

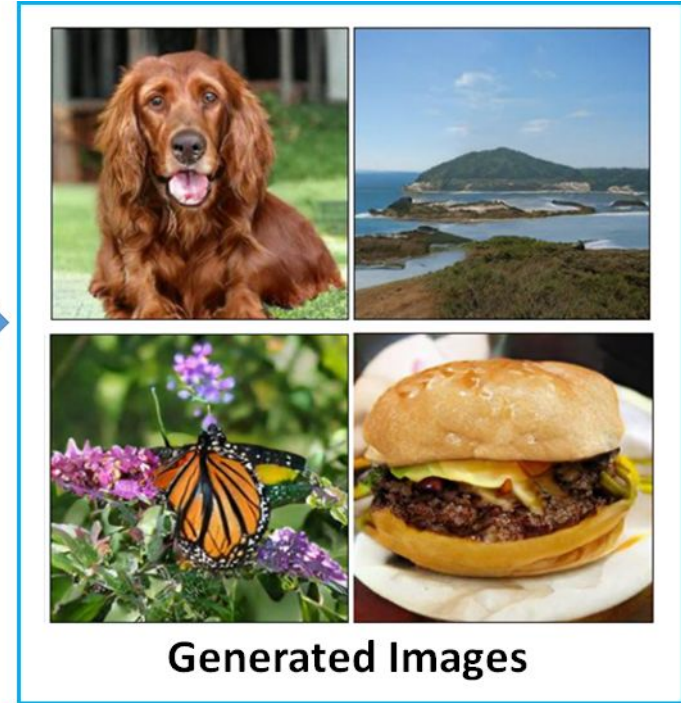


Bias and Generalization in Deep Generative Models

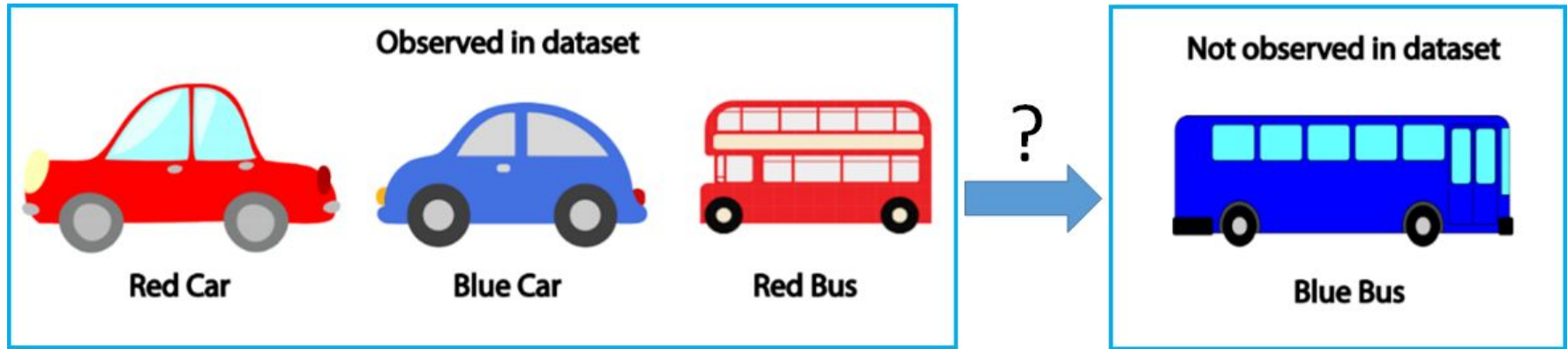
Shengjia Zhao*, Hongyu Ren*, Arianna Yuan, Jiaming Song,
Noah Goodman and Stefano Ermon

*equal contribution

Success in Generative Modeling of Images



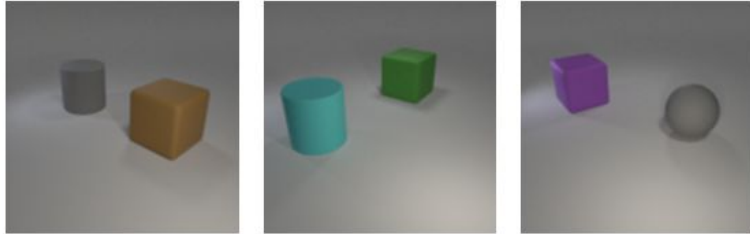
Goal: Understanding Generalization



How do generative models generalize?

Generalization Example: Object Count

All training images have 2 objects



2

2

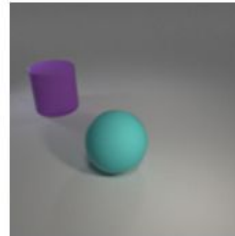
2

?

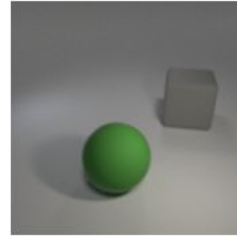


How many are in the generated images?

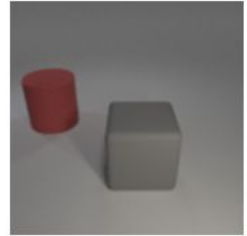
A



2

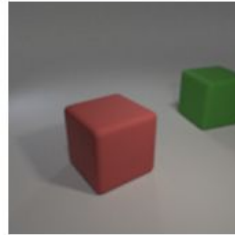


2

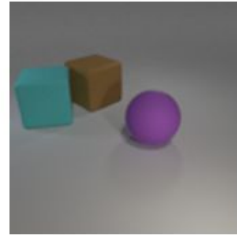


2

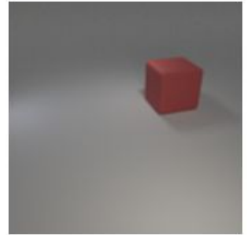
B



2



3



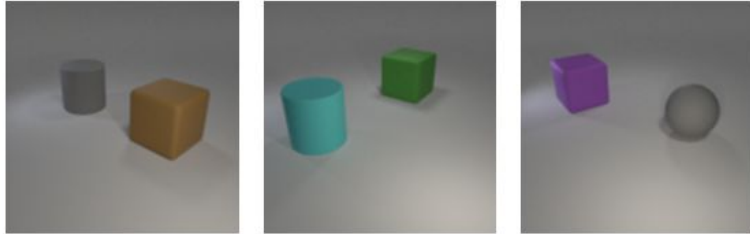
1

Empirical Study of Generalization: Method

- Design datasets
- Train generative models (VAE, GAN, PixelCNN)
- Observe generalization behavior
- Find common patterns

Generalization Example: Object Count

All training images have 2 objects



2

2

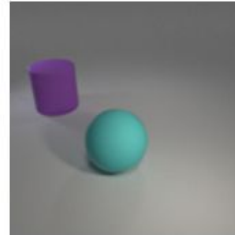
2

?

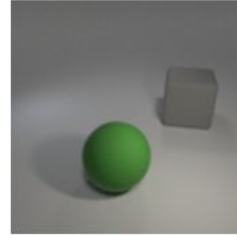


How many are in the generated images?

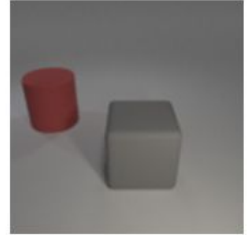
A



2

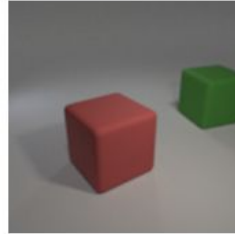


2

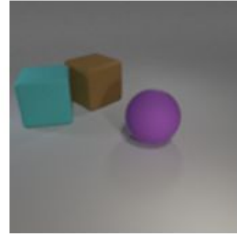


2

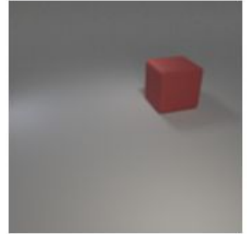
B



2

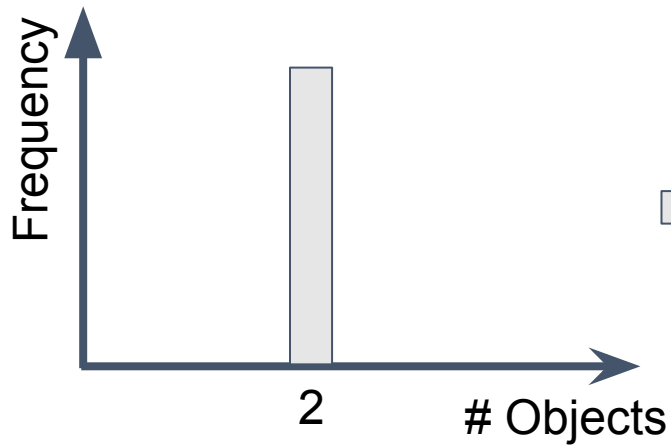


3

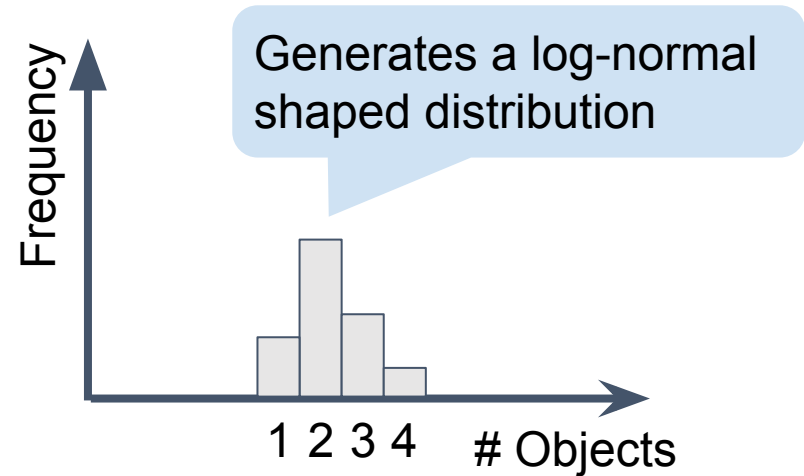


1

Generalization in Feature Space: Numerosity

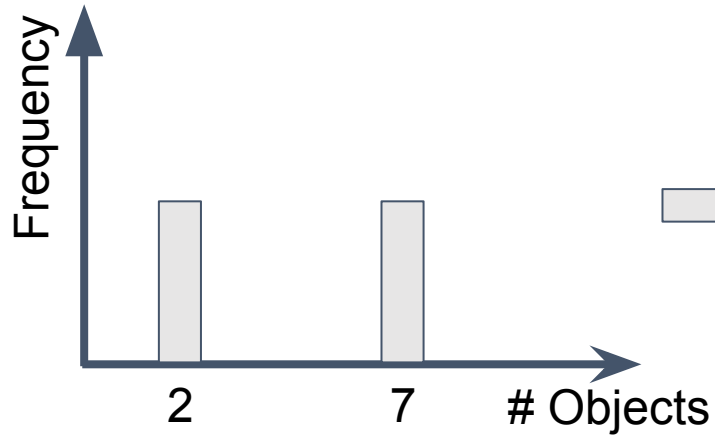


Training Distribution



Generated Distribution
(Observed)

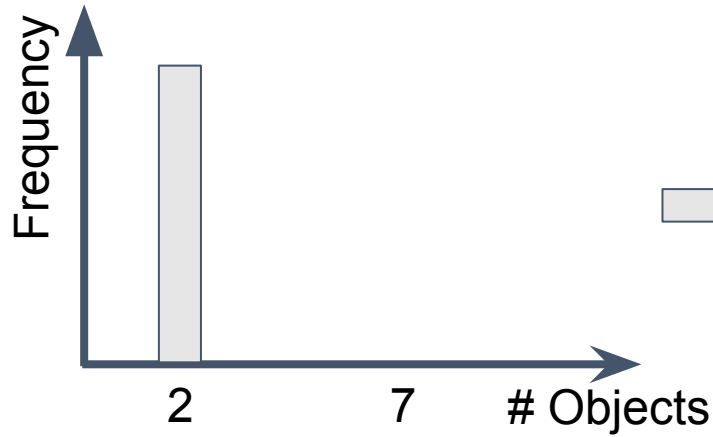
Multiple Numerosities



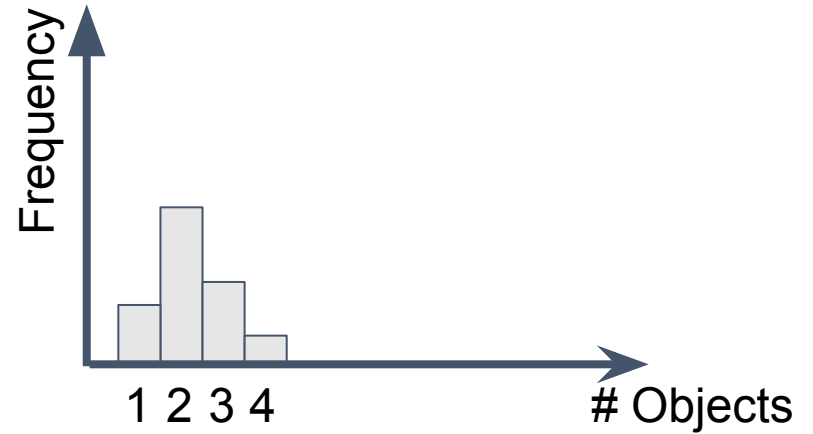
Training Distribution

?

Multiple Numerosities: Only 2

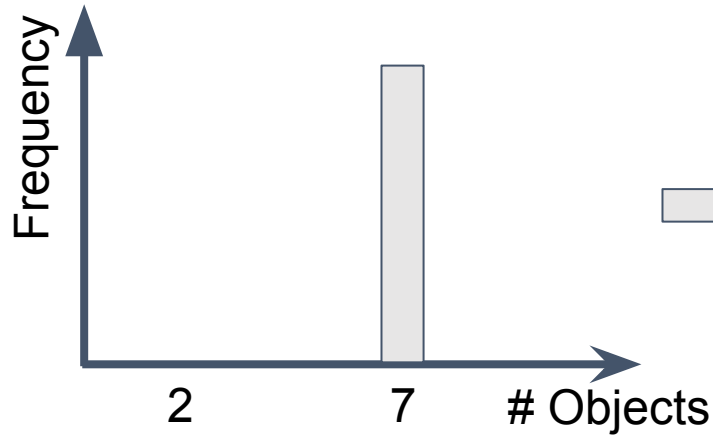


Training Distribution

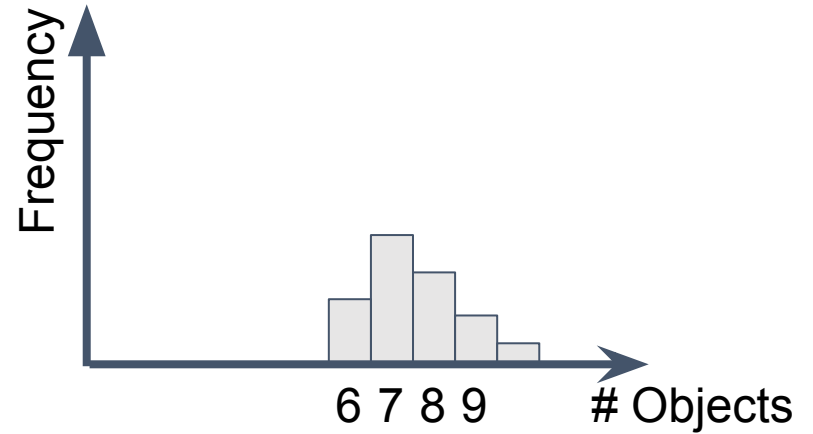


Generated Distribution

Multiple Numerosities: Only 7

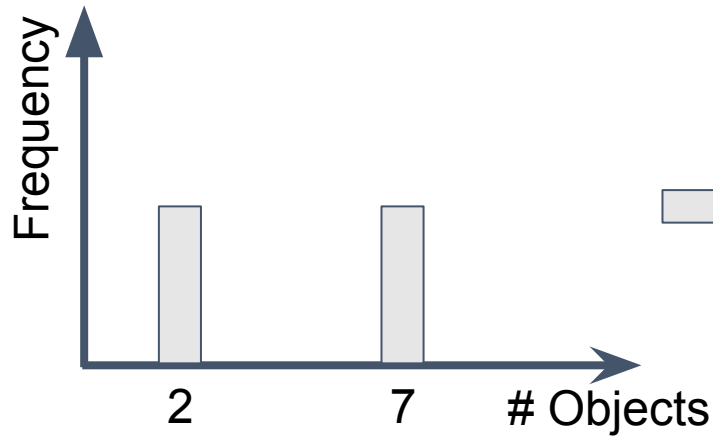


Training Distribution

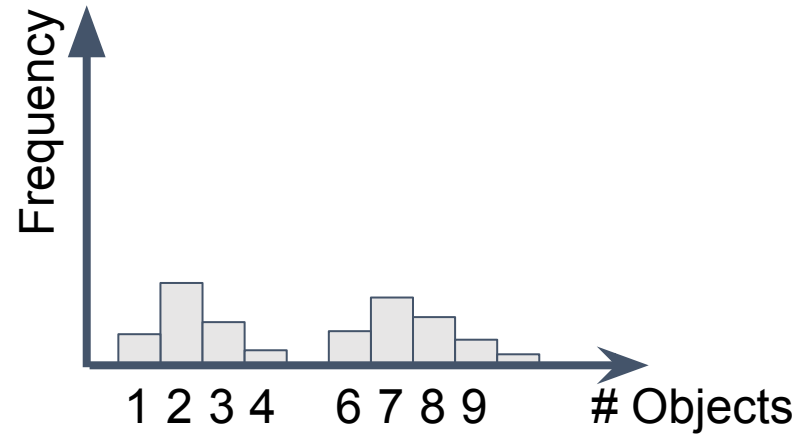


Generated Distribution

Multiple Numerosities: Additive Hypothesis

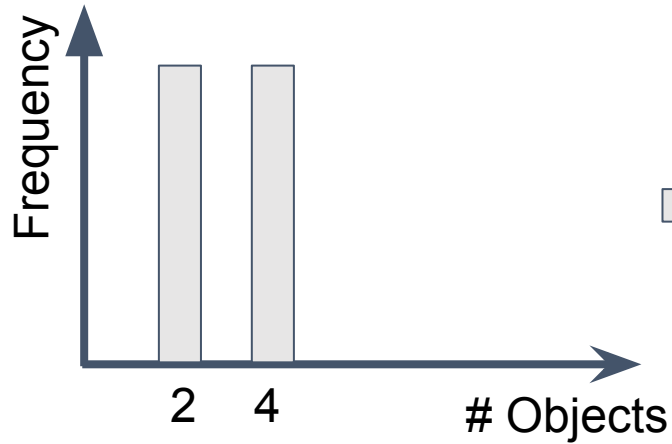


Training Distribution

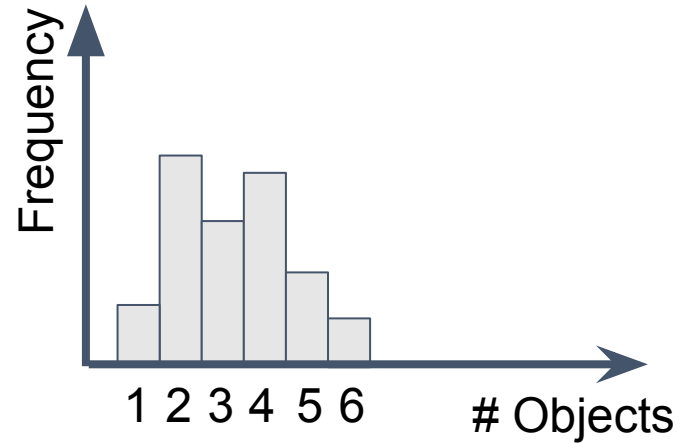


Generated Distribution
(**Observed**)

Additive Hypothesis with 2 and 4 Objects

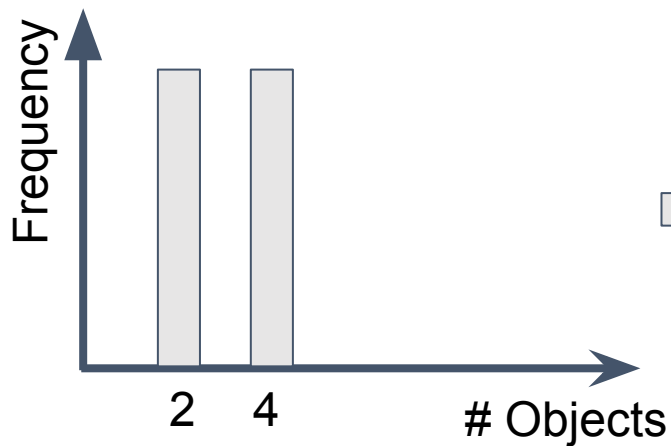


Training Distribution

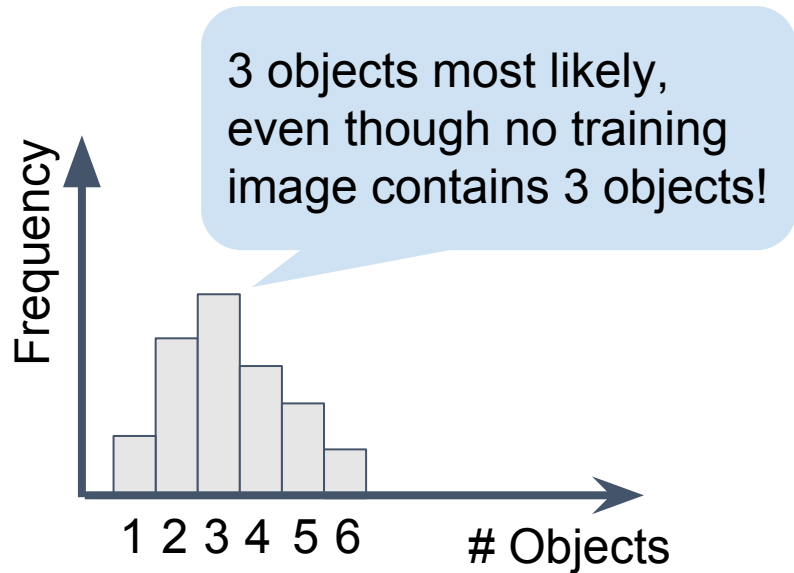


Generated Distribution
(**Hypothesized**)

Actual Result: Prototype Enhancement

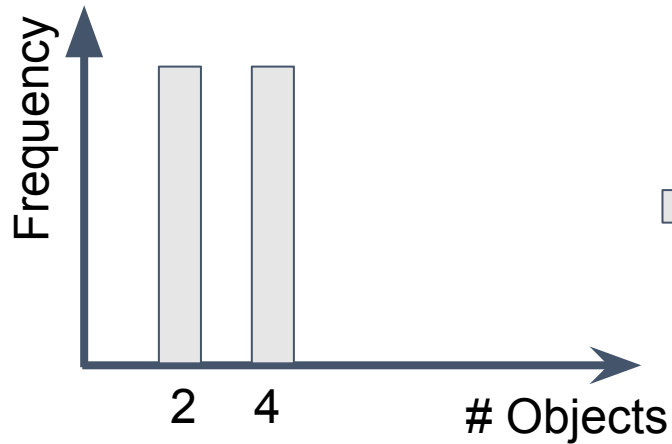


Training Distribution

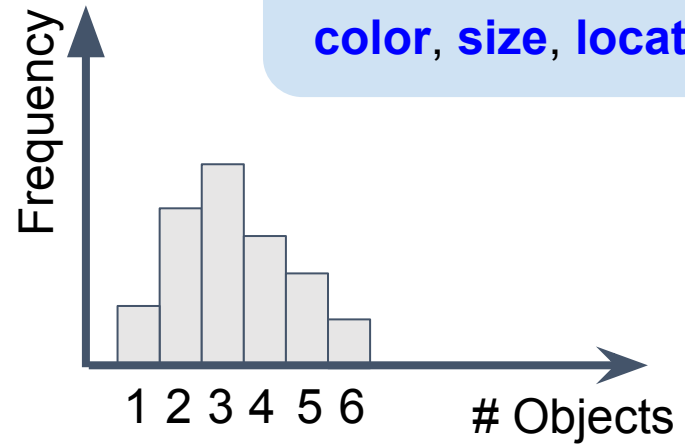


Generated Distribution
(**Observed**)

Prototype Enhancement



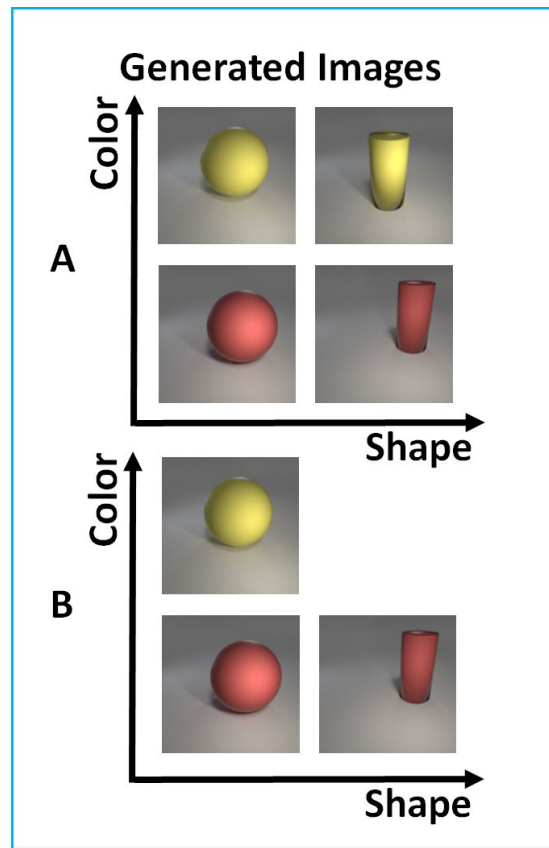
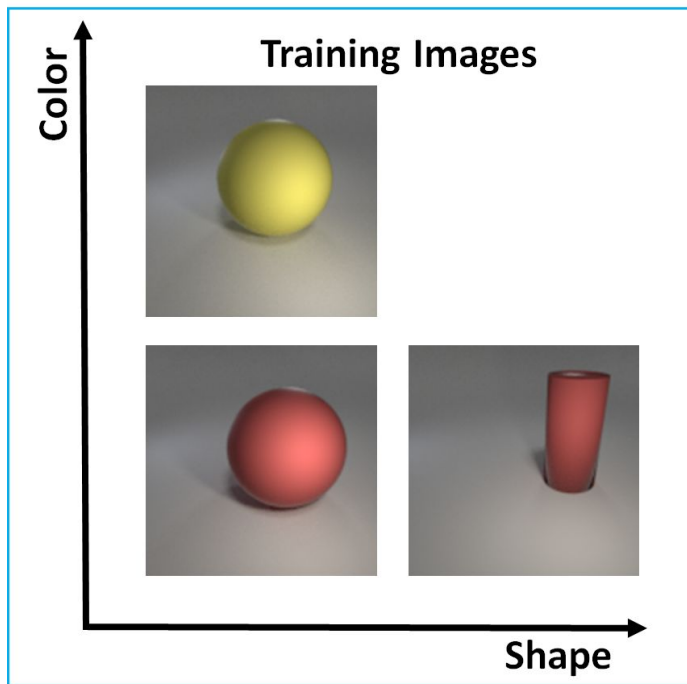
Training Distribution



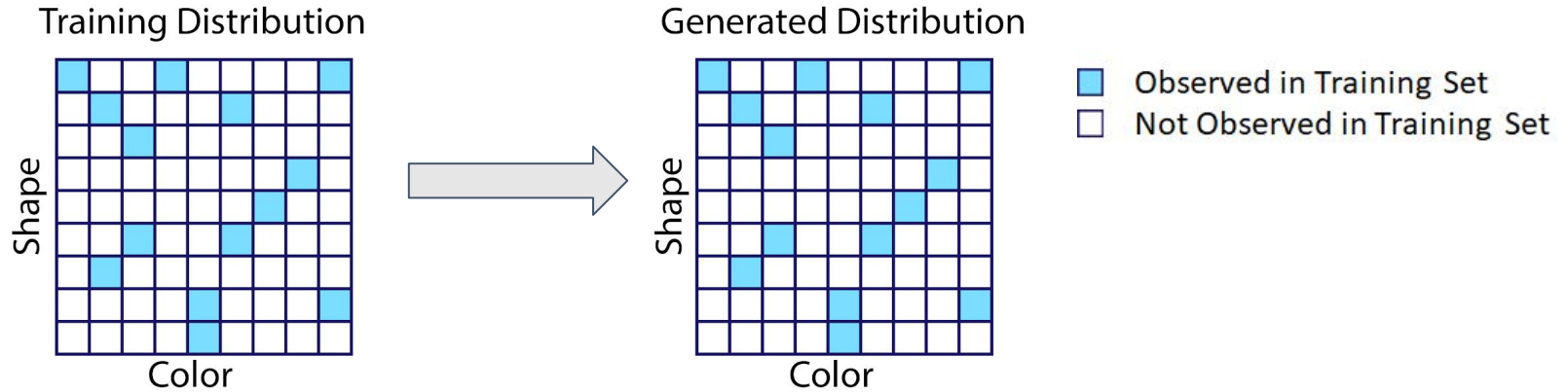
Generated Distribution
(Observed)

Similar pattern for
other features:
color, **size**, **location**

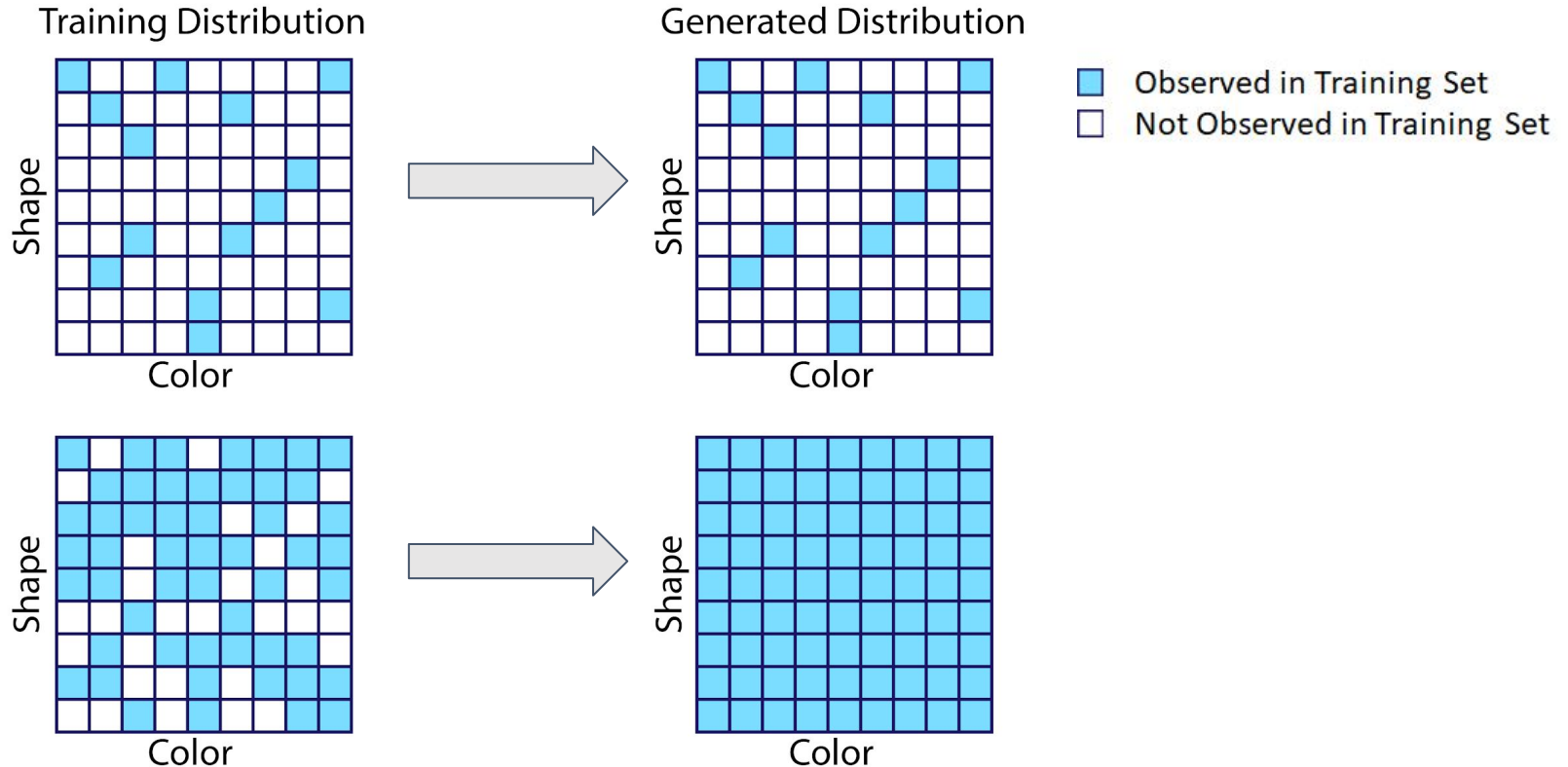
Multiple Features



Memorization vs. Generalization



Memorization vs. Generalization



Different Setups, Similar Results

- Different features (shape, color, size, numerosity, etc.)
- Different models: (VAE, GAN, PixelCNN, etc.)
- Different architectures (fully connected, convolutional, etc.)
- Different hyper-parameters (network size, learning rate, etc.)

Conclusion

- New methodology: design datasets to probe generative models
- Observed common patterns across different setups

Welcome to our poster session for further discussions!

Tuesday 5-7pm @ Room 210 & 230 AB #6

Code available at github.com/ermongroup/BiasAndGeneralization