# Crowd-Guided Ensembles | How Can We Choreograph Crowd Workers for Video Segmentation?

Alexandre Kaspar[1], Geneviève Patterson[2], Changil Kim[1],
Yagiz Aksoy[1,3], Wojciech Matusik[1] and Mohamed Elgharib[4]

[1]MIT CSAIL [2]MSR New England [3]ETH Zürich [4]HBKU QCRI

# Motivations

Movie post-production

(CC) Blender Foundation | Tears of Steel, see mango.blender.org

Video understanding

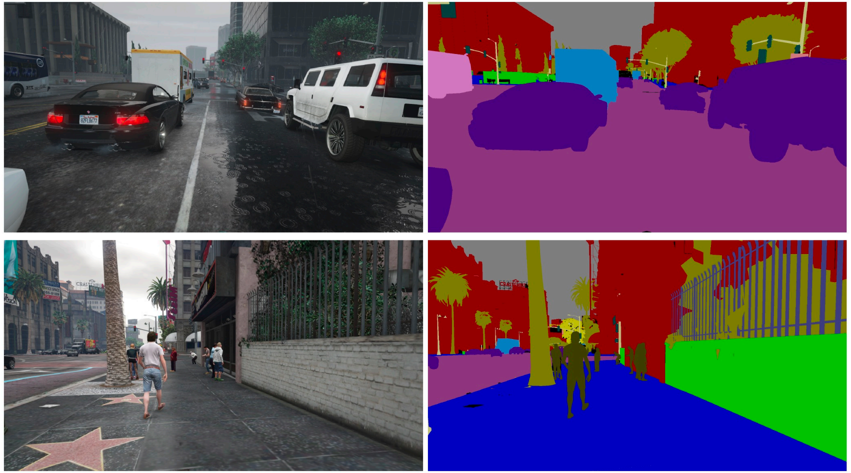
Playing for Data [Richter 2016]

# Motivations

Movie post-production



Video understanding

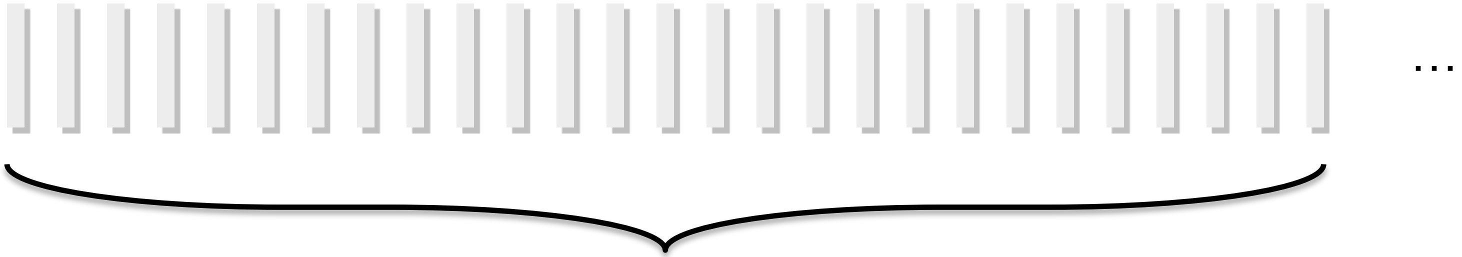


Playing for Data [Richter 2016]

Brief introduction to

# ROTOSCOPING

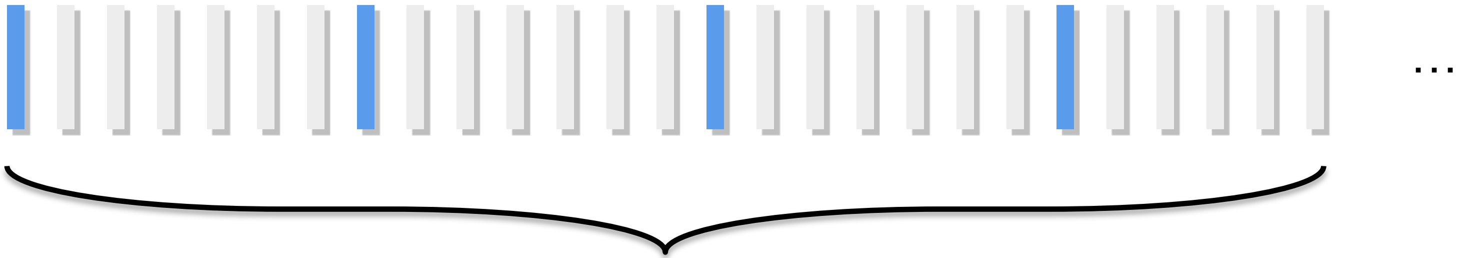
# Rotoscoping



Frames of video sequence

Artist

# Rotoscoping

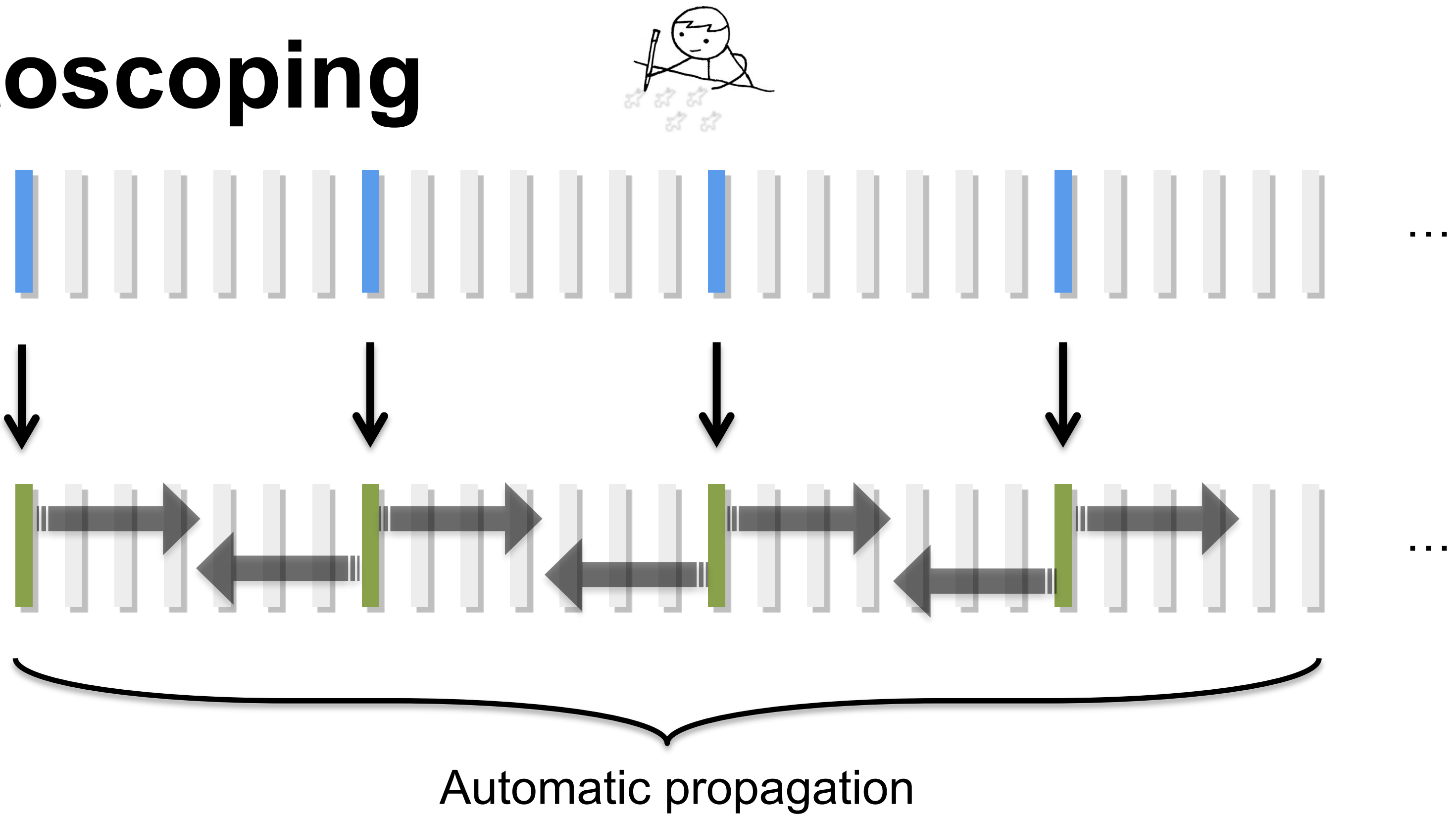


Frames of video sequence

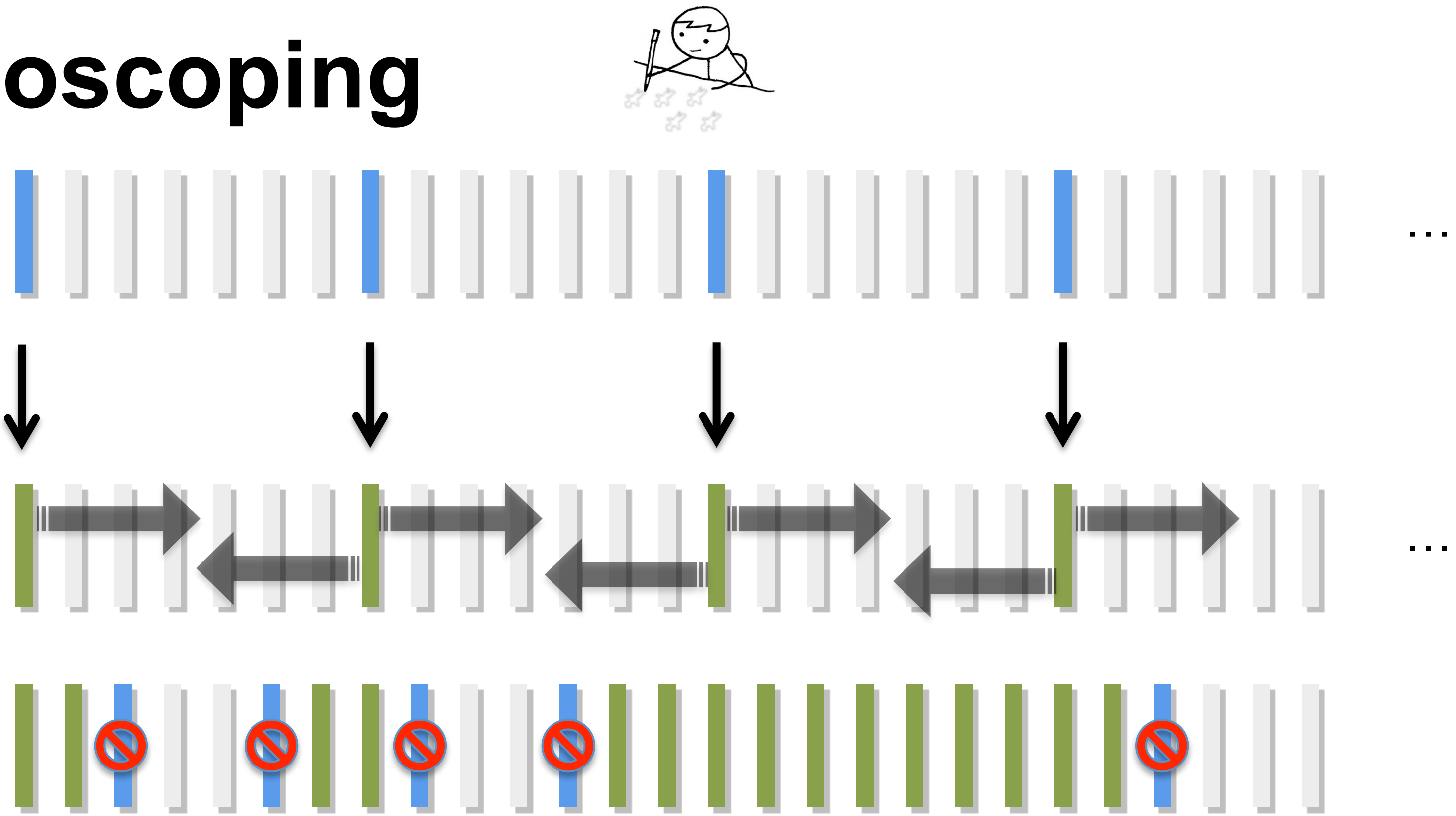
Keyframe: direct segmentation & manipulation

Remaining frame

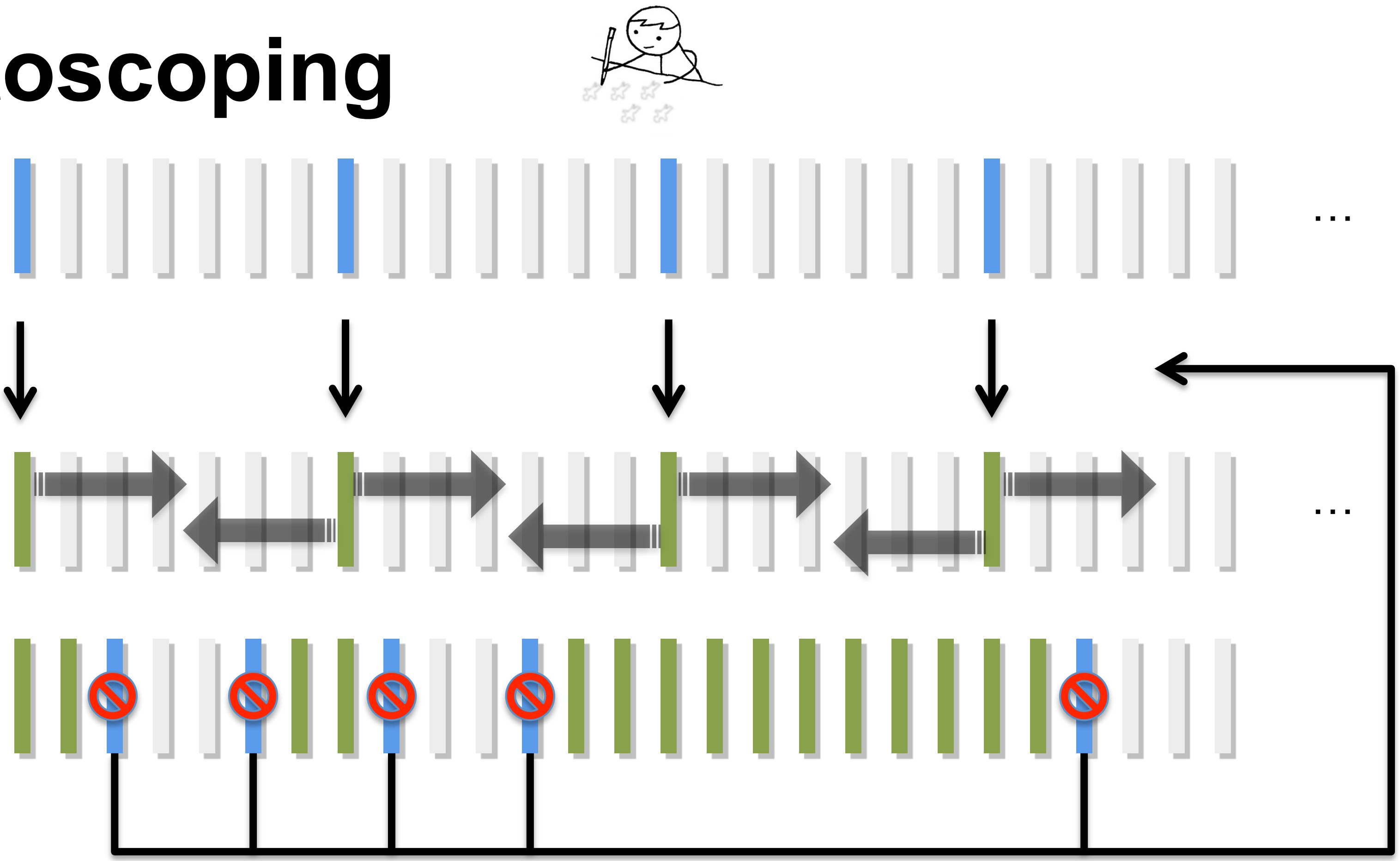
# Rotoscoping



# Rotoscoping



# Rotoscoping



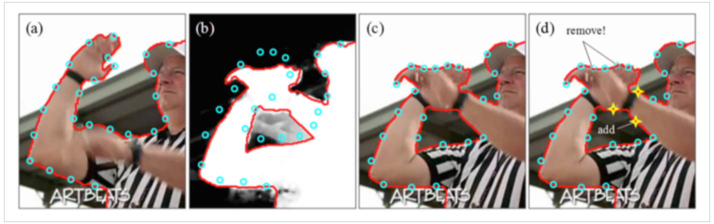
# Roto++ [Li 2016]

Professional Roto-Artist

Replay at x5

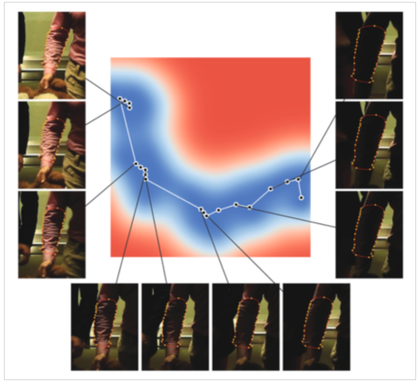
# RELATED WORK

# Related Work: Interactive Segmentation



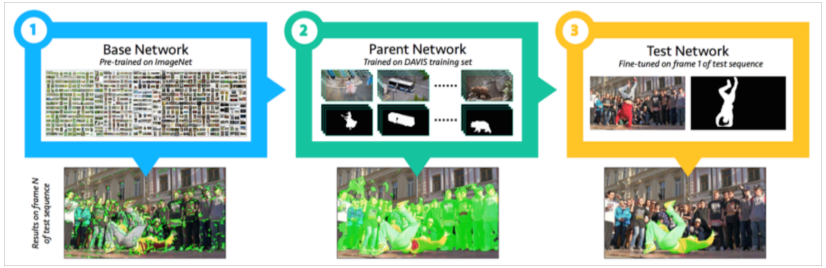

Video SnapCut: Robust Video Object Cutout Using Localized Classifiers [Bai 2009]


JumpCut: Non-Successive Mask Transfer and Interpolation for Video Cutout [Fan 2015]



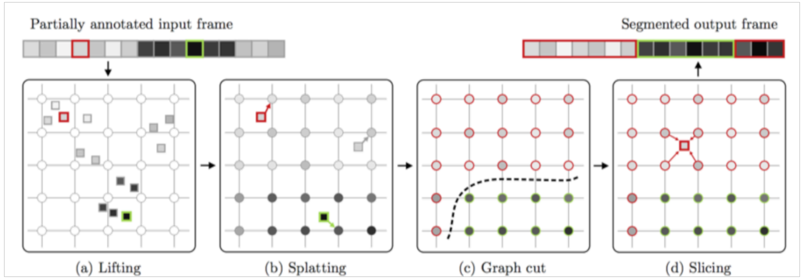
Roto++: Accelerating Professional Rotoscoping using Shape Manifolds [Li 2016]

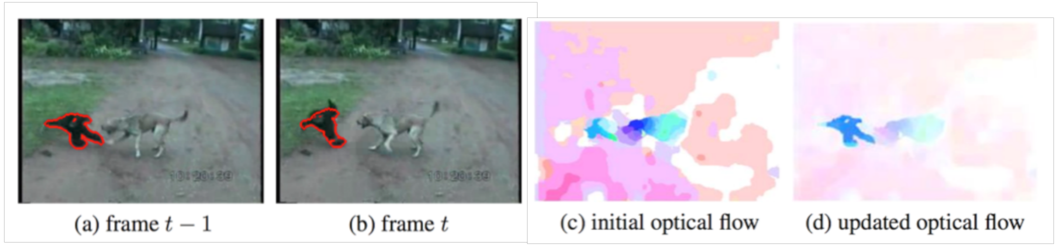
# Related Work: Computer Vision




One-Shot Video Object Segmentation [Caelles 2017]

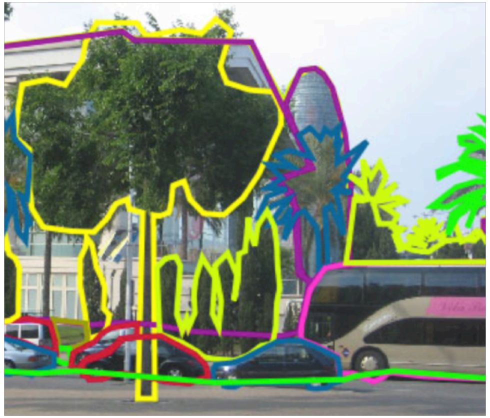
Bilateral Space Video Segmentation [Märki 2016]



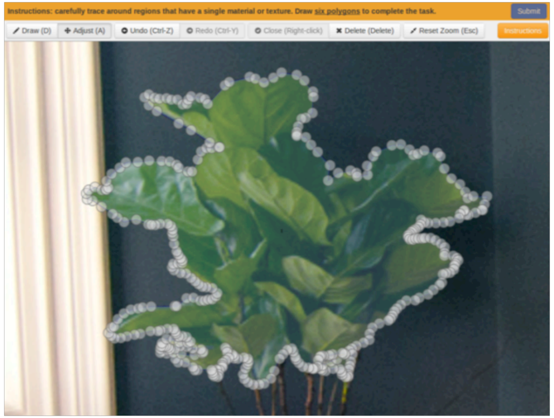

Video Segmentation via Object Flow [Tsai 2016]

# Related Work: Crowdsourcing



LabelMe [Russell 2007]

COCO [Lin 2014]

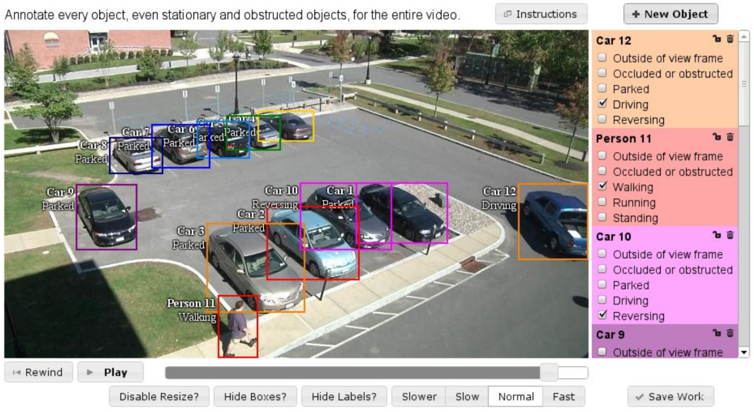



Open Surfaces [Bell 2013]

# Related Work: Video Annotation

LabelMe video: Vuilding a Video Database with Human Annotations [Yuen 2009]

Efficiently Scaling Up Crowdsourced Video Annotation [Vondrick 2012]

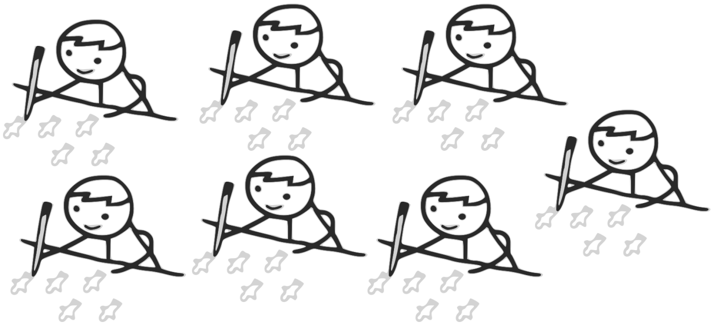


Crowdsourcing

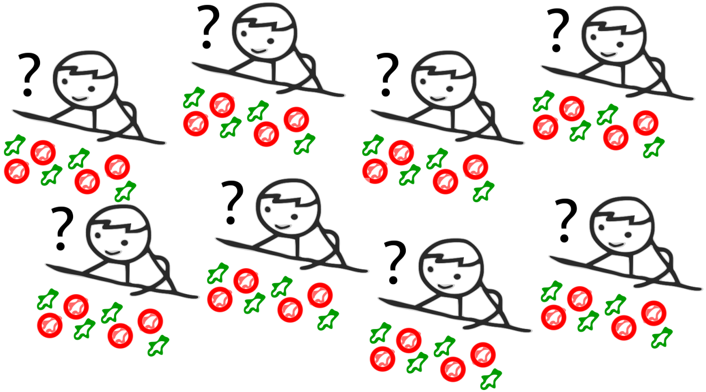
Higher-Quality Segmentation

# OVERVIEW

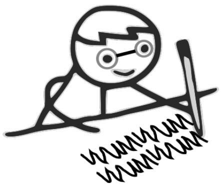
Segmentation



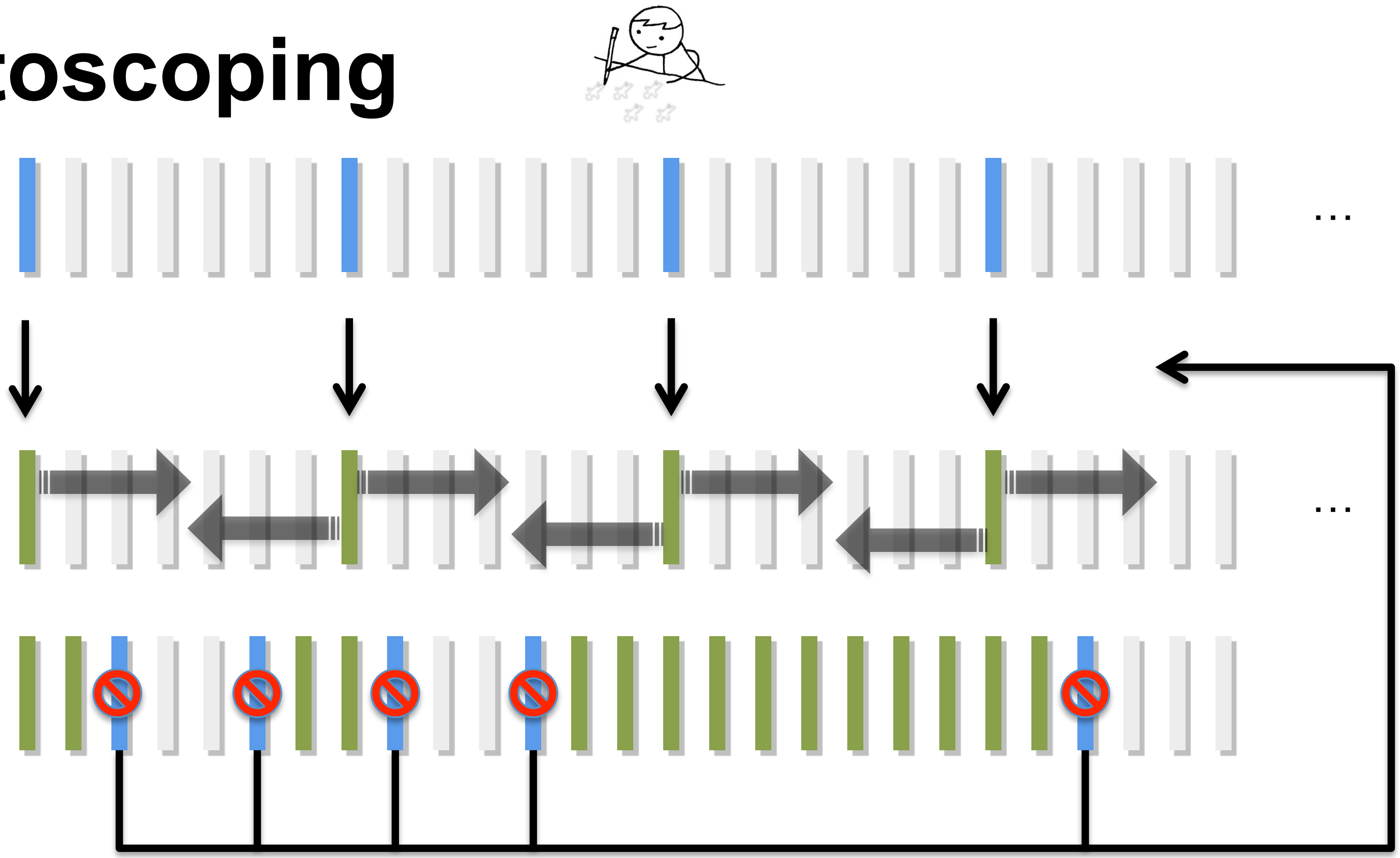
Propagation



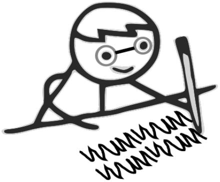
Requester



# Rotoscoping

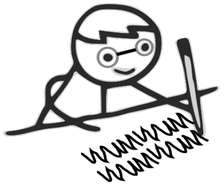


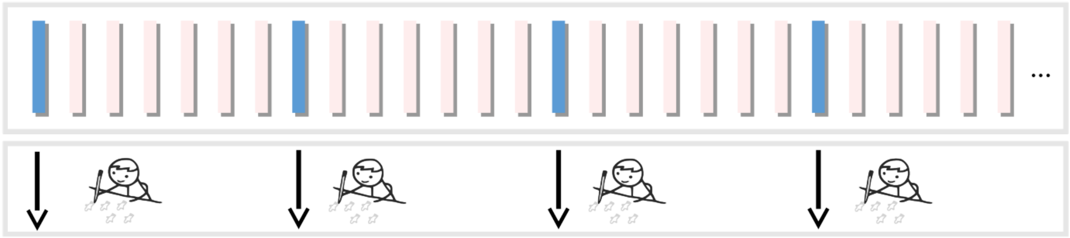
# Rotoscoping



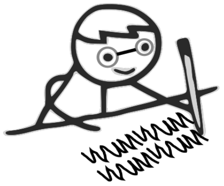
Requester

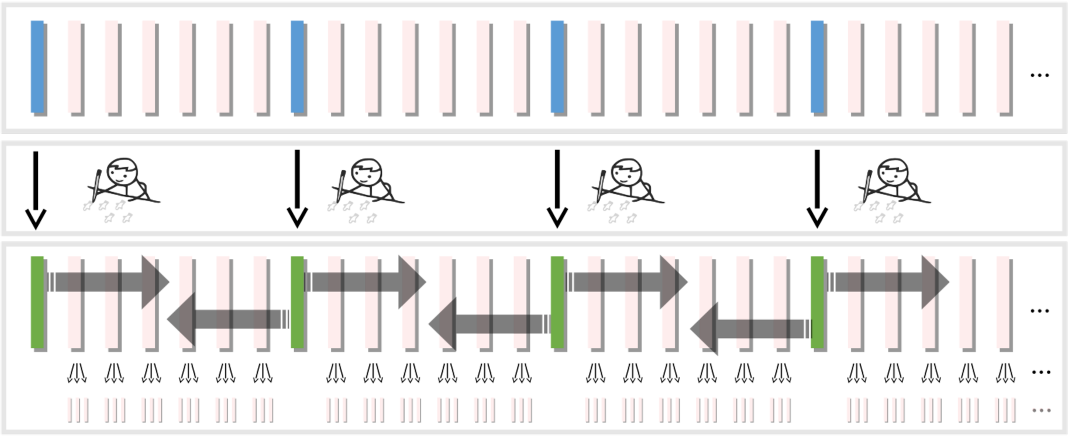
# Rotoscoping



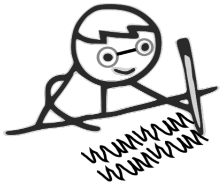


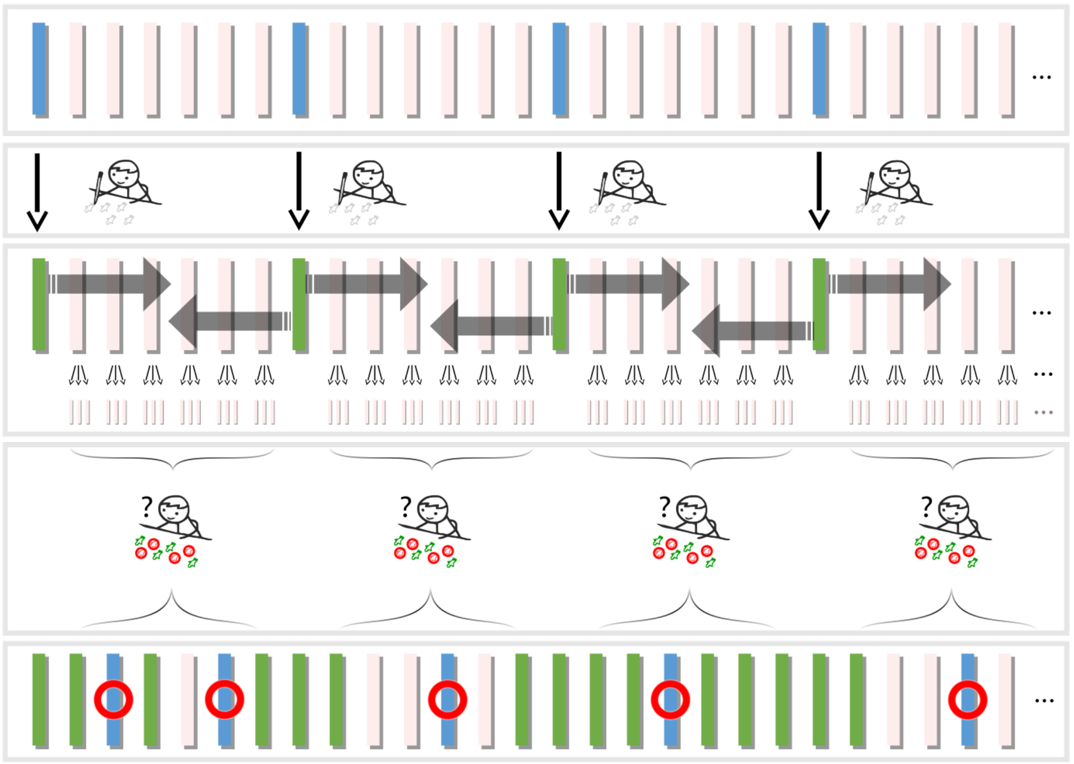
# Rotoscoping





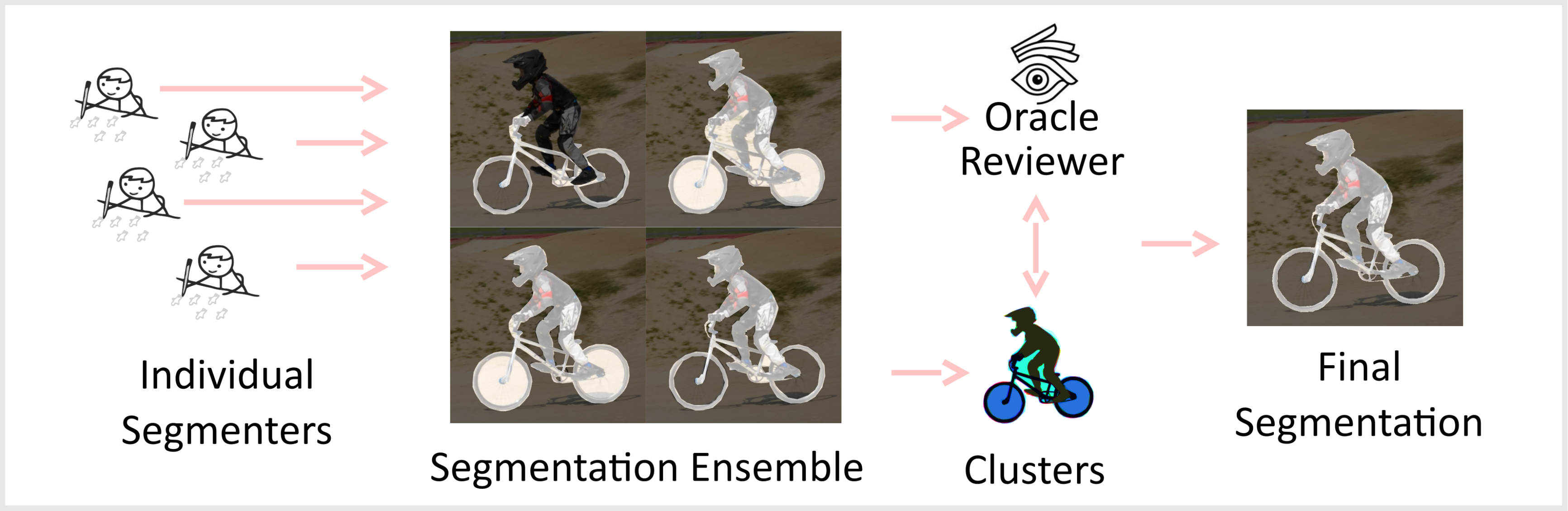
# Rotoscoping





# SEGMENTATION ENSEMBLE

# Overview


Individual Segmenters
Segmentation Ensemble
Oracle Reviewer
Clusters
Final Segmentation

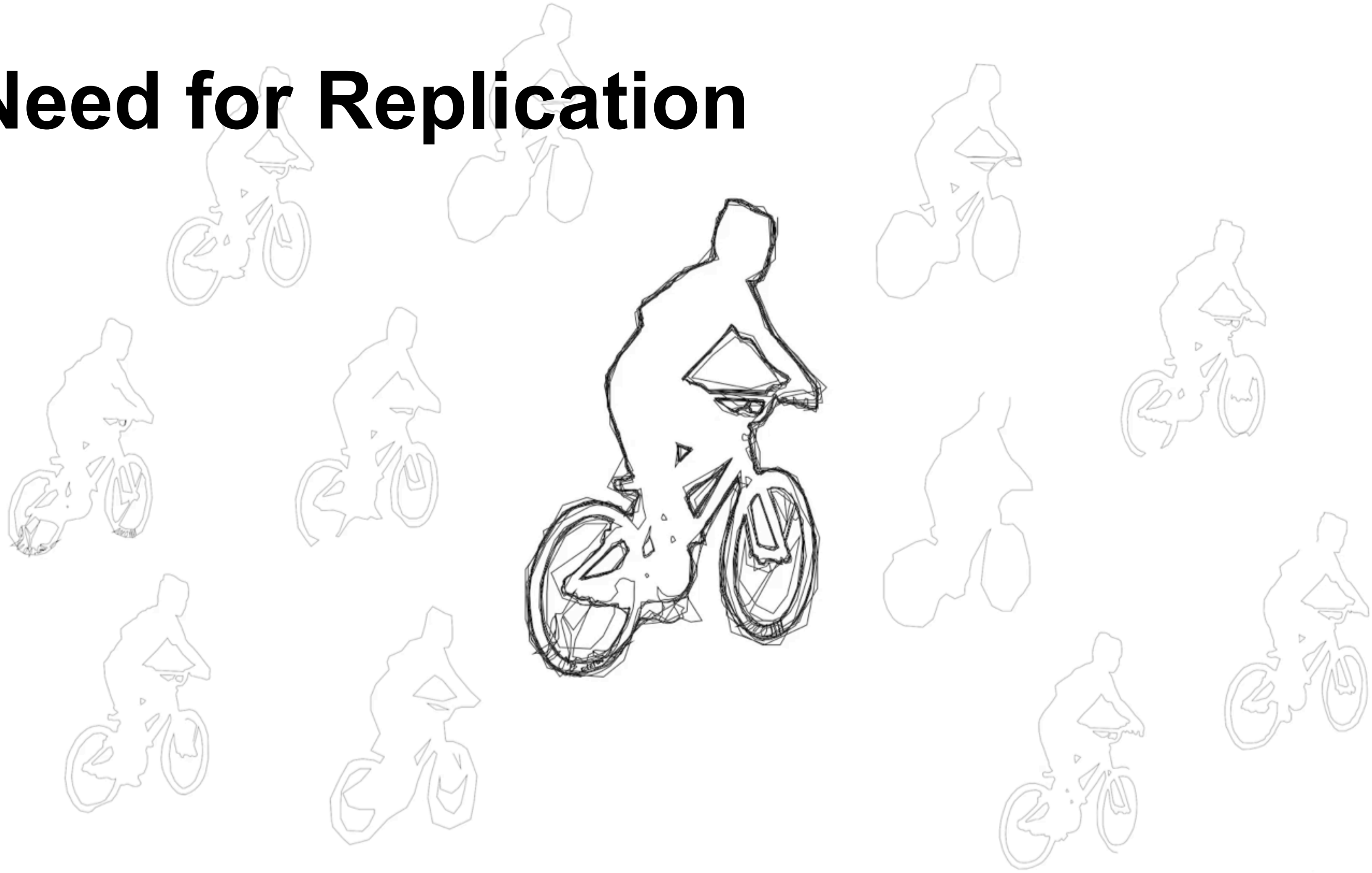
# Need for Replication

# Need for Replication

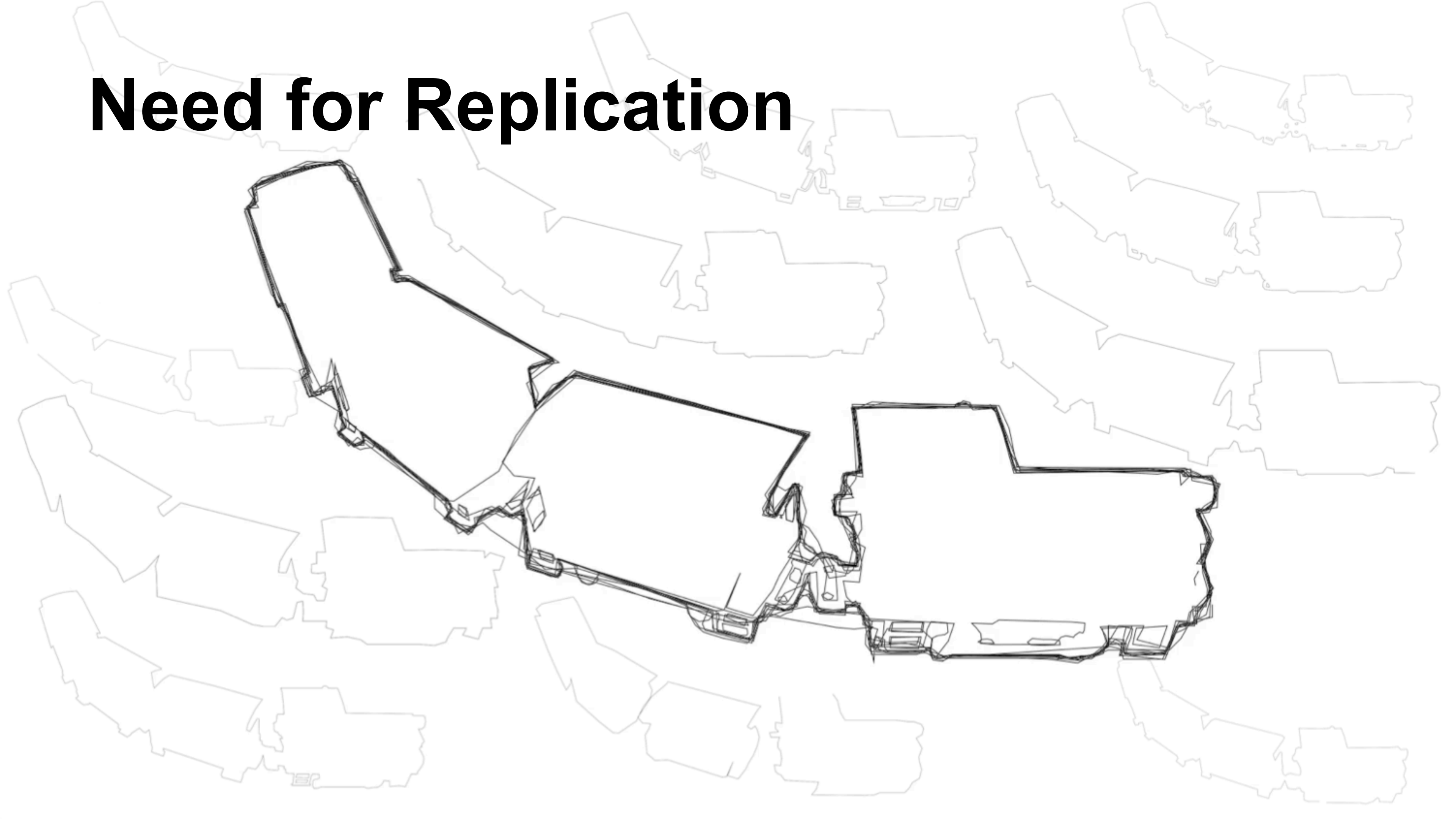
# Need for Replication

# Need for Replication

# Need for Replication



# Need for Replication



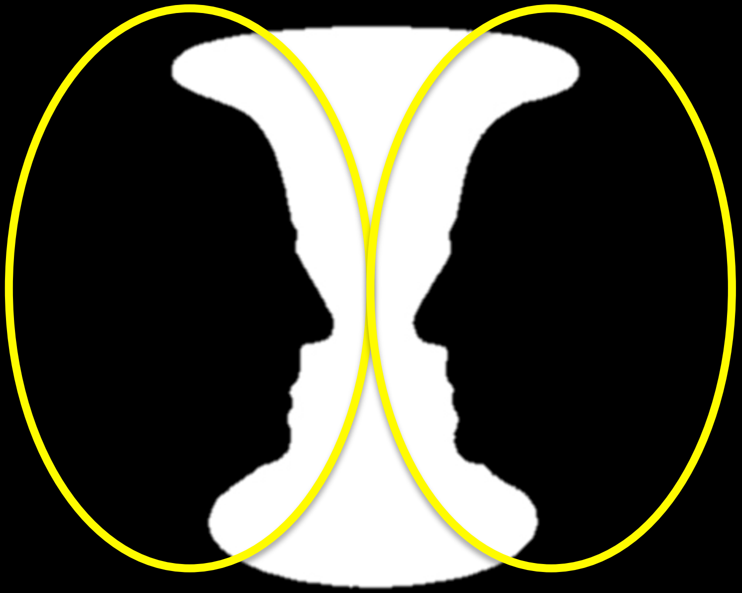
# Segmentation Ensemble

- Negative polygons
- Clustered merging
- Reviewer weighting
  (serving two purposes)

# Negative Space

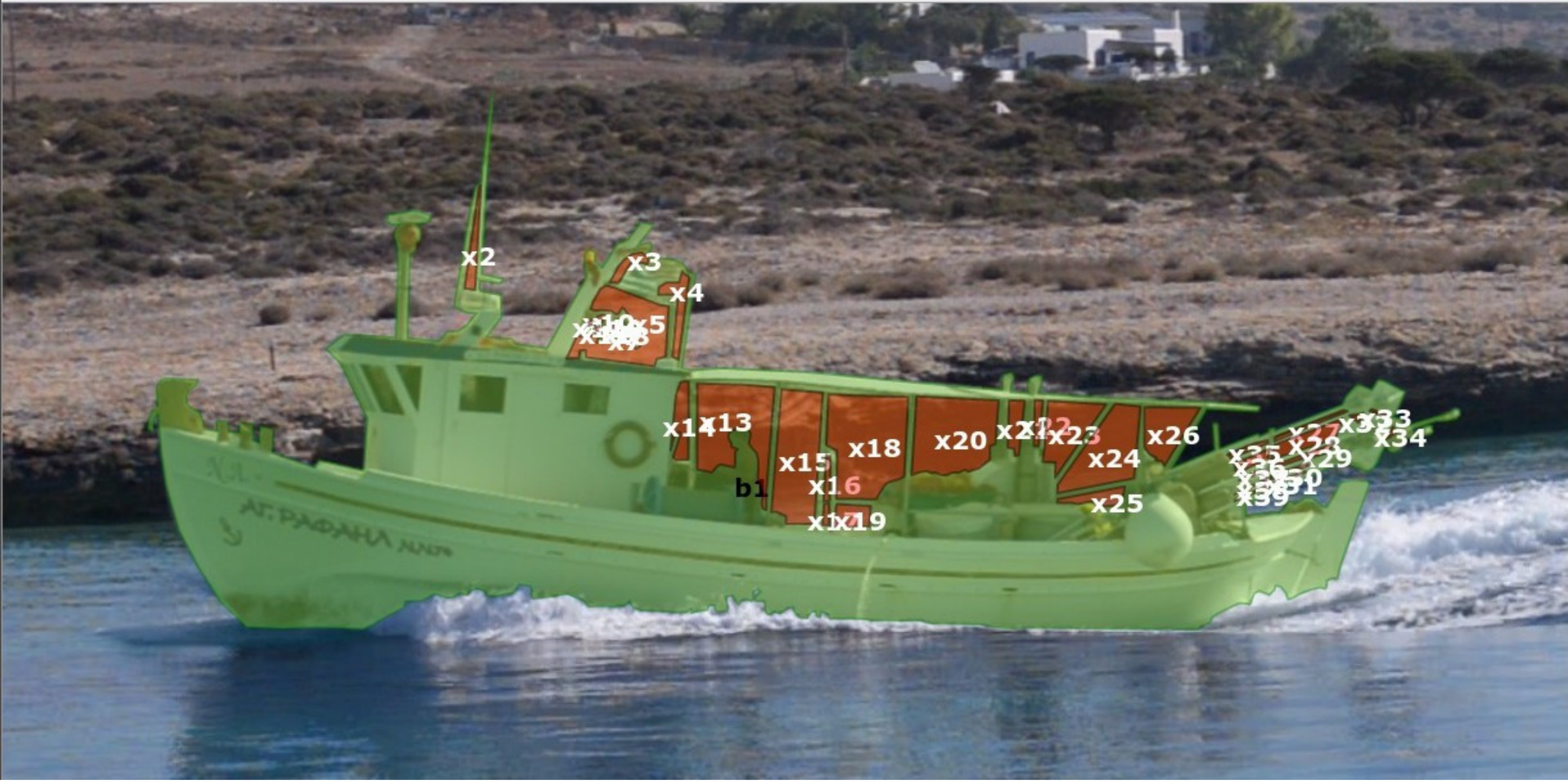
# Negative Space

# Negative Space



Instructions: carefully trace the contour of each element mentioned in the rotoscoping instructions. Assign the correct class to each polygon. **Estimated bonus**: 76¢ | Submit

Draw (D) | Adjust (A) | History | Actions | Class: Boat | Drag | Data | Video | Instructions



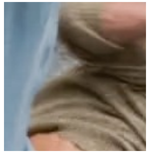

ΑΓ. ΡΑΦΑΗΛ

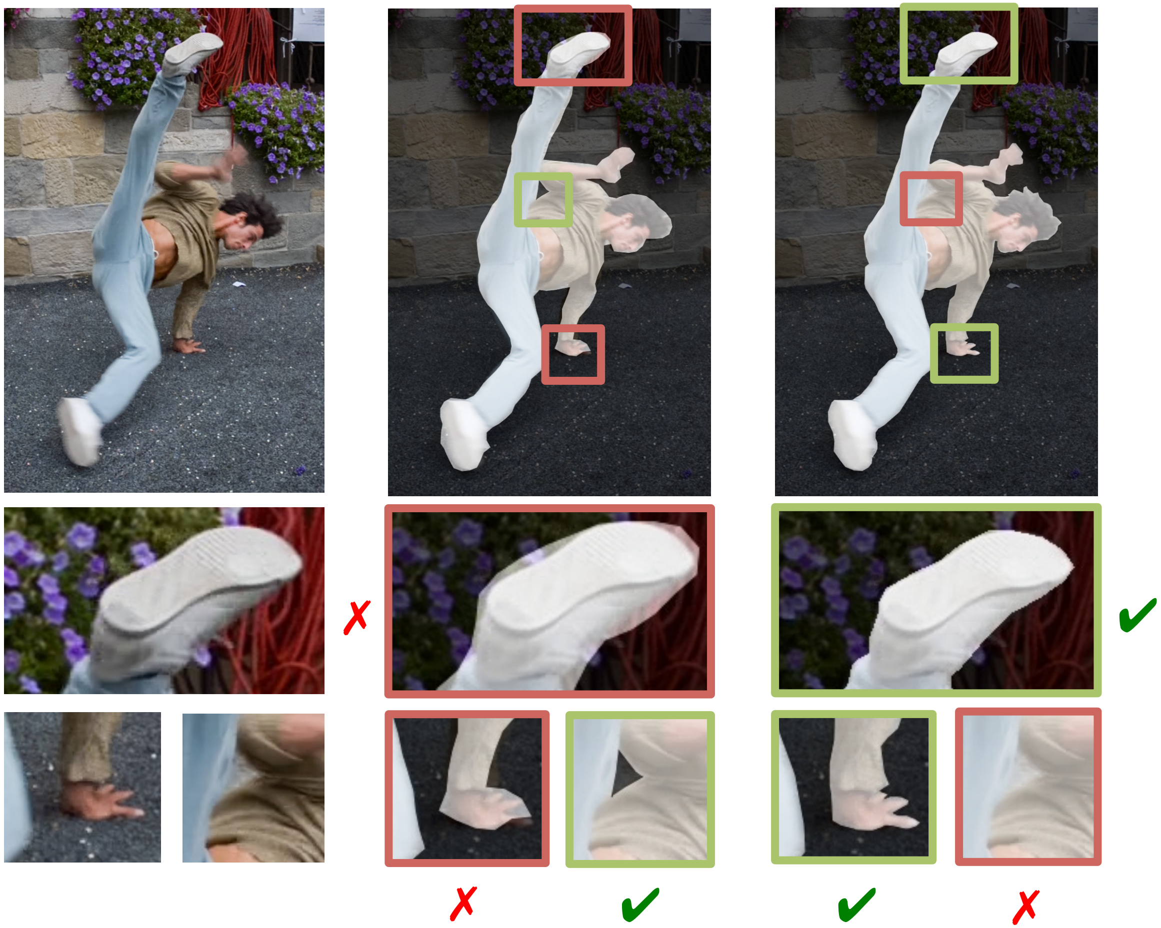
# Merging Segmentations

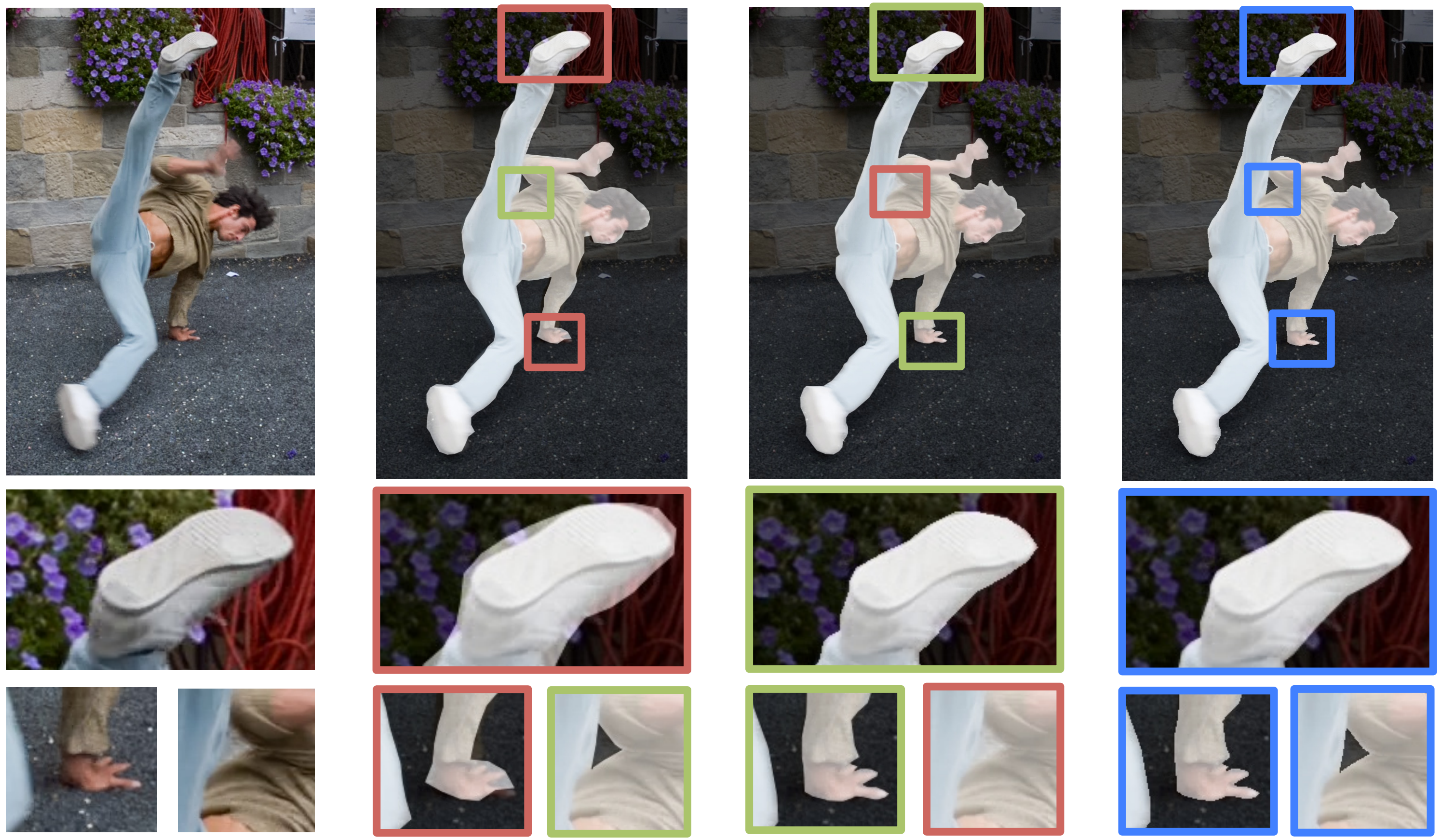
- **Clustering** by overlap + type of polygon

- **Weighted** combination







# Weights as bonuses

Total task reward:

$$R_i = ?$$

# Weights as bonuses

Total task reward:

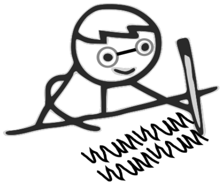
$$R_i = r + \beta_i$$

Weights and bonuses:

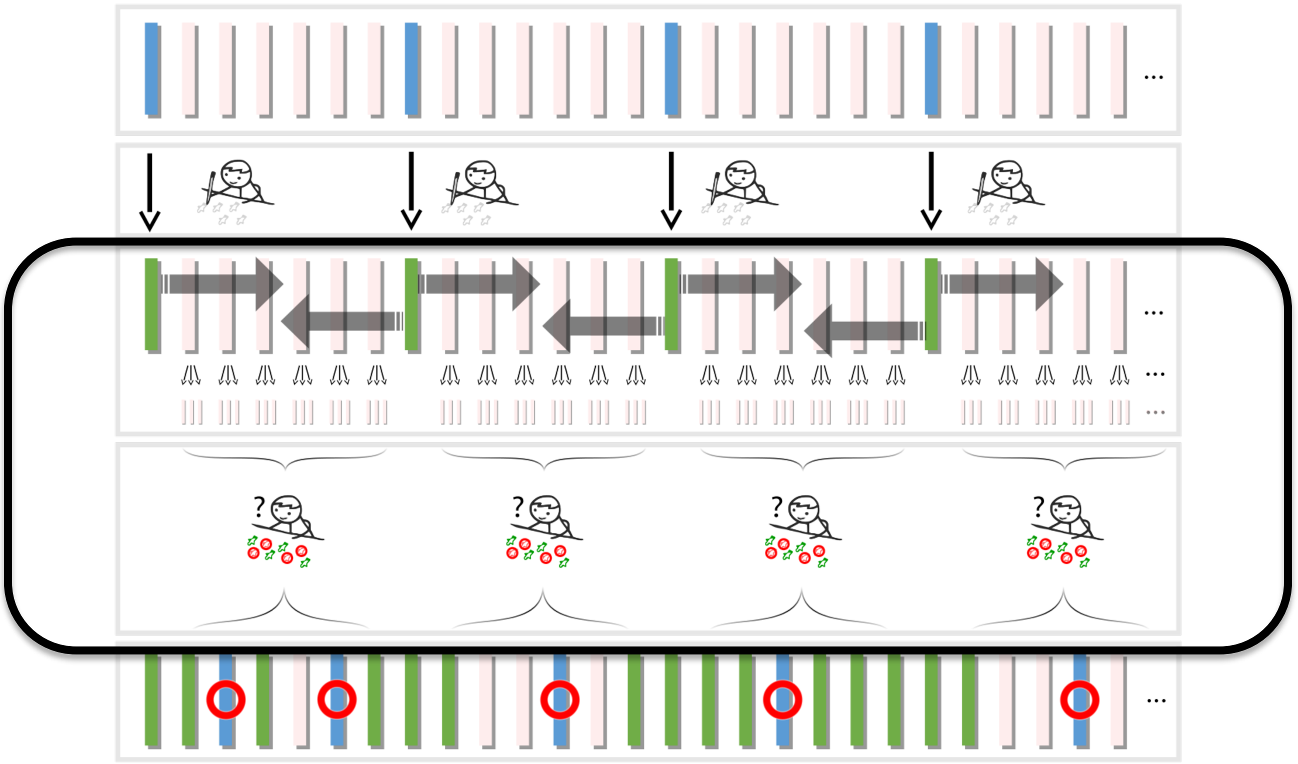
$$w_i = 1 + \beta_i$$

# PROPAGATION ENSEMBLE

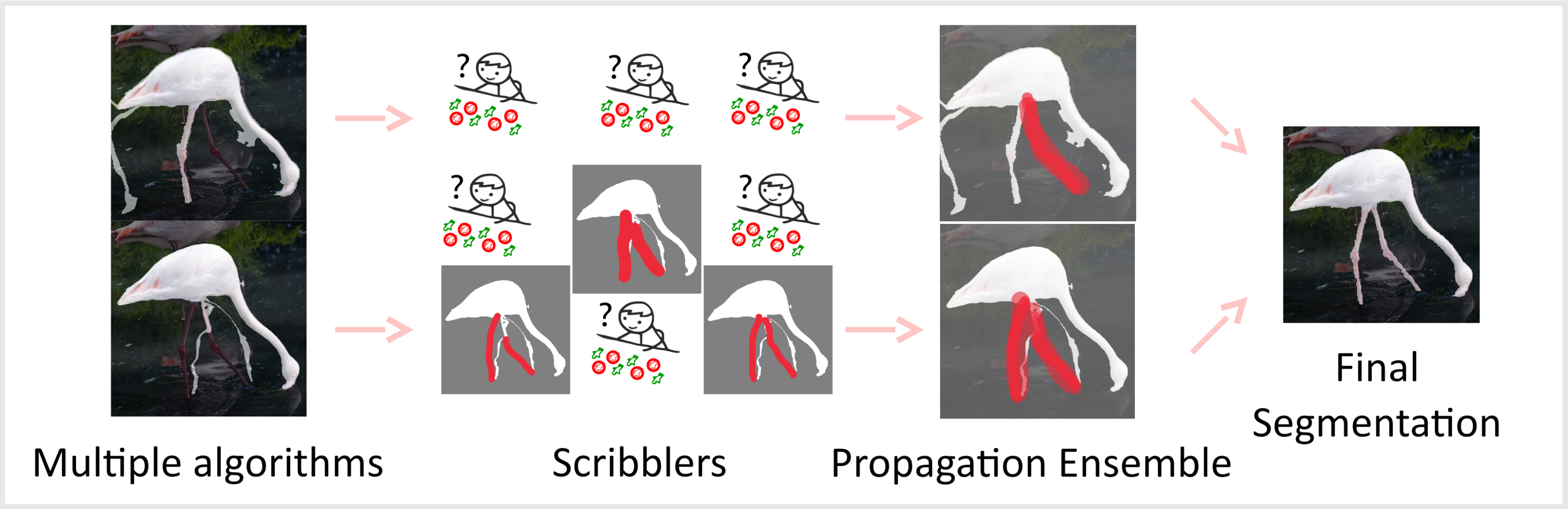
# Rotoscoping



Propagation Ensemble



# Overview



Multiple algorithms
Scribblers
Propagation Ensemble
Final Segmentation


Click on the images below to see the frame with more or less of the mask.

Click on the circles below to change the marker size.

< | play | > | undo

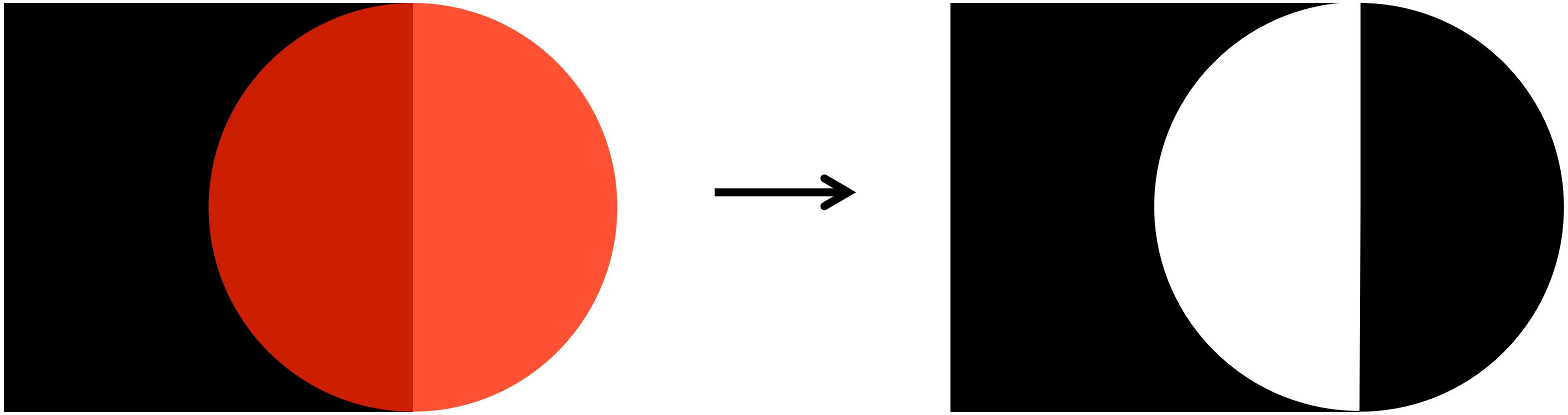
# Different Brush Scenarios

1. Corrective Fixing
   (inverting brushes)

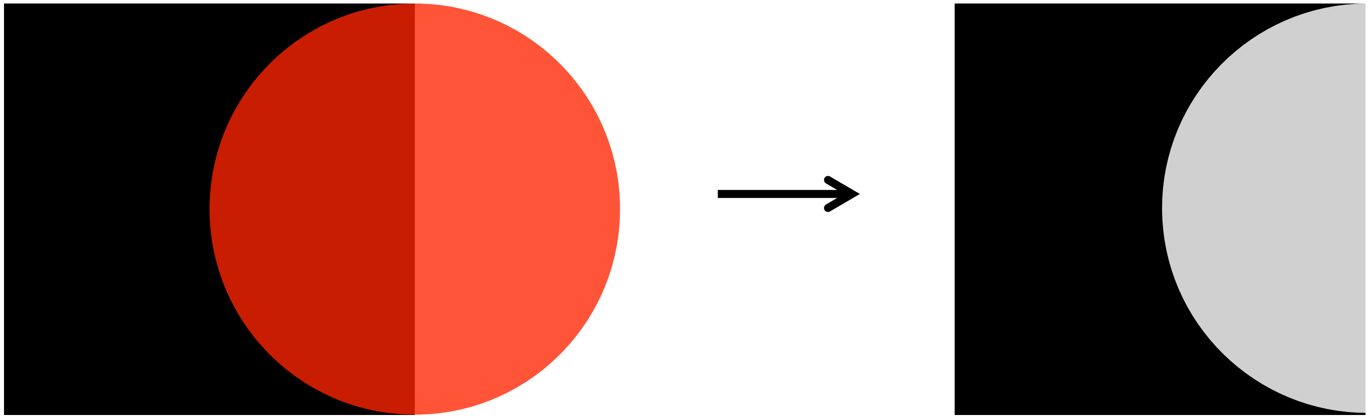
2. Penalty-based Weighting
   (weighting brushes)

3. Segmentation Refinement
   (2 brushes)

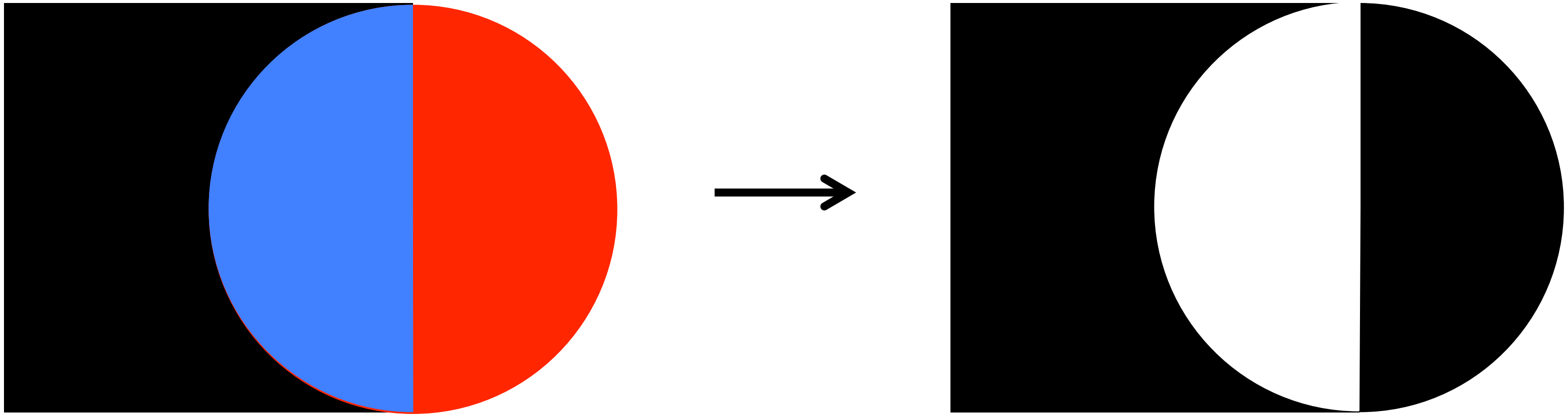
# Corrective Fixing



# Penalty-based Weighting

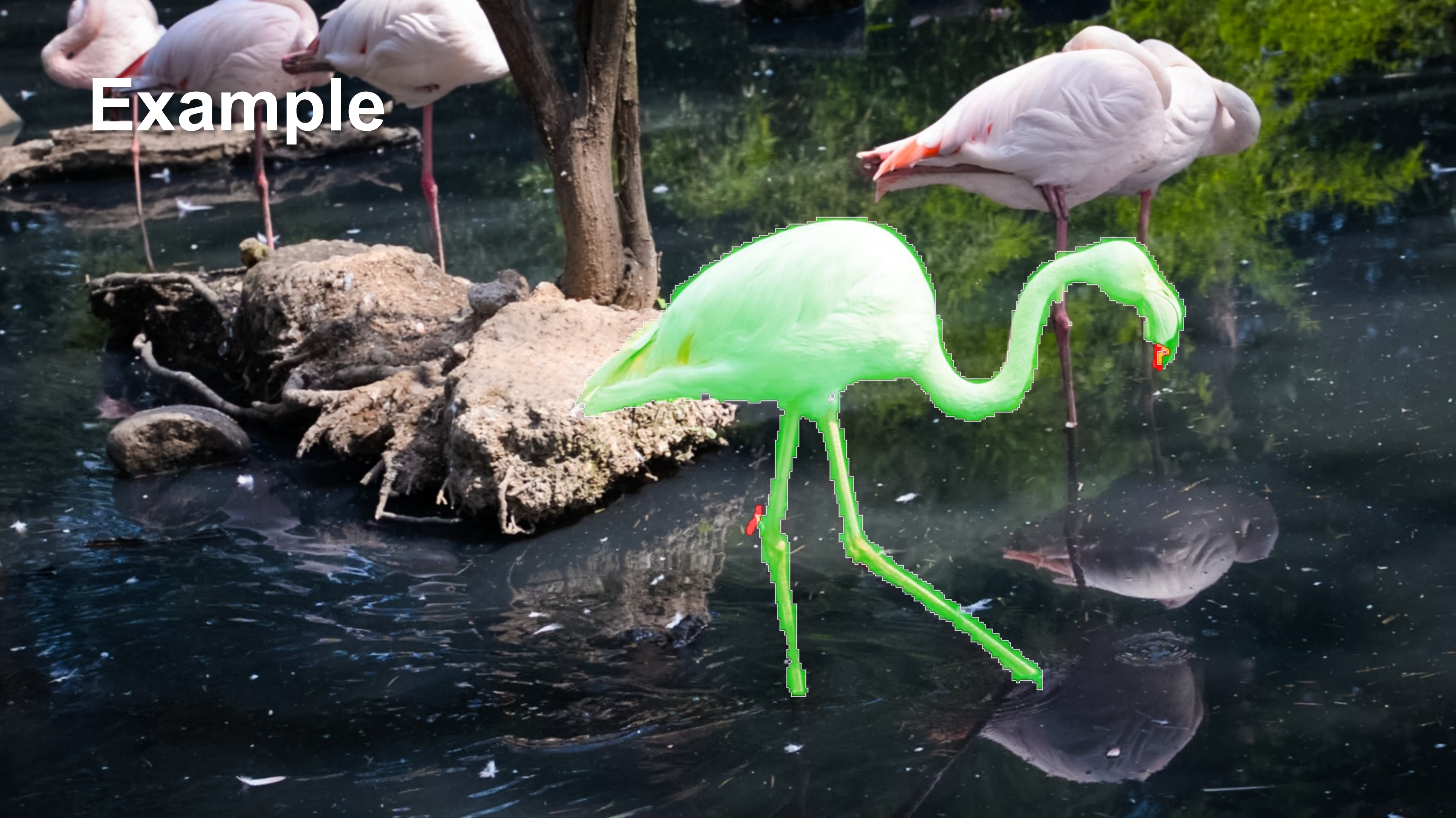


# Segmentation Refinement



# Example

# Example



# Example

# Example

# Example

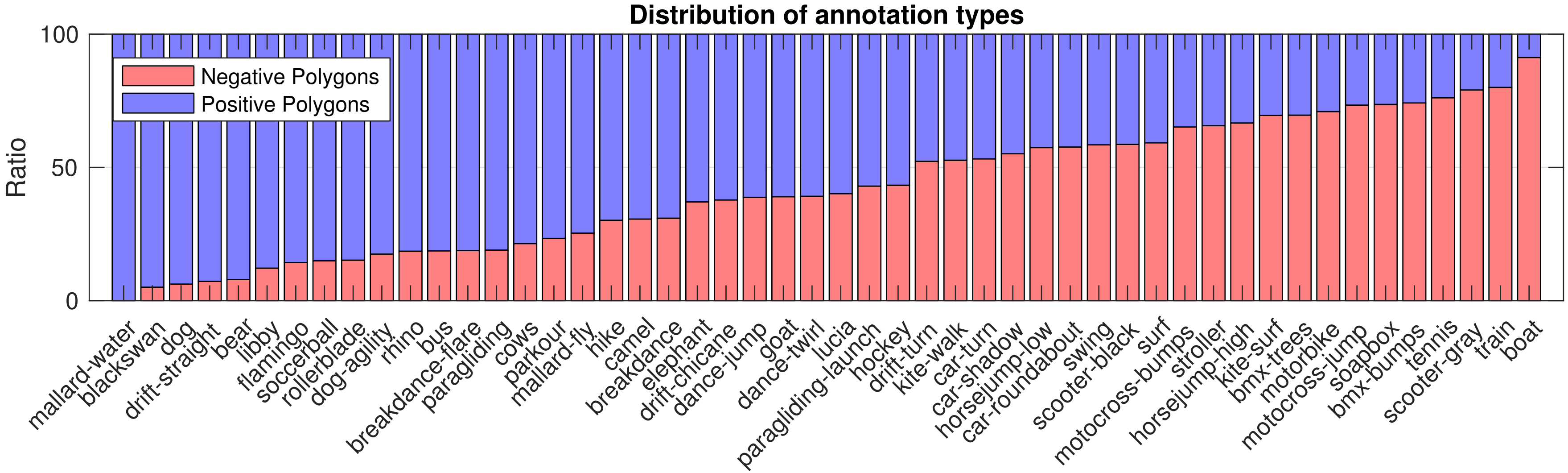
# Example

# EVALUATION


DAVIS Dataset

# Segmentation Ensemble

- Do workers use our negative polygons?
  - ✔ Yes. Extensively.



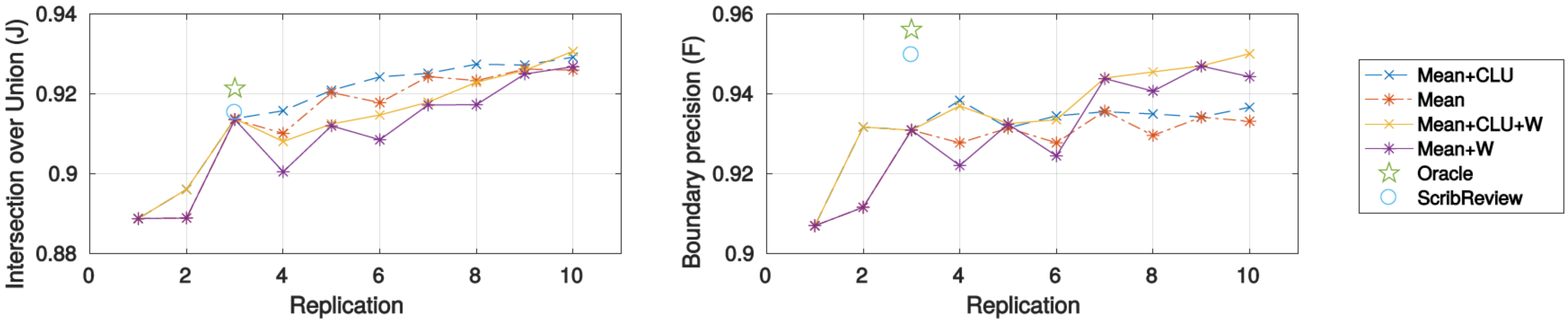
# Segmentation Ensemble

- Does our strategy produce higher quality?
  ✔ Yes. Notably for the boundaries.

| | Oracle | | Automatic | | |
|---|---|---|---|---|---|
| Metrics | **Clu/Mean** | Best | Clu/Mean | SDF | Worst |
| $J \uparrow$ | **0.917** | **0.914** | 0.903 | 0.875 | 0.842 |
| $F \uparrow$ | **0.952** | **0.939** | 0.928 | 0.880 | 0.879 |
| $T \downarrow$ | 0.380 | 0.372 | 0.359 | **0.350** | 0.483 |

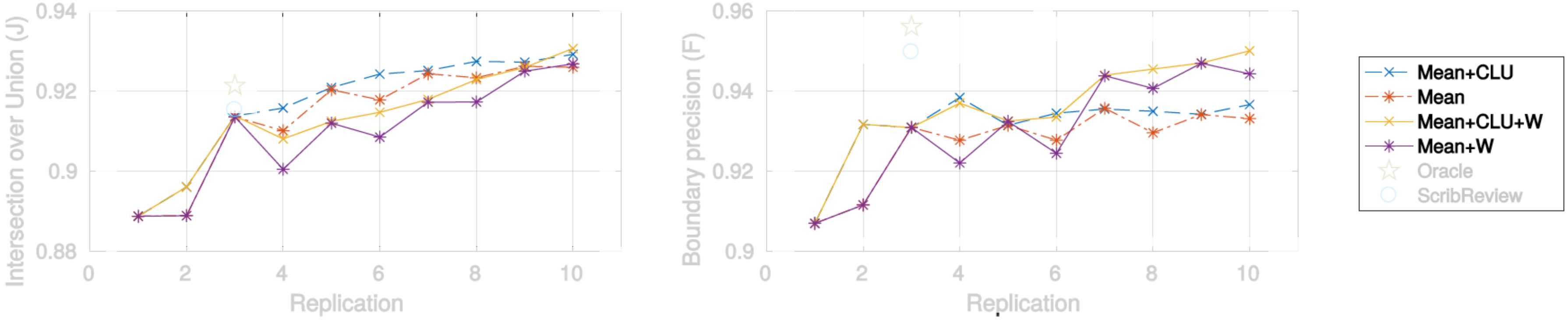
# Segmentation Ensemble

- How does replication impact the quality?
  - ✔ Positively!



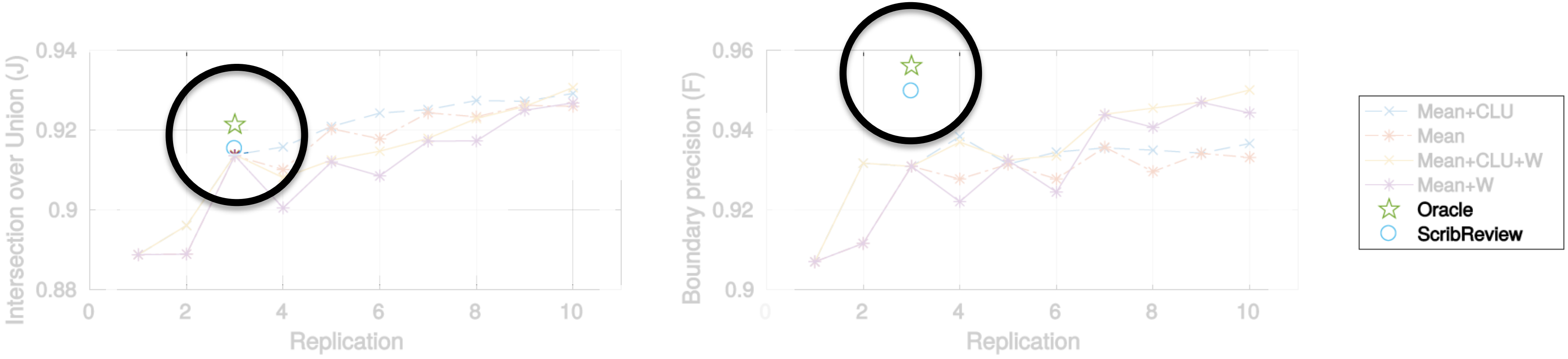
# Segmentation Ensemble

- How does replication impact the quality?
  - ✔ Positively!



# Segmentation Ensemble

- How does replication impact the quality?
  ✔ Positively!

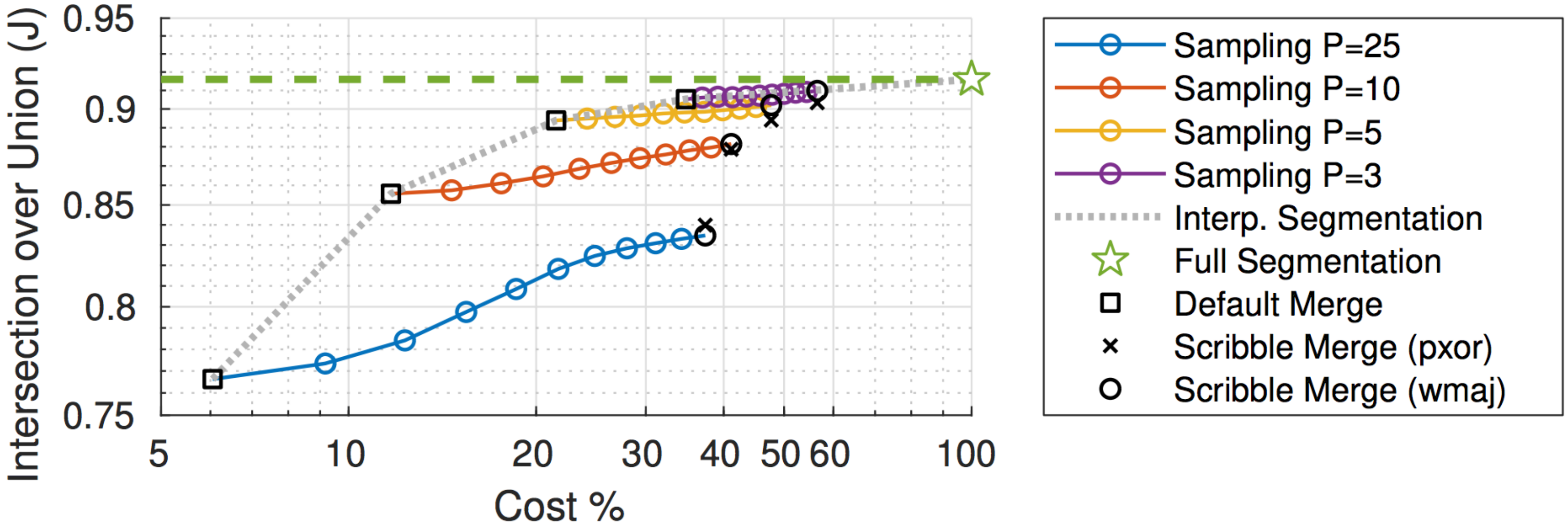


# Propagation Ensemble

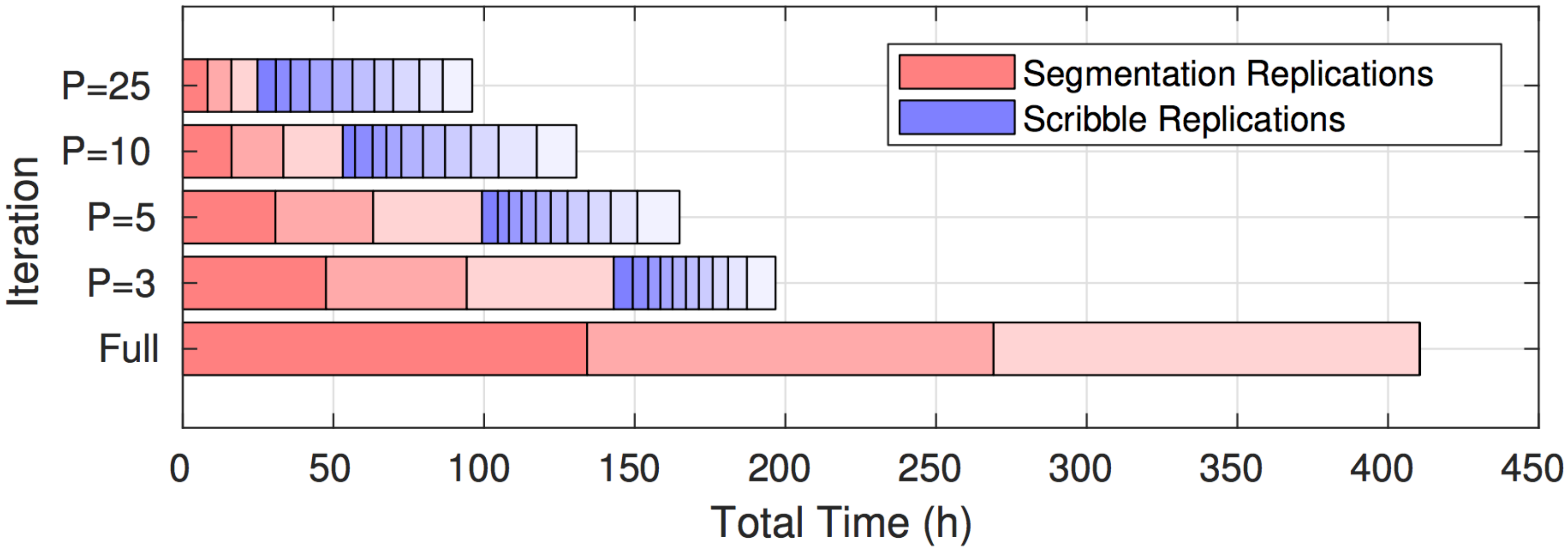
- Do the scribble improve quality?
  ✔ Yes, consistently.

- Are they efficient?
  ✗ Not for coarse samplings.
  ~ Maybe for fine samplings.

# Propagation Ensemble



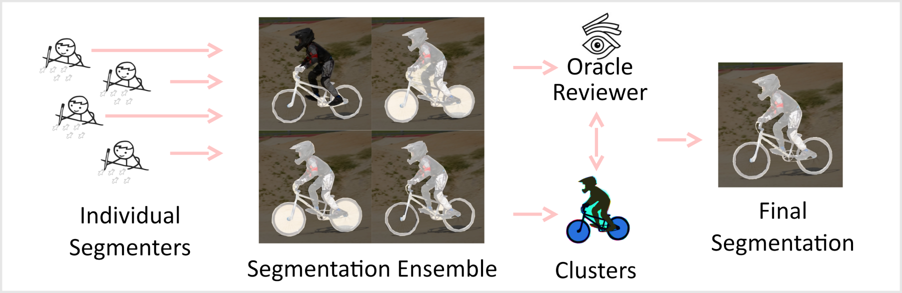
# Propagation Ensemble

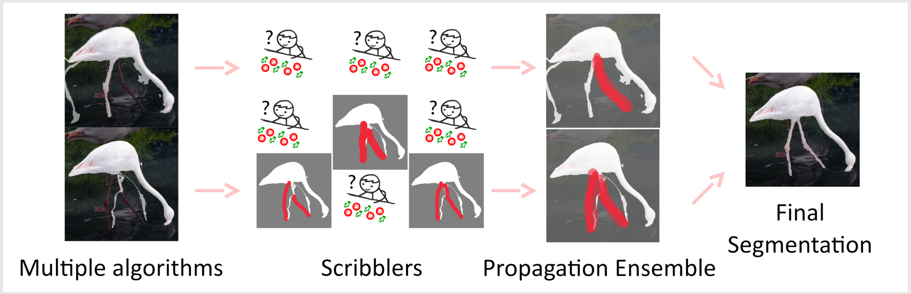


# Propagation Ensemble

- Which scribble scenario is most effective?
  As penalty weighting.

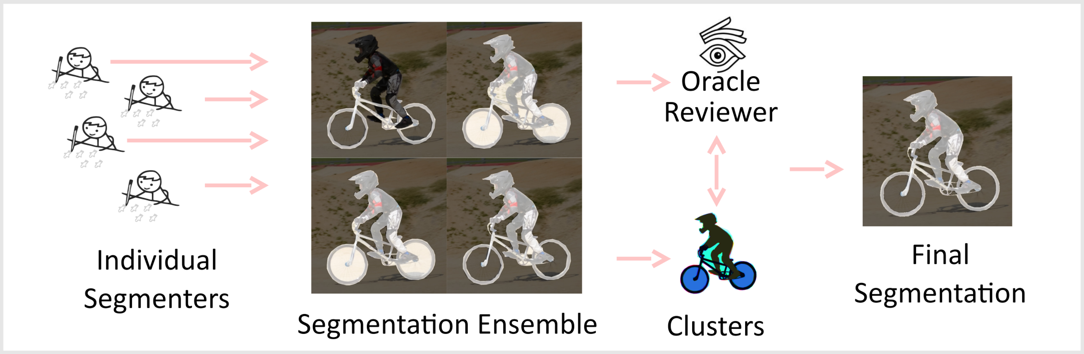
- Corrective fixing:
  - Unstable (see paper).
  - Single candidates cannot be improved.

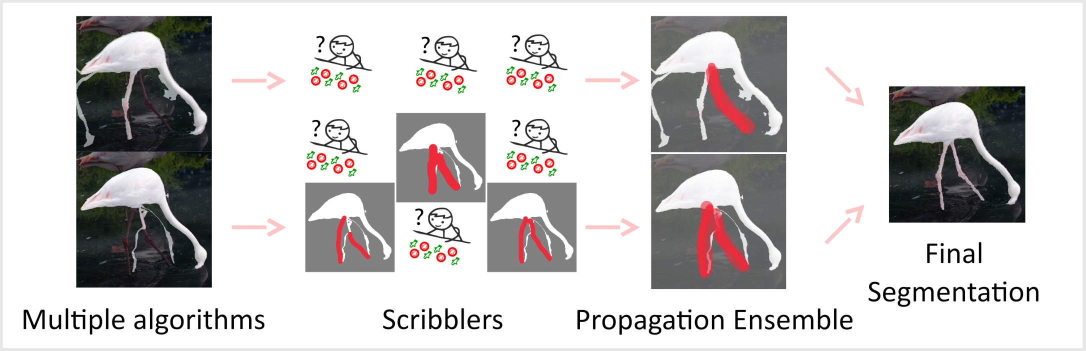



Crowd workers can create
(much) higher quality segmentation!

Tools and data will be made available online:
http://crowdensembles.csail.mit.edu




Tools and data will be made available online:

http://crowdensembles.csail.mit.edu

# Are the metrics meaningful?

| Method | Metric $J$ | $F$ | $T$ | $W$ |
|---|---|---|---|---|
| LDOF | 0.760 | 0.749 | 0.250 | **0.204** |
| DF+SVM | 0.764 | 0.782 | **0.250** | 0.192 |
| DF+SVM+Att | **0.822** | **0.823** | 0.317 | 0.198 |

**Table 7. Comparisons between the scores $W$ evaluated by crowd workers and the $J$, $F$, and $T$ metrics, as in Table 1. All values are averages over all sequences of the DAVIS dataset. The score $W$ is defined as the average number of positive evaluations for a frame segmentation.**