

YOLO X feedback.

Paper title	YOLOX: Exceeding YOLO Series in 2021
Authors	Zheng Ge*, Songtao Liu*, Feng Wang, Zeming Li, Jian Sun, Megvii Technology

Contents

1- What do you like about it?	2
2- What do you dislike?	2
3- What do you think are potential areas of improvement and what are the main contributions?	2
4- How do they differ from previous YOLO models?	3

1- What do you like about it?

This paper discussed many parts of object detection and tried various approaches from different papers and collected the best things together to achieve a new state-of-the-art model for object detection. From this, we can see the power of the literature review they have done to exploit many good ideas together.

They added the anchor-free manner and added an advanced label assignment algorithm and found that using a decoupled head idea improves the AP.

They used strong data augmentation: Mosaic, MixUp and used it carefully because it achieved good performance only in the large models but decreased the AP in the small versions like the Nano version (Yolox-Nano).

The authors have a good storytelling technique that makes me always pay attention while reading and they reference everything they use or talk about.

After testing their results with myself I can say that they don't claim any results.

2- What do you dislike?

The only thing I have noticed is that the authors used some pronouns such as we achieved ..., our model is ..., and so on and this is not good while writing a scientific paper.

3- What do you think are potential areas of improvement and what are the main contributions?

1- Anchor free

The anchor-based method has some disadvantages such as how to determine the anchor shapes. we need to set this configuration manually or by clustering algorithm and this is not the optimal solution because it depends on the dataset and doesn't generalize to other datasets also this increases the complexity of heads and the number of predictions, so this makes it difficult to run on the edge devices.

The Anchor free manner overcomes these problems, now the vector result of each cell will be $(5+n_classes)$ instead of $(5*n_anchors+n_classes)$

They used the object center points as reference points for predicting bounding box coordinates and object classes. This method can be more effective than anchor-based methods when detecting objects of varying sizes and shapes.

2- Decoupled head

The authors proved by experiments that dealing with the object detection problem with classification and regression in the same vector causes conflict between the tasks and does not achieve good performance, they tried to make the model deal with this as two problems by adding 3 head different heads (classification, regression, objectness score)

3-Strong data augmentation

The authors exploited the power of the new data augmentation ideas such as Mosaic and MixUp which gives a good improvement in similar tasks such as YOLOv3, YOLOv4, and YOLOv5 training.

But they found that this technique is good for large models only and this decreases the AP in the Nano version, so in the Nano version they remove the mix-up augmentation and weaken the mosaic (reduce the scale range from [0.1, 2.0] to [0.5, 1.5])

The authors say that they don't need to use ImageNet pre-training anymore after applying these augmentation methods.

They edited the implementation of the original MixUp by jittering both images by a randomly sampled scale factor before mixing up them, and they found that the scale jittering is a qualified replacement for Copypaste when no instance mask annotation is available.

4-Multiple positive

According to anchor free manner, there is only one cell that contains the center object differs from the anchor-based method which assumes that all the anchor's cell which have $\text{IOU} > 0.5$ with the ground truth is a positive cell, so they applied the multiple positive manners which assumes that positive cell is the 3×3 .

5-SimOTA

How to determine these 3×3 positive cells? They used the OTA algorithm for the assignment of the label which treats the problem as an optimal transport assignment which uses the Sinkhorn-Knopp Iteration algorithm to solve and find the best assignment.

The original OTA increased the time of the training by 25%, so they edited it with only dynamic top-k.

4- How do they differ from previous YOLO models?

The previous YOLO versions use the anchor-based manner and the used anchor free.

They used decoupled heads (classification, regression, abjectness) instead of one head.