SpatialGAIN: Spatial Generative Adversarial Imputation Networks and its Application to Zero-Shot Pixel Imputation

Justin Young
justiny@stanford.edu

Problem Statement

Pixel Imputation

Existing Methods

● Leading approaches are successful but computationally very expensive to train (e.g. Vision Transformers, He et al. [2021])
● Deterministic imputations do not allow for inference (e.g. DAE)
● Zero-shot methods that use no training data are inflexible.

Goals

We propose Spatial Generative Adversarial Networks (SpatialGAIN):
1. Computationally Feasible
2. Yields Sampling Distribution of Imputations
3. Zero-shot methods that use no training data are inflexible.

SpatialGAIN

Same set-up but extends framework to image data by using CNNs, a different masking matrix, and slightly altered hint generator.

Methods

Generative Adversarial Networks (GAIN)

Shown as a improvement to a wide array of discriminative and generative models (Yoon et al. [2018]). Allows for incomplete training data, unlike autoencoder models!

SpatialGAIN

Same set-up but extends framework to image data by using CNNs, a different masking matrix, and slightly altered hint generator.

Experiment 1

Baseline. We compare GAIN to SpatialGAIN to see whether we reap benefits on image vs. tabular data. This provides a first sanity check.

Dataset. We use UCI’s Spambase dataset and Imagenette, a downsampled version of ImageNet. We perform a 80/20 train/test split on 4601 observations from Spambase (tabular data) and 1350 160x160 grayscale images from Imagenette (downsampled ImageNet).

Methods. In order to use SpatialGAIN on tabular data, we artificially reshape tabular data. In order to use GAIN on image data, we flatten the image into a vector.

Analysis. Significant degradation of GAIN on image data while that of SpatialGAIN on tabular data is relatively modest. From a relatively small training set (<1000), this suggests SpatialGAIN is able to learn spatial dependencies and what constitutes a fish and its salient surroundings.

Note from the heat map that SpatialGAIN struggles at borders.

Future Work

and its Application to Zero-Shot Pixel Imputation

Experiment 2

Baseline. In the zero-shot setting, we compare SpatialGAIN to matrix completion to see whether a deep learning model can work.

Dataset. We fix one randomly chosen 320x320 from Imagenette.

Methods. We consider 20% MCAR and MNAR missingness in a localized region. We construct a training set from surrounding subsets.

Analysis. We see SpatialGAIN performs poorly relative to matrix completion but still reasonable. We pruned the “train” set based on pixel similarity, helping enforce learning of spatial dependence.