Understanding Dog Behavior through Visual and Aural Sensing Using Deep Learning

Amy Lin
Princeton High School
Princeton, NJ
March 2023

Research Motivation

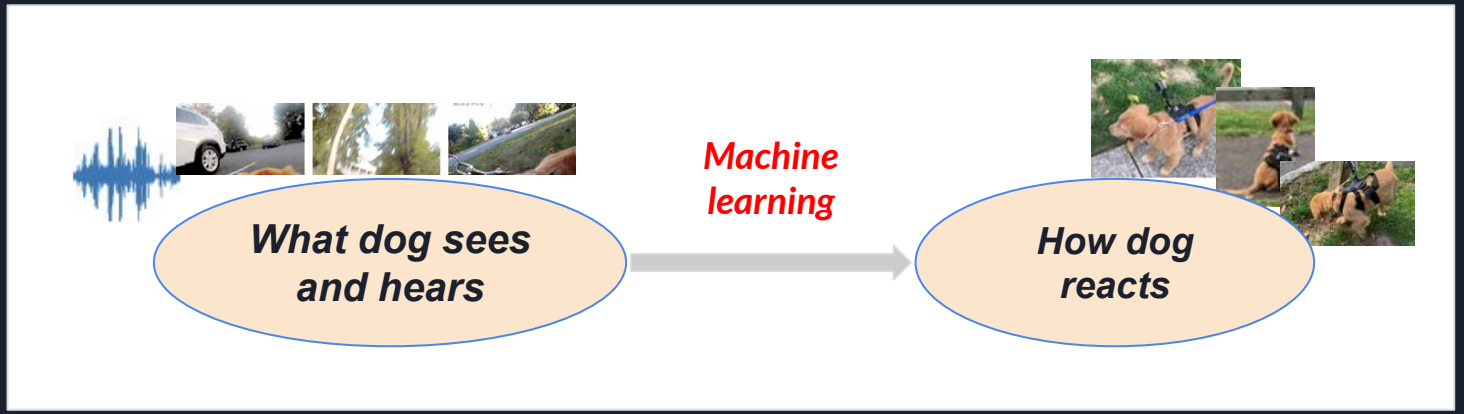
- Artificial vs. natural intelligence
  - AI's dependency on large amounts of data,
  - AI tends to make mistakes trivial to humans
- What can we learn from animal intelligence?
- How do dogs behave and respond to their environments?
- New idea: *Use ML to Model Dog Behavior through Visual and Aural Sensing*

Objectives

- Understand dog behavior and reaction to different environmental stimuli using machine learning
  - Visual stimuli
  - Auditory stimuli
  - Stimuli perceived from dog's egocentric perspective

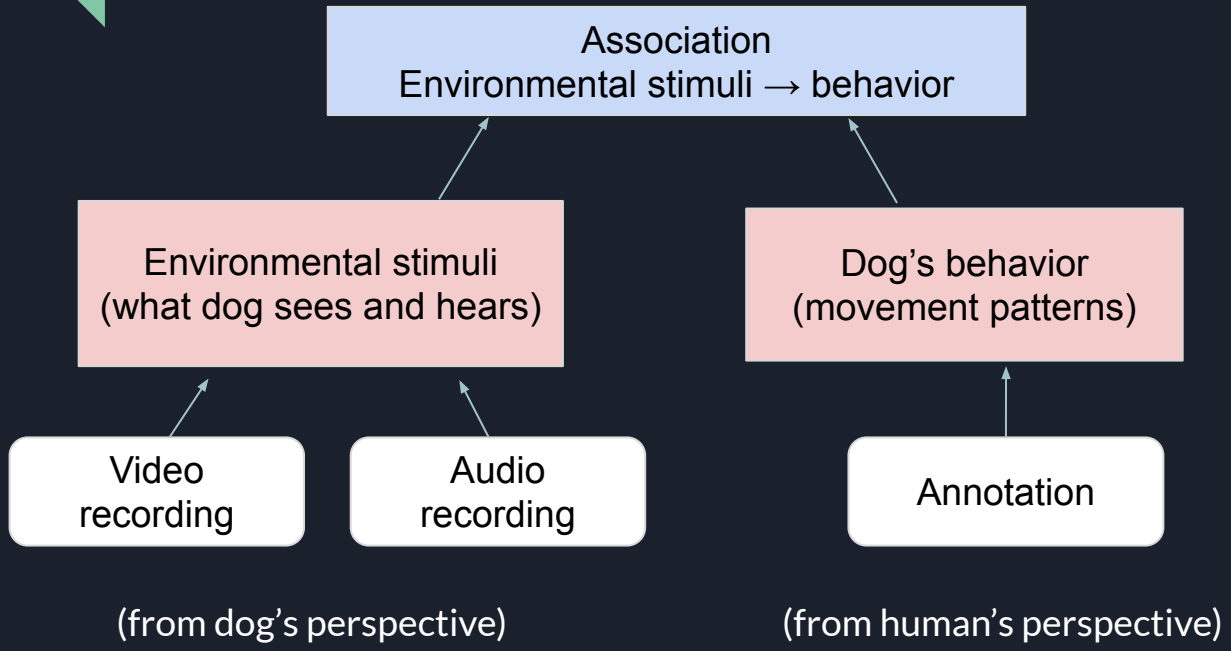
- Potential applications
  - Help develop new AI technologies (e.g. robot dog)
  - Create new ways of working with dogs: training environment customized to dog's natural reaction
  - Dog training: service dogs, military dogs, police dogs, rescue dogs, companion dogs
  - Understanding animal intelligence provides insights into human intelligence

Proposed Methodology



Learn and model the ***association*** between dog's perceived ***visual and audio stimuli*** and dog's ***reaction***

Proposed Methodology



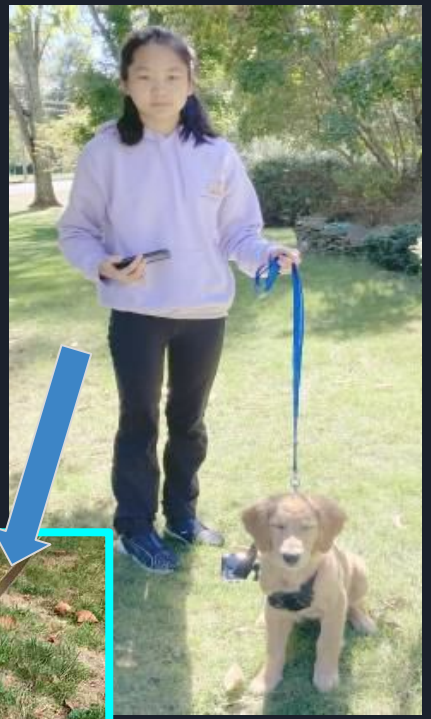
Main steps:

- *Sensing*
- *Computer vision*
- *Signal processing*
- *Machine learning*

What's new:

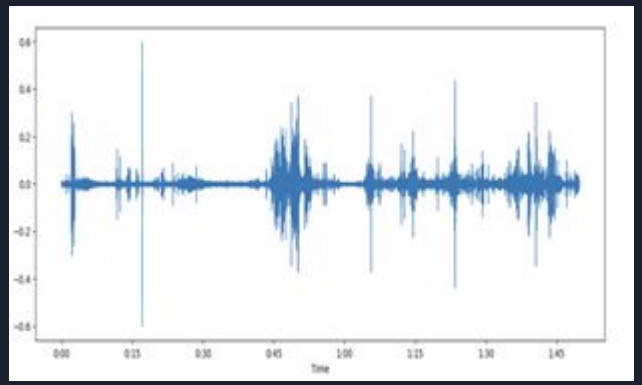
- *Multi-modal stimuli*
- *Sensing from dog's egocentric perspective*

Proposed Methodology: *Data collection*



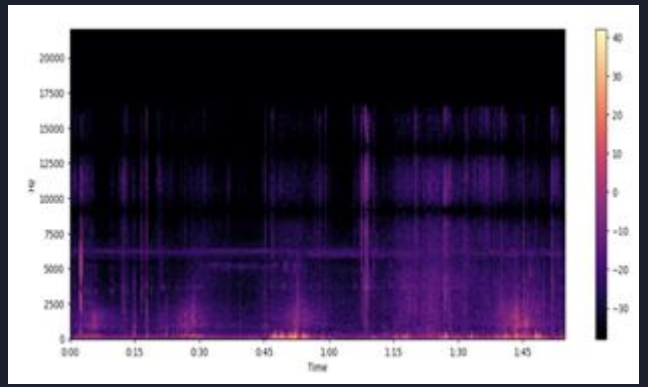
Proposed Methodology: *Audio Signal Analysis*

Audio signal: Short-Term Fourier Transform (STFT)



(Audio signal)

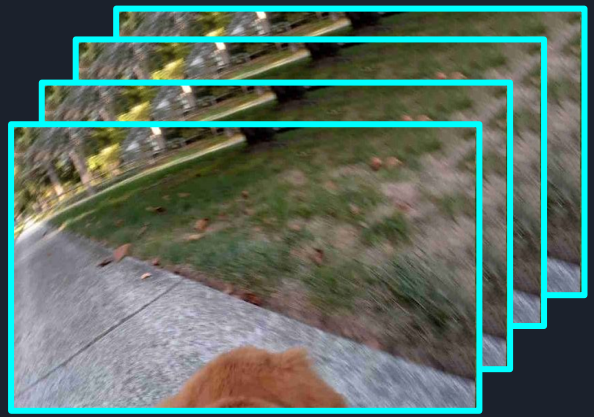
STFT



(STFT spectrogram)

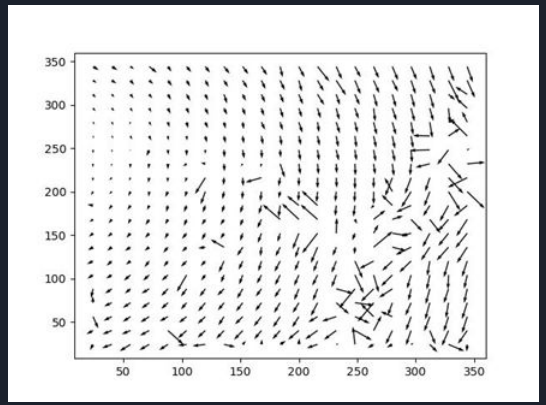
Proposed Methodology: *Image Motion Analysis*

Image motion: motion estimation by template matching



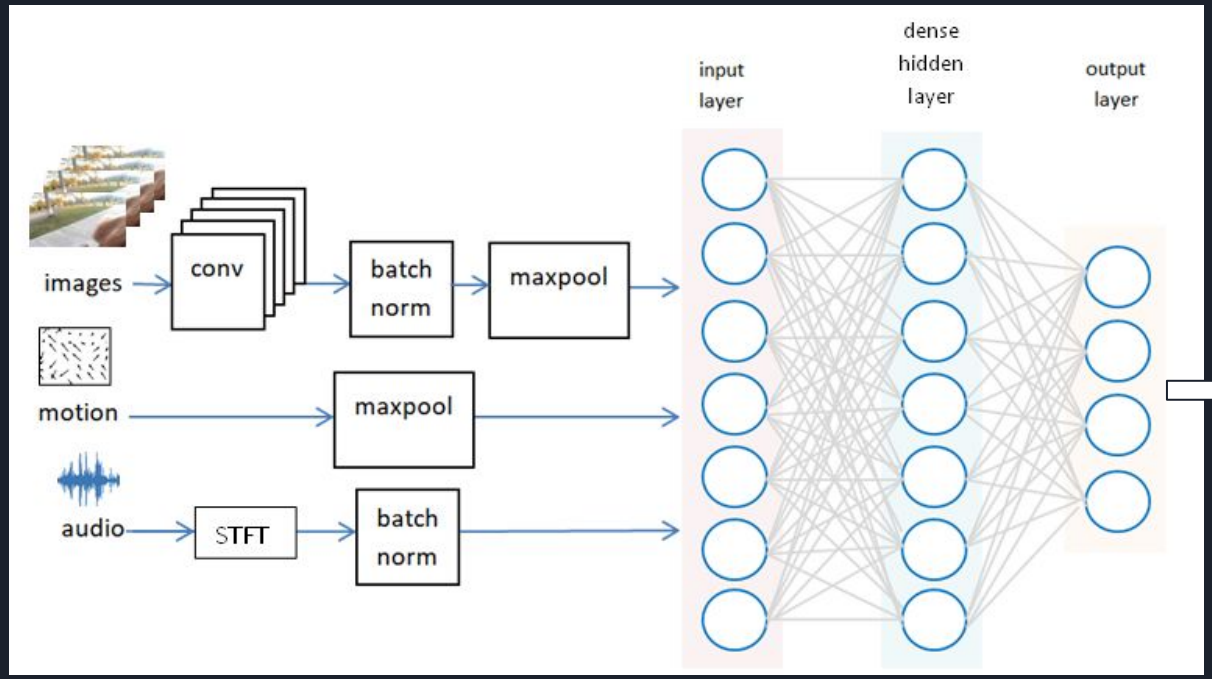
(image frame)

Motion estimation



(motion field)

Proposed Methodology: *eCNN model*



Multi-modal input

(extended Convolutional Neural Network model)

Experiments: *Setup*

- Data split into training set (70%), validation set (10%), testing set (20%)
- eCNN model structure:
  - Image input: 32 convolution filters of size 7x7 → batch normalization over color channels → max pooling with pool size 3x3
  - Audio input: STFT → batch normalization
  - Motion input: max pooling with pool size 9x3
- Train eCNN model over 40 hyper-epochs and 4 hyper-batches
- Performance on validation set used to select hyperparameters
- Train eCNN model on single-modal inputs (image only, audio only, motion only) to evaluate how single-modal information is perceived by dog

Experiments: *Results*

Overall prediction accuracy: 79.02%

	Training	Validation	Testing
Number of Samples	3505	458	954
Number of Samples: class *Sit*	338	55	102
Number of Samples: class *Stand*	521	65	149
Number of Samples: class *Walk*	1574	207	427
Number of Samples: class *Smell*	1072	131	276
Overall Accuracy	94.34%	79.47%	79.02%
Accuracy of class *Sit*	99.11%	88.00%	84.21%
Accuracy of class *Stand*	95.59%	72.73%	78.87%
Accuracy of class *Walk*	96.19%	78.95%	78.66%
Accuracy of class *Smell*	86.47%	78.20%	74.33%

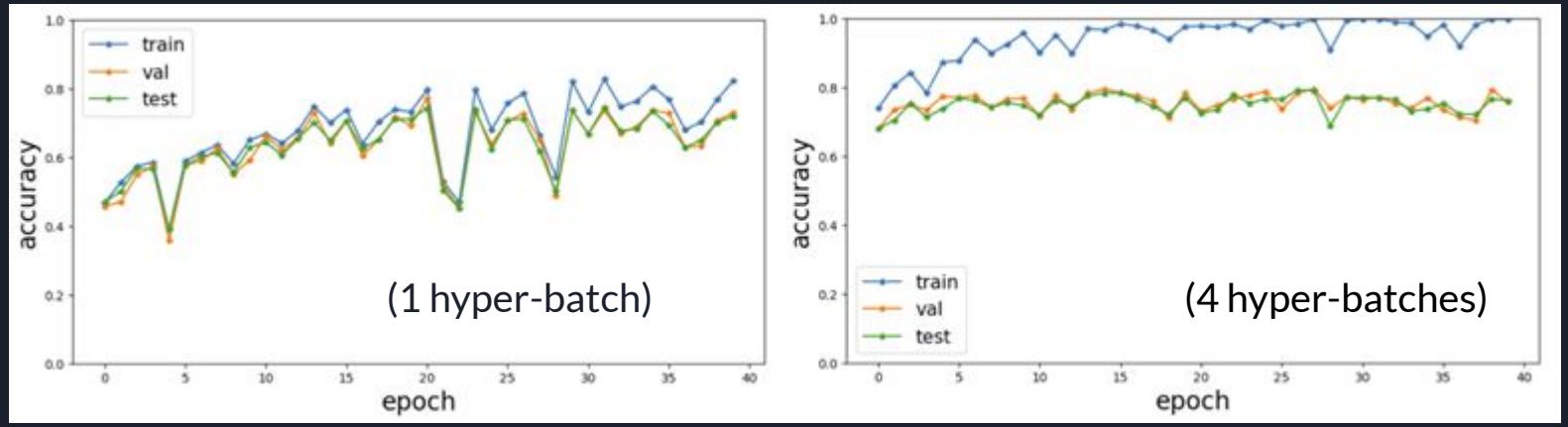
Experiments: *Results*

Confusion matrix on testing set

Ground truth \ Prediction	sit	stand	walk	smell
sit	80	5	9	1
stand	3	112	23	4
walk	15	26	328	48
smell	4	6	67	223

Experiments: *Results*

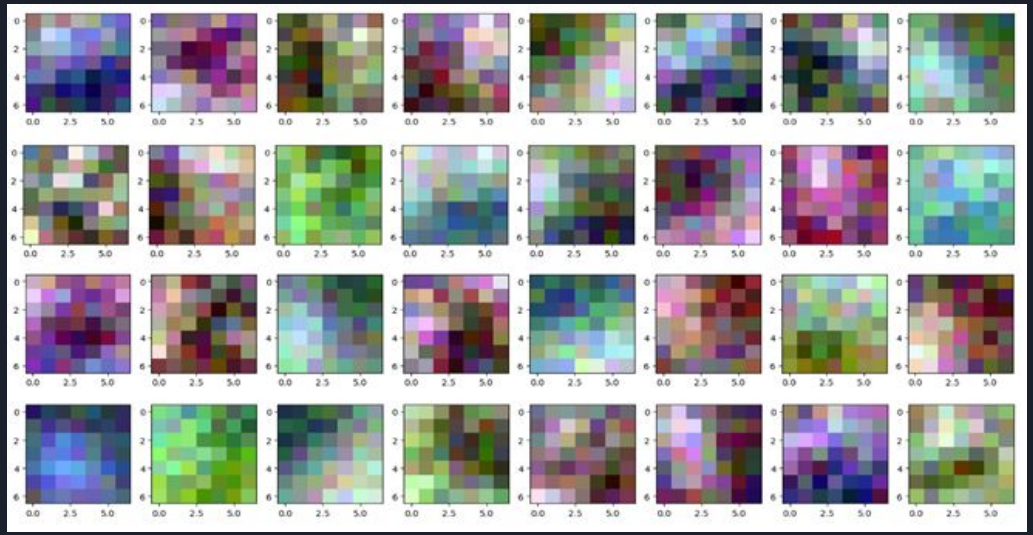
- *Observation: use of hyper-batches achieves optimal performance in less number of epochs*



(Testing hyperparameters: number of hyper-batches, number of hyper-epochs)

Experiments: *Results*

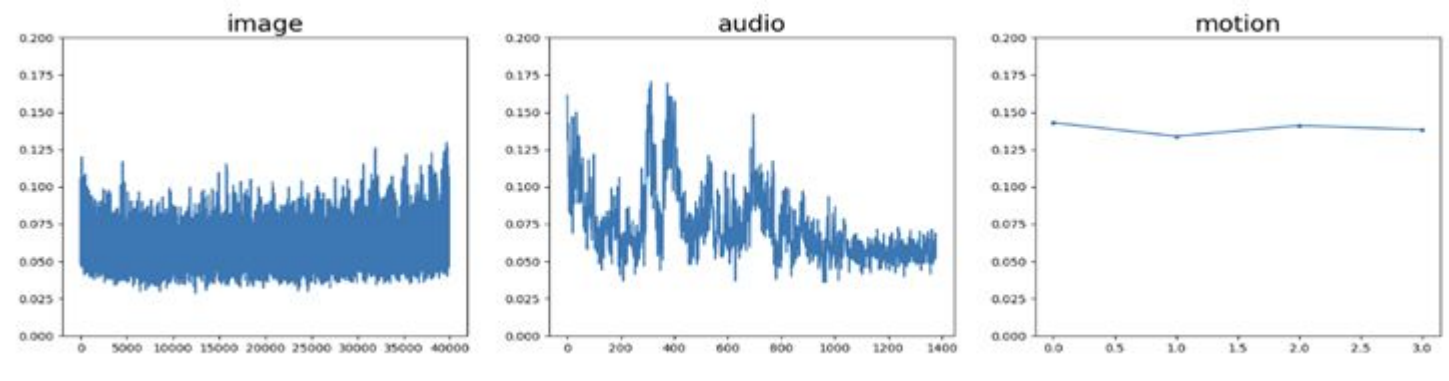
- *Observation: dog reacts to various color patterns and color contrasts!*



(convolution filters learned by eCNN)

Experiments: *Results*

- *Observation: dog reacts strongly to some low frequency components in what it hears!*



(average weights on image inputs, audio inputs, and motion inputs)

Conclusion

- We proposed an eCNN model to learn and model the association between a dog's perceived visual and audio stimuli and the dog's behavior
- eCNN model showed promising results in predicting dog's behavior
- Dog seems to react to various color patterns and color contrasts, as well as some low frequency components in the sound it hears
- The insights gained in this project can potentially create new ways of training service dogs for rescue work, companionship, and more

Future Work

- Add infrared sensors to study if and how dogs react to temperature
- Test sequence models such as Recurrent Neural Networks for potential performance improvements
- Extend data collection to study how a dog reacts to unfamiliar situations, human voices, other dogs barking, music, and much more
- Extend the study to different dogs and understand the general and individual behavior of dogs