IEEE International Conference on Image Processing

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, China

[2]School of Information and Communication Technology, Griffith University, Australia

[3]College of Computer Science, Zhejiang University, China

IEEE ICIP, 2020

Outline

1. **Introduction**
   - Object tracking methods
   - Problem statement
   - What is hyperspectral image
   - Material-based object tracking example

2. **Motivations**
   - Overview of HSI object tracking algorithm
   - Re-arranging the band images into three-channel groups
   - Deep neural network transforms hyperspectral images

3. **Our BAE-Net Method**

4. **Experimental Results**

5. **Conclusions & Future Work**

Introduction

◘ Object tracking methods

➢ Hand-crafted feature based correlation filters

- ◆ KCF: KCF is a kernelized version of correlation filter in which kernel trick is applied to achieve non-linear classification boundaries.
- ◆ CSR-DCF: By introducing the channel and spatial reliability , DCF tracking is efficient and seamless integration in the filter update and the tracking process.

➢ Deep Learning Feature

- ◆ Dsiam: DSiam is equipped with a fast general transformation learning model to consider the temporal variations of both foreground and background during online tracking.
- ◆ VITAL: On the basis of MDNet, GAN network is used to expand positive and negative samples.

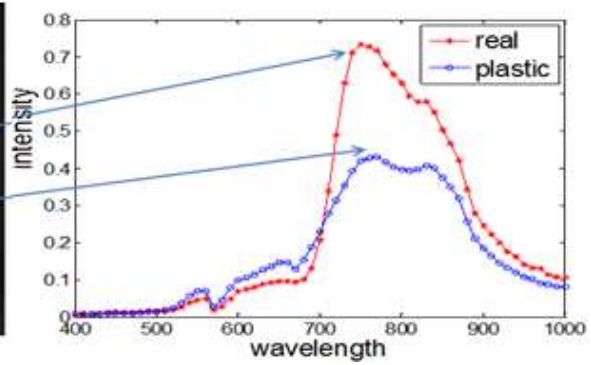
Introduction

◘ Problem Statement

- Tracking in color videos has intrinsic limitation in describing the physical properties of target, making the tracker vulnerable in complex scenarios with cluttered background and significant object shape change.

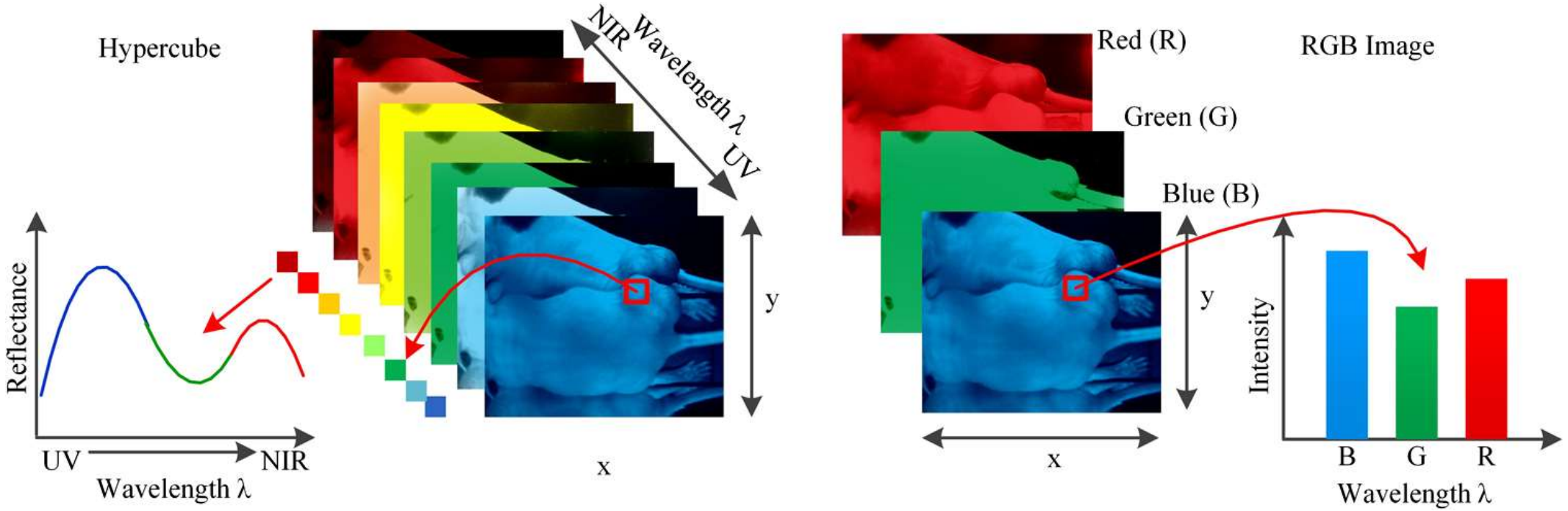
- **As shown in the picture on the right, can the tracking effect be guaranteed when the color background and texture are very similar ?**

(e.g. similar appearance and texture.)



Introduction

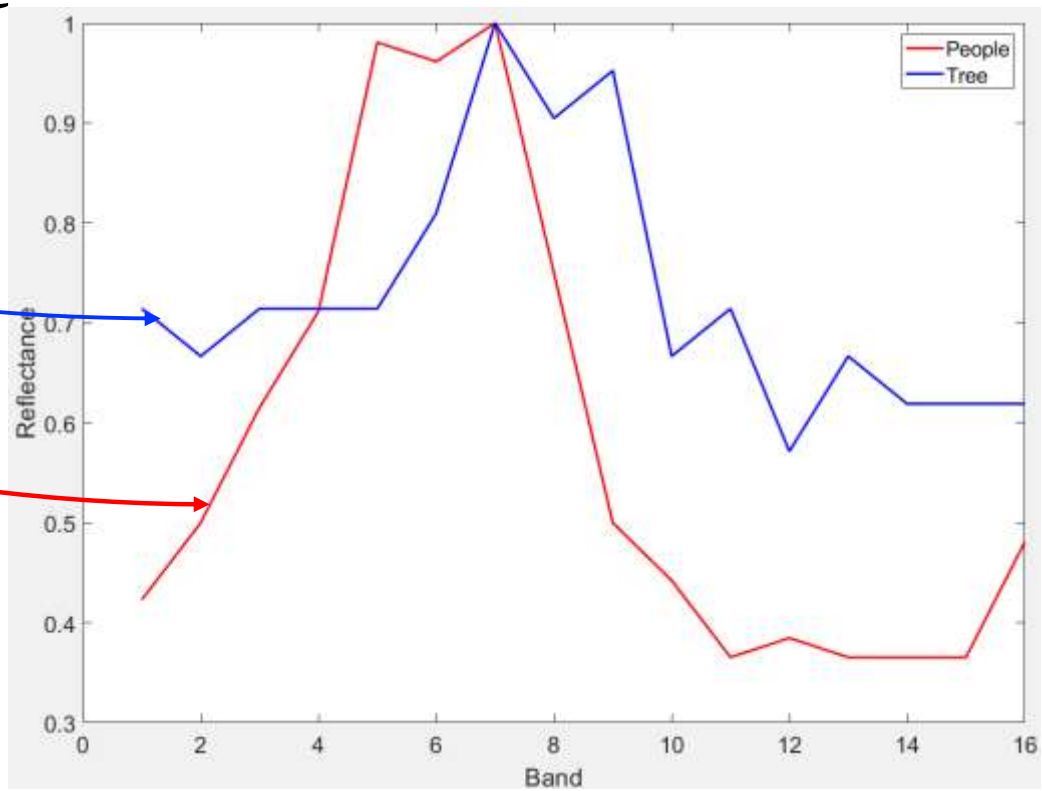
What is hyperspectral image?



- Hyperspectral images (HSIs) record continuous spectral reflectance of targets in light wavelength indexed band images.
- The spectral information enables material identification which dramatically increases the discriminative capability of HSIs and benefits object tracking.

Introduction

- Material-based object tracking example



The spectral curves of trees and people are very different.

Motivations

◻ Overview of HSI object tracking algorithm

➢ Hand-crafted features

- ◆ Qian et al. extracted a set of patches in every band as convolutional kernels but ignored spectral correlation in the image.
- ◆ Xiong et al. proposed a material based hyperspectral tracker which employs a histogram of multi-dimensional oriented gradients and abundances.

✓ **Hand-crafted features can not sufficiently describe the inherent nature of the data.**

➢ Deep features

- ◆ Due to limited HSI training data, Uzkent et al. proposed a deep kernelized correlation filter method in which HSIs are converted into a false-color image and then passed through VGGNet.

✓ **The converted false-color image loses valuable spectral information, compromising the tracking performance.**

Hyperspectral datasets are scarce but there are many RGB data sets. How to make better use of the models trained on RGB datasets?

Motivations

◘ Re-arranging the band images into three-channel groups is a key issue.

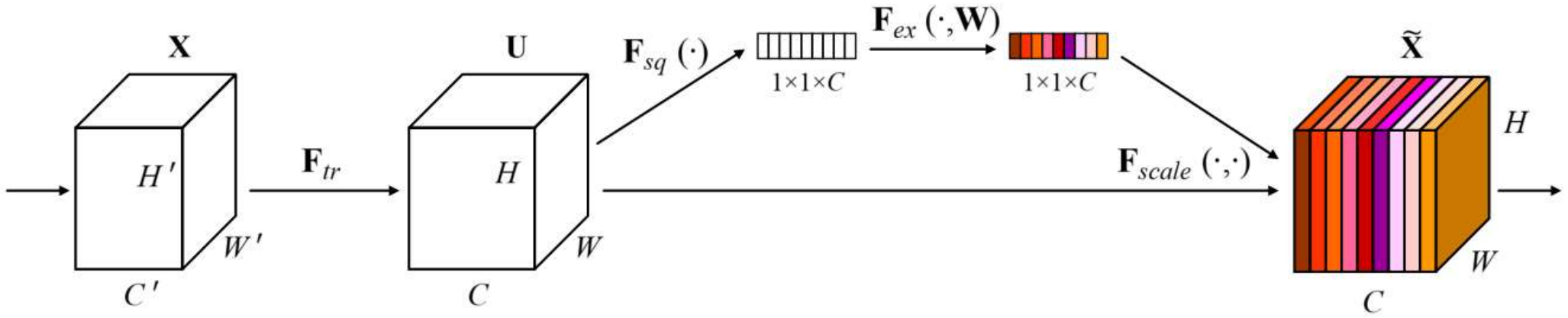
- A naive option is to group adjacent bands sequentially based on their indices. But adjacent bands are normally highly correlated, producing redundant bands in a group.

- Band grouping can be done according to the importance of each band.
  - ✓ Ranking based band selection selects a number of top-ranked bands in order of their importance.

- Driven by the success of deep learning and nonlinear features learning.
  - ✓ Deep attention mechanism is also adopted and trained in an end-to-end manner to model the nonlinear relationship between spectral bands.

Motivations

- **Deep neural network transforms hyperspectral images**



$$\mathbf{u}_c = \mathbf{v}_c * \mathbf{X} = \sum_{s=1}^{C'} \mathbf{v}_c^s * \mathbf{x}^s$$

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z}))$$

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j)$$

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, s_c) = s_c \, \mathbf{u}_c$$

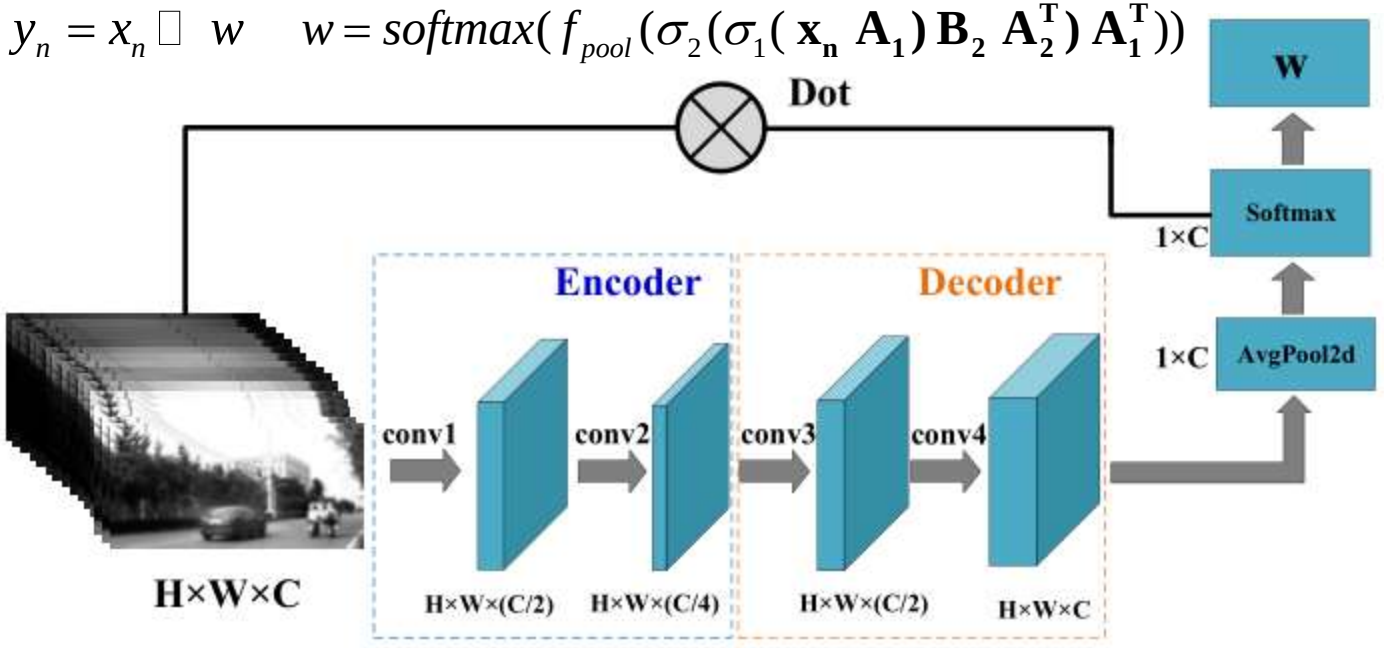
Motivations

Ideas

- **Autoencoder-like architecture is employed to learn the nonlinear relationship among bands and generate their significance while reducing the influence of noises.**

- **Achieve a balance between accuracy and efficiency**

$$y_n = x_n \square\ w \qquad w = softmax(f_{pool}(\sigma_2(\sigma_1(\mathbf{x_n}\ \mathbf{A_1})\mathbf{B_2}\ \mathbf{A_2^T})\mathbf{A_1^T}))$$

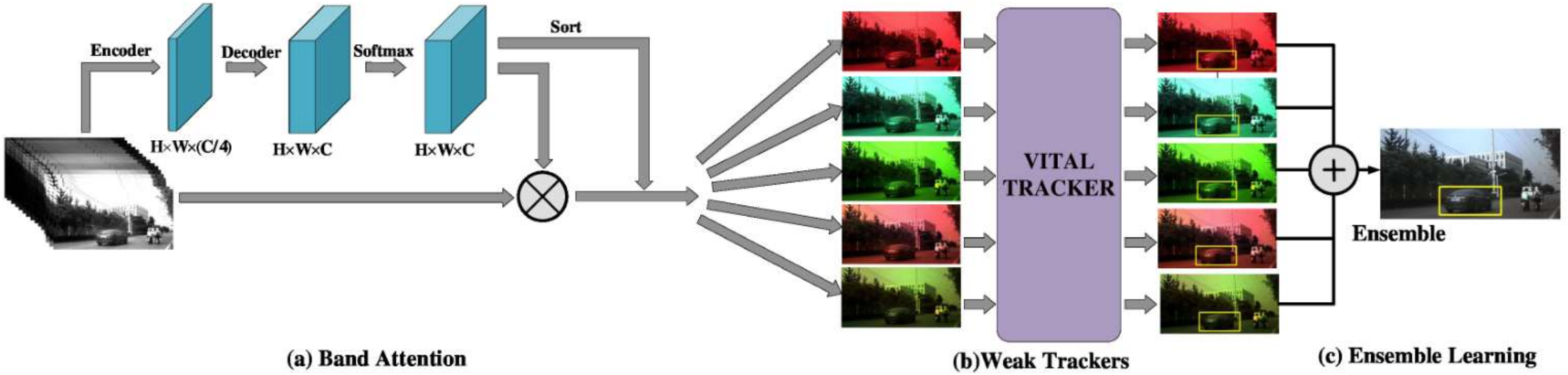


Q: How can the multiple sets of results produced after HSI conversion be integrated into the final result?

A: Ensemble Learning! We choose the mean ensemble method.

Our Approach

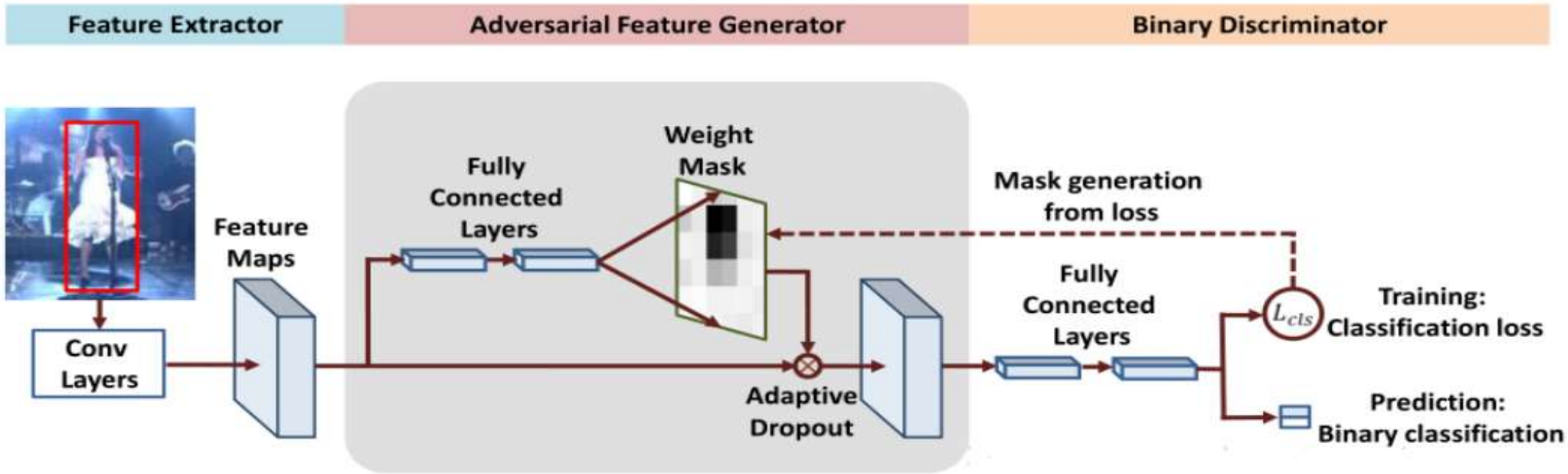
Our architecture



The architecture of BAE-Net. 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.

Our Approach

VITAL Architecture



Our Approach

Loss Function

$$L_{VITAL} = \min_{G} \max_{D} E_{(F,W)\sim P_{(F,W)}}[S_1 \cdot \log D(W \cdot F)] + E_{F\sim P_{(F)}}[S_2 \cdot \log(1 - D(G(F) \cdot F))] + \lambda E_{(F,W)\sim P_{(F,W)}} \left\| G(F) - W \right\|^2$$

$$L = \frac{1}{\lfloor L/3 \rfloor} \sum_{i=1}^{\lfloor L/3 \rfloor} L_i$$

M : the actual mask.

• : the dropout.

G(**F**) : the output features after the generator operates the feature **F**.

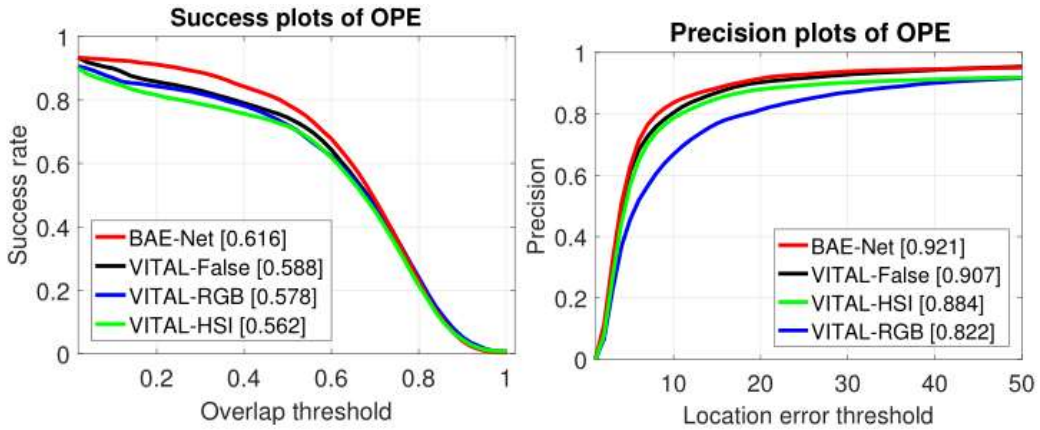
D(**N**) : the output features after the discriminator operates the feature **N**.

S_1 and S_2 : training sample loss.

L: number of bands.

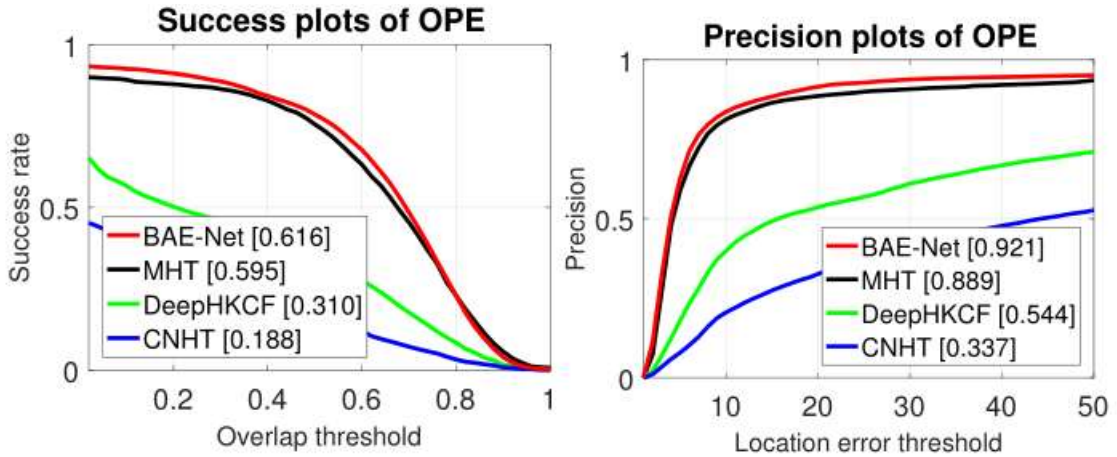
Experimental Results

Evaluation on HSI Datasets (Quantitative)



(a) Success plot (b) Precision plot

Fig. 3. Comparison with baseline VITAL tracker with respect to precision plot and success plot.



(a) Success plot (b) Precision plot

Fig. 4. Comparison with hyperspectral trackers.

Experimental Results

Evaluation on HSI Datasets (Quantitative)

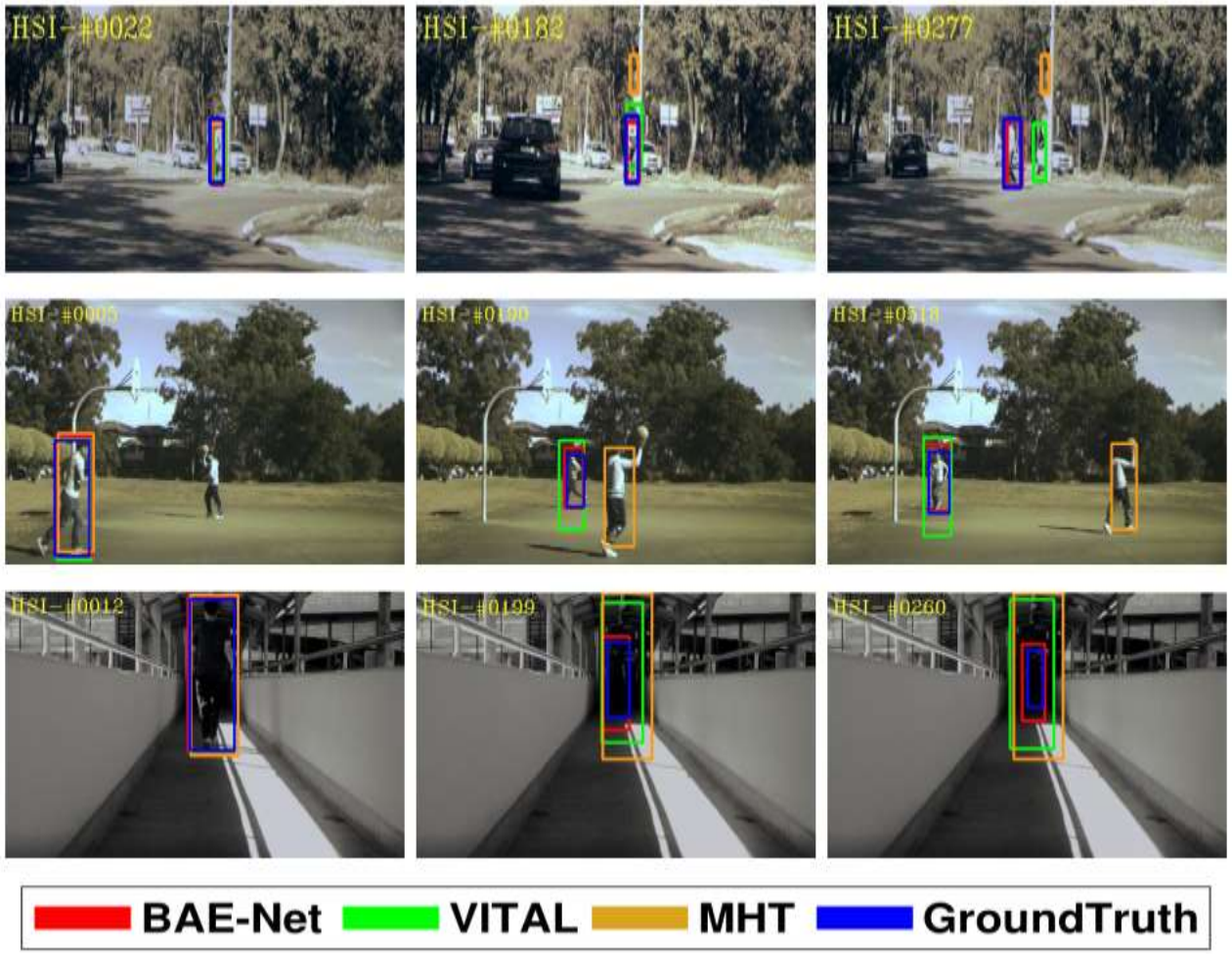
Comparison of AUC with deep color trackers. Red and blue mark the top two values.

Video	BAE-Net	C-COT [24]	ECO [25]	CFNet [7]	DSiam [6]	DeepSRDCF [26]	TRACA [27]
Color	n/a	**0.618**	0.586	0.590	0.552	**0.594**	0.566
Hyperspectral/False-color	**0.616**	0.568	**0.587**	0.519	0.464	0.569	0.517

As shown in Table, almost all trackers provide better AUCs on color videos because hyperspectral or false-color videos are of lower spatial resolution and contain more noises. The proposed BAE-Net achieves the best AUC on hyperspectral videos thanks to the embedded band attention module and ensemble learning.

Experimental Results

◘ Evaluation on HSI Datasets (Quantitative)



Conclusion & Future Work

Conclusions

- **Band ranking can suppress the uninformative bands and enhance learning from dominative informative bands.**
- **BAE-Net achieves better results than deep color trackers and hyperspectral trackers.**

Future Work

- **To focus on deep ensemble learning to increase tracking accuracy.**
- **To improve network operation efficiency and reduce running time.**

Reference

[1] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” IEEE TPAMI, vol. 37, no. 3, pp. 583–596, March 2015.

[2] Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, “Discriminative correlation filter tracker with channel and spatial reliability,” IJCV, 2018.

[3] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang, “Learning dynamic siamese network for visual object tracking,” in IEEE ICCV, Oct 2017.

[4] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R. W. H. Lau, and M. Yang, “VITAL: Visual tracking via adversarial learning,” in IEEE CVPR, 2018.

[5] W. Sun and Q. Du, “Hyperspectral band selection: A review,” IEEE GRSM, vol. 7, no. 2, pp. 118–139, June 2019.

[6] J. Wang, J. Zhou, and W. Huang, “Attend in bands: hyperspectral band weighting and selection for image classification,” IEEE JSTARS, vol. 12, no. 12, pp. 4712–4727, 2019.

[7] Y. Cai, X. Liu, and Z. Cai, “BS-Nets: An end-to-end framework for band selection of hyperspectral image,” IEEE TGRS, vol. 58, no. 3, pp. 1969–1984, 2020.

[8] L. Mou and X. X. Zhu, “Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification,” IEEE TGRS, vol. 58, no. 1, pp. 110–122, Jan 2020.

[9] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” IEEE TPAMI., pp. 1–1, 2019.

Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, and S. Wang, “Learning dynamic siamese network for visual object tracking,” in IEEE ICCV, Oct 2017.

[10] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, “End-to-end representation learning for correlation filter based tracking,” in IEEE CVPR, July 2017.

[11] M. Danelljan, A. Robinson, F. K. Shahbaz, and M. Fels- berg, “Beyond correlation filters: Learning continuous convolution operators for visual tracking,” in ECCV, 2016.

[12] M. Danelljan, G. Bhat, F. Shahbaz K., and M. Felsberg, “ECO: Efficient convolution operators for tracking,” in IEEE CVPR, July 2017.

[13] M. Danelljan, G. Hager, F. Shahbaz K., and M. Fels-berg, “Convolutional features for correlation filter based visual tracking,” in IEEE ICCV Workshops, 2015.

[14] J.Choi, H.J.Chang, T.Fischer, S.Yun, K.Lee, J.Jeong, Y. Demiris, and J. Y. Choi, “Context-aware deep feature compression for high-speed visual tracking,” in IEEE CVPR, 2018.

Thanks!
Q&A