Combining audio and visual features to play GeometryDash

P. Bakker, J. R. T. E. van der Hout, J. N. de Vries, R. van der Wal, and M. M. van der Wel

Introduction

Many deep learning networks have been trained to teach game agents how to play games. These networks are usually trained using images. It has been shown that game immersion improves for players when using audio [3]. A game agent might also benefit from using audio to learn how to play the game.
This project aims to investigate if an audio network can be trained that achieves a high AUC and accuracy, and if combining image and audio networks will improve the AUC and accuracy.
This will be tested using the game Geometry Dash, which has been previously played using an image network by Li and Rafferty [1].

Research

Research question:

Will adding audio to an image network increase accuracy and AUC?

Hypothesis:

We believe combining an audio and image network will increase accuracy by a maximum of 5%.

Experimental setup

Networks trained using imitation learning [2]. Audio and image datasets acquired by recording gameplay, automatically removing bad `death frames' . Raw audio network standard architecture [4]. Image network based on network by Li and Rafferty [1].

Image network:
- Dropout 0.25 on all layers
- 1000 epochs
- Train:validation ratio is 9:1
- Test created separately
- RMSprop
- Learning rate 0.05

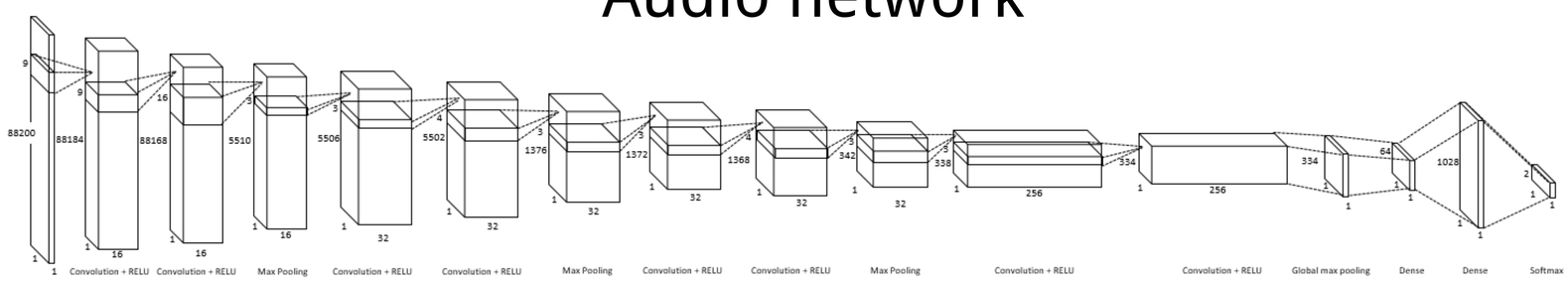
Audio network:
- Dropout 0.1 on all but last block (0.2)
- 10 epochs
- Train:validation 9:1
- Test created separately
- Adam
- Learning rate 0.001

Image data:
- 15 000 images
- Downscaled to 60x80
- Greyscale
- Prediction every 0.1s

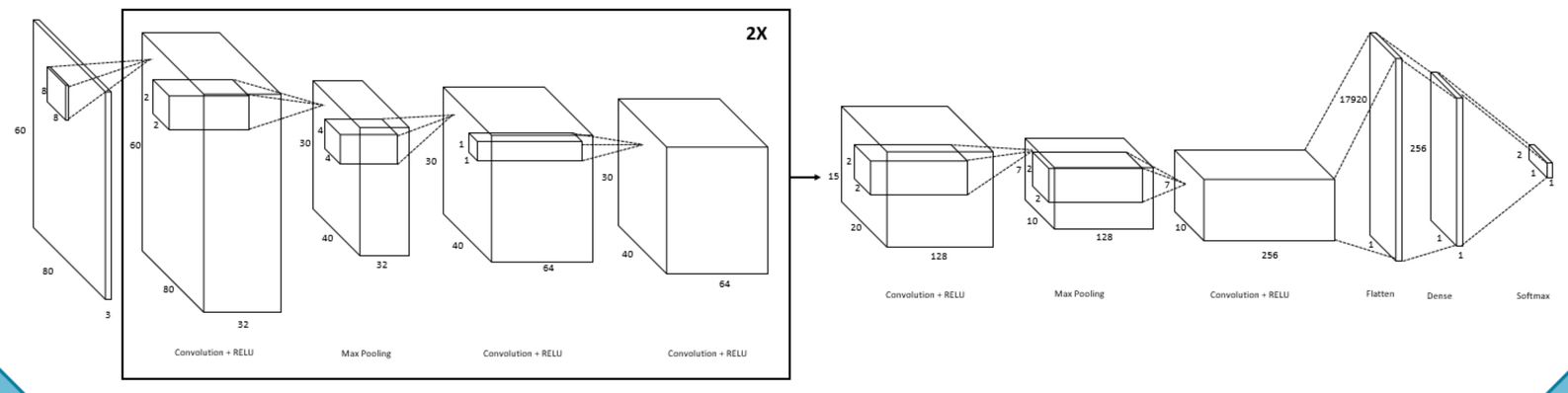
Audio data:
- 1500 audio samples of 2 seconds
- 2 second sliding-window
- Raw audio
- Prediction every 0.1 s

Networks

Audio network

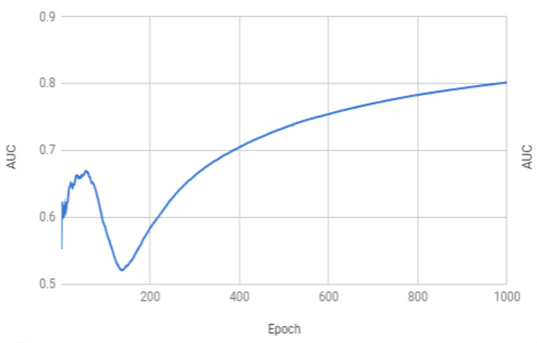


Visual network



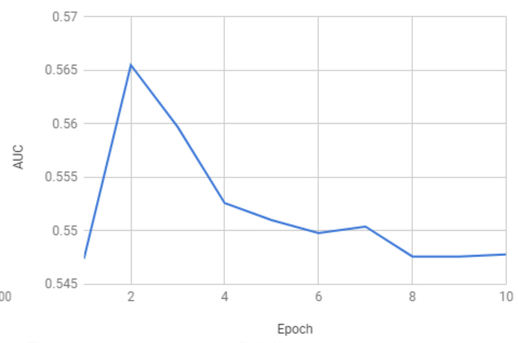
Experimental results

Visual network - Validation AUC over time



The classification accuracy plotted for all the epochs on the training data. Notice the dip around 100 epochs. This is most likely caused by the drop-out.

Audio network - Validation AUC over time



The area under the ROC curve plotted for all the epochs on the validation data. There is one deviation at the second epoch, but it is very small. We were unable to do more epochs due to time constraints.

Final results table

Network	Train	Validation	Test
	Accuracy		
Image	0.80	0.80	0.63
Audio	0.62	0.55	0.50
Both	-	-	0.78

Accuracy during trainig, validation and testing. For testing: the image network used 3000 images, the audio network used 1000 files and for the combined score 1850 instances were used. Test accuracy of the combined network is high due to the combined network classifying everything as no jump and having more no jump frames.

Discussion

- There is a desync between recording and action which impacts the quality of the training data
- Audio network could improve by replacing raw audio with MFCC/Beat detection
- Real time play requires small networks
- High validation AUC does not lead to good in-game performance
- The networks are currently combined by averaging, this could be improved by training a classifier or network

Conclusion

The audio network developed in this project does not show good results, as it tends to classify according to the prior of the training data. Likely factors that cause this tendency are a lack of training data, inaccurate training data and time delays. Because of this the combined network does not benefit from the addition of audio, but further improvements to both networks can still be made. Further refinement of both the audio and image network needs to be done to be able to conclude if adding audio can prove beneficial.

References

[1] Li, T., & Rafferty, S. (2017). Playing Geometry Dash with Convolutional Neural Networks (p. 7). Stanford University
[2] Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. In Advances in Neural Information Processing Systems
[3] Collins, K. (2013). Playing with sound: a theory of interacting with sound and music in video games. Mit Press.
[4]Beginner's Guide to Audio Data | Kaggle. Retrieved from https://www.kaggle.com/fizzbuzz/beginner-s-guide-to-audio-data