# **DEEP LEARNING**

Lecture 10: Generative Models

Dr. Yang Lu

Department of Computer Science and Technology

luyang@xmu.edu.cn





# Image Translation







Image source: Choi, Yunjey, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. "Stargan v2: Diverse image synthesis for multiple domains." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8188-8197. 2020.

# Image Translation







Image source: Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2414-2423. 2016.

## Scene Generation







Image source: Tang, Hao, Dan Xu, Yan Yan, Philip HS Torr, and Nicu Sebe. "Local class-specific and global image-level generative adversarial networks for semantic-guided scene generation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7870-7879. 2020.

# Facial Attribute Manipulation







Image source: Geng, Zhenglin, Chen Cao, and Sergey Tulyakov. "3d guided fine-grained face manipulation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9821-9830. 2019.

# Facial Attribute Manipulation









5

Image source: Liu, Ming, Yukang Ding, Min Xia, Xiao Liu, Errui Ding, Wangmeng Zuo, and Shilei Wen. "Stgan: A unified selective transfer network for arbitrary image attribute editing." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3673-3682. 2019.

# Gaze Correction



Image source: Zhang, Jichao, Meng Sun, Jingjing Chen, Hao Tang, Yan Yan, Xueying Qin, and Nicu Sebe. "Gazecorrection: Self-guided eye manipulation in the wild using self-supervised generative adversarial networks." arXiv preprint arXiv:1906.00805 (2019).

# Image Animation







7

Image source: Siarohin, Aliaksandr, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. "Animating arbitrary objects via deep motion transfer." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2377-2386. 2019.

# Image Inpainting









Image source: Yu, Jiahui, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. "Generative image inpainting with contextual attention." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5505-5514. 2018.

# Image Inpainting







Image source: Yu, Jiahui, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. "Free-form image inpainting with gated convolution." In Proceedings of the IEEE International Conference on Computer Vision, pp. 4471-4480. 2019.

# Image Blending







Image source: Wu, Huikai, Shuai Zheng, Junge Zhang, and Kaiqi Huang. "Gp-gan: Towards realistic high-resolution image blending." In Proceedings of the 27th ACM International Conference on Multimedia, pp. 2487-2495. 2019.

# Image Super-Resolution



SRGAN (21.15dB/0.6868)





original







Image source: Ledig, Christian, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken et al. "Photo-realistic single image super-resolution using a generative adversarial network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681-4690. 2017. https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

# Makeup



Image source: Li, Tingting, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. "Beautygan: Instance-level facial makeup transfer with deep generative adversarial network." In Proceedings of the 26th ACM international conference on Multimedia, pp. 645-653. 2018.

12

# Text-to-image generation



"a hedgehog using a calculator"

"a corgi wearing a red bowtie and a purple party hat"

"robots meditating in a vipassana retreat"

"a fall landscape with a small cottage next to a lake"





Image source: AlexanderQuinnNichol,PrafullaDhariwal,AdityaRamesh,PranavShyam,PamelaMishkin,BobMcgrew,IlyaSutskever,andMarkChen.2022. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In International Conference on Machine Learning. 16784–16804.

# Video generation

# Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators







# Text-to-3D generation



A blue tulip



An elephant skull



A delicious croissant





Michelangelo style statue of dog reading news on a cellphone



# Music Generation



Generated music samples: https://soundcloud.com/openai audio





### Generative vs. Discriminative

- In machine learning, two main approaches are called the generative approaches and the discriminative approaches.
- Given an observable variable *X* and a target variable *Y*:
  - A generative model is a statistical model of the data distribution P(X) or the joint probability distribution on  $X \times Y$ : P(X, Y).
  - A discriminative model is a model of the conditional distribution of Y given X: P(Y|X = x).



### **Discriminative Approaches**

- Most supervised learning methods fall into discriminative approaches.
  - Given data: (x, y), x is data, y is label.
  - Goal: Learn a function to map  $x \rightarrow y$ , namely posterior probability P(Y|X = x).
  - Examples: Classification, regression, object detection, face recognition, sentiment classification, etc.









### Generative Approaches

- Given training data, generate new samples from same distribution.
- Objectives:
- 1. Learn  $p_{model}(x)$  that approximates  $p_{data}(x)$ .
- 2. Sample a new x from  $p_{model}(x)$ .



### **Generative Models**







- Variational Autoencoder
- Generative Adversarial Nets
- Diffusion Model





# VARIATIONAL AUTOENCODER

22

### Autoencoder



loss =  $|| \mathbf{x} - \hat{\mathbf{x}} ||^2 = || \mathbf{x} - \mathbf{d}(\mathbf{z}) ||^2 = || \mathbf{x} - \mathbf{d}(\mathbf{e}(\mathbf{x})) ||^2$ 





### 23

Image source: https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73



The autoencoder is solely trained to encode and decode with as few loss as possible, no matter how the latent space is organized.



- A variational autoencoder can be defined as being an autoencoder whose training is:
  - regularised to avoid overfitting;
  - ensure that the latent space has good properties that enable generative process.





# Instead of encoding an input as a single point, we encode it as a distribution over the latent space.



Regularize the organization of the latent space by making the distributions returned by the encoder close to a standard normal distribution.









Image source: https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

- The loss of the autoencoder is to minimize both:
  - the reconstruction loss (how similar the autoencoder's output to its input);
  - the latent loss (how close its hidden nodes were to a normal distribution).
- It doesn't guarantee the quality of the generated image. A major drawback of VAEs is the blurry outputs that they generate. VAE models tend to produce unrealistic, blurry samples.



Human faces generated by VAEs





Image source: https://medium.com/sciforce/generative-models-under-a-microscope-comparing-vaes-gans-and-flow-based-models-344f20085d83#:~:text=A%20major%20drawback%20of%20VAEs,to%20produce%20unrealistic%2C%20blurry%20samples

# GAN



### GAN

# GAN was proposed by Ian Goodfellow in 2014.

- Yann LeCun described GANs as "the most interesting idea in the last 10 years in machine learning".
- Ian presented and explained his paper in NIPS 2016 with a 2-hour presentation.

<u>I Goodfellow</u>, <u>J Pouget-Abadie</u>... - Advances in neural ..., 2014 - proceedings.neurips.cc ... We propose a new framework for estimating **generative** models via **adversarial nets**, in which we simultaneously train two models: a **generative** model G that captures the data ...

☆ Save 57 Cite Cited by 61677 Related articles All 61 versions ≫

lan in NIPS 2016





Generative adversarial nets

### GAN: How to Do

- GAN is composed by a generator and a discriminator. They are both neural networks.
  - Generator network: try to fool the discriminator by generating real-looking images.
  - Discriminator network: try to distinguish between real and fake images.



### GAN: How to Do

- Generator and discriminator tells each other where it was wrong.
  - Generator tells discriminator how I fool you.
  - Discriminator tells generator how I detect you.





### GAN: How to Learn

- Given a prior on input noise variables  $z \sim p_z(z)$ , generator  $G_{\theta_g}(z)$  take z as input and map it into data space.
- Discriminator  $D_{\theta_d}(x)$  takes the data  $x \sim p_{data}$  as input and output probability that x came from the real data rather than generated data  $G_{\theta_a}(z)$ .
- $D_{\theta_d}$  and  $G_{\theta_g}$  have different goals (1 for real, 0 for fake):
  - Generator wants:  $D_{\theta_d}\left(G_{\theta_g}(\mathbf{z})\right) \to 1.$
  - Discriminator wants:  $D_{\theta_d}(x) \to 1$ ,  $D_{\theta_d}(G_{\theta_g}(z)) \to 0$ .




#### GAN: How to Learn

By maximizing the log-likelihood, the overall objective is to simultaneously train over all x with random generated z:

• train 
$$G_{\theta_g}$$
 to minimize  $\log\left(1 - D_{\theta_d}\left(G_{\theta_g}(\mathbf{z})\right)\right)$ ;

• train 
$$D_{\theta_d}$$
 to maximize  $\log D_{\theta_d}(x)$  and  $\log \left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)$ .

In other words,  $D_{\theta_d}$  and  $G_{\theta_a}$  play the following two-player minimax game:

$$\min_{\theta_{g}} \max_{\theta_{d}} \left[ \mathbb{E}_{x \sim p_{data}} \log D_{\theta_{d}}(x) + \mathbb{E}_{z \sim p_{z}(z)} \log \left( 1 - D_{\theta_{d}} \left( G_{\theta_{g}}(z) \right) \right) \right]$$
  
Discriminator output  
for real data  $x$   
Discriminator output for  
generated fake data  $G_{\theta_{g}}(z)$ .  
$$\bigotimes_{x \in [0, \infty]} \mathbb{E}_{x \sim p_{data}} \log \mathbb{E}_{x \sim p_{data}} \log \mathbb{E}_{x \sim p_{z}(z)} \log \left( 1 - D_{\theta_{d}} \left( G_{\theta_{g}}(z) \right) \right) \right]$$

#### GAN: How to Learn

Objective:

$$\min_{\theta_g} \max_{\theta_d} \left[ \mathbb{E}_{\boldsymbol{x} \sim p_{data}} \log \frac{D_{\theta_d}(\boldsymbol{x})}{D_{\theta_d}(\boldsymbol{x})} + \mathbb{E}_{\boldsymbol{z} \sim p_z(\boldsymbol{z})} \log \left( 1 - \frac{D_{\theta_d}\left(G_{\theta_g}(\boldsymbol{z})\right)}{D_{\theta_d}(\boldsymbol{z})} \right) \right]$$

Alternate between:

Gradient ascent on discriminator:

$$\max_{\theta_d} \left[ \mathbb{E}_{\boldsymbol{x} \sim p_{data}} \log D_{\theta_d}(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})} \log \left( 1 - D_{\theta_d} \left( G_{\theta_g}(\boldsymbol{z}) \right) \right) \right]$$

Gradient descent on generator:

$$\min_{\theta_g} \left[ \mathbb{E}_{z \sim p_z(z)} \log \left( 1 - \frac{D_{\theta_d}}{G_{\theta_g}(z)} \right) \right]$$





#### GAN: How to Learn

- After several steps of training, if G and D have enough capacity, they will reach a point at which both cannot improve because  $p_g = p_{data}$ .
  - Generator can generate real image.
  - Discriminator is unable to differentiate between the two distributions, i.e.  $D_{\theta_d}(\mathbf{x}) = 1/2$ .

discriminative distribution data distribution generative distribution



Image source: Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." In Advances in neural information processing systems, pp. 2672-2680. 2014.

#### GAN: Algorithm

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k, is a hyperparameter. We used k = 1, the least expensive option, in our experiments.

for number of training iterations do

for k steps do

- Sample minibatch of m noise samples  $\{z^{(1)}, \ldots, z^{(m)}\}$  from noise prior  $p_g(z)$ .
- Sample minibatch of m examples  $\{x^{(1)}, \ldots, x^{(m)}\}$  from data generating distribution  $p_{\text{data}}(x)$ .
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[ \log D\left( \boldsymbol{x}^{(i)} \right) + \log \left( 1 - D\left( G\left( \boldsymbol{z}^{(i)} \right) \right) \right) \right].$$

end for

- Sample minibatch of m noise samples  $\{z^{(1)}, \ldots, z^{(m)}\}$  from noise prior  $p_g(z)$ .
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

#### end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.





#### GAN: Result



Rightmost column shows the nearest training example of the neighboring sample, in order to demonstrate that the model has not memorized the training set.

**厦門大學信息学院(特色化示范性软件学院)** School of Informatics Xiamen University (National Characteristic Demonstration Software School)



Image source: Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." In Advances in neural information processing systems, pp. 2672-2680. 2014.

#### GAN: Result

	The second			the same second second second second	
Contract Strength and the strength of the stre	a contraction of a station of	in case where we go as	the state of the second s	the support of the local day with	IN THE R. P. LEWIS CO., NAME
and the second	1000		and the second second second second	Carl Contraction of the	The second second second
and the second second second	THE REAL PROPERTY OF	COLUMN TWO IS NOT THE OWNER.	and the second second second	and the second se	and the second se
THE PARTY OF THE P	THE OWNER AND A DESIGNATION OF	and the second second	A CONTRACT OF A	the state of the state of the	Variation of the Alaska
Contraction of the second second second	And the second second second	COMPACTOR AND A	A DESCRIPTION OF A DESC	A CONTRACTOR OF	Concernence and second second
	Contraction of the second s	States of the second second second	A REAL PROPERTY AND	and the second se	and the second se
and the second second	and the second	the state of the second st	and the second second	COMPANY AND ADDRESS	and the second second second
the same of the same of the same	the second second second second	A DE LA DE L	CONTRACTOR OF THE OWNER OF	A REAL PROPERTY OF THE REAL PR	CONTRACTOR OF A DESCRIPTION OF A DESCRIP
AND ADDRESS OF ADDRESS OF	A DESCRIPTION OF A DESC	COLUMN TWO IS NOT	CONTRACTOR NO. 1	A CONCEPT ON THE	THE COMPANY OF THE
AND DESCRIPTION OF ADDRESS.	COLOR & BUTCHES, MARKING	COLUMN TAXABLE AND A	A PERSON AND A PROPERTY AND A	THE R PROPERTY AND ADDRESS OF	CONTRACTOR OF STREET,
ALL COME THE DESIGNATION OF	THE R. D. LEWIS CO., LANSING, MICH.	CONTRACTOR OFFICE	A REAL PROPERTY AND ADDRESS OF \$1,000	NAME OF TAXABLE PARTY.	A DOMESTIC AND IN THE OF
	A REAL PROPERTY AND ADDRESS	And in case of the local division of the loc	No. of Concession, Name of Street, and	STATUTE AND ADDRESS OF ADDRESS OF	COMPANY OF THE OWNER OF THE OWNER OF
A REAL PROPERTY AND A REAL	THE OWNER WATER AND A DESCRIPTION OF A D	CONTRACTOR OF THE OWNER.	other the places of the second second	A DECEMBER AND A DEC	CONTRACTOR OF A CONTRACTOR
and the second second second	1 2 2 2 2 2 2 2 2 2 2 2	and the second second second	TODOLO CONTRACTOR	A DECEMBER OF	the second s
CARL CR. LANS	A COLUMN TWO IS NOT	COMPANY AND A REAL PROPERTY OF	THE REPORT OF A	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	AND A PART OF BUT AND
IN MY & CASE & CHI I	the second read of the second	OT LOD BOLLING CO. N. D.	A COLUMN TWO IS NOT THE OWNER.	Reading and a group of the	STREET, STREET,
			the second s	Contraction of the second second	
A DO N. HALF THE LOCAL	CONTRACTOR OF A DESCRIPTION OF A DESCRIP	CONTRACTOR OF ALL PROPERTY OF	NAME AND ADDRESS OF STREET	A DAY TO CHE AND	THE R. LEWIS CO., LANS.
and the second second	A DE LET A L'ANDER	The short of	CONTRACTOR OF A	CONTRACTOR OF THE OWNER	and a second second
		and the second se	and the second se	CONTRACTOR OF A DESCRIPTION OF A DESCRIP	Contraction of the second second
COLUMN TWO IS NOT	Distant and the state of the local distance	AND A MELLINE AND	and the second second second	The second second second	NAME AND ADDRESS OF TAXABLE
And a stand of the second second	A CONTRACTOR OF THE OWNER OWNER OF THE OWNER OWNE OWNER OWNE	and the second second second	COLUMN TWO IS NOT A TO	AN ADDRESS OF A DESCRIPTION OF A DESCRIP	And a set of the set of the set
surface of the based on a line of the	A REAL PROPERTY AND A REAL PROPERTY AND A	A DECEMBER OF STREET	A DECK OF A DECK OF A DECK	the second se	Date of the second s
and the second se	Contraction of the Contraction o	the second second second	Support of the Local Division of the local d	Contraction of the local division in the	And A showing and the second
Contraction of the second second	and the second second	COLUMN STREET, SALES	THE REAL PROPERTY OF	Contraction of the second second	COLUMN AND ADDRESS
A ROUGHLESS AND	A CONTRACTOR OF A PARTY	THE OWNER WANTED IN CO.	of the second second second second	THE R. L. LOW, M. L.	THE R. LEWIS CO., NAMES
Street of Long Street of Street	and the second states of the s	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	and the second second	Carl Start Party and and	and the second se
CONTRACTOR OF A DESCRIPTION OF A DESCRIP	In the Cold Street B Line and Lines.	COLUMN TWO IS NOT THE OWNER.	CALLSREE STATES OF THE OWNER OF	COMPANY OF A DESCRIPTION OF	A REAL PROPERTY AND ADDRESS
the second second	and the second second second	COLUMN TO A DOWN	and the second second second	The second second second	and the second second
and the second second	100 CT 10 C 10 M	the second second second	and the second second	CONTRACTOR OF A DECK	and the second second
County in a second second	Distants Property States of	and a state of the	A ST. ABOVE CREATER A	COLUMN ALL OCTAVA	A REAL PROPERTY OF A REAL PROPERTY OF
and the second second second second		And the second sec	100 D 20 V 10 V 10 V	and the second second second	Carlo Martin State and
the set of the state of the set	State of the second second	And I Real Property lies in the	and the second second	CONTRACTOR OF THE OWNER OF	THE OWNER AND ADDRESS OF
A CONTRACTOR DECIMATION	A CONTRACTOR OF A CONTRACT OF	A TAXABLE PARTY AND A DESCRIPTION OF A D	A DOMESTIC AND DESCRIPTION OF	NAMES AND ADDRESS OF TAXABLE	AT LONG BUILD AND A
The second se				International Addition of the local	
the style of the set of	COLUMN TO A VALUE OF	CALCULATION OF CALCULATIO	THE OWNER AND CARD AND ADDRESS.	CONTRACTOR OF THE OWNER OF	THE OWNER WATER AND ADDRESS OF
A REAL PROPERTY AND	A STREET FOR ALL	A COLOR OF A COLOR OF	states and shares a state	And the Area of the Area	THE R. P. LEWIS CO., LANSING, MICH.
and the state of the second	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	CONTRACTOR AND A REAL PROPERTY OF	A DAY AND A DAY NOT	A DECK AND A DECK AND A	The second second
Constant and the second second	and the second se	a contract of the second	A REAL PROPERTY AND A REAL	A REAL PROPERTY OF A REAL PROPER	10 10 10 10 10 10 10 10 10 10 10 10 10 1
Contraction of the local division of the loc	Charles and the State of State	and the second second	and the second second second second	the standard in the second	And the state of the
A CARLES AND A CARLES	Contraction of the second	100 June 18 18 19	PERSONAL PROPERTY AND ADDRESS	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	A DECEMBER OF
COLUMN STORY (COLUMNS)	and the second second	and the second sec	The Party of the state of the	and the second second second	
the part of the local day which	and the second second	and the first state	and the second second second	the second state and	and the second second
and the second second second	No. 10 Conception of the	CALL STREET, ST	and the second second second	A REAL PROPERTY AND A REAL PROPERTY AND A	Concernance of the second
The set of the local loc	THE R LOUGH COMMENT	A REAL PROPERTY AND INCOME.	a break the second at	A CONTRACTOR OF A	A R. M. THE MONTH & MILLING
Contraction of the local	Contraction of the local division of the loc	or chains that can be	THE COLORED STORE DECK	the state of the last of	AND A DECKER AND A D
and particular of a lot of	the state of the second second	CANES AND A REAL PROPERTY.	CALMONIA CONTRACTOR	in the same of the same of the	And the Local and the second
and the second second	A REAL PROPERTY.	and the second second	CALL STREET, ST	CONTRACTOR AND A	Carlos and a state
I THE REAL PROPERTY AND	the second second second	2 10 10 10 color	A COLUMN AND AND	NAME OF TAXABLE PARTY.	
A COMPANY OF ANY OTHER DO	THE R. LEWIS CO., LANSING MICH.	COLUMN TWO IS NOT THE OWNER.	and the second second	the base of the local division of the local	CONTRACTOR OF STREET
and the second second	And the second sec	the second s	and the second second		A
a second second second	The last of the second	ALCOLUMN AND ADD	Strate State State State	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	THE R. LEWIS CO., LANSING, MICH.
and the second se	A REAL PROPERTY AND	THE REPORT OF THE	A DECEMBER OF	A DESCRIPTION OF A DESC	Contraction of the local division of the loc
and the second second	the second states and	A CONTRACTOR OF STREET	and the second second second	Charles and a second	and an other and
And in the local day in such	The second second second	COLUMN STREET, SAL	100 C 100 - 100 - 100	Contraction of the local division of the loc	CONTRACTOR OF STREET
Control of the local distance of the local d	TO DECIMAL COMMON	Contraction of the local sector	State of the local data and the	Contraction of the local sector	and the second second second
	Long and the second second	STATISTICS.	A REAL PROPERTY OF THE REAL PROPERTY.	the second se	States of the second
the second of the second second	THE OWNER WAS ADDRESS OF	AND REAL PROPERTY.	Carlos and and and	the second second second second	and the second second second
CONTRACTOR OF A DECISION		and the second second	10 2 2 2 2 2 2 2 2 2	and the second second	100 C 200 C 200 C
and the second se	A DESCRIPTION OF THE OWNER OWNER OF THE OWNER OWNER OF THE OWNER OWNER OF THE OWNER OWNE	COLUMN TO THE R. P. N.	A REAL PROPERTY AND A REAL	and the second s	and the second sec
such that we have a second	the second second second	Carl State State		a man is a star of	the state of the second
and the second second	Contraction of the local division of the loc	A DESCRIPTION OF THE OWNER OWNER OF THE OWNER OWNER OF THE OWNER	COLUMN TWO IS NOT THE OWNER.	A REAL PROPERTY OF A REAL PROPER	A DESCRIPTION OF THE OWNER OWNER OF THE OWNER OWNER OF THE OWNER OWNER OWNER OF THE OWNER OWN
A DECEMBER OF A	AND A COMPANY AND A PROPERTY	the second second second	the second second second	100 M	1
and water and a good of	the state of the state of the	Charles and the state of the st	and a state of the	of the second	and the second second second
Statement of the statement	NAME AND ADDRESS OF TAXABLE PARTY.	THE OWNER WATER OF THE OWNER OF T		States of the local division of the local di	State of the local day in the local day
of the other of the local data	the second se	the second se	The second se	THE R. P. LEWIS CO., LANSING MICH.	A REAL PROPERTY OF A READ PROPERTY OF A REAL PROPER
	A REAL PROPERTY AND ADDRESS OF TAXABLE PARTY.	THE R. LEWIS CO., LANSING, MICH.	the second se	CONTRACTOR OF A DATE	the second se
A REPORT OF A REPORT OF	Contraction of the second	and the second	A PROPERTY OF A	The subscription of the local division of the local division of the local division of the local division of the	THE R. LEWIS CO.

1.4%	15,625	建取得	Sec. Asp	無法な	(Sec.2)
			2.74	(The second	- States
2.3.3 2	1 1 1		1. 1. 1. 2.	1.2.2.2	1. 5. 6. 6.
10.24		34 A 4	12025	ETERIA	1 Star
民主任	and the second				The second
So they advantage	A REAL PROPERTY AND	Name of Street Street, St.	No. of the state of the	an owner with	9 8 8 M
开始使	122	100	122/2		1.200
1.1.1					15.1
	Contraction of the	Contract for a contract	CONTRACTOR OF STREET	10. Test 10. Date:	and any of the second second
1712	1.1	1 Sugar	12/21/2	i netr	and the second
Contract of the second	COLUMN BOLIES	O seine de l		CO. In Charles	Contraction of the
12/2/2/	12/2/2/	and a second	1410	1274	States and
51.12			Same	1.1.2	
The Course of the			Constraint State of St		29-20 -01 -01
17-11-12		2.44	and states	1000	Sec. 2
1.1.4.21	12 1/2/	e ser se	1.1.1.1	1992	1222
100 MP 100 PF	S 5 6 6 12	Contraction of the	COLOR OF STREET		Street Artist

**SVHNs** 





epartment of Computer Science and Technology, Xiamen University

MNIST

#### GAN Starts an Era



2023





Image source: Ian Goodfellow. Samples from Goodfellow et al., 2014, Radford et al., 2015, Liu et al., 2016, Karras et al., 2017, Karras et al., 2018, Ho et al. 2020, Zhang et al. 2023

 Unsupervised representation learning with deep convolutional generative adversarial networks

 A Radford, L Metz, S Chintala - arXiv preprint arXiv:1511.06434, 2015 - arxiv.org

 ... REPRESENTATION LEARNING FROM UNLABELED DATA Unsupervised representation learning

 Learning ... A classic approach to unsupervised representation learning is to do clustering on the ...

 ☆ Save 切 Cite Cited by 16206 Related articles All 5 versions ≫

- Vanilla GAN simply uses MLP, rather than CNN in both generator and discriminator.
- CNN can be easily applied to discriminator.
- Now the problem is: how can CNN be used as a generator?
  - Pooling leads to downsampling, how to upsampling?





























- Fractionally-strided convolutions are also called transposed convolutions.
  - PyTorch: torch.nn.ConvTranspose2d.
  - TensorFlow: tf.keras.layers.Conv2DTranspose.
- Some researchers are used to call deconvolutions. However, true deconvolutions are the inverse operation of convolution, which is not the same as fractionally-strided convolutions.



Transposed convolution with stride is equivalent to convolving with zero-padding and inserting zeros.





More linear algebra details: https://arxiv.org/pdf/1603.07285.pdf

48

#### DCGAN



Architecture guidelines for stable Deep Convolutional GANs

- Replace any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator).
   stride=2 everywhere
- Use batchnorm in both the generator and the discriminator.
- Remove fully connected hidden layers for deeper architectures.
- Use ReLU activation in generator for all layers except for the output, which uses Tanh.
- Use LeakyReLU activation in the discriminator for all layers.

**厦門大學信息学院(特色化示范性软件学院)** School of Informatics Xiamen University (National Characteristic Demonstration Software School)



#### **DCGAN:** Visual Results



The generated bedrooms look very nice (at that time)!





#### DCGAN: Walking in the Latent Space

If walking in this latent space results in semantic changes to the image generations (such as objects being added and removed), we can reason that the model has learned relevant and interesting representations.



Interpolation between a series of 9 random points in Z show that the space learned has smooth transitions.





#### **DCGAN: Vector Arithmetic**



For each column, the Z vectors of samples are averaged. Arithmetic was then performed on the mean vectors creating a new vector Y.

厦門大學信息学院(特色化示范性软件学院) School of Informatics Xiamen University (National Characteristic Demonstration Software School)



## DCGAN: Use as Feature Extractor

- Train on Imagenet-1k and then use the discriminator's convolutional features from all layers.
- Maxpooling each layers representation to produce a 4  $\times$  4 spatial grid.
- These features are then flattened and concatenated to form a 28672 dimensional vector.
- A regularized linear L2-SVM classifier is trained on top of them.

信息学院(特色化示范性软

Model	Accuracy	Accuracy (400 per class)	max # of features units
1 Layer K-means	80.6%	63.7% (±0.7%)	4800
3 Layer K-means Learned RF	82.0%	70.7% (±0.7%)	3200
View Invariant K-means	81.9%	$72.6\%~(\pm 0.7\%)$	6400
Exemplar CNN	84.3%	$77.4\%~(\pm 0.2\%)$	1024
DCGAN (ours) + L2-SVM	82.8%	73.8% (±0.4%)	512







## • We can't control what we generate from the vanilla GAN.

- Noise is the only input and it is totally random.
- How can we tell GAN what we want it to generate?
- Straightforward solution: replace data distribution by conditional distribution.

$$p(\mathbf{x}) \rightarrow p(\mathbf{x}|\mathbf{y}).$$

- Now, the problem becomes:
  - Generator: generate a sample for class y.
  - Discriminator: distinguish the real sample in class y and the generated sample in class y.





#### CGAN

Both generator and discriminator are conditioned on some extra information y:

$$\min_{\theta_g} \max_{\theta_d} \left[ \mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(\boldsymbol{x} | \boldsymbol{y}) + \mathbb{E}_{z \sim p(z)} \log \left( 1 - D_{\theta_d} \left( G_{\theta_g}(\boldsymbol{z} | \boldsymbol{y}) \right) \right) \right].$$

- y could be any kind of auxiliary information, such as class labels or data from other modalities.
  - E.g. the speech of saying that class.











Generate MNIST digits by directly feeding one-hot class label.





56

## pix2pix

 Image-to-image translation with conditional adversarial networks

 P Isola, JY Zhu, T Zhou... - Proceedings of the IEEE ..., 2017 - openaccess.thecvf.com

 ... In analogy to automatic language translation, we define automatic image-to-image translation as the problem of translating one possible representation of a scene into another, given ...

 ☆ Save 勁 Cite Cited by 20433 Related articles All 24 versions ≫

## Given a pair of images, transfer the style of one image to another.



Image source: Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.

#### pix2pix

- The model is based on CGAN.
- As an improvement, the generator is tasked to not only fool the discriminator but also to be near the ground truth output.
- L<sub>1</sub> penalization is added to the loss of CGAN to encourage less blurring:

$$L_{L_1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1].$$



Image source: Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.

## pix2pix: More Applications







Background removal



by Kaihu Chen

Sketch  $\rightarrow$  Pokemon



by Bertrand Gondouin



Image source: Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.

## CycleGAN

#### **Unpaired image-to-image** translation using cycle-consistent adversarial networks

<u>JY Zhu, T Park, P Isola, AA Efros</u> - Proceedings of the IEEE ..., 2017 - openaccess.thecvf.com ... losses on domains X and Y yields our full objective for **unpaired image-to-image** translation. ... We first compare our approach against recent methods for **unpaired image-to-image** ... ☆ Save 50 Cite Cited by 20082 Related articles All 26 versions ≫

- Paired examples can be expensive to obtain.
- ■Can we translate from X ↔ Y in an unsupervised manner?



#### Paired vs. unpaired examples





## CycleGAN

- Two generators:
  - $G: X \to Y;$
  - $F: Y \to X.$
- Two discriminators:
  - D<sub>X</sub> aims to distinguish between images {x} and translated images {F(y)};
  - $D_Y$  aims to discriminate between  $\{y\}$  and  $\{G(x)\}$ .







#### CycleGAN

If we can go from X to  $\hat{Y}$  via G, then it should also be possible to go from  $\hat{Y}$  back to X via F.

Cycle consistency loss is added to the original adversarial loss:

 $L_{\rm cyc}(G,F) = \mathbb{E}_{x}\left[\left\|F(G(x)) - x\right\|_{1}\right] + \mathbb{E}_{y}\left[\left\|G(F(x)) - y\right\|_{1}\right].$ 



Image source: Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.

62



Image source: Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.

# **DIFFUSION MODEL**



#### **Denoising diffusion probabilistic models**

<u>J Ho</u>, <u>A Jain</u>, <u>P Abbeel</u> - Advances in neural information ..., 2020 - proceedings.neurips.cc ... This paper presents progress in **diffusion probabilistic models** [53]. A **diffusion probabilistic** 

model (which we will call a "diffusion model" for brevity) is a parameterized Markov chain ...  $\therefore$  Save  $\mathfrak{D}$  Cite Cited by 4825 Related articles All 6 versions  $\gg$ 



Reference: Hung-yi Lee, Diffusion model (https://speech.ee.ntu.edu.tw/~hylee/ml/ml2023-course-data/DiffusionModel%20(v2).pdf



Reference: Hung-yi Lee, Diffusion model (https://speech.ee.ntu.edu.tw/~hylee/ml/ml2023-course-data/DiffusionModel%20(v2).pdf )







67

Reference: Hung-yi Lee, Diffusion model (https://speech.ee.ntu.edu.tw/~hylee/ml/ml2023-course-data/DiffusionModel%20(v2).pdf





Reference: Hung-yi Lee, Diffusion model (https://speech.ee.ntu.edu.tw/~hylee/ml/ml2023-course-data/DiffusionModel%20(v2).pdf )







Image source: Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851.



#### Unconditional CIFAR10 progressive generation





Image source: Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851.
## **Diffusion Model**



Latent  $\mathbf{x}_{750}$ 

Decodings  $\mathbf{x}_0 \sim p_{\theta}(\mathbf{x}_0 | \mathbf{x}_{750})$ 





Image source: https://hojonathanho.github.io/diffusion/



## How can we generate desired image?







Reference: Hung-yi Lee, Diffusion model (https://speech.ee.ntu.edu.tw/~hylee/ml/ml2023-course-data/DiffusionModel%20(v2).pdf

## **Diffusion Model**



Reference: Hung-yi Lee, Diffusion model (https://speech.ee.ntu.edu.tw/~hylee/ml/ml2023-course-data/DiffusionModel%20(v2).pdf

High-resolution image synthesis with latent diffusion models
<u>R Rombach, A Blattmann</u>, D Lorenz... - Proceedings of the ..., 2022 - openaccess.thecvf.com
… To lower the computational demands of training diffusion models towards high-resolution **image synthesis**, we observe that although diffusion models allow to ignore perceptually ...
☆ Save 切 Cite Cited by 3600 Related articles All 8 versions ≫







# Layout-to-image synthesis on COCO

Text-to-image *LDM* model for userdefined text prompts







#### Semantic synthesis of landscape images







#### ImageNet $64 \rightarrow 256$ super-resolution





79



#### Image inpainting with latent diffusion





80

## **Stable Diffusion**

Stable diffusion is developed by researchers from the CompVis Group at Ludwig Maximilian University of Munich and Runway with a compute donation by Stability AI and training data from non-profit organizations.



#### Stable Diffusion

Get involved with the fastest growing open software project. Download and join other developers in creating incredible applications with Stable Diffusion as a foundation model.

Try Stable Diffusion

Download Code





### DALL·E



Image source: Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. "Hierarchical text-conditional image generation with clip latents." arXiv preprint arXiv:2204.06125 1, no. 2 (2022): 3.

#### Imagen



"A Golden Retriever dog wearing a blue checkered beret and red dotted turtleneck."









Sprouts in the shape of text 'Imagen' coming out of a A photo of a Shiba Inu dog with a backpack riding a A high contrast portrait of a very happy fuzzy panda fairytale book. bike. It is wearing sunglasses and a beach hat. There is a painting of flowers on the wall behind him.











Teddy bears swimming at the Olympics 400m Butter-fly event. A cute corgi lives in a house made out of sushi.

A cute sloth holding a small treasure chest. A bright golden glow is coming from the chest.







A brain riding a rocketship heading towards the moon. A dragon fruit wearing karate belt in the snow.

A strawberry mug filled with white sesame seeds. The mug is floating in a dark chocolate sea.









Reference: https://imagen.research.google/

## Midjourney



These fake images of the Pope and Donald Trump have gone viral in recent weeks, both were created on Midjourney v5.





Image source: https://petapixel.com/2023/03/30/midjourney-ends-free-trials-after-fake-ai-images-go-viral/

## ControlNet

- A new concept called zero convolution layers, with weights initialized to zeros so that they progressively grow during the training.
- This architecture ensures that harmful noise is not added to the deep features of the large diffusion model.



Adding conditional control to text-to-image diffusion models L Zhang, <u>A Rao</u>, <u>M Agrawala</u> - Proceedings of the IEEE/CVF ..., 2023 - openaccess.thecvf.com ... Learning conditional controls for large text-to-image diffusion ... network architecture that learns conditional controls for large ... for learning diverse conditional controls. The trainable copy ... ☆ Save 奶 Cite Cited by 440 Related articles All 3 versions ≫

Condition  $c_{\rm f}$ 

zero convolution

(F)

SD Encoder Block A

64×64 (trainable copy)

SD Encoder Block B

32×32 (trainable copy)

×3

×3

85

Prompt&Time

×3

Prompt  $c_t$  Time t

Time

Encoder

Input  $z_t$ 

SD Encoder Block A

64×64

SD Encoder Block B

32×32

Text

Encoder

Source: Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3836-3847. 2023.

After this lecture, you should know:

- What is a generative model?
- How generator and discriminator improve each other?
- How does transposed convolution work?
- How to design application specific loss to train with adversarial loss?
- What is the main principle of diffusion model?





## Suggested Reading

## Adversarial Nets Papers

## Tips and tricks to make GANs work

■ <u>令人拍案叫绝的Wasserstein GAN</u>

■<u>李宏毅讲Diffusion Model</u>







## Assignment 4 is released. The deadline is 18:00, 11st December.







- Any question?
- Don't hesitate to send email to me for asking questions and discussion. ③



