



Unsupervised Co-part Segmentation through Assembly

Qingzhe Gao^{1,2}

Bin Wang²

Libin Liu³

Baoquan Chen³

¹Shandong University

²AICFVE, Beijing Film Academy

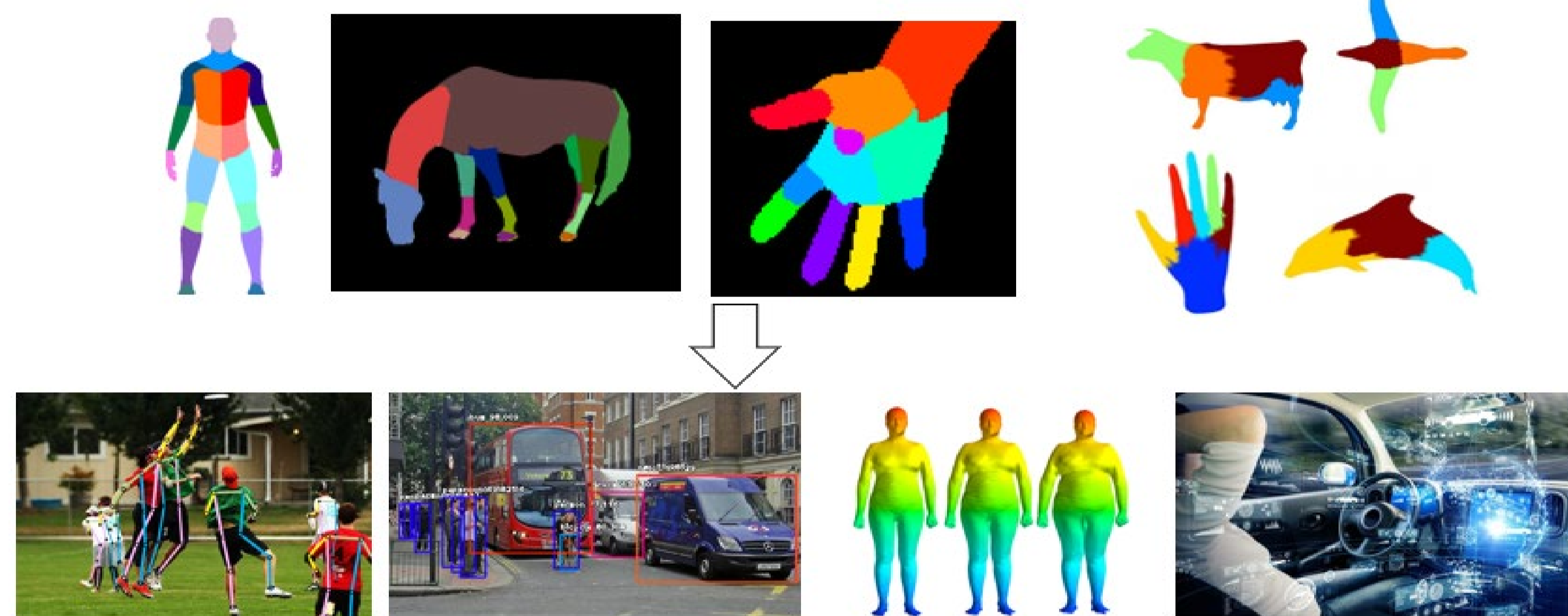
³CFCS, Peking University



ICML
International Conference
On Machine Learning

Part structure

- Robust representations to variations.
- Serving for many higher-level tasks.



Pose estimation Object detection 3d reconstruction Autopilot

Challenge

- Supervised: part annotations are expensive.
- Unsupervised: hard to achieve meaningful and consistent segmentation

Intuition: videos provide structural and motion-related information

- Each frame of a video can be segmented into several parts.

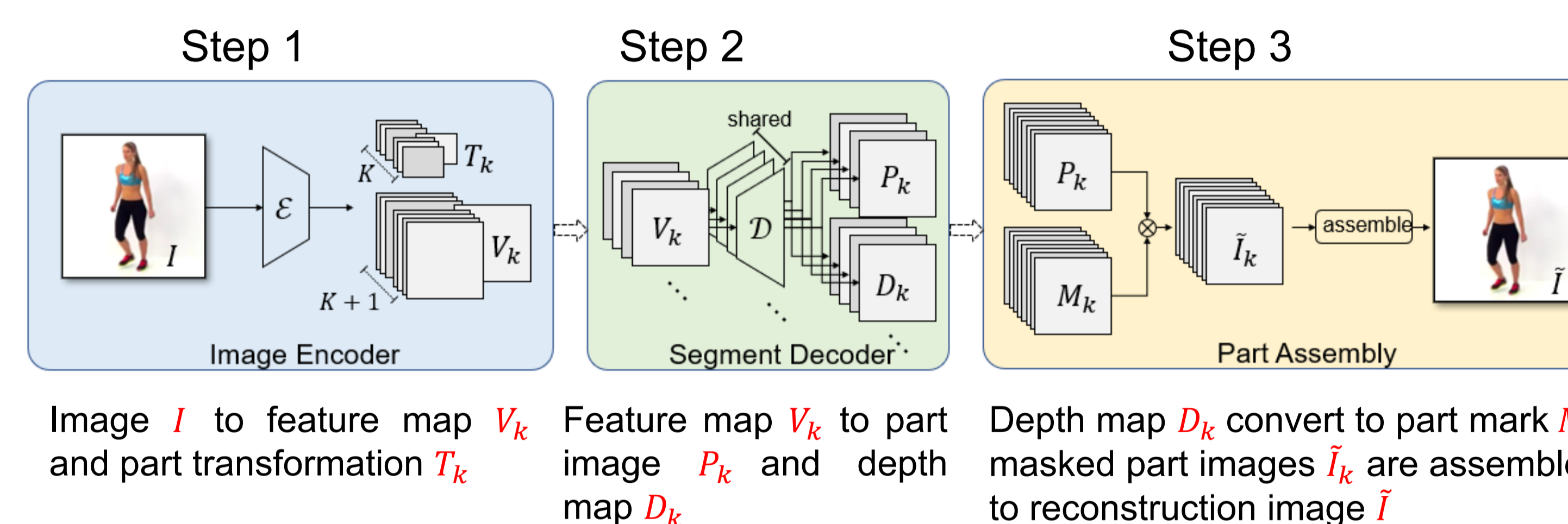


- These parts can be transformed using affine transformations.
- Then, the transformed parts can be assembled to reconstruct another frame.

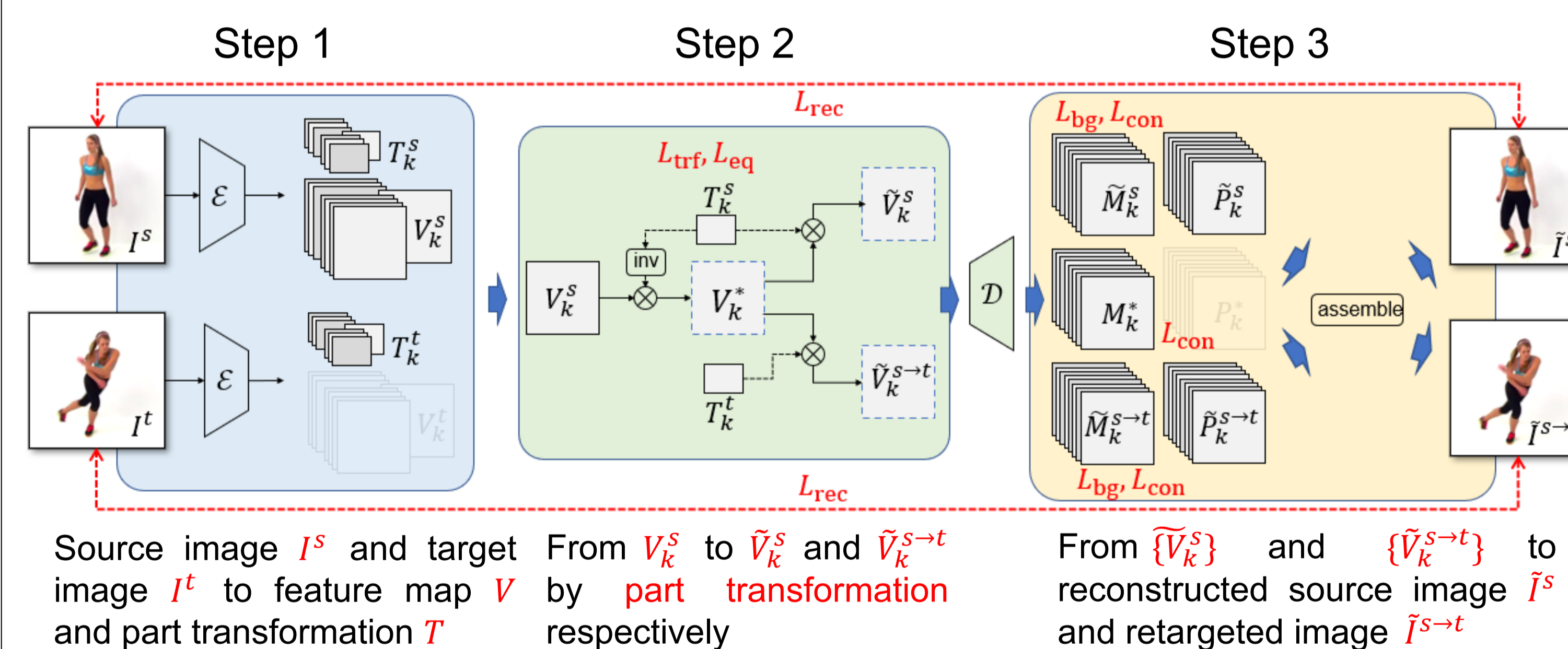
Contribution

- Introducing part-assembly as a dual-procedure of part-segmentation.
- Parts and their transformations have explainable physical meaning.
- Not require any field-based global warping operation.

Framework



End-to-end Training



Result

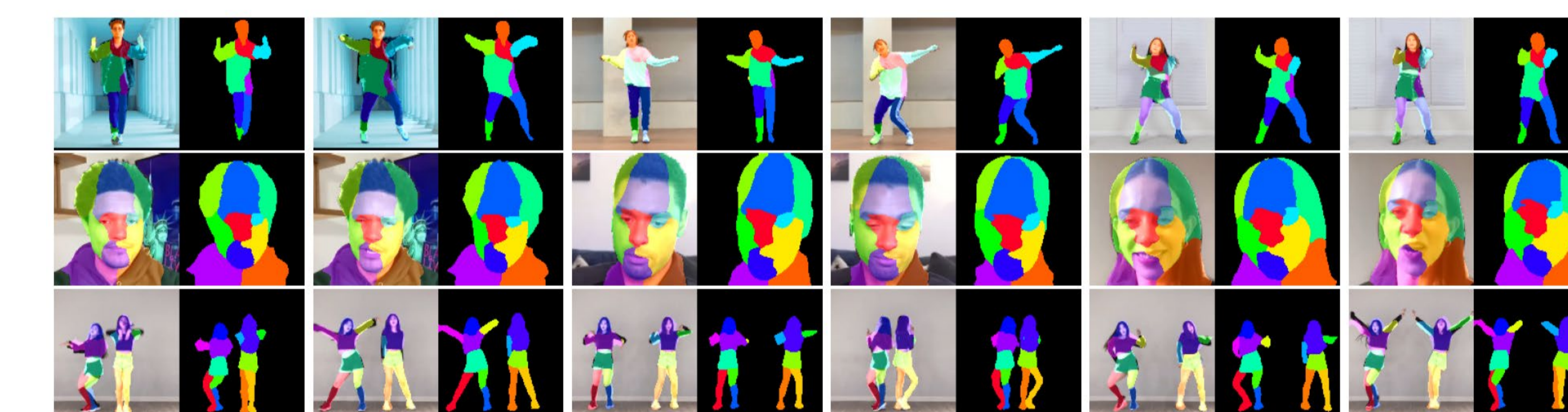
- Visualization of the segmentation results.



Visual results tested on different scenarios.



- Visual results tested on unseen wide videos (human, face) and trained on a two characters video.



Comparison with SCOPS and Motion Co-part.



Quantitative evaluation.

Dataset	Metrics	SCOPS	MCPS	Ours
Tai-Chi-HD	Landmarks	411.38	389.78	326.82
	IoU	0.5485	0.7686	0.8724
VoxCeleb	Landmarks	663.04	424.96	338.98
	IoU	0.5045	0.9135	0.9270