

Human Action Recognition without Human

Yun He^{1,2}, Soma Shirakabe^{1,2}, Yutaka Satoh^{1,2}, Hirokatsu Kataoka¹

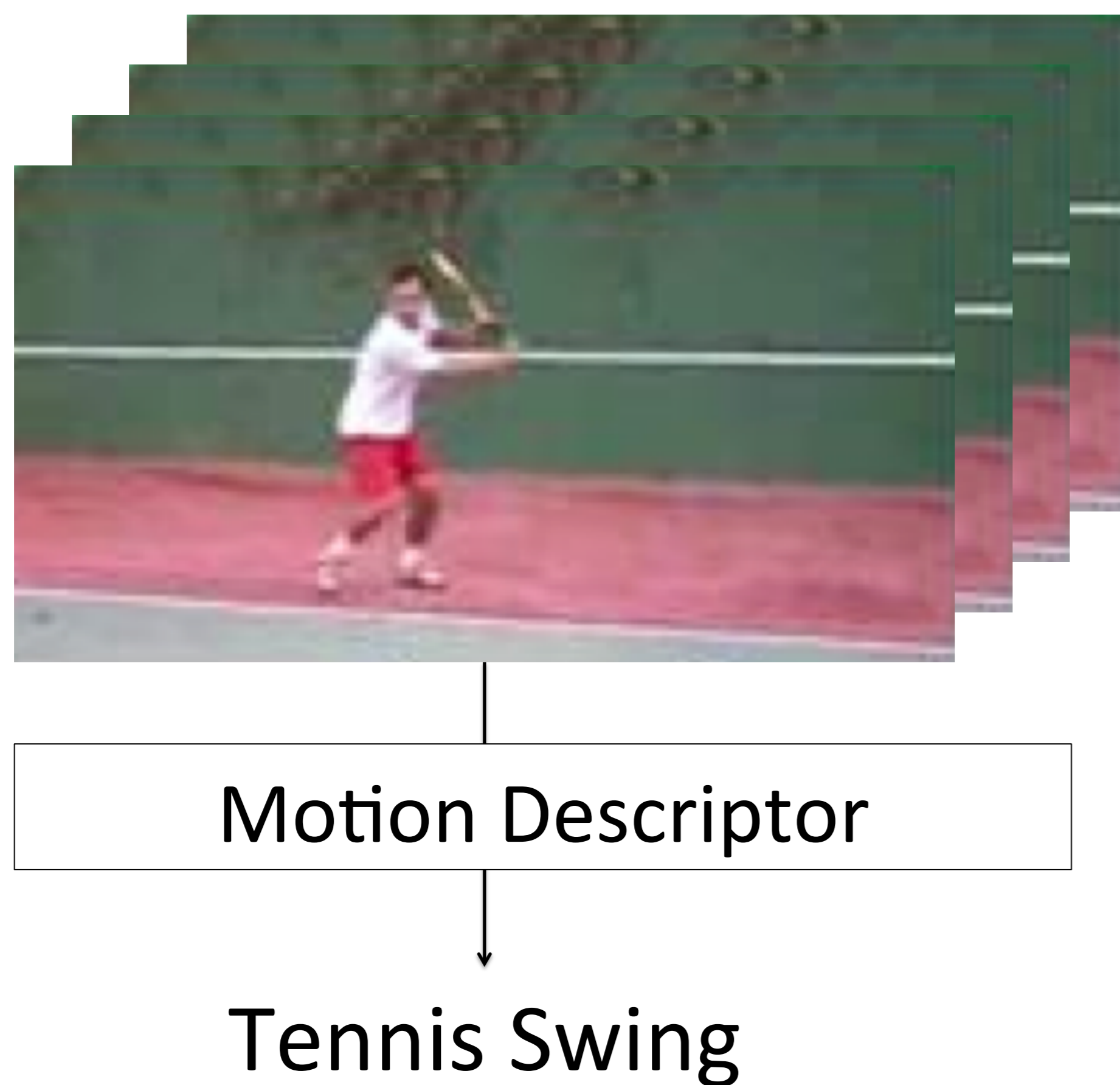
¹Computer Vision Research Group, AIST, Japan

²Human-Centered Vision Lab., University of Tsukuba, Japan

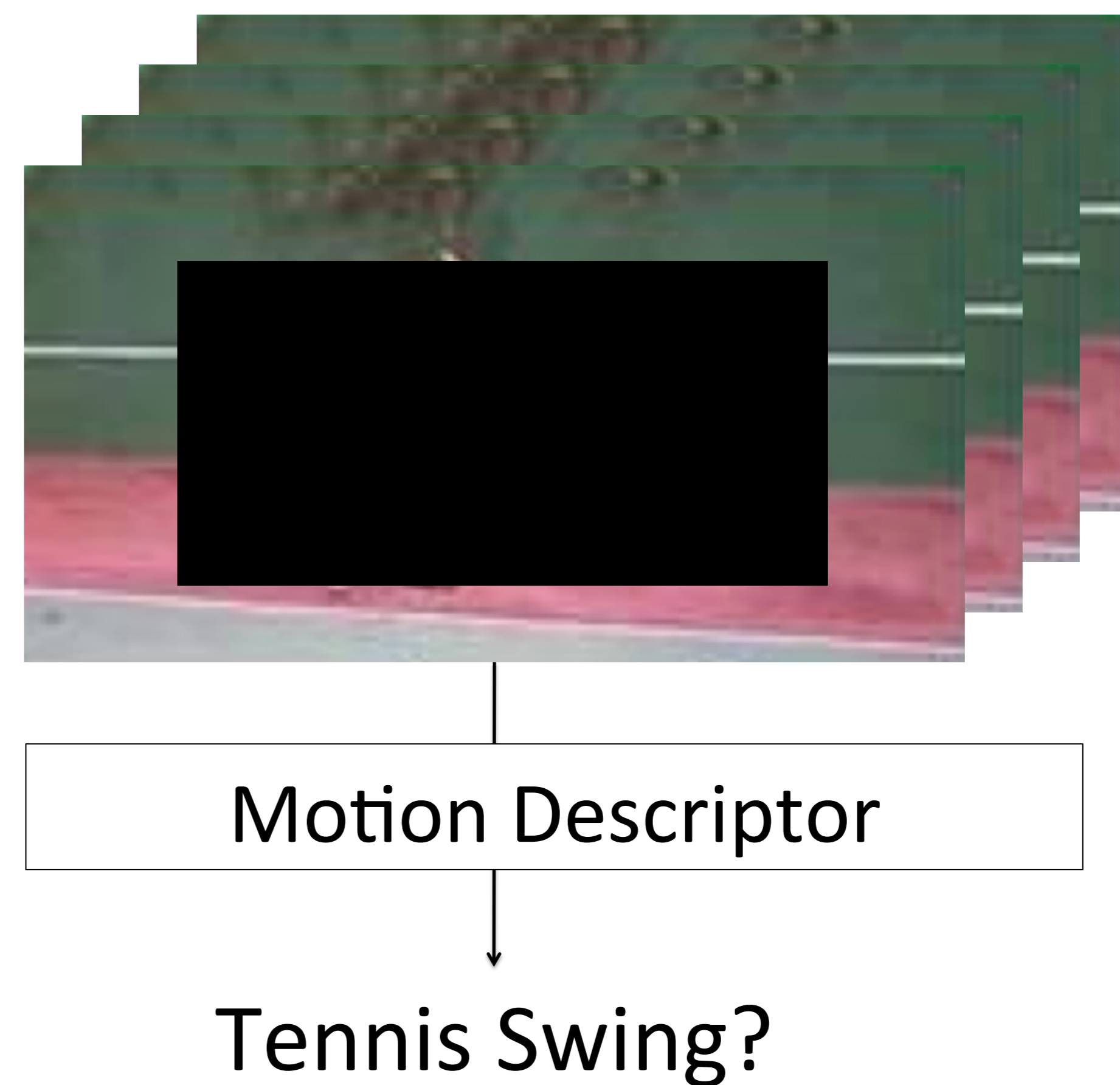
Do we need a human?

- ❑ Some features from the background could be too strong!
 - 84.5% (TSN^[1]) with RGB sequence on the UCF101
- ❑ Human action recognition can be done just by analyzing motion the the background?

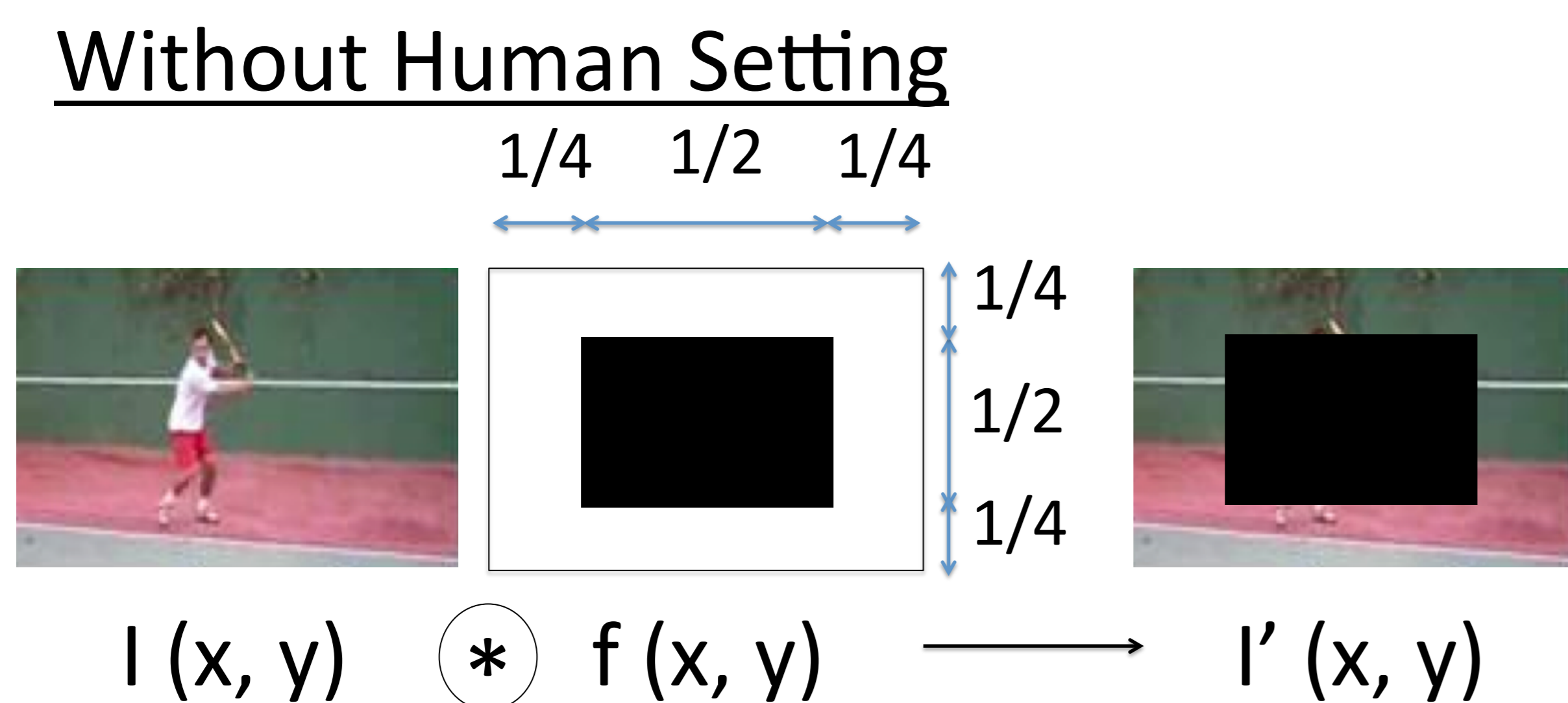
Human Action Recognition



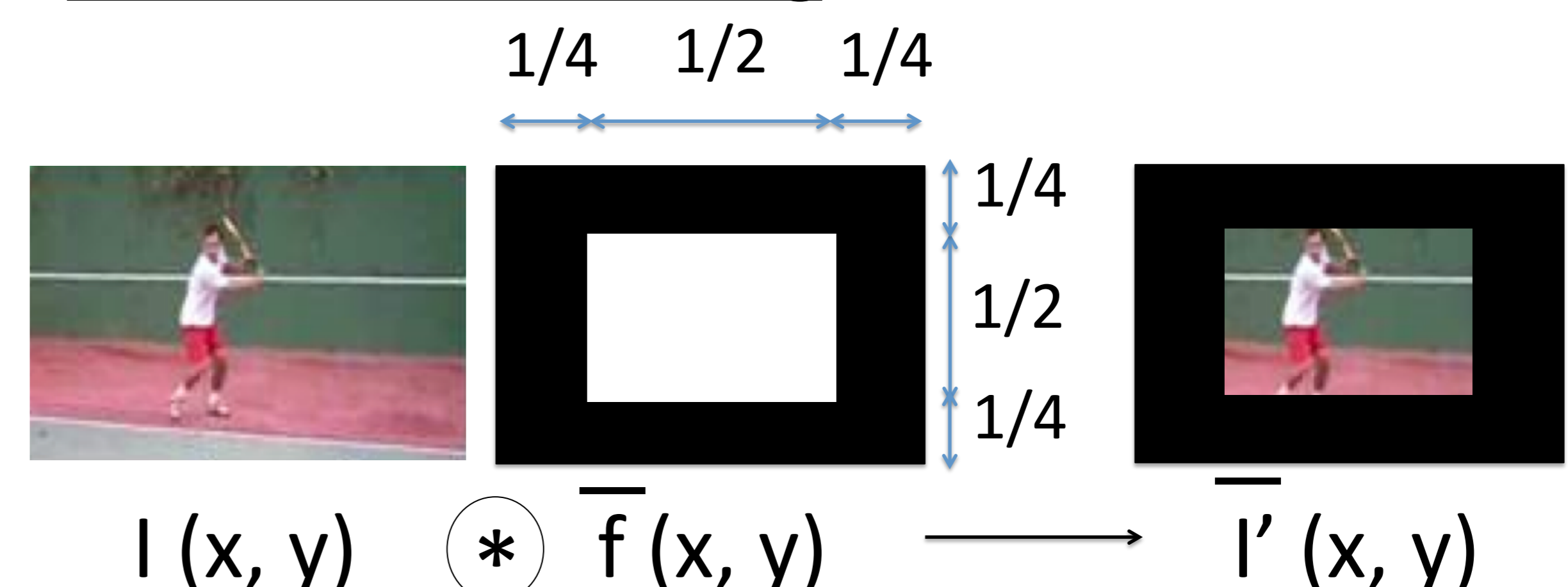
Human Action Recognition *without* Human



Without Human Setting



With Human Setting



Comparison

- ❑ Baseline: Very deep two-stream CNN [2]
 - 16-layer, UCF101 pre-trained model
- ❑ The baseline performs 84.30% on the original UCF101

W/ or w/o a human	Stream	% on UCF101 split 1
With Human	RGB	51.26
	Flow	40.50
	Two-stream	56.91
Without Human	RGB	45.33
	Flow	26.80
	Two-stream	47.42

❑ w/o human is close to w/ human (-9.49%)

Conclusion

- ❑ The first study of “human action recognition without human”
 - Motion representation from a bg sequence is too strong
- ❑ Human-centric and CNN-based motion representation are urgent!
 - We are preparing for motion only database

Reference

- [1] L. Wang, et al. Temporal segment networks: Towards good practices for deep action recognition. in ECCV, 2016.
- [2] L. Wang et al. Towards Good Practices for Very Deep Two-Stream ConvNets. In arXiv, 2015.