

CVPDL HW3

洪郡辰

R11944050

May 21, 2023

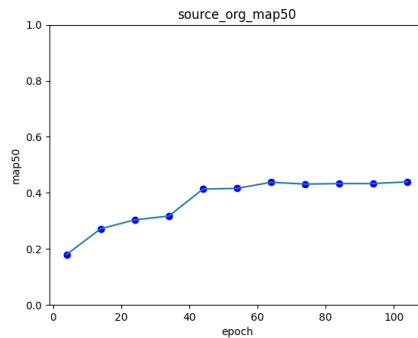
Problem

1.a

Table 1: Map of trained source model in the source validation set

Model	mAP@50	mAP@75	mAP@[50:5:95]
DETR-resnet50	0.4393	0.2537	0.4251

1.b



1.c

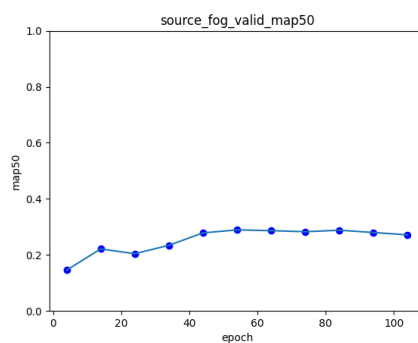
Conditional DETR - <https://arxiv.org/abs/2108.06152>

2.a

Table 2: Map of trained source model in the target validation set

Model	mAP@50	mAP@75	mAP@[50:5:95]
DETR-resnet50	0.2715	0.1773	0.2831

2.b



2.c

Conditional DETR - <https://arxiv.org/abs/2108.06152>

3.

以下是兩個兩篇 paper 的 reference:

- Improving Transferability for Domain Adaptive Detection Transformers(O2net): <https://arxiv.org/pdf/2204.14195.pdf>
- ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing: <https://arxiv.org/pdf/2210.11539.pdf>

接下來先介紹 O2net:

- O2net 是一個 detr-based 的 model，它是透過 Object-Aware Alignment(OAA) 及 Optimal Transport based Alignment(OTA) 兩種方法結合針對 CNN output 及 final output 進行 aligning 以及透過 Sequence Feature Alignment(SFA) 對每一層 encoder 及 decoder 進行 aligning 達到 domain adaptation。
- OAA 是透過 aligns foreground regions identified by pseudo-labels in the backbone outputs 進而找到 domain-invariant base features。OTA 是透過使用 sliced Wasserstein distance 最大化 retention of location information while minimizing the domain gap in the final outputs。
- SFA 是 token-wise feature alignment(TDA) 及 domain query-based feature alignment(DQFA) 的結合。DQFA 負責減少 domain discrepancy in global feature and object relation。TDA 負責 reduces domain gaps in local and instance-level feature representations in encoder and decoder by aligning token features。

接下來介紹 ConfMix:

- ConfMix 是一個 CNN-based 的 model，它是透過 novel sample mixing strategy(NSMS) 進行 domain adaptation。
- NSMS 是一個基於 region-level detection confidence for adaptive object detector learning 的一種策略。
- 首先透過 mix the local region of the target sample that corresponds to the most confident pseudo detections with a source image and apply an additional consistency loss term to gradually adapt towards the target data distribution。

透過 novel pseudo labelling scheme 設定 filter 篩選 pseudo target detections based on confidence metric，filter 會隨著 training 越來越嚴格。

confidence metric 是一個基於 detector-dependent confidence and the bounding box uncertainty 的指標。

4.

以下是 O2net 的優缺點:

- pros
 - 模型表現好
 - robustness great
 - 相較於 CNN-based 是一種新的 approach，具有更多發展性
- cons
 - 收斂慢
 - 適用於較大規模的資料集
 - source model 訓練時間長
 - 需要大量的運算資源
 - 複雜的 alignment 導致 domain adaptation 訓練時間長

以下是 ConfMix 的優缺點:

- pros
 - 可輕鬆根據需求替換 yolo model
 - 小規模資料集也能取得一定成效
 - 訓練速度快且簡單
 - 較小的 Inference Time
- cons
 - robustness 較差
 - alignment 方法單一
 - model performance 較差

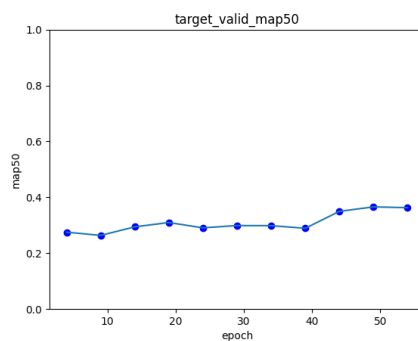
兩個方法相較之下我認為差異最大的部分在於 source model 本身對於學習的方法有差異，導致了 detr-based 能夠延伸更多不同於 CNN-based 的 domain adaptation 方法。在 ConfMix 只透過 novel sample mixing strategy 進行 domain adaptation，設定 pseudo labelling 篩選 pseudo target detection 進行學習。O2net 在每個 layer、encoder、decoder、Output 部分都有使用不同的 domain adaptation 方法，在 output 部分也有使用到類似 pseudo labelling 的方法，但在 layer 與 layer 之間又多了許多不同的 domain adaptation 方法。

5.a

Table 3: Map of adapted model in the target validation set

Model	mAP@50	mAP@75	mAP@[50:5:95]
DETR-resnet50	0.3655	0.2201	0.3691

5.b



5.c

在 map50 上升約 0.094

reference

- O2net:<https://github.com/BIT-DA/O2net>
- ConfMix:<https://github.com/giuliomattolin/ConfMix>