Generating Realistic Facial Expressions through Conditional Cycle-Consistent Generative Adversarial Networks (CCycleGAN)

Gino Tesei
gtesei@stanford.edu

[YouTube video]

1 Introduction

Photorealistic facial expression synthesis can be helpful in

- face recognition
- entertainment
- virtual and augmented reality
- computer graphics

Novelty: GANs in facial expression generation conditioning on the emotion of facial expression, in absence of paired examples, is a green field.

CCycleGAN = CGAN + CycleGAN
Novel approach for learning to translate a face image of a person conditioned on a given emotion of facial expression (e.g. *joy*) to the same domain but conditioned on a different emotion (e.g. *surprise*)

2 Data

FER2013
(unpaired dataset)

Figure 1: data sample

- **Train:** 28,709
- **Test:** 7,178
- **48x48** pixel **grayscale**
- **Normalization** in data pre-processing
- Good **trade-off accuracy** & **complexity**

3 Methods

$$\mathcal{L}(G,D,X,Y;\lambda_{cyc},\lambda_{cl}) = \mathcal{L}_{\mathcal{RF}}(G,D,X,Y) + \lambda_{cl}\mathcal{L}_{\mathcal{CL}}(G,D,X,Y) + \lambda_{cyc}\mathcal{L}_{cyc}(G,X,Y)$$

Three Loss Terms

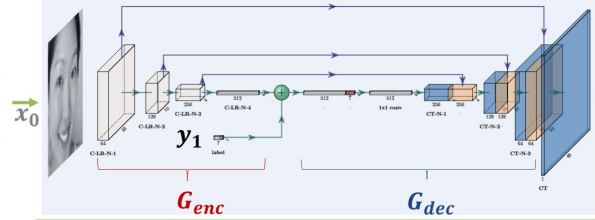
$$\mathcal{L}_{\mathcal{RF}}(G,D,X,Y) = \mathbb{E}_{x\sim p(x)}[\log D_{rf}(x)] + \mathbb{E}_{x,y\sim p(x,y)}[\log(1 - D_{rf}(G(x|y)))]$$

$$\mathcal{L}_{\mathcal{CL}}(G,D,X,Y) = -\mathbb{E}_{x,y\sim p(x,y)}[l_{cl}(x,y)] - \mathbb{E}_{x\sim p(x),y\sim p(y)}[l_{cl}(G(x|y),y)], \text{ where } l_{cl}(x,y) = \sum_{0\le i\le k} -\mathbf{1}_{\{y_i=1\}}\log(D_{cl}(x)_i)$$

$$\mathcal{L}_{cyc}(G,X,Y) = \mathbb{E}_{x,y\sim p(x,y)}\|G_{dec}(G_{enc}(x)|y) - x\|_1$$

(G_{enc}, G_{dec}) - (D_{rf}, D_{cl})

U-NET Generator



Multi-Task Discriminator

$x_1 = G_{dec}(G_{enc}(x_0)|y_1)$

4 Results

- **Adam** solver with $\beta_1 = 0.5, \beta_2 = 0.999$
- **Frechet Inception Distance** (FID) to find **optimal values** of remaining **hyperparameters**

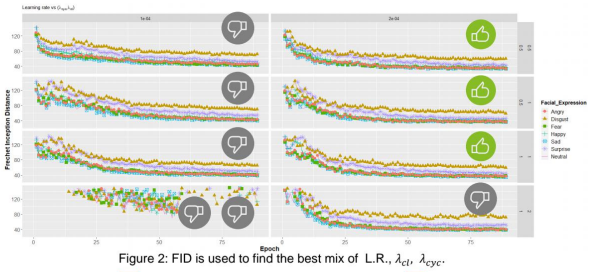


Figure 2: FID is used to find the best mix of L.R., λ_{cl}, λ_{cyc}.

Figure 3: results with Adam solver, batch size of 64 and optimal hyperparameters, i.e. $\lambda_{cl} = \lambda_{cyc} = 1$, L.R. = 0.0002, $\beta_1 = 0.5, \beta_2 = 0.999$

5 Conclusion

- ✓ **Introduced CCycleGAN**: *a novel approach for the synthesis of realistic face images conditioning on the emotion of facial expression in absence of paired examples*
- ✓ **Good qualitative results**
- ✓ **Quantitative justification** for **optimal hyperparameters**