

Extending CTAB-GAN with StackGAN Architecture

"Can the CTAB-GAN+ performance, as measured by ML Utility and SSD, be improved by stacking multiple networks under varying privacy budgets?"

Orhan Rauf Akdemir, Zilong Zhao, Lydia Chen
TU Delft

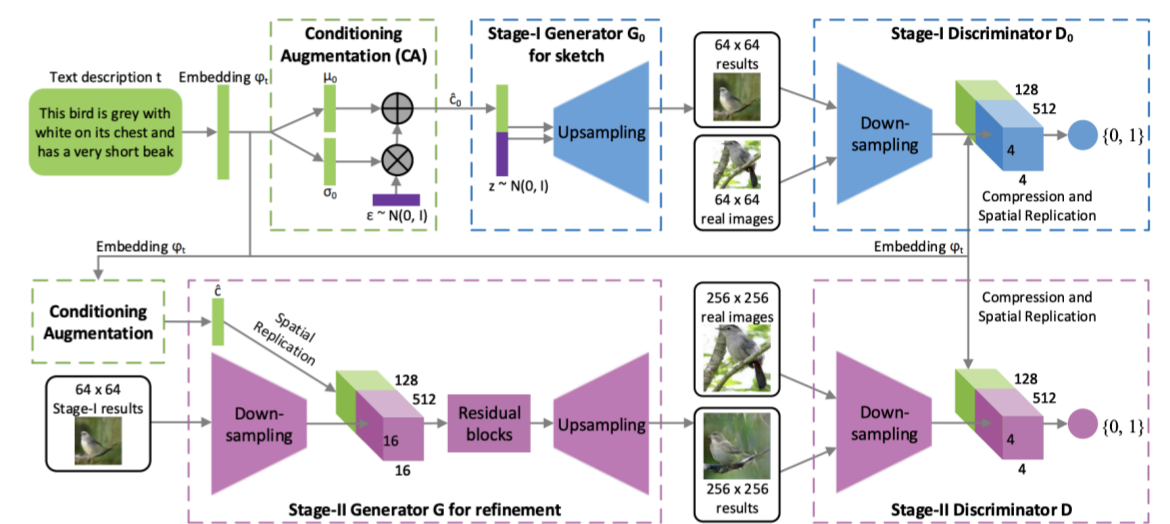


Background and Motivation - Borrowing concepts from text-to-image for tabular data synthesis

- Privacy regulation prevents valuable data transfers
- Data synthesis provides a way to share captured information in original data sets without sharing personal details
- Aligning model distribution and data distribution is difficult due to high dimensionality of data
- Stacking GANs helps with high dimensional data that is common for tabular data synthesis

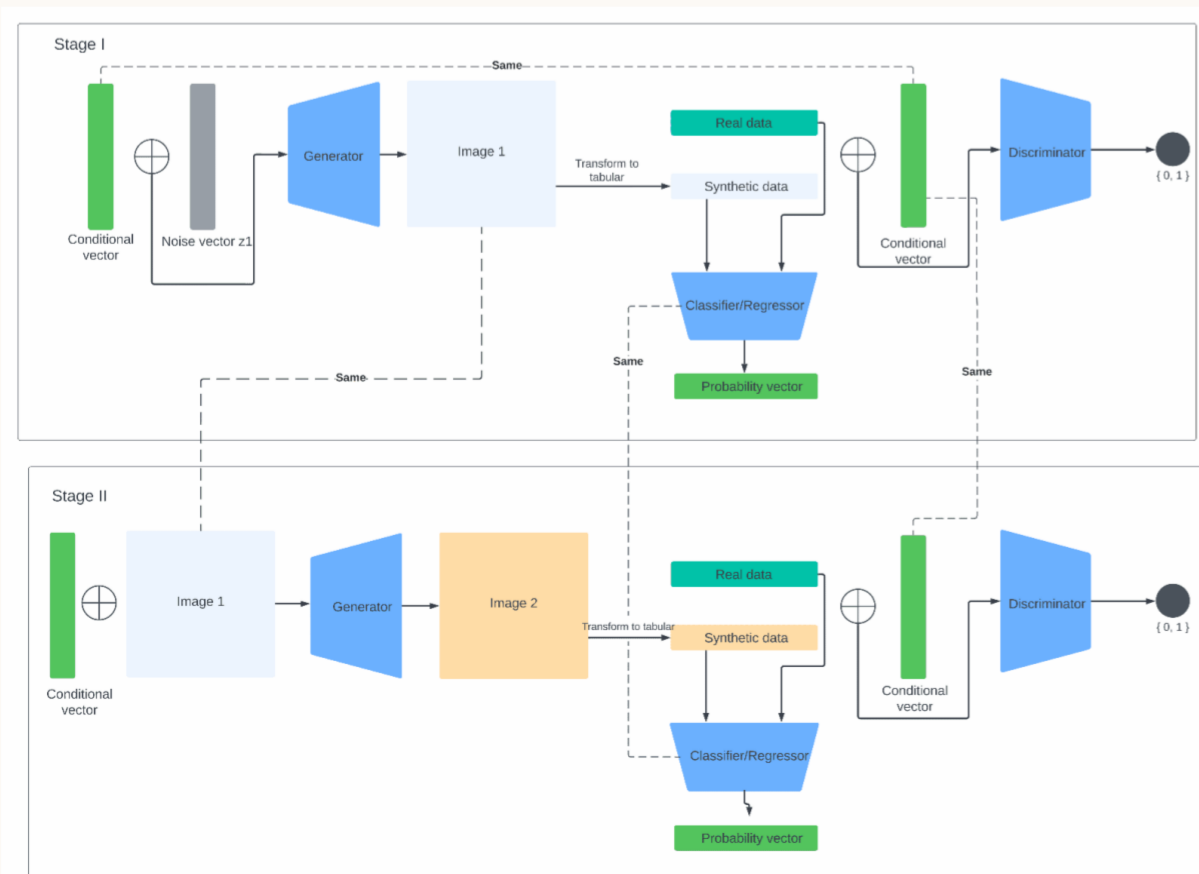
StackGAN Architecture

- Uses first layer to generate low quality image from text embedding
- Reuses the same conditional vector to train and produce second high-definition



Method

1. CTAB-GAN is trained with its generator, information loss and classification loss
2. Generated images and their conditional vectors are used as input vector to second CTAB-GAN, which helps with model alignment. This second CTAB-GAN has a fully connected generator, which helps rectify defects from the first CTAB-GAN.



Result

Using a fully connected generator in the second layer of the stack results in an improved performance based on some metrics, while functioning under the same privacy conditions. Stacking CTAB-GAN with the same generator resulted in worse results overall.

ML Utility Results			
Metrics	CTAB-GAN	Stacked CTAB-GAN (FCG)	Stacked CTAB-GAN
Accuracy	6.89%	8.14%	11.28%
F1-score	0.124	0.116	0.197
AUC	0.165	0.147	0.231

Statistical Similarity Results			
Metrics	CTAB-GAN	Stacked CTAB-GAN (FCG)	Stacked CTAB-GAN
Average WD	0.031	0.028	0.033
Average JSD	0.068	0.051	0.140
Correlation Distance	1.911	5.655	3.032

Privacy Preservability Results			
Metrics	CTAB-GAN	Stacked CTAB-GAN (FCG)	Stacked CTAB-GAN
DCR between R&S	1.126	1.400	1.409
DCR of R	0.383	0.382	0.383
DCR of S	0.982	1.266	0.777
NNDR between R&S	0.710	0.757	0.800
NNDR of R	0.439	0.451	0.439
NNDR of S	0.672	0.608	0.482