

# Fast Online Object Tracking and Segmentation: A Unifying Approach

Alessandro, Vish, and Mel

# Arguments for SiamMask

- Real-time tracking at 55fps
  - Mask Track R-CNN has no speed information
  - Real-time performance directly translates to efficacy of real-world application
- Citation count

## [Fast online object tracking and segmentation: A unifying approach](#)

[Q Wang](#), [L Zhang](#), [L Bertinetto](#)... - Proceedings of the ..., 2019 - openaccess.thecvf.com

... In Table 2 we compare the two variants of **SiamMask** with MBR strategy and **SiamMask-Opt** against five recently published state-of-the-art trackers on the VOT-2018 benchmark. Unless ...

☆ Save 📄 Cite Cited by 1335 Related articles All 11 versions 🔗

## [Video instance segmentation](#)

[L Yang](#), [Y Fan](#), [N Xu](#) - Proceedings of the IEEE/CVF ..., 2019 - openaccess.thecvf.com

... In addition, we propose a novel algorithm called **MaskTrack R-CNN** for this task. Our new method introduces a new tracking branch to Mask **R-CNN** to jointly perform the detection, ...

☆ Save 📄 Cite Cited by 394 Related articles All 6 versions 🔗

- Stronger datasets
  - Better generalization
  - YouTube VIS has 40 vs YouTube VOS 90 categories

# Video Instance Segmentation (VIS)

**ICCV 2019**

Lingjie Yang, Yuchen Fan, Ning Xu

Presented by Amit, Michael, and Jun

# Debate

## 1. New Benchmark Dataset (YouTube-VIS)

- 2.9k videos and 40 object categories
- (SiamMask: no new dataset)

## 1. No Initialization needed

- An instance that appears in a middle frame can be segmented and tracked.
- (SiamMask: need initialization)

## 1. New Task

- Perform multiple instance segmentations simultaneously
- (SiamMask: overestimated for inference speed (1 instance: 60fps -> 3 instances: 20fps))

# Video Instance Segmentation

70 papers with code · 8 benchmarks · 8 datasets

The goal of video instance segmentation is simultaneous detection, segmentation and tracking of instances in videos. In words, it is the first time that the image instance segmentation problem is extended to the video domain.

To facilitate research on this new task, a large-scale benchmark called YouTube-VIS, which consists of 2,883 high-resolution YouTube videos, a 40-category label set and 131k high-quality instance masks is built.

## Benchmarks

[Add a Result](#)

These leaderboards are used to track progress in Video Instance Segmentation

Trend	Dataset	Best Model	Paper	Code	Compare
	YouTube-VIS validation	UNINEXT-H			<a href="#">See all</a>
	OVIS validation	DVIS(Swin-L, Offline)			<a href="#">See all</a>
	YouTube-VIS 2021	RefineVIS (Swin-L, online)			<a href="#">See all</a>
	BDD100K val	PCAN			<a href="#">See all</a>
	Youtube-VIS 2022 Validation	CTVIS (Swin-L)			<a href="#">See all</a>
	HQ-YTVIS	VMT (Swin-L)			<a href="#">See all</a>
	YouTube-VIS	STC			<a href="#">See all</a>
	Youtube-VIS (trained with no video masks)	MaskFreeVIS			<a href="#">See all</a>