Using Temporal Contrast to Detect Small Vessels in Low Resolution Optical Satellite Imagery

WenXin Dong

CS231N: Convolutional Neural Networks for Visual Recognition

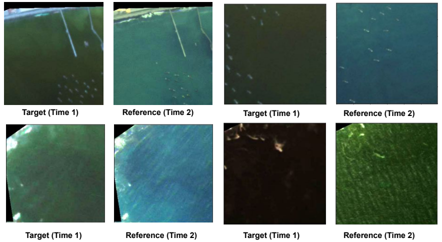
Problem

Goal: Detect small vessels in low-resolution satellite imagery

Motivation: Small vessels occupy a few pixels. Even humans easily confuse small ocean objects, such as cloud, wave highlight and reef, with small vessels. Having reference images aid human labelers, can they also aid CNN models?

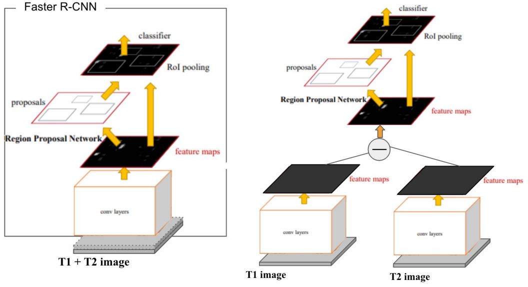
Task: Modify Faster R-CNN to intake a target image and a reference image, and produce bounding boxes for vessels in the target image.

Data



- After preprocessing, 592 target-reference image pairs of size (299, 299, 3), from 5 different locations along the Peruvian coast.
- Use 5-fold cross validation. Train on 4 locations and test on unseen location.
- Each pixel in T1 and T2 share the exact same geographical coordinates
- Data manually curated and labeled.

Methods



(a) Early Fusion

(b) Late Fusion

Methodology 1 - Early Fusion
Input 6-channel image to Faster R-CNN. Modify transformation later and first conv layer, and copy pretrained weights to new channels.

Methodology 2 - Late Fusion
Run backbone twice. Aggregate two feature maps into one as input to region proposal network.

Baselines

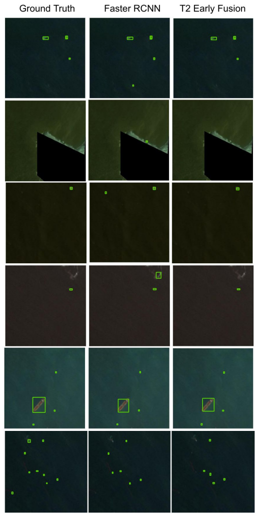
- *Vanilla Faster R-CNN*. ResNet-50 backbone, COCO pretraining.
- *Random Early Fusion*. Use a random reference image
- *Identify Early Fusion*. Use the target image itself as the reference image

Experiments

- *T2 Early Fusion*. Use target's true reference image.
- *Diff Early Fusion*. Use pixel-wise absolute difference between target and reference as the reference image.
- *Diff Late Fusion*. Use element-wise difference between two feature maps as the aggregated feature map.

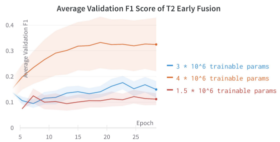
Results

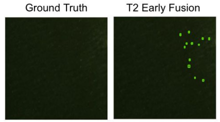
Method	AP	F1	Precision	Recall	Small Vessel Recall
Faster R-CNN	**0.35**	**0.50**	0.49	**0.65**	**0.53**
Random Early Fusion	0.20	0.31	0.29	0.48	0.41
Identity Early Fusion	0.33	0.47	0.49	0.63	0.52
T2 Early Fusion	0.27	0.41	**0.53**	0.60	0.44
Diff Early Fusion	0.15	0.26	**0.55**	0.30	0.17
Diff Late Fusion	0.23	0.37	**0.55**	0.41	0.22



- Using transfer learning, Faster R-CNN is a strong baseline.
- Fusion methods produce less false positives than vanilla Faster R-CNN at the expense of recall.
- Late Fusion underperforms Early Fusion, may due to information loss in the feature maps.
- Early Fusion methods were prone to overfit to noisy temporal features. A common mistake in Early Fusion models is detecting shifted wave highlights as vessels.
- Both Early Fusion and Late Fusion benefit from training end-to-end instead of freezing backbone or partial backbone, this may due to large input domain shift.

Average Validation F1 Score of T2 Early Fusion





Misclassifying shifted wave highlights as vessels

Discussion and Future Work

- *More data.* Our current 592-images dataset is too small to examine the full potential of Early Fusion and Late Fusion, especially when training a 4 million parameter model.
- *Attention-based CNN.* While our focus is to explore the potential of Faster R-CNN, we could benchmark with attention-based CNNs.
- *Non-linear aggregation.* We tried using convolutional layer to aggregate feature maps, which resulted in significantly more overfitting compared to the Diff Late Fusion method. However, non-linear aggregation might work better for a larger dataset.