# GAP: Quantifying the Generative Adversarial Set and Class Feature Applicability of Deep Neural Networks

Edward Collier

*Louisiana State University*

Supratik Mukhopadhay

*Louisiana State University*

# Outline

- Generative Adversarial Networks, Transferability and Applicability
- Contributions
- GAN Applicability
- Experimental Setup
- Results
- Conclusions

# Generative Adversarial Networks (GAN)

- GANs are a semi supervised neural network model that consists two competing networks; a generator and a discriminator.
- The Generator’s objective is to produce products almost identical to the ground truth.
- The Discriminator’s objective is to discriminate between the ground truth and the generator’s output.
- The two competing networks eventually converge at the Nash equilibrium.

# Transferability

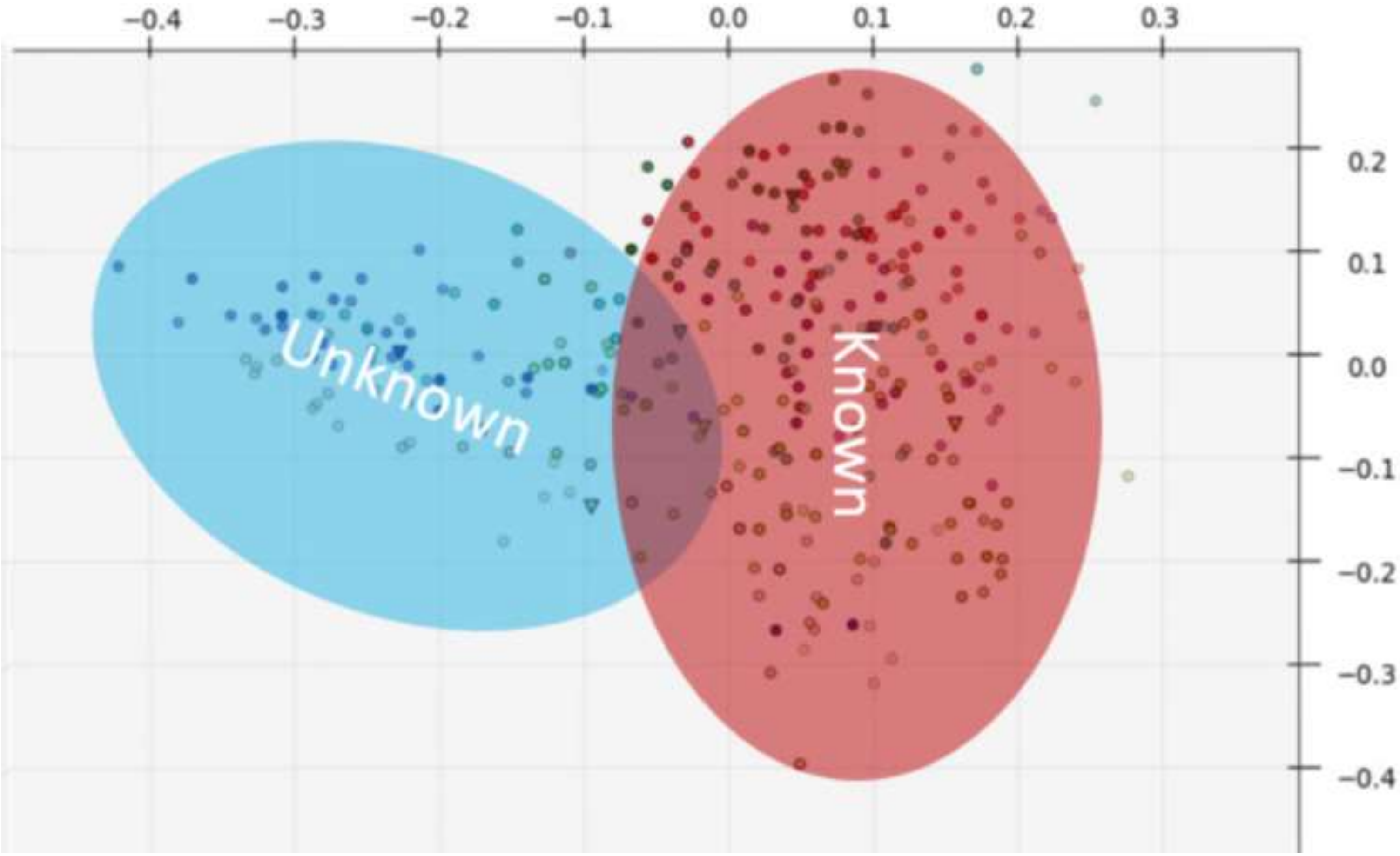
- Humans are able to discern what previous knowledge might apply to new problems.
- The brain will ”fit” previous knowledge to new tasks where it applies.
- This same idea can apply to neural networks, which we know as transferability.
- Transferability utilizes the idea that certain features overlap many classes.

# Applicability

- We define applicability as how well known features at a specific layer can be used to differentiate.
- In neural networks this can be broken down into three sub-groups
  1. Set applicability: How well does a network apply to a whole task?
  2. Class applicability: How well can the known features be used to differentiate an input class from all other input classes?
  3. Input applicability: How well can the known features be used to differentiate a single input from all other inputs?

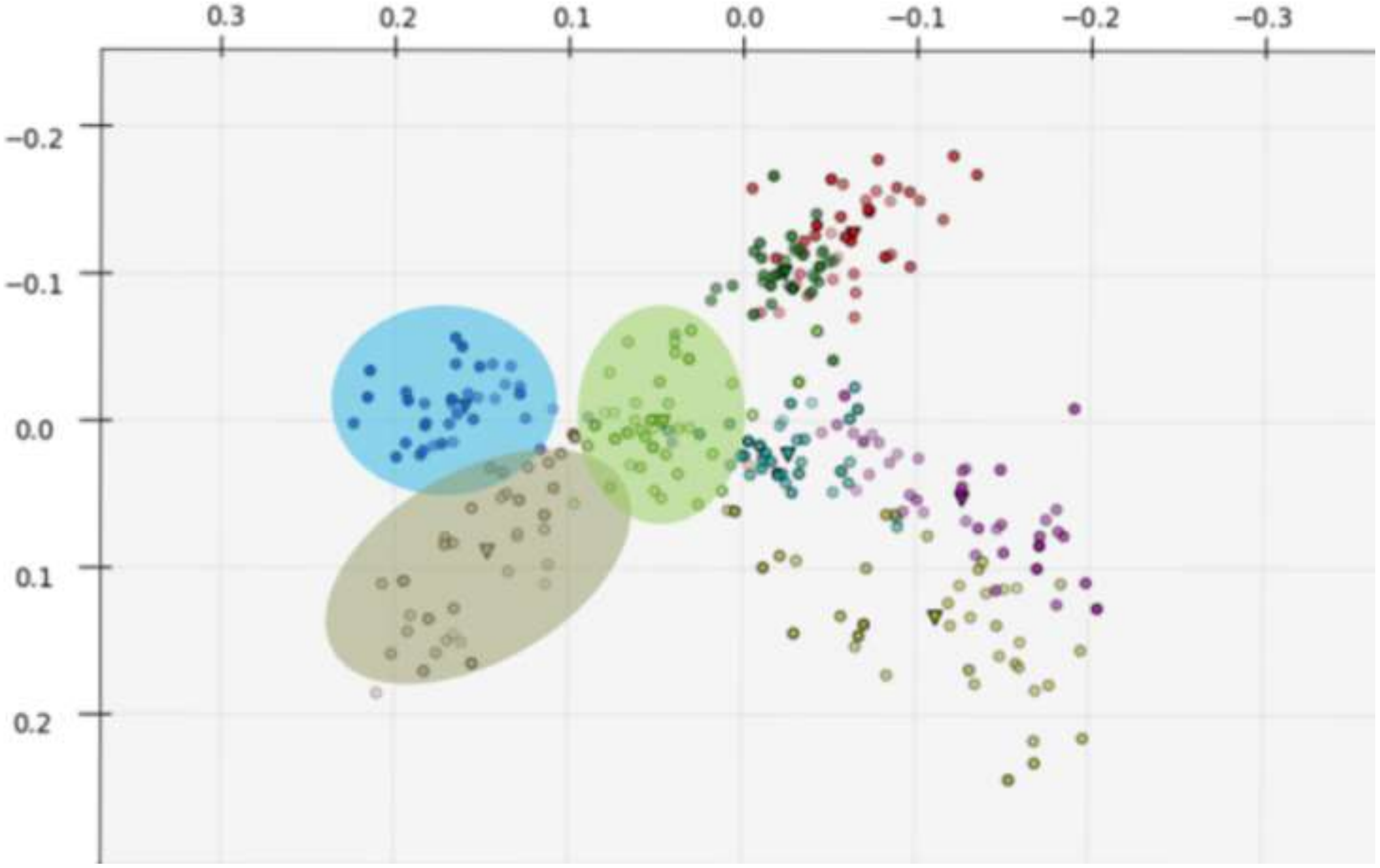
LSU

# Set Applicability




LSU

# Class Applicability



# Measuring Applicability

- By freezing and retraining layers we measure how well a network *N* can separate input *x* from a class from the unknown set $un_j$ at layer $n_i$.

$$\xi_j = N((x, un_j), n_i)$$

- Class applicability is then the average separability between *x* and all $un_j$.

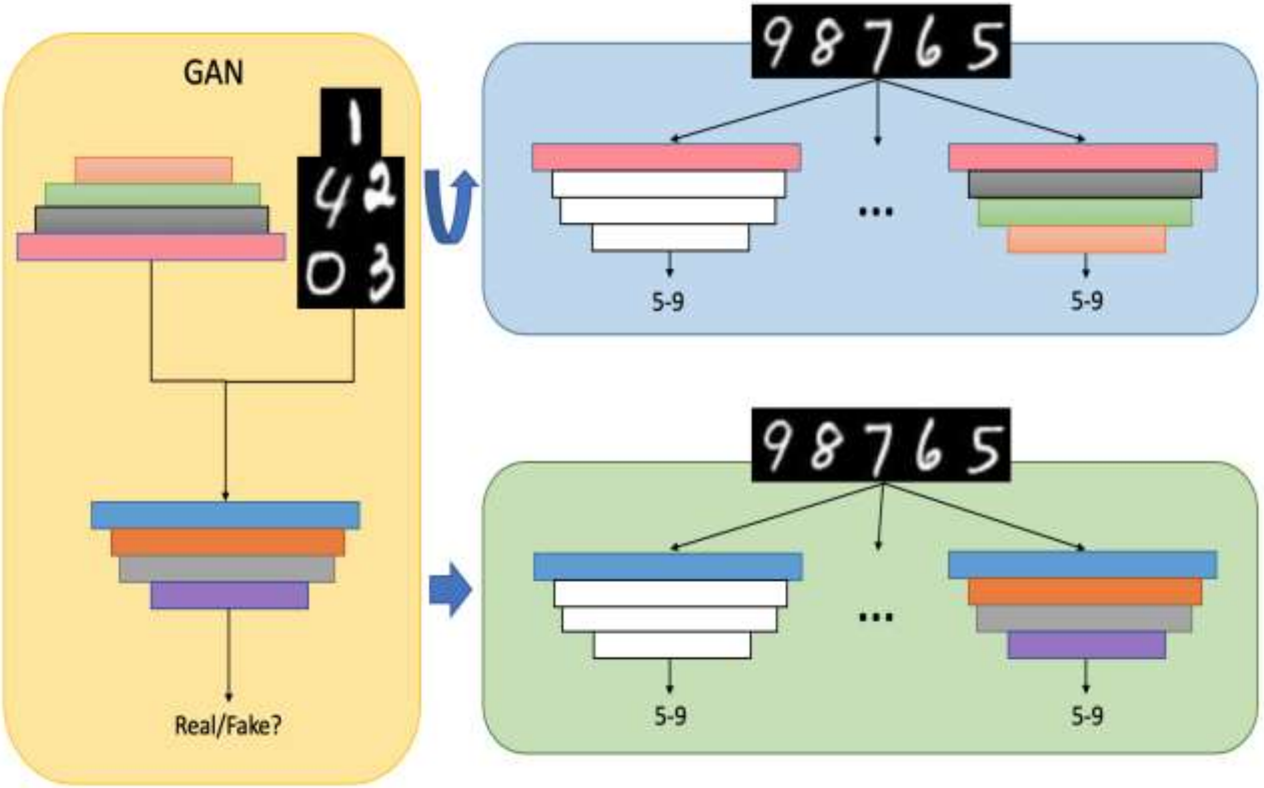
$$App_x = \frac{\sum_{j=1}^{z} \xi_j}{z}$$

# Contributions

- Explore the applicability of features between discriminative networks non-adversarially trained classification networks.
- Demonstrate the differences between the learned features in a discriminative and a classification process.
- Transferability of features to a GAN is judged by measuring the applicability of features to the generator and discriminator.

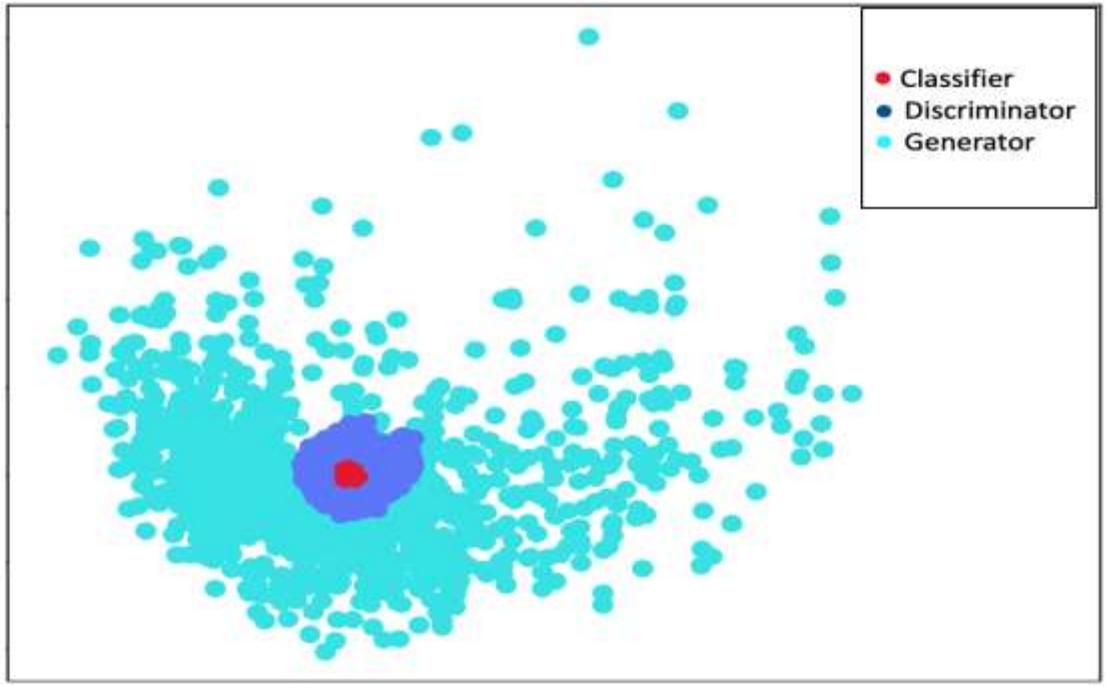
# Experimental Setup

- Classification is a good task to measure differentiation and thus applicability.
- Applying the features learned by the generator and discriminator to classification allows for measuring their applicability at the different levels.

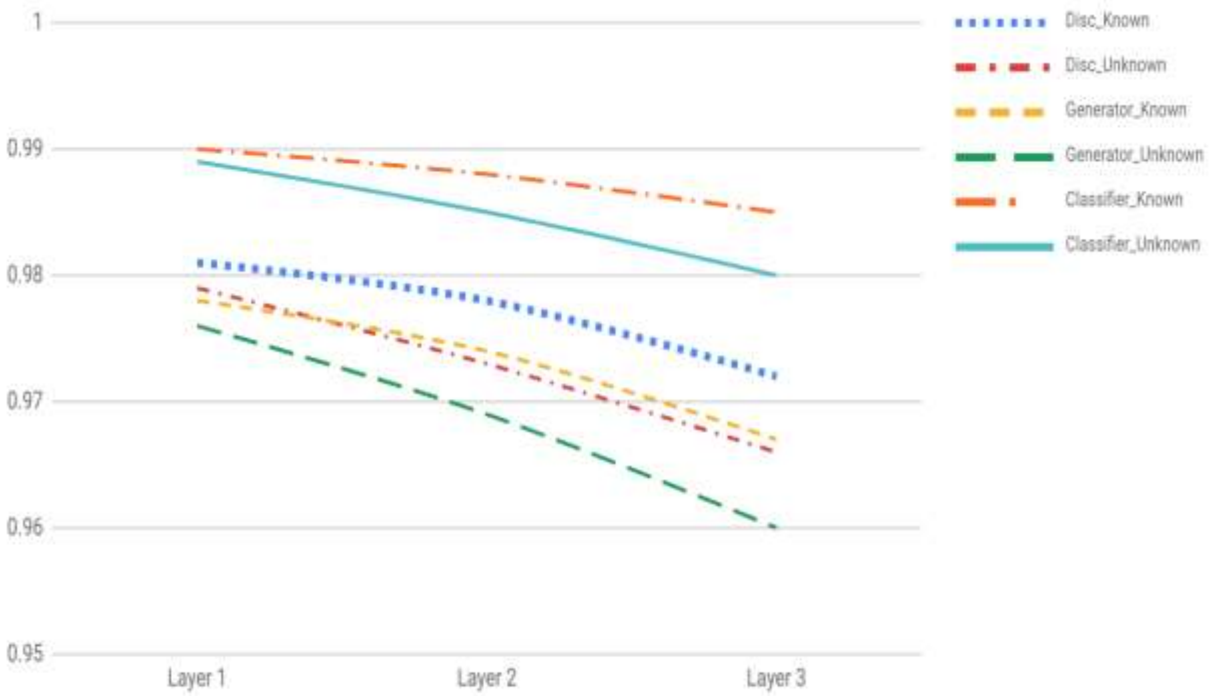


# GAN Set Applicability

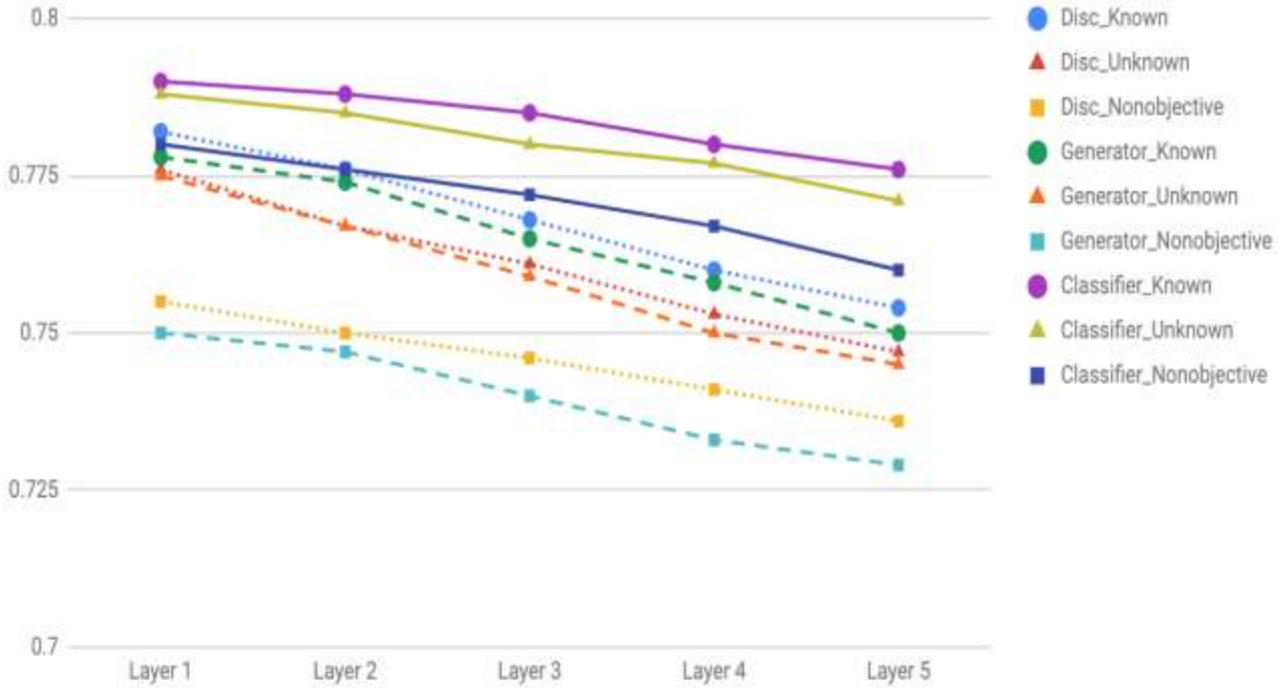
- Apply the learned weights from both Generator G and Discriminator D to classification of the unknown data set.
- Weights are transferred layer by layer.
- The set applicability for each network is then the ability to use those features at each layer to differentiate known from unknown.



# GAN Class Applicability

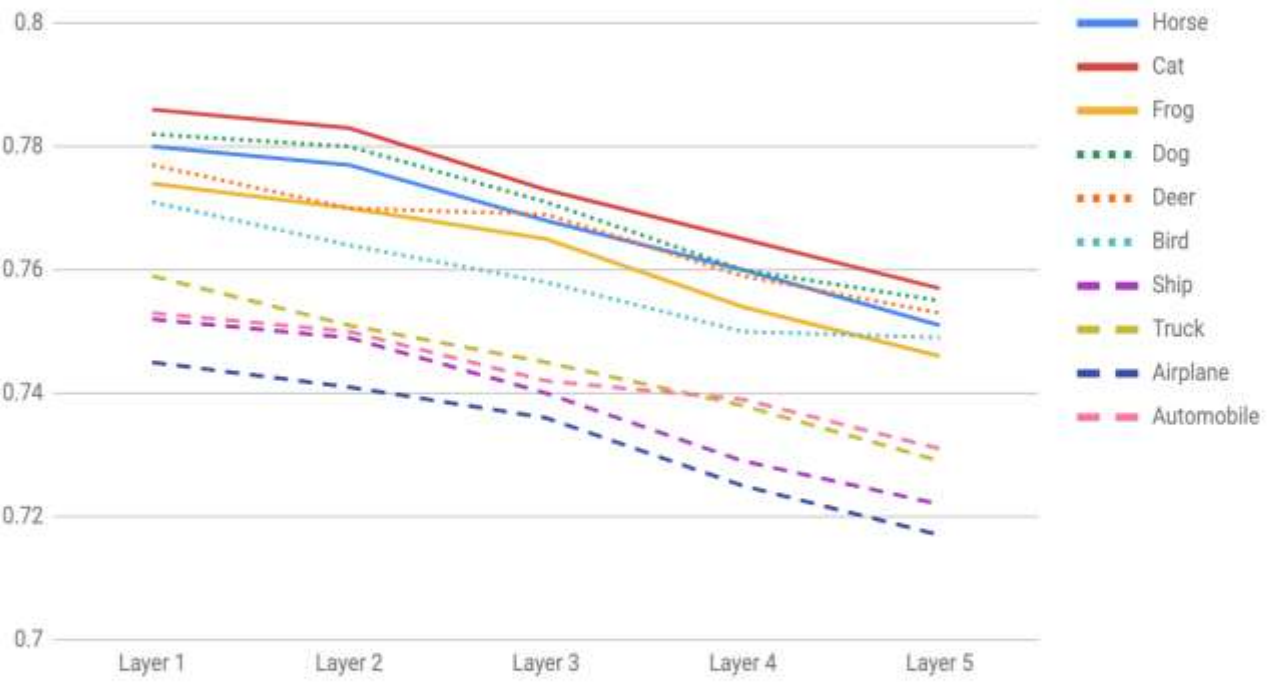



MNIST

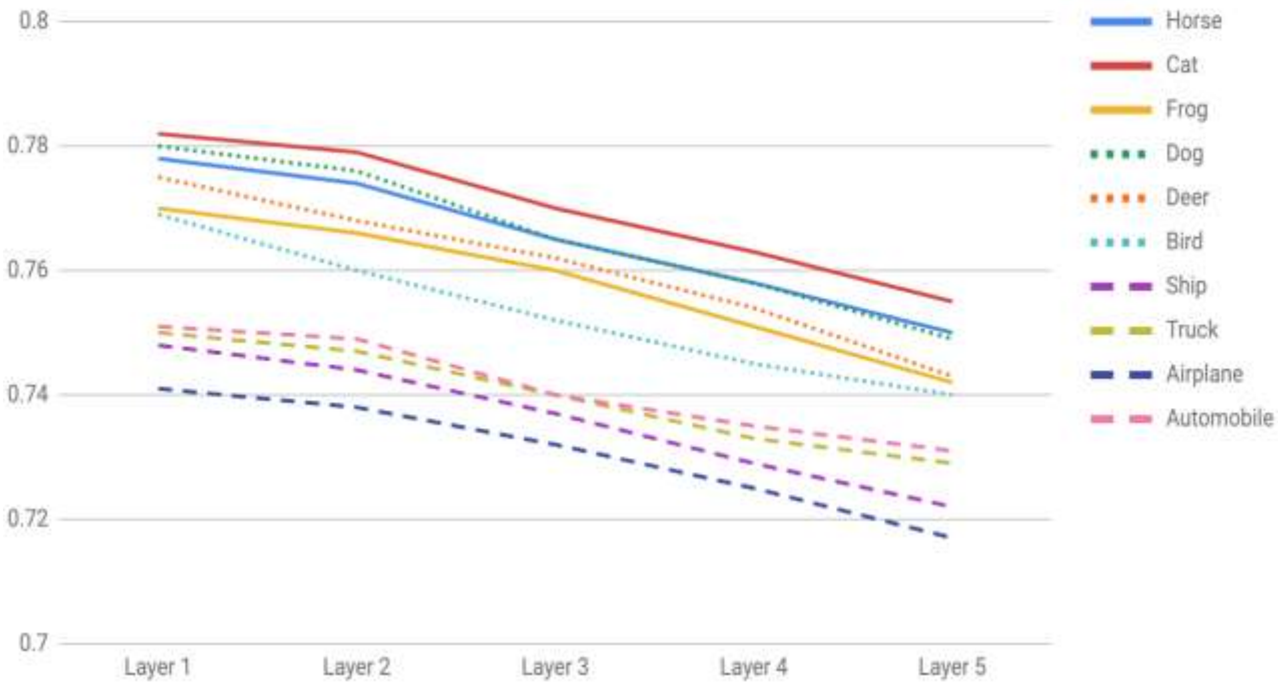



CIFAR

# GAN Layer Applicability



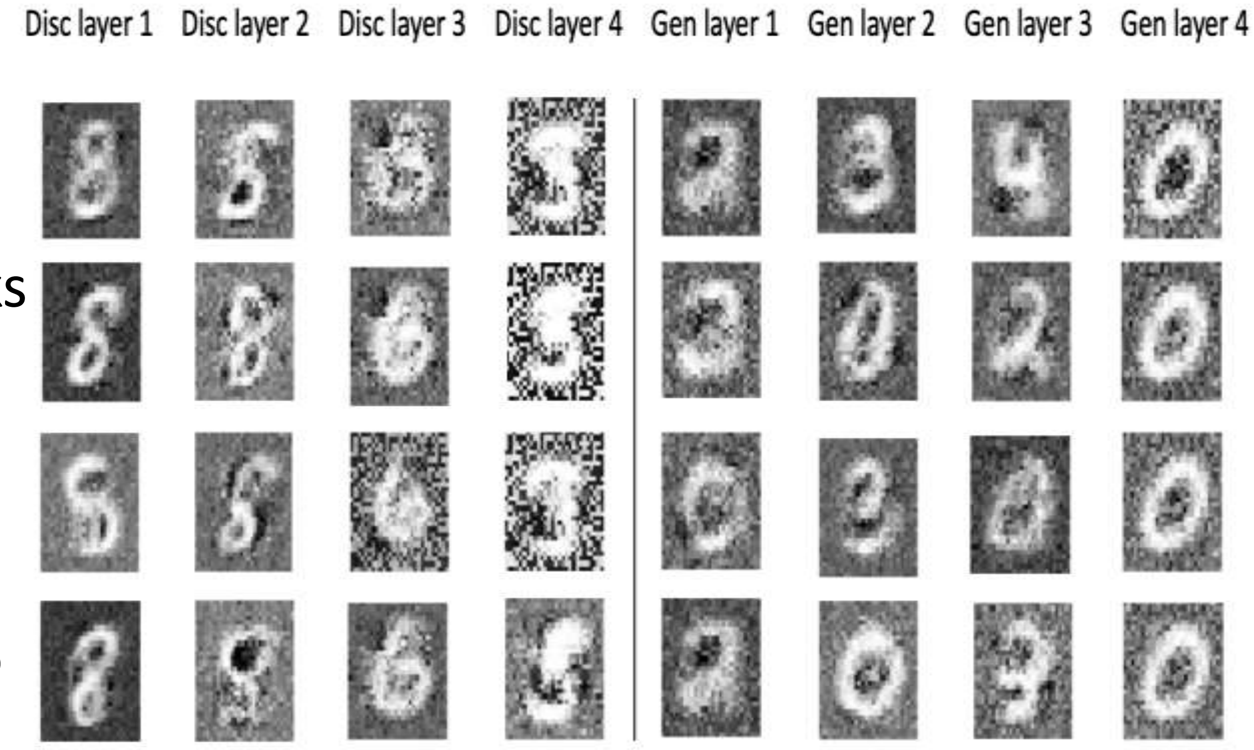

Discriminator




Generator

# GAN Transfer

- We have seen that there is high applicability between the generator and discriminator for similar tasks.
- Given the relationship between the tasks it goes to reason that they learn similar features and apply them differently.
- We swapped roles of the generator and discriminator to measure how well the learned features could be transferred to the opposite task.



# Conclusions

- Presented, to the best of our knowledge, the first results on evaluation of feature applicability and transferability in generative adversarial networks.
- Demonstrated a discriminator and a generator can both be applicable to classification tasks.
- Provided resented insights into how applicable they are to a classification task from both a set and class applicability perspective.

# Thank You!

LSU