Bayesian Image Completion with Deep Generative Models

Christopher Krapu

Duke University Department of Civil & Environmental Engineering, Department of Statistical Science

Problem statement

Machine learning algorithms are able to generate realistic completions of images with missing regions for increasingly complex images. **However, the problem of generating a diverse set of possible completions has not received as much attention**. Our application of interest is in the environmental sciences; given a 2D topographical map of the Earth's surface (equivalently, an image with a single channel), with missing patches, we want to estimate the posterior distribution over the missing data.

We desire an approach which satisfies the following requirements:

- **Complex image features** consistent with nearby regions of data are represented in completions
- The image completions are drawn from a **coherent probability model**
- A range of **varied features are present** in the posterior completions

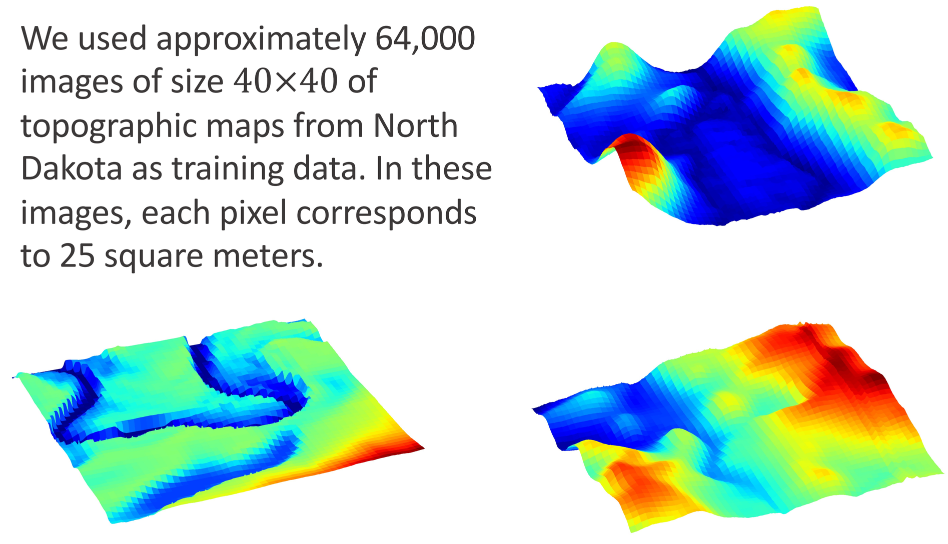
Methods

We employ the following approach to draw posterior samples of completed images:

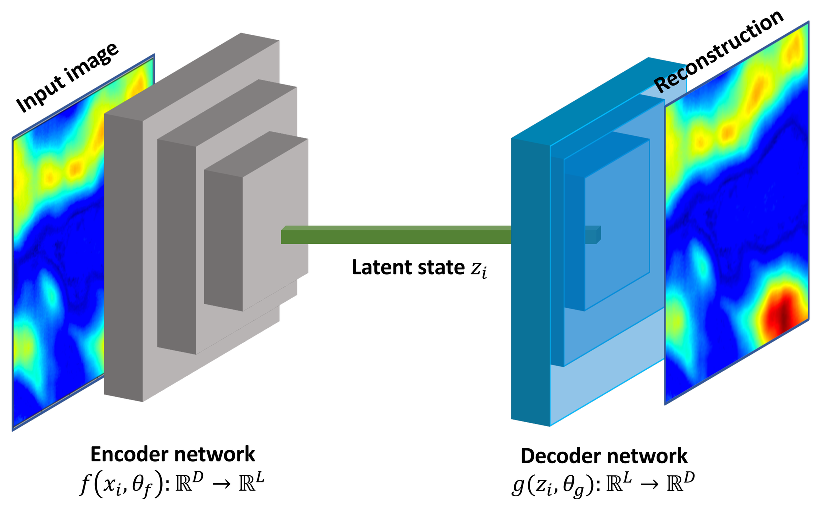
1. We **train a convolutional variational autoencoder** (VAE) in Keras and PyMC3 to learn a compressed latent state representation z_i of each image x_i.
2. For each image with missing regions, we **condition on the observed portion** to define a posterior density over the per-image latent state. This density is calculated using the decoder half of the VAE.
3. We use Hamiltonian Monte Carlo, a version of Markov chain Monte Carlo (MCMC), to **sample from the high-dimensional density over latent states** in PyMC3.
4. The image completions are generated **by passing the latent state samples back through the decoder** network.

Data

We used approximately 64,000 images of size 40×40 of topographic maps from North Dakota as training data. In these images, each pixel corresponds to 25 square meters.



Variational autoencoder



An autoencoder is a network designed to learn a **low-dimensional representation of the data** by forcing its connections through a small bottleneck. In the process, a *generative model* is learned via the decoder network which, if trained correctly, can **generate plausible exemplars of complex images** given samples drawn from the latent state's prior. Here, we assume an isotropic multivariate normal prior $z_i \sim MVN(0, I)$ over the latent states. We use a variational Bayesian approximation to fit the neural network's parameters.

Updating the prior

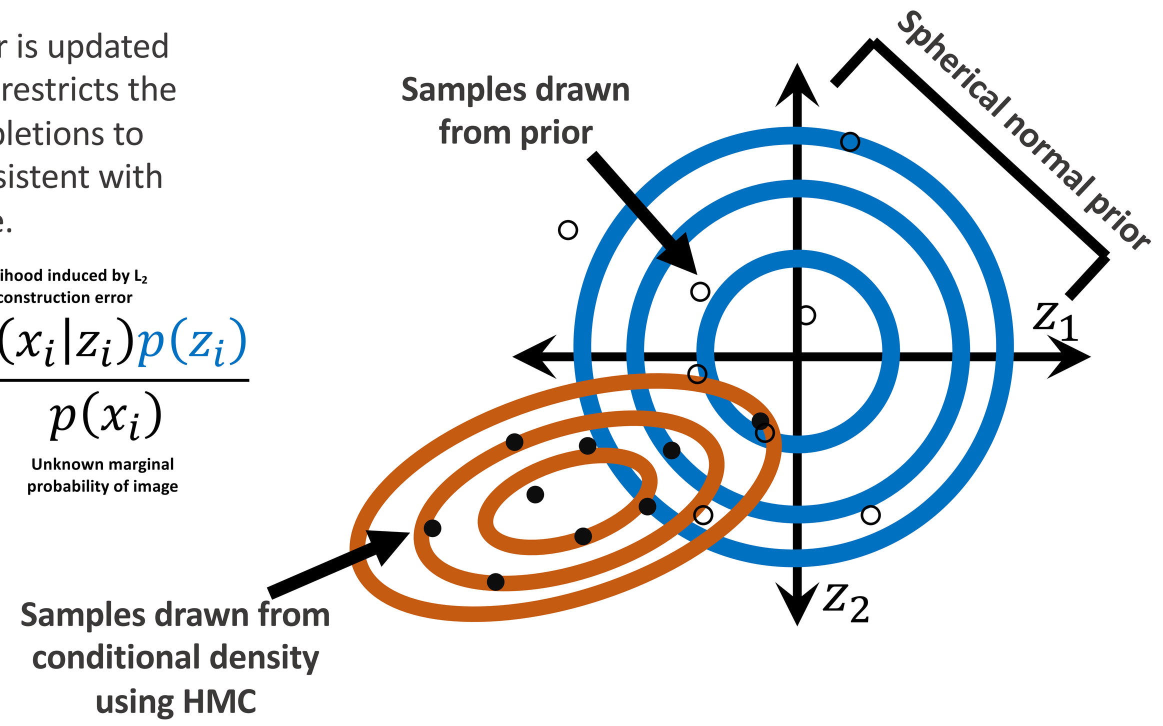
Our latent state prior is updated to a posterior which restricts the possible image completions to those which are consistent with the rest of the image.

Likelihood induced by L_2 reconstruction error

$$p(z_i|x_i) = \frac{p(x_i|z_i)p(z_i)}{p(x_i)}$$

Target posterior density for MCMC

Unknown marginal probability of image



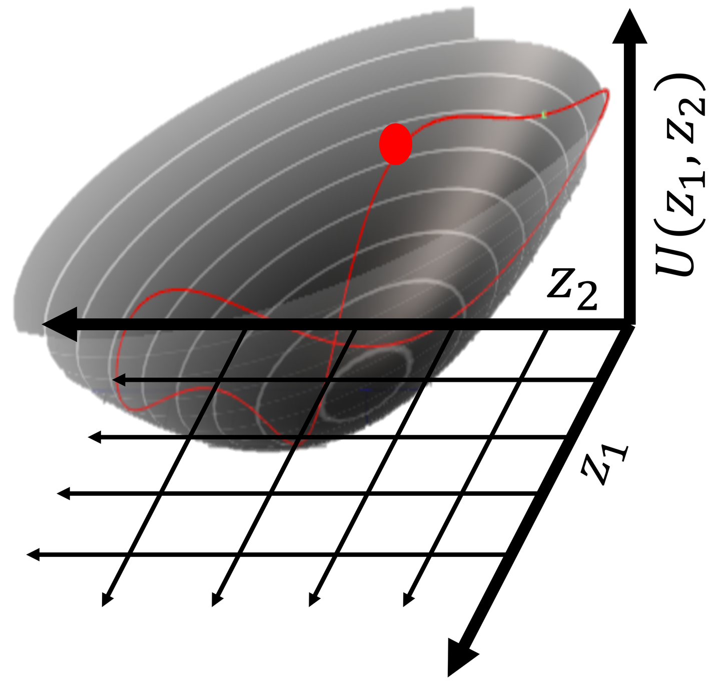
Hamiltonian MC

- HMC is a version of Markov chain Monte Carlo that makes more efficient proposals by incorporating knowledge of the posterior geometry via the gradient of the its density.
- In our work, **we fixed the neural network parameters** and used HMC to sample the latent states.
- By randomizing the initial velocity of the sampler**, different levels of $p(z_i|x_i)$** will be explored.

$$\underbrace{\nabla_{z_i} U(z_i, \theta_g)}_{\text{Gradient of log likelihood with regard to latent states}} = \underbrace{\nabla_{z_i} \log p(z_i)}_{\text{Gradient of latent state prior}} + \underbrace{\nabla_{z_i} \|g(z_i) - x_i\|_2^2}_{\text{Gradient of per-pixel squared error loss}}$$

Location update: $z_{new} = z_{old} + \dot{z}_{old} \cdot dt$

Velocity update: $\dot{z}_{new} \propto \dot{z}_{old} + \nabla_z U$



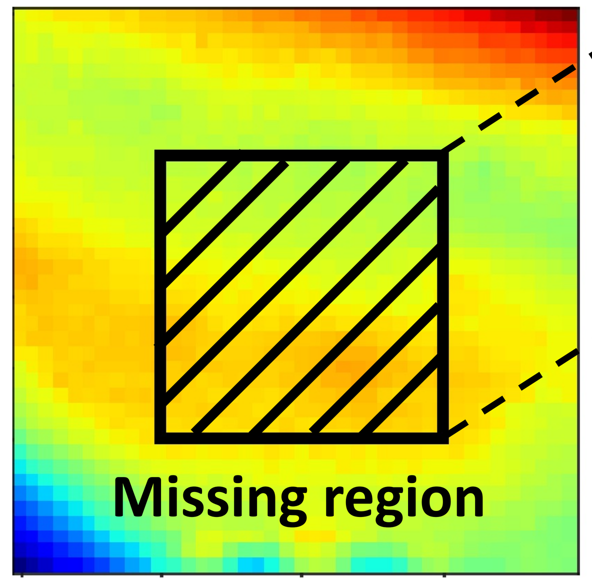
Limitations

- **Drawing samples is slow**, even with GPU acceleration; upwards of 2-3 seconds is frequently required to draw a single sample per image.
- This combination of pointwise network training and posterior sampling of latent states **does not allow for a full quantification of uncertainty due to network parameters.**

Results

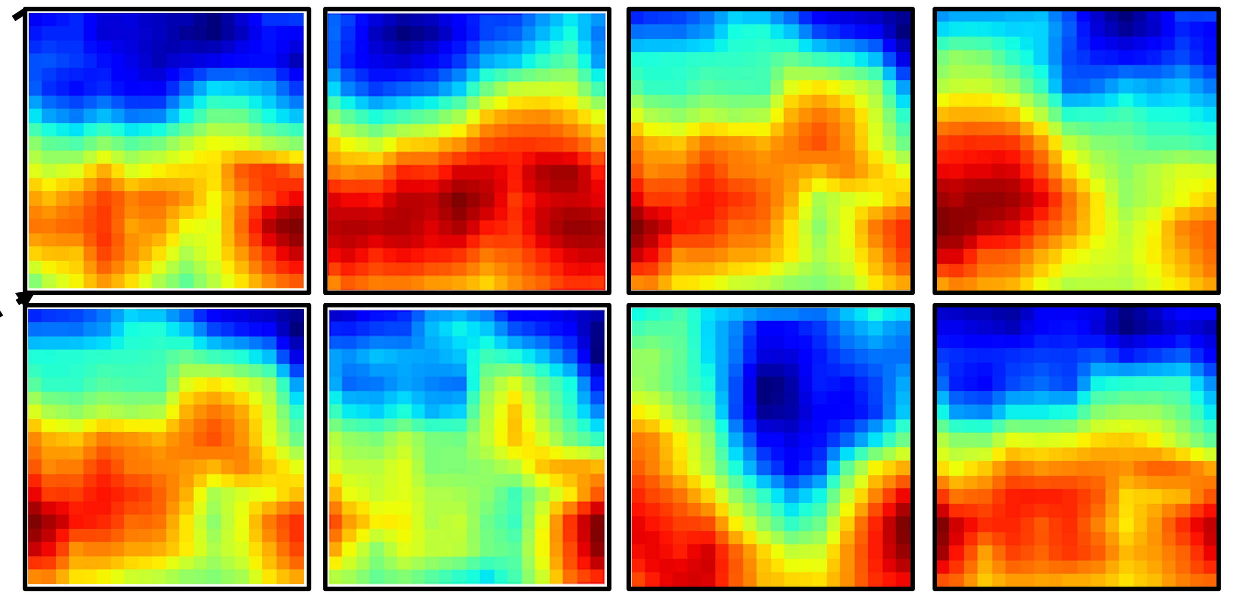
- For images in the held-out test set, we were able to generate a diverse set of posterior estimates of infilled regions.
- We also conducted evaluations of posterior **coverage probabilities** of several of the image completions and found these to be within 5—10% of nominal probabilities.
- Small-scale landscape features such as raised berms or narrow ditches were not captured in the posterior completions.

Original image



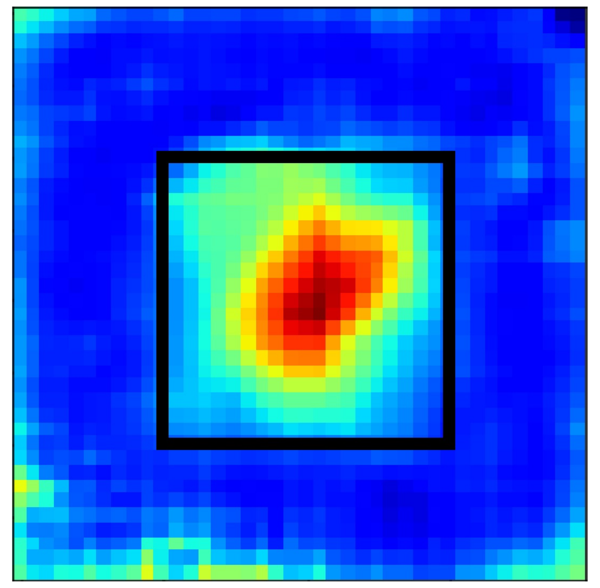
Sampled completions

(Colors rescaled for contrast)



Posterior variance

Warmer colors indicate higher uncertainty



Acknowledgements

This work was funded by NASA through an Earth and Space Sciences Fellowship as well as by the Duke Wetland Center. Computational resources and hardware were provided by grants from Nvidia and Amazon Web Services.

NASA

Duke NICHOLAS SCHOOL OF THE ENVIRONMENT