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
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
Input: Unlabeled Videos
Output: Objectness Segmentation
Our Task: Unlabeled Videos $\implies$ Object Segmentation

Given optical flow detector

(Our RCF Segmentation)
Existing Methods Rely on Common Fate
Objectness = What Move Together Belong Together
Two Failure Modes From Gestalt Law of Common Fate

- Articulation
- Reflection

- Image
- Optical Flow
- AMD+
- OCLR

- Fragmented Objectness
- Excessive Objectness
Two Failure Modes From Gestalt Law of Common Fate

Articulation

Reflection

Our Solution

Relaxed Common Fate

Visual Grouping
Our Approach Addresses Both Caveats

Articulation

Reflexion

Image | Optical Flow | AMD+ | OCLR | Ours

[Images showing the results of articulation and reflection for different methods]
Existing Methods: Three Camps

1. Motion Segmentation

Motion Grouping

OCLR
Existing Methods: Three Camps

1. Motion Segmentation
2. Motion Guided Segmentation

Motion Grouping: ICCV. 2021.
OCLR. NeurIPS 2022.
AMD. NeurIPS 2021.
Existing Methods: Three Camps

1. Motion Segmentation
2. Motion Guided Segmentation
3. Motion and Segmentation Jointly Learned
Insight 1: Dealing with Articulation

Free-form Flow \times \text{Pred. Mask (Being optimized)} \Rightarrow \text{Constant/Affine Flow w.r.t. Mask} \approx \text{Free-form Flow}
Insight 1: Dealing with Articulation

Free-form Flow \times \text{Pred. Mask (Being optimized)} \Rightarrow \text{Constant/Affine Flow w.r.t. Mask} \approx \text{Free-form Flow}
Insight 1: Dealing with Articulation

Free-form Flow × Pred. Mask (Being optimized) ⇒ Constant/Affine Flow w.r.t. Mask ≈ Free-form Flow

Common Fate
Insight 1: Fitting Flow with *Relaxed* Common Fate

Free-form Flow × Pred. Mask (Being optimized) ⇒ Constant/Affine Flow w.r.t. Mask + Relaxed Common Fate

Pred. Mask

ResNet 50

Video Frame

Residual Flow (Intra-mask motion)

≈

Free-form Flow
Insight 2: Dealing with Artifacts from Motion

Let motion and appearance complement each other for supervision.

Motion Supervision Only
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: Dealing with Reflection

Let motion and appearance complement each other for supervision.

Motion Supervision Only
Insight 2: Visual Grouping Based on Semantics across Images

Let motion and appearance complement each other for supervision.

Motion Supervision Only

Motion + Appearance

Iteratively minimize the normalized cut of DINO feature
## Advantages of Our RCF to Previous Methods

<table>
<thead>
<tr>
<th>UVOS Method</th>
<th>Motion Grouping</th>
<th>Emergence of Objectness</th>
<th>Guess What Moves</th>
<th>Our work (RCF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources of supervision</td>
<td>Motion</td>
<td>Motion (Frame Warping)</td>
<td>Motion</td>
<td>Motion + Appearance</td>
</tr>
<tr>
<td>Segment stationary objects</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Handle articulated/deformable objects</td>
<td>—</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Label-free hyperparameter</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>
Our RCF: SOTA on Unsupervised Object Segmentation

![Bar chart showing 7% mIoU gain without post-processing]
No post-processing applied: results can be further enhanced with post-processing.
No post-processing applied: results can be further enhanced with post-processing
Thank you!

Code, Model Zoo, and Demos Available