Visual Recommender System with Adversarial Generator-Encoder Networks

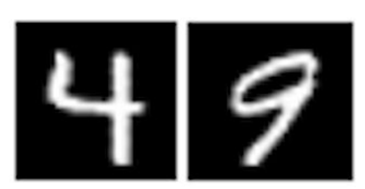
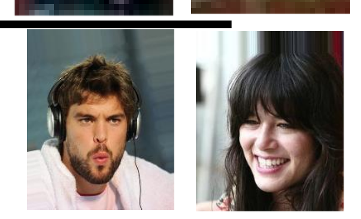
CS231N Final Project - Spring 2017

Bowen Yao, Yilin Chen, Mingxiang Chen

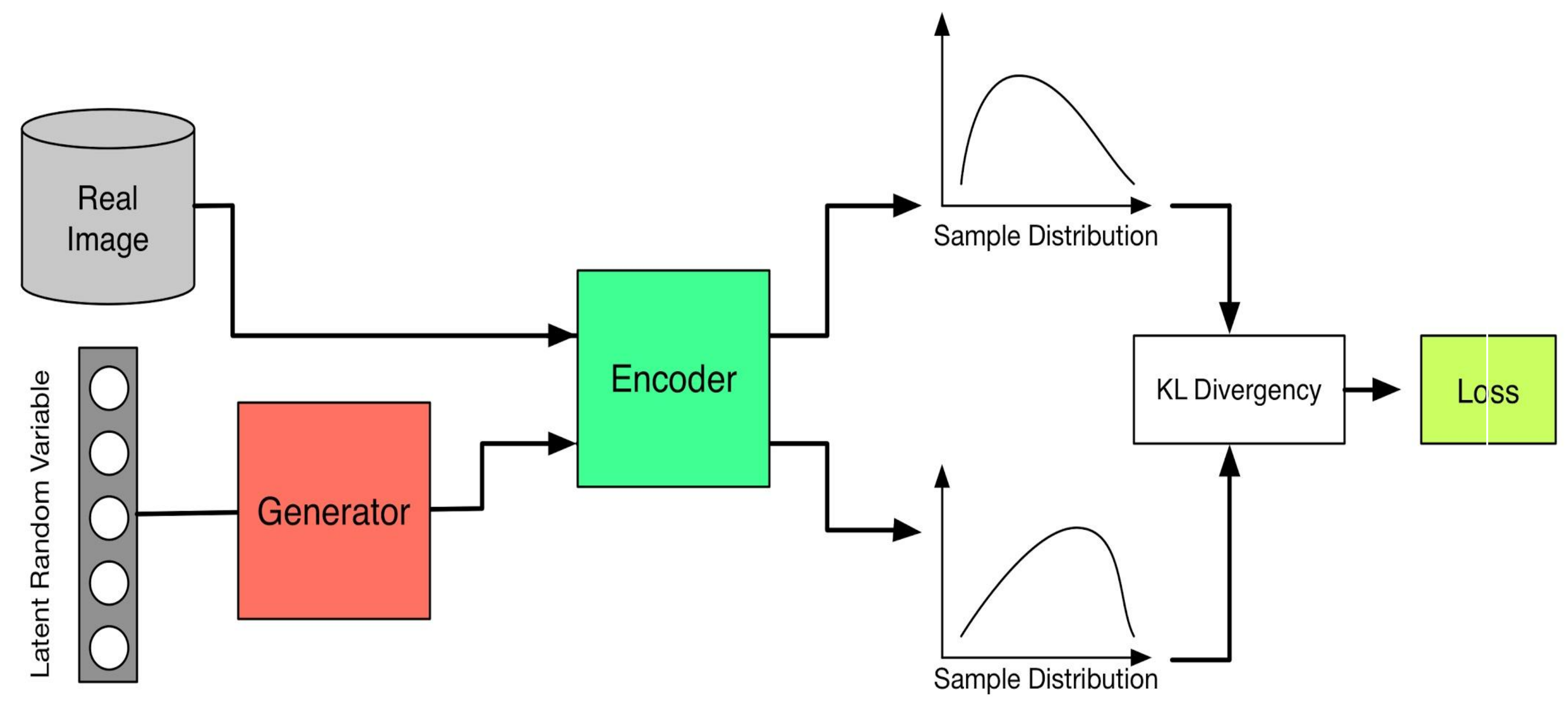
Overview

- We build a deep-learning-based visual recommender system in an unsupervised fashion.
- Such system will be most useful for E-commerce companies where visual recommendation can be used to alleviate cold start issue of common non-deep-learning-based recommender system.
- We use Adversarial Generative-Encoder Network to learn embeddings for images and then K-nearest neighboring images of the query image in the embedding space is output as recommendation results.
- We show both qualitative and quantitative results of our model.

Data:

	Format	#Train	#Val	#Test	Example
MNIST	(N, 28, 28, 3)	45K	5K	10K	
SVHN	(N, 32, 32, 3)	70K	10K	16K	
CIFAR10	(N, 32, 32, 3)	40K	10K	10K	
CelebA	(N, 218, 178, 3)	9K	0.5K	0.5K	
Tiny Imgenet	(N, 64, 64, 3)	100K	10K	10K	

Model



The overall AGE Network model includes a generator and an encoder, which define the mapping between a given distribution in latent space and the data distribution. The generator will try to generate images as indistinguishable from the real data as possible, while the encoder will try to distinguish them from real data. During the "battle" between the generator and the encoder, the joint model gradually "learn" the optimal mappings between the given distribution and the data distribution. After training this adversarial network, we then use this mapping to build our recommendation system.

Quantitative results

SVM classification using features from AGE (on SVHN)

Average training accuracy	23.54%
Average validation accuracy	19.46%

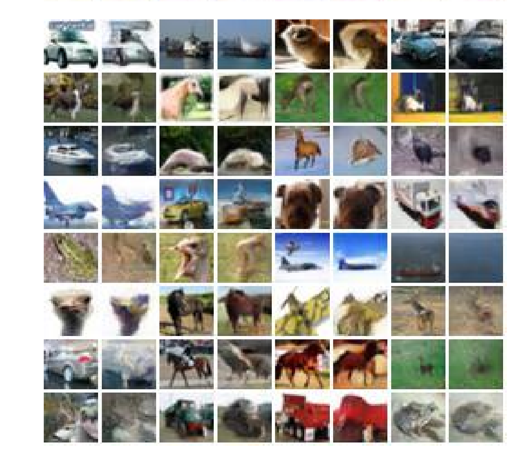
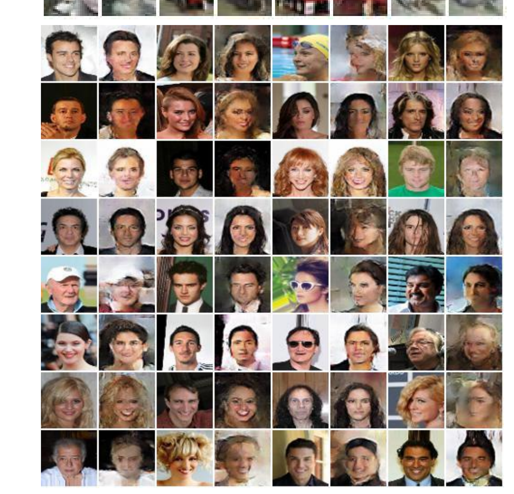
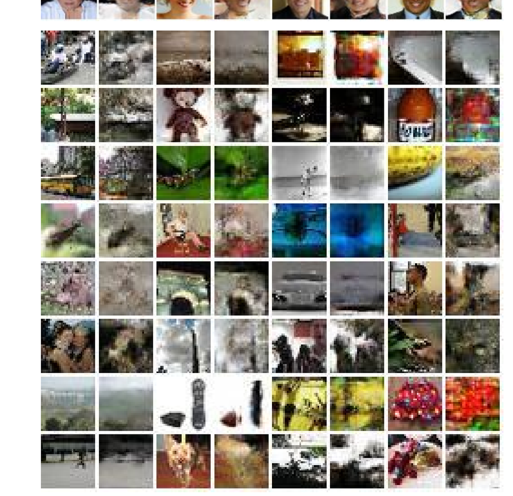
SVM classification using raw pixel informations (on SVHN)

Average training accuracy	46.49%
Average validation accuracy	12.79%

KNN classification using features from AGE (K=3)

Average training accuracy	43.4%
Average validation accuracy	16.2%

Qualitative Results

	Sample	Reconstruction
SVHN		
CIFAR10		
CelebA		
Tiny Imagenet		

Conclusion

- Our AGE network is able to generate and reconstruct good quality images across various datasets
- The embeddings that we learn is better than raw pixels and show meaningful results
- The embeddings are not good enough to build a recommender system on top of it.

Stanford University