

Finding the Fun: Game stream highlight detection using Deep Learning

Charlie Ringer

IGGI Conference 2018



Engineering and Physical Sciences
Research Council



i G G i

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

What's in a stream?



- 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



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?

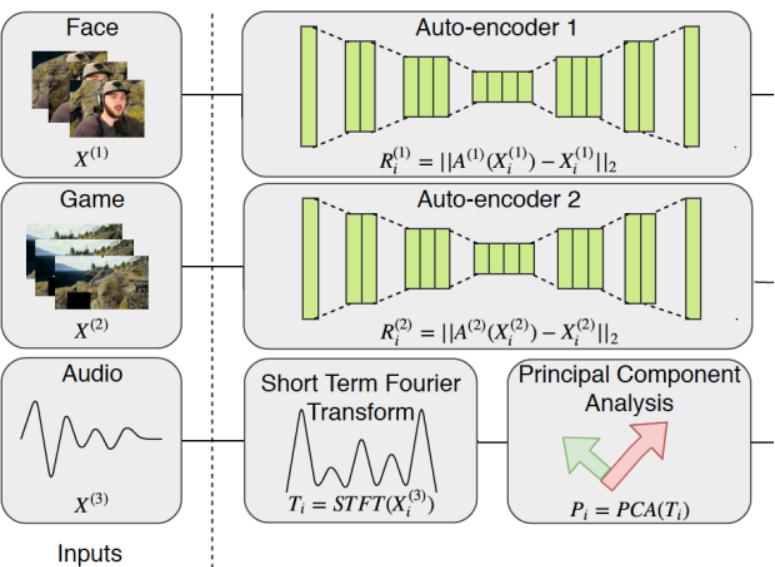
Problems:

- 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

Unsupervised Feature Extraction



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

Audio features (frequencies) extracted with a STFT

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

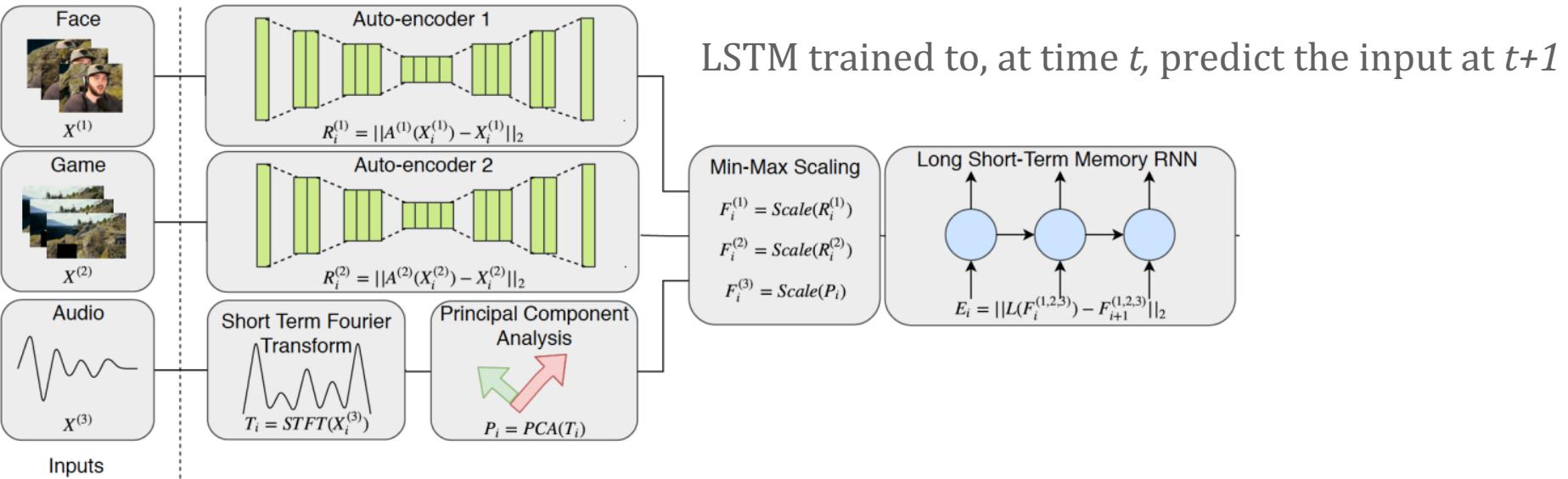
Frequencies are then decomposed using PCA:

- Decomposed feature then used for fusion

Feature Scaling & Multi-View Late Fusion

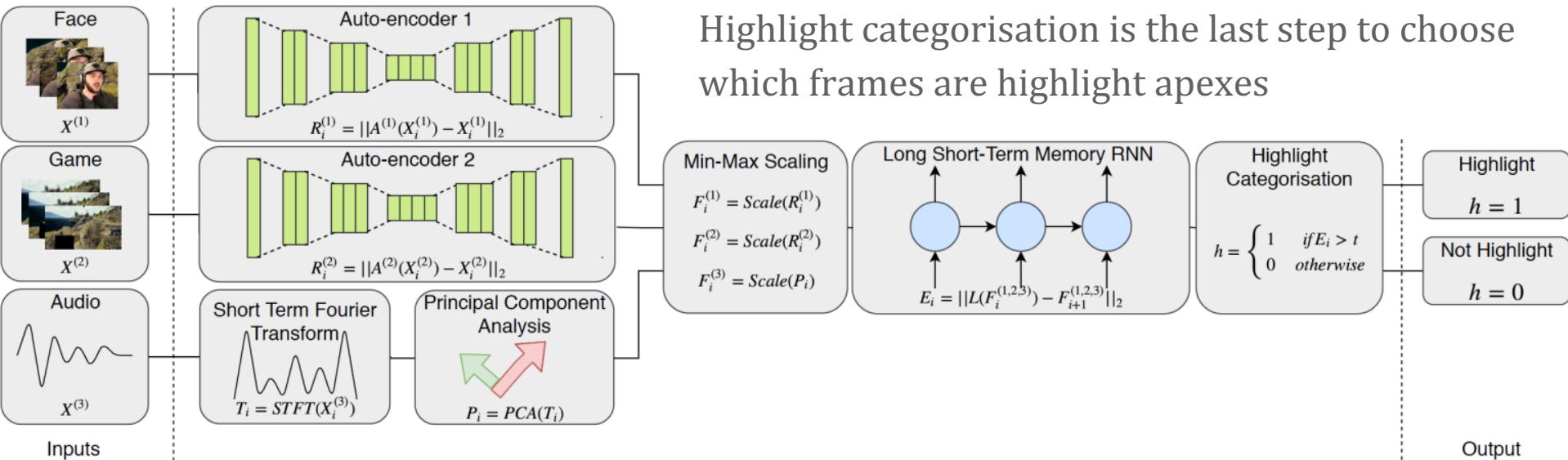
LSTM RNN used for multi-view late fusion

- Small network – two LSTM layer and one dense layer



Complete Multi-view model for unsupervised highlight detection

LSTM error treated as final measure of novelty



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

- Most novel 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
-

What highlights did we find?

Categorisation

FUNNY



INTERACTION



ACTION



NO HIGHLIGHT

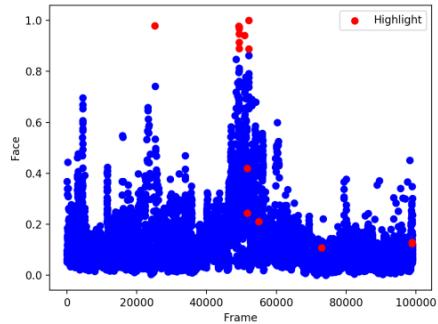


Comparison of various combinations of views

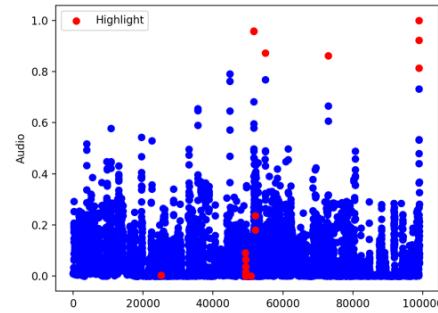
	No. videos	Highlights	Highlights by category		
		Total	Funny	Action	Interaction
Face, Game & Audio	98	77%	27%	26%	24%
Face & Audio	95	74%	22%	23%	28%
Face Only	96	52%	14%	14%	24%
Audio Only	126	56%	8%	29%	18%
Game Only	94	29%	4%	18%	4%

Comparison of views

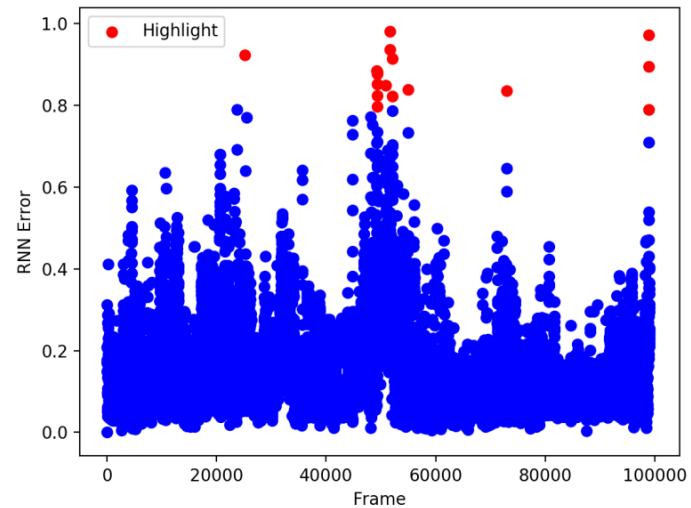
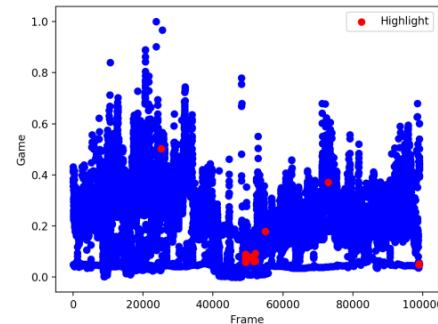
Face



Audio

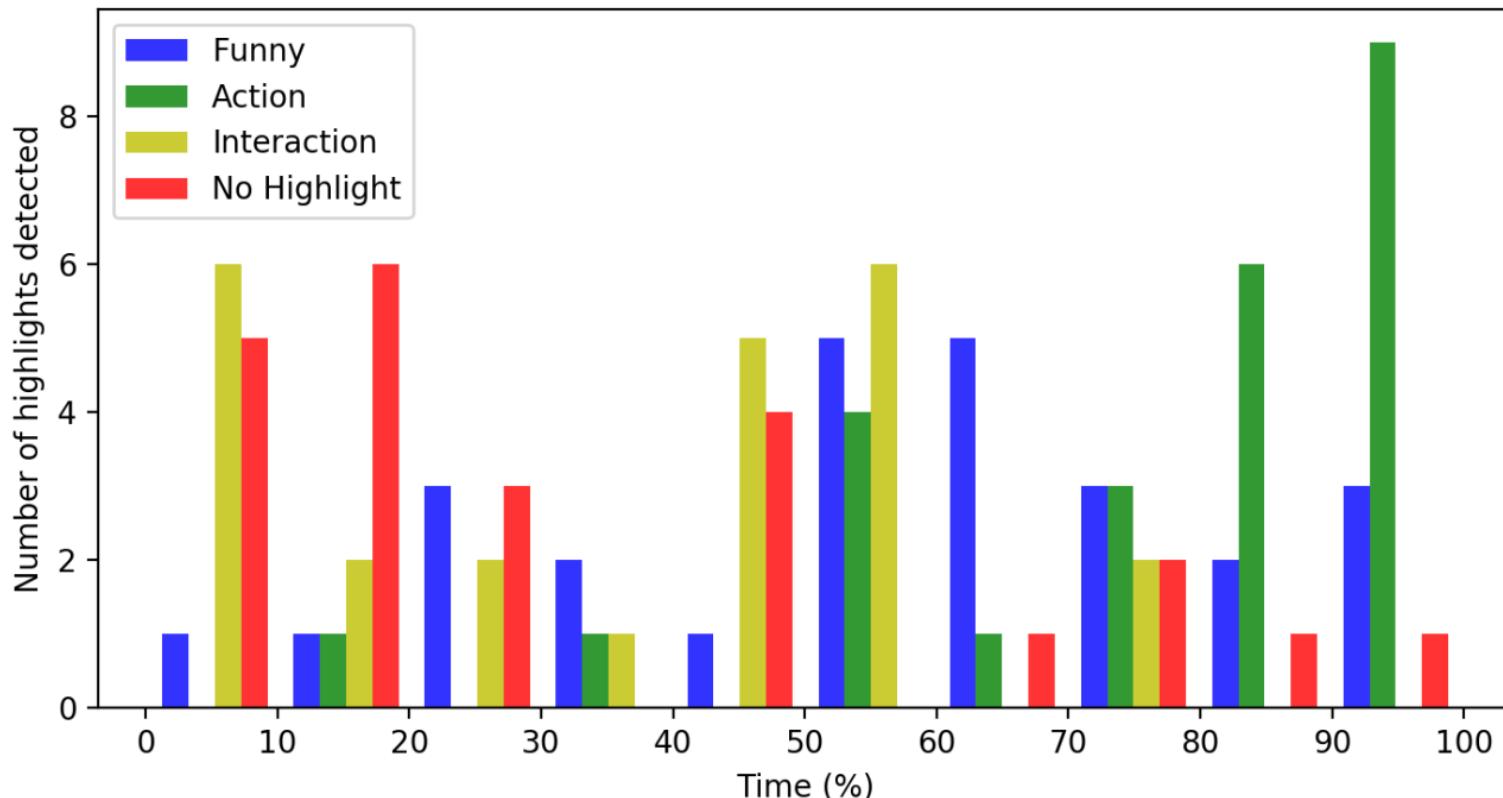


Game



Fused views

Highlight occurrence over time



Where is this research going?

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

We Need More Data

A larger study is needed:

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

Have gathered some data already:

- 5 games (Hearthstone, League of Legends, PUBG, No Man's Sky, Slay the Spire)
- 19 different streamers
- Estimated 40-80 hours of footage

We also aim to annotate this data (a big task!)

- Emotional/Behavioural streamer annotations
- Salient game play moment annotations

We Need A Better Model

- Can computational models of affect help detect highlights?
- Does a richer, multi-view latent space representation tell us anything?
- Can context be understood through chat messages?
- How can we make our model generalisable across streamers, game and genre?
- How can computer vision techniques help us to understand what is going on in a game scene?

Last Thing

I am looking for a placement

Thank You

c.ringer@gold.ac.uk
@chardr27
charlieringer.com