

Yolo is a new approach to object detection that frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities solved by a singular end-to-end optimized neural network that performs at 45(normal) or 155(fast) frames per second. The algorithm “learns” general representations of objects and outperforms other methods including DPM and R-CNN generalizing for natural images or art.

Yolo presents itself as very simple; a regression problem on a scaled down image where the image is only viewed once and performs real-time with < 25ms latency with < .1/2 the error of R-CNN.

The YOLO algorithm is a CNN with 24 convolutional layers which extract features while two fully connected layers predict output probability and coordinates. The network is trained in imageNet for about a week and achieves 88% accuracy. The output is optimized for SSE, and to compensate for confidence errors, loss is increased for bounding box coordinate predictions and confidence loss is decreased. The results from each bounding box are normalized so that small and large boxes have comparable confidence levels. Since the algorithm predicts multiple boxes in a grid the best are chosen. Dropout and augmentation are used to reduce overfitting.