



Conv
Feature
extractor
+FPN

RPN

$P_1: 256 \times 144 \times 196$

$P_2: 256 \times 72 \times 98$

$P_3: 256 \times 36 \times 49$

$256 \times 7 \times 7$

Per-
region
net

FC-81

$\hat{c}_j^{(i)}$

FC-4

$\hat{t}_j^{(i)}$

Mask
FCN

$\hat{m}_j^{(i)}$