Object Detection

- Mohit Raj Aryal
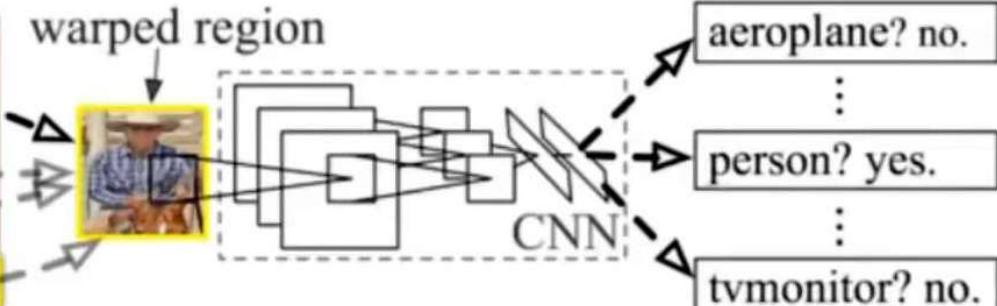
RCNN

warped region
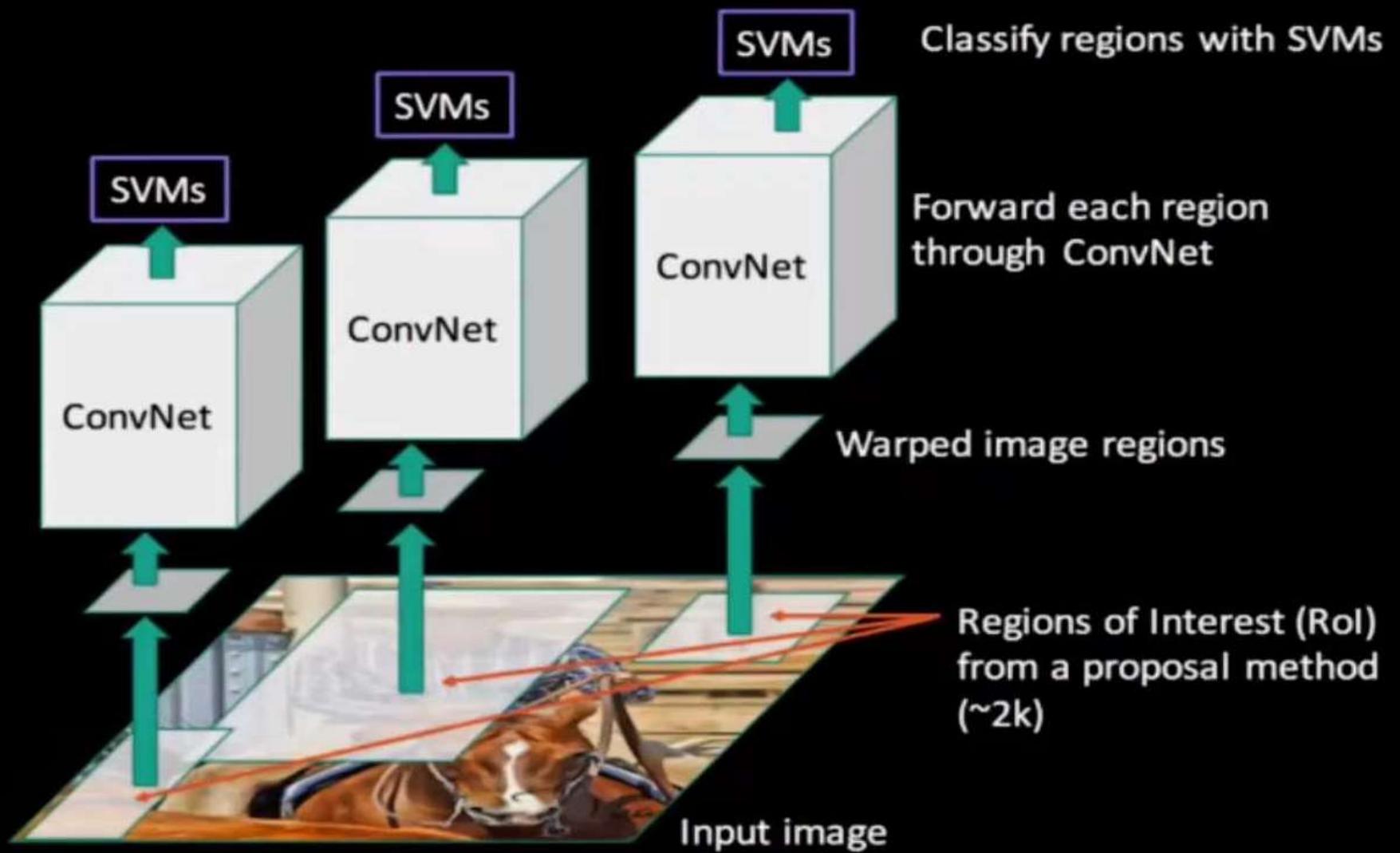
1. Input image **2.** Extract region proposals (~2k) **3.** Compute CNN features **4.** Classify regions
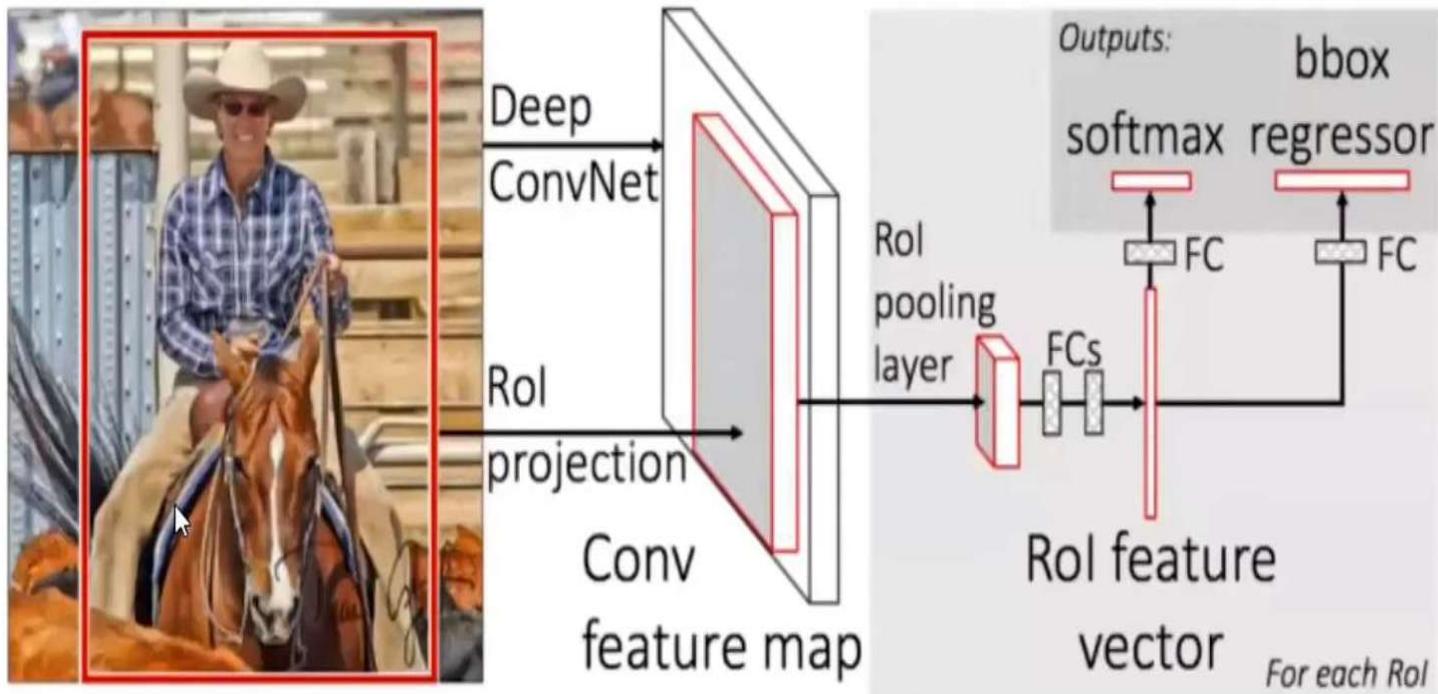
aeroplane? no.

person? yes.

tvmonitor? no.

CNN

SVMs — Classify regions with SVMs

ConvNet — Forward each region through ConvNet

Warped image regions

Regions of Interest (RoI) from a proposal method (~2k)

Input image

Girshick et al. CVPR14.
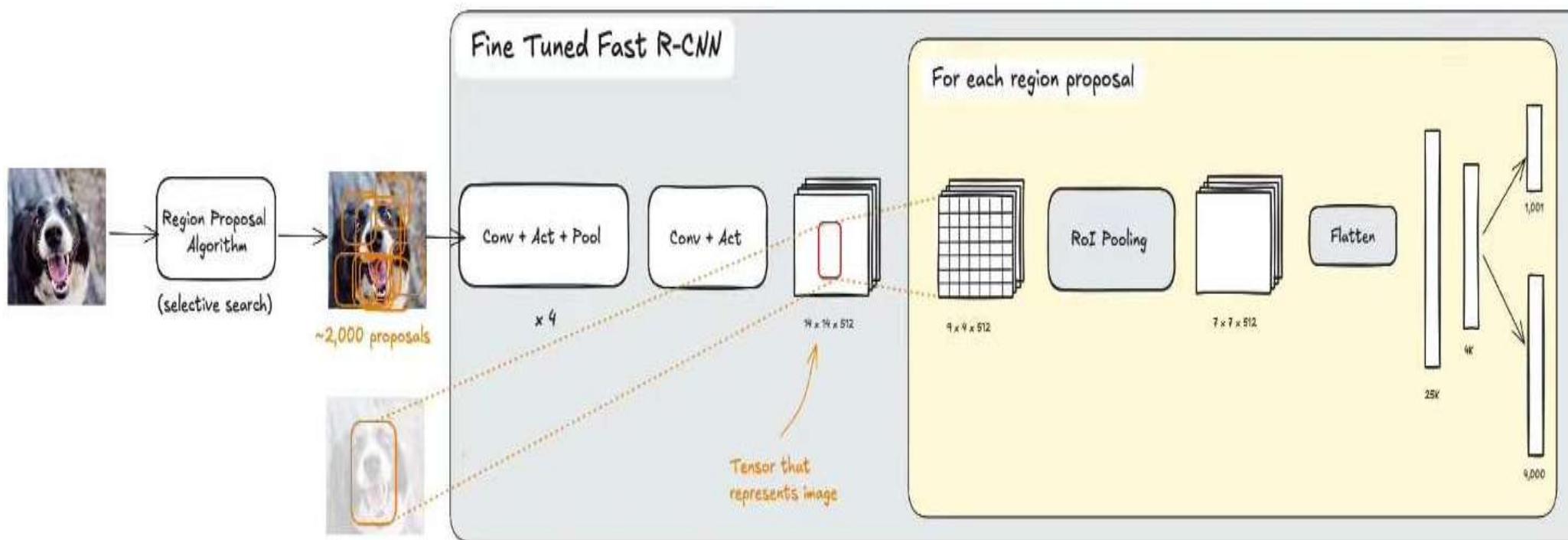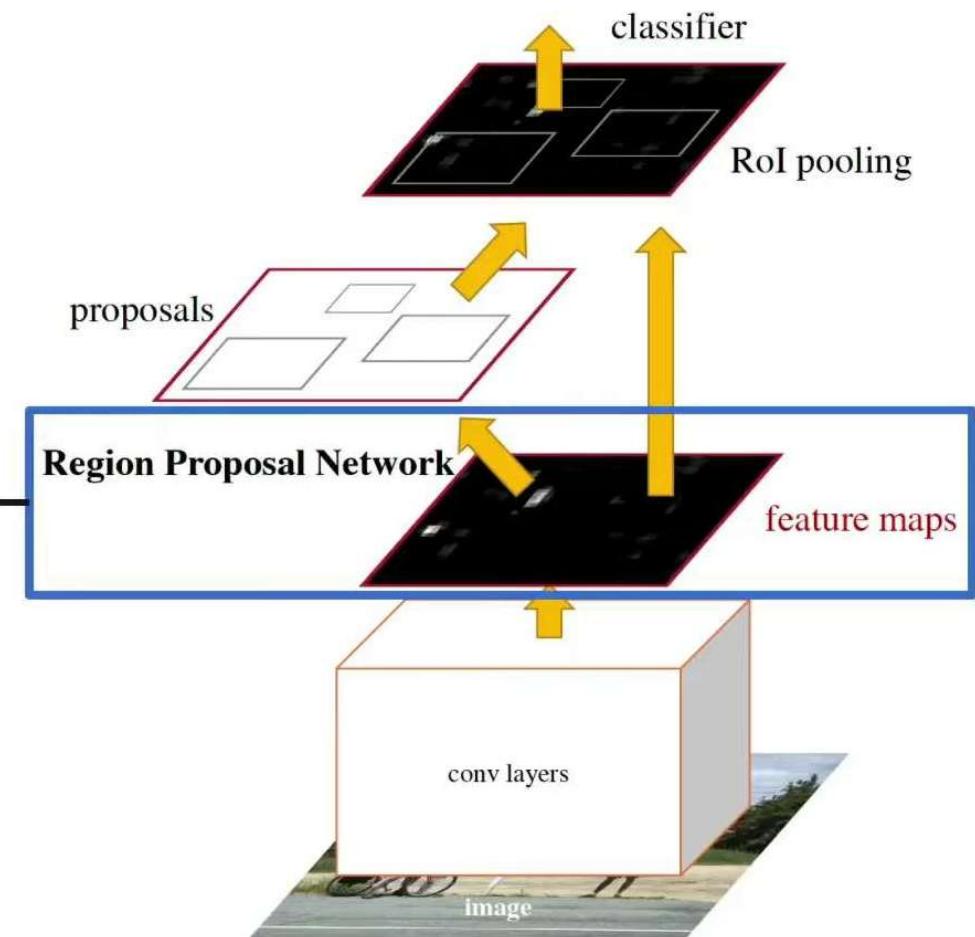
FAST RCNN

In 2015, Fast R-CNN was introduced to speed up inference

Fine Tuned Fast R-CNN

Region Proposal Algorithm (selective search) → ~2,000 proposals

Conv + Act + Pool → Conv + Act → ×4

14 x 14 x 512

Tensor that represents image

For each region proposal

9 x 9 x 512 → RoI Pooling → 7 x 7 x 512 → Flatten

25K → 4K → 1,001 / 4,000
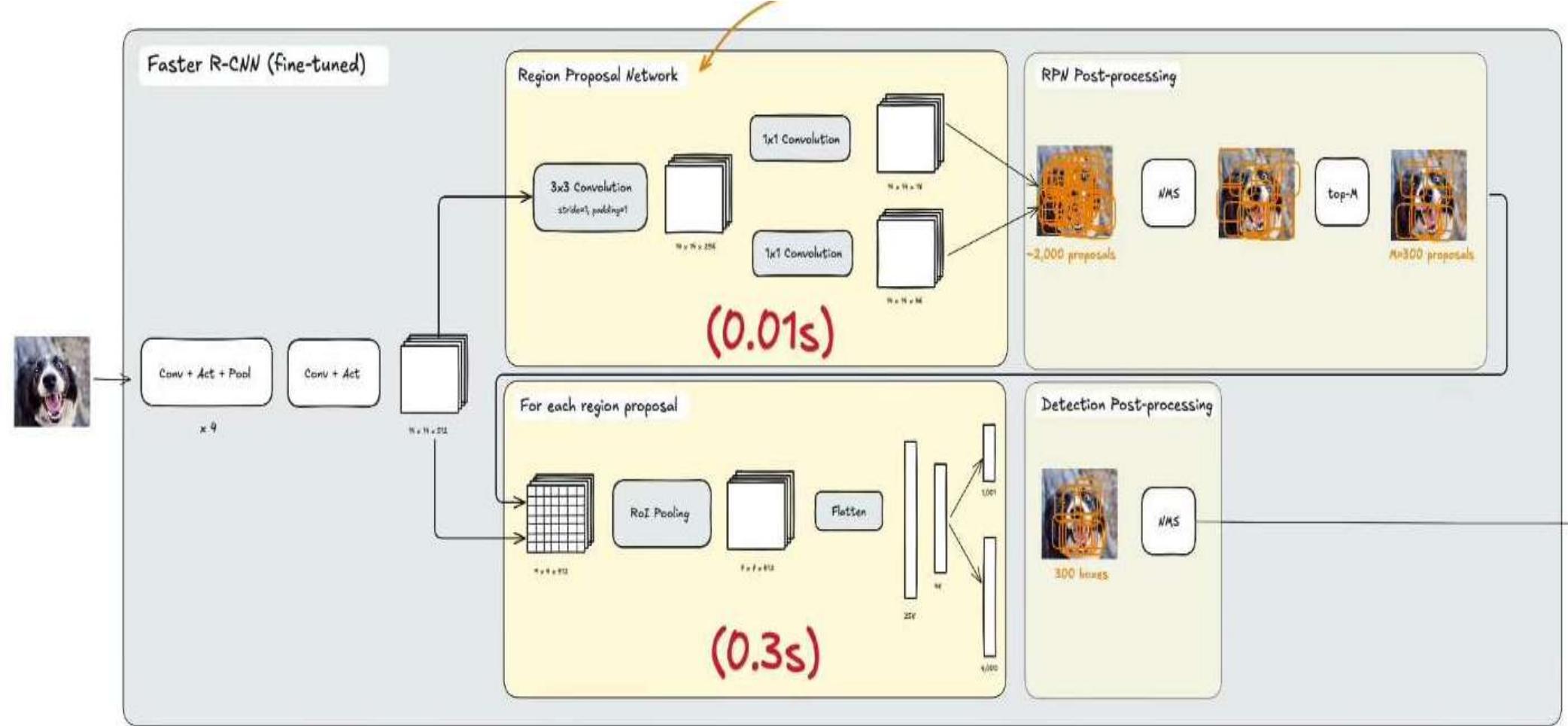
Faster R-CNN

Replaces the region proposal algorithm with a significantly faster neural network that can learn to propose better regions for the task at hand.

classifier

RoI pooling

proposals

Region Proposal Network

feature maps

conv layers

image

Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).

Faster R-CNNs with sub 1/2 second inference can be used in real time systems