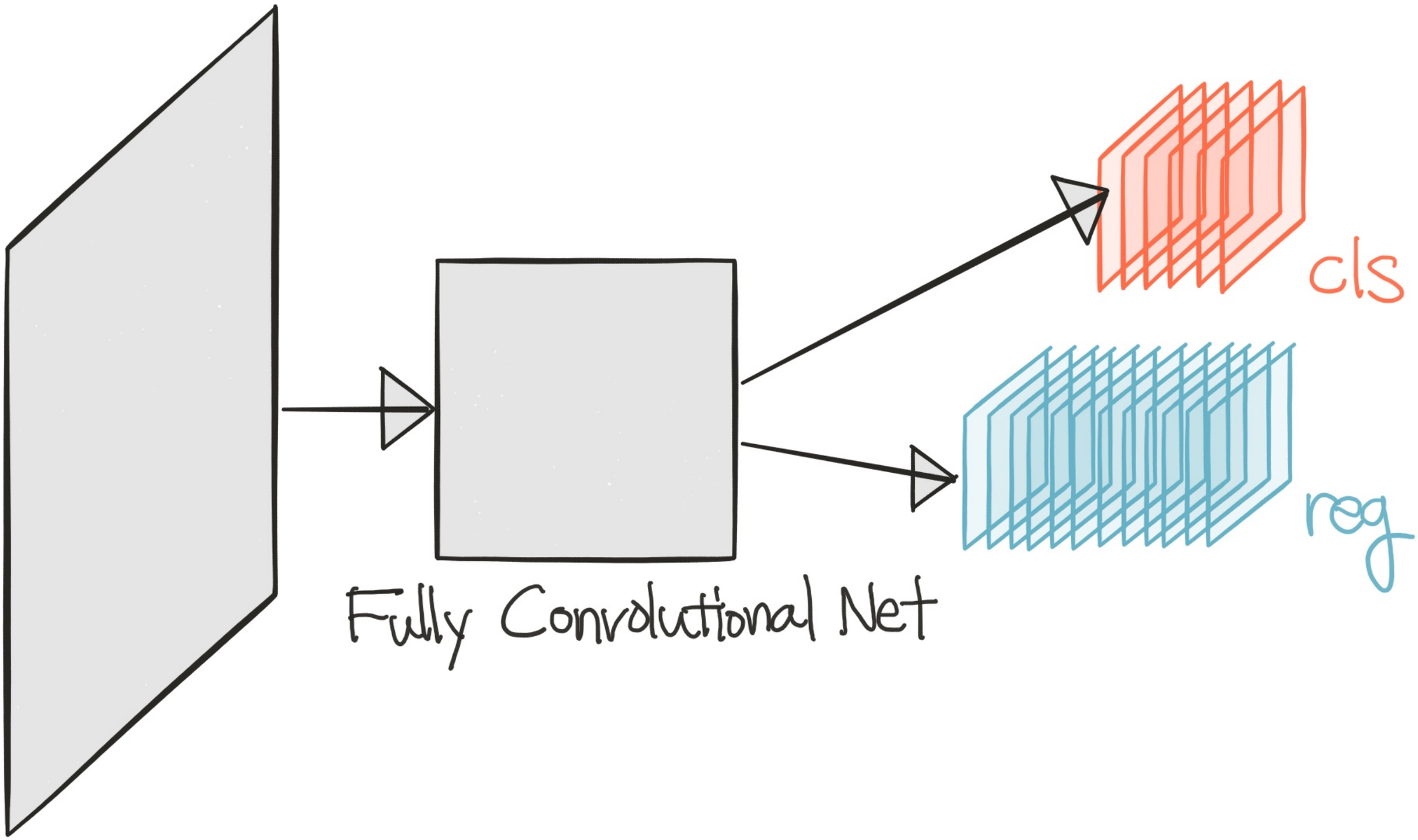


2 Images x 64 RoIs

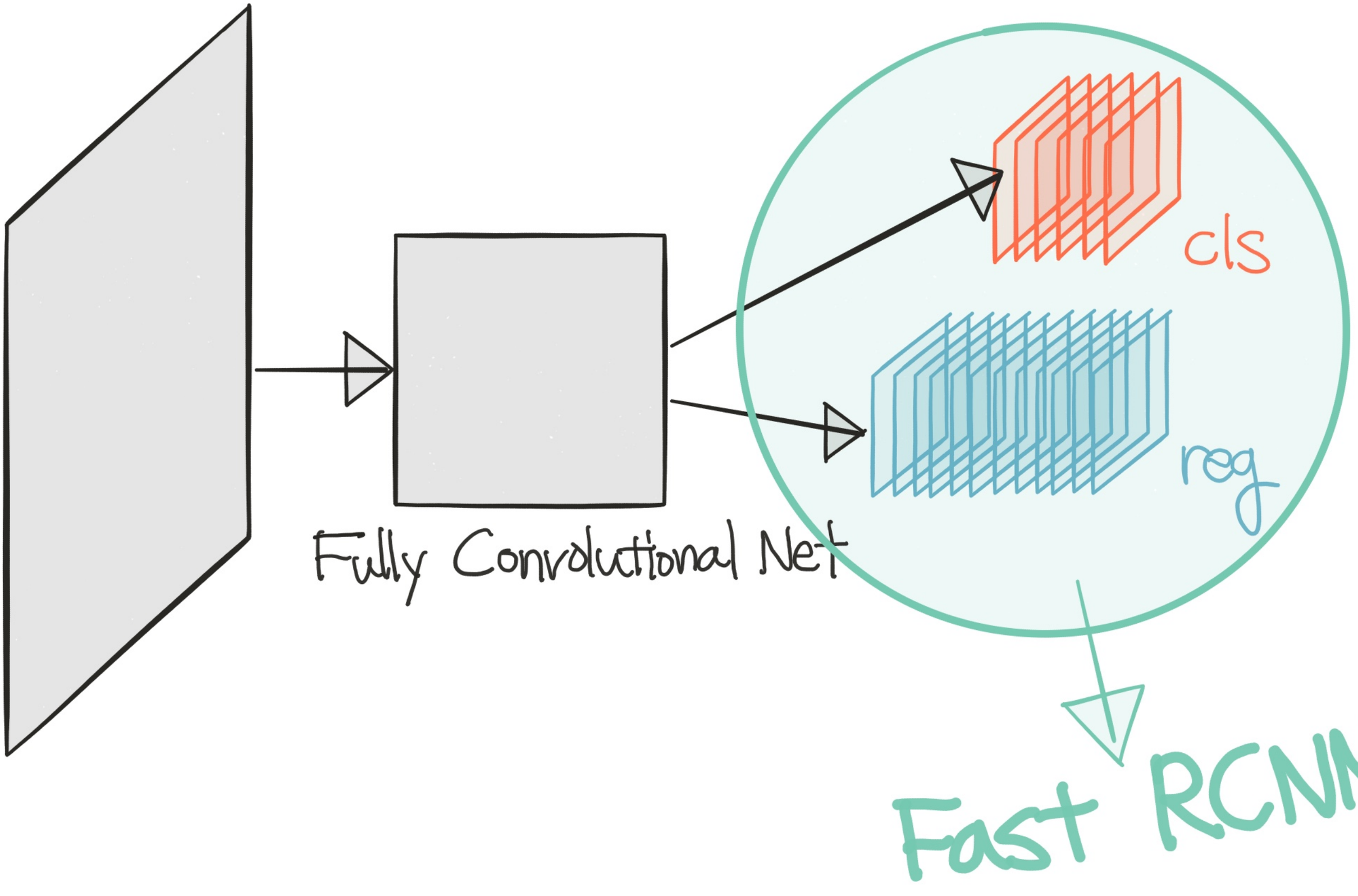
if IoU > 0.5 ⟶ Positive (with label)

else ⟶ Background

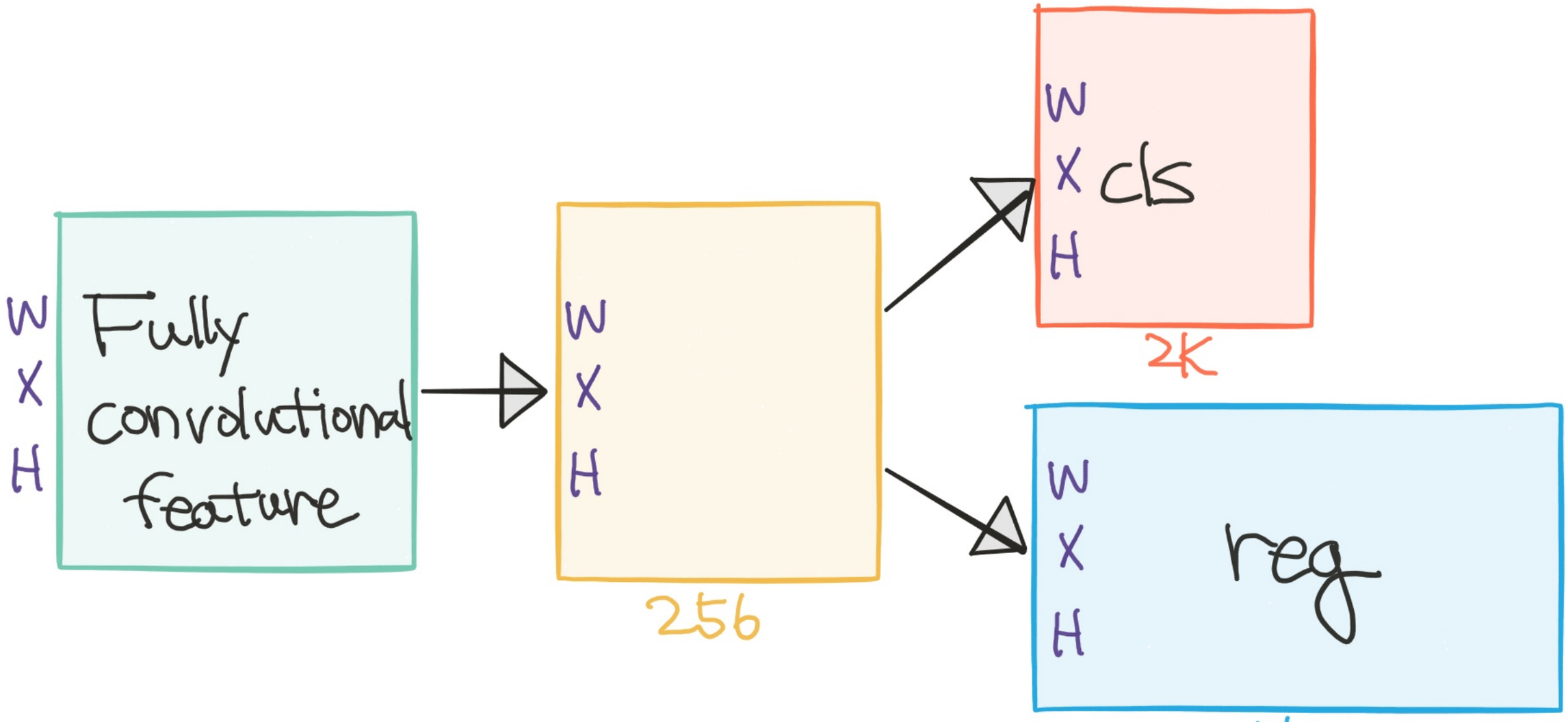
Training Fast RCNN



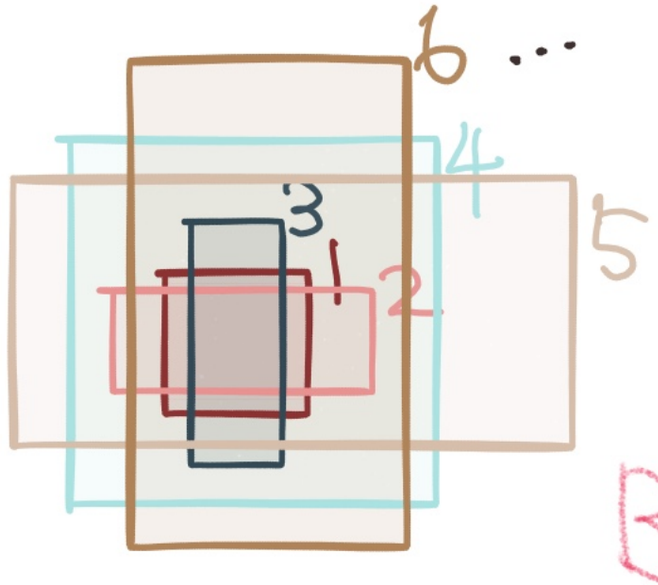
Faster R CNN!



Faster R CNN!



k=9



Bounding Box Regression

Region Proposal Network