

1 Description of the Problem

Artificial neural networks (ANN) are inspired by their biological counterparts. One function of biological neural networks is their ability to store and retrieve memories. What makes them unique is the fact that these memories are stored within the network structure. In comparison, artificial neural network structures approximate functions using their parameters, rather than saving data in them.

We wish to design an ANN that also saves memories in its structure. The way we've thought of accomplishing this is through the use of an invertible variational autoencoder. Because it's invertible, it will learn a bijective map that allows a "memory" to be almost perfectly reconstructed. The reconstruction will primarily use information from the network structure. Because the ANN is invertible, its output can be fed backward through the network instead of through a separate decoder.

This problem is interesting for many reasons. We would like to try playing with the latent space variables to create "false memories" (a common task with autoencoders, like creating fake celebrity pictures). We are also interested to see if this can be used as a form of data compression. We were able to find a paper from 2019 where the authors implemented an invertible autoencoder, and it showed promising results (see [\[1\]](#)). We believe we can improve their results through use of different loss functions and by varying the architecture. We are also interested in the information-theoretic applications of an invertible autoencoder. One question we are curious about is if we can construct some information-theoretic norm such that the INN output is continuous with respect to its latent space variables.

2 How will Deep Learning Be Used?

Invertible neural networks (INNs) and variational autoencoders (VAEs) are deep learning architectures that have been the focus of much study. We will design, implement, and train an INN-VAE hybrid architecture.

3 Data

We will use CIFAR-10 and CelebA to train and test our INN-VAE. We won't need to collect any data on our own, as we will only use these easily available datasets.

4 Training

We hope to train our INN-VAE using Google Colab. If Colab is insufficient, Taylor has access to the BYU supercomputer, so we can use the GPUs there.

5 Distribution of Labor

This will be a group project between Jake Callahan and Taylor Paskett. Taylor will train the basic models (similar to the structure from [\[1\]](#)). Jake will train our improved models. We will work together on the theoretic question.