Bidirectional Alignment for Domain Adaptive Detection with Transformers

Liqiang He, Wei Wang, Albert Chen, Min Sun, Cheng-Hao Kuo, and Sinisa Todorovic

Previous Transformer-based methods do not explicitly consider differences between source and target domains for image patches and object tokens.

Contribution 1: We explicitly design a token-wise domain specific embedding, at the image level in the encoder, and at the object level in the decoder.

Motivation 1: Different image patches and different object tokens may exhibit different domain characteristics.

Contribution 2: We design two attention modules to align the two domains bidirectionally.

→ This is seamlessly integrated in existing attention modules.

Motivation 2: The resulting domain aligned image/object features contain both domain-invariant and domain-specific features.

Existing Transformer-based methods use GRL (gradient reverse layer) to reduce the domain gap in the image/object features.

Existing Transformer-based methods do not explicitly consider differences between source and target domains for image patches and object tokens.

Contribution 1: We explicitly design a token-wise domain specific embedding, at the image level in the encoder, and at the object level in the decoder.

Motivation 1: Different image patches and different object tokens may exhibit different domain characteristics.

→ They need to be aligned individually!

Contribution 2: We design two attention modules to align the two domains bidirectionally.

→ This is seamlessly integrated in existing attention modules.

Motivation 2: The resulting domain aligned image/object features contain both domain-invariant and domain-specific features.

Existing Transformer-based methods use GRL (gradient reverse layer) to reduce the domain gap in the image/object features.

Contribution 1: We explicitly design a token-wise domain specific embedding, at the image level in the encoder, and at the object level in the decoder.

Motivation 1: Different image patches and different object tokens may exhibit different domain characteristics.

→ They need to be aligned individually!

Contribution 2: We design two attention modules to align the two domains bidirectionally.

→ This is seamlessly integrated in existing attention modules.

Motivation 2: The resulting domain aligned image/object features contain both domain-invariant and domain-specific features.

An overview of BiADT

Comparison with SOTA on the Cityscapes — FoggyCityscapes domain shift.

The t-SNE visualization of object features of the test images in the Cityscapes — FoggyCityscapes setting.

An example of the predicted domain masks by BiADT encoder.

Method | Backbone | Detector | Pseudo-Label | person | rider | car | truck | train | motor | bike | mAP
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---
MTTrans [37] ECCV’22 | R50 | Deform-Detr | Yes | 47.7 | 49.9 | 65.2 | 25.8 | 45.9 | 33.8 | 32.6 | 46.5 | 43.4
PT | I [40] ECCV’22 | V16 | Faster R-CNN | Yes | 43.2 | 52.8 | 63.4 | 33.4 | 56.6 | 37.8 | 41.3 | 48.7 | 47.1
TDD [35] ICCV’22 | R50 | Faster R-CNN | Yes | 50.7 | 53.7 | 68.2 | 35.1 | 53.0 | 45.1 | 38.9 | 49.1 | 49.2
AT [44] CVPR’22 | V16 | Faster R-CNN | Yes | 45.5 | 55.1 | 64.2 | 35.0 | 56.3 | 54.3 | 38.5 | 51.9 | 50.9
AT* [44] CVPR’22 | V16 | Faster R-CNN | Yes | 44.1 | 54.2 | 62.7 | 33.6 | 54.4 | 51.9 | 39.2 | 49.2 | 49.5
PDN [32] ACC ’21 | R101 | Faster R-CNN | No | 32.8 | 44.4 | 60.6 | 31.0 | 46.1 | 33.8 | 29.9 | 35.7 | 36.6
ICCR-VDD [17] ECCV’21 | R50 | Faster R-CNN | No | 33.4 | 44.0 | 51.7 | 33.0 | 52.0 | 34.7 | 34.2 | 36.8 | 40.0
SFA [20] ACM MM ’21 | R50 | Deform-Detr | No | 46.5 | 48.6 | 62.6 | 25.1 | 46.2 | 29.4 | 28.3 | 44.0 | 41.3
MGADA [10] CVPR ’22 | R101 | FCOS | No | 43.1 | 47.3 | 61.5 | 30.2 | 53.2 | 50.3 | 27.9 | 36.9 | 43.8
SIGMA [63] CVPR’22 | R50 | FCOS | No | 44.0 | 43.9 | 60.3 | 31.6 | 50.4 | 51.5 | 31.7 | 40.6 | 44.2
AQ [55] IJCAI’22 | R50 | Deform-Detr | No | 49.3 | 52.3 | 64.4 | 27.7 | 53.7 | 46.5 | 36.0 | 46.4 | 47.1
AQ* [55] IJCAI’22 | R50 | DAB-Deform-Detr | No | 49.8 | 54.2 | 65.8 | 29.0 | 56.2 | 37.5 | 38.9 | 48.2 | 47.4
BiADT | R50 | DAB-Deform-Detr | No | 50.3 | 56.4 | 66.5 | 32.5 | 52.3 | 47.8 | 40.1 | 48.3 | 49.3
BiADT+AQT | R50 | DAB-Deform-Detr | No | 50.1 | 55.4 | 67.9 | 31.5 | 56.1 | 46.8 | 38.6 | 49.3 | 49.6
BiADT+TS | R50 | DAB-Deform-Detr | Yes | 52.2 | 58.9 | 69.2 | 31.7 | 55.0 | 45.1 | 42.6 | 51.3 | 50.8

Acknowledgement: USDA NIFA award No.2021-67021-35344 (AgAID AI Institute), Amazon Lab126