Caltech

Unsupervised RGB-to-Thermal Domain Adaptation via Multi-Domain Attention Network

Lu Gan, Connor Lee, and Soon-Jo Chung

ARCLab Autonomous Robotics and Control Lab

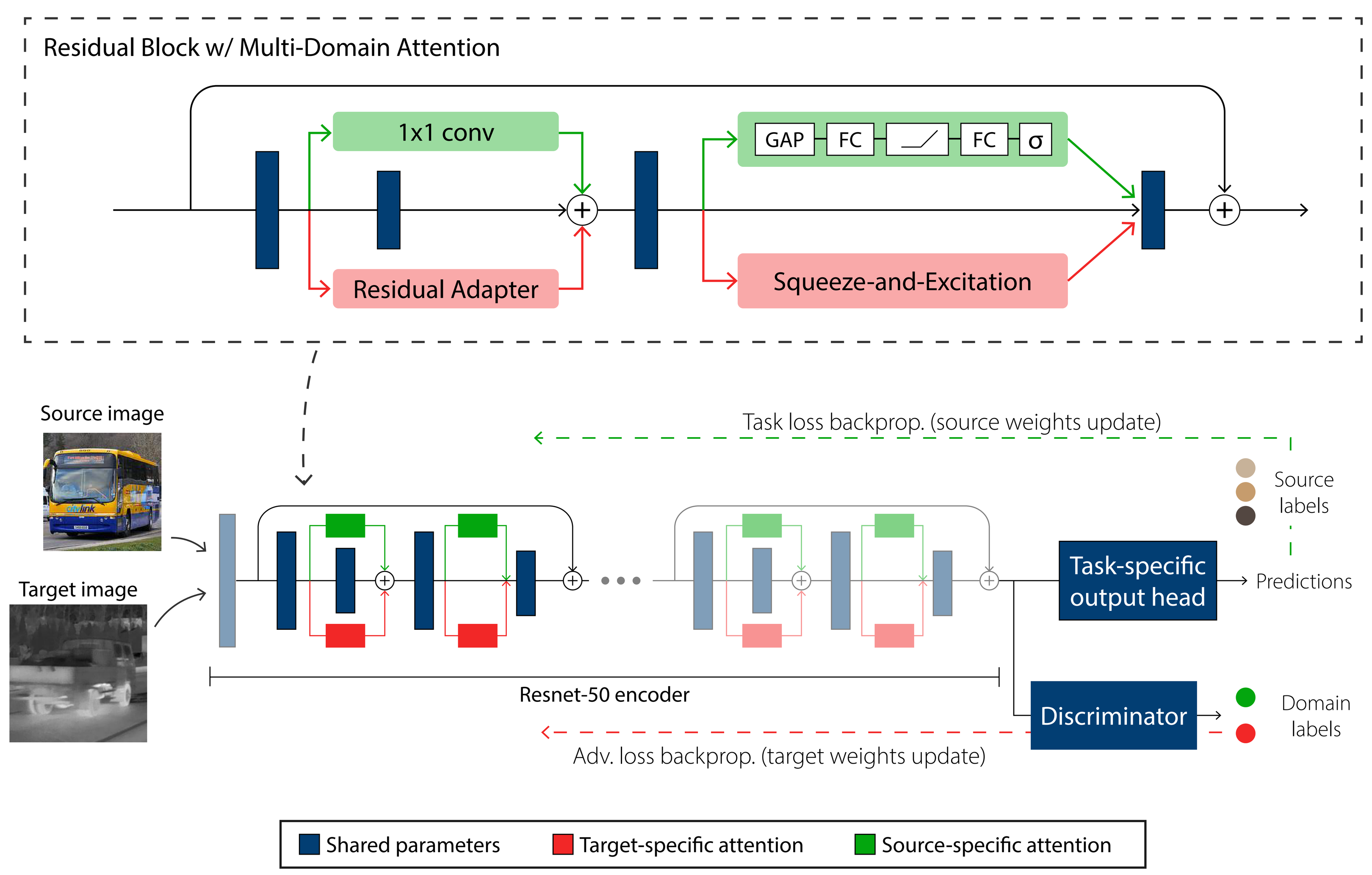
Background

- Lack of large, annotated thermal datasets prevent adoption of deep learning algorithms for nighttime robotic operations.
- Current unsupervised domain adaptation (UDA) methods align images/features across domains but struggle with cross-modal data where not all features transfer.
- We use shared CNN encoders with domain-specific attention modules that mitigate *forced transfer* by attending to domain-invariant and domain-specific features.

Multi-Domain Attention Network Architecture

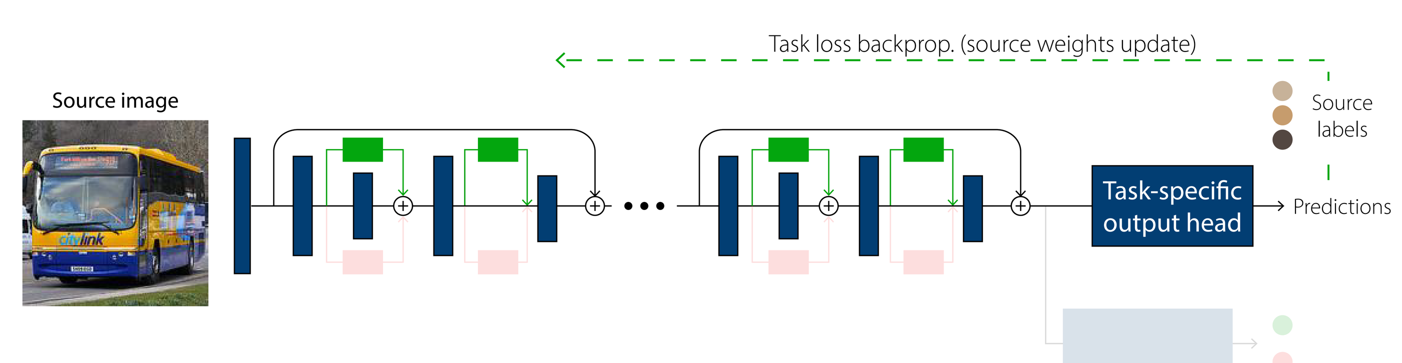
We insert source- and target-domain residual adapters and squeeze-and-excitation blocks into basic CNN blocks to enable RGB-T DA. This requires:

- Source (RGB) images use only shared and source weights.
- Target (thermal) images use only shared and target weights.

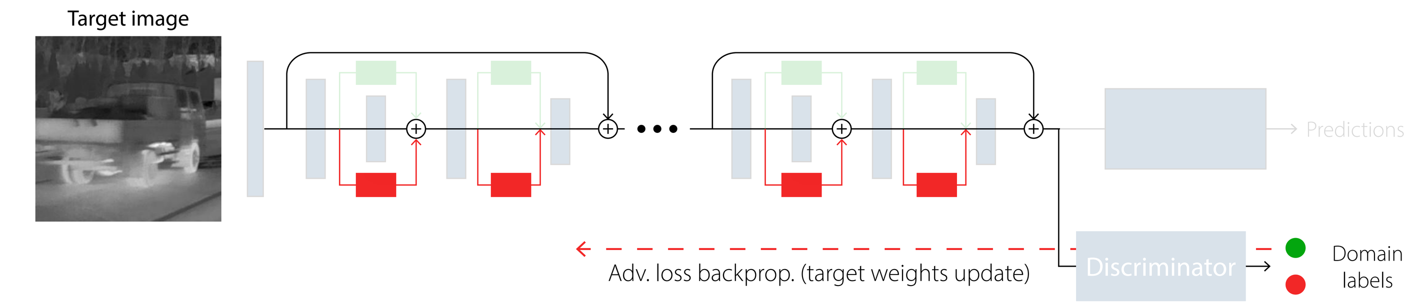


Unsupervised Training Algorithm

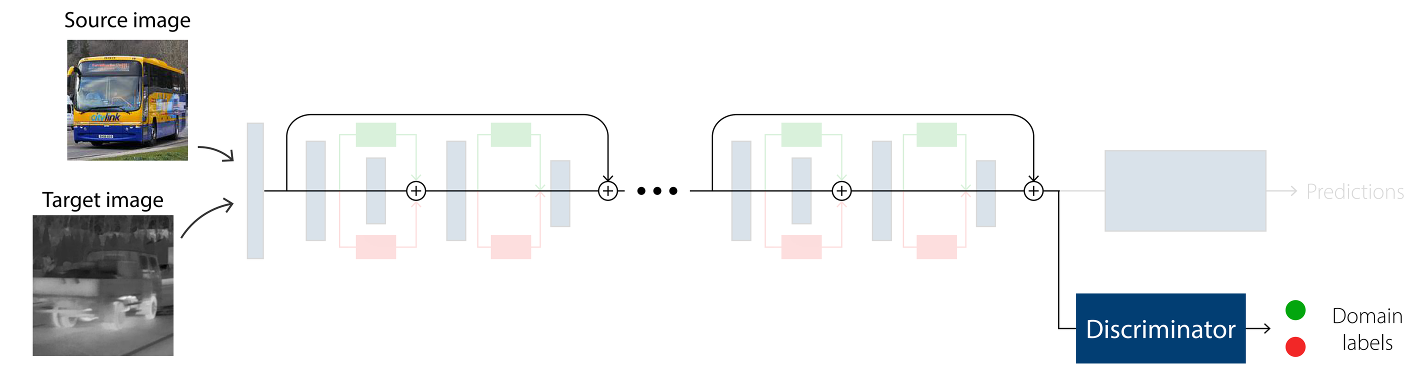
Step 1: Update shared and source weights with source task loss



Step 2: Update target weights via adversarial target domain confusion loss using a fixed discriminator



Step 3: Fix encoder and update domain discriminator. Repeat steps 1-3 until converged.



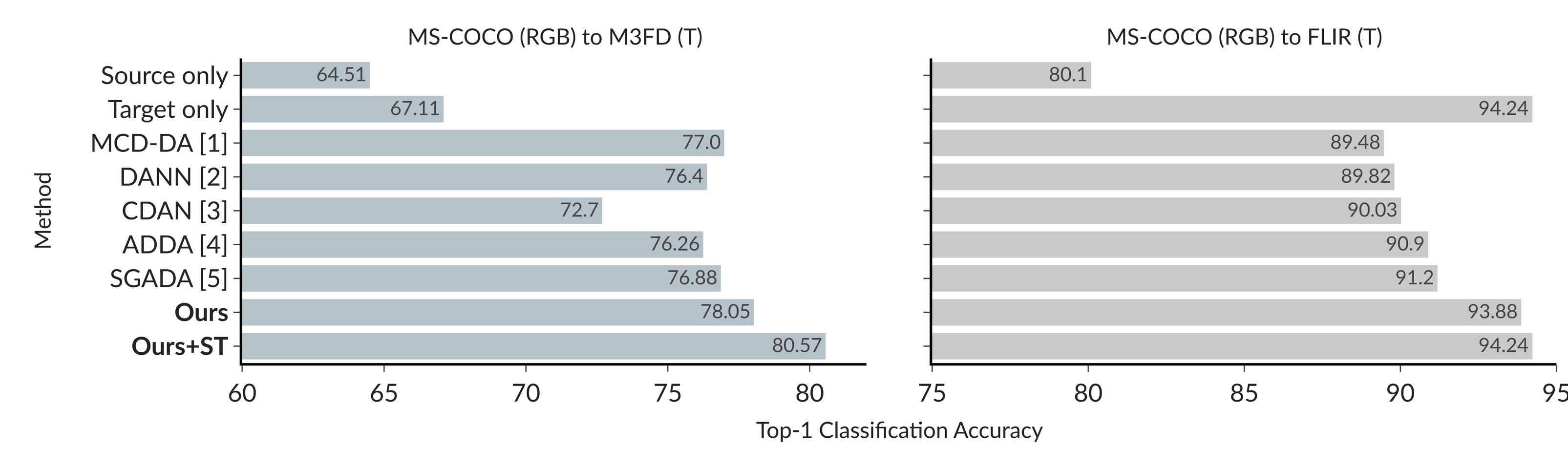
Step 4 (optional): Self-train on target domain data

- Finetune target weights using pseudolabels of target samples.
- *Good* pseudolabel = confused discriminator + high confidence.

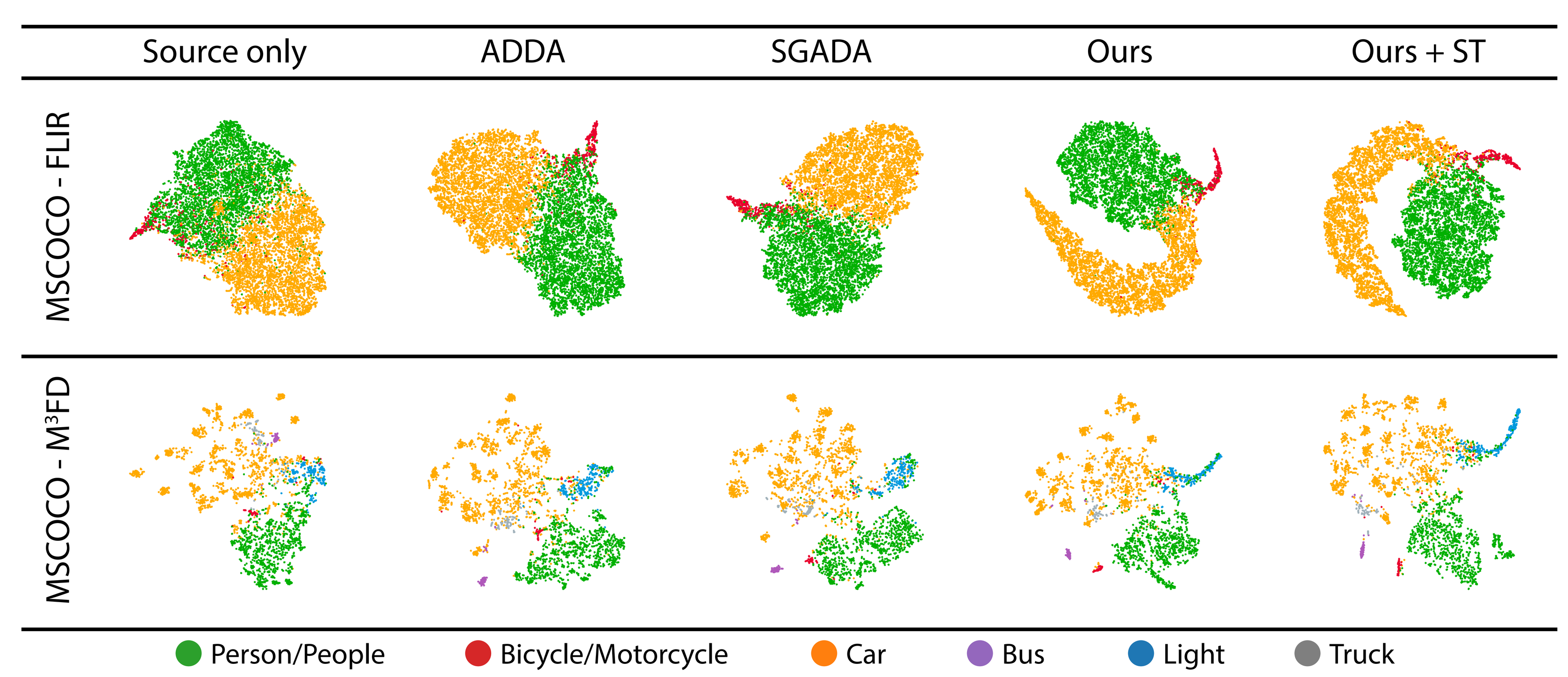
Results: Unsupervised Thermal Image Classification

We benchmark our method on image classification, using bounding box crops from object detection datasets.

- **Source data:** RGB images from MSCOCO
- **Target data:** Thermal images from (1) FLIR and (2) M^3FD datasets



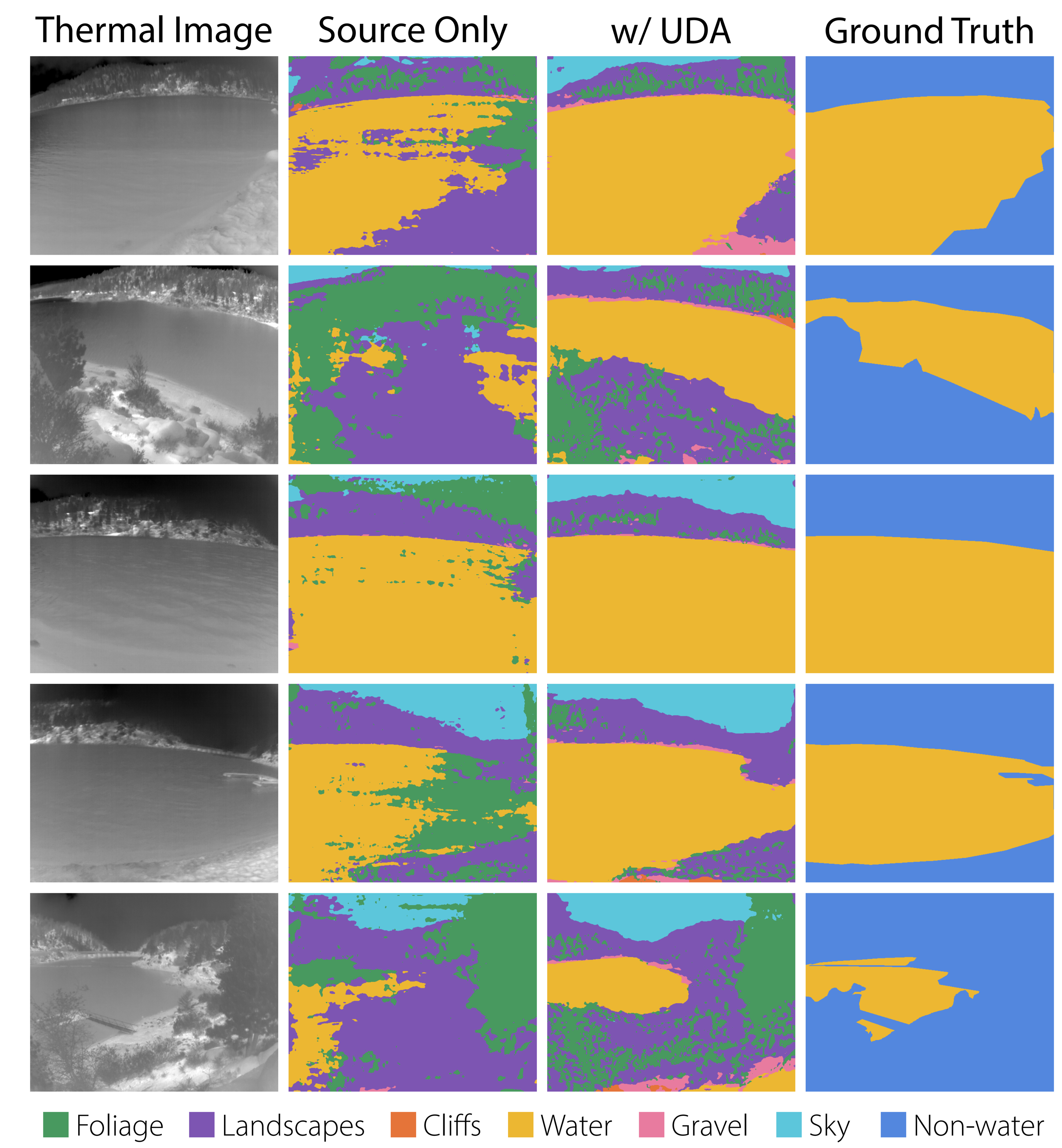
T-SNE analysis shows improved class distinction in target test samples.



Results: Unsupervised Thermal Water Segmentation

We train a DeepLabV3+ network to perform thermal water segmentation in order to assist nighttime littoral robotic operations.

- **Source data:** Synthetic grayscale images of rivers/lakes from Microsoft AirSim.
- **Target data:** Thermal lake images captured at Big Bear Lake, California



Conclusions

- RGB-T UDA using modular multi-domain attention and adversarial training.
- Outperformed other RGB-T DA methods in two classification benchmarks.
- Easily extends to other networks and tasks, like semantic segmentation.

References

[1] Ibrahim Batuhan Akkaya, Fazil Altinel, and Ugur Halici. Self-training guided adversarial domain adaptation for thermal imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4322–4331, June 2021.

[2] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

[3] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. 31, 2018.

[4] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. pages 3723–3732, 2018.

[5] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. pages 7167–7176, 2017.