



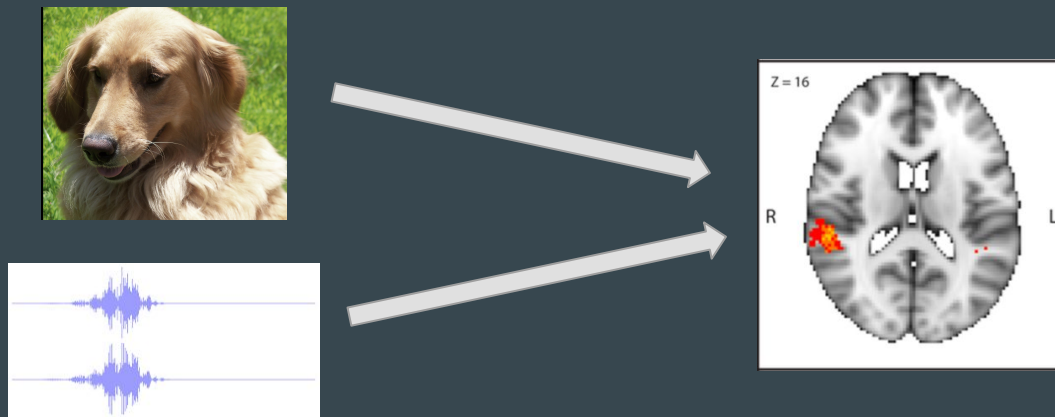
# The Brain Sees, The Brain Hears

...

Michelle Huntley, Armin Bazarjani, Dan  
Garvey

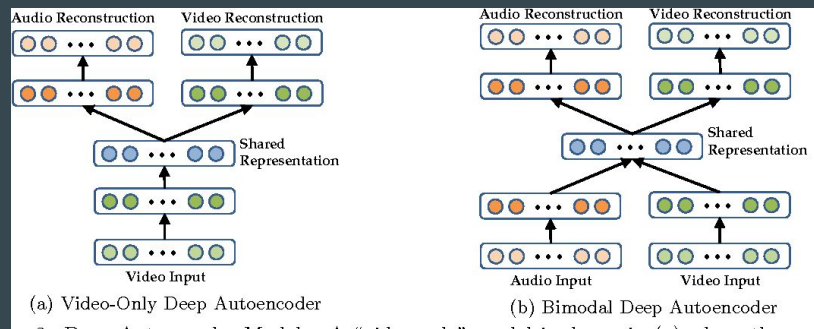
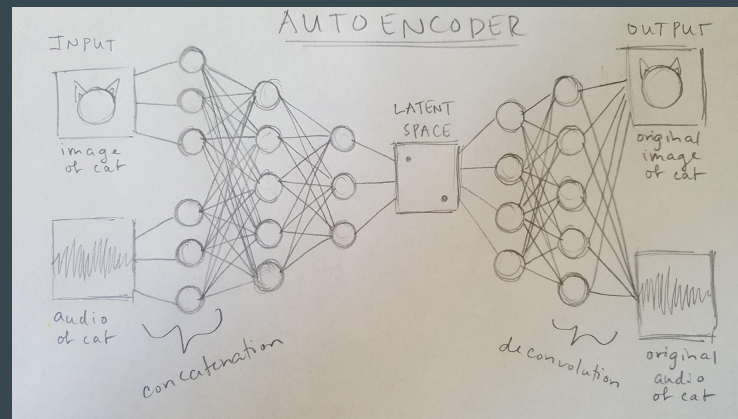
# Inspiration/Background

- Professor Kingson Man → Brain and Creativity Institute
- Mapping sensory input to areas of the brain
- GOAL: learn unified representation of multi-modal sensory input
  - In a neural network rather than the human brain



# Initial Steps

- Concepts
  - Autoencoders
  - PCA (latent space)
- Tools
  - Keras
    - Variational autoencoder framework
  - Moments in Time
    - Datasets of labeled videos
  - Deepmind
    - Research papers
    - Andrew Ng's 2015 project



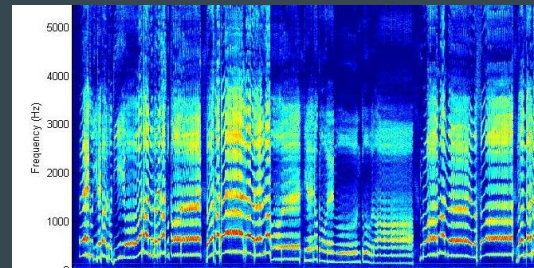
# Obstacles Faced

- New concept: autoencoder
- Little prior experience
- Open-ended project
- Finding an adequate dataset

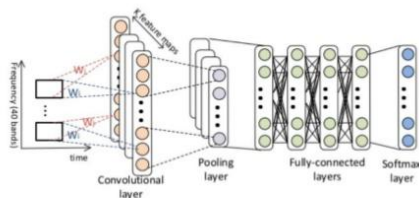


# Current stage

- Building two CNNs to connect to the autoencoder
  - 2D for spectrograms from audio
  - 3D for video
- Extract mp3s/WAV from videos in dataset
- Generate spectrograms from audio files



Convolutional Neural Network (CNN)



# Next Steps

- Continue building CNNs and autoencoders
- Integrate other/more models
  - GAN
    - Popular in industry
  - RNN
    - Recurrent nature could work well with audio
- Utilize dataset in different ways
- Read more research papers
- Continue meeting/communicating with Kingson