Stanford | ENGINEERING
Computer Science

# Predictive BMI with Conditional Adversarial Autoencoder

Alejandro Bravo, Edward Guzman, John Romano
{abravo16, eguzman3, jromano}@stanford.edu

## OVERVIEW

### Motivation

Generative Adversarial Networks (GAN) and related architectures such as Conditional Adversarial Autoencoders (CAAE) have been widely used for predictive tasks such as facial aging and rejuvenation. In addition to facial aging, such generative images are potentially useful for a number of health applications, such as predictive Body Mass Index (BMI) image generation. However, we were not able to find any models that performed this task. We present a model that uses a CAAE in order to generate facial images at different BMI levels. We achieved a mean BMI difference of 6.44 between generated images and their labels.

## DATA

- Initial model pre-trained on ~23,000 face images (UTKFace Dataset)
- Additional augmented version of Face-to-BMI dataset with ~42,000 images (originally ~4,200 images)

Examples of data augmentation (horizontal flipping, recoloring, contrast adjustments)

## METHODS

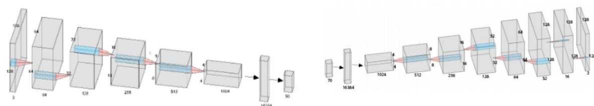
### Models

#### Encoder

5 convolutional layers and a fully connected layer. Face images of dimension 128x128x3 are converted to latent vectors of configurable size (default size of 50).

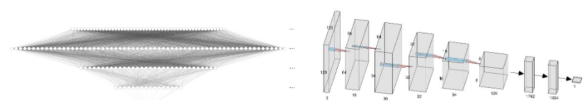
#### Generator

7 deconvolutional layers and a fully-connected layer. Labeled Z vectors in a latent space are transformed into face images of dimensions 128x128x3.



### Discriminators

One discriminator on Z with 4 fully-connected layers, and a second on images with 4 convolutional layers and 2 fully connected layers.



### Features

Our conditional adversarial autoencoder uses an encoder network that turns an input image into a 100-dimensional vector (labeled Z) in latent space. These latent vectors are transformed to get different BMIs. The decoder network then generates images from latent vectors.

## RESULTS

Evaluated qualitatively and quantitatively. Using the Face-to-BMI classifier, we computed the mean difference between what BMI an image was classified to have and what BMI range an image belongs to. This mean difference was 6.44 for our generated images, only 2.5 points more than the mean difference for the original input images.

## DISCUSSION AND FUTURE

We had best results when input images were clear and aligned with subjects facing the camera. The model outputted better results when generating images that increase an individual's BMI rather than decrease it, likely due to substantially more data of high BMI images.

Future study could explore tuning of hyperparameters to improve the model. Because we found that the amount of publicly available BMI-labeled face images was limited, collecting more data would likely increase performance.

## REFERENCES

1. Kocabey, E., et al.(2017). Face-to-BMI: Using Computer Vision to Infer Body Mass Index on Social Media. Retrieved from https://arxiv.org/pdf/1703.03156.pdf
2. Antipov, G., Baccouche, M., & Dugelay, J. (2017). Face Aging With Conditional Generative Adversarial Networks. Retrieved from https://arxiv.org/abs/1702.01983
3. Wang, Z., Tang, X., Luo, W., & Gao, S. (n.d.). Face Aging with Identity-Preserved Conditional Generative Adversarial Networks. Retrieved from http://openaccess.thecvf.com/content_cvpr_2018/papers/Wang_Face_Aging_With_CVPR_2018_paper.pdf