VTT: Long-term Visual Tracking with Transformers

Presentation, ICPR 2020

Tianling Bian
Shanghai Jiao Tong University, China

Yang Hua
Queen’s University Belfast, UK

Tao Song
Shanghai Jiao Tong University, China

Zhengui Xue
Shanghai Jiao Tong University, China

Ruhui Ma
Shanghai Jiao Tong University, China

Neil Robertson
Queen’s University Belfast, UK

Haibing Guan
Shanghai Jiao Tong University, China

Neil Robertson
Queen’s University Belfast, UK
Goal

- Long-term visual tracking
- Difficulties
  - Occlusion
  - Scale and ratio variation
  - Distractors
Related Work

• GlobalTrack*
  • Unable to handle Occlusion, scale and ratio variation
  • Inferior performance against instance level distractors

* GlobalTrack: A Simple and Strong Baseline for Long-Term Tracking. Huang, Zhao, and Huang. AAAI 2020
Our Method

• Transformers* were first proposed for machine translation

* Attention is all you need. Ashish, et al. NeurIPS 2017
Our Method

• We adapt this architecture for tracking
Our Method

• End to end trainable simple architecture
Experiment: Results

• Significant improvement compared to GlobalTrack on LaSOT* as well as other long-term tracking datasets.

* Lasot: A high-quality benchmark for large-scale single object tracking. Heng, et al. CVPR 2019
Experiment: Encoder Attention

- Lower level attention focus on visuals
- Higher level attention focus on semantics
- Attend to disconnected regions
Experiment: Decoder Attention

Occlusion

Large scale

Extreme aspect ratio
Experiment: Distractors

• Compared to GlobalTrack, our method is significantly better
VTT: Long-term Visual Tracking with Transformers

Presentation, ICPR 2020

Tianling Bian  
Shanghai Jiao Tong University, China

Yang Hua  
Queen’s University Belfast, UK

Tao Song  
Shanghai Jiao Tong University, China

Zhengui Xue  
Shanghai Jiao Tong University, China

Ruhui Ma  
Shanghai Jiao Tong University, China

Neil Robertson  
Queen’s University Belfast, UK

Haibing Guan  
Shanghai Jiao Tong University, China