Bounding Box Prediction

Transformers

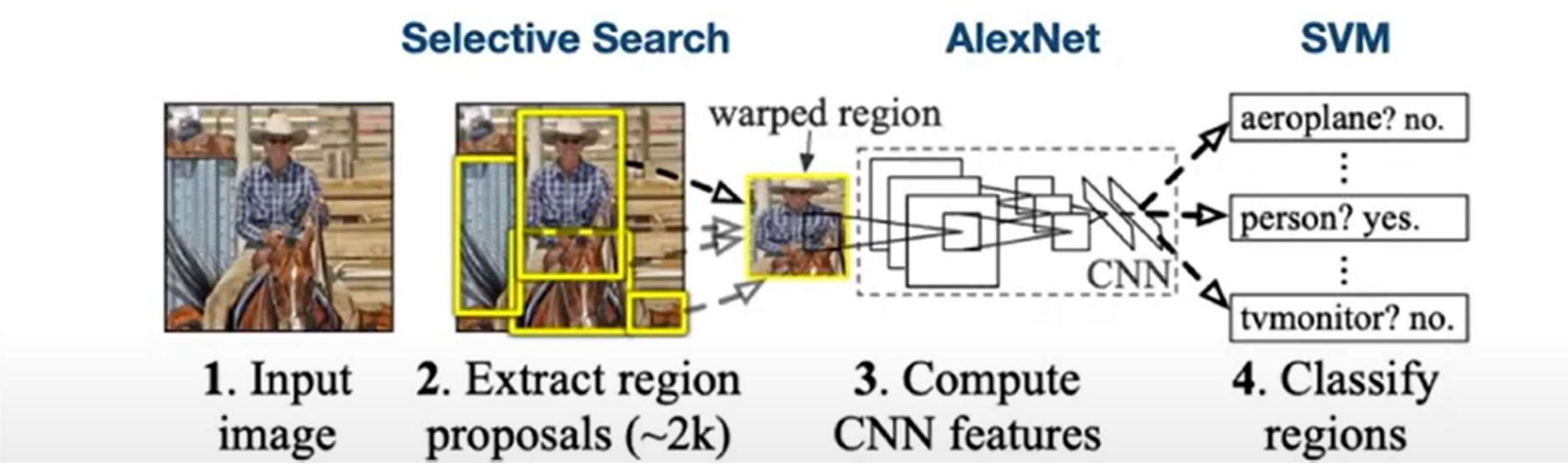
Object Detection

How to sample boxes?

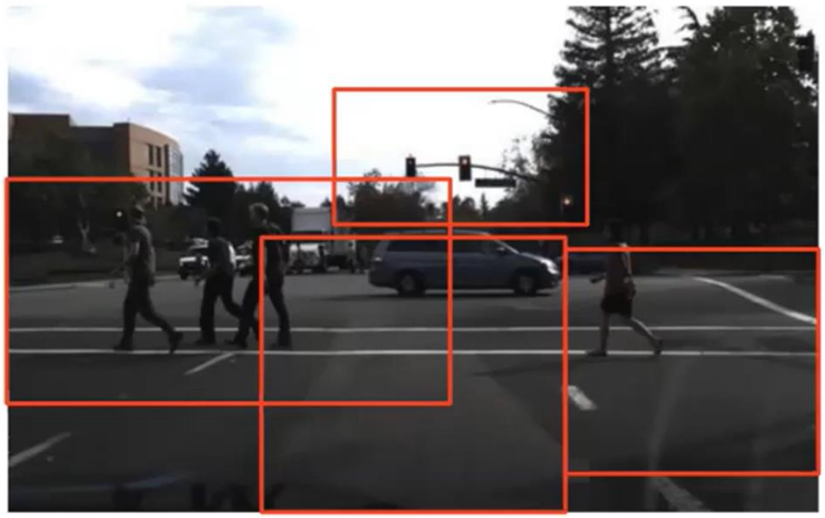
1. sliding window — expensive!
2. region proposal

Approach #1: R-CNN

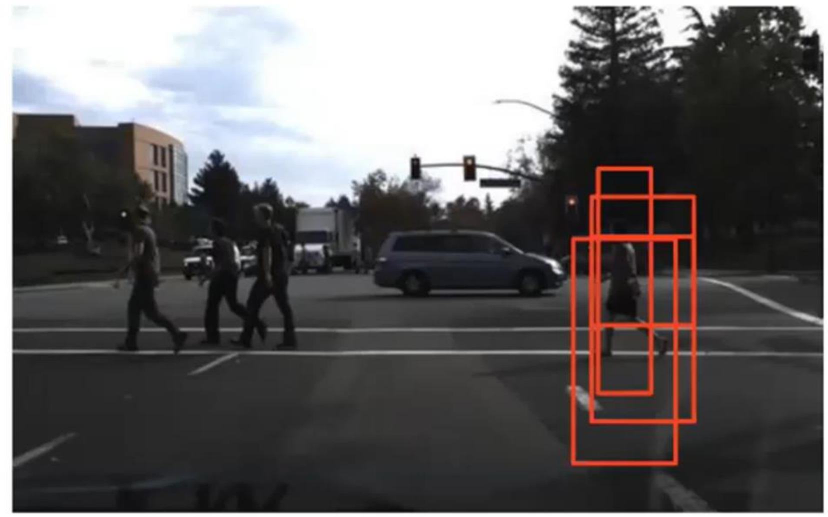
R-CNN (arXiv: 1311.2524)



Auxiliary Methods



Region Proposal



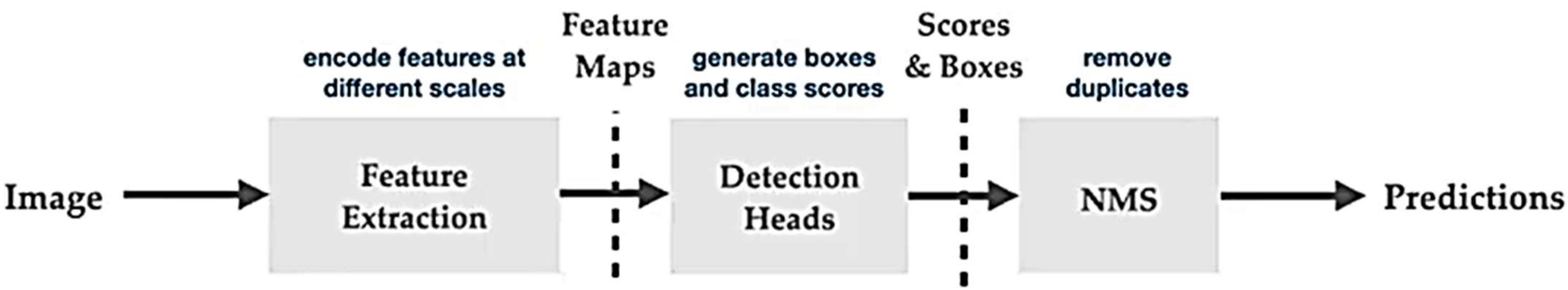
Non maximum suppression

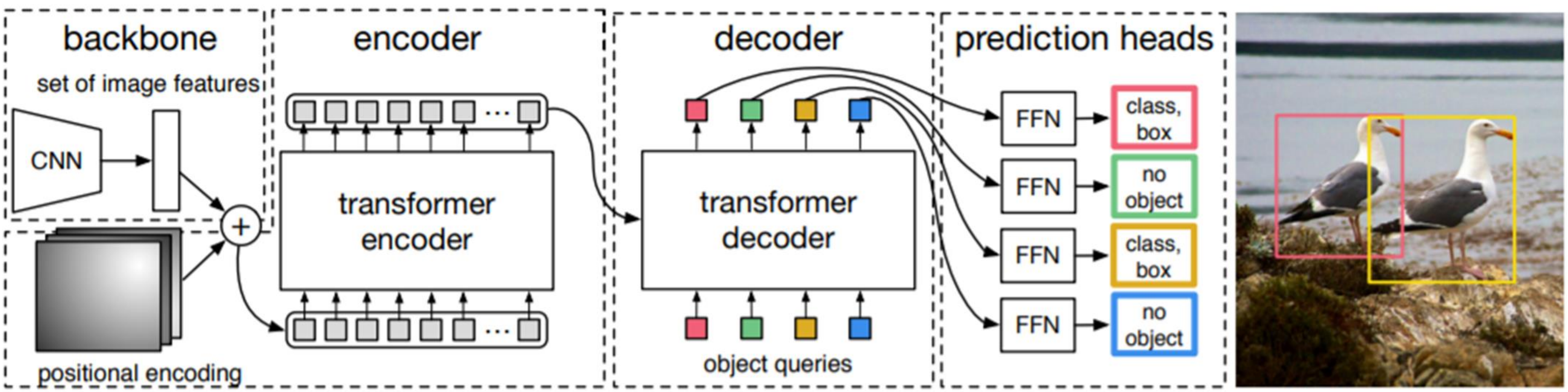
Transformers does not use these auxiliary methods for object detection like many other existing architectures.

They use Single Shot Detection (SSD) algorithm for bounding box predictions.

- Some alternatives:
  - Fast(er) R-CNN — end-to-end version of R-CNN
  - YOLO
  - Single Shot Multibox Detector

SSD Algorithm





How to match?

Prediction :
Class,bx, by, bh, bw

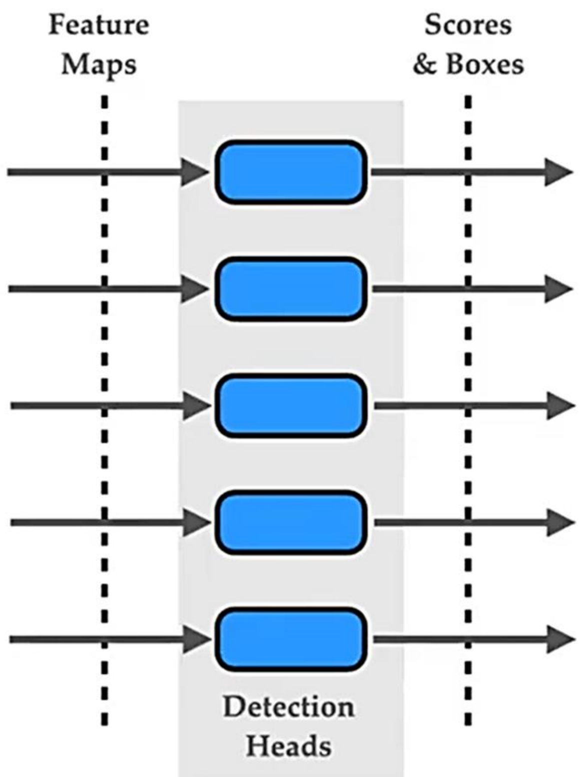
SSD matches a set of default bounding boxes. For an image size of 300x300, we have 8732 bounding boxes.

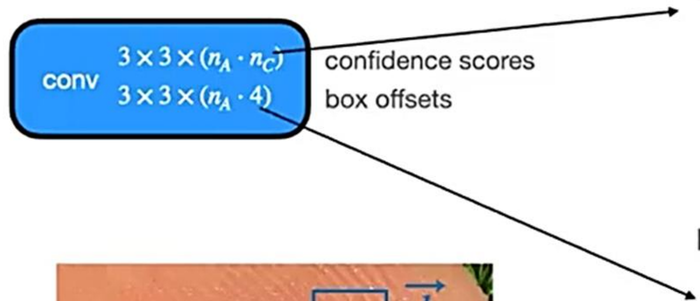
It uses IoU metric with a threshold (usually 50%) to find the accurate boxes.

Out of 8732 bounding boxes, SSD selects top 200 boxes and perform non-maximal suppression to predict the final boxes.

SSD Architecture

Multibox detector



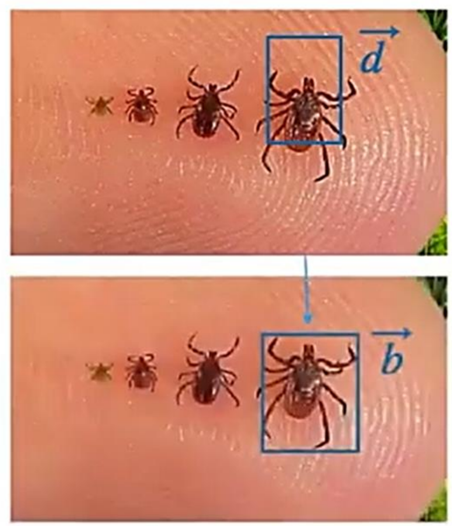


Confidence scores

background: 0.5
tick: 0.7
...

Box predictions

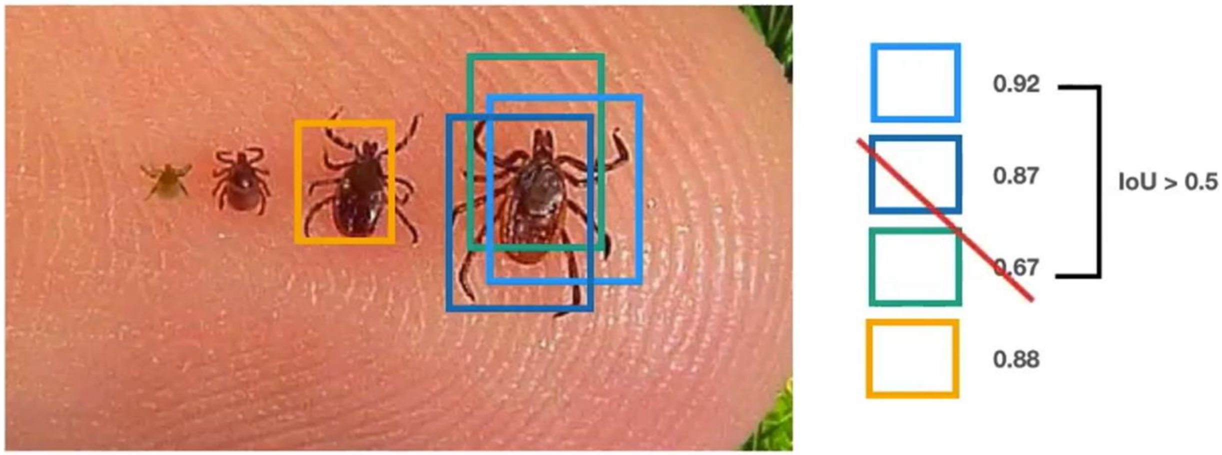
NN Output: $(\beta_x, \beta_y, \beta_w, \beta_h)$
Default box: (d_x, d_y, d_w, d_h)
Prediction: (b_x, b_y, b_w, b_h)



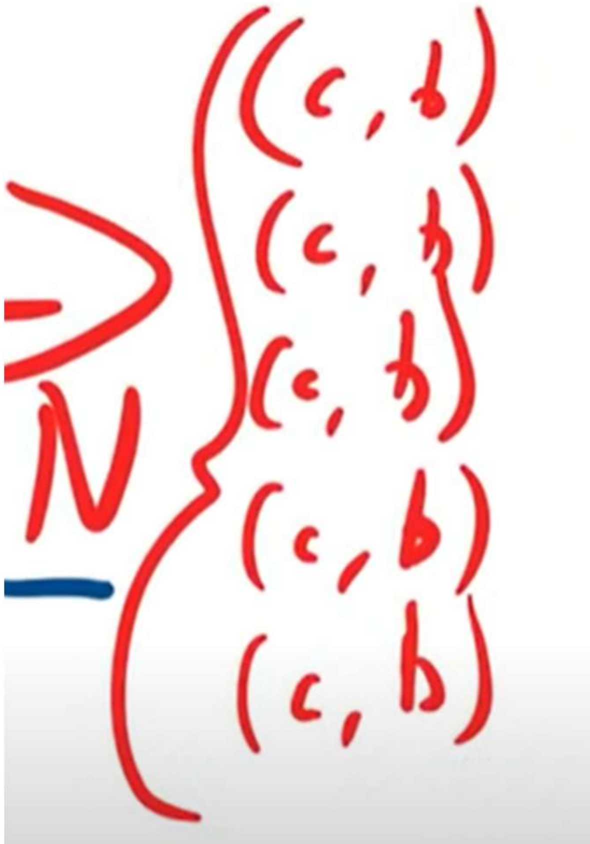
$$b_x = d_x + d_w \beta_x \qquad b_w = d_w e^{\beta_w}$$

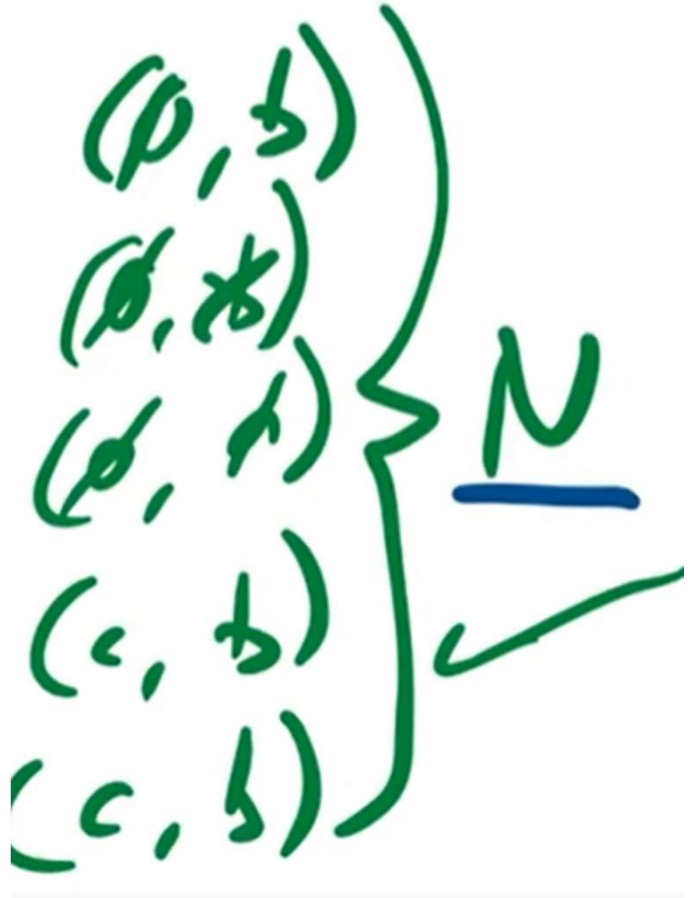
$$b_y = d_y + d_h \beta_y \qquad b_h = d_h e^{\beta_h}$$

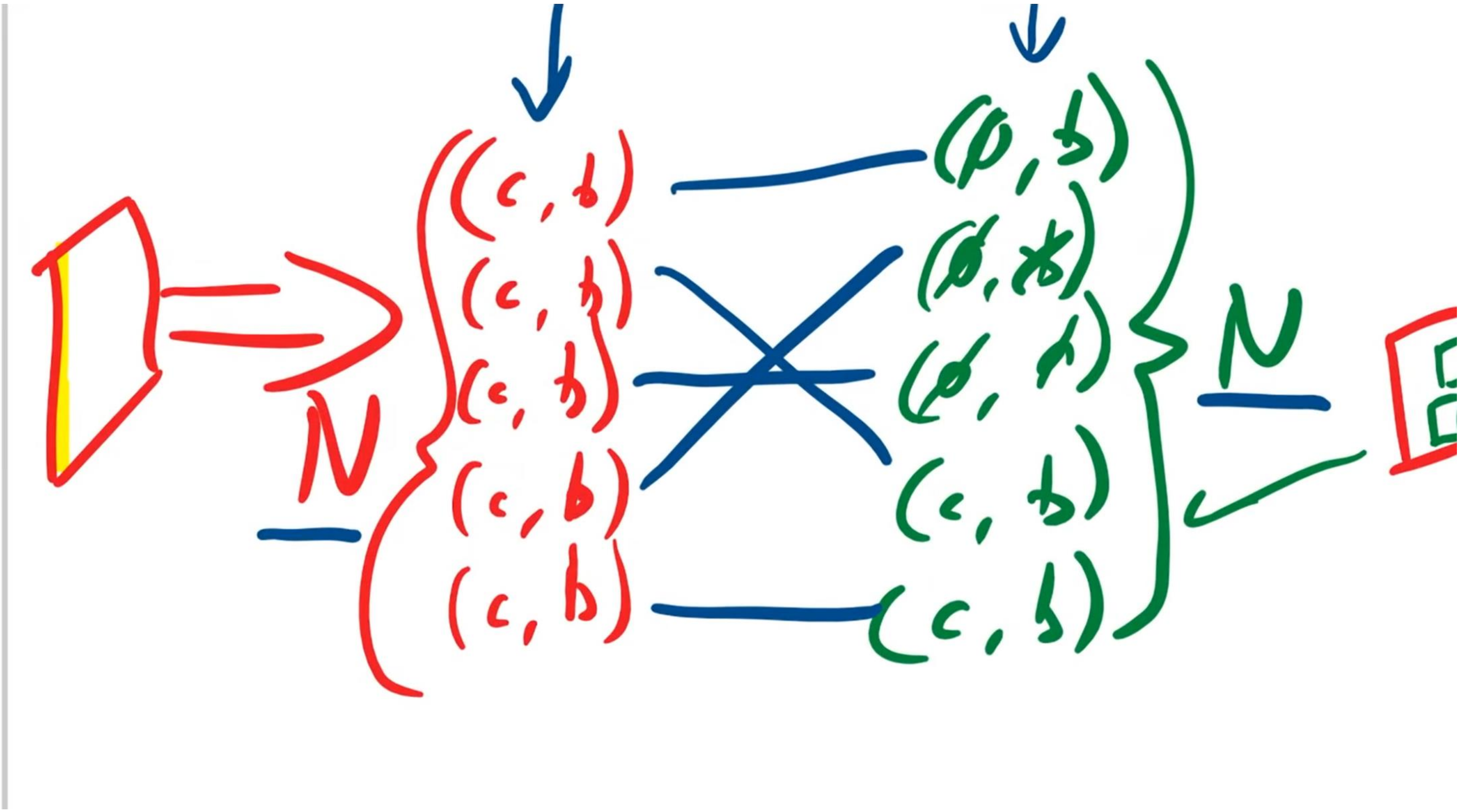
NMS



Bipartite Matching Loss







Hungarian Algorithm is responsible to get such matching.

Bipartite Loss:

$$\hat{\sigma} = \underset{\sigma \in \mathfrak{S}_N}{\arg\min} \sum_{i}^{N} \mathcal{L}_{\mathrm{match}}(y_i, \hat{y}_{\sigma(i)}),$$

$$\mathcal{L}_{\mathrm{Hungarian}}(y, \hat{y}) = \sum_{i=1}^{N} \left[-\log \hat{p}_{\hat{\sigma}(i)}(c_i) + \mathbb{1}_{\{c_i \neq \varnothing\}} \mathcal{L}_{\mathrm{box}}(b_i, \hat{b}_{\hat{\sigma}}(i)) \right]$$

Bounding box Loss:

$\lambda_{\mathrm{iou}} \mathcal{L}_{\mathrm{iou}}(b_i, \hat{b}_{\sigma(i)}) + \lambda_{\mathrm{L1}} ||b_i - \hat{b}_{\sigma(i)}||_1$ where $\lambda_{\mathrm{iou}}, \lambda_{\mathrm{L1}} \in \mathbb{R}$ are hyperparameters.