

Motivation

360° Video Saliency Detection

- One of the core problems in 360° video understanding
- Directly related to practical scenarios
- 360° video summarization
- Dynamic rendering for VR

Problem 1: Distortion & Discontinuity

 \rightarrow Architecture design for panoramic videos



Ignore geometric property [1] ☺ Simple

Not scalable



Tailored for spherical input

• Not transferrable



Additional modules & training [4]

- Transferrable
- (:) Overhead, error accumulation
- \rightarrow **Our solution**: Local patch-based modeling
- Format-agnostic geometry-aware approximation
- (Transferrable from most ViT variants
- (a) No additional training for geometric adaptation

Problem 2: Ambiguity and Subjectivity

 \rightarrow Modeling saliency in panorama without supervision \rightarrow **Our solution**: Leverage features from NFoV domain Rich and readily available pretrained knowledge Spatio-temporal feature consistency suffices ⊕ Not depend on optical flow, class activation, etc.

Panoramic Vision Transformer for Saliency Detection in 360° Videos

Use designated architecture [2,3]





