CVPDL Homework 2

Generic Object Detection

1. (10%) How does DETR [1] achieve end-to-end object detection using Transformers and bipartite matching? Briefly discuss the advantages and limitations of this approach compared to traditional convolutional detectors.

2. (10%) How does the DINO [2] model improve training efficiency and accuracy over DETR? Briefly explain how "Contrastive DeNoising," "Mixed Query Selection," and "Look Forward Twice" enhance small object detection and overall AP on the COCO dataset.

Practical Issues 1: Lightweight Computer Vision

3. (10%) How does the EdgeViT [3] model achieve efficient performance on mobile devices compared to traditional Vision Transformers (ViTs) and CNNs, and what key architectural features contribute to this efficiency?

4. (10%) What is the main design strategy of the Lite DETR [4] model to improve efficiency in Transformer-based object detection, and how does it achieve computational savings while maintaining performance?

Practical Issue 2: Data Imbalance and Domain Adaptation

5. (10%) Based on the concepts discussed in *“Class-Balanced Loss Based on Effective Number of Samples”* [5], explain the "effective number of samples" concept and how it helps address challenges associated with long-tailed data distributions.

6. (10%) How does Adversarial Domain Alignment [6] improve model performance in domain adaptation tasks?

Practical Issue 3: Weakly-/ Semi-supervised Learning

7. (10%) What is the primary difference between weakly supervised and semi-supervised learning in terms of annotation requirements for training in computer vision tasks?

8. (10%) In the context of Multiple-Instance Learning (MIL) applied to weakly supervised object detection (WSOD) [7], what is the purpose of using positive and negative "bags," and what challenges does this approach present?

9. (10%) In the MixMatch [8] algorithm for semi-supervised learning, how does the use of data augmentation and label guessing contribute to reducing the reliance on labeled data, and why are these techniques effective in leveraging unlabeled data?

Practical Issue 4: Self-supervised Learning

10. (10%) SimCLR's contrastive learning framework achieves high performance by leveraging a particular loss function known as NT-Xent. Explain the mechanism behind the NT-Xent loss and discuss why it is particularly effective for self-supervised learning.

Submission and Rules

Deadline: 2024/11/20 (Wed.) 23:59

Late policy: Refer to the previous announcement on late submission policy.

Submission Guidelines: Please submit your file to NTU Cool using the following format: hw2_<student_id>.pdf.

Helps

If you have any questions, contact TAs via this email: cvpdl.ta.2024@gmail.com

Reference

1. End-to-End Object Detection with Transformers [ECCV 2020]
2. DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection [ICLR 2023]
3. EdgeViTs: Competing Light-weight CNNs on Mobile Devices with Vision Transformers [ECCV 2022]
4. Lite DETR : An Interleaved Multi-Scale Encoder for Efficient DETR [CVPR 2023]
5. Class-Balanced Loss Based on Effective Number of Samples [CVPR 2019]
6. Adversarial Discriminative Domain Adaptation [CVPR 2017]
7. Weakly Supervised Learning in Computer Vision [ECCV 2020 tutorial]
8. MixMatch: A Holistic Approach to Semi-Supervised Learning [NeurIPS 2019]
9. A Simple Framework for Contrastive Learning of Visual Representations [ICML 2020]