YOLOv3

→ 320x320, 22ms, 28.2 mAP
→ It is kind of a "Tech Report"

Deal

BBox Prediction

↳ dimension clusters as anchor boxes.
4 predictions for each bbox: t_x, t_y, t_w, t_h

$b_x = \sigma(t_x) + c_x$
$b_y = \sigma(t_y) + c_y$ > cell offset from top of the img
$b_w = e^{t_w} p_w$
$b_h = e^{t_h} p_h$ > bbox prior sizes w, h respectively

→ sum of squared errors loss.
↳ $\hat{t}_* - t_*$ → t_* calculated by inverting equations above

→ Calculate objectness score for each bbox using logistic regression. It is 1 if bbox prior overlaps a ground truth obj by more than any other bbox prior. Even if it's not the best and IoU passes some threshold, it is ignored.

→ System assigns only one bbox prior for each gt obj. If a bbox is not assigned, then it incurs no loss for coordinate or class preds.

Class Prediction

↳ each bbox does multilabel classification.
NO softmax. USES indep. logistic classifiers.
Loss: binary cross-entropy

Across Scales

↳ Predicts bboxes at 3 diff. scales.
↳ extract features using smtg. like feature pyramid networks
↳ Last tensor: $N \times N \times [3 \cdot (4+1+80)]$ (scales; objectness pred; bbox offsets; class preds)
↳ then apply upsampling (p2) for two times

→ Use k-means clustering for bbox priors.

Feature Extractor

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

← Darknet-53

→ Train w/
multi-scale
batch normalization
lots of data augmentation
↳ check YOLOv2 for training details

→ COCO's 'weird' mean AP metric?

- logistic activation = sigmoid
- linear (identity) activation ⇒ $f(x) = x$