Deep Unsupervised Multi-View Detection of Video Game Stream Highlights

Charles Ringer & Mihalis A. Nicolaou

Foundations of Digital Games 2018

EPSRC
Engineering and Physical Sciences Research Council

Goldsmiths
UNIVERSITY OF LONDON

Why highlight detection?

Introduction to twitch.tv

Twitch.tv is a live-streaming platform used mostly for games

Three main stream categories:

- eSports/Mind Sports - e.g. LCS, Overwatch League, Magic: the Gathering
- **Individuals streaming game play sessions**
- IRL channels

Extremely popular:

- Average 1.15m viewers and 39.7k channels
- 43 billions minutes watched per month

Stats from https://twitchtracker.com

What’s in a stream?

THEY STEALING OUR DESERT !Giveaway 1080ti or Noblechair · …

PLAYERUNKNOWN'S BATTLEGROUNDS

46 0 Share

47:42 kimilim2000 : uhh that internet speed is BAD😒..... I thought Germany was well up there with good internet, in sorry to hear and Hope you can get fiber for a decent prize !

47:43 farmingskill : you should be able to get 100/100 fiber anywhere in developed europe lol

47:47 aliik6 : morning pawny and chat

47:48 Stormy1991 : @P4wnyhof can u Show ur Sound Settings

47:56 whatsluck : @flexioso going to try that. thank you

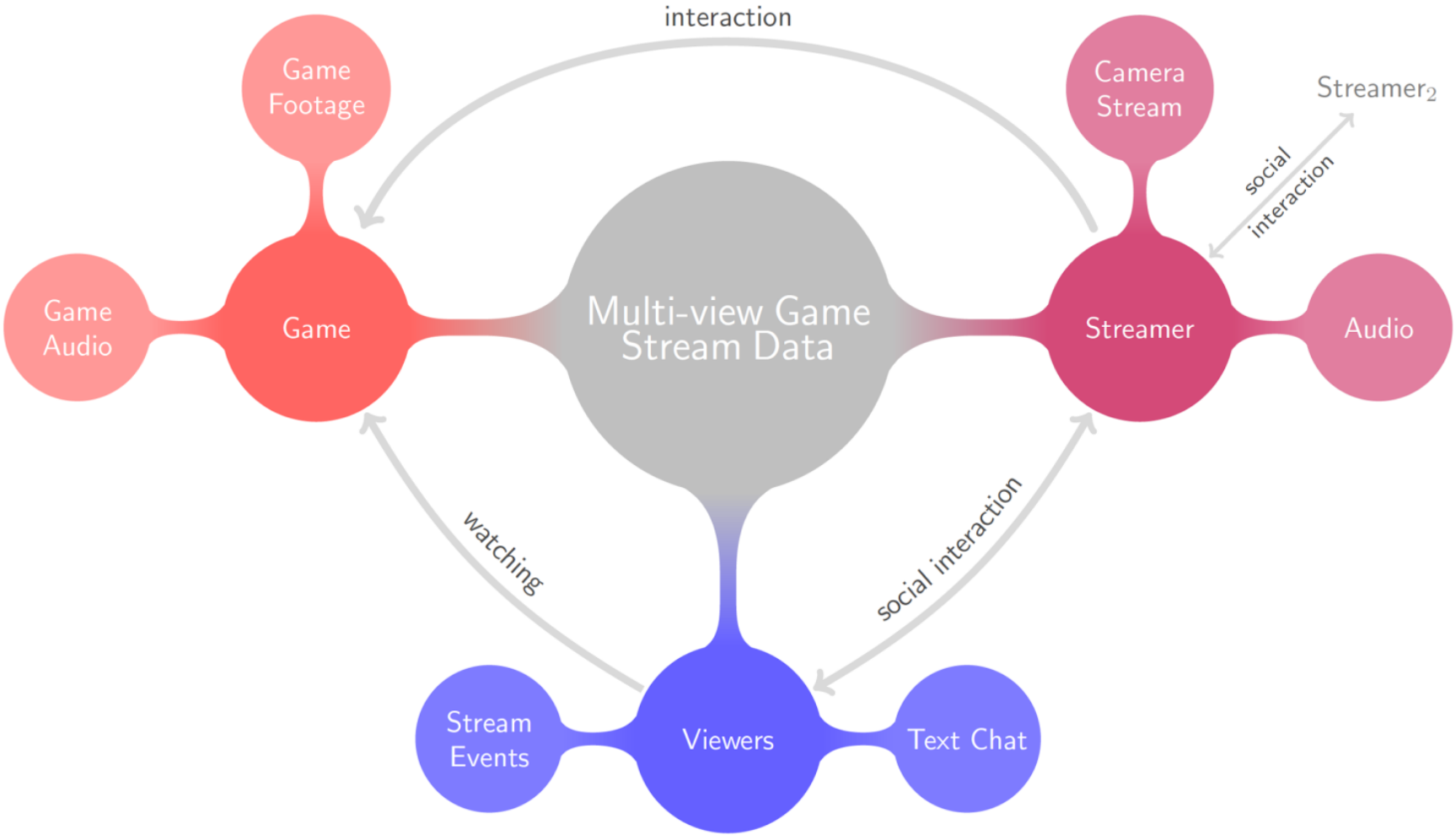
48:02 SentinelUK : !giveaway

48:02 P4wnyhof : @SentinelUK, Check our June Giveaway: http://bit.ly/2J80Cmu

48:03 daZimi : Moin @P4wnyhof built your House on Helgoland so you can defend it

48:05 Visa1223 : sex dungeon

48:08 loqoo_ : @tempeter10 well i'm from belgium and the best upload you can get is only 20 mbit/s


interaction
Game Footage
Camera Stream
Streamer$_2$
social interaction
Game Audio
Game
Multi-view Game Stream Data
Streamer
Audio
watching
social interaction
Stream Events
Viewers
Text Chat

What is a highlight?

"An outstanding part of an event or period of time"
- Oxford English Dictionary

What is a highlight?

"An outstanding part of an event or period of time"
- Oxford English Dictionary

Why are highlights interesting?

Advertisement

- Advertise the Game, Streamer, and Streaming Service

Entertainment

- Why watch a whole stream when you can just watch the highlights?

Already being created in different ways:

- Twitch clips - user made
- Nvidia Shadowplay Highlights - game data based

How can we model highlights?

How can we model highlights?

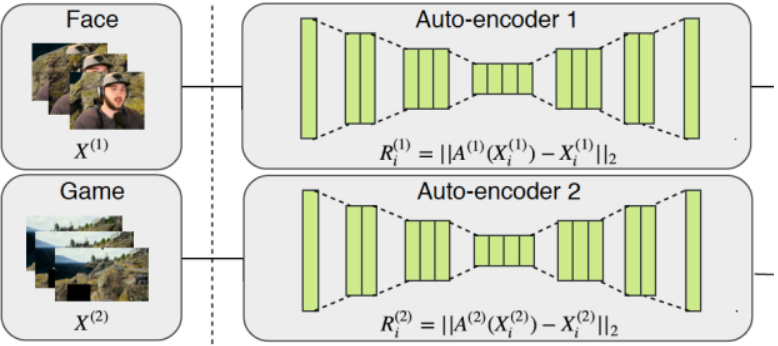
Problems:

- Detecting highlights is difficult as they are subjective and require context
- Computers are bad at context

However:

- Highlights are also novel moments in the stream
- Novelty can be considered a proxy
- Novelty is modelably!
- We can use multiple views: Face, Game & Audio

Face and Game: Convolutional Autoencoders



Frames sampled at 10fps

Images used to train convolutional autoencoders:

- **Encoder**: Takes an input and transforms it to a smaller latent space Z encoding
- **Decoder**: Takes this Z encoding and transforms it back to the input dimensions

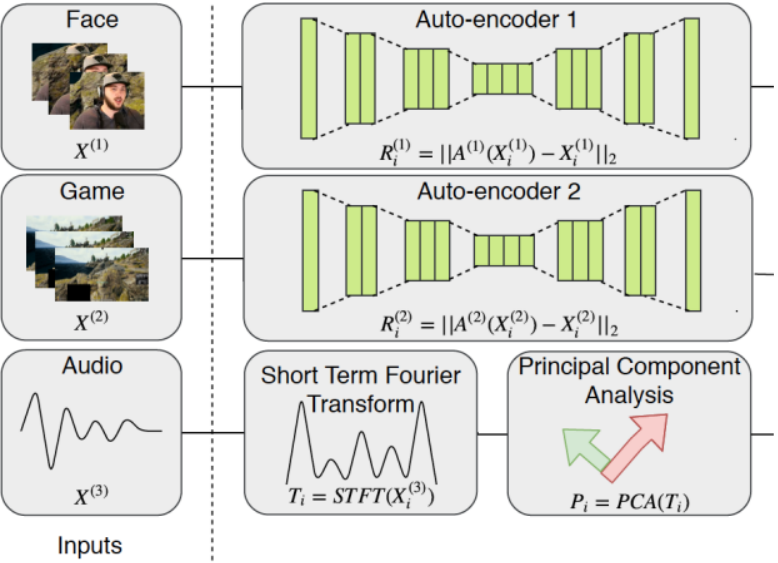
Neural Network Loss :

- Measured by Reconstruction Error (input-output)
- When trained Reconstruction Error is a measure of input novelty

Architecture:

- Two stacked VGG16 networks (without dense layers)
- Decoder has layer order flipped

Audio: Short-term Fourier Transform and Principal Component Analysis



Sliding window samples of 0.4 seconds generated

- Larger window to account for ”anticipatory coarticulation”

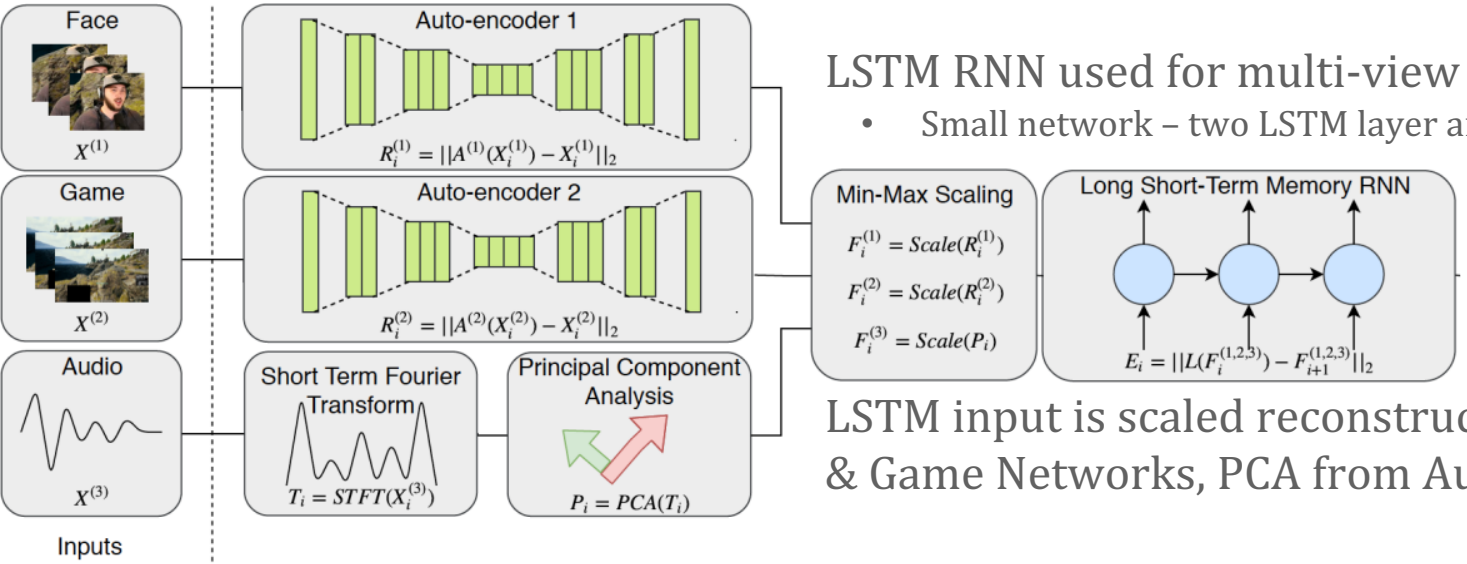
Audio features (frequencies) extracted with a STFT

- Also transformed from complex numbers to real numbers
- High and low frequencies discarded to isolate the streamers voice

Frequencies are then decomposed using PCA:

- Decomposed feature then used for fusion

Feature Scaling & Multi-View Late Fusion



Mix-Max scaling applied to all features

- Gives equal weighting to each view

LSTM RNN used for multi-view late fusion

- Small network – two LSTM layer and one dense layer

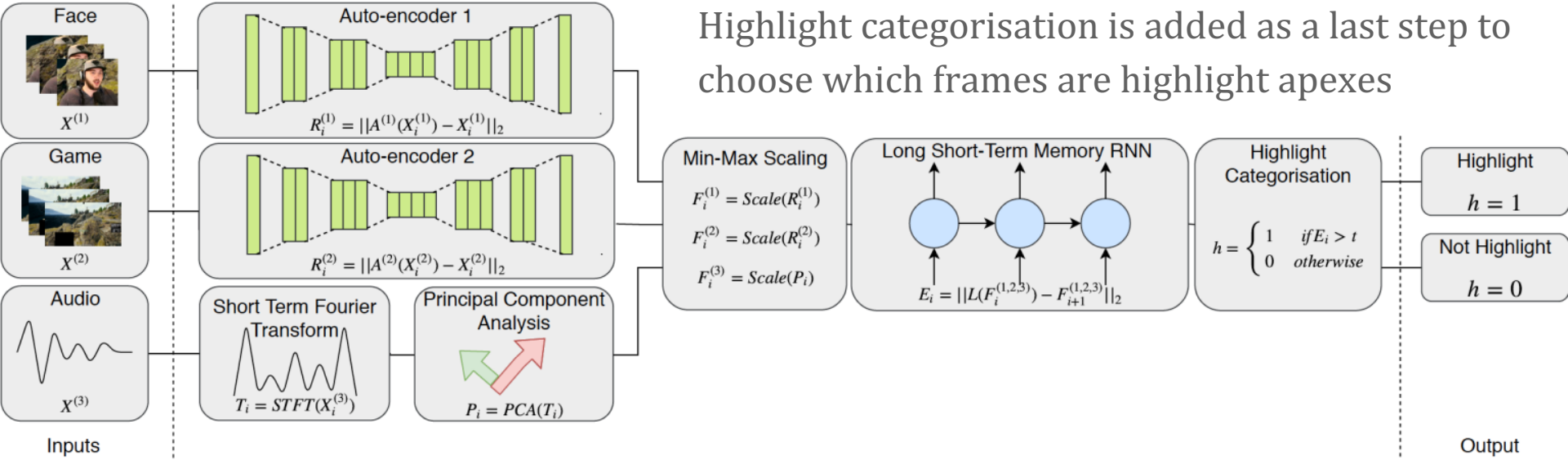
LSTM input is scaled reconstruction error from Face & Game Networks, PCA from Audio

LSTM trained to, at time *t*, predict the input at *t+1*

Complete Multi-view model for unsupervised highlight detection

LSTM error treated as final measure of novelty

Highlight categorisation is added as a last step to choose which frames are highlight apexes



Data

- 2 streamers, both male
- Playing *Player Unknown's Battleground*
- Streams were in English only
- Over 5 hours of footage
- 11 distinct games.

Highlight Generation

- Top 0.01% of frames selected a "Highlight Apexes"
- Each apex is linked to other temporally close Apexes
- Highlight is built using the preceding 10 seconds and succeeding 5 seconds for an Apex/Apexes

Results

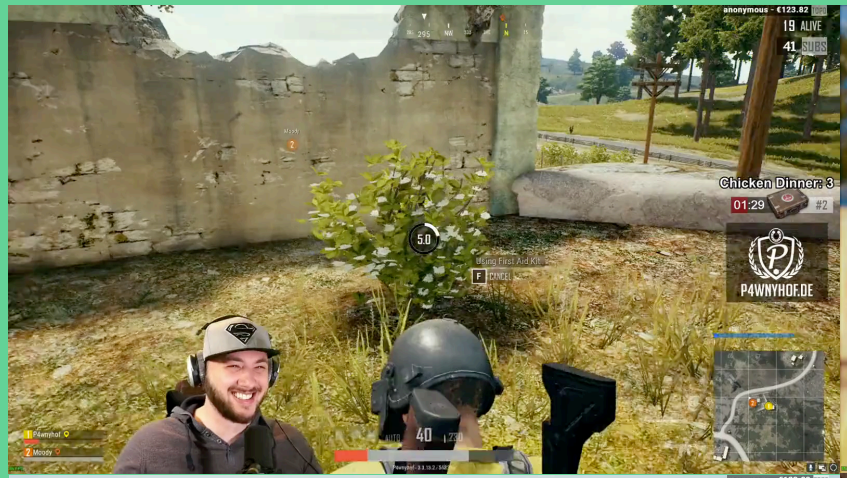
Categorisation

FUNNY

ACTION

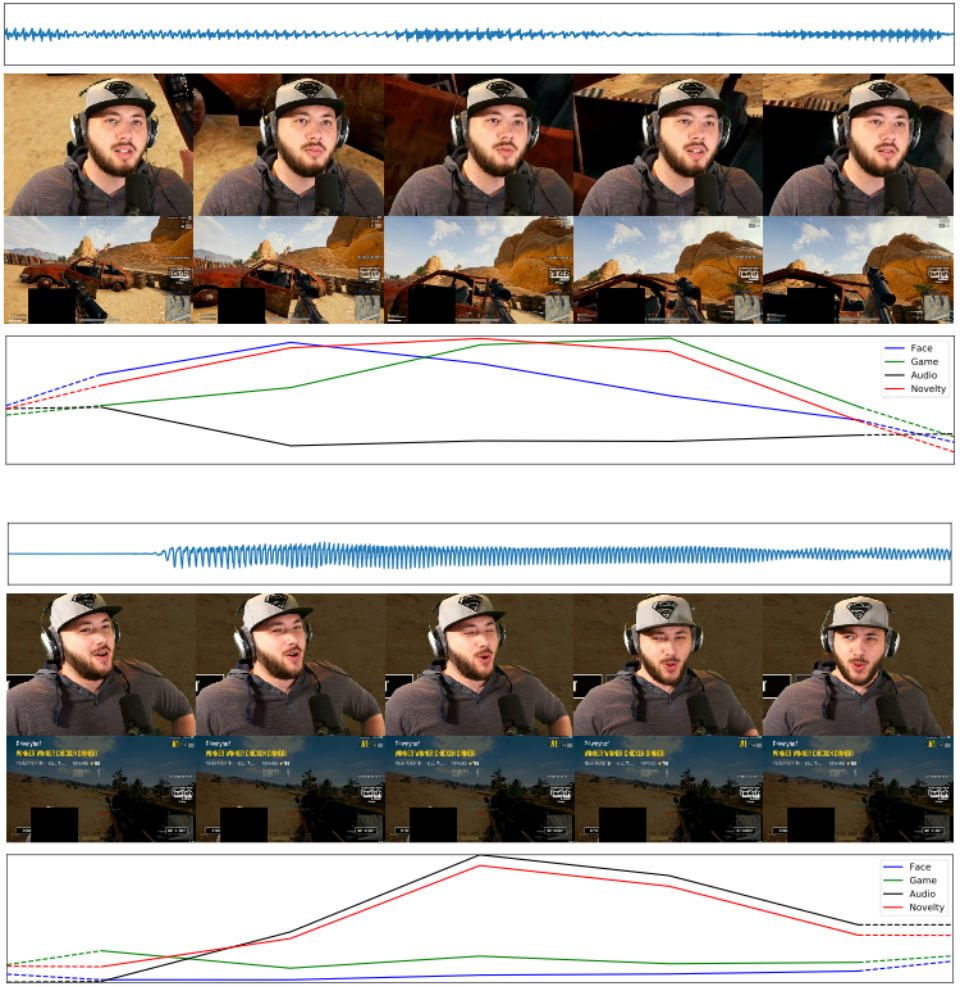
INTERACTION

NO HIGHLIGHT



Comparison of various combinations of views

	No. videos	Highlights	Highlights by category		
		Total	Funny	Action	Interaction
Face, Game & Audio	98	**77%**	**27%**	26%	24%
Face & Audio	95	74%	22%	23%	**28%**
Face Only	96	52%	14%	**14%**	24%
Audio Only	**126**	56%	8%	**29%**	18%
Game Only	94	**29%**	**4%**	18%	**4%**

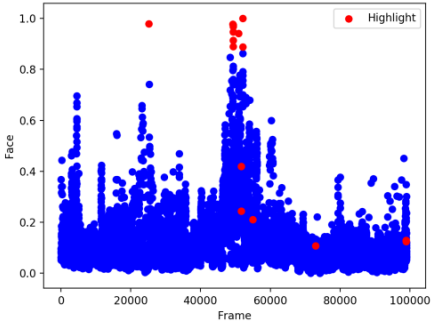


Example highlight apexes

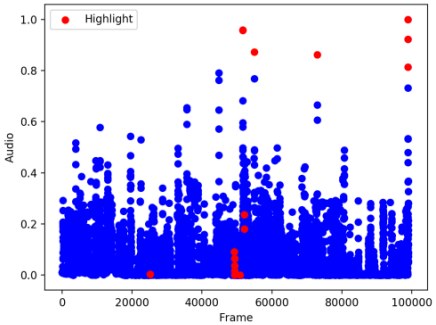
Top: Raw Audio
Upper Middle: Face
Lower Middle: Game
Bottom: Views and novelty

Comparison of views

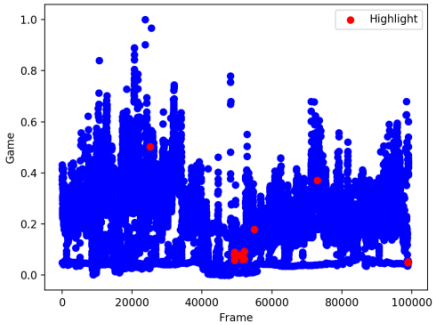
Face

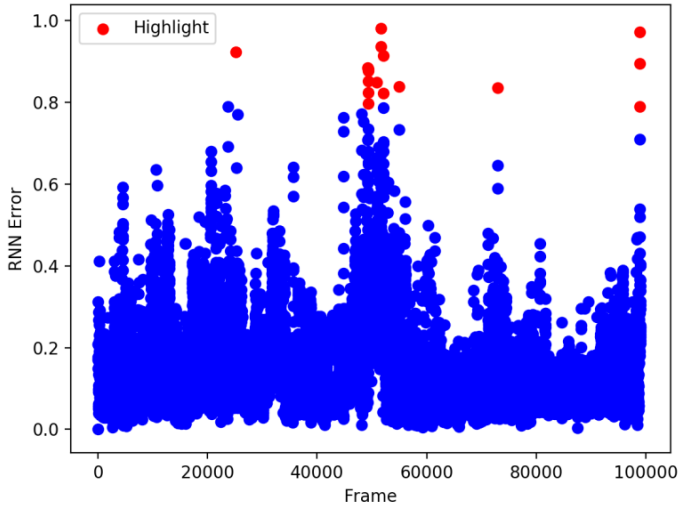


Audio



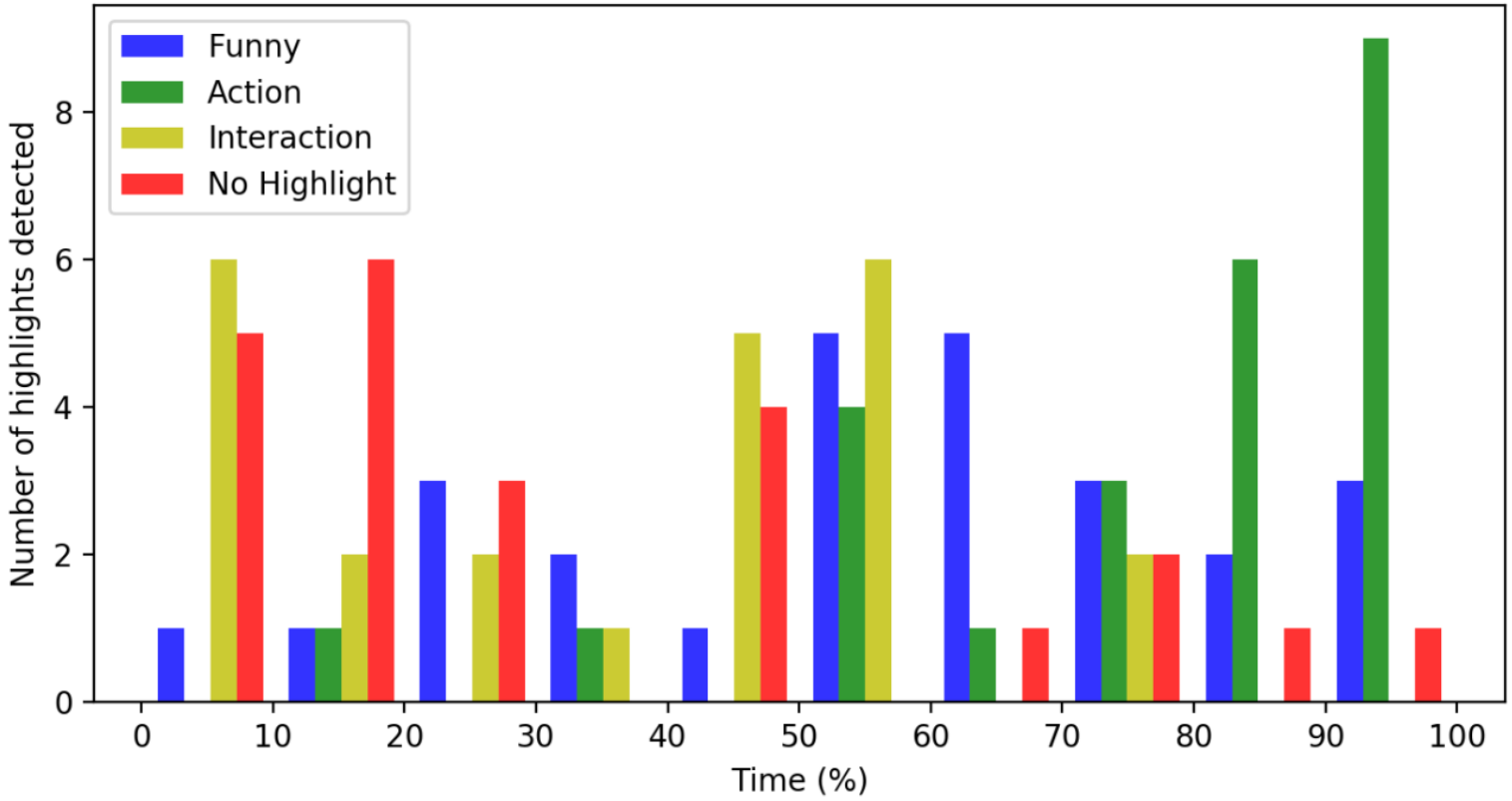
Game





Fused views

Highlight occurrence over time



Conclusions

Recap

- Novelty appears to be a reasonable proxy for highlights and the approach presented show promise
- Highlights fall into 3 broad categories: funny (streamer led), action (game led), and interaction (community led)
- Our current approach to game data does not work as well as face/audio because of a lack of baseline
- There are still a large number of incorrectly detected moments because context is not understood

Future work

A larger study is needed:

- More streamers - wider demographic
- More games - wider range of expected reactions

Can the approach be improved?:

- Game modelling is poor and contributes little, how can this be improved?
- Does a richer, multi-view latent space representation tell us anything?
- Can context be understood through chat messages?
- (Semi-)Supervised learning using an annotated database of streamers

Thank You

Contact: c.ringer@gold.ac.uk