

Modelling Complex Perception-Action Choices

Ruohan Zhang¹, Jake A Whritner¹, Zhuode Liu², Luxin Zhang³,
Karl S. Muller¹, Mary M. Hayhoe¹, Dana H. Ballard¹

¹The University of Texas at Austin

²Google Inc.

³Peking University

July 13, 2018

Research Problem

- Can we build a general model to predict human gaze and action behaviors in Atari video games?¹²
- Video games are more restrictive than natural tasks, but still contain rich amount of visual features and have different dynamics.



Screenshots of 8 Atari games.

¹Thanks to Geisler Lab@UT-Austin with data collection

²Bellelmare et al, The Arcade Learning Environment: An evaluation platform for general agents. JAIR 2013

1 Gaze Modeling: A Data-Driven Approach

2 Action Modeling: Gaze-Conditioned Prediction

Gaze Prediction: Traditional Models

- Bottom-up: Saliency models are driven by image features, such as the contrast of color, intensity, and orientation³

³Itti and Koch, Computational modelling of visual attention. Nature reviews neuroscience 2001

⁴Johnson et al, Predicting human visuomotor behaviour in a driving task. Phil. Trans. R. Soc. 2014

Gaze Prediction: Traditional Models

- Bottom-up: Saliency models are driven by image features, such as the contrast of color, intensity, and orientation³
- Top-down: Task features such as reward and uncertainty⁴

³Itti and Koch, Computational modelling of visual attention. Nature reviews neuroscience 2001

⁴Johnson et al, Predicting human visuomotor behaviour in a driving task. Phil. Trans. R. Soc. 2014

Gaze Prediction: Another Perspective

- Neural networks are data-driven and require minimum effort in searching for and crafting these features.

⁵Bylinskii et al. Where should saliency models look next?. ECCV 2016.

Gaze Prediction: Another Perspective

- Neural networks are data-driven and require minimum effort in searching for and crafting these features.
- Gaze data from bottom-up research: Subjects view **static** images in a **free-viewing** manner

⁵Bylinskii et al. Where should saliency models look next?. ECCV 2016.

Gaze Prediction: Another Perspective

- Neural networks are data-driven and require minimum effort in searching for and crafting these features.
- Gaze data from bottom-up research: Subjects view **static** images in a **free-viewing** manner
- Gaze data from top-down research: Subjects perform a particular **task** in an **interactive** environment

⁵Bylinskii et al. Where should saliency models look next?. ECCV 2016.

Gaze Prediction: Another Perspective

- Neural networks are data-driven and require minimum effort in searching for and crafting these features.
- Gaze data from bottom-up research: Subjects view **static** images in a **free-viewing** manner
- Gaze data from top-down research: Subjects perform a particular **task** in an **interactive** environment
- Deep neural networks are shown to be very successful modeling the first type of data⁵

⁵Bylinskii et al. Where should saliency models look next?. ECCV 2016.

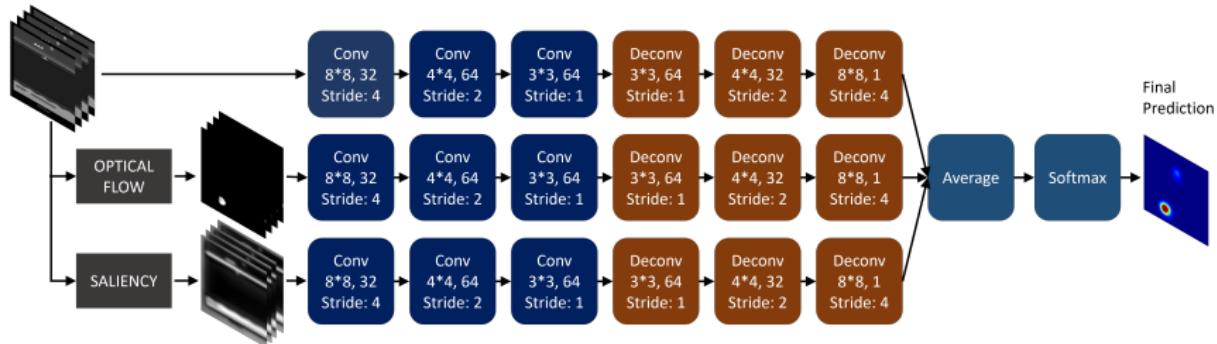
Gaze Prediction: Another Perspective

- Neural networks are data-driven and require minimum effort in searching for and crafting these features.
- Gaze data from bottom-up research: Subjects view **static** images in a **free-viewing** manner
- Gaze data from top-down research: Subjects perform a particular **task** in an **interactive** environment
- Deep neural networks are shown to be very successful modeling the first type of data⁵
- Can they be used to model task-driven gaze data, as in Atari games?

⁵Bylinskii et al. Where should saliency models look next?. ECCV 2016.

The Gaze Prediction Network

- Inputs:
 - 4 consecutive images
 - Motion: Optical flow between 2 frames⁶
 - Image saliency: Itti-Koch model⁷
- Output: Predicted probability distribution of gaze
- For each game, trained on 90-minute gaze data and tested on 30-minute data via standard supervised learning.



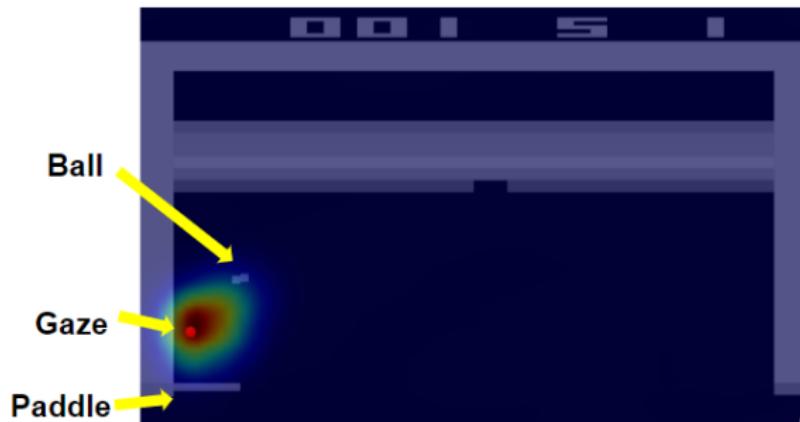
⁶Farnebck, Two-frame motion estimation based on polynomial expansion. Image analysis 2003

⁷Itti and Koch, Computational modelling of visual attention. Nature reviews neuroscience 2001.

Gaze Prediction Results I: Game Breakout (AUC = 0.970)

Gaze Prediction Results I: Game Breakout

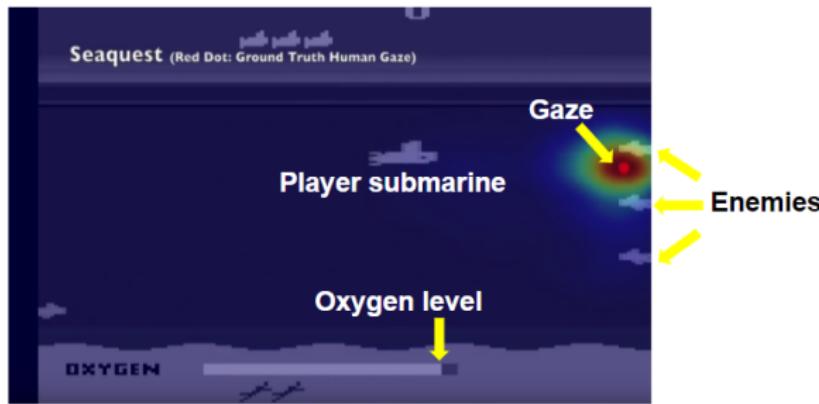
- The model predicts gaze behaviors that are difficult to capture using hand-defined features, e.g., predictive eye movements



Gaze Prediction Results II: Game Seaquest (AUC = 0.964)

Gaze Prediction Results II: Game Seaquest

- The model can tell visually identical task-relevant object for the moment from others.



Gaze Prediction: Quantitative Results

- Average area under the receiver operating characteristic curve (AUC) results across 8 games on testing dataset

	Random	Saliency(S)	Motion(M)	Image(I)	I+S	I+M	I+S+M
AUC↑	0.500	0.509	0.707	0.963	0.962	0.965	0.965

1 Gaze Modeling: A Data-Driven Approach

2 Action Modeling: Gaze-Conditioned Prediction

Action Prediction: Motivation

- Predict a human player's action given a game frame



Action Prediction: Motivation

- Predict a human player's action given a game frame (with gaze)



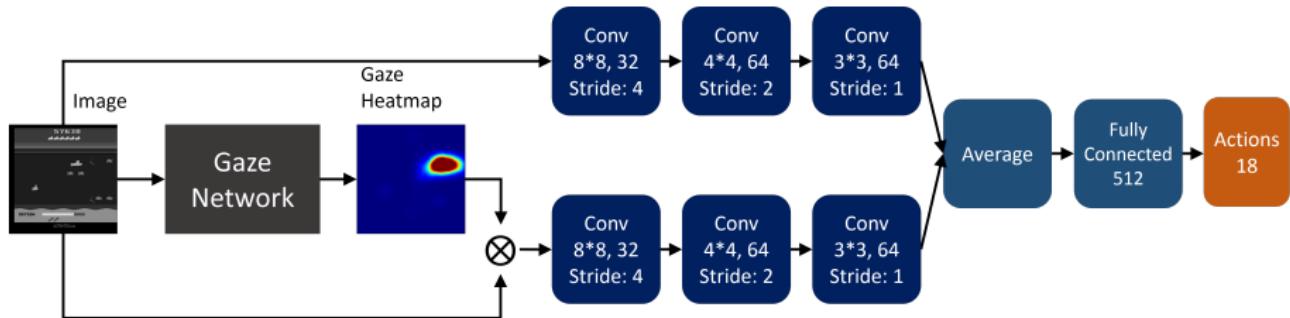
Action Prediction: Motivation

- Predict a human player's action given a game frame (with gaze)



The Action Network

- Including predicted gaze information into action prediction⁸
- Mask the image with gaze heatmap to highlight the object(s) being fixated at.



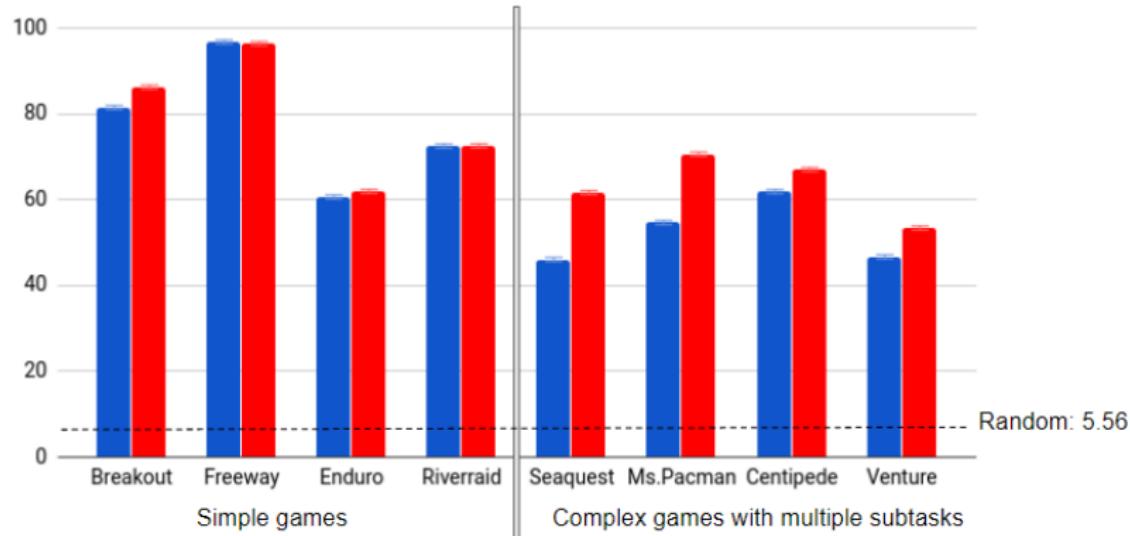
⁸Ruohan Zhang, Zhuode Liu, Mary Hayhoe, and Dana Ballard, Attention Guided Deep Imitation Learning. Cognitive Computational Neuroscience, 2017

Performance: Action Prediction Accuracy

- Including gaze information improves action prediction accuracy

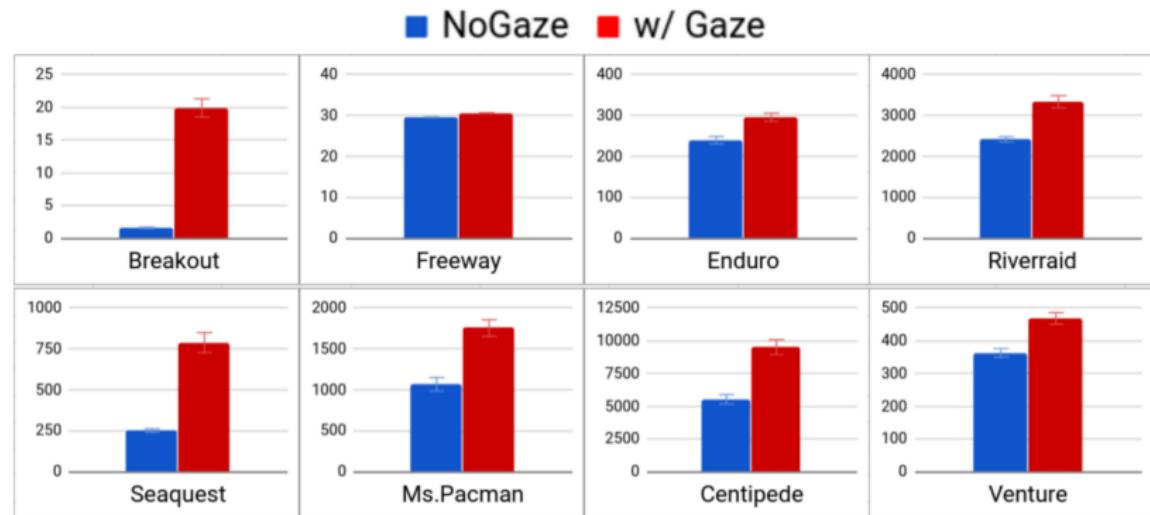
Action Prediction Accuracy (%)

■ NoGaze ■ w/ Gaze



Performance: Game Scores

- Including gaze information leads to better performance



Conclusions

- Multi-channel convolutional neural networks can be trained to accurately predict human gaze in video games.
- Actions are conditioned on gaze hence incorporating gaze information into action modeling improves the performance.
- Being able to jointly model human gaze and action can lead to exciting new models and algorithms in AI research.

Thank You