



Wanyi Zhuang, Qi Chu\*, Zhen Tao Tan, Qian Kun Liu, Hao Jie Yuan, Chang Tao Miao, Zi Xiang Luo, and Neng Hai Yu  
CAS Key Laboratory of Electromagnetic Space Information, University of Science and Technology of China

## Introduction

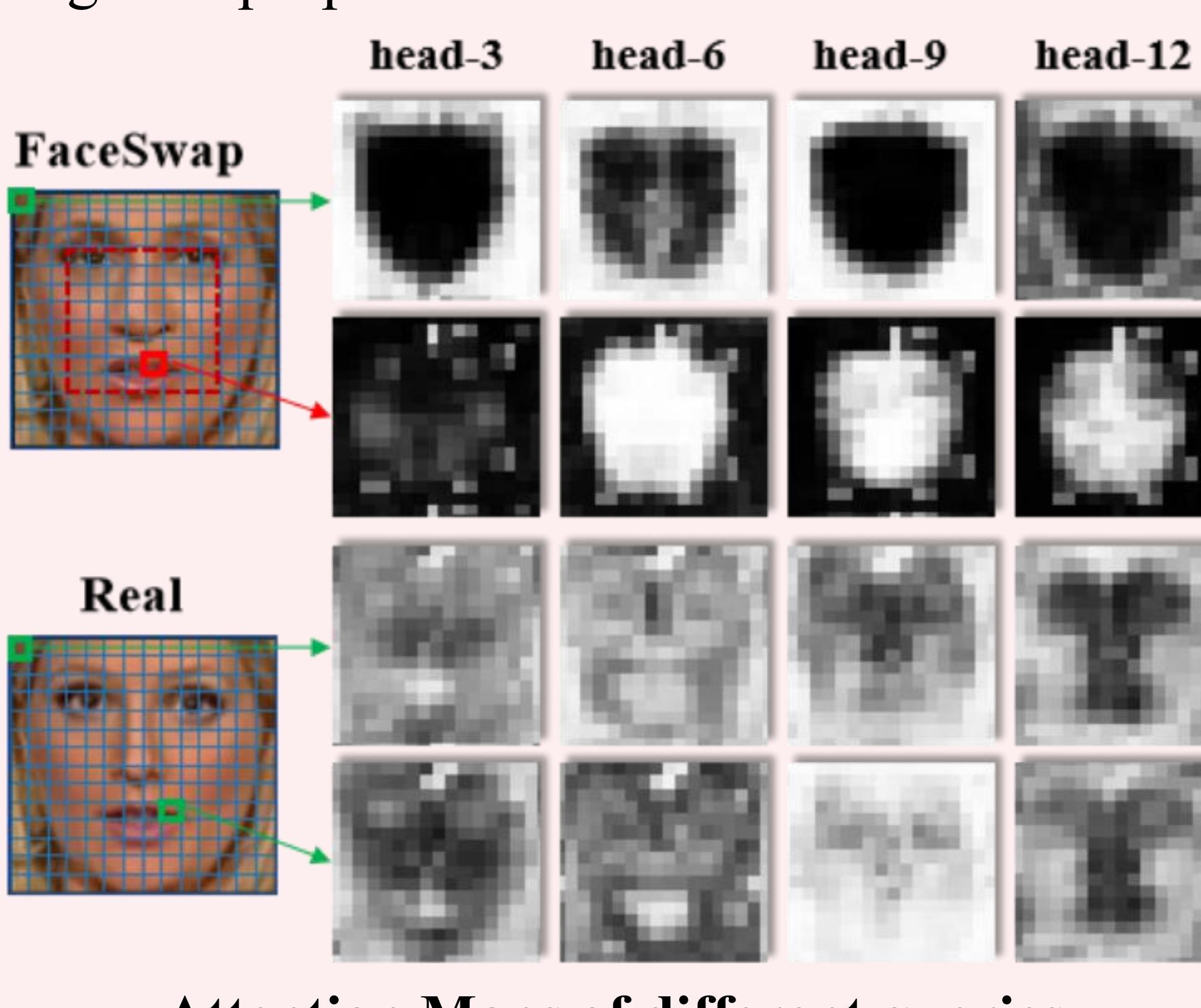
**Intra-frame inconsistency** has been proved to be effective for the generalization of face forgery detection. However, existed inconsistency-aware methods require extra pixel-level forged location annotations, which is non-trivial.

### How to locate forgery region without pixel-level supervision?

Our proposed **UIA-ViT** only makes use of video-level labels and can learn inconsistency-aware feature without pixel-level annotations. We propose two key components:

**UPCL**(Unsupervised Patch Consistency Learning) is a training strategy for learning the consistency-related representations through an unsupervised forgery location method. We approximately estimate forgery location maps by comparing the Mahalanobis distances between the MVGs of real/fake features and the patch embedding from the middle layer of ViT.

**PCWA**(Progressive Consistency Weighted Assemble) is a feature enhancement module and can take full advantage of the consistency representation through the proposed UPCL module.



## Contact

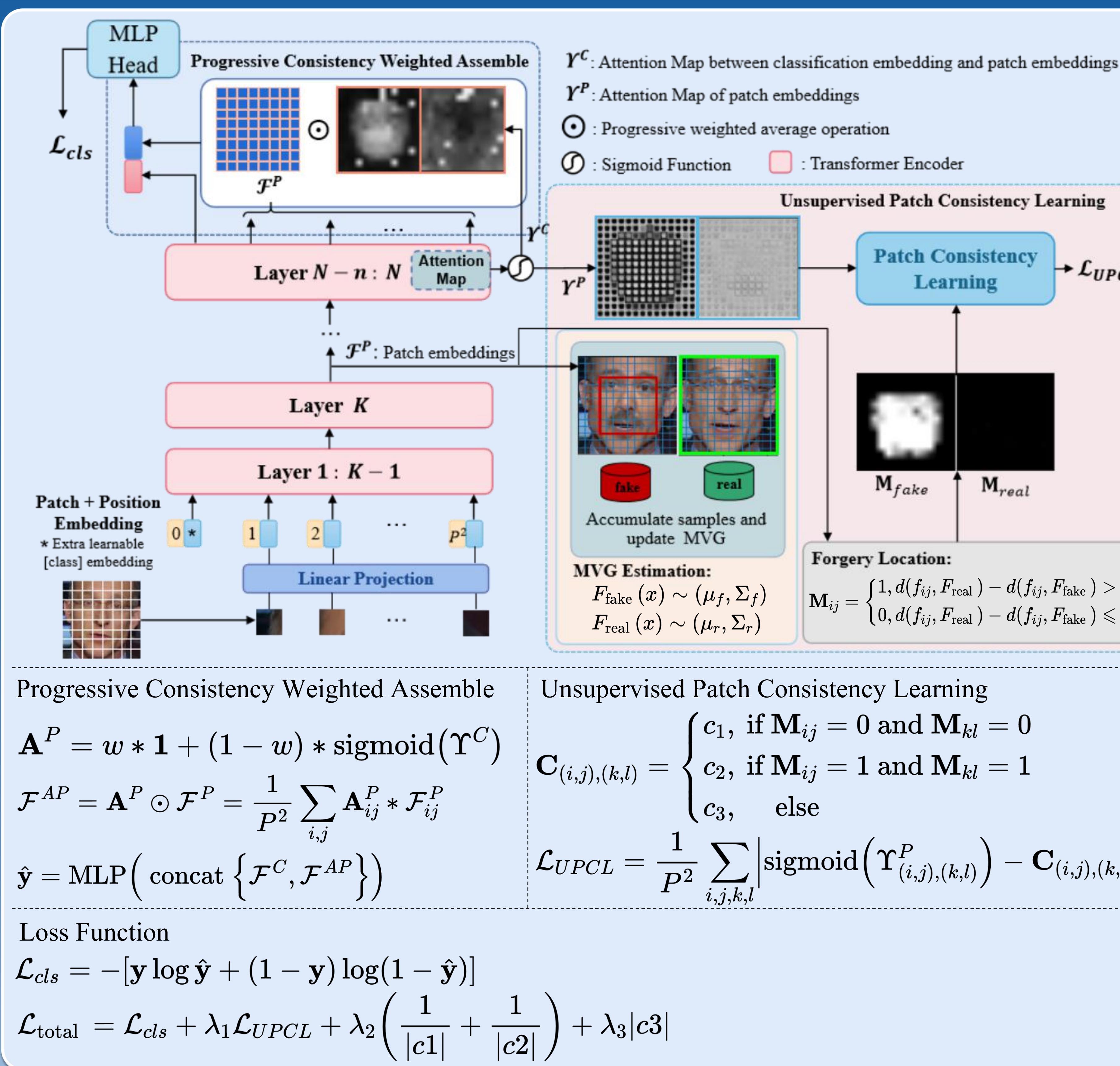
Qi Chu : qchu@ustc.edu.cn

Wanyi Zhuang : wy970824@mail.ustc.edu.cn

code github: <https://github.com/wany0824/UIA-ViT>

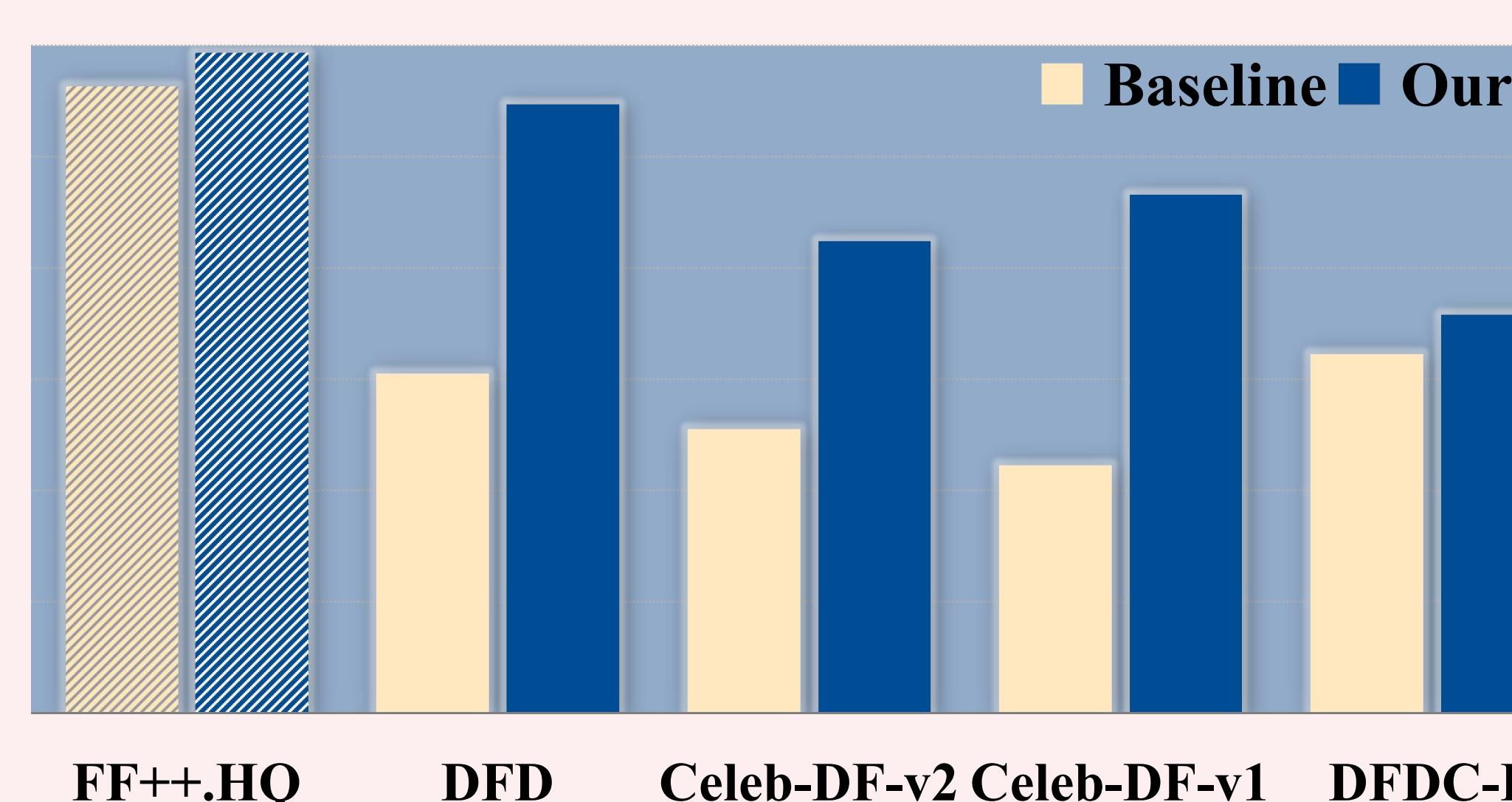
## for Face Forgery Detection

## Methods



## Results

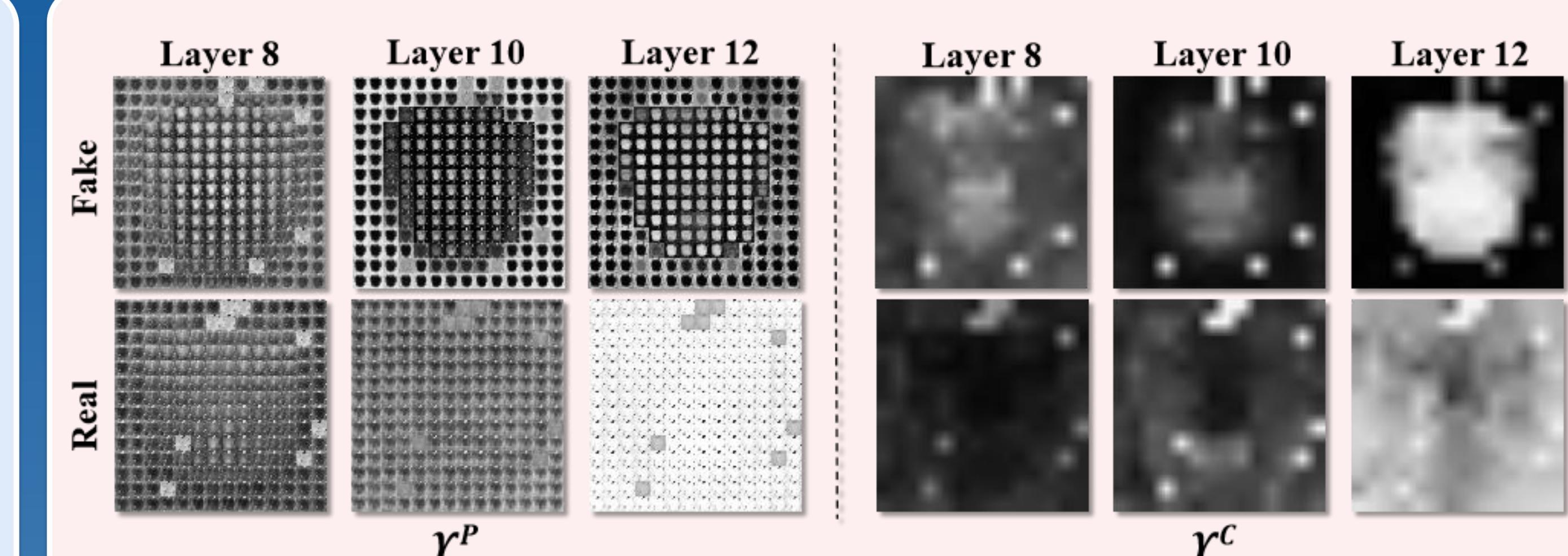
Train on FF++, test on other unseen datasets



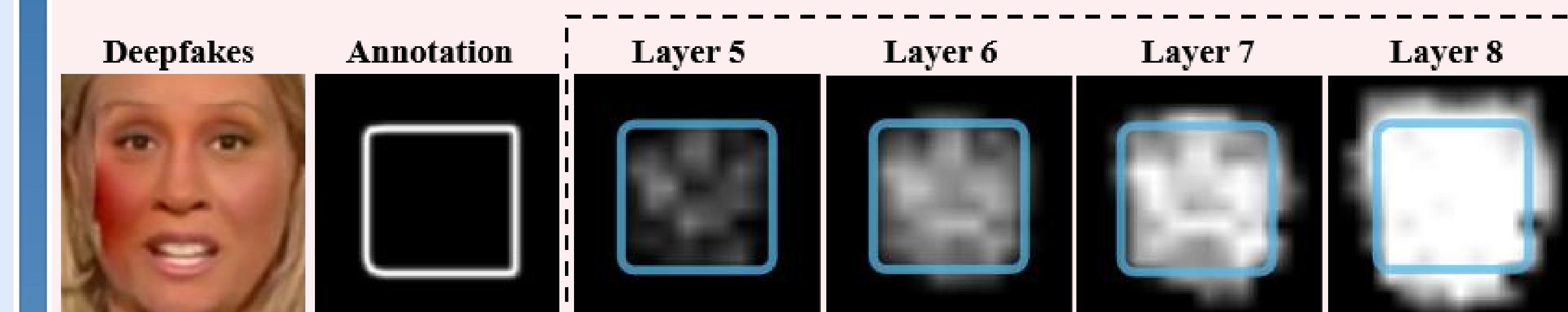
## Ablation Study

	AUC(%)	CDF-v1	CDF-v2
baseline	75.32	76.25	
+ UPCL	82.96	80.86	
+ UPCL + PCWA	<b>86.59</b>	<b>82.41</b>	

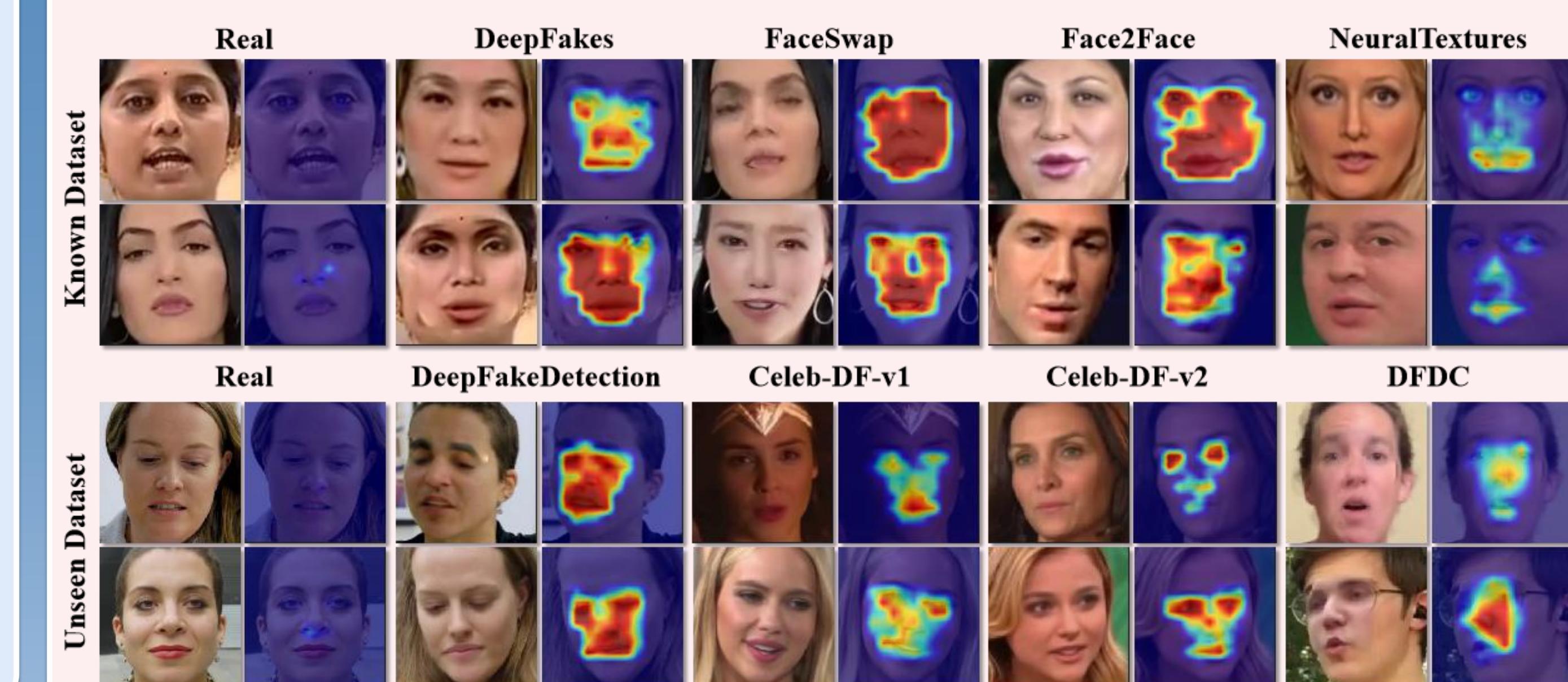
## Visualization



Attention Map of layer 8/10/12 (averaged from multiple heads).



Predicted Location Map of different layers.



Predicted Location Map of different datasets.

## Conclusion

- Due to the self-attention mechanism, Vision Transformer is suitable for the consistency representation learning.
- For face forgery detection, learning inconsistency-aware model can be independent of forgery location annotation.
- GFR (General Forgery Region) is defined as the general manipulated region among different forgery datasets. The assignment of manipulated region might be redundant. We will further explore how to realize unsupervised approximate forgery location without the specified GFR.