



Object Detection and Tracking: Camera-based 2D Object Detection (YOLOX)

Lee Xuan Le, Chloe 22410022

Valen Tang Wenwen 22410023

黃翔 12110426

钟靖宸 12110320

CONTENT

01

Background & Significance

02

Current Research Status

03

Completed Work

04

Research Plan & Expected Results

05

Potential Challenges & Solutions



The background of the slide features a complex, abstract digital network. It consists of numerous glowing green and blue nodes, which are interconnected by a web of thin, white lines. This network is set against a dark, almost black, background, creating a strong contrast. The nodes vary in size and intensity of their glow, suggesting a hierarchy or active status within the network. Some nodes are larger and more prominent, while others are smaller and less intense. The overall effect is one of a dynamic, interconnected system, possibly representing a global network or a complex data structure.

[11]

Background & Significance

What is Object Detection?

- **Purpose:** Identifies and classifies object positions in images
- **Driving force:** Powered by deep convolutional neural networks (CNNs) for real-time feature extraction

Popular Models

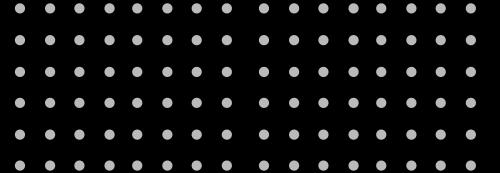
- **YOLO:** You Only Look Once
- **R-CNN:** Region-based Convolutional Neural Networks

Applications of Object Detection

- Robotics and automation
- Self-driving cars
- Facial recognition systems

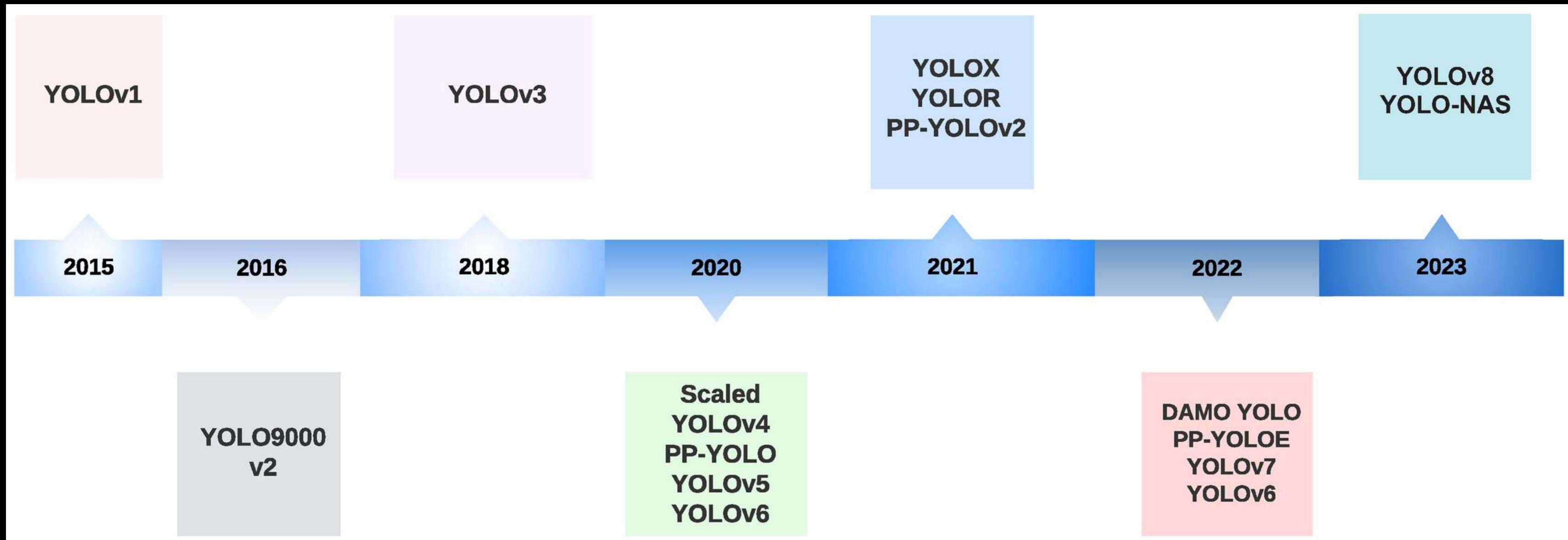
Challenges & Ethical Concerns

- **Potential misuse:** Deepfake technology
 - Misleading content
 - Non-consensual videos
- **Threats:** Privacy and security issue

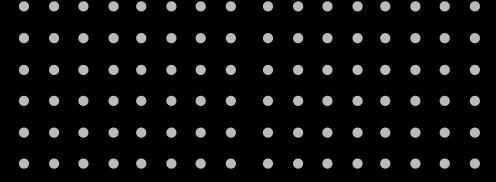


YOLUX

Origins of YOLOX

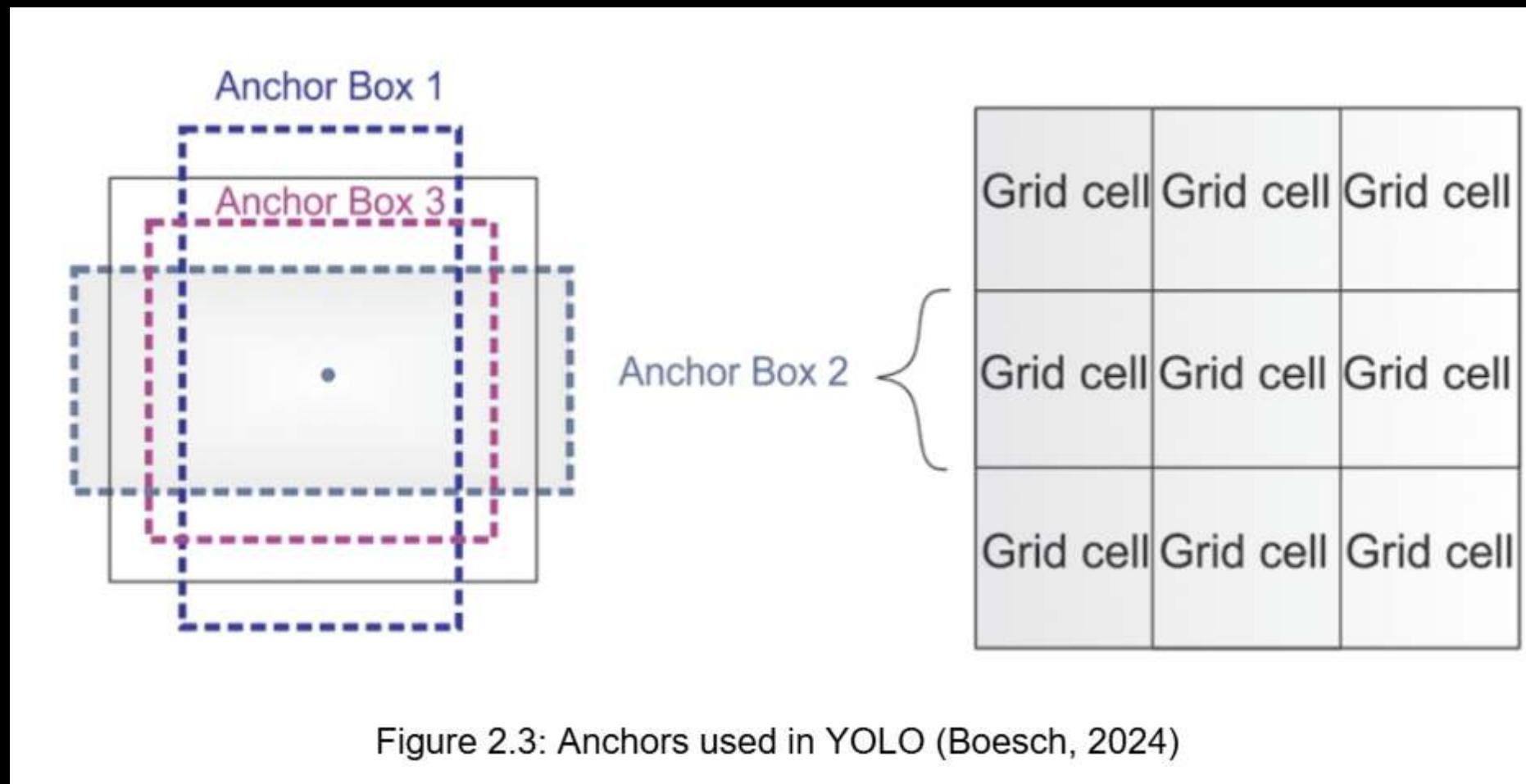
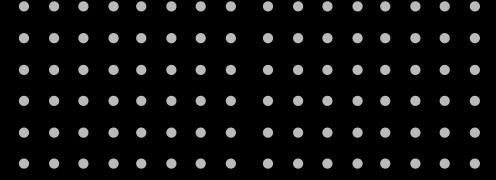


YOLOX Features



- **Architecture**
 - **Backbone:** Extract features from the input image
 - **Neck:** Combines features from multiple scales in the backbone
 - **Head:** Utilises the extracted features to conduct predictions and classification
- **Anchor free-design**
- **SimOTA label assignment**
- **Decoupled head**

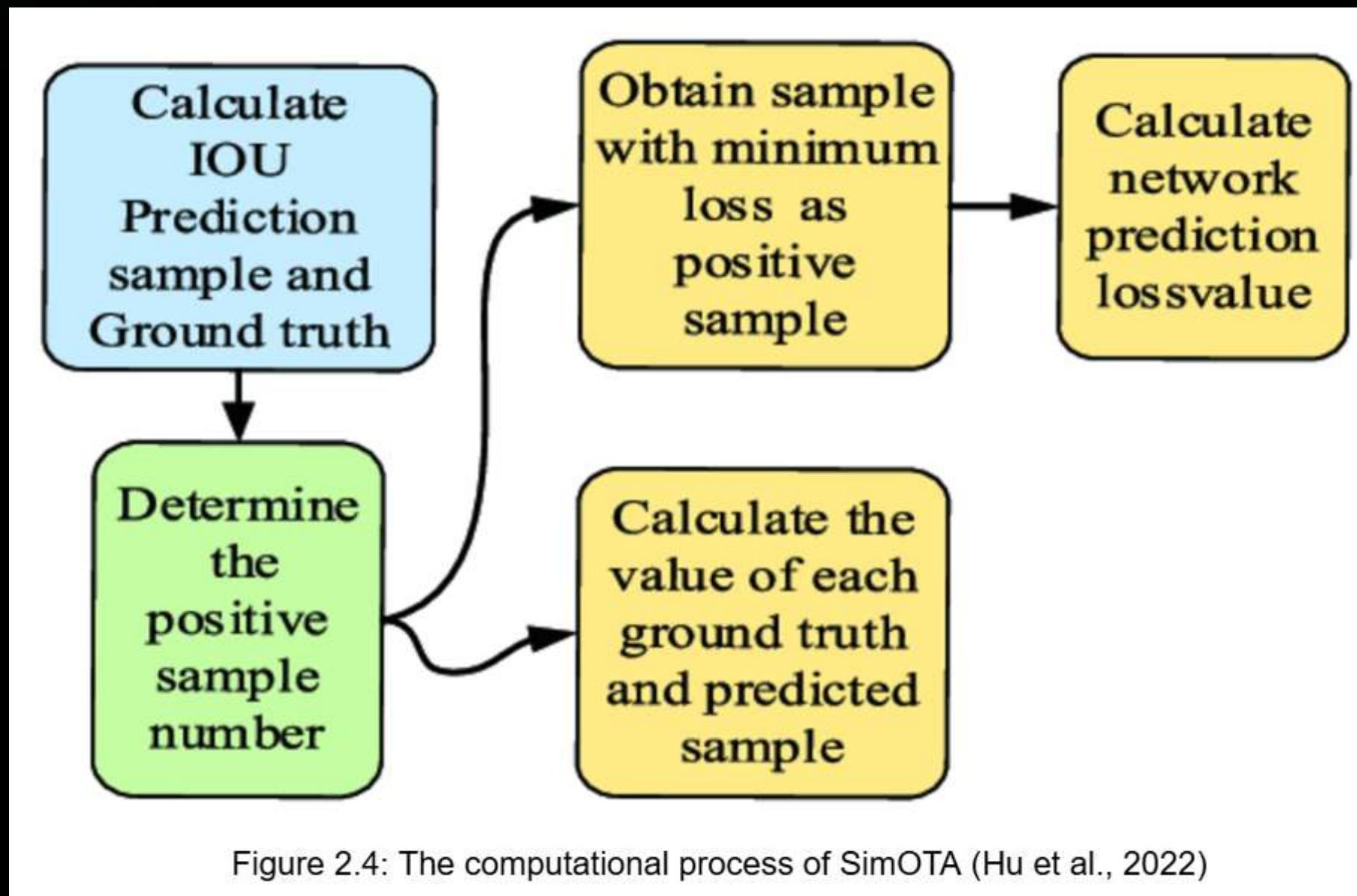
YOLOX Feature: Anchor free-design



Benefits

- **Greater flexibility:** Better adapt to objects of different shapes and sizes
- **Greater efficiency:** Reduced required predictions
- **Enhanced overall processing speed**

YOLOX Feature: SimOTA Label Assignment

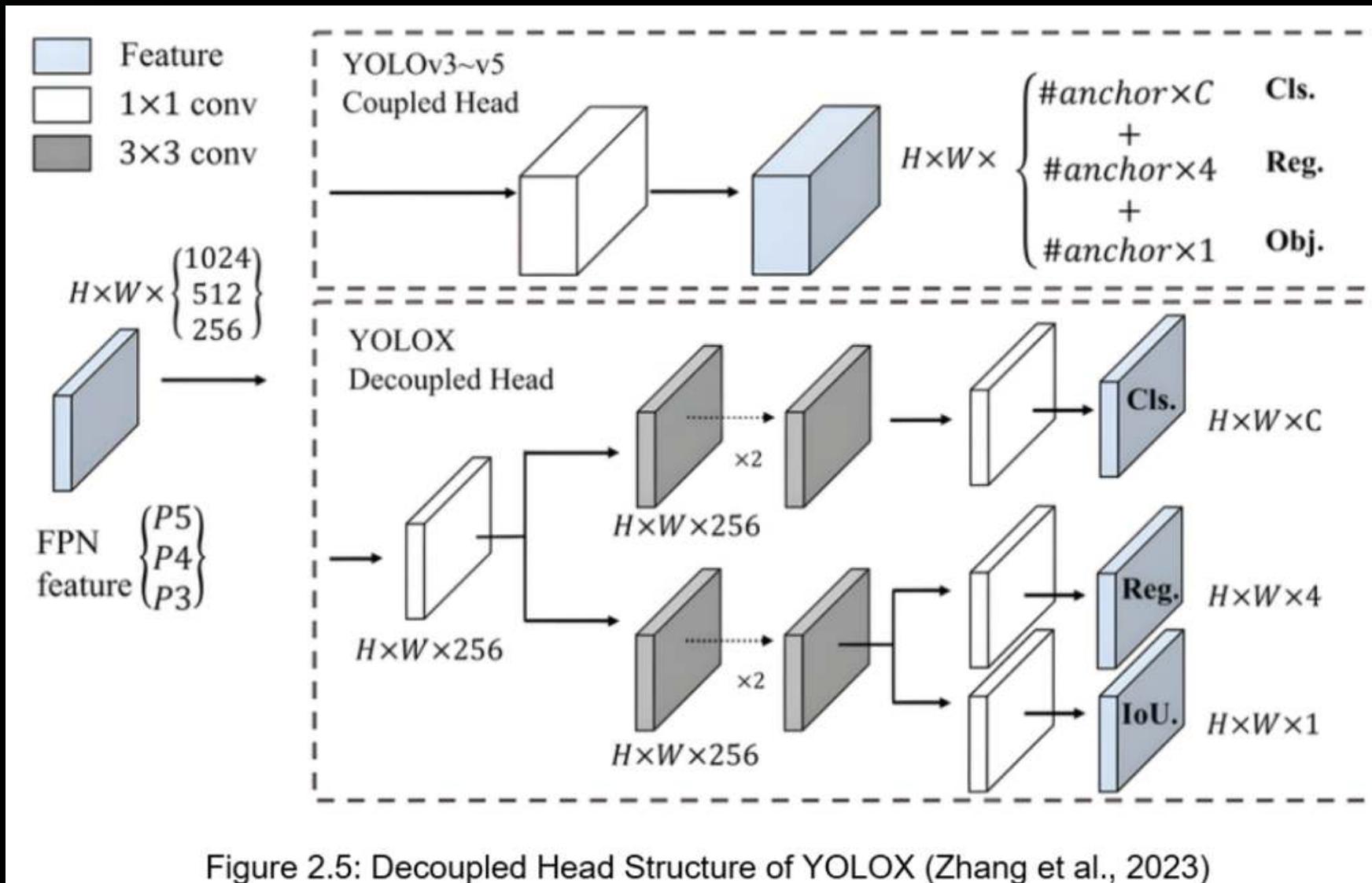
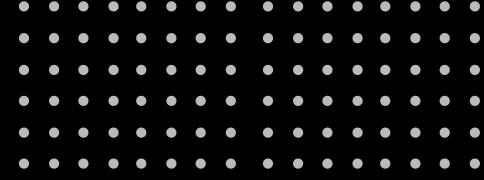


Benefits

- Simplified version as compared to OTA
- Requires only a single iteration to approximate the assignments
- Reduced training time

Figure 2.4: The computational process of SimOTA (Hu et al., 2022)

YOLOX Feature: Decoupled Head



Benefits

- Effectively target classification and regression tasks
- Separates the tasks for optimality
- Utilise fewer parameters
- Decreasing its computational expenses and overhead



Current Research Status

Case Study: Agricultural Industry

Agricultural Use: Weed Detection Model

- **Purpose:** Identify and classify crops, pests and diseases
- **Key features:**
 - **Deep network connection lightweight attention mechanism:** Recognise the different types of weeds
 - **Deconvolution layer:** Within the residual module
 - **Generalised Intersection over Union (GIoU):** Boost learning capability
- **Results**
 - **Average detection accuracy:** Almost 95% in the performance assessment,
 - **F1 value:** 0.07% improvement
 - **AP value:** 1.16% improvement
 - **Detection rate:** >88% for the detection rate of maize seedlings and weed recognition

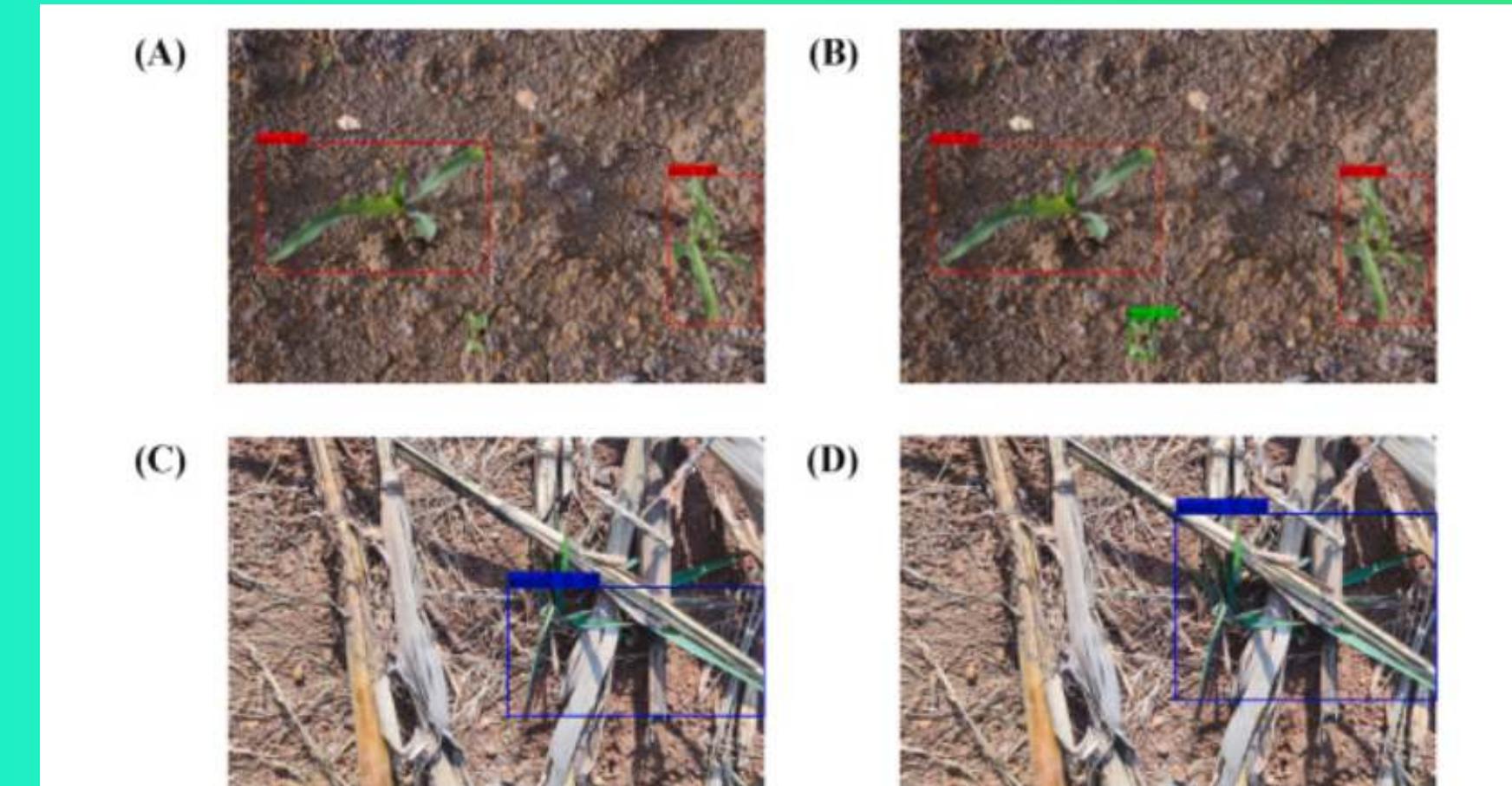


Figure 2.7: Side-by-side Comparison of Original YOLOX versus Improved YOLOX Detection Results (Zhu et al., 2024)

Case Study: Algal Bioprocessing

Chemical Use: Algae Detection Model

- **Purpose:** Identify and classify mixed algae species
- **Key features:**
 - **Starting model input:** Multi-level and morphology microalgae image dataset
 - **Focal loss:** Solve the quantity disparity challenge
 - **DIoU loss:** For regression loss, shortened processing time
- **Results**
 - **Precision:** >95%
 - **Recall:** >93%
 - **mAP value:** 3.33% improvement compared to earlier YOLOX-s
- **Learning points**
 - **YOLOX paired with other existing tools:** Could improve results
 - **Research opportunities:** Many combinations of tools

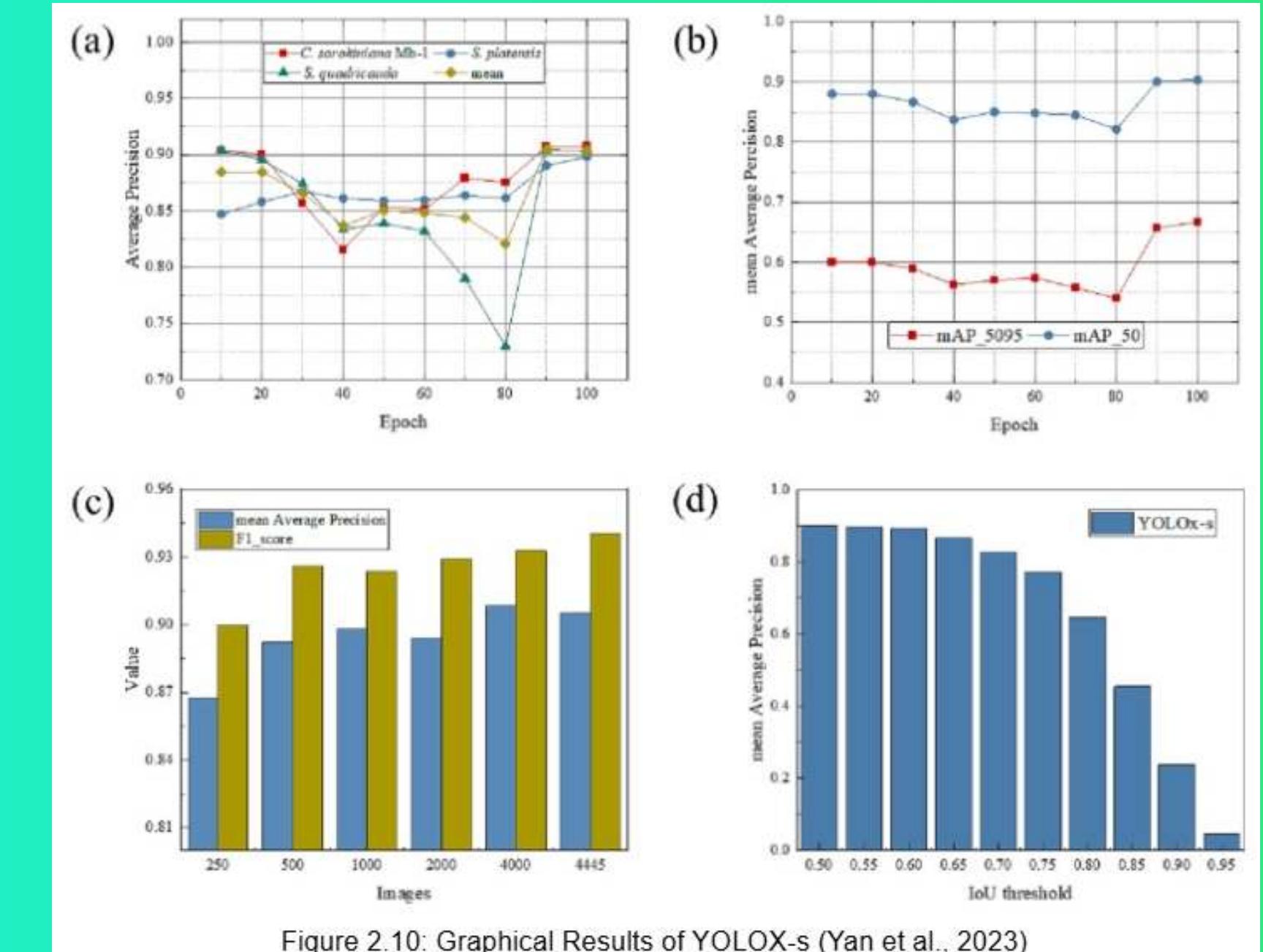
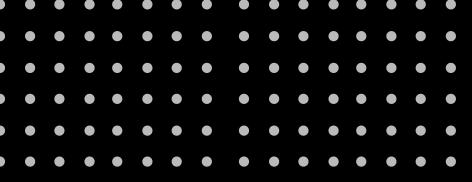


Figure 2.10: Graphical Results of YOLOX-s (Yan et al., 2023)

Research challenges



- **Limited effectiveness:** Dependency on dataset characteristics
- **Time-consuming:** Requires appropriate training data for new scenarios
- **Challenges with object detection:** Struggles with very small or overly large objects
- **Limited comparative analysis:** Lack of benchmarking against other models
- **Narrow focus:** Does not account for industry standards, economic dynamics and regulatory problems

A futuristic, glowing green digital interface with a hexagonal button containing the number 13. The background features a complex, glowing green and blue circuit board or data visualization.

13

Completed Work

SimOTA equation

$$c_{ij} = L_{cls}^{ij} + \lambda L_{reg}^{ij}$$

- c_{ij} : Total cost of matching a ground-truth object (g_i) with a predicted prediction (p_j)
- L_{cls}^{ij} : Measures the error in predicting the class of the ground-truth object (g_i)
- L_{reg}^{ij} : Measures the error in predicting the location and size of the ground-truth object (g_i)
- λ : Balancing coefficient, controls the relative importance of classification (L_{cls}^{ij}) and regression loss (L_{reg}^{ij})

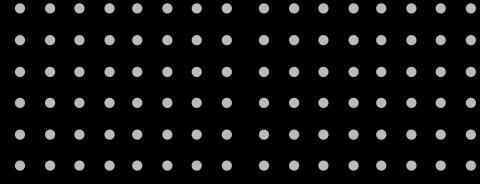
Aim:

1. Calculate the cost (c_{ij}).
2. Select the top k predictions (p_j) with the lowest costs for each ground-truth object (g_i).

Outcome:

- Find the **best match** between predictions (p_j) and ground-truth objects (g_i)
- Ensure accurate label assignment for training.

Analysis of Baseline Results

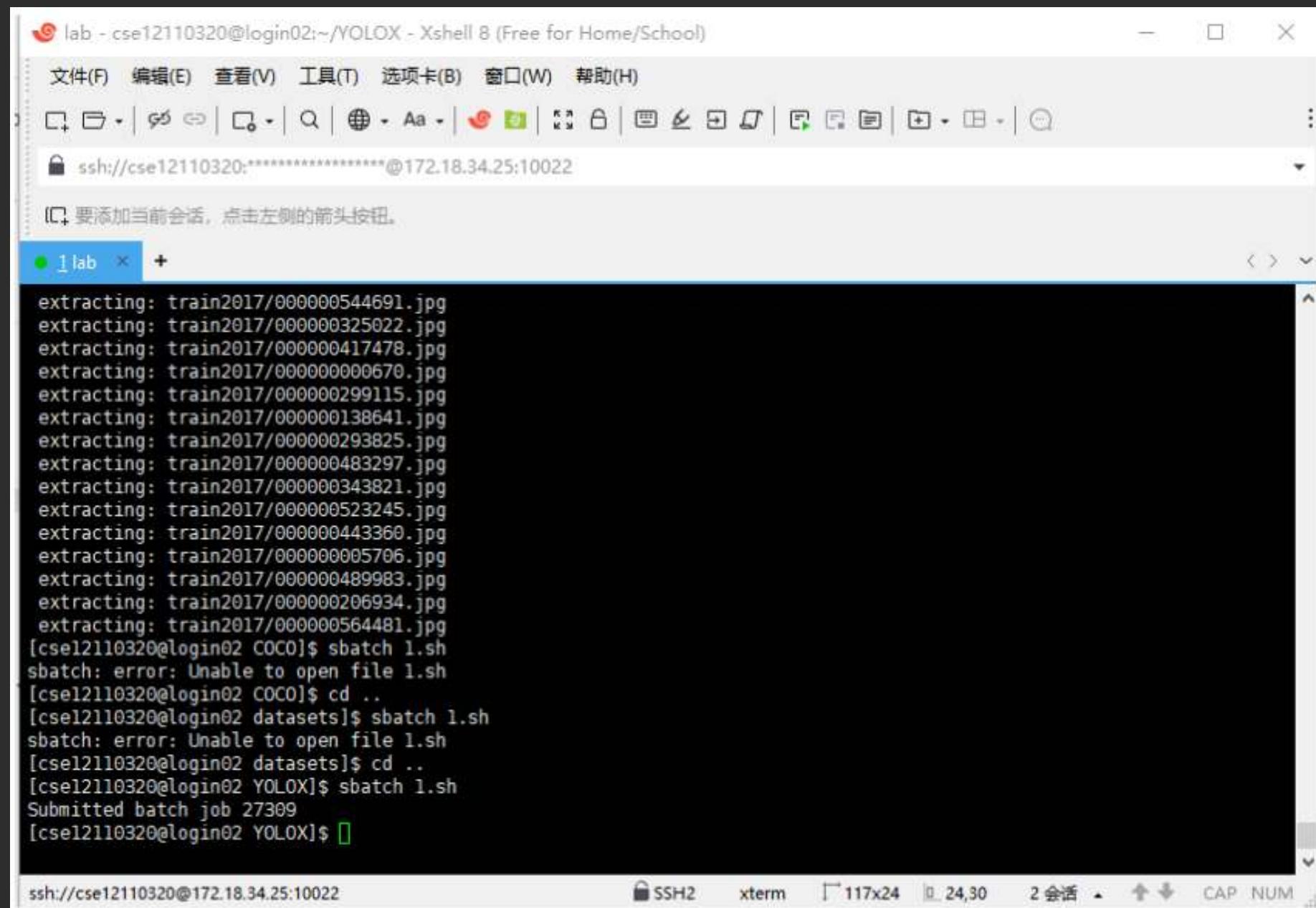


Models	AP (%)	Parameters	GFLOPs
YOLOv4-Tiny [30]	21.7	6.06 M	6.96
PPYOLO-Tiny	22.7	4.20 M	-
YOLOX-Tiny	32.8 (+10.1)	5.06 M	6.45
NanoDet ³	23.5	0.95 M	1.20
YOLOX-Nano	25.3 (+1.8)	0.91 M	1.08

- **YOLOX-Tiny has an AP of 32.8%:** Much higher than its counterparts, implies better accuracy
- **YOLOX-Nano has an AP of 25.3%:** Achieves balance between efficiency and accuracy
- YOLOX-Tiny has **moderately sized parameters (5.06M)** vs YOLOX-Nano has extremely **lightweight parameters (0.91M)**, indicating lower model complexity.
- **Computational cost:** YOLOX-Tiny has GFLOPs of **6.45**, which is relatively **efficient** for its accuracy, while YOLOX-Nano has GFLOPs of **1.08**, making it much more **efficient** but with slightly **lower accuracy**.

Deployment Results

- **Dataset: COCO 2017 dataset to run a YOLOX-s model**
- **Step 1: Connect to the server and run the model**

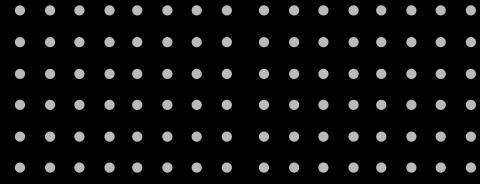


```
lab - cse12110320@login02:~/YOLOX - Xshell 8 (Free for Home/School)
文件(F) 编辑(E) 查看(V) 工具(T) 选项卡(B) 窗口(W) 帮助(H)
ssh://cse12110320:*****@172.18.34.25:10022

要添加当前会话，点击左侧的箭头按钮。
lab x +
extracting: train2017/000000544691.jpg
extracting: train2017/000000325022.jpg
extracting: train2017/000000417478.jpg
extracting: train2017/000000000670.jpg
extracting: train2017/000000299115.jpg
extracting: train2017/000000138641.jpg
extracting: train2017/000000293825.jpg
extracting: train2017/000000483297.jpg
extracting: train2017/000000343821.jpg
extracting: train2017/000000523245.jpg
extracting: train2017/000000443360.jpg
extracting: train2017/000000005706.jpg
extracting: train2017/0000000489983.jpg
extracting: train2017/000000206934.jpg
extracting: train2017/000000564481.jpg
[cse12110320@login02 COCO]$ sbatch l.sh
sbatch: error: Unable to open file l.sh
[cse12110320@login02 COCO]$ cd ..
[cse12110320@login02 datasets]$ sbatch l.sh
sbatch: error: Unable to open file l.sh
[cse12110320@login02 datasets]$ cd ..
[cse12110320@login02 YOLOX]$ sbatch l.sh
Submitted batch job 27309
[cse12110320@login02 YOLOX]$
```

ssh://cse12110320@172.18.34.25:10022 SSH2 xterm 117x24 24,30 2 会话 CAP NUM

Deployment Results



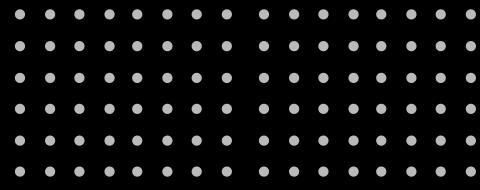
- **Step 2: Execute commands**

```
1 [6].sh - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

#!/bin/bash
#SBATCH -o job.%j.out      # 脚本执行的输出将被保存在当job.%j.out文件下, %j表示作业号;
#SBATCH --partition=titan    # 作业提交的指定分区队列为titan
#SBATCH --qos=titan          # 指定作业的QOS
#SBATCH -J myFirstGPUJob    # 作业在调度系统中的作业名为myFirstJob;
#SBATCH --nodes=1             # 申请节点数为1,如果作业不能跨节点(MPI)运行, 申请的节点数应不超过1
#SBATCH --ntasks-per-node=6   # 每个节点上运行一个任务, 默认情况下也可理解为每个节点使用一个核心;
#SBATCH --gres=gpu:1          # 指定作业的需要的GPU卡数量, 集群不一样, 注意最大限制;

python -m yolox.tools.train -n yolox-s -d 1 -b 64 --fp16 -o
```

Deployment Results



- ### • Step 3: Obtain results

```
2024-12-08 04:46:19 | INFO | yolox.core.trainer:133 - exp value:  
| keys | values |  
| seed | None |  
| output_dir | '/YOLOX_outputs' |  
| print_interval | 10 |  
| eval_interval | 10 |  
| dataset | None |  
| num_classes | 80 |  
| depth | 0.33 |  
| width | 0.5 |  
| act | 'silu' |  
| data_num_workers | 4 |  
| input_size | (640, 640) |  
| multiscale_range | 1.5 |  
| data_dir | None |  
| train_ann | 'instances_train2017.json' |  
| val_ann | 'instances_val2017.json' |  
| test_ann | 'instances_test2017.json' |  
| mosaic_prob | 1.0 |  
| mixup_prob | 1.0 |  
  
2024-12-08 04:54:37 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 120/1849, gpu_mem: 21921MB, mem: 6.6Gb, iter_time: 2.964s, data_time: 2.472s, total_loss: 14.7, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 7.9, ch_loss: 2.2, lr: 1.685e-06, size: 512, ETA: 2024-12-19 04:46:19 | INFO | yolox.core.trainer:132 - args Namespace(experiment_name='yolox_s', name='yolox-s', dist_backend='nccl', dist_url=None, batch_size=64, devices=1, exp_file=None, resume=False, ckpt=None, start_epoch=None, num_machines=1, machine_rank=0, fp16=True, cache=None, occupy=True, logger='tensorboard', optis=[])  
2024-12-08 04:46:19 | INFO | yolox.core.trainer:133 - exp value:  
2024-12-08 04:57:38 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 150/1849, gpu_mem: 21921MB, mem: 6.9Gb, iter_time: 5.058s, data_time: 4.504s, total_loss: 15.9, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 9.1, ch_loss: 2.2, lr: 2.633e-06, size: 576, ETA: 2024-12-08 04:58:29 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 160/1849, gpu_mem: 21921MB, mem: 6.7Gb, iter_time: 4.978s, data_time: 3.903s, total_loss: 19.1, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 12.3, ch_loss: 2.1, lr: 2.995e-06, size: 736, ETA: 2024-12-08 04:59:26 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 170/1849, gpu_mem: 21921MB, mem: 6.9Gb, iter_time: 5.719s, data_time: 5.314s, total_loss: 14.8, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 7.9, ch_loss: 2.3, lr: 3.301e-06, size: 512, ETA: 2024-12-08 05:00:13 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 180/1849, gpu_mem: 21921MB, mem: 6.9Gb, iter_time: 4.742s, data_time: 4.264s, total_loss: 14.6, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 7.5, ch_loss: 2.1, lr: 3.751e-06, size: 512, ETA: 2024-12-08 05:00:43 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 190/1849, gpu_mem: 21921MB, mem: 6.9Gb, iter_time: 3.038s, data_time: 2.278s, total_loss: 18.1, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 11.3, ch_loss: 2.1, lr: 4.224e-06, size: 768, ETA: 2024-12-08 05:01:26 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 200/1849, gpu_mem: 21921MB, mem: 6.8Gb, iter_time: 4.274s, data_time: 3.797s, total_loss: 15.4, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 8.6, ch_loss: 2.1, lr: 4.680e-06, size: 512, ETA: 2024-12-08 05:02:18 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 210/1849, gpu_mem: 21921MB, mem: 7.0Gb, iter_time: 5.170s, data_time: 4.686s, total_loss: 14.6, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 7.5, ch_loss: 2.5, lr: 5.160e-06, size: 512, ETA: 2024-12-08 05:02:46 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 220/1849, gpu_mem: 21921MB, mem: 6.9Gb, iter_time: 2.816s, data_time: 2.178s, total_loss: 17.4, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 10.7, ch_loss: 2.0, lr: 5.663e-06, size: 672, ETA: 2024-12-08 05:03:41 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 230/1849, gpu_mem: 21921MB, mem: 7.0Gb, iter_time: 5.462s, data_time: 4.880s, total_loss: 15.9, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 9.1, ch_loss: 2.2, lr: 6.189e-06, size: 608, ETA: 2024-12-08 05:04:29 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 240/1849, gpu_mem: 21921MB, mem: 7.0Gb, iter_time: 4.752s, data_time: 4.298s, total_loss: 14.1, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 7.3, ch_loss: 2.2, lr: 6.739e-06, size: 480, ETA: 2024-12-08 05:04:55 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 250/1849, gpu_mem: 21921MB, mem: 7.0Gb, iter_time: 2.650s, data_time: 2.133s, total_loss: 15.3, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 8.4, ch_loss: 2.3, lr: 7.313e-06, size: 544, ETA: 2024-12-08 05:05:15 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 260/1849, gpu_mem: 21921MB, mem: 7.1Gb, iter_time: 4.738s, data_time: 4.202s, total_loss: 14.8, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 8.1, ch_loss: 2.2, lr: 7.909e-06, size: 576, ETA: 2024-12-08 05:06:41 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 270/1849, gpu_mem: 21921MB, mem: 7.2Gb, iter_time: 5.856s, data_time: 4.860s, total_loss: 16.2, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 9.3, ch_loss: 2.1, lr: 8.529e-06, size: 704, ETA: 2024-12-08 05:07:45 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 280/1849, gpu_mem: 21921MB, mem: 7.3Gb, iter_time: 6.384s, data_time: 5.928s, total_loss: 14.7, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 8.0, ch_loss: 2.2, lr: 9.173e-06, size: 480, ETA: 2024-12-08 05:08:35 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 290/1849, gpu_mem: 21921MB, mem: 7.3Gb, iter_time: 5.046s, data_time: 4.461s, total_loss: 16.0, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 9.3, ch_loss: 2.1, lr: 9.840e-06, size: 640, ETA: 2024-12-08 05:09:03 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 300/1849, gpu_mem: 21921MB, mem: 7.4Gb, iter_time: 2.801s, data_time: 2.212s, total_loss: 15.2, iou_loss: 4.5, n_lloss: 0.0, conf_loss: 8.5, ch_loss: 2.1, lr: 1.053e-05, size: 608, ETA: 2024-12-08 05:09:45 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 310/1849, gpu_mem: 21921MB, mem: 7.4Gb, iter_time: 4.185s, data_time: 3.627s, total_loss: 14.7, iou_loss: 4.5, n_lloss: 0.0, conf_loss: 8.0, ch_loss: 2.3, lr: 1.124e-05, size: 576, ETA: 2024-12-08 05:10:48 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 320/1849, gpu_mem: 21921MB, mem: 7.4Gb, iter_time: 5.231s, data_time: 4.652s, total_loss: 16.3, iou_loss: 4.5, n_lloss: 0.0, conf_loss: 9.6, ch_loss: 2.1, lr: 1.190e-05, size: 608, ETA: 2024-12-08 05:11:00 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 330/1849, gpu_mem: 21921MB, mem: 7.4Gb, iter_time: 2.286s, data_time: 1.525s, total_loss: 17.3, iou_loss: 4.6, n_lloss: 0.0, conf_loss: 10.6, ch_loss: 2.0, lr: 1.274e-05, size: 768, ETA: 2024-12-08 05:10:23 | INFO | yolox.core.trainer:270 - epoch: 1/300, iter: 340/1849, gpu_mem: 21921MB, mem: 7.4Gb, iter_time: 4.185s, data_time: 3.627s, total_loss: 14.7, iou_loss: 4.5, n_lloss: 0.0, conf_loss: 8.0, ch_loss: 2.3, lr: 1.124e-05, size: 576, ETA:
```



14

Research Plan & Expected Results

Week	Task	Expected outcomes
12	<ul style="list-style-type: none"> • Topic selection • Identify the purpose and direction of the project • Conduct secondary research on the current problem 	<ul style="list-style-type: none"> • Conduct a thorough literature review of YOLOX to identify existing gaps and establish a baseline understanding of the problem. • Identify the advantages and limitations of YOLOX in object detection.
13	<ul style="list-style-type: none"> • Construct dataset • Train the results • Complete and submit report and ppt 	<ul style="list-style-type: none"> • Gather a diverse dataset for testing, ensuring the dataset includes challenging scenarios like small objects. • Train the YOLOX model on the dataset using existing architecture as a baseline to establish starting performance metrics.

Week	Task	Expected outcomes
14	<ul style="list-style-type: none"> • Improve the accuracy of working code • Dataset refinement 	<ul style="list-style-type: none"> • Implement proposed enhancements, such as fine-tuning hyperparameters to improve detection accuracy. • Clean and label datasets, ensuring proper annotations for bounding boxes and object identities.
15	<ul style="list-style-type: none"> • Collect and analyse results • Visualise possible trends • Further refinement of code 	<ul style="list-style-type: none"> • Evaluate the improved model on validation datasets, analyse performance metrics, and compare with baseline results to measure progress. • Generate visualisations of key trends to identify remaining issues and areas for further refinement. • Enhance code efficiency by reducing unnecessary processing and improving how the system handles tasks in real time.

Week	Task	Expected outcomes
16	<ul style="list-style-type: none">• Summary of key results• Presentation of results	<ul style="list-style-type: none">• Summarise the research findings, highlighting the performance gains, improvements made and any unresolved challenges.• Present the final outcomes, including comparisons to the baseline and implications for future work.



LIS

Potential Challenges & Solutions

Challenge 1

Dataset lacks diversity

Solutions

- Apply developed **data augmentation techniques** (e.g. brightness adjustments) to simulate different conditions
- **Synthetic data** could be produced to generate scenarios that are less or underrepresented in the real dataset
- Additional **real-world datasets** can be collected to improve representation

Challenge 2

Training YOLOX on large training datasets
presents significant computational challenges

Solutions

- Leverage **cloud platforms** (e.g. AWS) to access high-performance computing resources
- Start with a **small dataset** initially and gradually increase the sample size
- Implementation of **model pruning** to reduce unnecessary parameters
- **Mixed-precision training**, using lower-precision data types without sacrificing model performance

Challenge 3

Model may fail to effectively generalise to fresh data

Solutions

- **Dropout:** Randomly set a proportion of the weights to zero during training
- **Weight decay:** Adds a penalty to the loss function, discouraging excessively or overly large weights
- **Label smoothing:** Softening the target labels during training
- **Class imbalance:** Mitigated through oversampling underrepresented classes or using weighted loss functions, higher weights are assigned to less frequent classes



Thank
you!