Deep Learning for Neuro-visualization and Continuous Control in Autonomous Systems

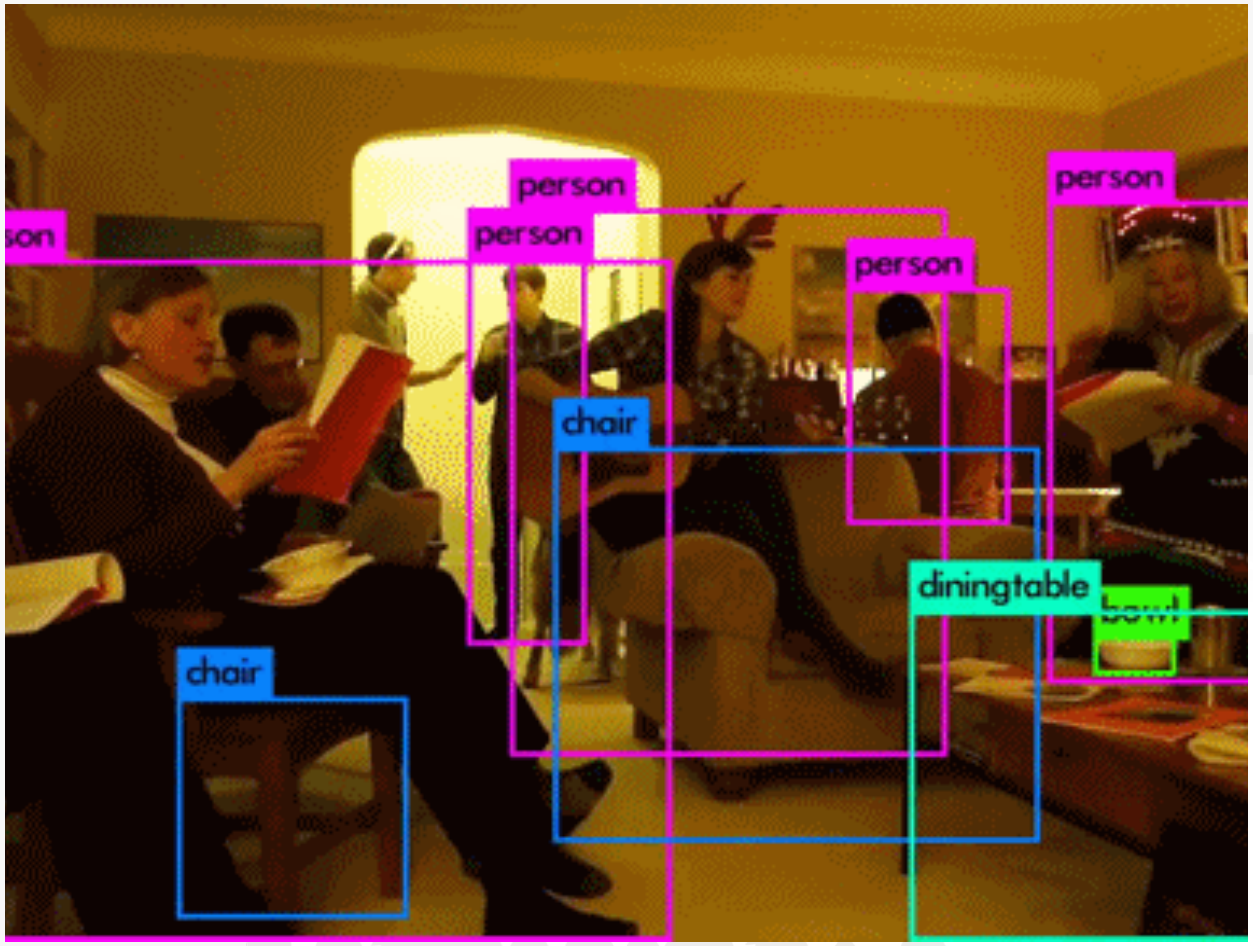
James Ecker(presenter)

james.e.ecker@nasa.gov

Goddard AI Workshop

27 November 2018

Yolo Real-time Image Recognition

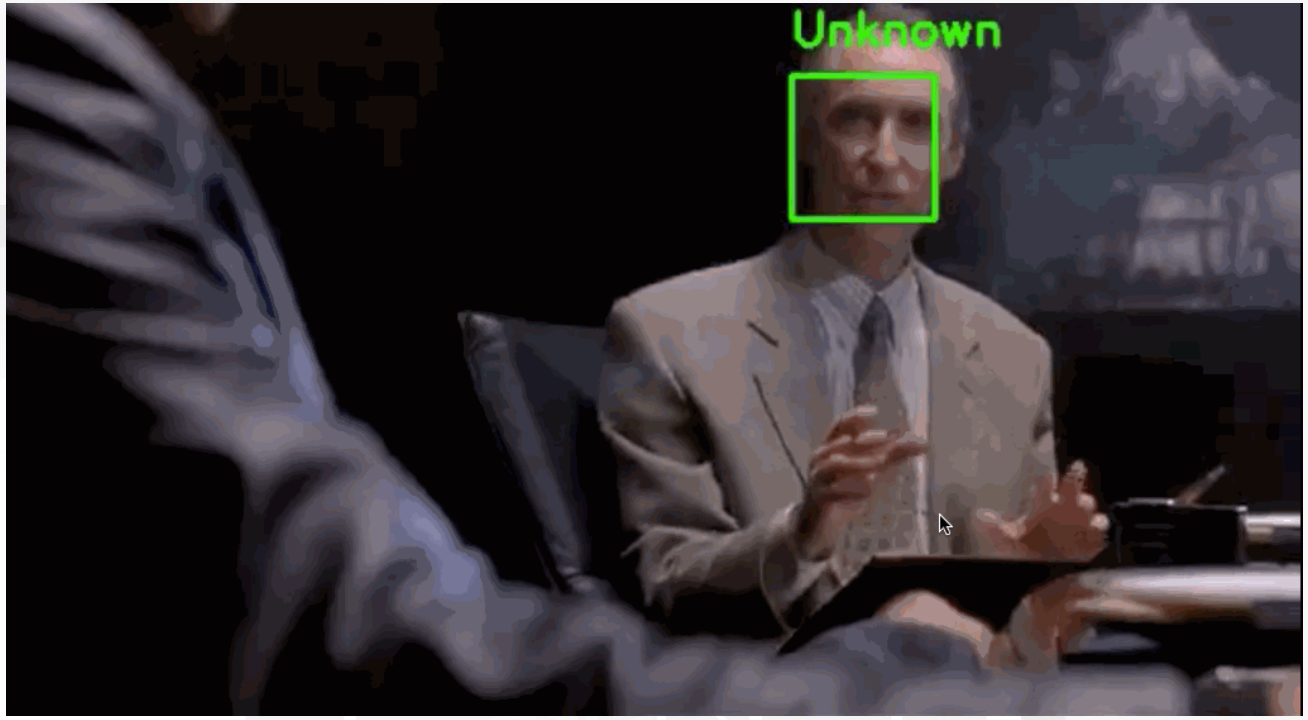


Source: [1] prosthetic knowledge @ tumblr.com

- Object granularity at class level
- Identify group density via bounding box overlap
- Specificity beyond out-of-the-box classes requires significant training

Individual Face Recognition

- Each individual is now a class
  - Requires large amount of class instance data
  - Training data explodes with number of faces to be detected
  - Photos of individual at various angles
  - Not appropriate for time-sensitive search and rescue ops
- Unknown faces detected introduce uncertainty in recognizing known faces

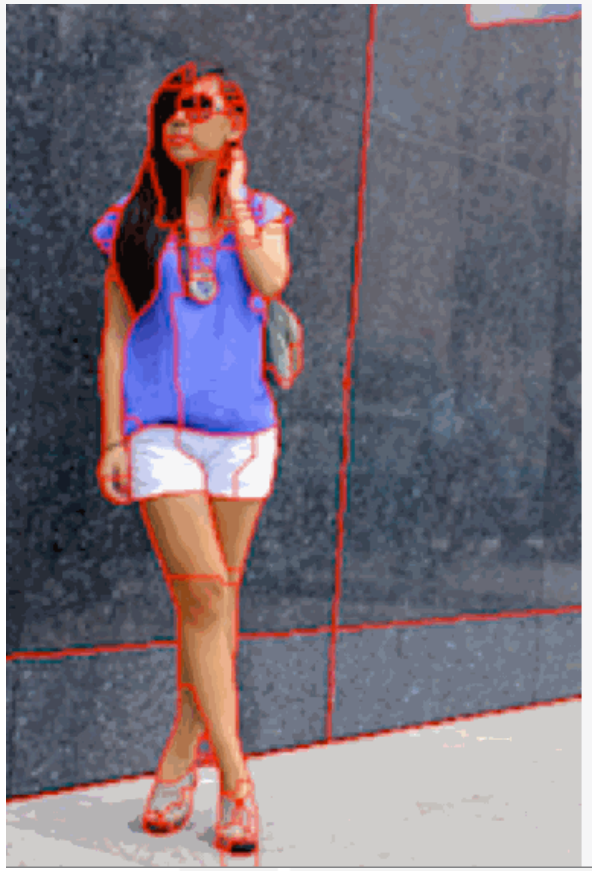


Source: [2] Adrian Rosebrock

Parse Clothing in Scene

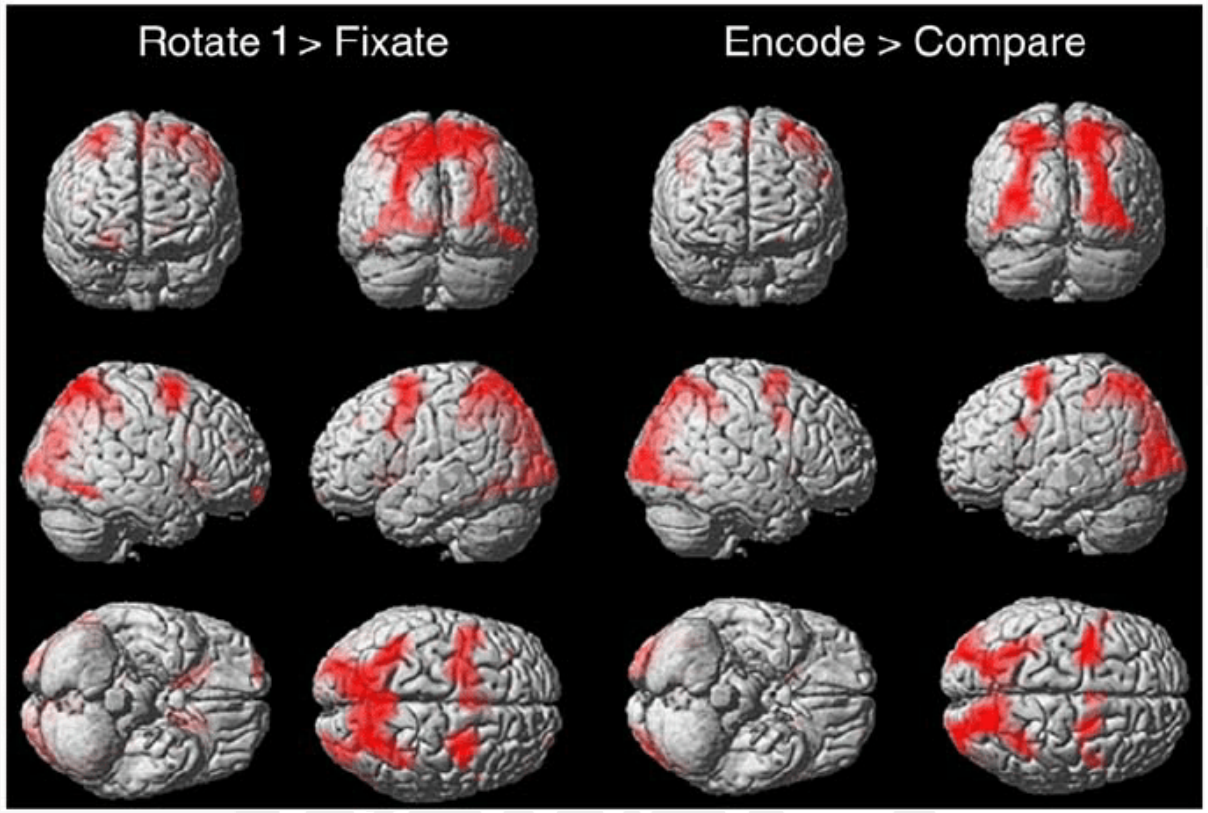
[3] Clothing Parser

- Too computationally expensive for real-time decision analysis
  - State of the art industry parsers run somewhere on order of 5-10 seconds per image
  - Bottleneck in the segmentation algorithm
- Large memory footprint
  - 2-3 GB for parser
  - 5-7 GB to train pose estimator
- Can be extremely useful for explainability or augmenting proposed system decisions



Source: [3] Kota Yamaguchi et al.

Mental Imagery in Humans

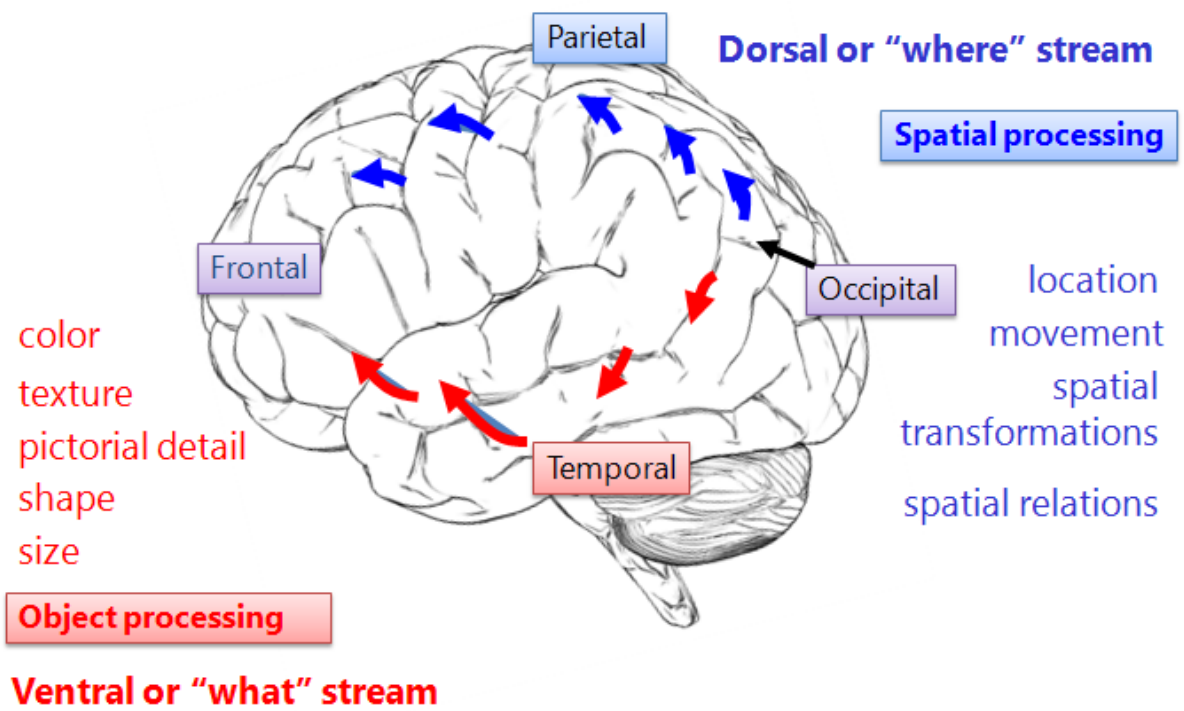


Source: [5] Claus Lamm et al.

- Studies using fMRI have shown that the lateral geniculate nucleus and the V1 area of the visual cortex are activated during mental imagery tasks [4]
  - [4] found that focusing attention on features of the imagined faces (e.g., eyes, lips, or nose) resulted in increased activation in the right intraparietal sulcus (IPS) and the right inferior frontal gyrus (IFG)
  - Results in [4] suggest differential effects of memory and attention during the generation and maintenance of mental images of faces
- Representational Form
  - Dual Code Theory
    - **Analogue codes**
    - **Symbolic codes**
  - Propositional Theory
    - Stored as propositions rather than as images
  - The Functional-equivalency Hypothesis
    - Model of object/concept is stored exactly as it is perceived

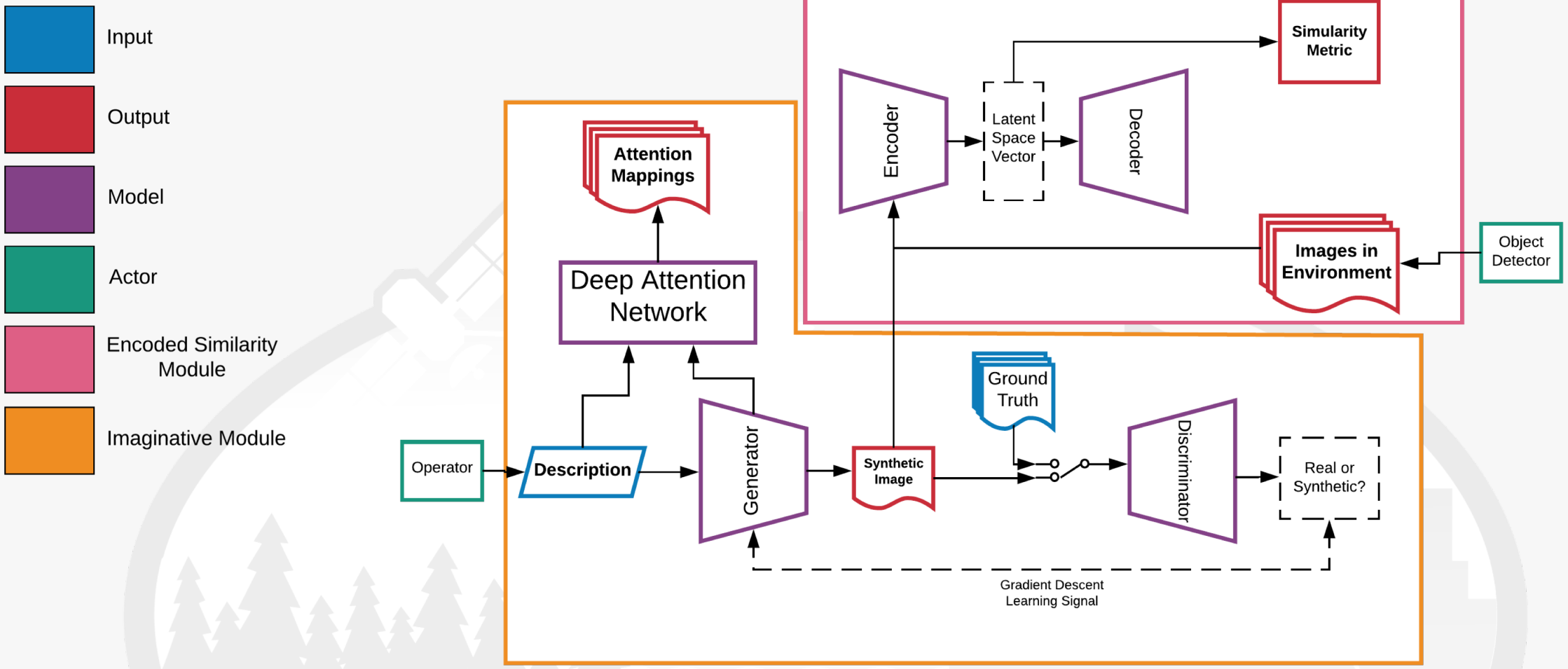
Object Visualization

- Results in [7] show object visualization skills grow with age, indicating link between object visualization, memory, and learned representations of objects
- [8] shows a positive link between object visualization and vocabulary acquisition
- [9] shows that individuals with higher ability to visualize objects recognize objects at lower resolution than those with less ability
- [10] shows that visualizing tasks help students gain the motor skills associated with the task at higher rates than those who don't



Source: [6] Maria Kozhevnikov et al.

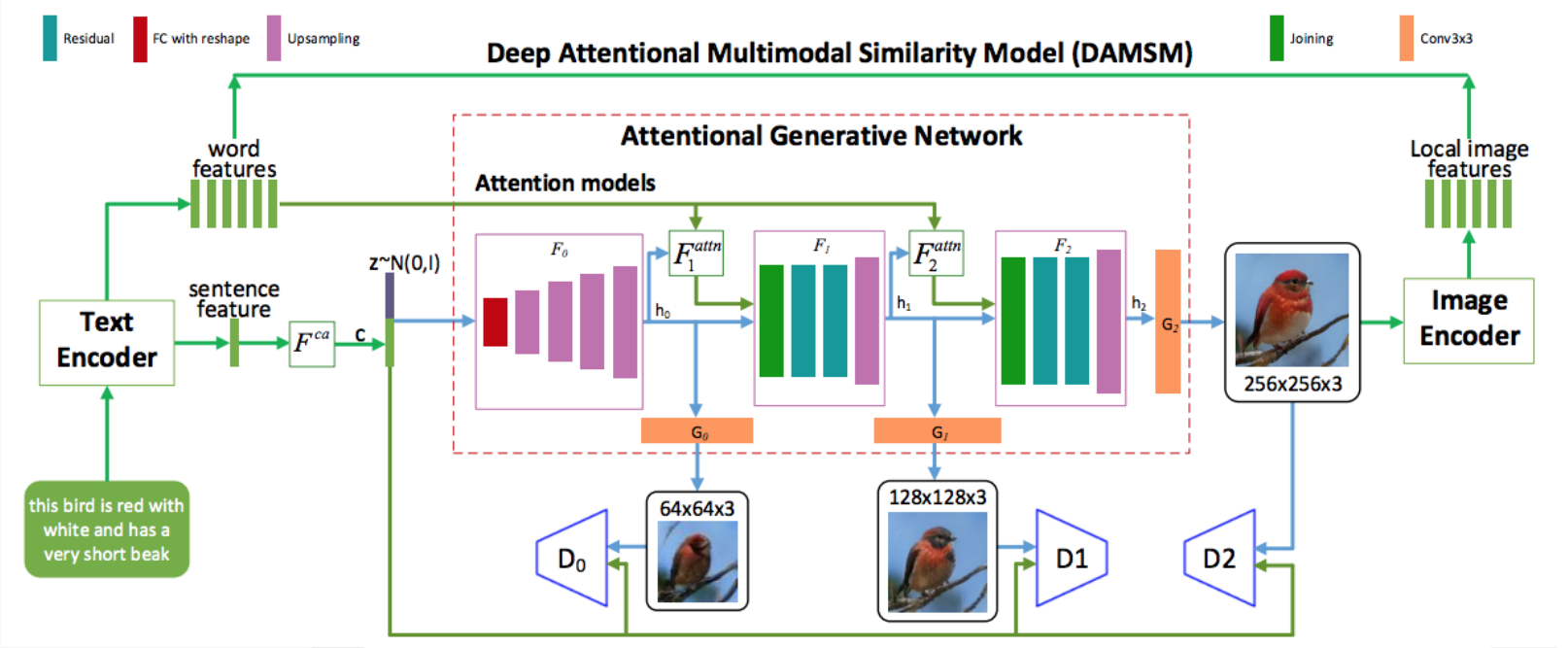
Proposed System



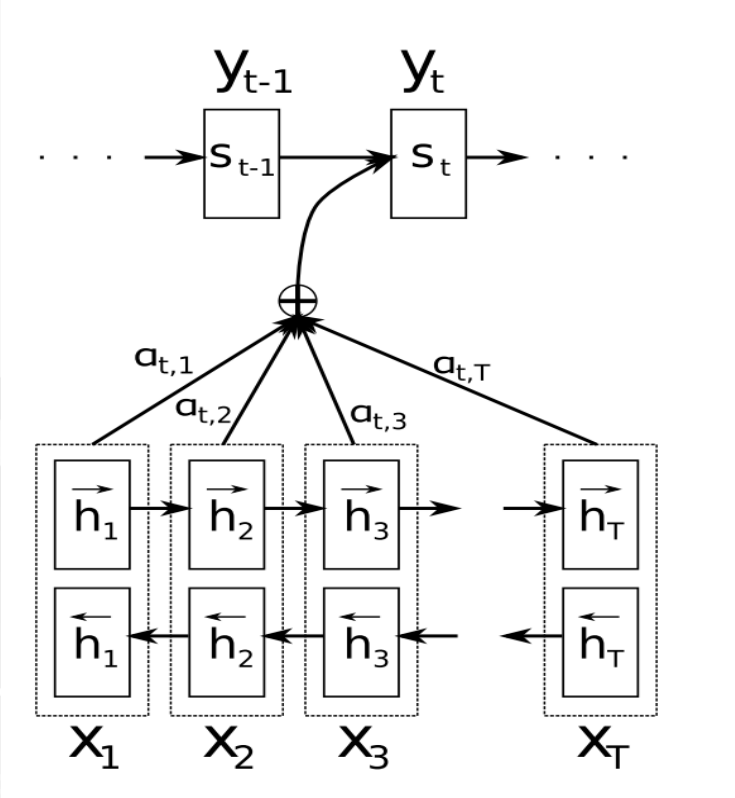
Imaginative Module

- Natural language to image generative network
  - Generates pixel-space representation of human described object
  - Surrogate for mental imagery via functional-equivalency hypothesis
  - Attention mechanisms are key to proper NL-image generation

Imaginative Module cont'd



Source: [11] Tao Xu et al.



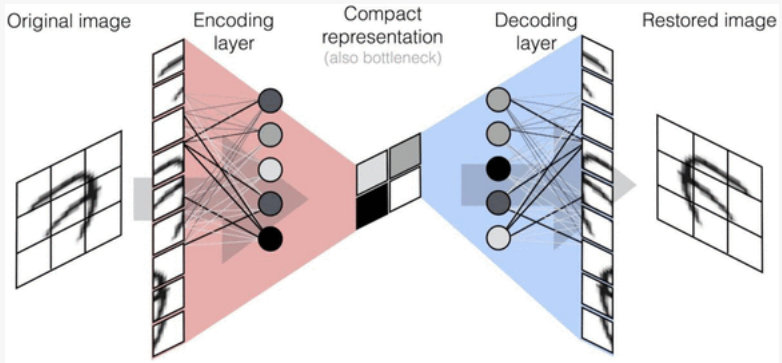
Source: [12] Pranoy Radhakrishnan.

- AttnGAN [11] system diagram – Top
- Attention Mechanism [12] – Right
  - Iterate over the input in discrete timesteps
  - Learning a context vector for each timestep instead of encoding a single fixed-length context vector over the entire input
  - This context vector allows the model to learn where to attend based on the input data current to its respective timestep

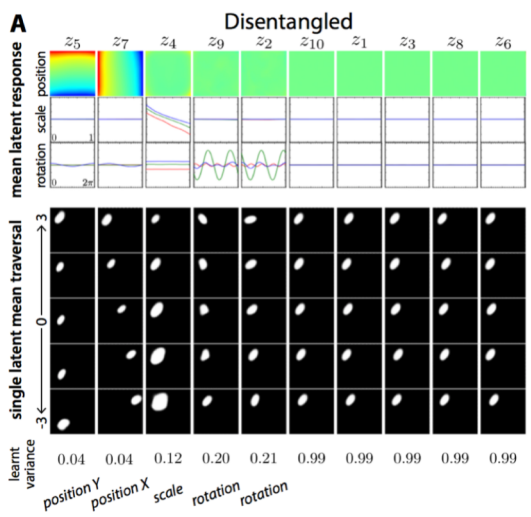
Encoded Similarity Module

- Tasked with converting target and observed image pairs to their respective feature vectors via a pre-trained model
  - Disentangled variational autoencoder
  - Last layer in Generator in GAN
- Cosine similarity metric

$$\sigma = \frac{\sum_{i=1}^{n} T_i O_i}{\sqrt{\sum_{i=1}^{n} T_i^2}\sqrt{\sum_{i=1}^{n} O_i^2}}$$



Source: [12] William Jones et al.



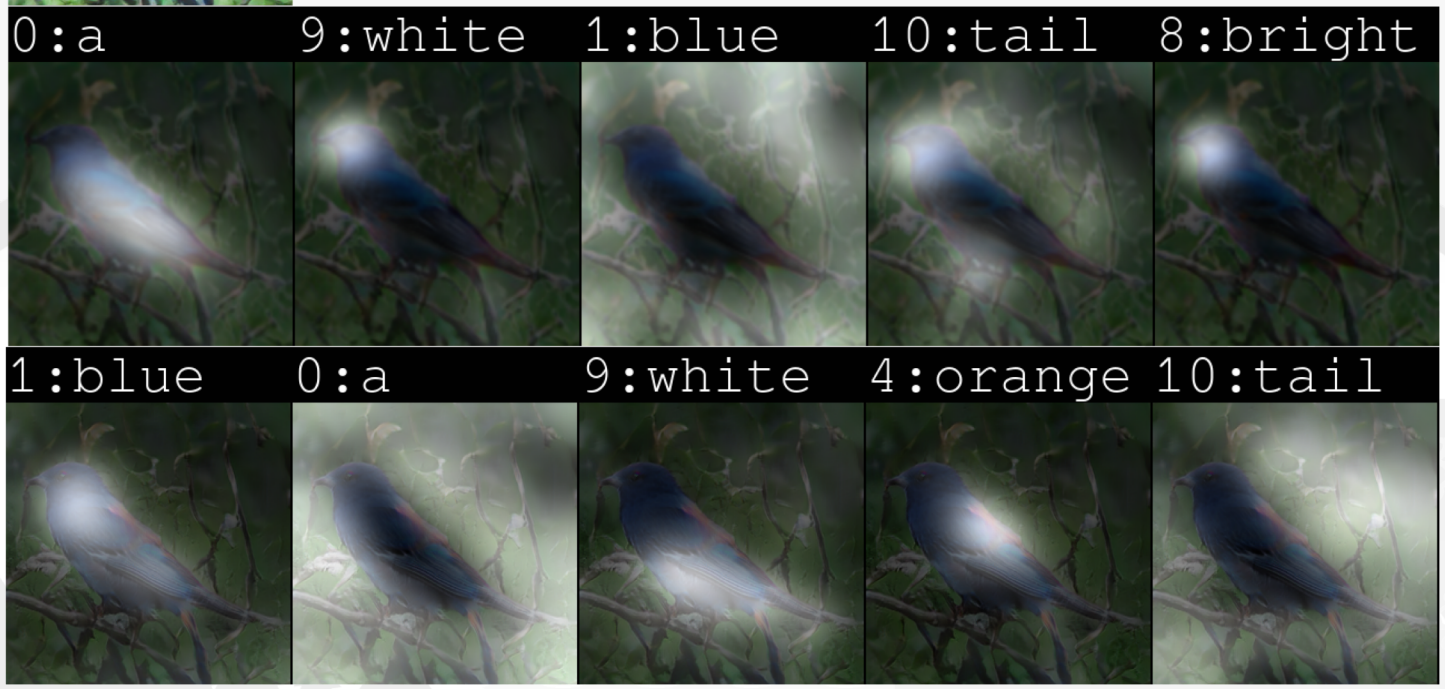
Higgins et al.

Gregor et al.

Source: [13] Joe Marino.

Explainability Mock-up

A blue bird with orange wings and a bright white tail



Preliminary Findings

- Good results rely on two main characteristics in the training data
  - Description Context
  - Localization of object
- CUB data
  - Object specific descriptions with little regard to other information in image
  - Bounding box meta-data for the bird being described

Preliminary Findings

- COCO dataset
  - Descriptions contain information holistic to the image
    - Low object specificity
  - Segmentation masking meta-data available
    - AttnGAN does not exploit this

A Boy Playing Baseball

Preliminary Findings

Synthetic | Ground Truth

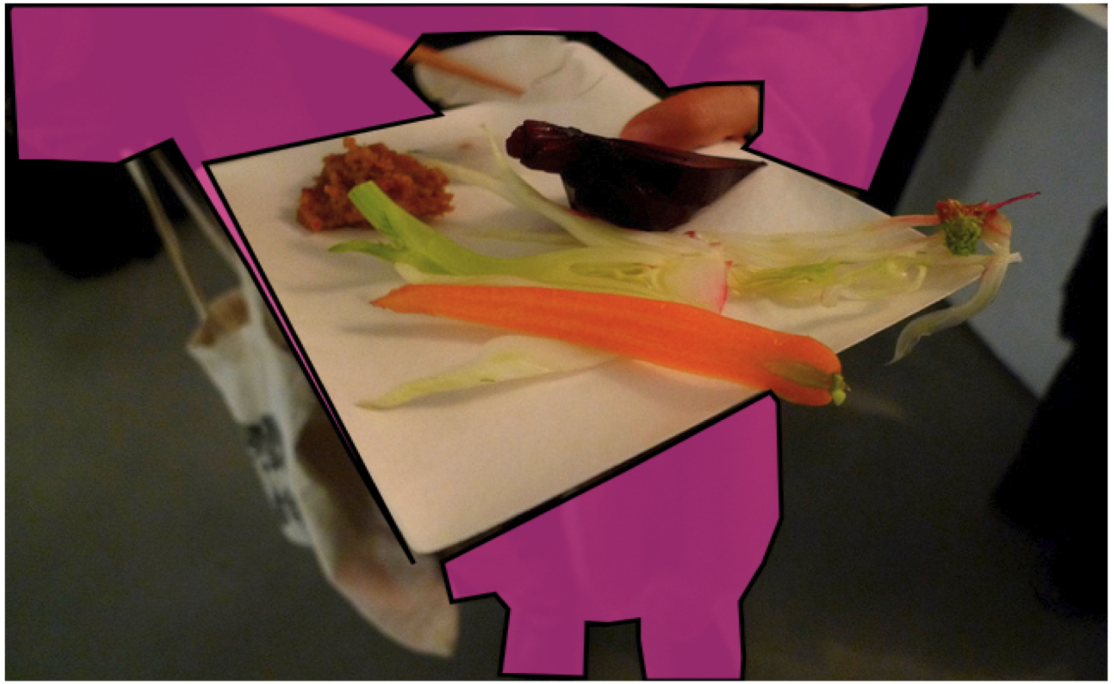
this is a small black bird with a white spot on its nape, it has a very large head and bill for its body.

- Reconstruction from ground truth
  - Shows highly accurate but imperfect recall
  - Generalization of distribution high on both datasets
  - High class stratification in COCO dataset contributes to low spatial learning
    - Subset of COCO
      - only images containing people

Preliminary Findings

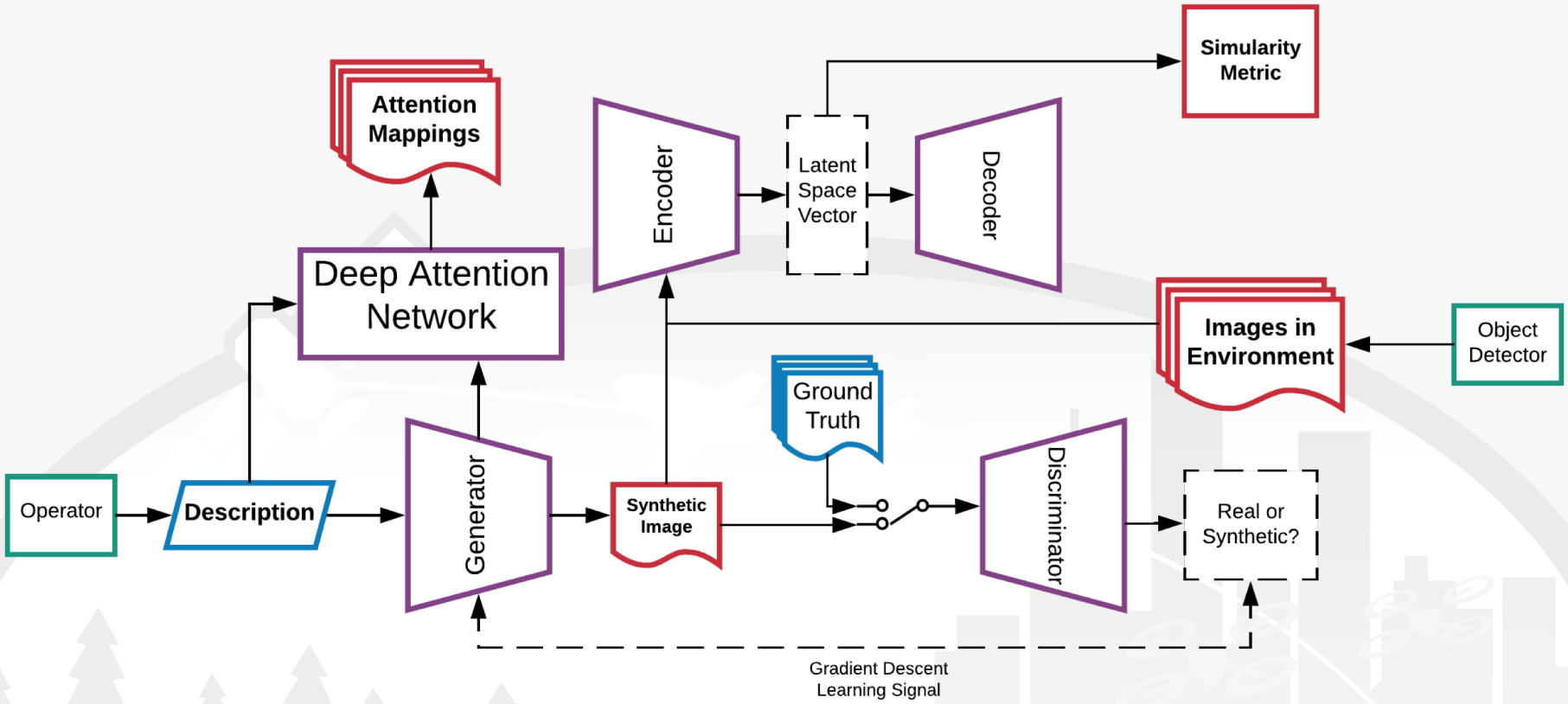
several motorcycles and cars sitting on a grassy lot
large motorcycle sitting on a grassy area in a line.
older motorcycle displayed on grass along with several old cars.
a motorcycle that is parked in a field.
a number of motorbikes and cars parked in the field

a view of a plate of food , that looks very decorative.
a small plate with some vegeatables being held by a person
some food that is sitting on a napkin.
a square plate with all kinds of vegetables on the plate
a person is holding a board with sliced vegetables.

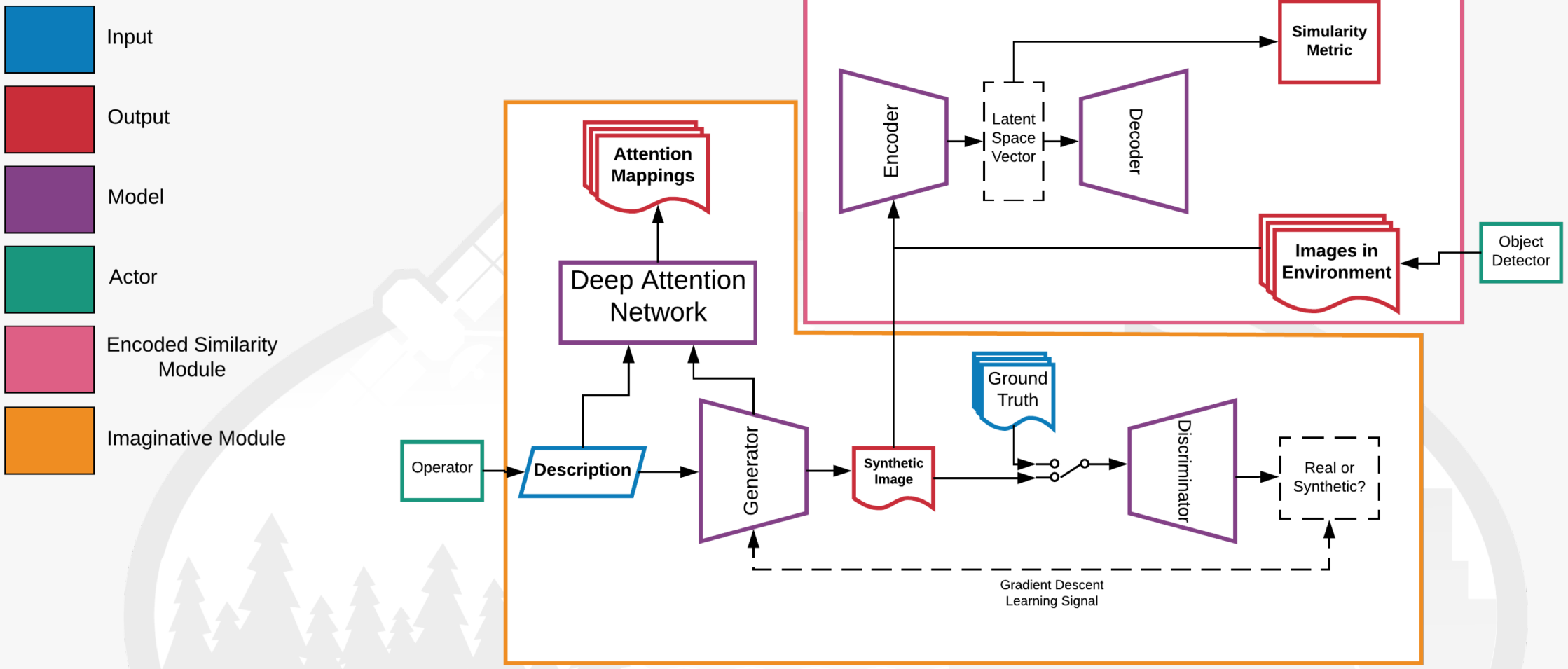


- People subset
  - Some images contain people but subject is some other object
  - Descriptions either not related to the people in the image (left) or relating to person not well represented in image (right)

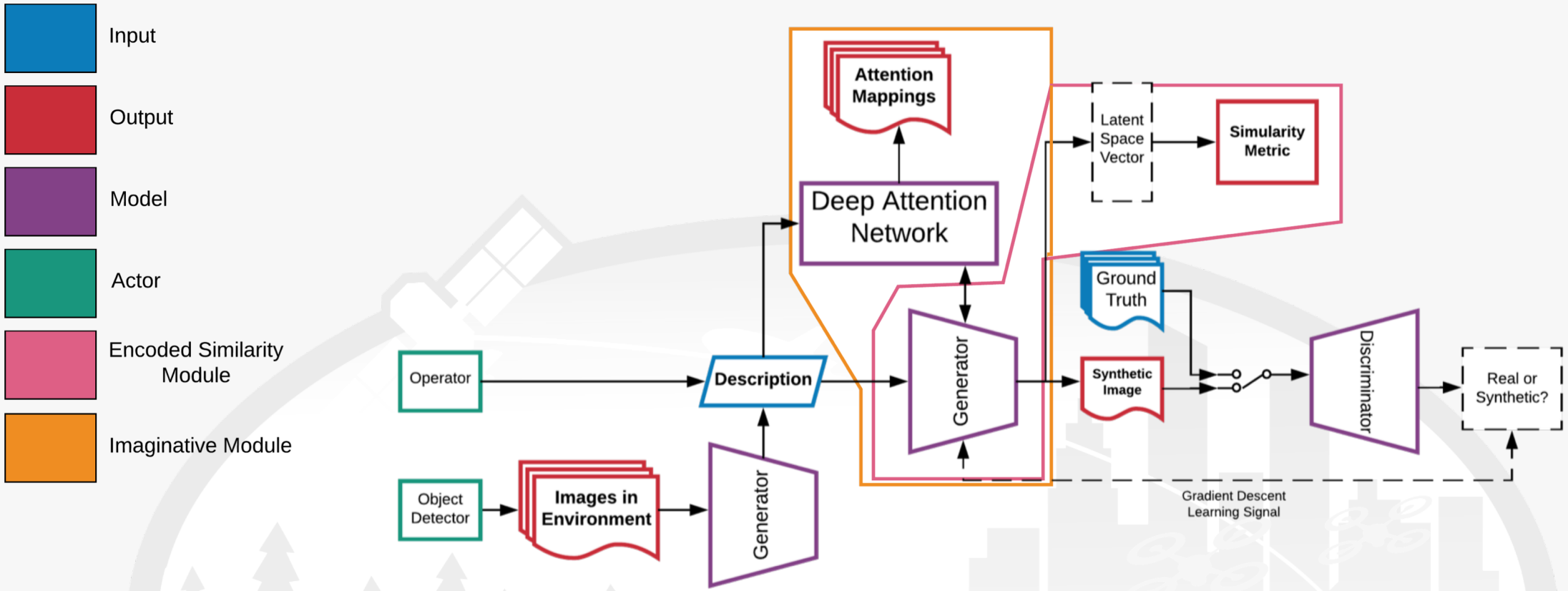
Future Considerations



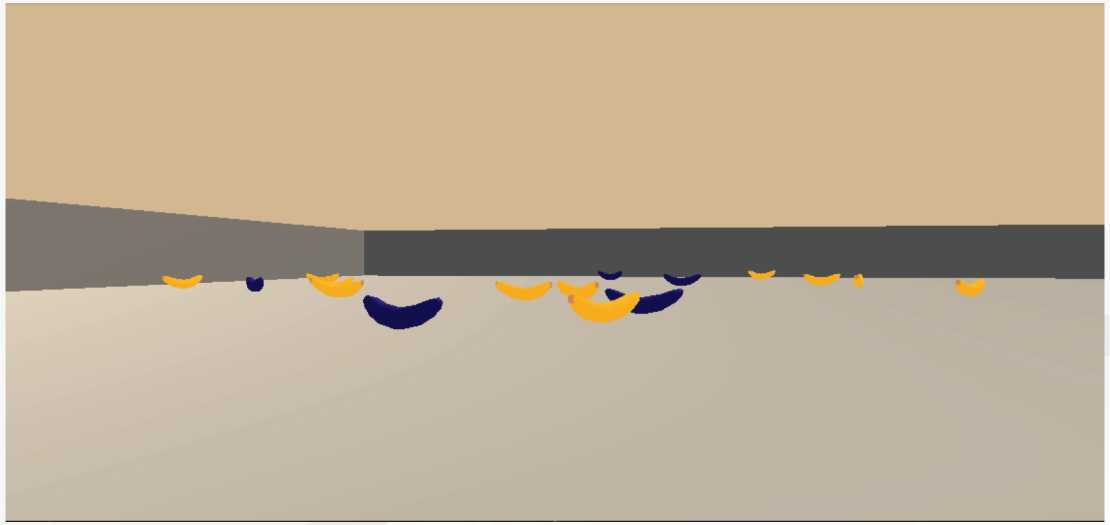
Future Considerations

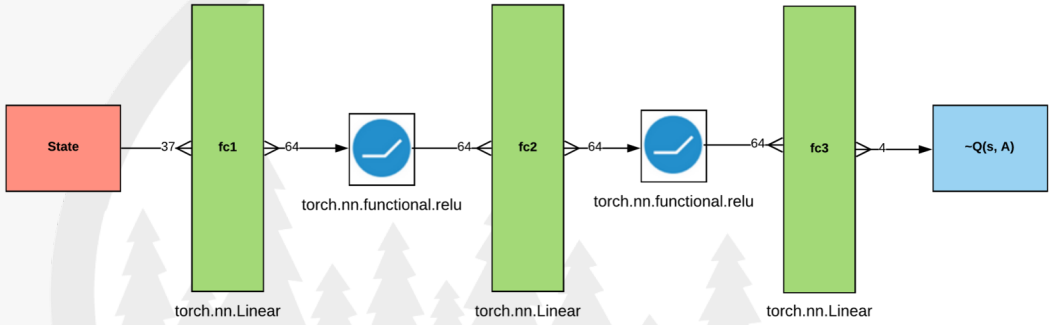


Future Considerations



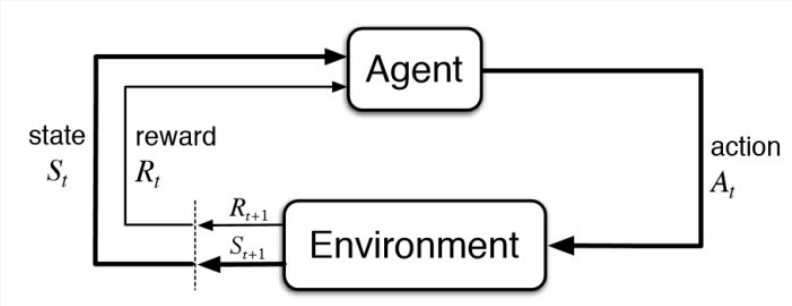
Distributed Deep RL in Unity

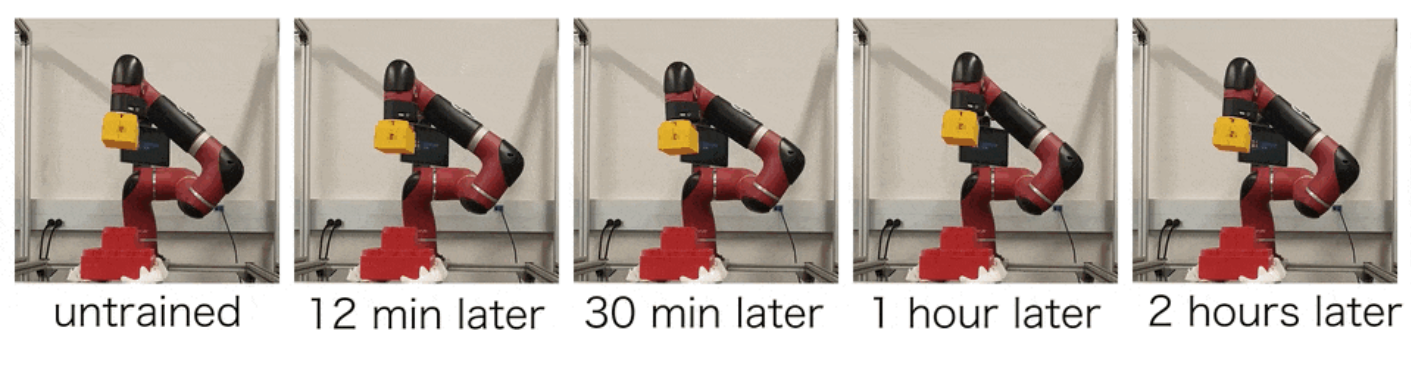


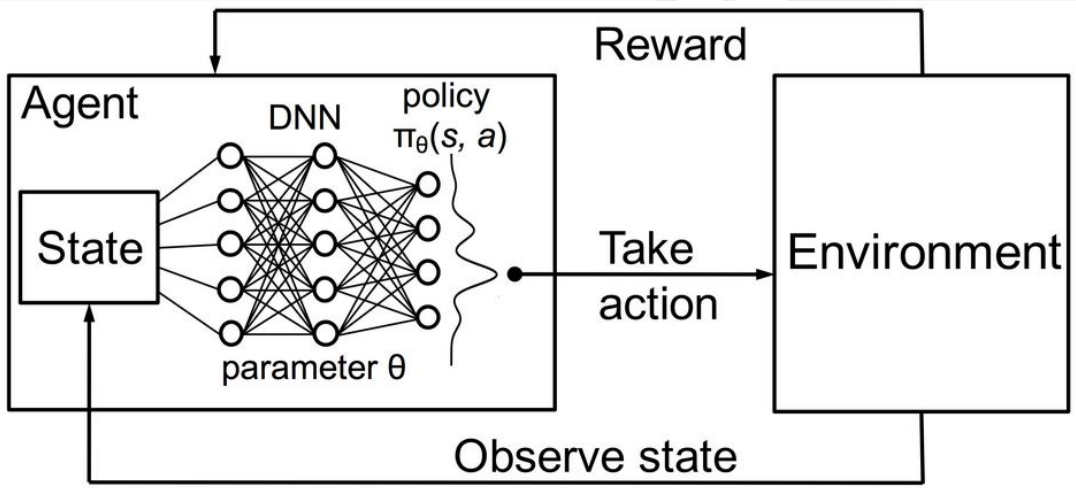


- Uses Unity ML Librarys
- State Space
  - 37 Dimensional Continuous Data
  - Agent Velocity
  - Ray Traces to objects in agent's forward path
- Action Space
  - 4 Dimensional Discrete Data
    - 0 – Forward
    - 1 – Backward
    - 2 – Left
    - 3 – Right
- Rewards
  - + 1 for each Yellow Banana
  - - 1 for each Blue Banana
- Goal
  - Learn to navigate
    - Avoid blue bananas
    - Collect yellow bananas
  - Solved when agent gets average score of 13 over 100 episodes

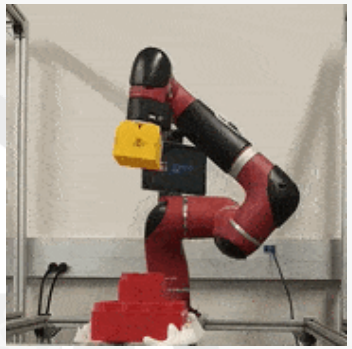
Deep Reinforcement Learning for Continuous Control in Autonomous Systems





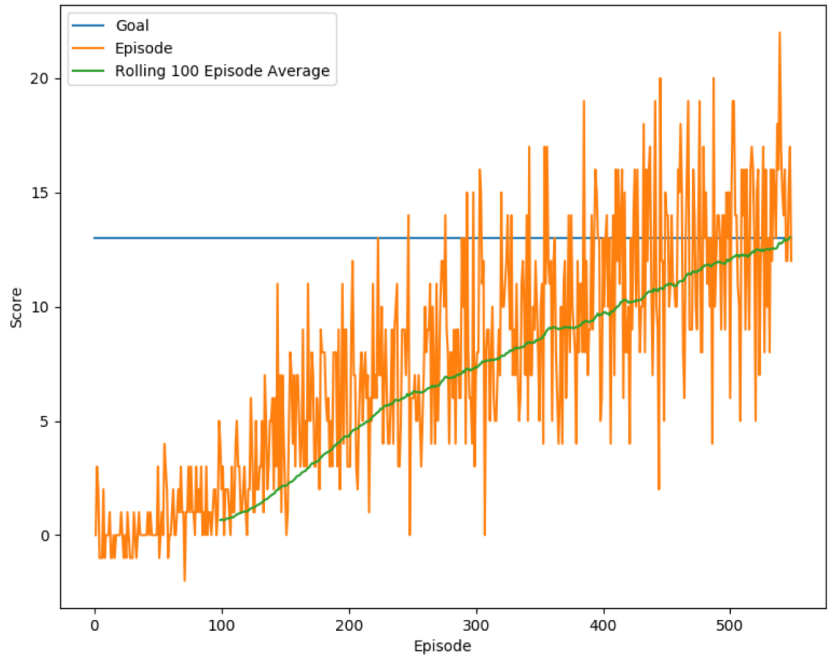


credit: pVoodoo
Deep Reinforcement for Trading? NOPE! - Blogspot



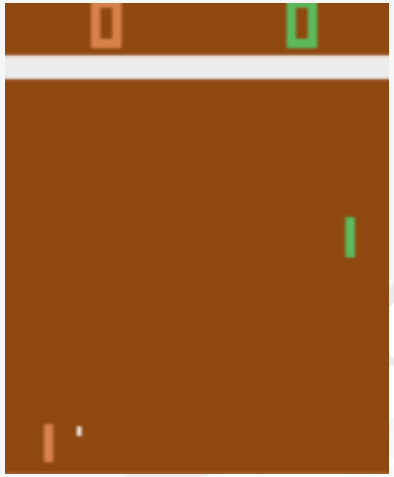
credit: Aurick Zhou
Berkeley Artificial Intelligence Research

Distributed Deep RL in Unity

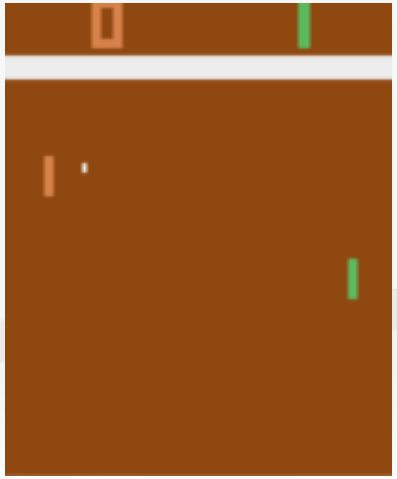


Architecture	Episodes	Wall Time (minutes)
Single GPU	1800	35
Heterogenous	624	18
Homogenous	482	10

Deep RL in Continuous Action Space using only Pixels



After 500 Episodes



After 2500 Episodes

- State Space
  - 84 x 84 pixels
- Action Space
  - Continuous
    - 0 in center of player side
    - + pixels – Up
    - - pixels – Down
- Rewards
  - + 1 for each win
  - - 1 for each loss
- Goal
  - 21 points

Questions?

Sources

1. Prostheticknowledge. 2018. prostheticknowledge. (March 2018). Retrieved June 21, 2018 from http://prostheticknowledge.tumblr.com/post/172309765681/yolo-v3new-release-of-object-recognition-framework

2. Adrian Rosebrock. 2018. Face recognition with OpenCV, Python, and deep learning. (June 2018). Retrieved June 21, 2018 from https://www.pyimagesearch.com/2018/06/18/face-recognition-with-opencv-python-and-deep-learning/

3. Parsing clothing in fashion photographs Kota Yamaguchi, Hadi Kiapour, Luis E Ortiz, Tamara L Berg Compute Vision and Pattern Recognition 2012.

4. Ishai, Alumit et al. "Visual imagery of famous faces: effects of memory and attention revealed by fMRI." *NeuroImage* 17 4 (2002): 1729-41.

5. Lamm, Claus & Windischberger, Christian & Moser, Ewald & Bauer, Herbert. (2007). The functional role of dorso-lateral premotor cortex during mental rotation: an event-related fMRI study separating cognitive processing steps using a novel task paradigm. NeuroImage. 36. 1374-86.

6. Kozhevnikov, Maria et al. "Trade-off in object versus spatial visualization abilities: restriction in the development of visual-processing resources." *Psychonomic bulletin & review* 17 1 (2010): 29-35.

7. Blazhenkova, O., Kozhevnikov, M., Becker, M. (in press, available online Dec 2, 2010). Object-Spatial Imagery and Verbal cognitive styles in children and adolescences. Learning and Individual Differences

8. Cohen, Marisa. (2009). The Effectiveness of Imagery Interventions on the Vocabulary Learning of Second Grade Students. NERA Conference Proceedings 2009.

9. Vannucci, Manila, et al. "Object Imagery and Object Identification: Object Imagers Are Better at Identifying Spatially-Filtered Visual Objects." *Cognitive Processing*, vol. 9, no. 2, 2008, pp. 137–143., doi:10.1007/s10339-008-0203-5.

10. Schuster, Corina & Hilfiker, Roger & Amft, Oliver & Scheidhauer, Anne & Andrews, Brian & Butler, Jenny & Kischka, Udo & Ettlin, Thierry. (2011). Best practice for motor imagery: A systematic literature review on motor imagery training elements in five different disciplines. BMC medicine. 9. 75. 10.1186/1741-7015-9-75.

11. Xu, Tao et al. "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks." *CoRR* abs/1711.10485 (2017): n. pag.

12. Jones, William, et al. "Computational Biology: Deep Learning." *Emerging Topics in Life Sciences*, Portland Press Journals Portal, 14 Nov. 2017, www.emergtoplifesci.org/content/1/3/257.

13. Marino, Joe Louis. "Thoughts on Generative Models." *Thoughts on Generative Models*, joelouismarino.github.io/blog_posts/blog_VAE.html.

14. Lin, Tsung-Yi; Maire, Michael; Belongie, Serge; Hays, James; Perona, Pietro; Ramanan, Deva; Dollár, Piotr; Zitnick, Lawrence, *Microsoft COCO: Common Objects in Context.* European Conference on Computer Vision (ECCV), Zürich, 2014, (Oral).