JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Code, Model Zoo, and Demos

Bootstrapping Objectness from Videos by Relaxed Common Fate and Visual Grouping

Long Lian

UC Berkeley

Zhirong Wu

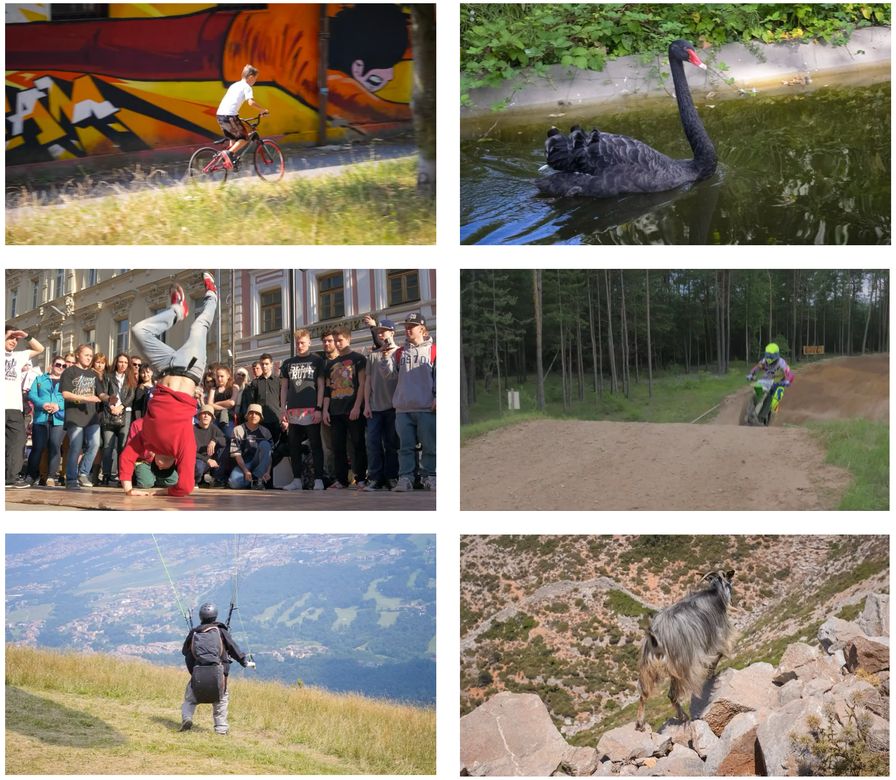
Microsoft Research Asia

Stella X. Yu

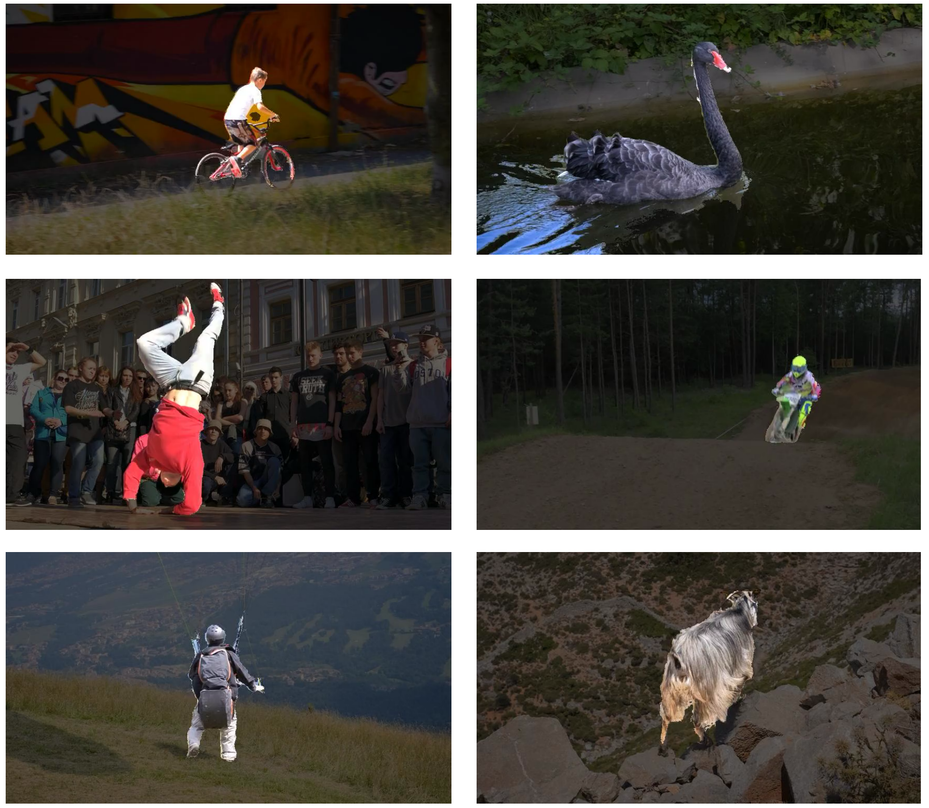
University of Michigan

Paper Tag: WED-PM-213. **Poster Number:** 213.

Our Task: Unlabeled Videos ⇨ Object Segmentation



Given optical flow detector



(Our RCF Segmentation)

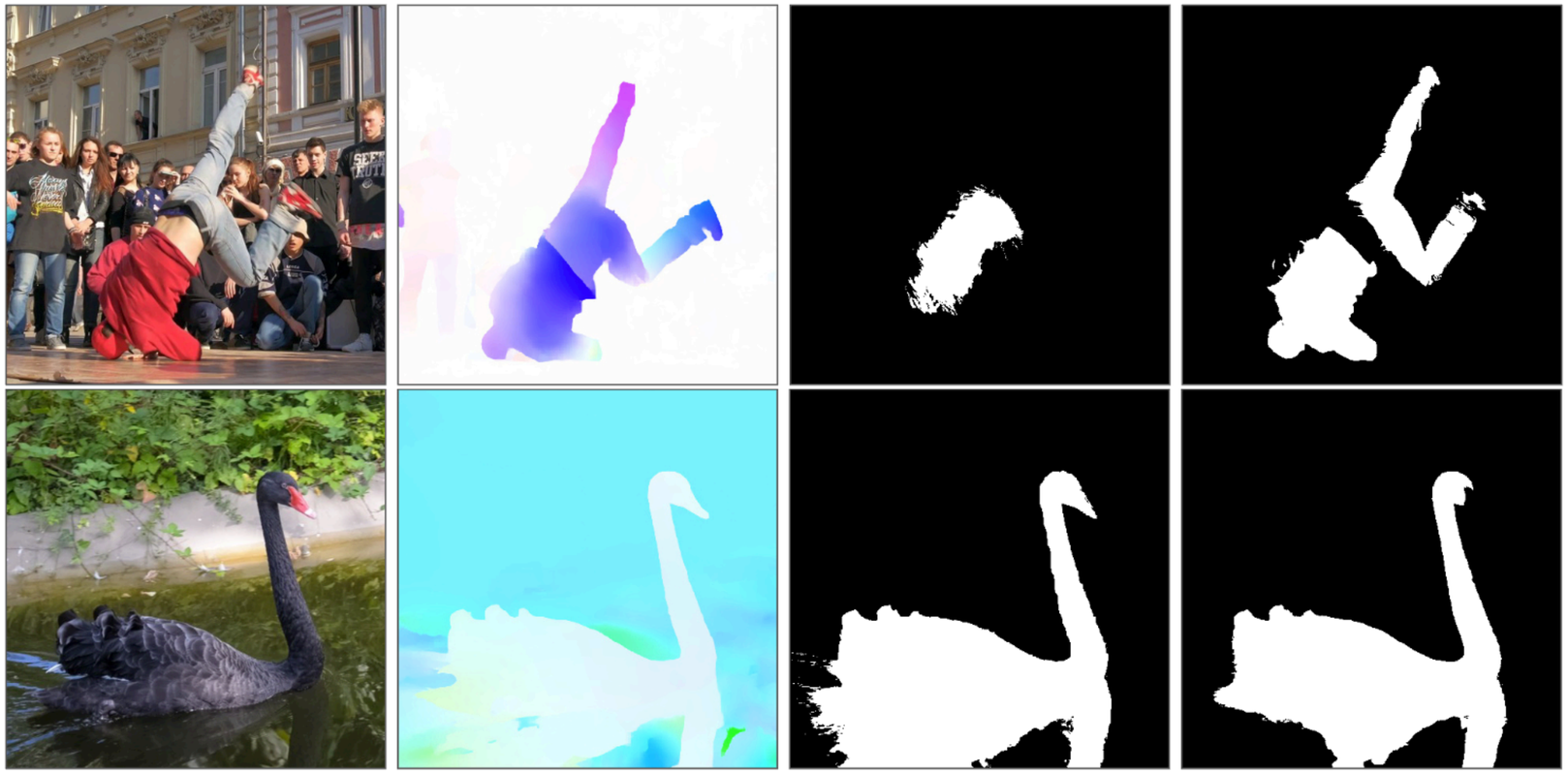
Existing Methods Rely on Common Fate

Motion Grouping. ICCV. 2021.
OCLR. NeurIPS 2022.
GWM. BMVC 2022.
AMD. NeurIPS 2021.

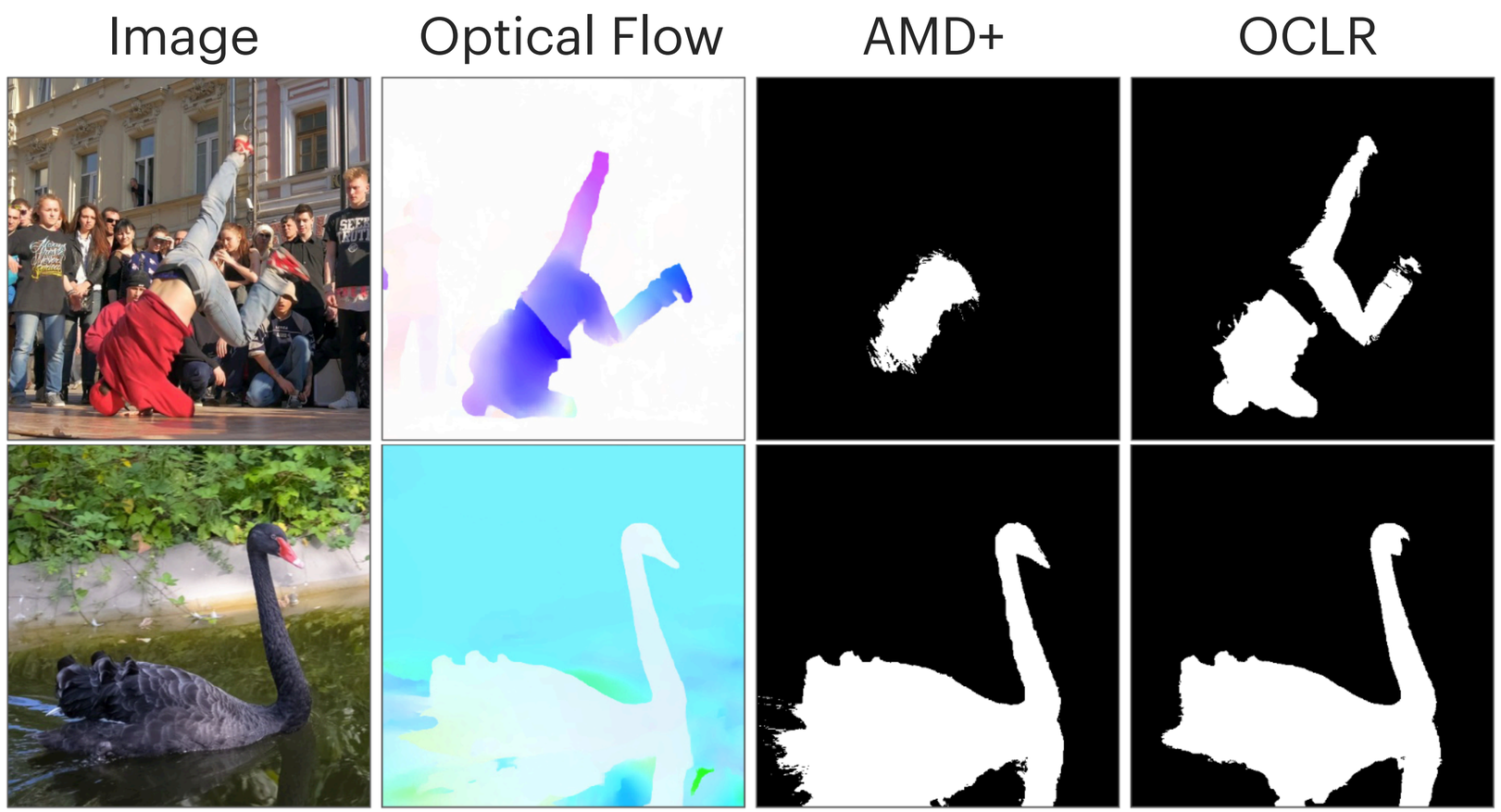
Objectness = What Move Together Belong Together

Motion Grouping. ICCV. 2021.
OCLR. NeurIPS 2022.
GWM. BMVC 2022.
AMD. NeurIPS 2021.

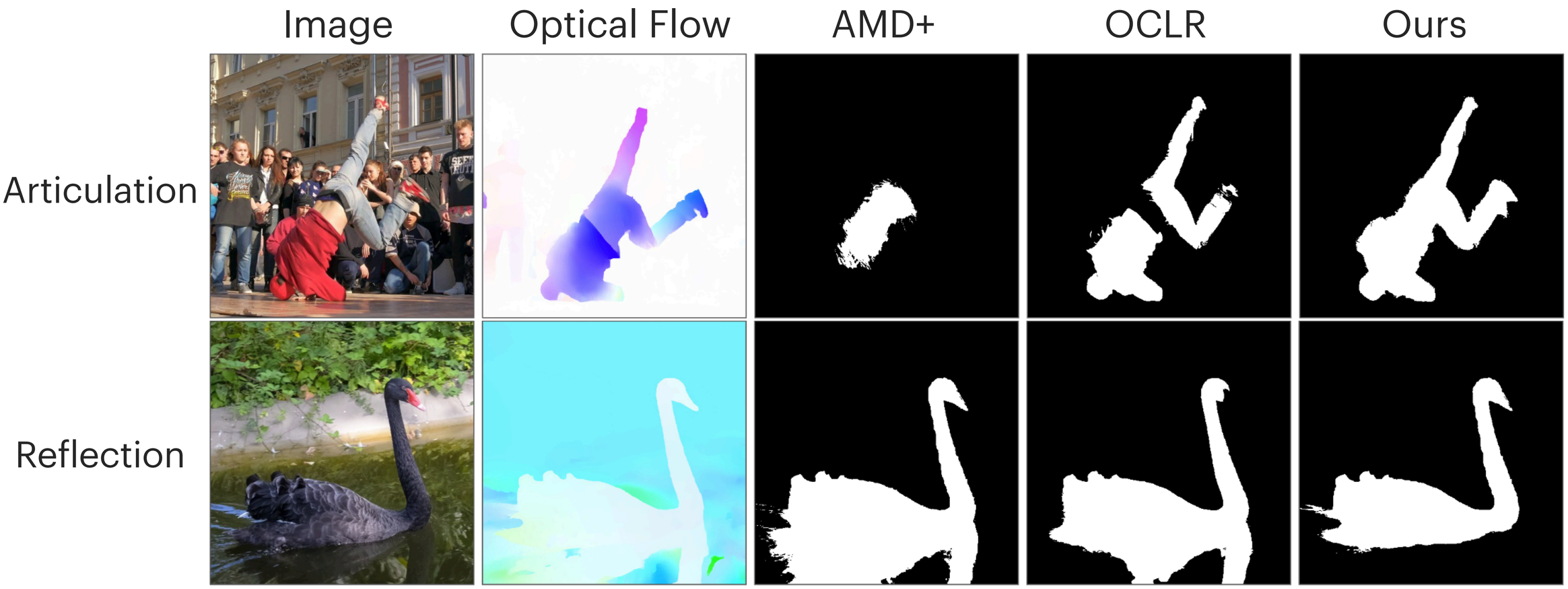
Two Failure Modes From Gestalt Law of Common Fate



Two Failure Modes From Gestalt Law of Common Fate



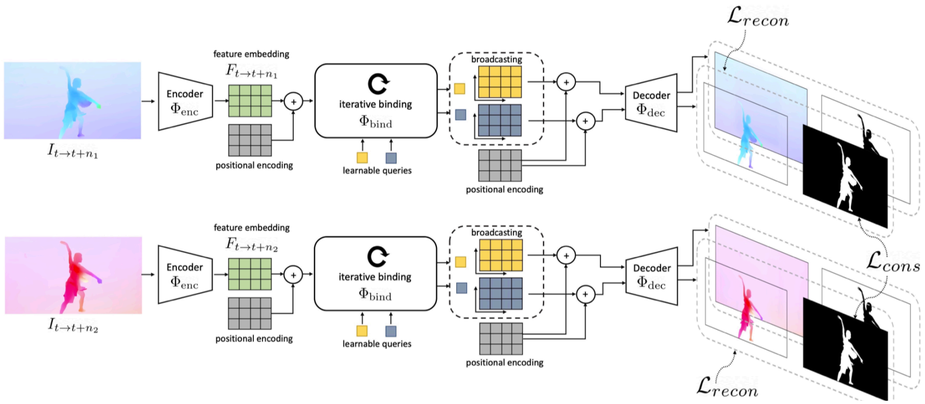
Our Approach Addresses Both Caveats



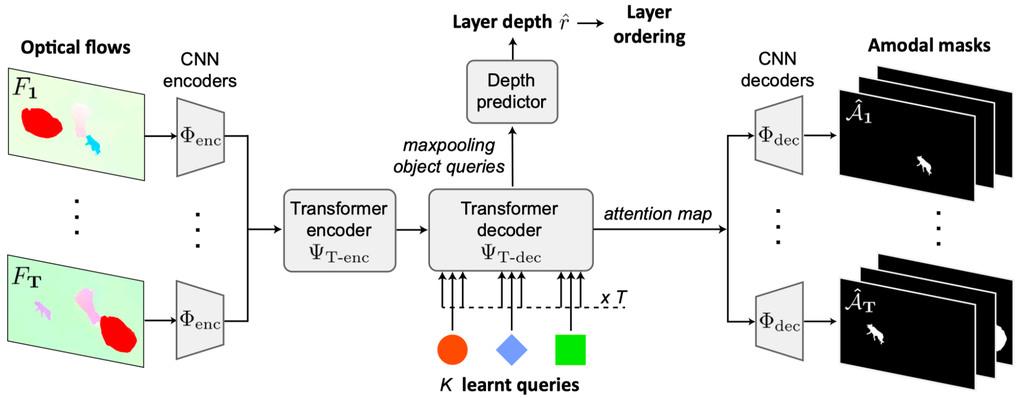
Existing Methods: Three Camps

1. Motion Segmentation

Motion Grouping



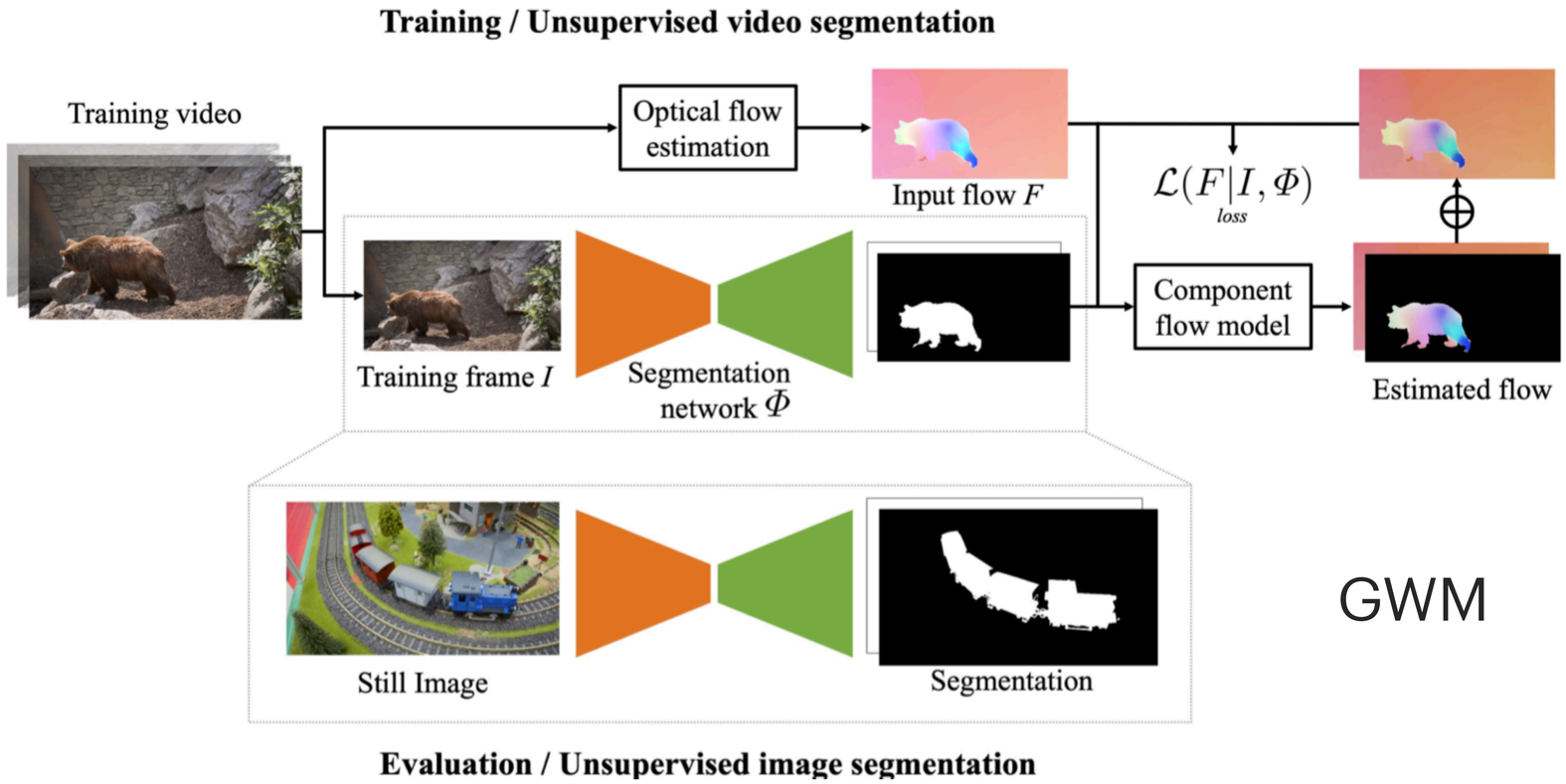
OCLR



Motion Grouping. ICCV. 2021.
OCLR. NeurIPS 2022.

Existing Methods: Three Camps

1. Motion Segmentation
2. Motion Guided Segmentation

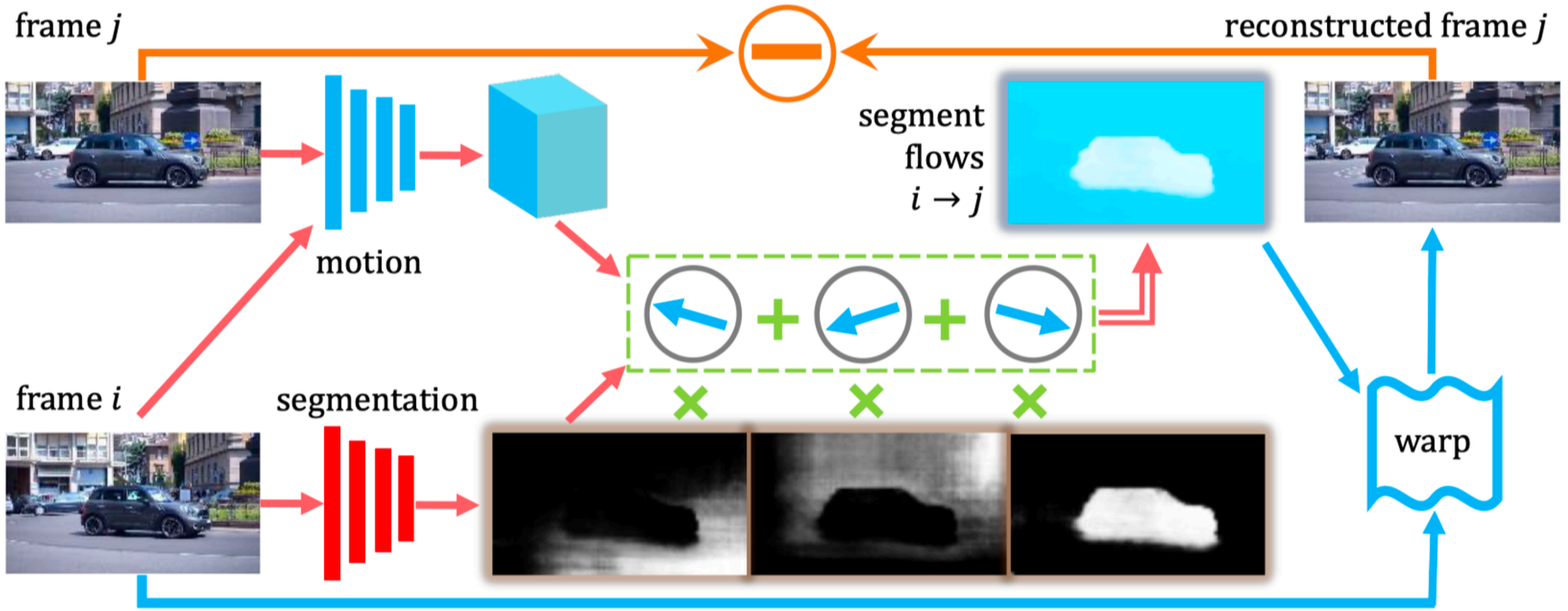


GWM

GWM. arXiv:2205.07844 (2022).

Existing Methods: Three Camps

1. Motion Segmentation
2. Motion Guided Segmentation
3. Motion and Segmentation Jointly Learned



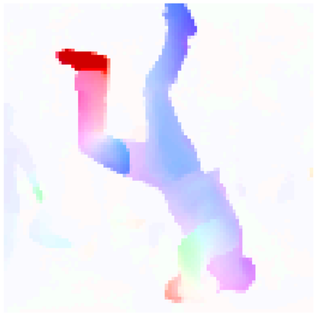
AMD

AMD. NeurIPS 2021.

Insight 1: Dealing with Articulation

Previous works:

Common Fate

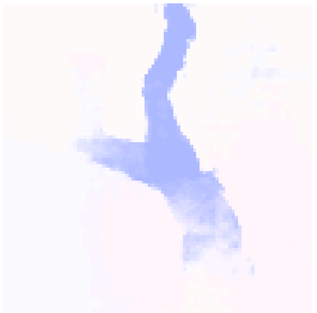


Free-form Flow

×

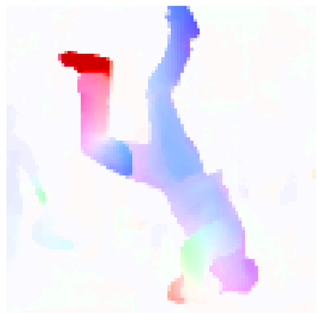
Pred. Mask
(Being optimized)

⇒



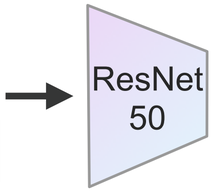
Constant/Affine Flow
w.r.t. Mask

≈



Free-form Flow

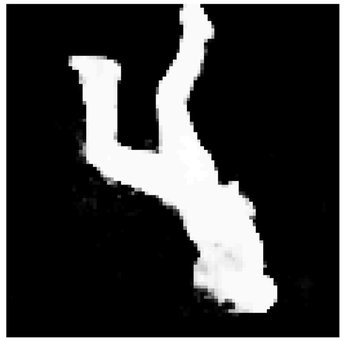
Video Frame



Insight 1: Fitting Flow with *Relaxed* Common Fate

Our work RCF:

Free-form Flow



Pred. Mask
(Being optimized)

Relaxed
Common Fate

Constant/Affine Flow
w.r.t. Mask

+

Residual Flow
(Intra-mask motion)

≈

Free-form Flow

Video Frame

ResNet
50

Insight 2: Visual Grouping within the Image

Let motion and appearance complement each other for supervision.

Motion Supervision Only

Bootstrapping from CRF

Motion + Appearance

Insight 2: Visual Grouping Based on Semantics across Images

Let motion and appearance complement each other for supervision.

Motion Supervision Only

Iteratively minimize the normalized cuts of DINO feature

Motion + Appearance

Extra Benefits: Label-free Hyperparameter Tuning

Using the normalized cuts of DINO feature as an unsupervised segmentation quality indicator for hyperparameter tuning.

High Normalized Cuts

(Mask boundary spans coherent features)

Low Normalized Cuts

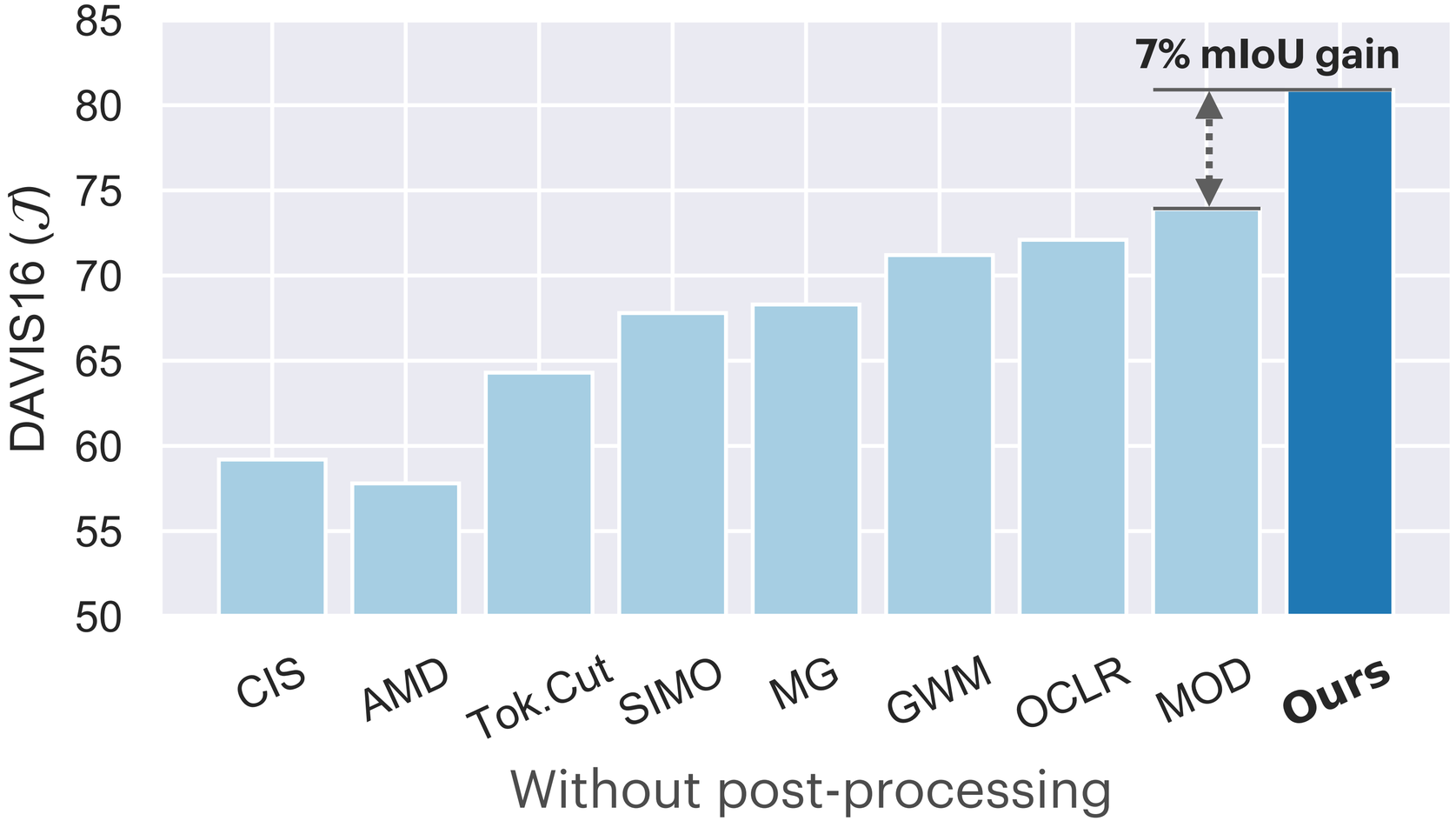
(Mask boundary does not span coherent features)

✓ Use this setting

Advantages of Our RCF to Previous Methods

UVOS Method	Motion Grouping	Emergence of Objectness	Guess What Moves	**Our work (RCF)**
Sources of supervision	Motion	Motion (Frame Warping)	Motion	Motion + Appearance
Segment stationary objects	✗	✓	✓	✓
Handle articulated/ deformable objects	–	✗	✗	✓
Label-free hyper-parameter tuning?	✗	✗	✗	✓

Our RCF: SOTA on Unsupervised Object Segmentation



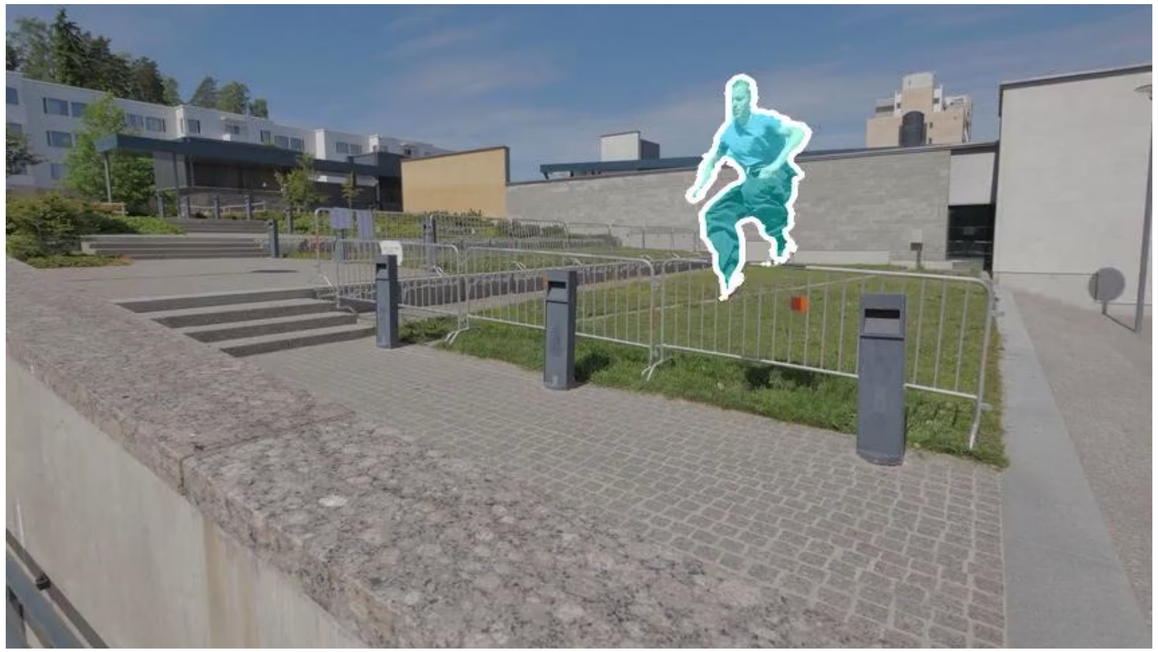
6.3% gain with post-processing
9.1% on STv2
4.5% on FBMS59

OCLR

RCF (***Ours***)

No post-processing applied: results can be further enhanced with post-processing

RCF (***Ours***)



No post-processing applied: results can be further enhanced with post-processing

Thank you!

Code, Model Zoo,
and Demos Available