BAE-NET: A BAND ATTENTION AWARE ENSEMBLE NETWORK FOR HYPERSPECTRAL OBJECT TRACKING

Zhuanfeng Li [1] , Fengchao Xiong [1,*] , Jun Zhou [2] , Jing Wang [2] , Jianfeng Lu [1] , and Yuntao Qian [3]
1 School of Computer Science and Engineering, Nanjing University of Science and Technology
2 School of Information and Communication Technology, Griffith University
3 College of Computer Science, Zhejiang University

Introduction

Tracking on color videos have made great progress, but it has intrinsic limitation in depicting the physical properties of target.

Compared to color images, hyperspectral images (HSIs) record continuous spectral reflectance of targets in light wavelength indexed band images. Qian and Xiong et.al proposed hand-crafted features for object tracking but cannot describe the inherent nature of the data. Uzkent et al. proposed a deep kernelized correlation filter method for tracking but losed valuable spectral information with HSI converted into 3-channel image.

Alternatively, an HSI can be divided into a series of three-channel images with band selection to select a number of top-ranked bands in order of their importance.

Purpose

Hyperspectral videos contain images with a large number of light wavelength indexed bands that can facilitate material identification for object tracking.

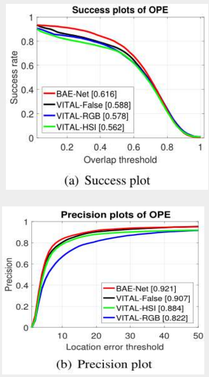
Most hyperspectral trackers use hand-crafted features rather than deep learning generated features for image representation due to limited training samples.

To fill this gap, this paper introduces a band attention aware ensemble network (BAE-Net) for deep hyperspectral object tracking, which takes advantages of deep models trained on color videos for feature representation.

Methods

- BAE-Net integrates the band attention mechanism into a VITAL tracker which enhances positive sample data using adversarial learning and deals with class imbalance using a cost-sensitive loss function to suppress negative samples.
- Specifically, an autoencoder-like architecture is employed to learn the nonlinear relationship among bands and generate their significance while reducing the influence of noises.
- Subsequently, the spectral bands are reordered according to their importance and partitioned into a number of three-channel images.
- These images are then passed through VITAL tracker, producing a set of weak trackers which are finally combined via ensemble learning to determine the location of target.

Results



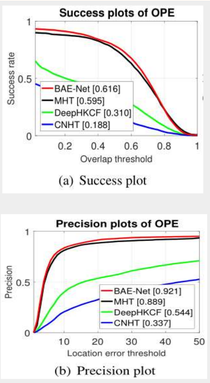
Comparison with baseline VITAL tracker with respect to precision plot and success plot.

- BAE-Net ranks top among all the trackers by achieving 0.616 on AUC metric.
- VITAL-False and BAE-Net outperform VITAL-HSI.
- Not all the bands contribute equally to tracking but band ranking can suppress the uninformative bands.

Video	BAE-Net	C-COT	ECO	TRACA
Color	n/a	0.618	0.586	0.566
Hyperspectral/False-color	0.616	0.568	0.587	0.517
	CFNet	DSiam	DeepSRDCF	
Color	0.590	0.552	0.594	
Hyperspectral/False-color	0.519	0.464	0.569	

Table 1. Comparison of AUC with deep color trackers. Red and blue mark the top two values.

- All these alternative trackers were run on both color videos and false-color videos.
- Almost all trackers provide better AUCs on color videos because HSI are of lower spatial resolution and contain more noises.
- BAE-Net achieves the best AUC on HSI videos thanks to the embedded band attention module and ensemble learning.



Comparison with hyperspectral trackers.

- BAE-Net ranks the top as it considers the importance of different bands before deep feature extraction and fuse the tracking results from weak trackers using ensemble learning.
- MHT obtains better results than DeepHKCF and CNHT since it considers spectral-spatial structure of HSIs.

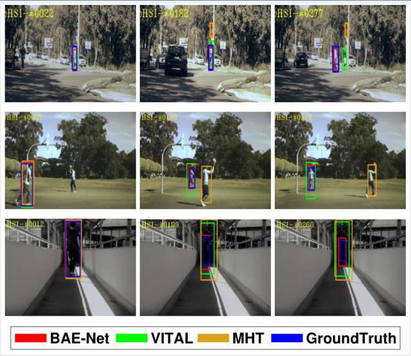
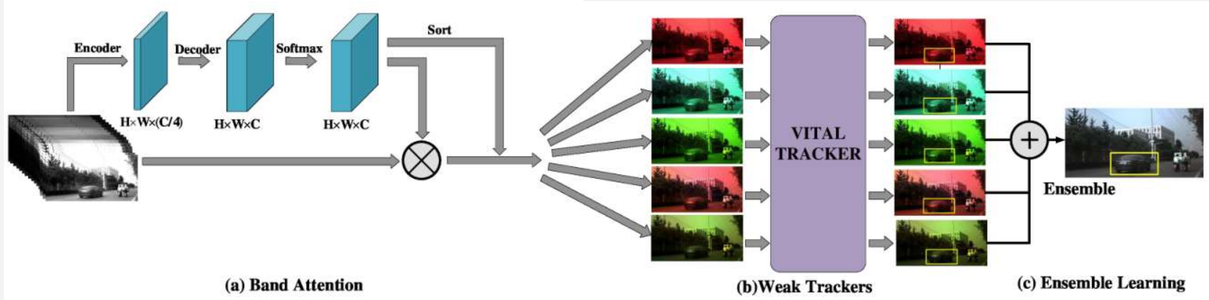
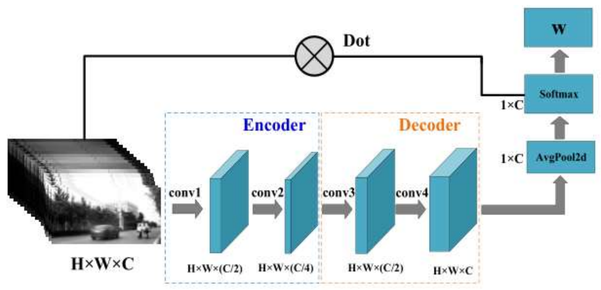


Fig. Demonstration of visual tracking results. HSIs, which visualizes the tracking results on pedestrian2, playground and student sequences.



An autoencoder-like architecture is used to produce band ranking. The sorted bands are split into a group of three-channel images, which are fed into VITAL to produce weak trackers. Ensemble learning combines weak trackers for object localization.



Band attention block

The encoder-decoder and average pooling respectively correspond to "Excitation" and "Squeeze". All the convolution operators are of size 1×1.

Conclusion

In this paper, we have introduced a BAE-Net for hyperspectral object tracking. Firstly, the HSIs are split into a group of three-channel images according to the band-wise importance learned by autoencoder-like band attention module. Then the split three-channel images are fed into a deep color tracking network to generate a set of weak trackers. These weak trackers are fused by ensemble learning. Finally, all parameters are learned in an end-to-end manner. Experimental results have proved that proposed method achieves better results than deep color trackers and hyperspectral trackers.