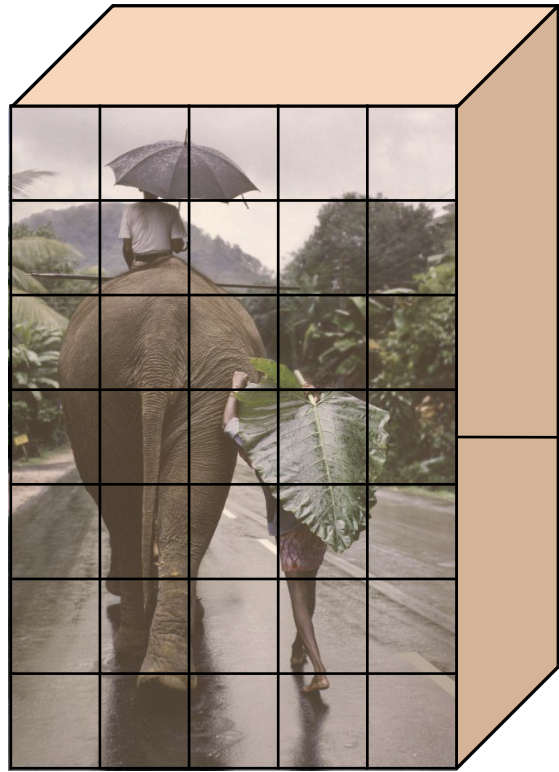$3 \times H \times W$

Convolutional feature extractor (+ FPN)

$C_L \times H_L \times W_L$

Detector head

anchor category
$\boldsymbol{C} \times \boldsymbol{k} \times H_L \times W_L$

objectness
$\boldsymbol{k} \times H_L \times W_L$

anchor correction
$4k \times H_L \times W_L$

$s_1, \ldots s_C$

o

$\Delta x, \Delta y, \Delta w, \Delta h$

k times