Generating Class-conditional Images with Gradient-based Inference

Bowen Xu, David Acuna, David Duvenaud

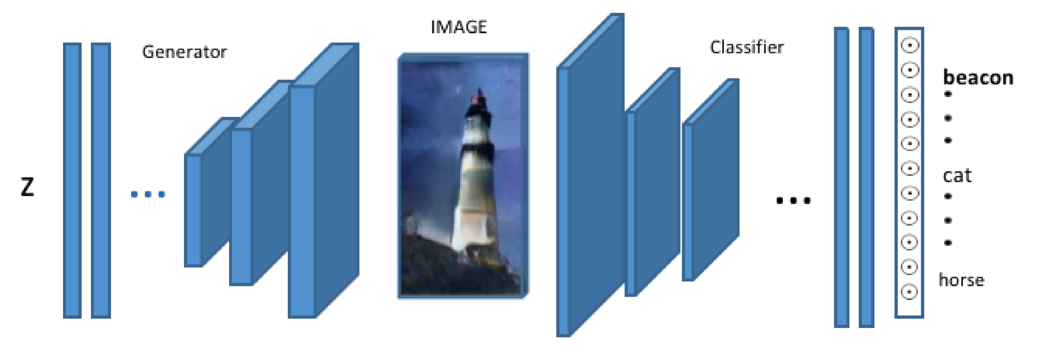
{xubo3, davidj, duvenaud}@cs.toronto.edu

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. [1] uses gradient-based optimization of images to do approximate MAP estimation, synthesizing large (227 x 227) class-conditional images.



$image = g_\theta(z)$ $\quad$ $p(class|image)$

- 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

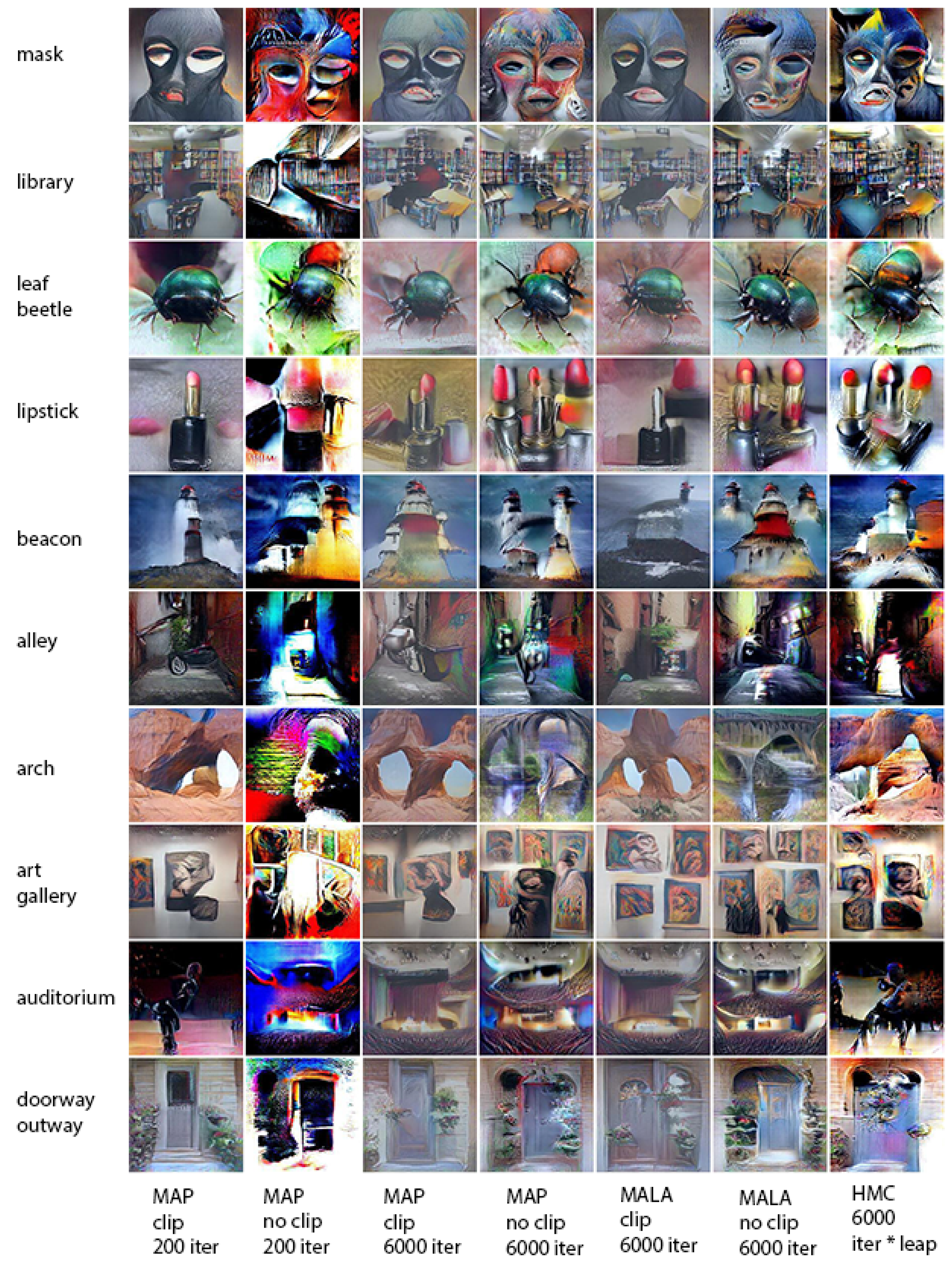
We use gradient-based MCMC methods to approximately sample from the class-conditional posterior:

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

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

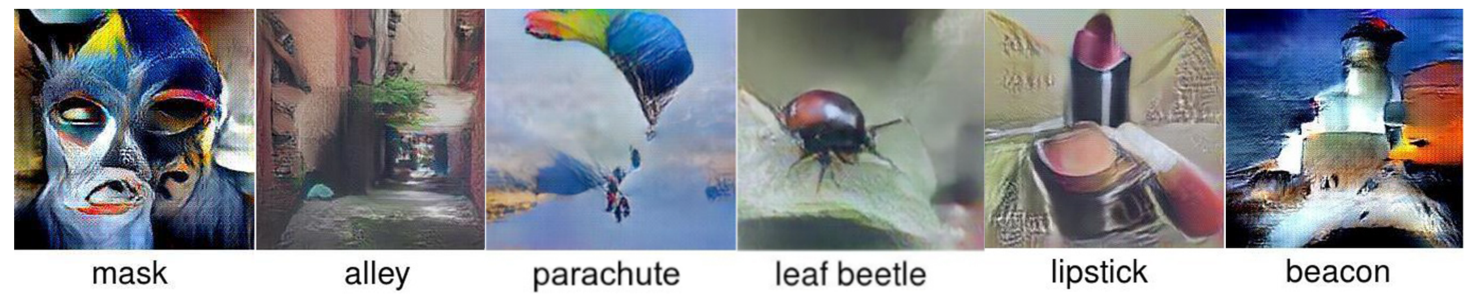


Automatically Evaluating Image Quality

We use a discriminator [2], 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

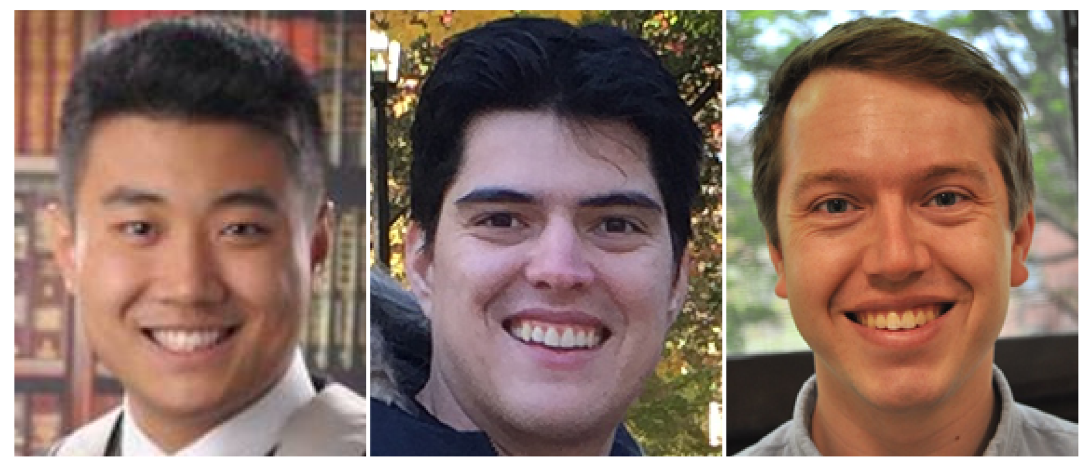
- Use the Hamiltonian Variational Inference (HVI) model [3].
- Generate images from text.
- Generate other media such as sound and video.

Bibliography

[1] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune, "Synthesizing the preferred inputs for neurons in neural networks via deep generator networks," *arXiv*, pp. 1–29, 2016.

[2] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv*, pp. 1–15, 2015.

[3] T. Salimans, D. P. Kingma, and M. Welling, "Markov Chain Monte Carlo and Variational Inference: Bridging the Gap," *ArXiv e-prints*, Oct. 2014.



Bowen Xu $\quad$ David Acuna $\quad$ David Duvenaud