THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

ECE 599/692 – Deep Learning

Lecture 12 – GAN – Design Ideas

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu

1

AICIP RESEARCH

Outline

- Lecture 11: GAN – Introduction and theoretic analysis
- Lecture 12: Design ideas
- Lecture 13: Conditional GAN
- Lecture 14: Implementation

THE UNIVERSITY OF TENNESSEE

2

THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

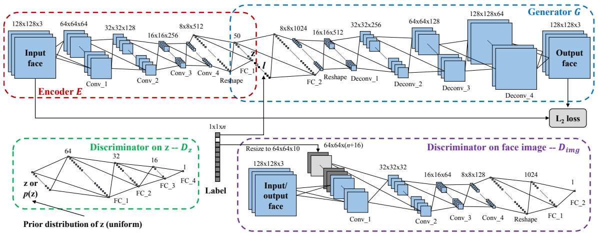
Case Study 1: Conditional Adversarial Autoencoder for Age Progression/Regression

Zhifei Zhang, Yang Song, Hairong Qi, "Conditional adversarial autoencoder for age progression/regression," CVPR, 2017

3

Conditional Adversarial Autoencoder - CAAE

AICIP RESEARCH



CAAE

AICIP RESEARCH

$$\min_{E,G} \max_{D_z, D_{img}} \lambda\mathcal{L}\left(x, G(E(x), l)\right) + \gamma TV(G(E(x), l))$$

D on *z*

$$+\mathbb{E}_{z^*\sim p(\mathbf{z})}\left[\log D_z(z^*)\right]$$
$$+\mathbb{E}_{x\sim p_{data}(\mathbf{x})}\left[\log(1 - D_z(E(x)))\right]$$

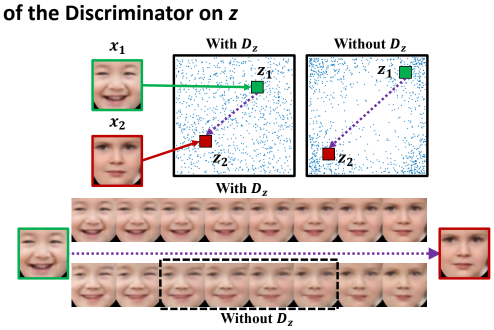
D on image

$$+\mathbb{E}_{x,l\sim p_{data}(\mathbf{x},\mathbf{l})}\left[\log D_{img}(x, l)\right]$$
$$+\mathbb{E}_{x,l\sim p_{data}(\mathbf{x},\mathbf{l})}\left[\log(1 - D_{img}(G(E(x), l)))\right],$$

where $TV(\cdot)$ denotes the total variation which is effective in removing the ghosting artifacts. The coefficients λ and γ balance the smoothness and high resolution.

AICIP RESEARCH

Effect of the Discriminator on *z*



THE UNIVERSITY OF TENNESSEE KNOXVILLE

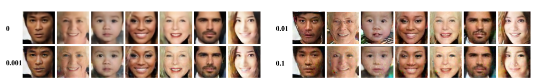
AICIP RESEARCH

Case Study 2: Decoupled Learning for Conditional Adversarial Networks

Zhifei Zhang, Yang Song, and Hairong Qi, “Decoupled learning for conditional adversarial networks,” WACV, 2018

AICIP RESEARCH

Motivation:



The conditional adversarial networks applied in existing works mainly consists of two parts:

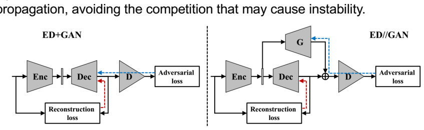
1) the encoding-decoding nets (ED)
2) the GANs, which are tied in the parts of decoder and generator.

Therefore, the reconstruction loss and adversarial loss interact/compete with each other, potentially causing unstable results as shown above.
Existing works have to introduce a weighting factor (e.g., the values in the figure) to balance the effect of the two losses. How to adaptively set an appropriate weight or completely remove the necessity of the weighting factor is a problem unexplored.

AICIP RESEARCH

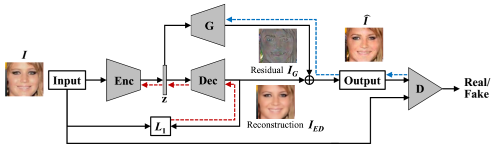
Main Idea:

Decouple the interaction between the reconstruction loss and adversarial loss in backpropagation, avoiding the competition that may cause instability.



- ED+GAN: the traditional structure
- ED//GAN: the proposed structure(decoupled learning)
- Enc and Dec: the encoder and decoder networks
- G and D: the generator and discriminator
- Black arrows: feedforward path
- Red arrows: backpropagation of reconstruction loss
- Blue arrows: backpropagation of adversarial loss

Approach:



The objective function:

$$\min_{Enc,Dec} \mathcal{L}_{const}(Enc, Dec) + \min_{G}\mathcal{L}_{adv}(G) + \min_{D}\mathcal{L}_{adv}(D).$$

There are no weighting parameters between the losses in the objective function, which relaxes the manual tuning.

Experimental Results:

Compare ED+GAN and ED//GAN:

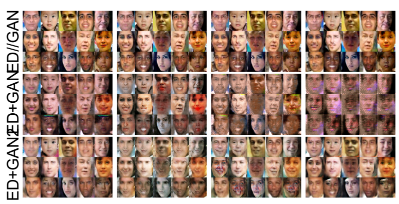


Table 2. NRDS with different weight settings and their std.

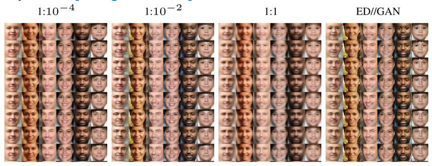
	0.001	0.01	0.1	1	std
ED+GAN	.1172	.1143	.1163	.0731	.0215
ED+GAN2	.1066	.1143	.1268	.1267	.0099
ED//GAN	.1432	.1434	.1458	.1466	.0017

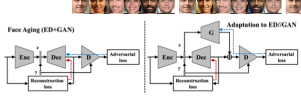
ED+GAN2 denote the structure with batch normalization

ED+GAN is sensitive to weight variation. By contrast, ED//GAN is robust to weight variation, relaxing the weight tuning.

Experimental Results:

Adapt CAAE [Zhang et al., 2017] to ED//GAN:





Method	ED+GAN			ED//GAN
	$1{:}10^{-4}$	$1{:}10^{-2}$	1:1	
NRDS	.2527	.2496	.2430	**.2547**

AICIP RESEARCH

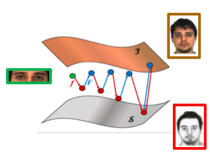
Case Study 3: Cross-domain Face Composite and Synthesis from Limited Facial Patches

Yang Song, Zhifei Zhang, Hairong Qi, "Cross domain face composition and synthesis from limited facial parts," *AAAI Conference on Artificial Intelligence (AAAI)*, New Orleans, LA, February 2018.

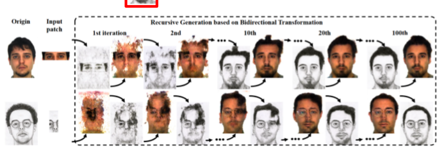
13

AICIP RESEARCH

The Recursive Bidirectional Transformation Network

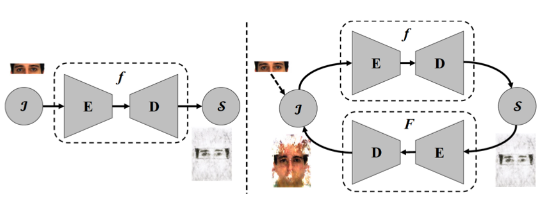


Assumption:
Whole Face and Sketch are lie on two manifolds;
The mapping from sketch domain ($\mathcal{S}$) to face domain ($\mathcal{I}$) is modeled by F;
The mapping from face domain ($\mathcal{I}$) to Sketch domain ($\mathcal{S}$) is modeled by f;
The face/sketch patches lie outside the manifold.



AICIP RESEARCH

Transformation Network

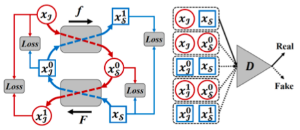


Uni-directional transformation vs. Bi-directional transformation network

AICIP RESEARCH

Training Stage

Training Stage: learn face2sketch mapping f and sketch2face mapping F in a bidirectional fashion



$$x_{\mathcal{S}}^0 = f(x_{\mathcal{I}}),\ x_{\mathcal{I}}^1 = F(x_{\mathcal{S}}^0) = F(f(x_{\mathcal{I}})),$$
$$x_{\mathcal{I}}^0 = F(x_{\mathcal{S}}),\ x_{\mathcal{S}}^1 = f(x_{\mathcal{I}}^0) = f(F(x_{\mathcal{S}})).$$

Objective Function:

$$\min_{f,F,D}\ \mathcal{L}_{adv} + \lambda\mathcal{L}_{rec},$$

$$\mathcal{L}_{rec} = \sum^{1} \left(\|x_{\mathcal{I}} - x_{\mathcal{I}}^i\|_1 + \|x_{\mathcal{S}} - x_{\mathcal{S}}^i\|_1\right), \quad (1)$$

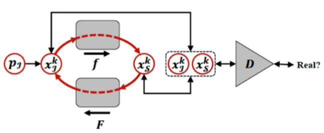
$$\mathcal{L}_{adv} = \mathbb{E}_{\omega\in\Omega}\left[\log D(\omega)\right] - \mathbb{E}_{\substack{x_{\mathcal{I}}\in\mathcal{I}\\ x_{\mathcal{S}}\in\mathcal{S}}}[\log D(x_{\mathcal{I}}, x_{\mathcal{S}})], \quad (2)$$

$$\Omega = \{(x_{\mathcal{I}}, x_{\mathcal{S}}^0)_j, (x_{\mathcal{I}}^1, x_{\mathcal{S}}^0)_j, (x_{\mathcal{I}}^0, x_{\mathcal{S}})_j, (x_{\mathcal{I}}^0, x_{\mathcal{S}}^1)_j\}$$

THE UNIVERSITY OF TENNESSEE

AICIP RESEARCH

Testing Stage: Given a patch $P_{\mathcal{I}}$, the mapping f and F trained on whole face/sketch will guide the patch to converge to a whole sketch/face. In order to keep the identity, the given patch area is kept as constant at each iteration.



$$x_{\mathcal{I}}^k \leftarrow x_{\mathcal{I}}^k \odot (1 - M) + p_{\mathcal{I}},$$
$$x_{\mathcal{S}}^k \leftarrow f\left(x_{\mathcal{I}}^k\right),$$
$$x_{\mathcal{S}}^k \leftarrow x_{\mathcal{S}}^k - \frac{\partial D(x_{\mathcal{I}}^k, x_{\mathcal{S}}^k)}{\partial x_{\mathcal{S}}^k},$$
$$x_{\mathcal{I}}^{k+1} \leftarrow F\left(x_{\mathcal{S}}^k\right).$$

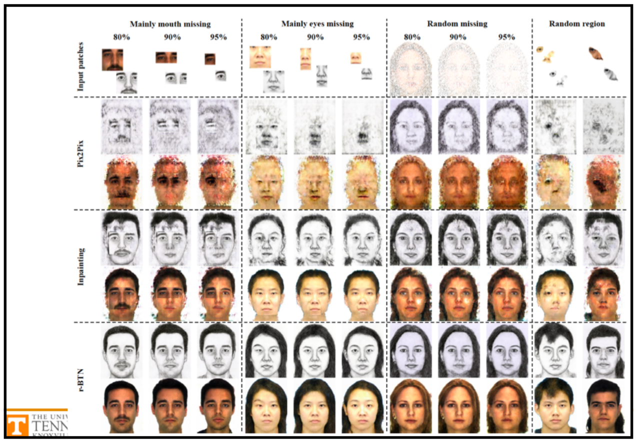
THE UNIVERSITY OF TENNESSEE

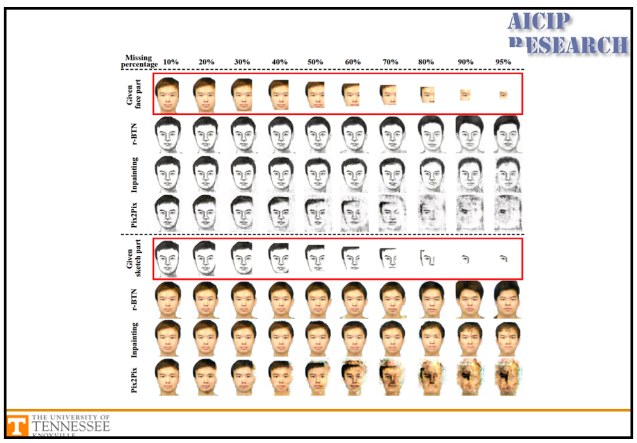
AICIP RESEARCH

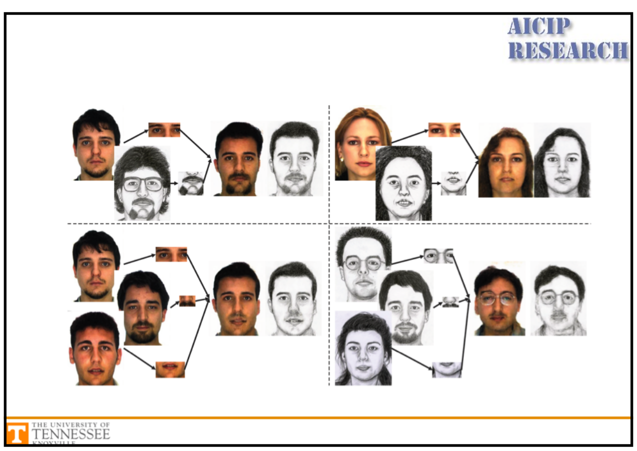
Data Collection

- Paired Dataset:
  - 1,577 face/sketch pairs from the datasets CUHK [24], CUFSF [28], AR [13], FERET [17], and IIITD
- Unpaired Dataset:
  - Image Crawl from google image search engineering
  - Transformation based on pre-trained network.

THE UNIVERSITY OF TENNESSEE


Mainly mouth missing
Mainly eyes missing
Random missing
Random region
80% 90% 95%
80% 90% 95%
80% 90% 95%
Input patches
Pix2Pix
Inpainting
r-BTN


AICIP
RESEARCH
Missing percentage
10% 20% 30% 40% 50% 60% 70% 80% 90% 95%
Given face part
r-BTN
Inpainting
Pix2Pix
Given sketch part
r-BTN
Inpainting
Pix2Pix
THE UNIVERSITY OF
TENNESSEE


AICIP
RESEARCH
THE UNIVERSITY OF
TENNESSEE

THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

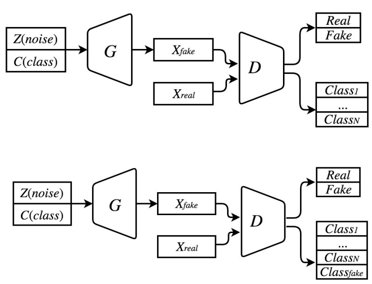
Case Study 4: FCGAN

Chengcheng Li, Zi Wang, Hairong Qi, "Fast-converging conditional generative adversarial networks for image synthesis," *IEEE International Conference on Image Processing (ICIP)*, October 2018.

22

AICIP RESEARCH

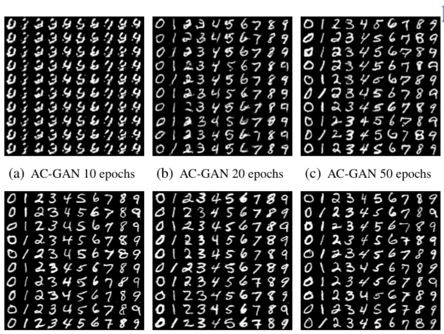
AC-GAN and FC-GAN



THE UNIVERSITY OF TENNESSEE

23

AICIP RESEARCH



(a) AC-GAN 10 epochs (b) AC-GAN 20 epochs (c) AC-GAN 50 epochs

(d) FC-GAN 10 epochs (e) FC-GAN 20 epochs (f) FC-GAN 50 epochs

THE UNIVERSITY OF TENNESSEE

24

Acknowledgement

AICIP RESEARCH

Most slides are taken from student presentations at AICIP group meetings, including mainly those from Zhifei Zhang, Yang Song and Chengcheng Li.

THE UNIVERSITY OF TENNESSEE KNOXVILLE 25