Challenge:
High-dimensional Data Generation

Hao Dong

Peking University

Challenge: High-dimensional data generation

Past	**Now**	**Next**
64x64	**1K, 2K**	**Retina Screen**

We use images for demonstration

Progressive Growing of GANs for Improved Quality, Stability, and Variation. *T. Karras, T. Aila, et al. ICLR. 2018.*

Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

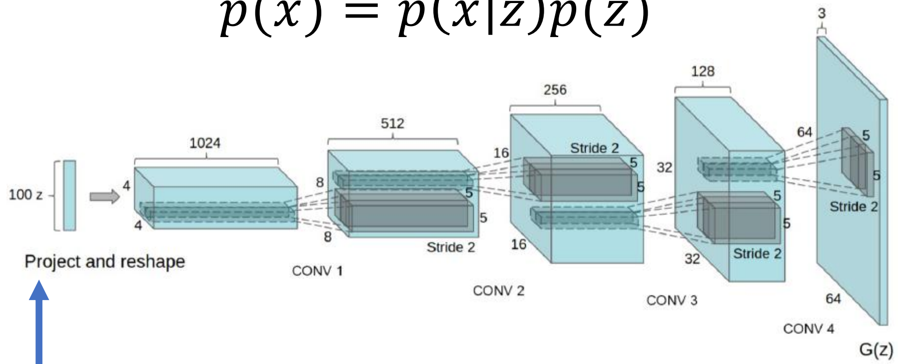
Formulation

Features/Latent Code
(e.g., the prior distribution, predefined features)

→

Large Scale(e.g. Resolution)
(e.g., image, video, …)

$$p(x) = p(x|z)p(z)$$



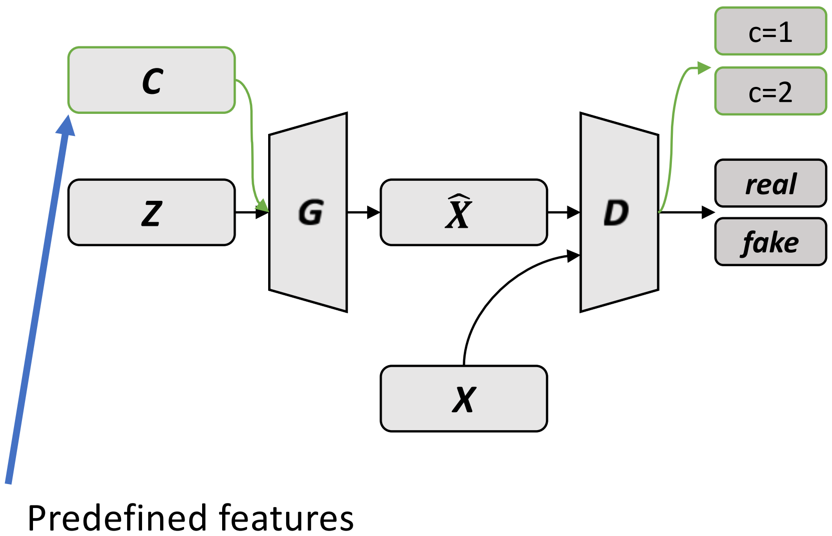
(Prior) Normal Distribution
z = 100 values

Formulation

Features
(e.g., the prior distribution, predefined features)

Large Scale(e.g. Resolution)
(e.g., image, video, …)

$$p(x|z,c)$$



- **Shape deformation! (Locally & Globally)**

monarch butterfly goldfinch daisy

Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

CG-based Methods

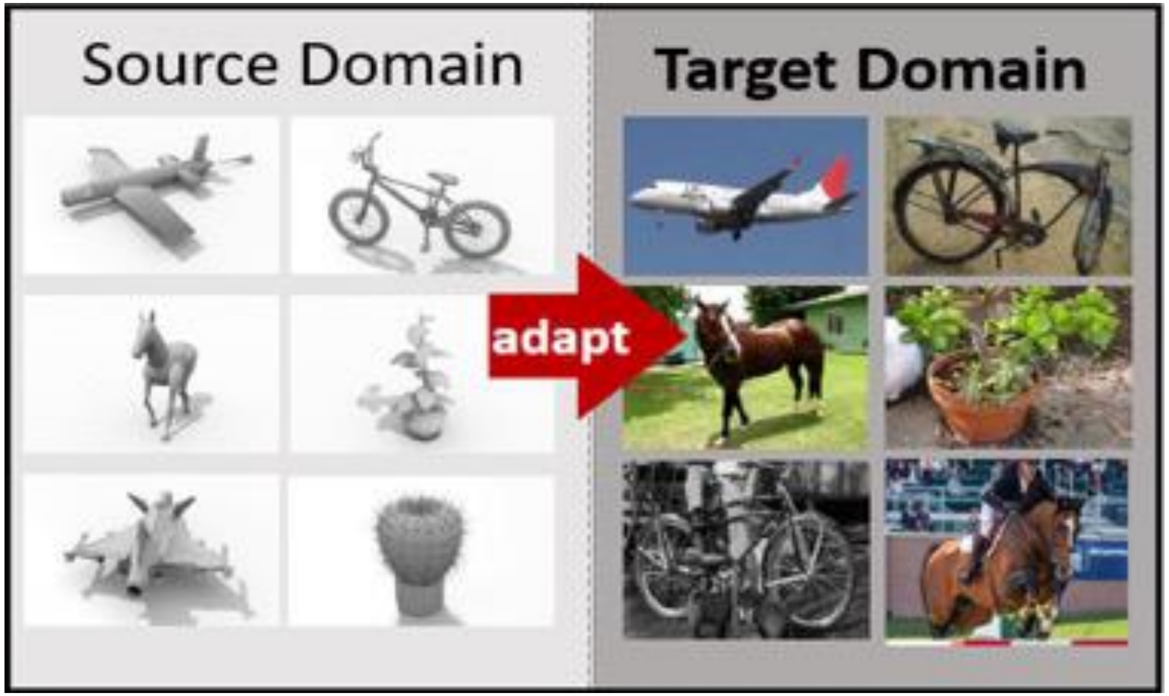
- Fully CG-based
- Hybrids

GTA 5

- **Pros:**
  - **Reasonable Structure**
  - **As “structure” is relatively more well-defined.**
- **Cons:**
  - **Distorted Details**
  - **We cannot well-define “What is a human face” or “What is real wall texture”.**

CG-based Methods

- Fully CG-based
- Hybrids
  - Computer Graphics + GAN



Peng, X., Usman, B et al. (2018). VisDA: A synthetic-to-real benchmark for visual domain adaptation. *CVPR Workshops*

CG-based Methods

- Fully CG-based
- Hybrids
  - Computer Graphics + GAN

Input GTA5 CG

Output image with German street view style

Unsupervised image-to-image translation networks. *M.Y. Liu, T. Breuel, J. Kautz. NIPS. 2017*
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *J. Zhu, T. Park et al. ICCV 2017.*

CG-based Methods

- Fully CG-based
- Hybrids
- Limitation
  - Need prior knowledge
    - Intensive Engineering…
  - Limited Generalisation
    - Artificially designed generation rule can only capture limited latent structure of domain
  - Improvement need more prior …
    - Anyway, automatically learning prior knowledge is necessary.

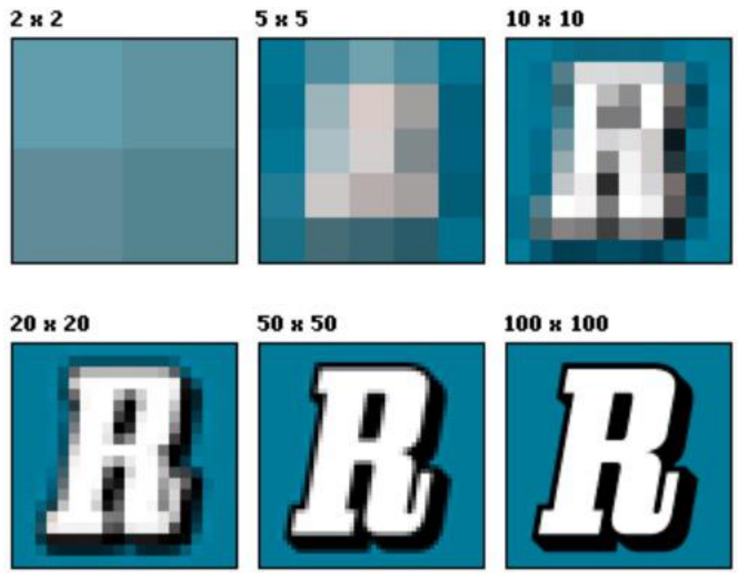
Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

Deep Methods

- As resolution grows, high-level information contained in same image grows much slower than low-level features

- As shown on the right From 50x50 -> 100x100
  - “High-level” information grows much slower
  - “low-level” information keeps growing
  - Intensively modelling of details



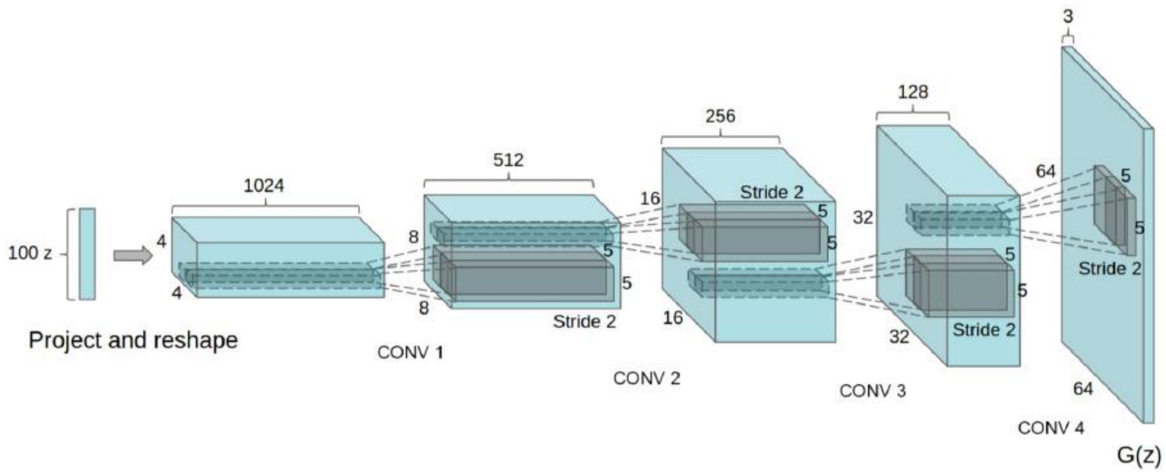
- Note that if we want to keep “R”’s structure, then we have to keep all pixels’ relative position fixed and average distance between each pixel-pair is proportional to resolution.
  - Long-range dependency problem

Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

Progressive GAN

- Recap: DCGAN



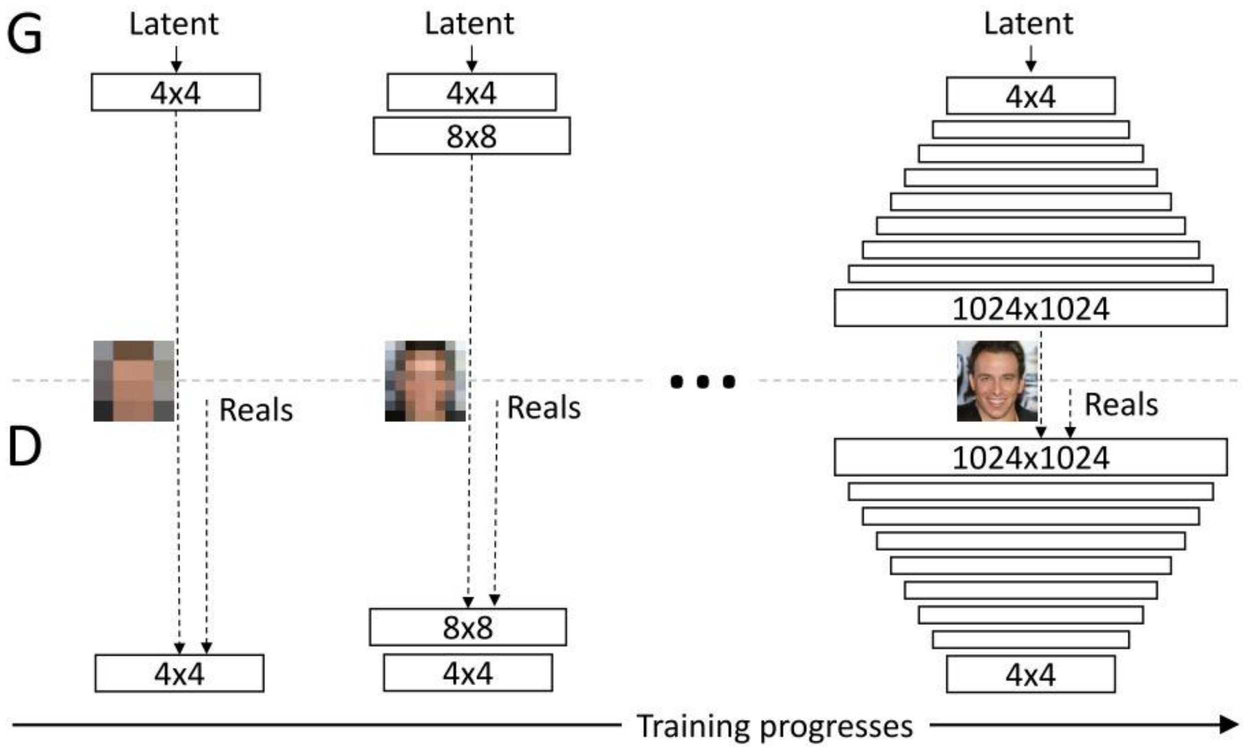
- Difficult to scale:
  - Unstable training
  - Computer memory constraints
  - High resolution images make the discriminator easier to discriminate the fake and real images, amplifying the gradient problem.

64x64 work !

1024x1024 fail ……

Progressive GAN

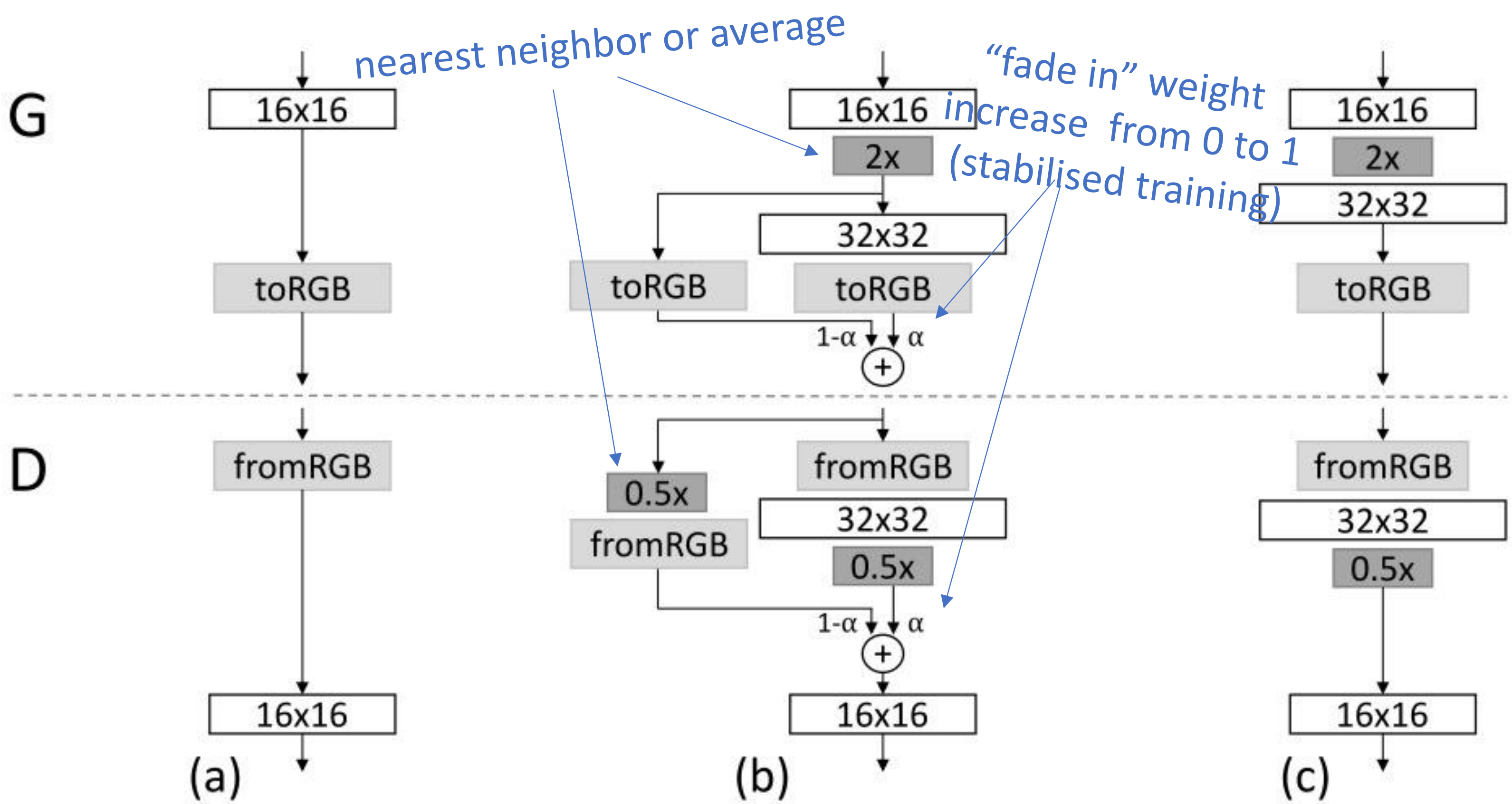
- From Coarse to Fine : 4x4 → 8x8 → 16x16 → 32x32 …. →1024x1024



Uality, Q., Tability, S., Ariation, V., & Karras, T. 2018 ICLR. Progressive Growing of GANs for Improved Quality, Stability, and Variation

Progressive GAN

- From Coarse to Fine : 4x4 → 8x8 → 16x16 → 32x32 …. →1024x1024



Uality, Q., Tability, S., Ariation, V., & Karras, T. 2018 ICLR. Progressive Growing of GANs for Improved Quality, Stability, and Variation

Progressive GAN

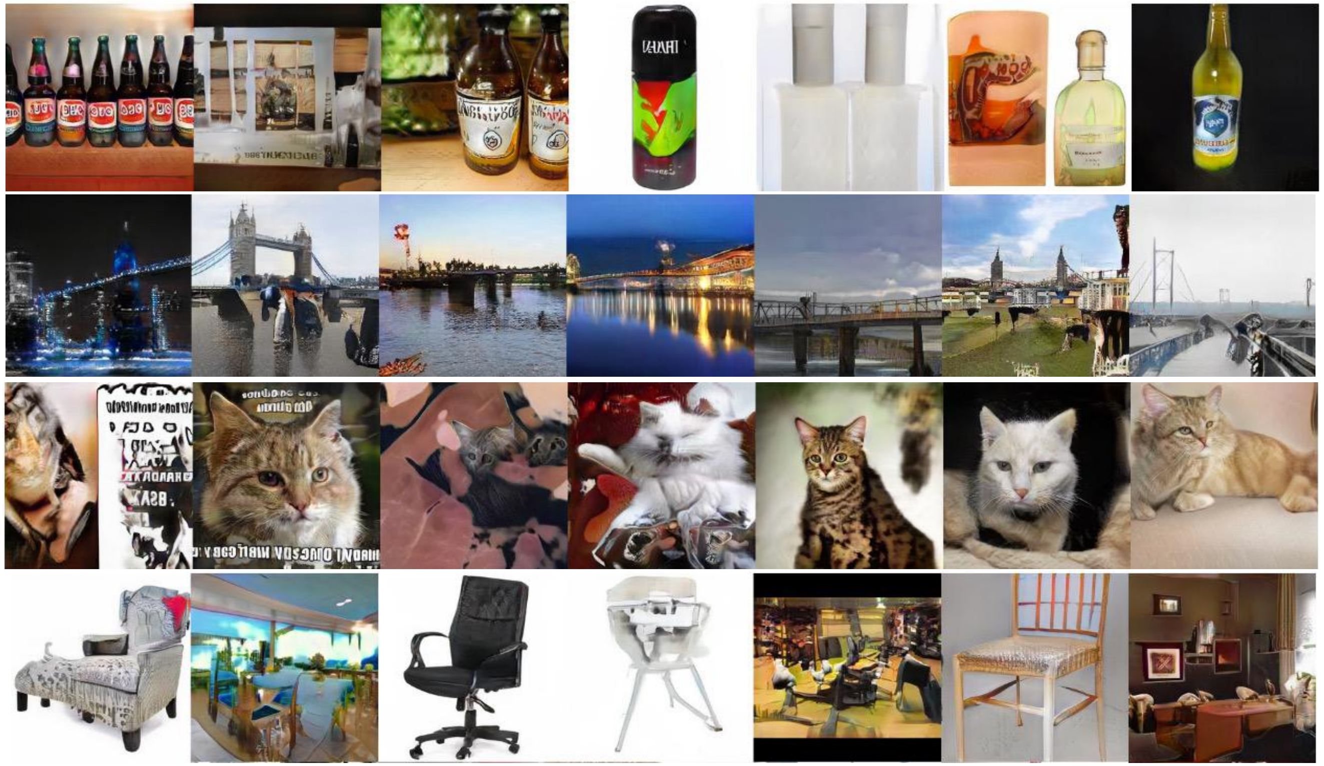
- From Coarse to Fine **with Condition**

Bottle

Bridge

Cat

Chair



Uality, Q., Tability, S., Ariation, V., & Karras, T. 2018 ICLR. Progressive Growing of GANs for Improved Quality, Stability, and Variation

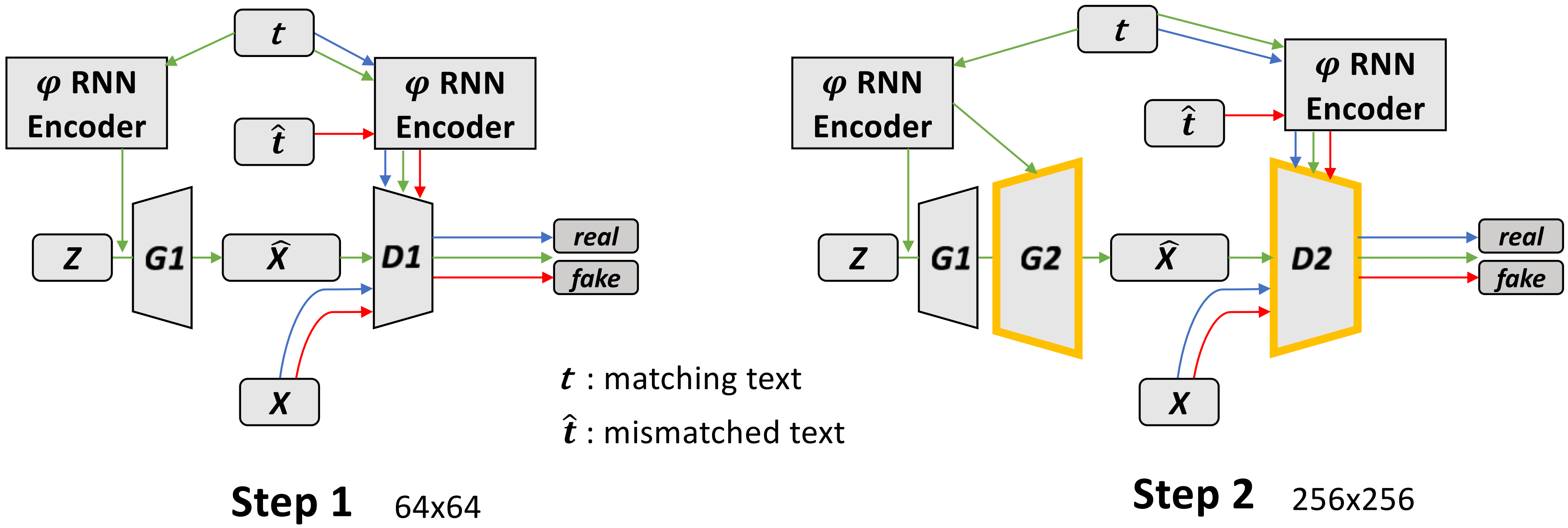
Progressive GAN → StackGAN

- From Coarse to Fine: **Text-to-Image Synthesis**

Zhang, H., Xu, T, te al. (2017) ICCV. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.

Progressive GAN → StackGAN

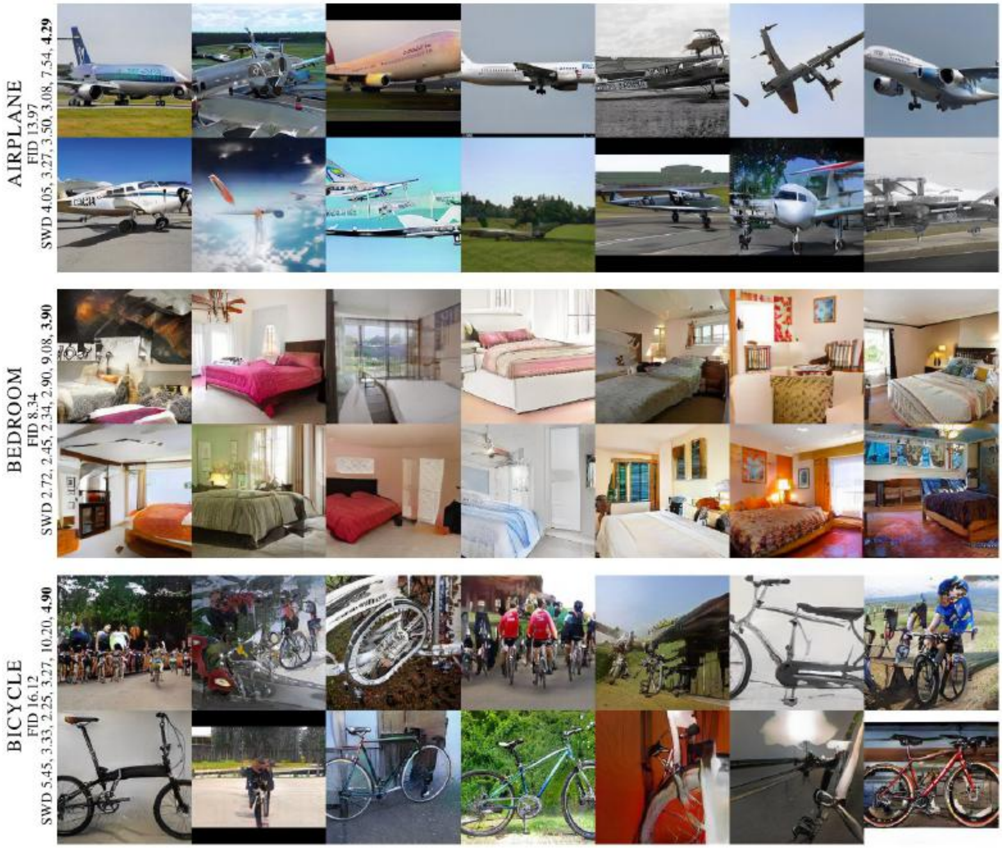
- From Coarse to Fine: Text-to-Image Synthesis



Zhang, H., Xu, T, te al. (2017) ICCV. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks.

Progressive GAN

- Question: Can Computer Graphic Generates This?



Uality, Q., Tability, S., Ariation, V., & Karras, T. 2018 ICLR. Progressive Growing of GANs for Improved Quality, Stability, and Variation

Challenge: High-dimensional data generation

PEKING UNIVERSITY

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

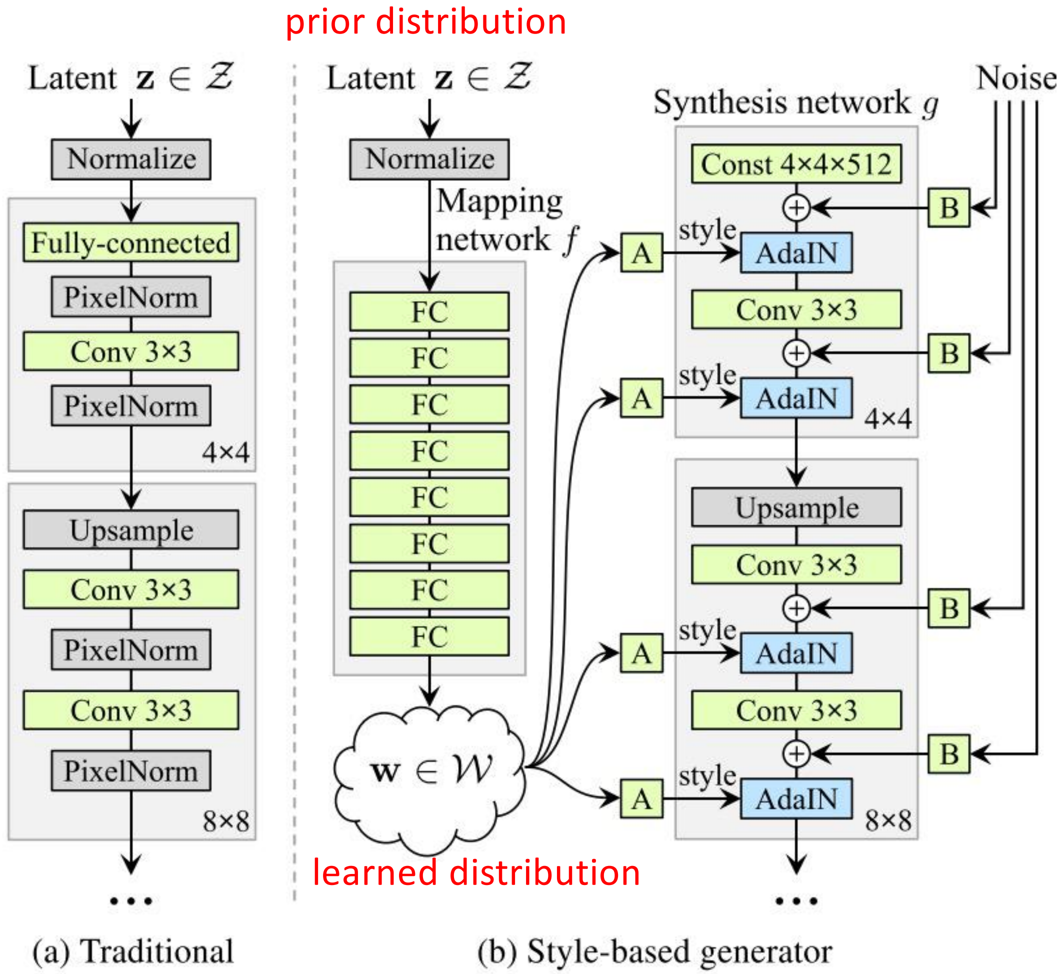
StyleGAN

- Insert Feature as Style Transfer

Adaptative Normalisation:

$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i}$$

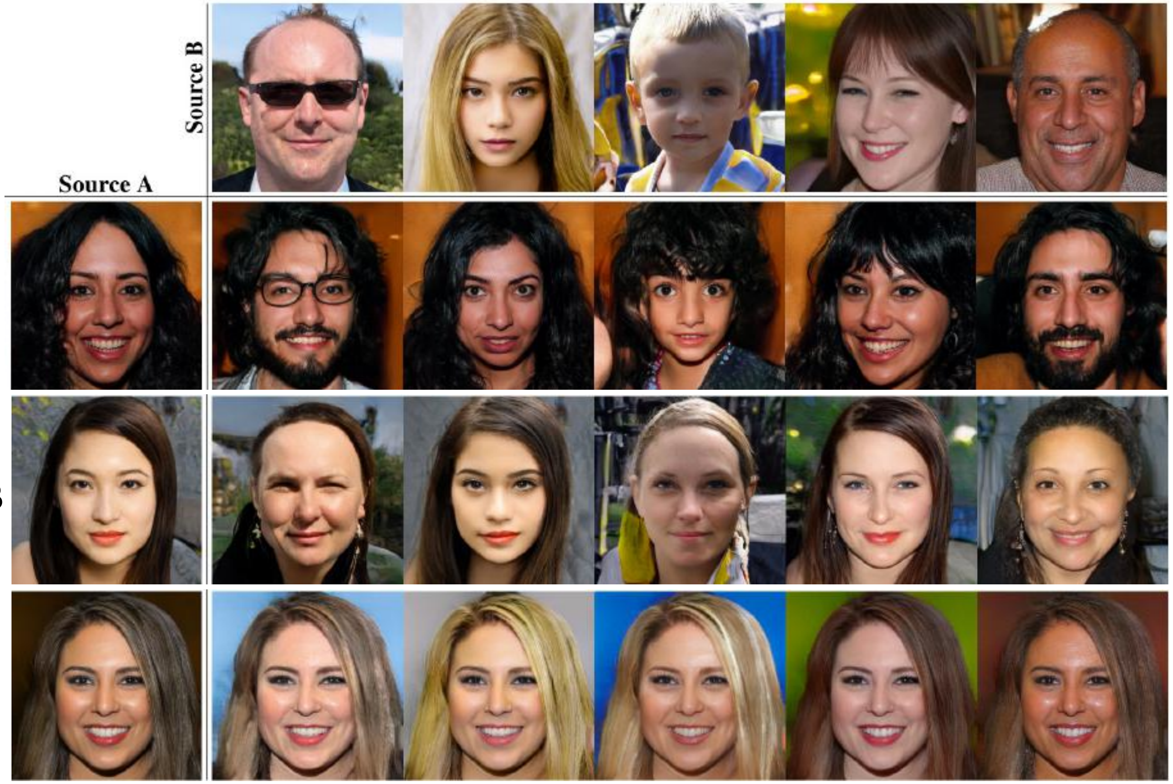
where each feature map x_i is normalised separately, and then scaled and biased using the corresponding scalar components from style y.



A Style-Based Generator Architecture for Generative Adversarial Networks. Ero Karras, Samuli Laine, Timo Aila. arXiv 2018

StyleGAN

- Hierarchical Latent Code



Coarse style from source B — Pose, hair style, face shape, eyeglasses

Middle style from source B — Eye open/close, facial feature

Fine style from source B — Color, microstructure

A Style-Based Generator Architecture for Generative Adversarial Networks. Ero Karras, Samuli Laine, Timo Aila. arXiv 2018

StyleGAN

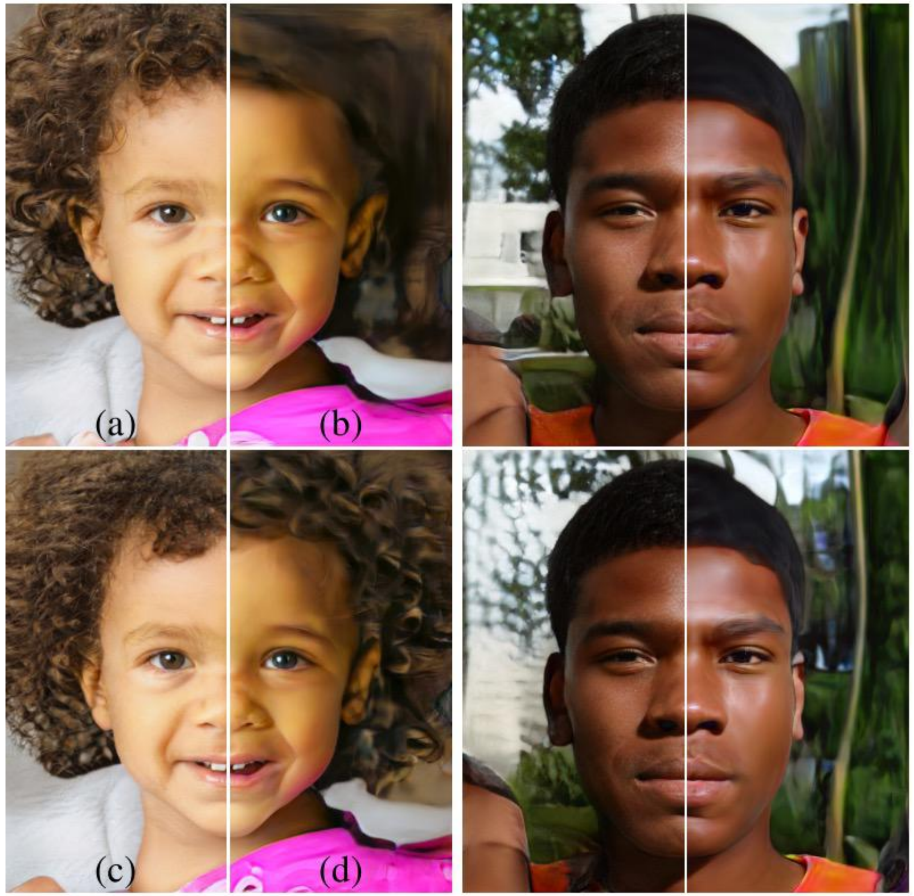
- Hierarchical Noise

(a) Noise is applied to all layers.

(b) No noise. look “smooth” (hair)

(c) Noise in fine layers only (64^2 – 1024^2). fine details

(b) Noise in coarse layers only (4^2 – 32^2). coarse details



A Style-Based Generator Architecture for Generative Adversarial Networks. Ero Karras, Samuli Laine, Timo Aila. arXiv 2018

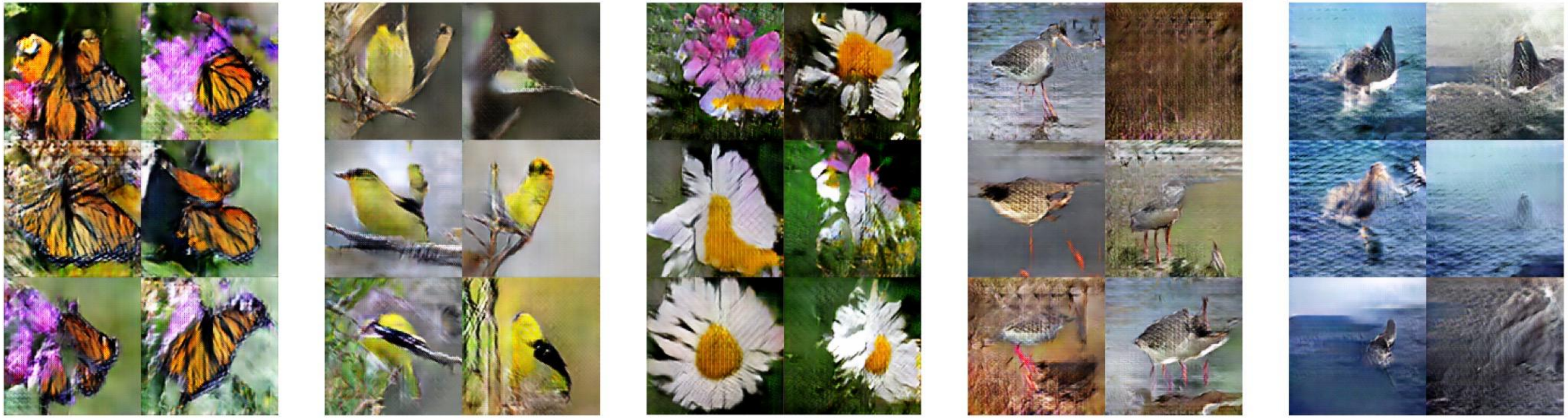
Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

SAGAN（Self-Attention GAN）

- Recap: Shape Deformation When Directly Scaling Up DCGAN
  - And recall that deep model’s challenges lie on
    - Intensively modelling details
    - Long range dependency

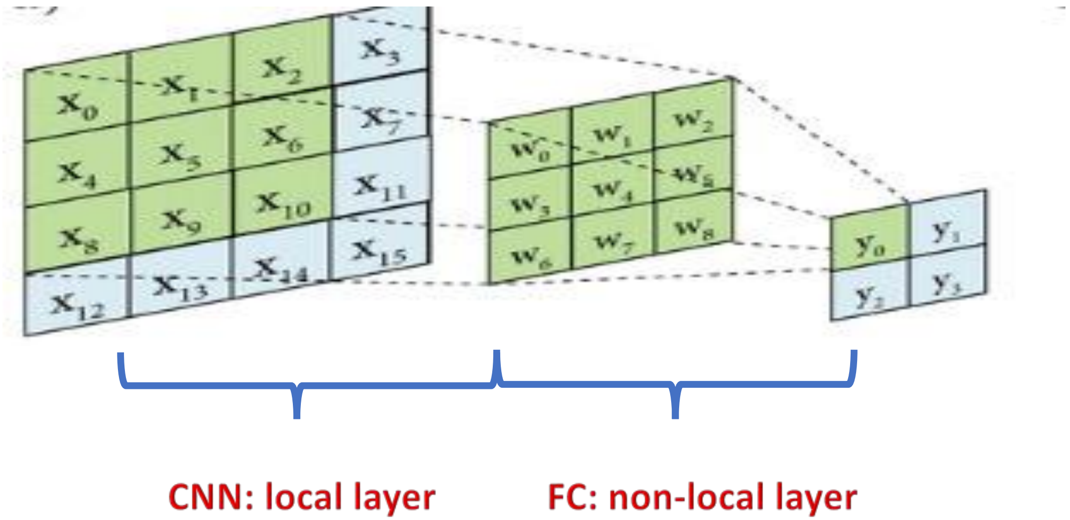
- CNN is a strong inductive bias to model natural details, but fails when modelling long range dependency.



monarch butterfly　　goldfinch　　daisy　　redshank　　grey whale

SAGAN

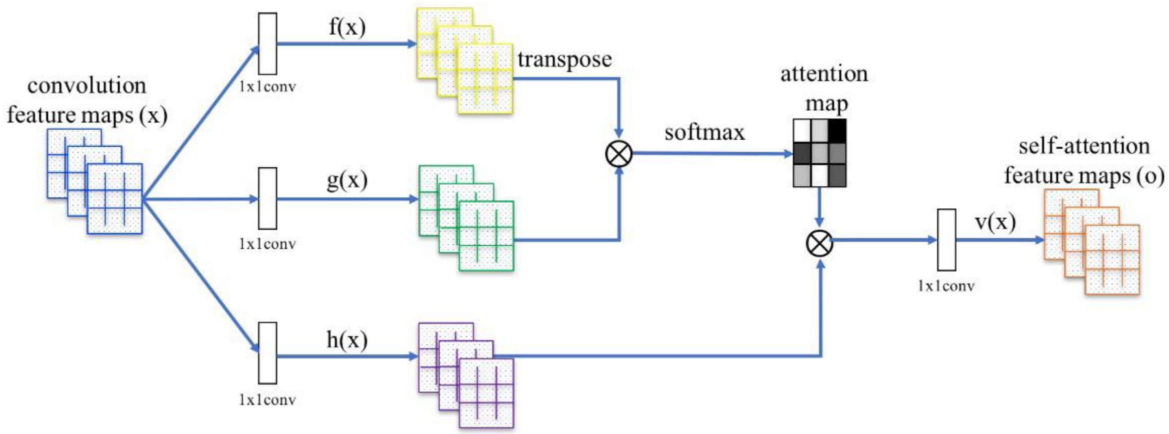
- Non-local layer vs Local layer
  - CNN is “local layer”, a neuron only observes part elements of the previous layer.

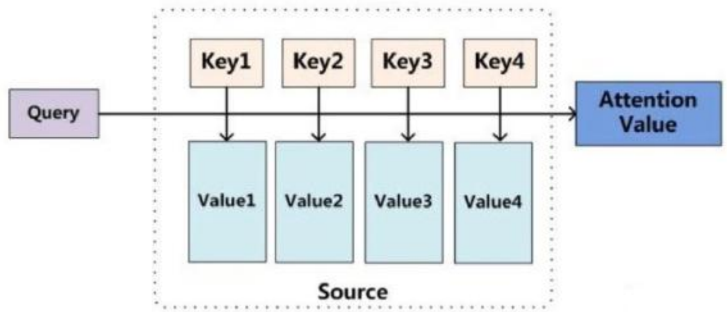


  - Which limits the network’s ability to capture global dependencies.

SAGAN

- SAGAN: Introduce attention layer into DCGAN backbone
  - Attention: have become an integral part of models that must capture global dependencies
  - Illustration of attention:

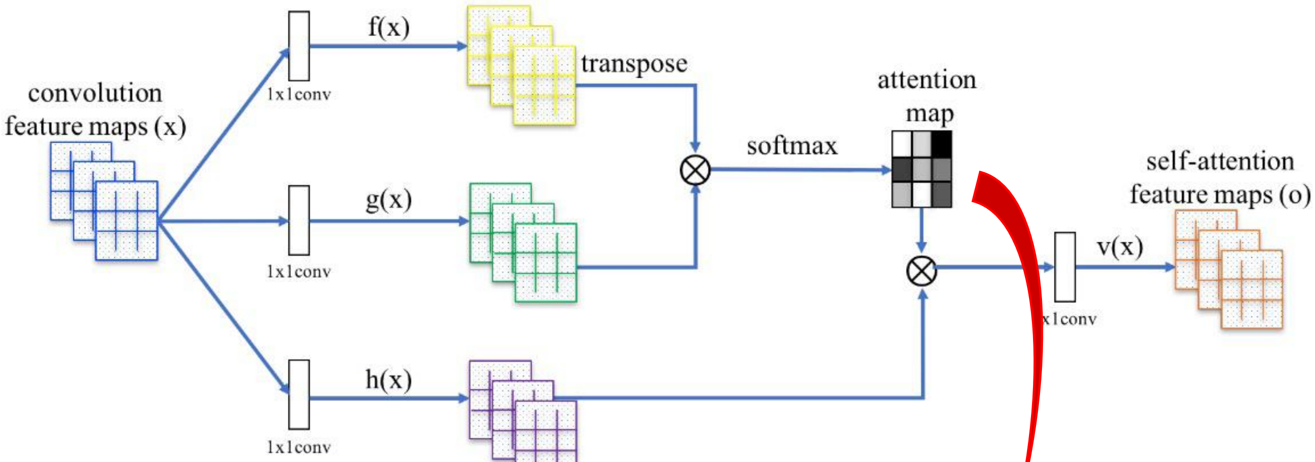


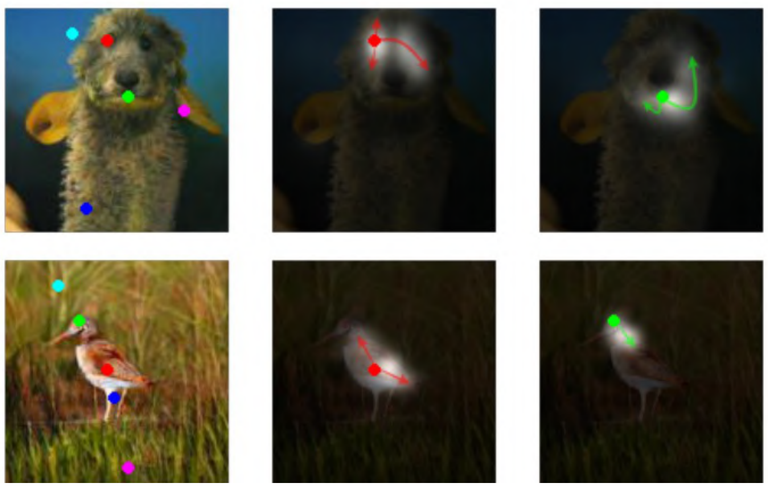


  - $Attention\ Value = \sum_{i=1}^{4} value_i \ * coefficient_i$ Convex combination of value w.r.t. coefficients
  - $coefficient_i = \frac{e^{-Key_i \circ Query}}{\sum_{j=1}^{4} e^{-Key_j \circ Query}}$ Correlation coefficient

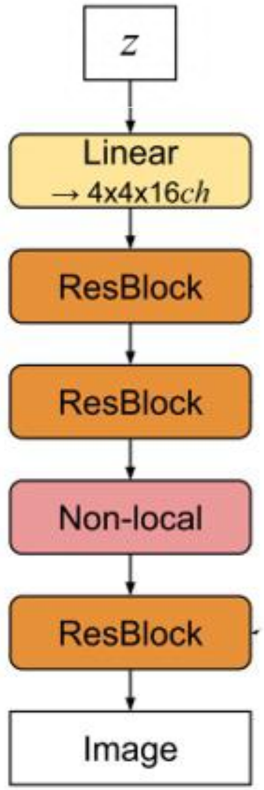
SAGAN

- SAGAN: Introduce attention layer into DCGAN backbone





Visualisation of attention
Coefficients (attention map)



Attention Block

SAGAN

- SAGAN: Introduce attention layer into DCGAN backbone

DCGAN

SAGAN

Challenge: High-dimensional data generation

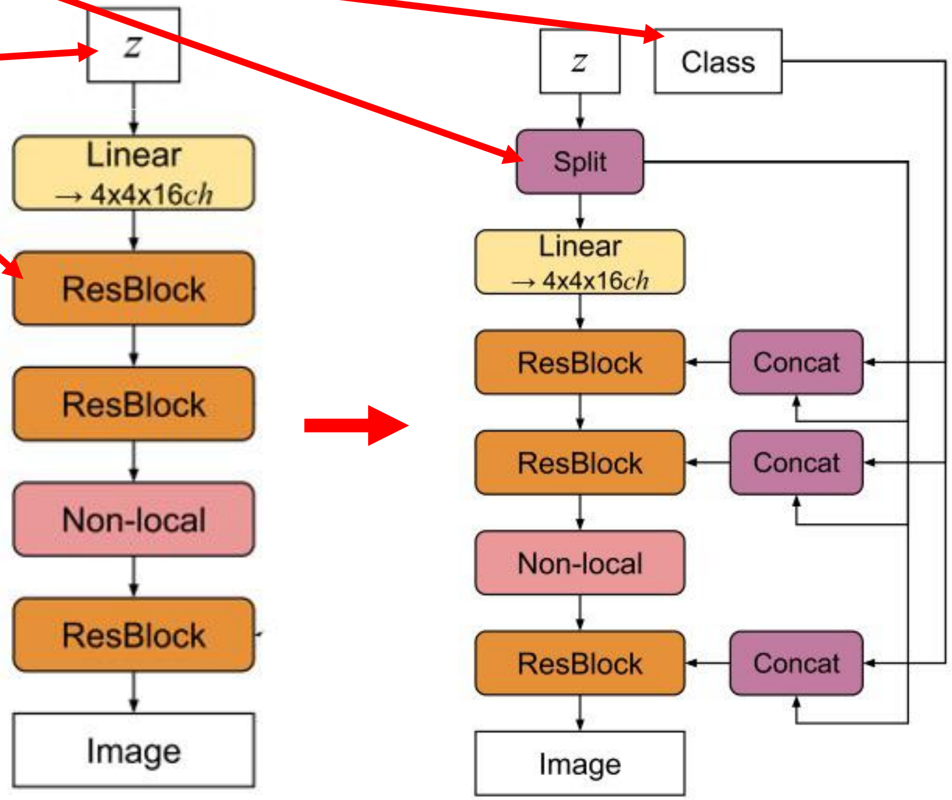
北京大學 PEKING UNIVERSITY

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

Big-GAN

- Big-GAN: Some novel tricks to scale up SAGAN + SAGAN backbone
  - 1. SAGAN -> conditional-SAGAN + skip-z
  - 2. 64x channel -> 96x channel
  - 3. 256x batch size -> 2048x batch size

- Ablation:
  - After applying 1:
    - Performance + 4%
    - Training speed + 18%
  - After applying 2:
    - IS + 21%
  - After applying 3:
    - IS + 50%



Big-GAN

- Big-GAN: Some novel tricks to scale up SAGAN + SAGAN backbone
  - 4. truncation trick
    - Using different latent distribution for sampling than used in training

Amenable to truncation

Not amenable

  - 5. orthogonal regularisation
    - Enforce Generator to be more amenable to truncation
    - Orthogonal regularisation can make G smoother

$$R_\beta(W) = \beta \|W^\top W \odot (\mathbf{1} - I)\|_{\mathrm{F}}^2,$$

Big-GAN

- Big-GAN: Some novel tricks to scale up SAGAN + SAGAN backbone
  - Samples generated by BigGAN at 256x resolution on ImageNet

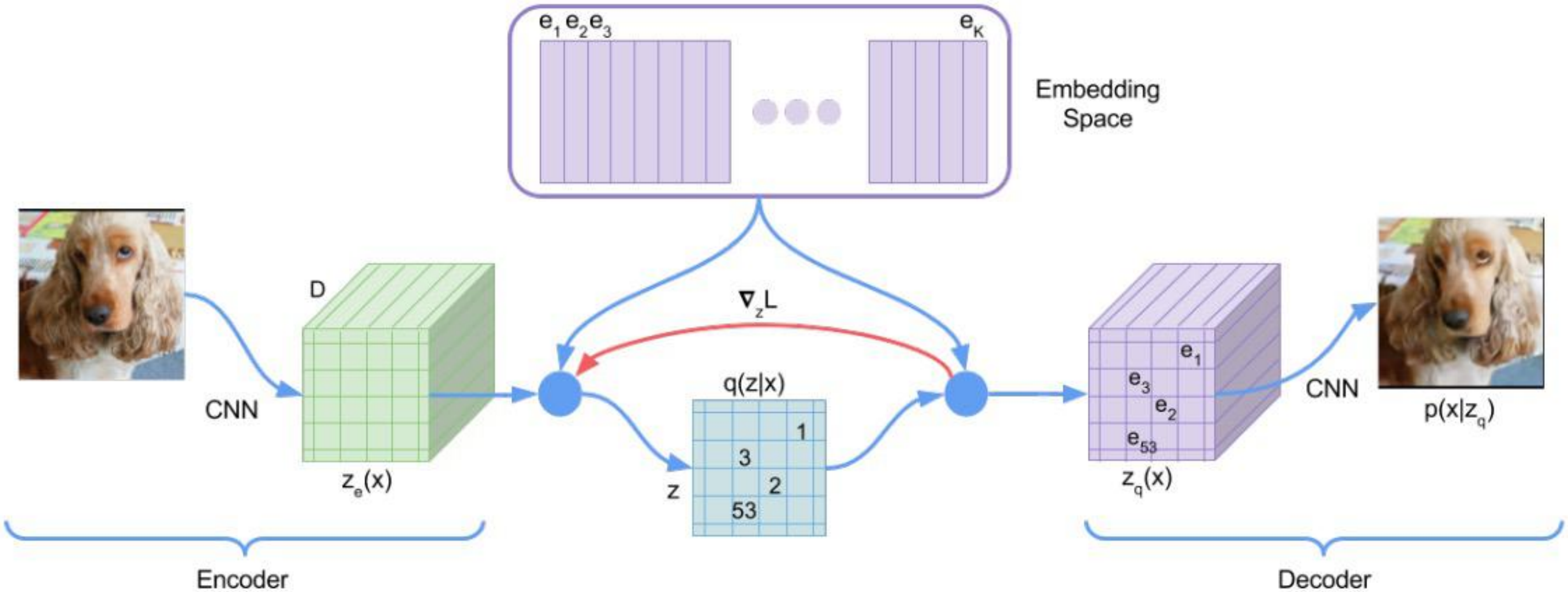
Challenge: High-dimensional data generation

北京大學 PEKING UNIVERSITY

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

VQ-VAE

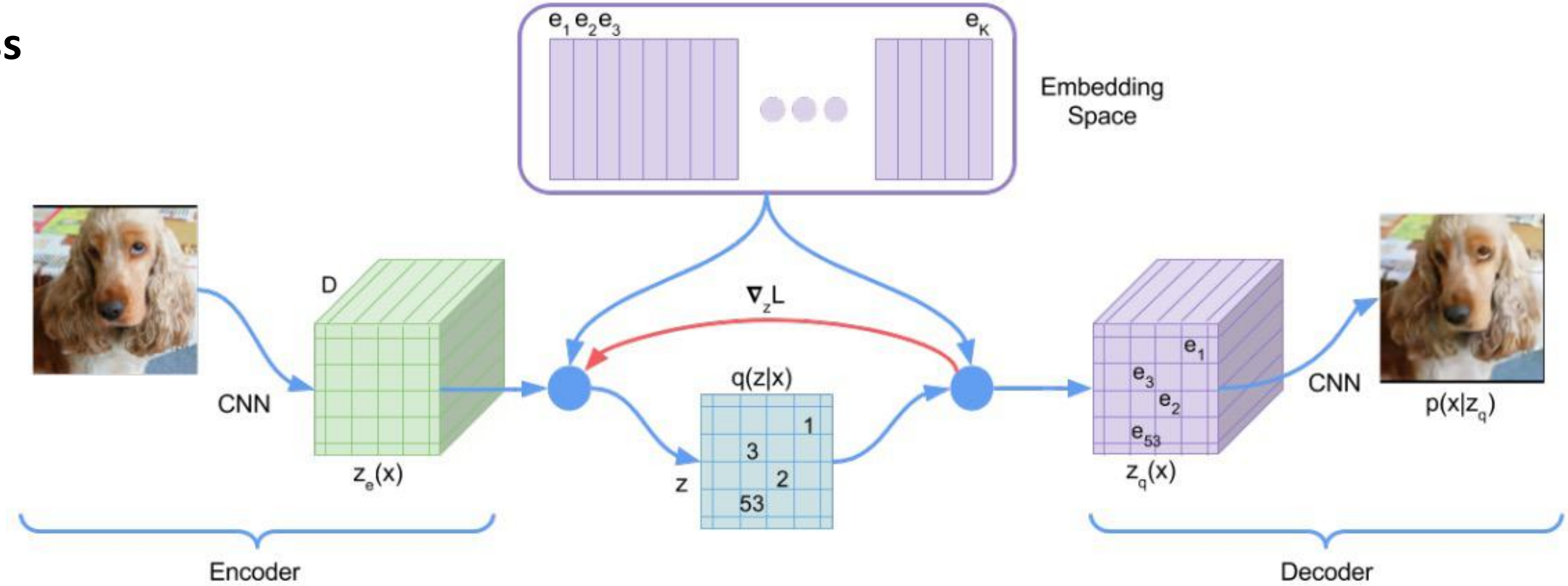
- **Straight-though Estimator**



VQ-VAE: Neural discrete representation learning. Van Den Oord, Aaron. Vinyals, Oriol. Kavukcuoglu, Koray. NeurIPS 2017

VQ-VAE

- **Loss**



$$L = \underbrace{\log p(x|z_q(x))}_{\text{Image reconstruction loss}} + \underbrace{\|\mathrm{sg}[z_e(x)] - e\|_2^2 + \beta\|z_e(x) - \mathrm{sg}[e]\|_2^2}_{\text{Make the quantised vector as close as the original vector}},$$

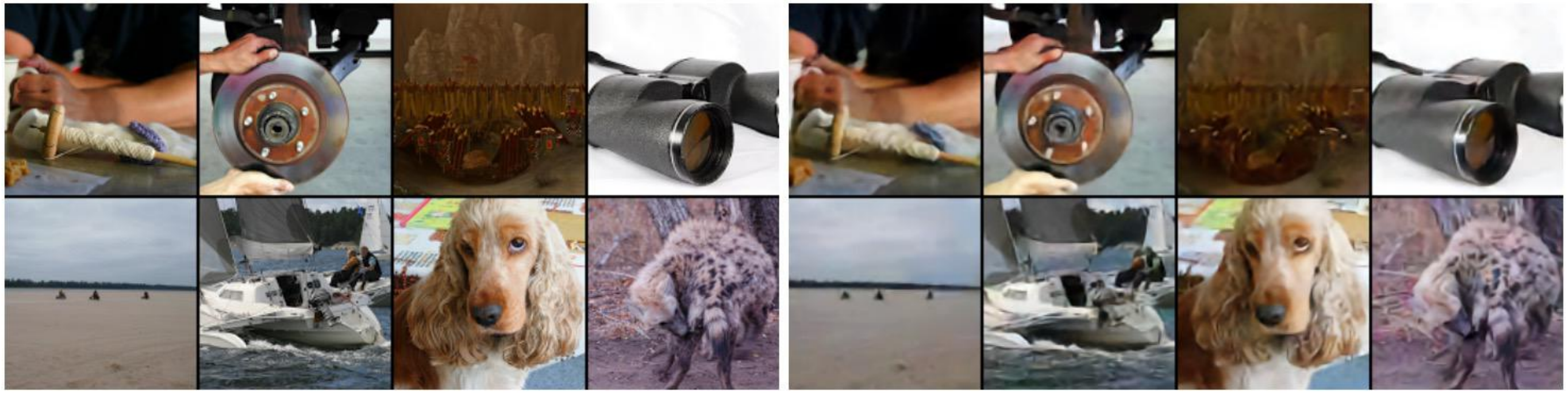
sg : the stop gradient operator that
e : quantised vector
$z_e(x)$: vector from the encoder output

VQ-VAE: Neural discrete representation learning. Van Den Oord, Aaron. Vinyals, Oriol. Kavukcuoglu, Koray. NeurIPS 2017

VQ-VAE

- **Results**

Input **Reconstruction**



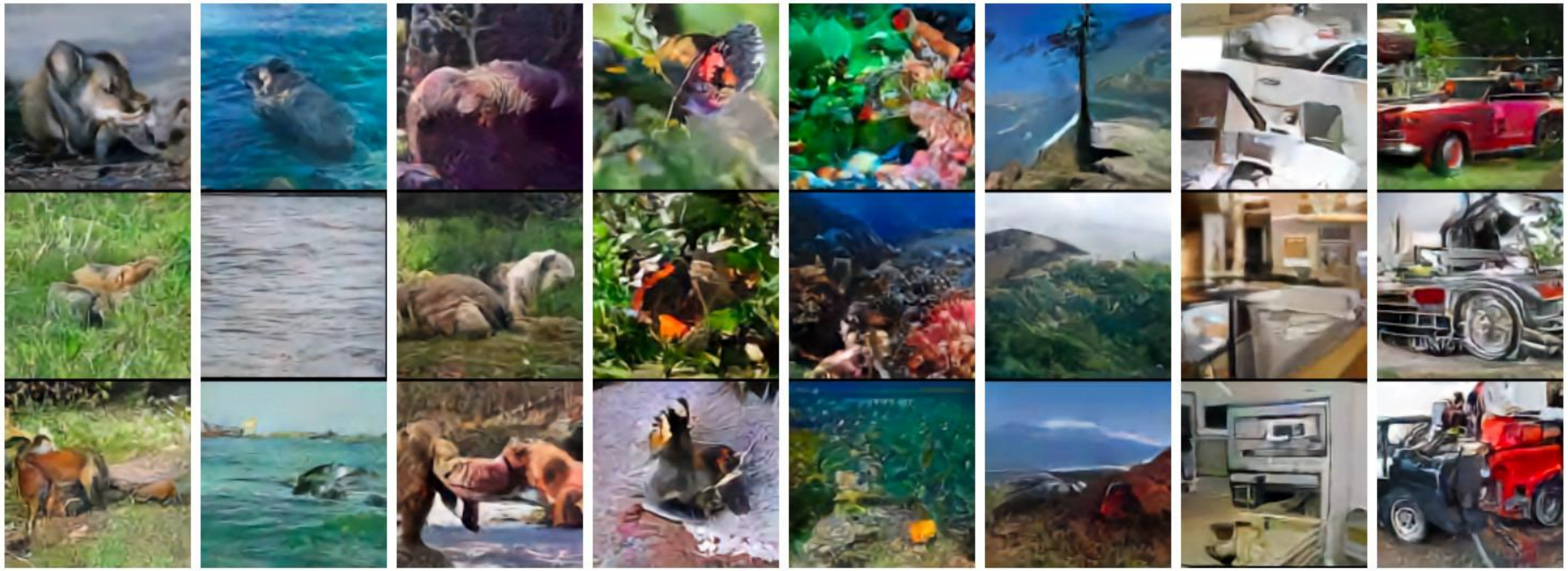
Images contain a lot of redundant information as most of the pixels are correlated and noisy, therefore learning models at the pixel level could be wasteful.

In this experiment we show that we can model $x = 128 \times 128 \times 3$ images by compressing them to a $z = 32 \times 32 \times 1$ discrete space (with K=512) via a purely deconvolutional $p(x|z)$. So a reduction of $\frac{128\times128\times3\times8}{32\times32\times9} \approx 42.6$ in bits. We model images by learning a powerful prior (PixelCNN) over z. This allows to not only greatly speed up training and sampling, but also to use the PixelCNNs capacity to capture the global structure instead of the low-level statistics of images.

VQ-VAE: Neural discrete representation learning. Van Den Oord, Aaron. Vinyals, Oriol. Kavukcuoglu, Koray. NeurIPS 2017

VQ-VAE

- **Results – Random Sampling**

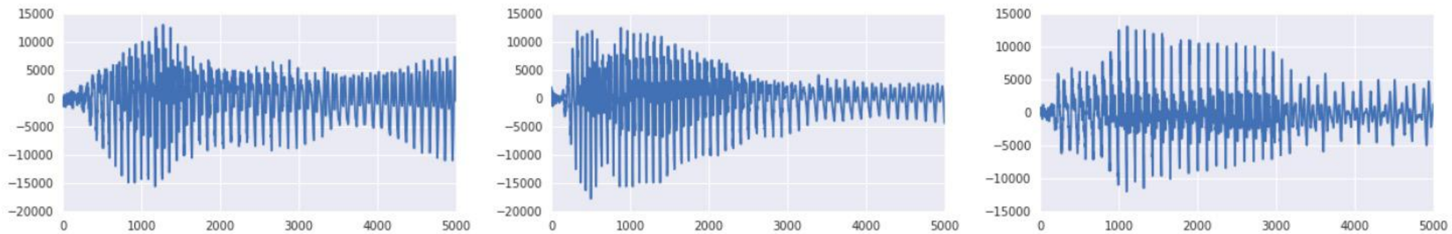


Samples (128x128) from a VQ-VAE with a PixelCNN prior trained on ImageNet images. From left to right: kit fox, gray whale, brown bear, admiral (butterfly), coral reef, alp, microwave, pickup.

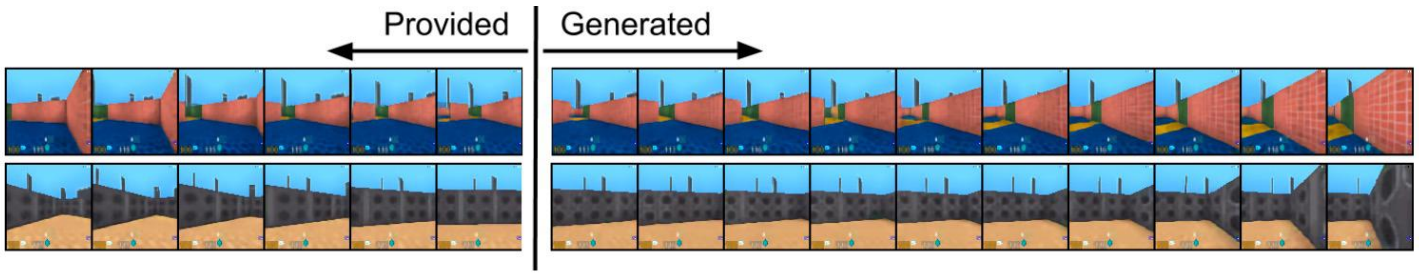
VQ-VAE: Neural discrete representation learning. Van Den Oord, Aaron. Vinyals, Oriol. Kavukcuoglu, Koray. NeurIPS 2017

VQ-VAE

- **Results – More Data Modalities**



Left: original waveform, middle: reconstructed with same speaker-id, right: reconstructed with different speaker-id. The contents of the three waveforms are the same.

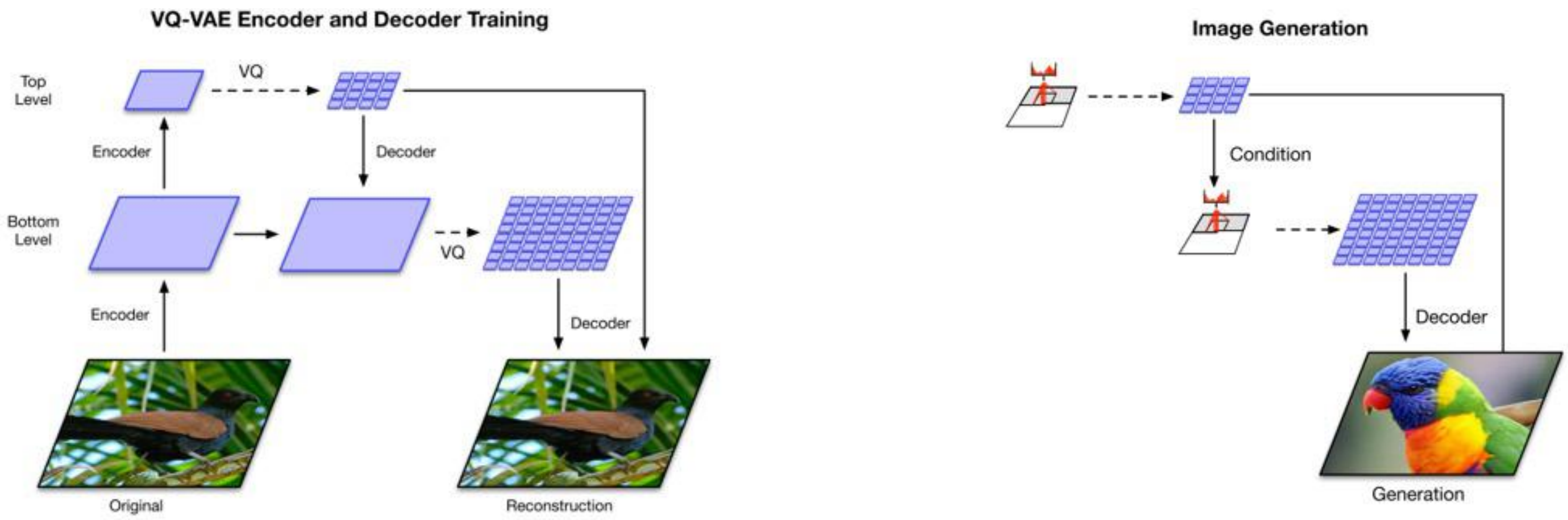


First 6 frames are provided to the model, following frames are generated conditioned on an action. Top: repeated action "move forward", bottom: repeated action "move right".

VQ-VAE: Neural discrete representation learning. Van Den Oord, Aaron. Vinyals, Oriol. Kavukcuoglu, Koray. NeurIPS 2017

VQ-VAE-2

- VQ VAE-2: Scale single-level VQ-VAE to hierarchical VQ-VAE
- Intuition: Different level's features encode different level's information



Generating Diverse High-Fidelity Images with VQ-VAE-2. Razavi, Ali Oord, Aaron van den Vinyals, Oriol. arXiv 2019.

VQ-VAE-2

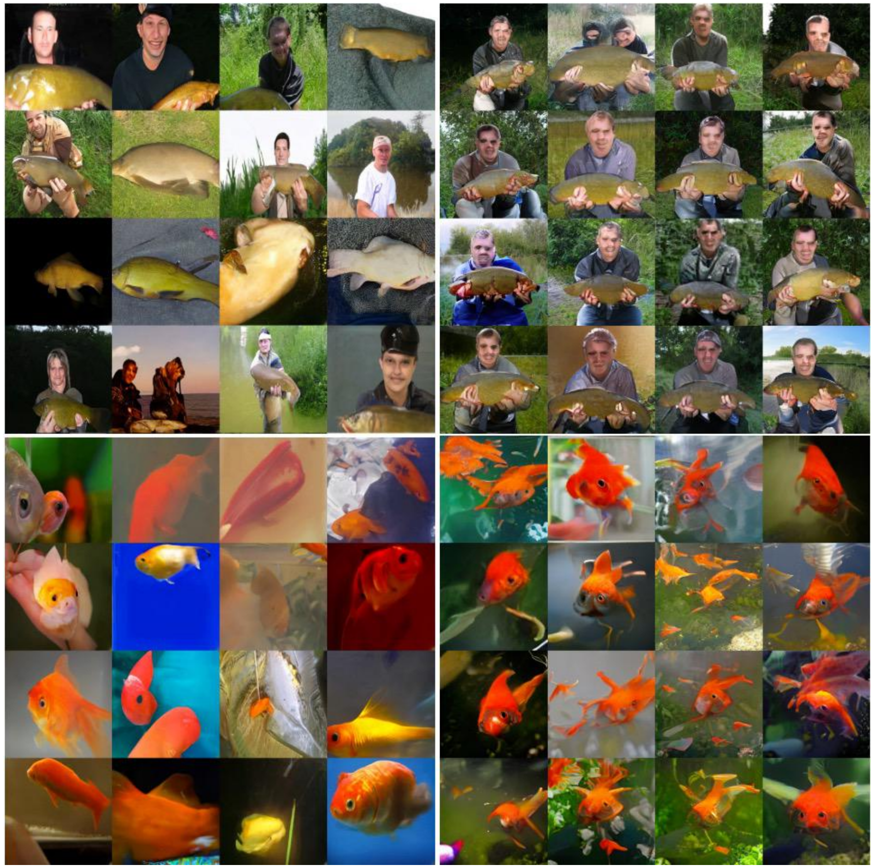
- An intuitive interpretation of different level's information

h_{top} | $h_{\text{top}}, h_{\text{middle}}$ | $h_{\text{top}}, h_{\text{middle}}, h_{\text{bottom}}$ | Original

Low-level features == details

Generating Diverse High-Fidelity Images with VQ-VAE-2. Razavi, Ali Oord, Aaron van den Vinyals, Oriol. arXiv 2019.

VQ-VAE-2



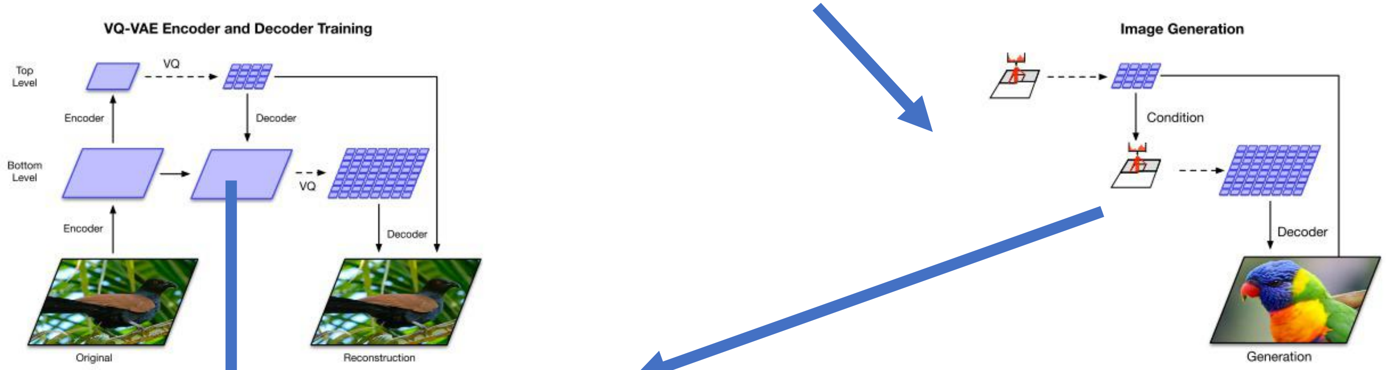
VQ-VAE-2 BigGAN

(more diverse)

Generating Diverse High-Fidelity Images with VQ-VAE-2. Razavi, Ali Oord, Aaron van den Vinyals, Oriol. arXiv 2019.

VQ-VAE-2 Limitation

- The latent representation is NOT a prior distribution, an additional deep model is required to model the latent distribution for sampling, it is not a “real encoding”



- For VQ-VAE-2, the hierarchical representations are not independent, we cannot change the hierarchical feature individually.

- For both VQ-VAE and VQ-VAE-2, the spatial representations (the features within a same latent map) are not independent, we cannot change the spatial feature individually.

VQ-VAE: Neural discrete representation learning. Van Den Oord, Aaron Vinyals, Oriol Kavukcuoglu, Koray. NIPS 2017.
Generating Diverse High-Fidelity Images with VQ-VAE-2. Razavi, Ali Oord, Aaron van den Vinyals, Oriol. arXiv 2019.

Challenge: High-dimensional data generation

- Challenges:
  - Formulation
  - For CG-based Methods
  - For Deep Methods
- Approaches:
  - Progressive-GAN
  - Style-GAN
  - SAGAN
  - Big-GAN
  - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
  - Ideal Generative Models

Discussion: Ideal Model

PEKING UNIVERSITY

- High-dimensional data generation
- Data encoding, explicit inverse x->z

Next Lecture

- More
  - Interpolation in latent space
  - Multi-modality
  - Mode collapse
  - Fast training
  - Disentanglement
  - Hierarchical representation with independent property
  - Spatial representation with independent property

北京大學
PEKING UNIVERSITY

Thanks