

**计算机英语**

#### 作业2： 课题调研报告

**团队成员：**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **姓名** | **专业班级** | **学号** | **贡献比** | **得分** |
| 庄佳强 | 计算2114 | 202121331104 | 100% |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**时间： 2023 年 10 月 15 日**

目录

[感兴趣的话题：计算机视觉 3](#_Toc20408)

[一、引言：背景，相关技术、概念、术语等简介 3](#_Toc5144)

[二、研究现状 3](#_Toc17368)

[三、总结与展望 4](#_Toc31913)

感兴趣的话题：计算机视觉

综述1：《DETRs Beat YOLOs on Real-time Object Detection》

摘要：

最近，基于transformer的端到端检测器(DETRs)取得了显著的性能。然而,DETRs的高计算成本限制了其实际应用并阻止他们充分利用广告优点是没有后处理，如非最大sup压力(NMS)。本文首先对nega进行了分析NMS对存在的准确性和速度的影响Ing实时目标检测器，并建立端到端的速度基准。为解决上述问题，我们提出pose一个实时检测TRansformer (RT-DETR)。据我们所知，这是第一个实时端到端目标检测器边缘。设计了一种高效的混合编码器为了有效地处理多尺度特征，通过解耦提出了尺度内交互和跨尺度融合iou感知的查询选择，以进一步提高性能通过向de提供更高质量的初始对象查询编码器。此外，我们提出的检测器支持flex通过使用不同的推理速度的Ible调整解码器层无需重新训练，其中fa简述了在各种实时sce中的实际应用narios。我们的RT-DETR-L在COCO上取得了53.0%的AP在T4 GPU上实现了114 FPS，而RT-DETR-X达到54.8%的AP和74 FPS，优于state最先进的YOLO检测器的速度相同和准确性。此外，我们的RT-DETR-R50实现了53.1% ap 和108fps outperforming DINO-DeformableDETR-R50的准确率提高2.2% AP，提高约21倍在FPS。

一、引言：背景，相关技术、概念、术语等简介

研究内容：DETRs虽然简化了目标检测 流水线，但由于模型本身的高计算成本，很难实现实时目 标检测。以上问题自然激发我们考虑是否可以将 DETR 扩展到实时场景，充分利用端到端检测器的优势，避免 NMS 对实时检测器造成的延迟。

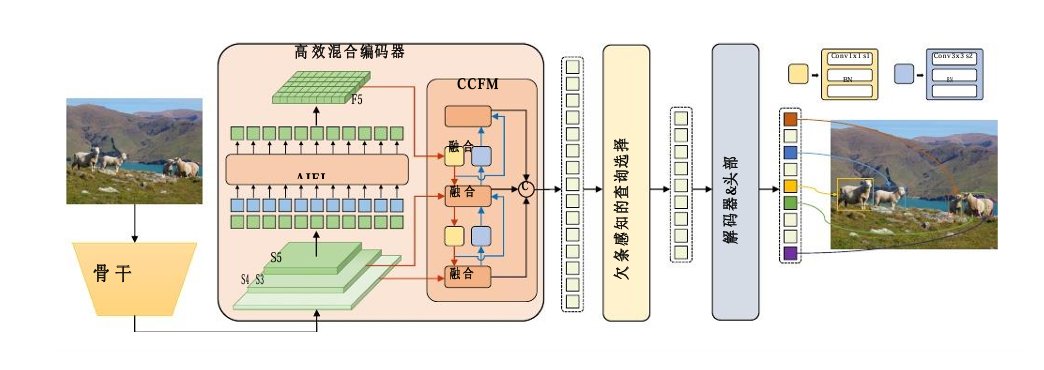
研究目的：将 DETR 扩展到实时场景，充分利用端到端检测器的优势，避免 NMS 对实时检测器造成的延迟。

研究意义：提升视觉检测的速度和检测精度。

主要挑战：transformer 编码器 由于计算成本高而成为模型的计算瓶颈。

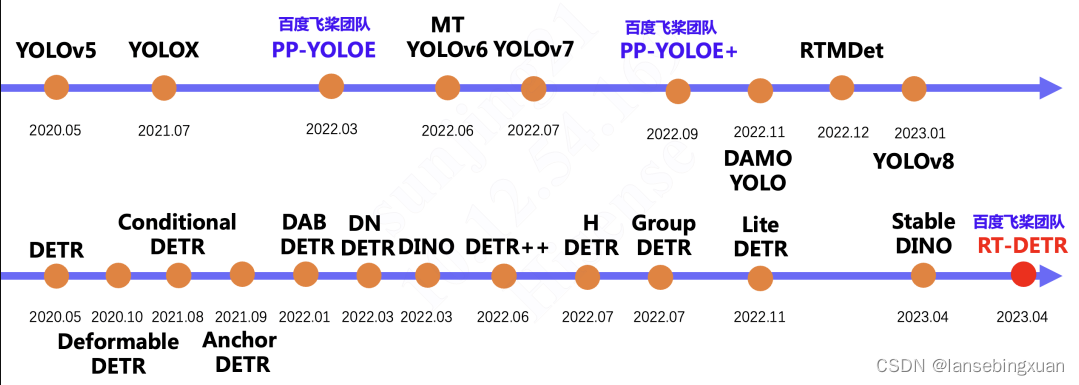
二、研究现状

总体算法步骤/流程图；



我们首先利用主干{S3, S4,S5}最后三个阶段的特征作为编码器的输入。高效混合编码器通过尺度内特 征交互(AIFI)和跨尺度特征融合模块(CCFM)将多尺度特征转化为图像特征序列。利用 IoU-aware 查询选择来选择固定数量 的图像特征作为解码器的初始对象查询。最后，带有辅助预测头的解码器迭代优化对象查询以生成框和置信度分数。

发展历程：



第一行是[YOLO系列](https://so.csdn.net/so/search?q=YOLO%E7%B3%BB%E5%88%97&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/lansebingxuan/article/details/_blank)算法的发展过程，第二行是DETR系列算法的发展过程。

输入、输出及预处理：输入为图片 输出为框和置信度分数。预处理：图像处理。

主要的方法分类（主要思想、优缺点及典型方法）：

把DETR和DINO基础上进行改进,首次实现实时端到端检测。

优点：用时更短，计成本更低，

可用的资源，如数据集与开源工具包

数据集：我们在微软 COCO 数据集上进行了实验。我们 在 COCO train2017 上进行训练，在 COCO val2017 数据 集上进行验证。我们使用标准 COCO AP 度量，并将单 个尺度图像作为输入。

三、总结与展望

未解决的问题：目前还不够成熟。

未来发展趋势：与其他实时检测器和类似大小的端到端检测器相比RT-DETR在速度和精度方面都达到了最先进的性能。此外，我们提出的检测器支持通过使用不同的解码器层灵活调整推理速度，而不需要重新训练，这促进了实时目标检测器的实际应用。未来的前景很大。

参考文献：

（采用APA 7th参考文献格式记录）

1. Alexey Bochkovskiy, Chien-Yao Wang, and Hong- Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. [1](#_bookmark0), [2](#_bookmark2)
2. Daniel Bogdoll, Maximilian Nitsche, and J Marius Zo¨llner. Anomaly detection in autonomous driving: A survey. In *Pro- ceedings of the IEEE/CVF conference on computer vision* *and pattern recognition*, pages 4488–4499, 2022. [1](#_bookmark0)
3. Yuxuan Cai, Yizhuang Zhou, Qi Han, Jianjian Sun, Xiang- wen Kong, Jun Li, and Xiangyu Zhang. Reversible column networks. *arXiv preprint arXiv:2212.11696*, 2022. [7](#_bookmark18)
4. Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delv- ing into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recogni-* *tion*, pages 6154–6162, 2018. [1](#_bookmark0)
5. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to- end object detection with transformers. In *Computer Vision– ECCV 2020: 16th European Conference, Glasgow, UK, Au- gust 23–28, 2020, Proceedings, Part I 16*, pages 213–229.

Springer, 2020. [1](#_bookmark0), [2](#_bookmark2), [7](#_bookmark18), [8](#_bookmark20)

1. *Qiang Chen, Xiaokang Chen, Gang Zeng, and Jingdong Wang. Group detr: Fast training convergence with de-coupled one-to-many label assignment. arXiv preprint* *arXiv:2207.13085, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3)*
2. Qiang Chen, Jian Wang, Chuchu Han, Shan Zhang, Zex- ian Li, Xiaokang Chen, Jiahui Chen, Xiaodi Wang, Shum- ing Han, Gang Zhang, et al. Group detr v2: Strong object detector with encoder-decoder pretraining. *arXiv preprint* *arXiv:2211.03594*, 2022. [7](#_bookmark18)
3. Cheng Cui, Ruoyu Guo, Yuning Du, Dongliang He, Fu Li, Zewu Wu, Qiwen Liu, Shilei Wen, Jizhou Huang, Xiaoguang Hu, Dianhai Yu, Errui Ding, and Yanjun Ma. Beyond self- supervision: A simple yet effective network distillation alter- native to improve backbones. *CoRR*, abs/2103.05959, 2021. [7](#_bookmark18)
4. Peng Gao, Minghang Zheng, Xiaogang Wang, Jifeng Dai, and Hongsheng Li. Fast convergence of detr with spatially modulated co-attention. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3621– 3630, 2021. [7](#_bookmark18), [8](#_bookmark20)
5. Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint* *arXiv:2107.08430*, 2021. [1](#_bookmark0), [2](#_bookmark2)
6. Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Nas-fpn: Learning scalable feature pyramid architecture for object de- tection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7036–7045, 2019. [3](#_bookmark3)
7. Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE inter- national conference on computer vision*, pages 1440–1448, 2015. [1](#_bookmark0)
8. Jocher Glenn. Yolov5 release v7.0. *[https://github.](https://github.com/ultralytics/yolov5/tree/v7.0) [com/ultralytics/yolov5/tree/v7.0](https://github.com/ultralytics/yolov5/tree/v7.0)*, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [4](#_bookmark8), [7](#_bookmark18), [8](#_bookmark20)
9. Jocher Glenn. Yolov8. *[https://github.com/](https://github.com/ultralytics/ultralytics/tree/main) [ultralytics/ultralytics/tree/main](https://github.com/ultralytics/ultralytics/tree/main)*, 2023. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [4](#_bookmark8), [7](#_bookmark18), [8](#_bookmark20)
10. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceed- ings of the IEEE conference on computer vision and pattern* *recognition*, pages 770–778, 2016. [7](#_bookmark18)
11. Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Jun- yuan Xie, and Mu Li. Bag of tricks for image classifica- tion with convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern* *recognition*, pages 558–567, 2019. [7](#_bookmark18)
12. Xin Huang, Xinxin Wang, Wenyu Lv, Xiaying Bai, Xiang Long, Kaipeng Deng, Qingqing Dang, Shumin Han, Qiwen Liu, Xiaoguang Hu, et al. Pp-yolov2: A practical object detector. *arXiv preprint arXiv:2104.10419*, 2021. [1](#_bookmark0), [2](#_bookmark2)
13. Chuyi Li, Lulu Li, Yifei Geng, Hongliang Jiang, Meng Cheng, Bo Zhang, Zaidan Ke, Xiaoming Xu, and Xiangxi- ang Chu. Yolov6 v3. 0: A full-scale reloading. *arXiv preprint* *arXiv:2301.05586*, 2023. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [4](#_bookmark8), [7](#_bookmark18), [8](#_bookmark20)
14. Feng Li, Ailing Zeng, Shilong Liu, Hao Zhang, Hongyang Li, Lei Zhang, and Lionel M Ni. Lite detr: An inter- leaved multi-scale encoder for efficient detr. *arXiv preprint arXiv:2303.07335*, 2023. [3](#_bookmark3)
15. Feng Li, Hao Zhang, Shilong Liu, Jian Guo, Lionel M Ni, and Lei Zhang. Dn-detr: Accelerate detr training by intro- ducing query denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13619–13627, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [7](#_bookmark18)
16. Junyu Lin, Xiaofeng Mao, Yuefeng Chen, Lei Xu, Yuan He, and Hui Xue. Dˆ 2etr: Decoder-only detr with com- putationally efficient cross-scale attention. *arXiv preprint* *arXiv:2203.00860*, 2022. [5](#_bookmark12)
17. Tsung-Yi Lin, Piotr Dolla´r, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyra- mid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recogni-* *tion*, pages 2117–2125, 2017. [3](#_bookmark3)
18. Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dolla´r. Focal loss for dense object detection. In *Pro- ceedings of the IEEE international conference on computer* *vision*, pages 2980–2988, 2017. [1](#_bookmark0)
19. Shilong Liu, Feng Li, Hao Zhang, Xiao Yang, Xianbiao Qi, Hang Su, Jun Zhu, and Lei Zhang. Dab-detr: Dynamic anchor boxes are better queries for detr. *arXiv preprint* *arXiv:2201.12329*, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [6](#_bookmark15), [7](#_bookmark18)
20. Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Pro- ceedings of the IEEE conference on computer vision and pat-* *tern recognition*, pages 8759–8768, 2018. [3](#_bookmark3), [5](#_bookmark12)
21. Xiang Long, Kaipeng Deng, Guanzhong Wang, Yang Zhang, Qingqing Dang, Yuan Gao, Hui Shen, Jianguo Ren, Shumin Han, Errui Ding, et al. Pp-yolo: An effective and ef- ficient implementation of object detector. *arXiv preprint* *arXiv:2007.12099*, 2020. [1](#_bookmark0), [2](#_bookmark2)
22. Depu Meng, Xiaokang Chen, Zejia Fan, Gang Zeng, Houqiang Li, Yuhui Yuan, Lei Sun, and Jingdong Wang. Conditional detr for fast training convergence. In *Proceed- ings of the IEEE/CVF International Conference on Com-* *puter Vision*, pages 3651–3660, 2021. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [7](#_bookmark18)
23. Rashmika Nawaratne, Damminda Alahakoon, Daswin De Silva, and Xinghuo Yu. Spatiotemporal anomaly de- tection using deep learning for real-time video surveillance. *IEEE Transactions on Industrial Informatics*, 16(1):393– 402, 2019. [1](#_bookmark0)
24. Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. [2](#_bookmark2)
25. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information process-* *ing systems*, 28, 2015. [1](#_bookmark0)
26. Tianhe Ren, Jianwei Yang, Shilong Liu, Ailing Zeng, Feng Li, Hao Zhang, Hongyang Li, Zhaoyang Zeng, and Lei Zhang. A strong and reproducible object detector with only public datasets, 2023. [7](#_bookmark18)
27. Byungseok Roh, JaeWoong Shin, Wuhyun Shin, and Sae- hoon Kim. Sparse detr: Efficient end-to-end object detection with learnable sparsity. *arXiv preprint arXiv:2111.14330*, 2021. [3](#_bookmark3)
28. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, San- jeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy,Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. International journal of computer vision, 115:211–252, 2015. [7](#_bookmark18)
29. Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chen- feng Xu, Wei Zhan, Masayoshi Tomizuka, Lei Li, Zehuan Yuan, Changhu Wang, et al. Sparse r-cnn: End-to-end ob- ject detection with learnable proposals. In *Proceedings of the IEEE/CVF conference on computer vision and pattern* *recognition*, pages 14454–14463, 2021. [1](#_bookmark0), [2](#_bookmark2)
30. Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern* *recognition*, pages 10781–10790, 2020. [3](#_bookmark3)
31. Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceed- ings of the IEEE/CVF international conference on computer* *vision*, pages 9627–9636, 2019. [1](#_bookmark0)
32. Chien-Yao Wang, Alexey Bochkovskiy, and Hong- Yuan Mark Liao. Scaled-yolov4: Scaling cross stage partial network. In *Proceedings of the IEEE/cvf conference on com- puter vision and pattern recognition*, pages 13029–13038, 2021. [2](#_bookmark2)
33. Chien-Yao Wang, Alexey Bochkovskiy, and Hong- Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* *preprint arXiv:2207.02696*, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [4](#_bookmark8), [7](#_bookmark18), [8](#_bookmark20)
34. Jianfeng Wang, Lin Song, Zeming Li, Hongbin Sun, Jian Sun, and Nanning Zheng. End-to-end object detection with fully convolutional network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern* *recognition*, pages 15849–15858, 2021. [1](#_bookmark0), [2](#_bookmark2)
35. Yingming Wang, Xiangyu Zhang, Tong Yang, and Jian Sun. Anchor detr: Query design for transformer-based detector. In *Proceedings of the AAAI conference on artificial intelli-* *gence*, volume 36, pages 2567–2575, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [6](#_bookmark15), [7](#_bookmark18)
36. Shangliang Xu, Xinxin Wang, Wenyu Lv, Qinyao Chang, Cheng Cui, Kaipeng Deng, Guanzhong Wang, Qingqing Dang, Shengyu Wei, Yuning Du, et al. Pp-yoloe: An evolved version of yolo. arXiv preprint arXiv:2203.16250, 2022. [1](#_bookmark0), [2](#_bookmark2), [4](#_bookmark8), [7](#_bookmark18), [8](#_bookmark20)
37. Jianwei Yang, Chunyuan Li, Xiyang Dai, and Jianfeng Gao. Focal modulation networks. *Advances in Neural Information* *Processing Systems*, 35:4203–4217, 2022. [7](#_bookmark18)
38. Zhuyu Yao, Jiangbo Ai, Boxun Li, and Chi Zhang. Efficient detr: improving end-to-end object detector with dense prior. *arXiv preprint arXiv:2104.01318*, 2021. [3](#_bookmark3), [6](#_bookmark15), [7](#_bookmark18)
39. Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443– 58469, 2020. [1](#_bookmark0)
40. Fangao Zeng, Bin Dong, Yuang Zhang, Tiancai Wang, Xi- angyu Zhang, and Yichen Wei. Motr: End-to-end multiple- object tracking with transformer. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, Octo- ber 23–27, 2022, Proceedings, Part XXVII*, pages 659–675.

Springer, 2022. [1](#_bookmark0)

1. Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel Ni, and Heung-Yeung Shum. Dino: Detr withimproved denoising anchor boxes for end-to-end object de- tection. In The Eleventh International Conference on Learn- ing Representations, 2022. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [5](#_bookmark12), [6](#_bookmark15), [7](#_bookmark18), [8](#_bookmark20)
2. Haoyang Zhang, Ying Wang, Feras Dayoub, and Niko Sun- derhauf. Varifocalnet: An iou-aware dense object detector. In *Proceedings of the IEEE/CVF Conference on Computer* *Vision and Pattern Recognition*, pages 8514–8523, 2021. [6](#_bookmark15)
3. Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *Computer Vision–ECCV 2022: 17th Eu- ropean Conference, Tel Aviv, Israel, October 23–27, 2022,*

*Proceedings, Part XXII*, pages 1–21. Springer, 2022. [1](#_bookmark0)

1. Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable trans- formers for end-to-end object detection. *arXiv preprint* *arXiv:2010.04159*, 2020. [1](#_bookmark0), [2](#_bookmark2), [3](#_bookmark3), [5](#_bookmark12), [6](#_bookmark15), [7](#_bookmark18)
2. Zhuofan Zong, Guanglu Song, and Yu Liu. Detrs with collaborative hybrid assignments training. *arXiv preprint arXiv:2211.12860*, 2022. [7](#_bookmark18)