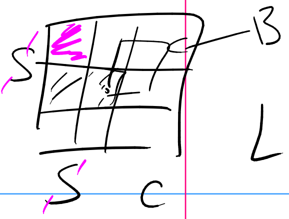


$$L = \lambda_1 \sum_{s}^{S\times S} \sum_{b}^{B} [bbox] \cdot \left((\sqrt{w}-\sqrt{\hat{w}})^2 + (\sqrt{h}-\sqrt{\hat{h}})^2 + (x-\hat{x})^2 + (y-\hat{y})^2 \right)^2$$

$S \times S \times (5B + C)$

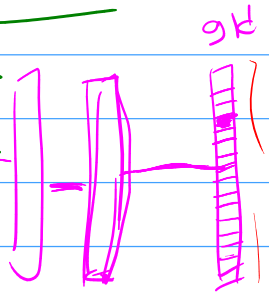
$$+ \lambda_2 \sum_{s} [bbox] \cdot L_{class}(c, \hat{c})$$

$$+ \lambda_3 \sum_{s} \sum_{b} [bbox] \cdot (1-p)^2 + \lambda_4 \sum_{s} \sum_{b} [\text{не } bbox] \cdot p^2$$

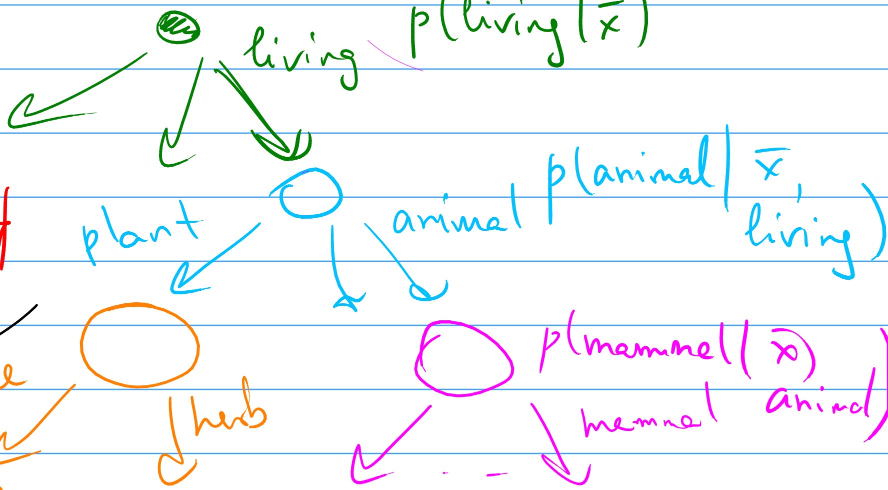
ImageNet – 120 пород собак

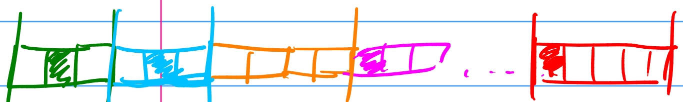
9000

WordNet

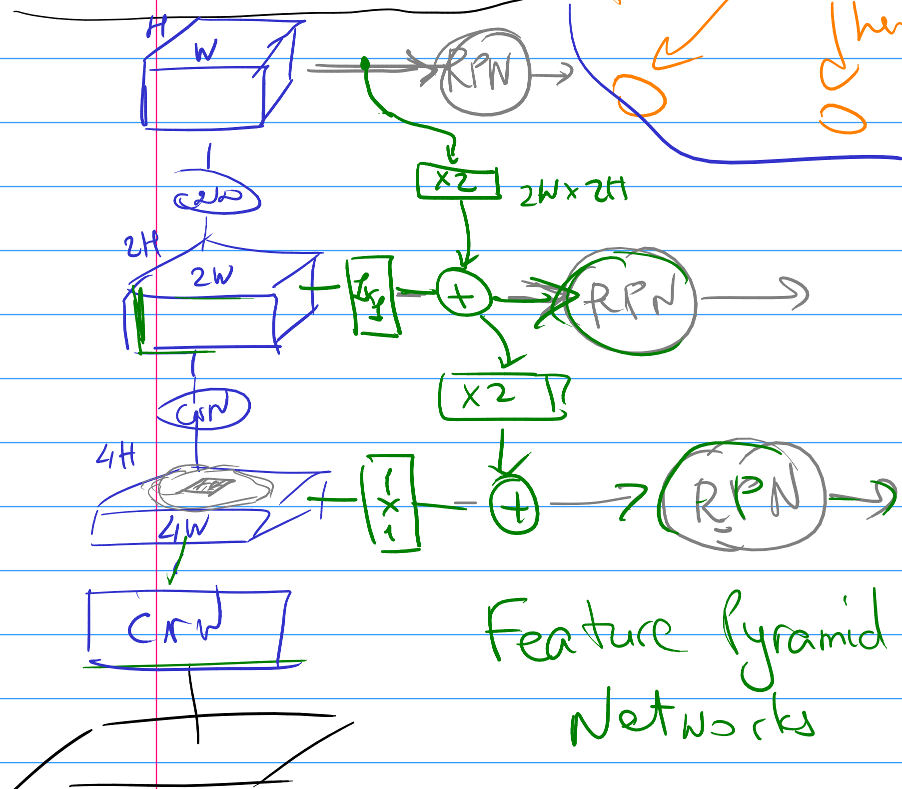


Hierarchical softmax

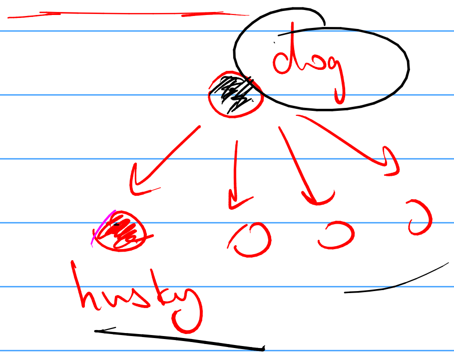


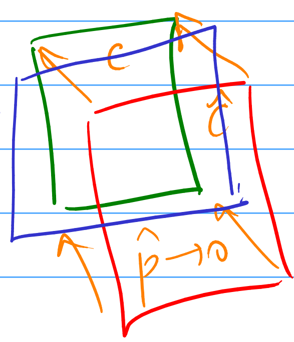


$L = l_1 + l_2 + l_4 + l_N$



Feature Pyramid Networks





$\rightarrow p \rightarrow 1$, ~~box~~ / $\rightarrow 0$ ~~box~~

$$BCE = -t \log p - (1-t)\log(1-p)$$

~~5000~~

5

$(1-p)^{\gamma}$

$p = 0.2 \quad BCE = -\log 4/5 = 0.32\ldots$

$p = 0.8 \quad BCE = -\log 1/5 = 2.32\ldots$

$$\text{Focal loss} = -t(1-p)^{\gamma}\log - (1-t)p^{\gamma}\log(1-p)$$