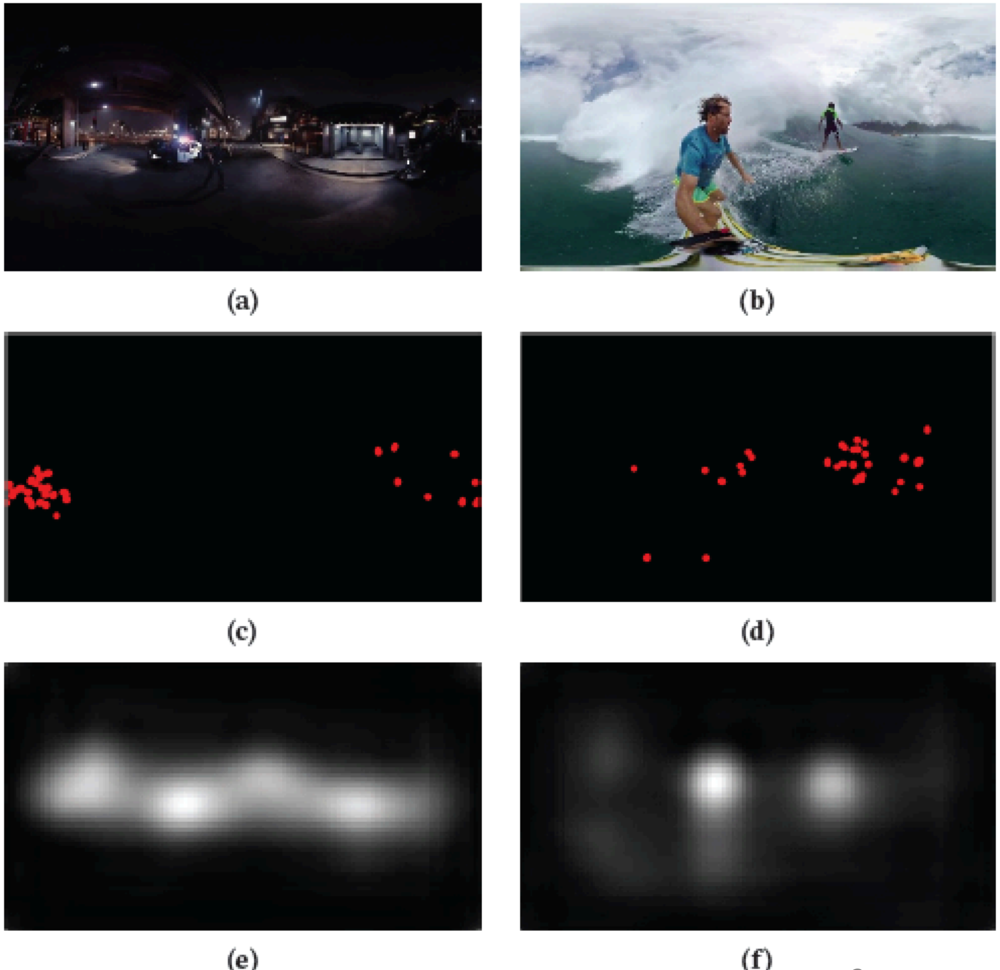
Introduction

- head movement prediction is one essential yet daunting task that needs to be addressed urgently
  - enable bandwidth-efficient 360-degree video streaming
  - significantly reduce the motion-to-photon delay
- Contributions
  - Building a dataset for panoramic saliency
  - Training a saliency detection model for 360-degree videos
  - Consolidating a head movement prediction model for 360-degree videos

Saliency for 360-degree Videos

- two intrinsic problems of traditional saliency models for regular images/videos
  - Central Bias
  - Multi-object Confusion



Dataset

- Deriving Head Orientation (2 public head datasets)
  - 18 videos viewed by 48 users in 2 experiments (9 videos from 1st experiment)
  - 5 videos viewed by 59 users (2 videos from them)
- Extracting Fixation (at a specific region for a short period of time)
  - removing *saccade*: velocity over 20°/s and acceleration over $50°/s^2$
- Creating Fixation Maps
  - DBSCAN clustering to remove noisy points
- Outputting Saliency Maps
  - Gaussian Filter to generalize and smooth these scattered user fixation points to a statistical region

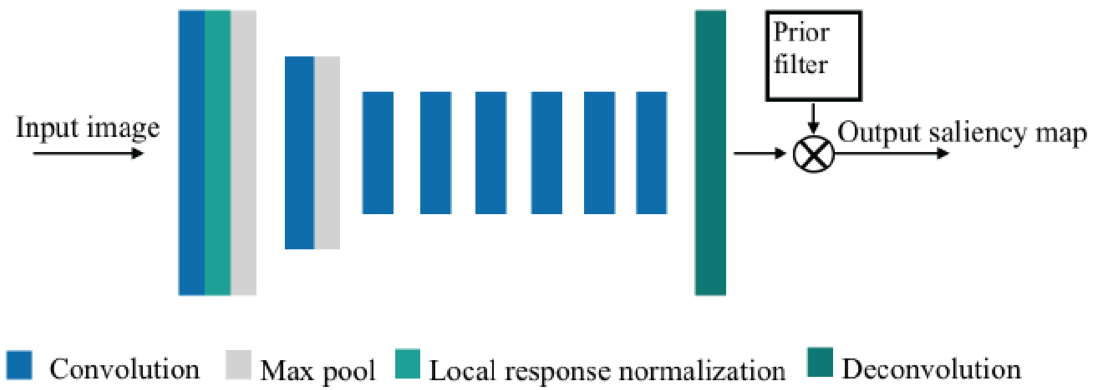
⇒ 11 videos and 7000 equirectangular frames with fixation maps and saliency maps

Dataset Evaluation

- Equator Bar
  - the saliency is linearly decreased from 1.0 to zero when the latitude is varied from 0° (equator line) to $\pm 90^\circ$ (two poles)
- Circle at Center
  - labels the frame center point as the highest saliency (1.0) and gradually decreases the saliency by expanding a circle around the center point

Model	sAUC	NSS	CC
Dataset Saliency	0.7966	1.9864	0.2521
Equator Bar	0.5012	0.8086	0.1078
Circle at Center	0.4462	0.3424	0.0487

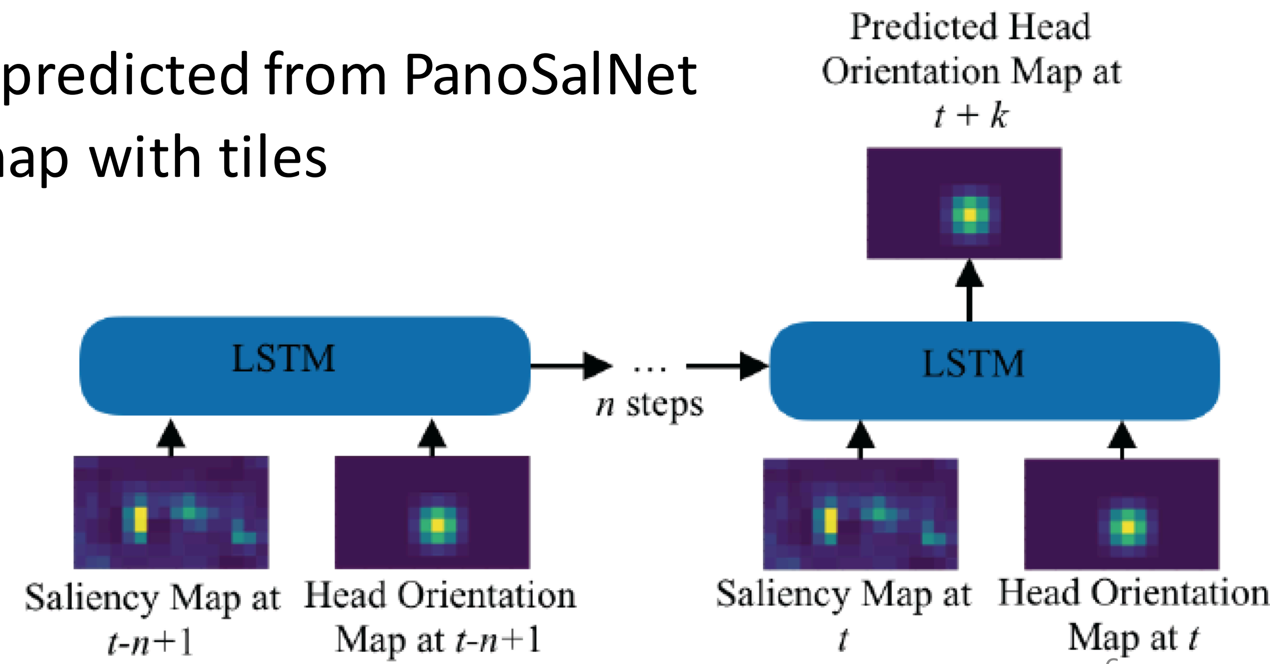
PanoSalNet



- Transfer learning from VGGNet and SALICON
- Model training
  - 400 pairs of video frames and saliency maps from the panoramic saliency dataset (concentrated fixation points)
  - Resolution: 512x288
  - 800 iterations to prevent overfitting

Head Movement Prediction in HMD

- Long Short-Term Memory (LSTM) model
  - Able to handle a large amount of temporal data and outperform other similar algorithms
- Input data
  - Saliency map predicted from PanoSalNet
  - Orientation map with tiles



Model Training

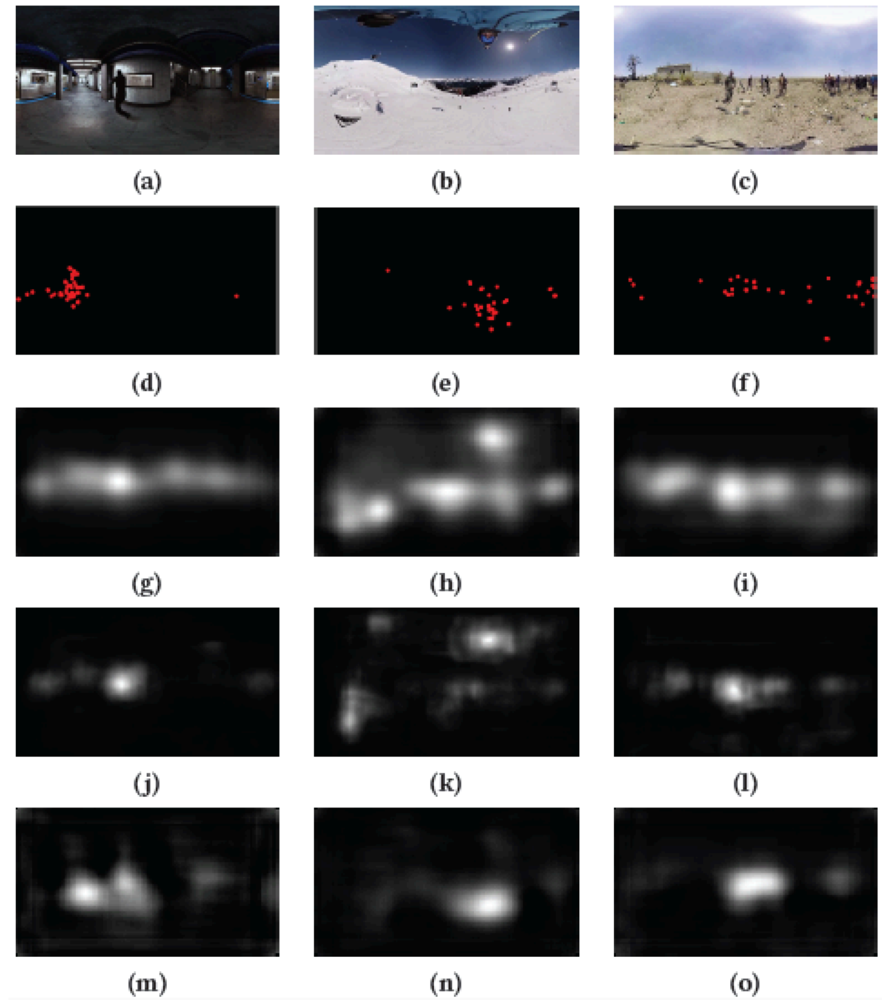
- 5 videos for training and 4 videos for validation
- 1 segment with a length of 20-45 second
  - One or more events happened and leads to a new saliency region and fast head movements of users
- 300,000 data samples from 432 time series using viewing logs of 48 users
- Loss function: Euclidean distance between predicted head orientation map and ground truth head orientation map
- Model parameters are updated with Root Mean Square Propagation (RMSprop) method

Saliency Detection Evaluation

- Randomly select 1000 frames from the dataset

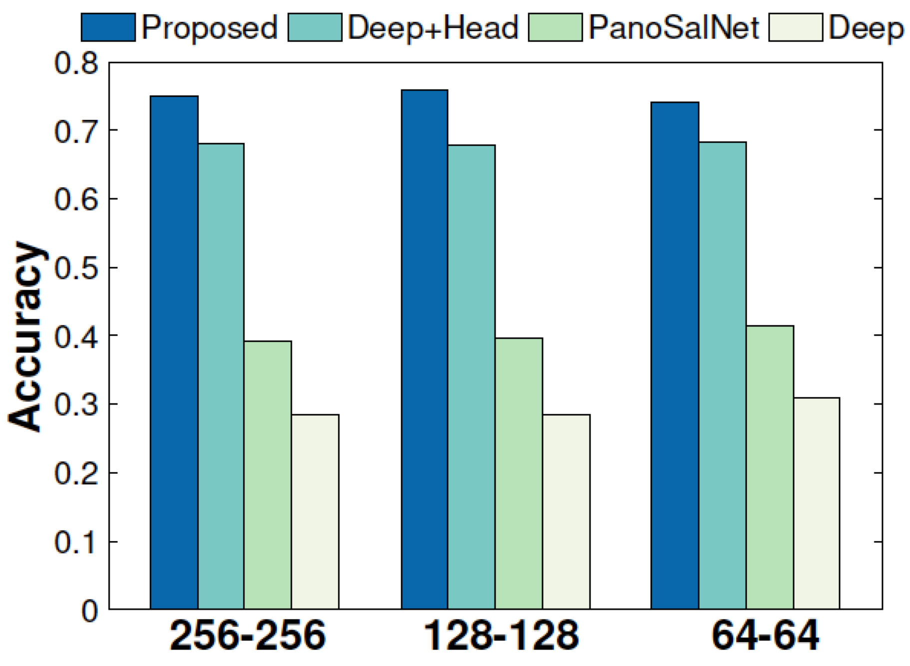
Model	sAUC	NSS	CC
Dataset Saliency	0.7966	2.4806	0.2885
Deep Convnet	0.6320	1.3256	0.1982
PanoSalNet	0.7112	1.9864	0.2521

- Deep Convnet suffers the problems of central bias and multi-object confusion

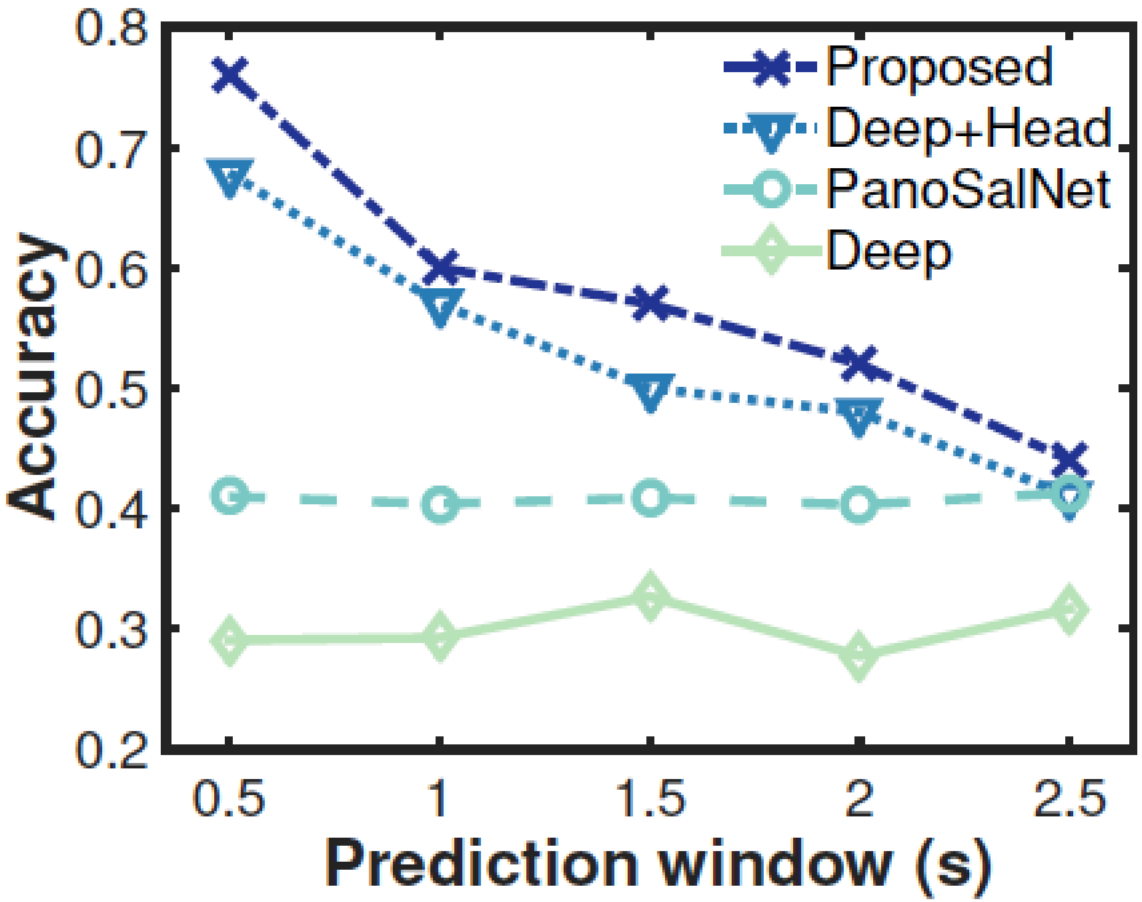


Head Movement Prediction Evaluation

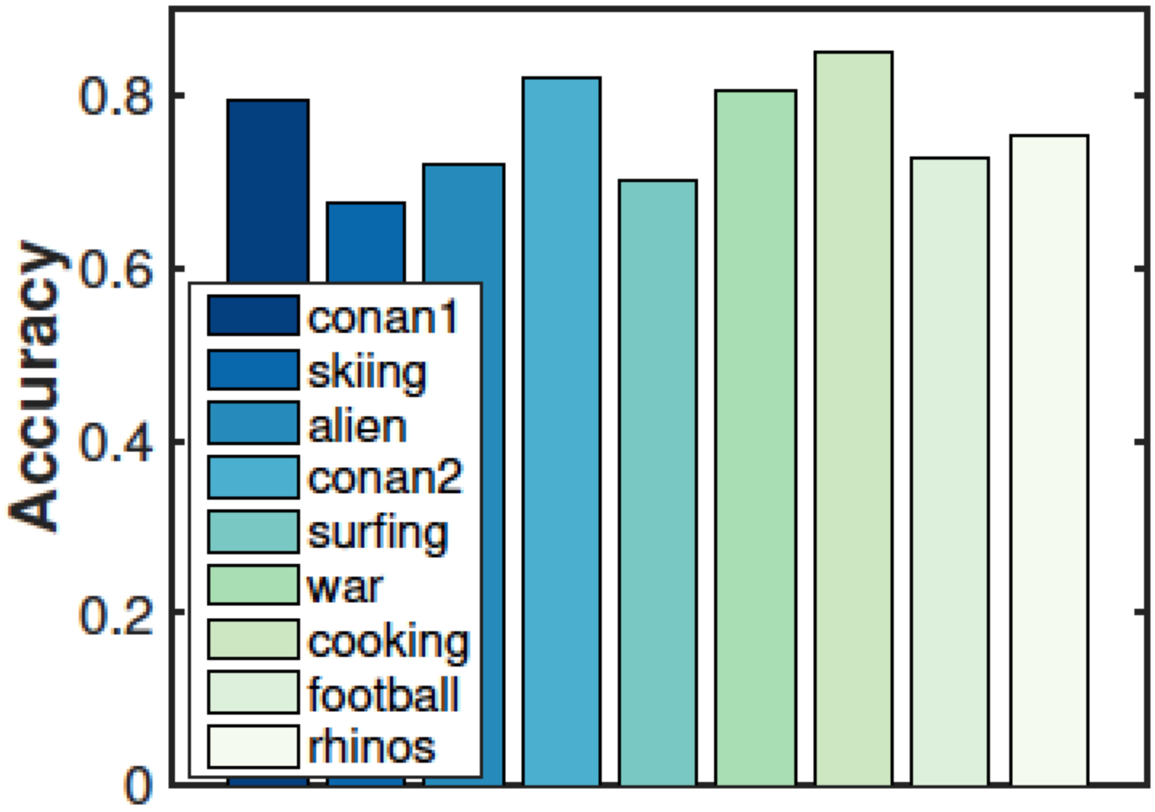
- the proposed model achieves an accuracy that is 1.9 times over PanoSalNet, 2.6 times over Deep, and 9% higher than Deep+Head
  - Saliency prediction for 360 video
  - head orientation and its temporal interplay with saliency is important
- Complicated and deeper model may not be needed



Impacts of Prediction Window



Impacts of Video Content



Discussion and Future Work

- Panoramic Saliency Dataset
- Content Trajectory Feature
- System Integration
  - stream a larger area or
  - image-based rendering can be exploited to compensate the missing tiles

Conclusion

- Unique panoramic saliency
- Head movement prediction framework
  - panoramic saliency
  - head orientation history