



# Generating Class-conditional Images with Gradient-based Inference

Bowen Xu, David Acuna, David Duvenaud

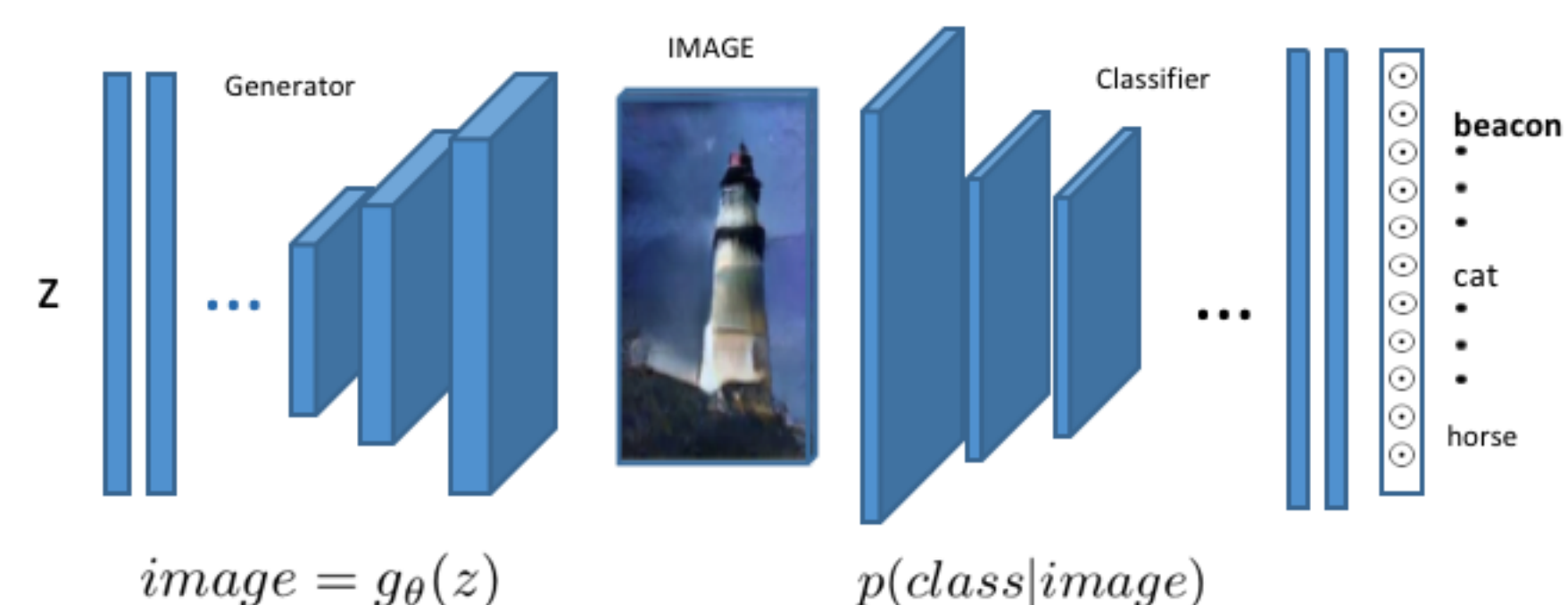
{xubo3, davidj, duvenaud}@cs.toronto.edu

## Main idea

- Gradient descent on images is a promising approach to generating class-conditional images.
- This method produces high-resolution class-conditional images with crisp details and coherent overall structure.
- Even better, we sample from  $p(\text{image}|\text{class})$  using gradient-based MCMC methods.
- Our approach produces realistic and content-diverse class-conditional images.
- Our method removes the need for *ad-hoc* tweaks to the objective when longer iterations are introduced.

## Previous Work

A previous approach by Nguyen et. al. [?] uses gradient-based optimization of images to do approximate MAP estimation, synthesizing large ( $227 \times 227$ ) class-conditional images.



- An initial image is specified by a random vector  $z \sim \mathcal{N}(0, I)$ .
- This vector is passed through a pre-trained image generation network  $g_\theta(z)$ .
- The generated image is fed into a pre-trained classification network.
- The latent vector is then optimized using backpropagation.

## Our contribution

We use gradient-based MCMC methods to approximately sample from the class-conditional posterior:

$$\hat{z} \sim p(z|\text{class}) \propto p(\text{class}|g_\theta(z))p(z) \quad \text{where} \quad p(z) = \mathcal{N}(0, I)$$

and  $p(\text{class}|g_\theta(z))$  is the classification network.

We adopt:

- Hamiltonian Monte Carlo (HMC)
- Metropolis-adjusted-Langevin-algorithm (MALA)

## Experiments and Results

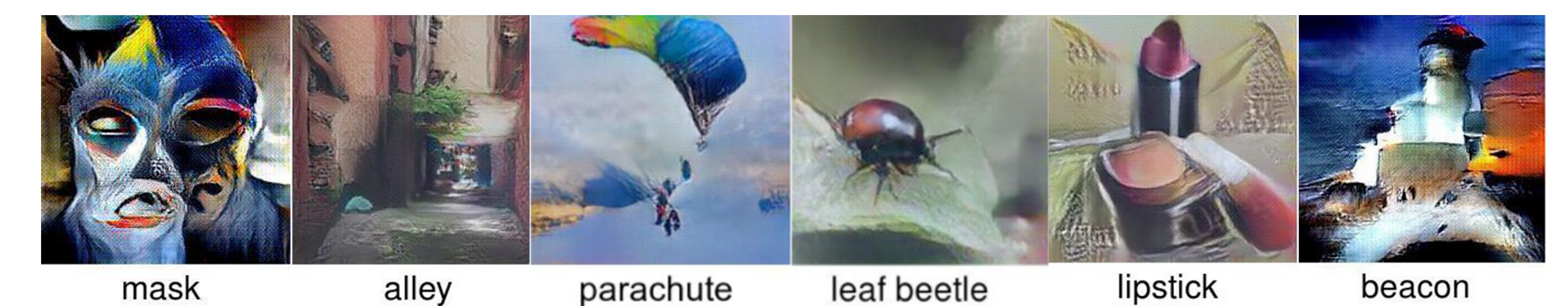


## Automatically Evaluating Image Quality

We use a discriminator [?], trained on ImageNet, to output the probability of an image being natural versus synthesized.

	MAP with class-specific regularization	MALA with class-specific regularization	MALA HMC
Average log-probability	-6.00	-5.53	-9.79 -7.96

## Diversity of Image Contents



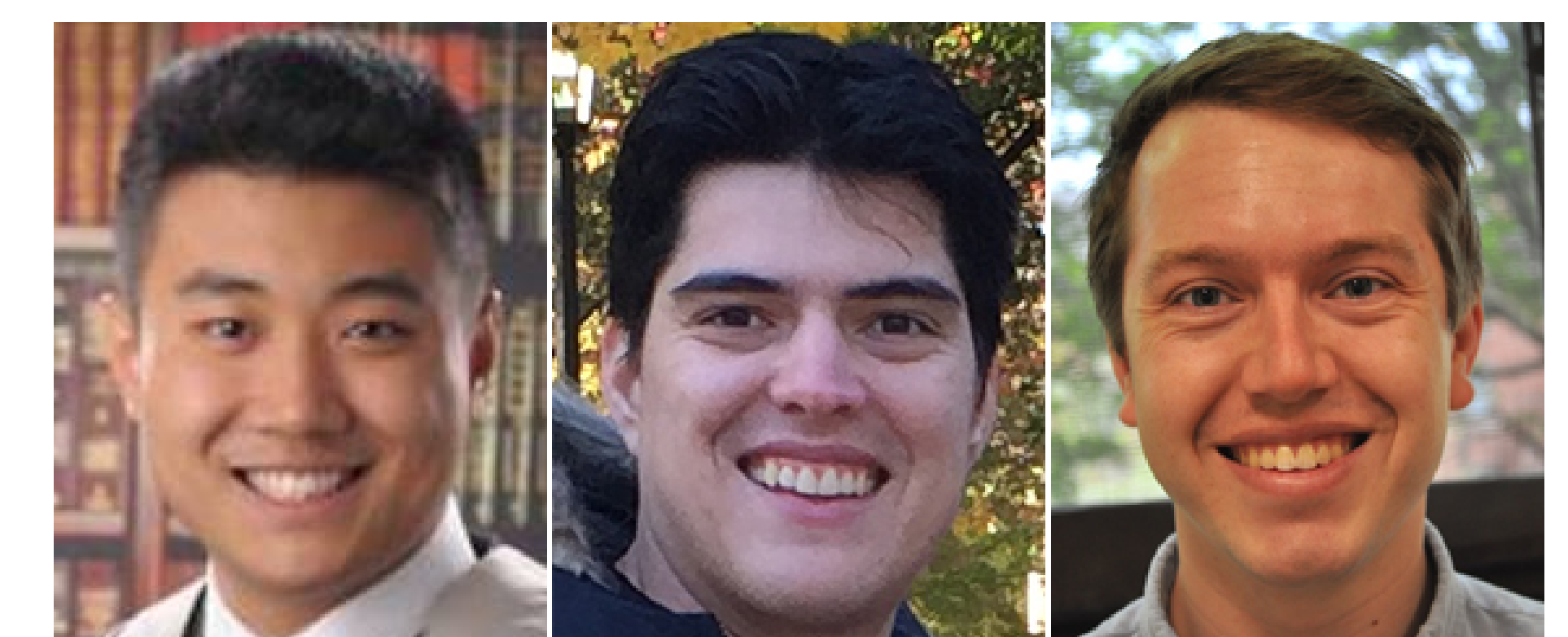
## Advantages and Limitations

- + Produces high-resolution images with crisp details and coherent overall structure.
- + Generates more content-diverse images.
- + Removes the need for class-specific regularization when longer iterations are introduced.
- Small step sizes and longer iterations are required.

## Future Work

- Use the Hamiltonian Variational Inference (HVI) model [?].
- Generate images from text.
- Generate other media such as sound and video.

## Bibliography



Bowen Xu

David Acuna

David Duvenaud