SICHUAN UNIVERSITY 四川大學

香港中文大學(深圳) The Chinese University of Hong Kong, Shenzhen

AIRS SHENZHEN INSTITUTE of ARTIFICIAL INTELLIGENCE AND ROBOTICS for SOCIETY 深圳市人工智能与机器人研究院

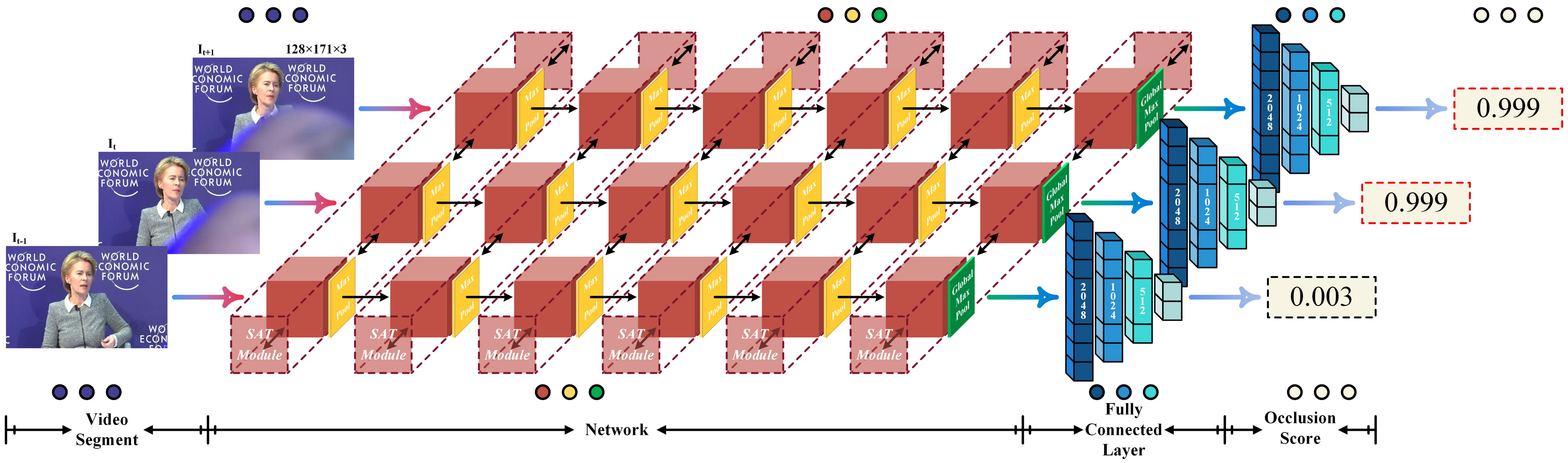
icassp 2022 Singapore, China, Online #1831

A Light Weight Model for Video Shot Occlusion Detection

Junhua Liao[1], Haihan Duan[2,3], Wanbin Zhao[1], Yanbing Yang[1,4], Liangyin Chen[1,4]

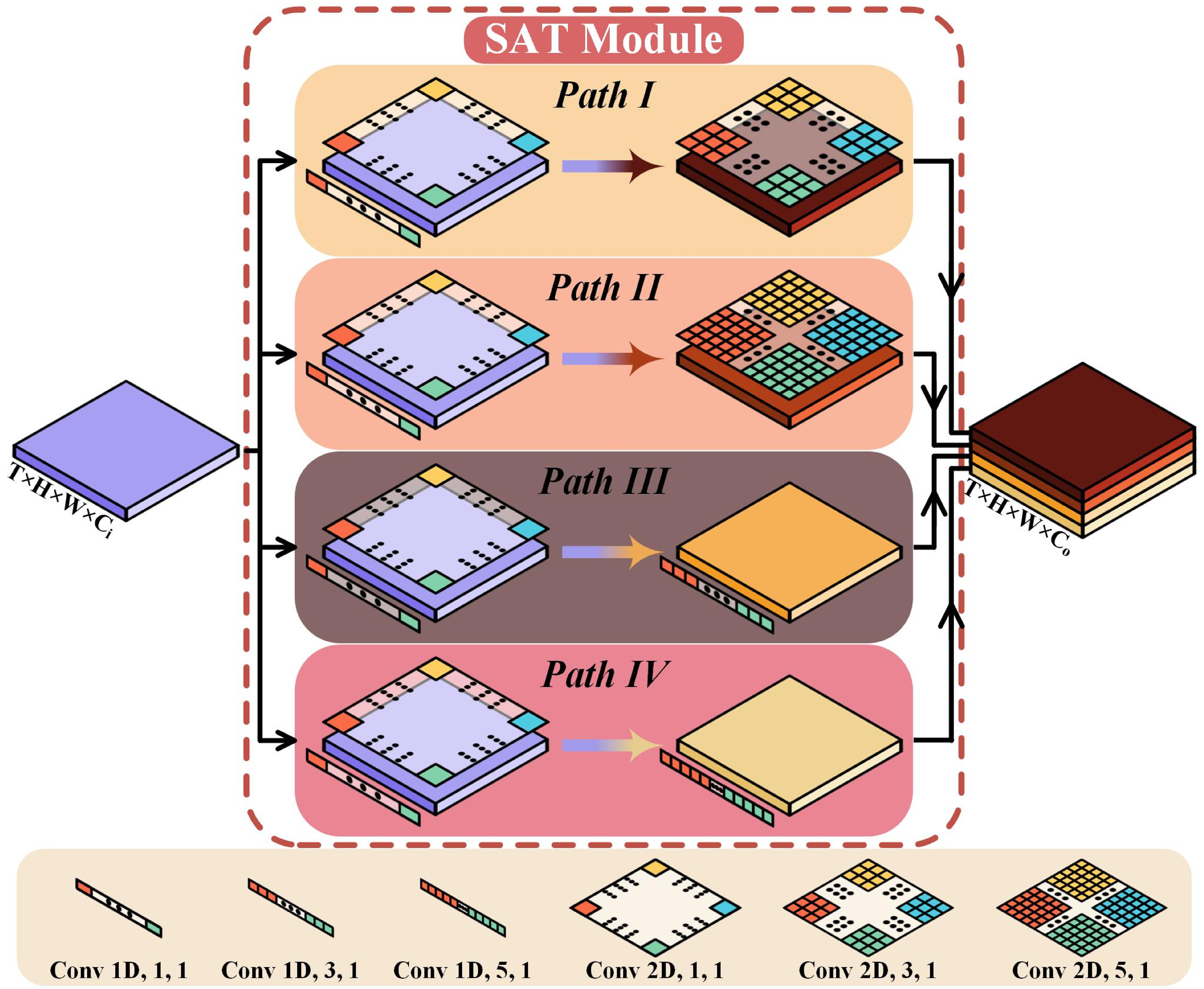
1. College of Computer Science, Sichuan University, Chengdu, China; 2. The Chinese University of Hong Kong, Shenzhen, China; 3. Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China; 4. The Institute for Industrial Internet Research, Sichuan University, Chengdu, China.

Architecture



The architecture of the proposed shot occlusion detection model

SAT Module



Structure of the SAT module

Loss Function

Firstly, we calculate the percentage $P_{occlusion}$.

$$P_{occlusion} = \frac{A_{occlusion}}{A_{frame}}$$

Where $A_{occlusion}$ is the area of occlusion and A_{frame} is the frame area.

Secondly, we calculate the maximum $P_{occlusion}$ in the video.

$$M_{occlusion} = max(P_{occlusion} \in V_i)$$

Where V_i represents the entire sequence of the i_{th} video.

Thirdly, we calculate the occlusion ratio $R_{occlusion}$.

$$R_{occlusion} = \frac{2 - (P_{occlusion} + \frac{P_{occlusion}}{M_{occlusion}})}{2\beta}$$

Where β is set to 10 as an equilibrium coefficient empirically.

Finally, we calculate and assign the weights.

$$L_{occlusion} = -e^{-R_{occlusion}}(t_j^i \log(p_j^i) + (1 - t_j^i)\log(1 - p_j^i))$$

Where t and p represent the tag and prediction results, respectively.

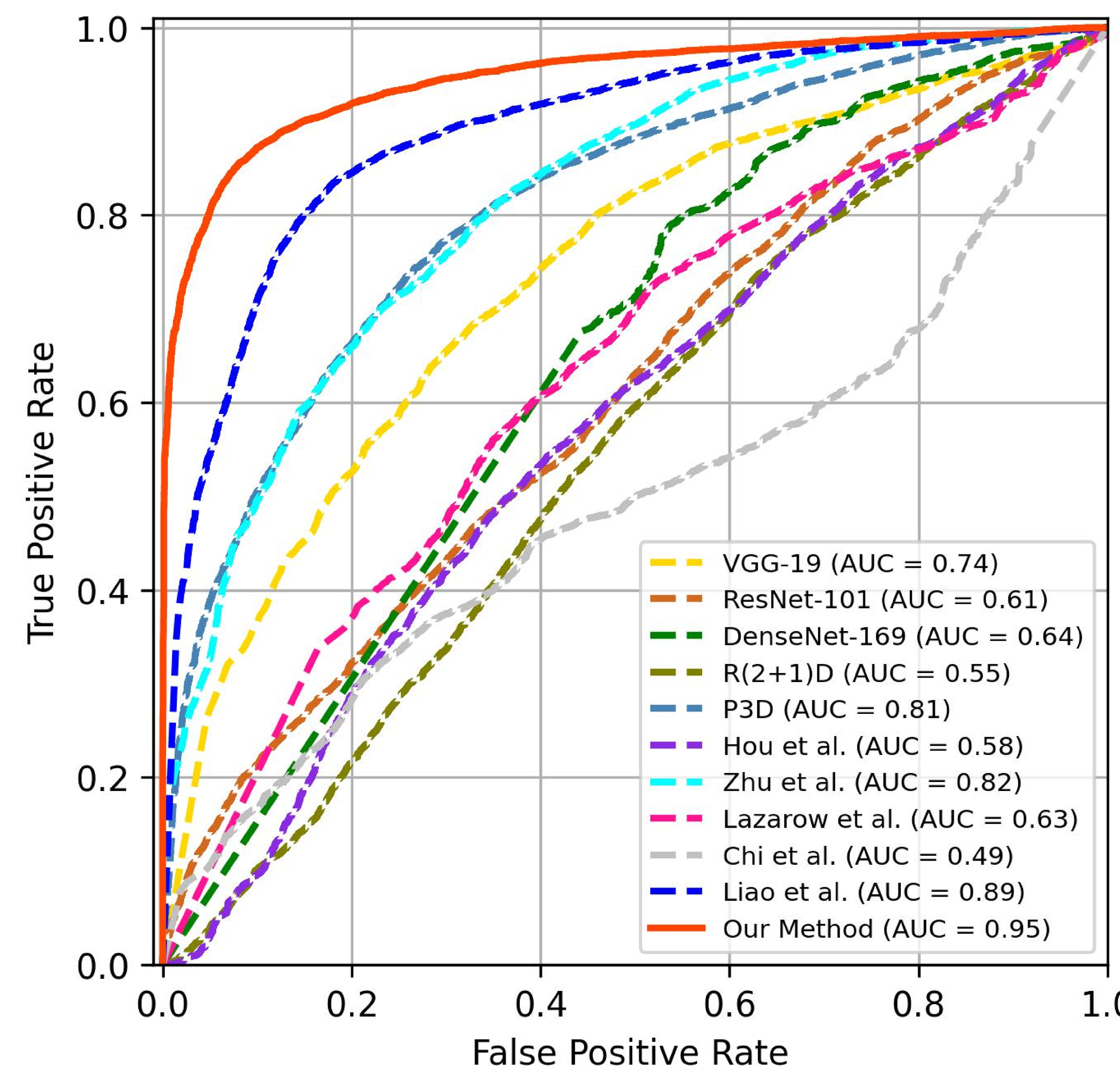
Contributions

- ★ We design a SAT module to extract spatio-temporal information instead of 3D convolution, and construct a new high-performance video shot occlusion detection framework based on this module.
- ★ We improve the existing occlusion detection loss function to more reasonably assign weights to occlusion frames, which significantly increases the accuracy of recognition.
- ★ The extensive experiments show that our shot occlusion detection method outperforms the state-of-the-art methods.

Result

Performance comparison with the state-of-the-art methods

Method	Parameters	Accuracy	FPS
VGG-19	139.59M	68.85%	70
ResNet-101	42.50M	61.06%	83
DenseNet-169	12.49M	65.56%	95
R(2+1)D	33.18M	59.10%	99
P3D	24.93M	74.09%	120
Hou et al.	23.51M	42.66%	60
Zhu et al.	15.76M	73.17%	61
Lazarow et al.	64.66M	62.26%	32
Chi et al.	40.78M	50.43%	33
Liao et al.	59.64M	82.70%	106
Our Method	**11.37M**	87.03%	**130**
Our Method+$L_{occlusion}$	**11.37M**	**88.25%**	**130**



ROC curves

Code and models are available at https://github.com/Junhua-Liao/ICASSP22-OcclusionDetection