Fader Networks: Manipulating Images by Sliding Attributes

1 Idea

The authors attempt to disentangle facial features from images and re-generate images after tuning (fader knobs) certain continuous-valued attributes of the image like age, expression, gender etc. This is an encoder-decoder architecture.

The major difference touted compared to existing methods is that adversarial training is used to learn the latent space, as opposed to the decoder output, thus, helping the latent space become invariant to the attributes (conditioning labels).

2 Method

- The attributes that are binary during train time, can be treated as continuous during image generation.
- The data set is pictures of actors with certain attributes, like ‘smile’, ‘glasses’, ‘mouth-open’ etc.
- The architecture comprises of 3 main components, the encoder, the discriminator and the decoder. The discriminator is the adversarial component.
- The discriminator is trained with a single objective in mind: to correct identify the attributes, given an encoded image representation
- The encoder-decoder is trained with 2 objectives in mind:

  - The decoder being able to reconstruct the original input, given the encoded representation and the true attributes.
  - The encoded representation making it difficult for the discriminator to ascertain which attributes are present in the original image.

- Without the adversarial component, the decoder learns to ignore the true attributes, and changing these at test time for conditioned generation makes no difference to the decoder output, which we don't want.

- The cost attributed by the discriminator to the encoder loss is gradually increased from 0 over the course of the training.

- The encoded image representation is generated by a convolutional network.

- Augmentation of the face images is done by flipping the images horizontally.

- The generated images were evaluated qualitatively and quantitatively for naturalness.

3 Observations

- The objective is to make the attributes the only source of information for the extra image attributes.

- Avoiding having an adversarial network as part of the decoder is that backpropagation can occur even for discrete objectives, like text sequence prediction.