

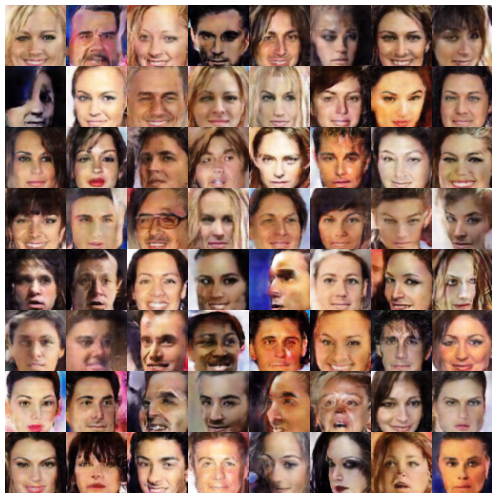
Challenge: High-dimensional Data Generation

Hao Dong

Peking University

Challenge: High-dimensional data generation

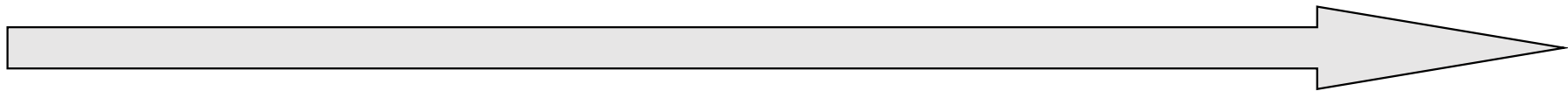
Past
64x64



Now
1K, 2K



Next
Retina Screen



We use images for demonstration

Challenge: High-dimensional data generation

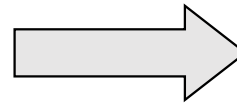
- Challenges:
 - Formulation
 - For CG-based Methods
 - For Deep Methods
- Approaches:
 - Progressive-GAN
 - Style-GAN
 - SAGAN
 - Big-GAN
 - VQ-VAE VQ-VAE-2 and Limitation
- Discussion:
 - Ideal Generative Models

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

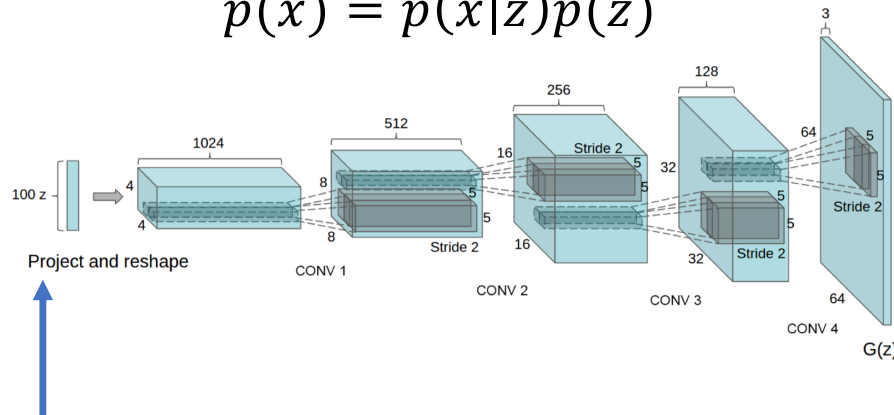
Formulation

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

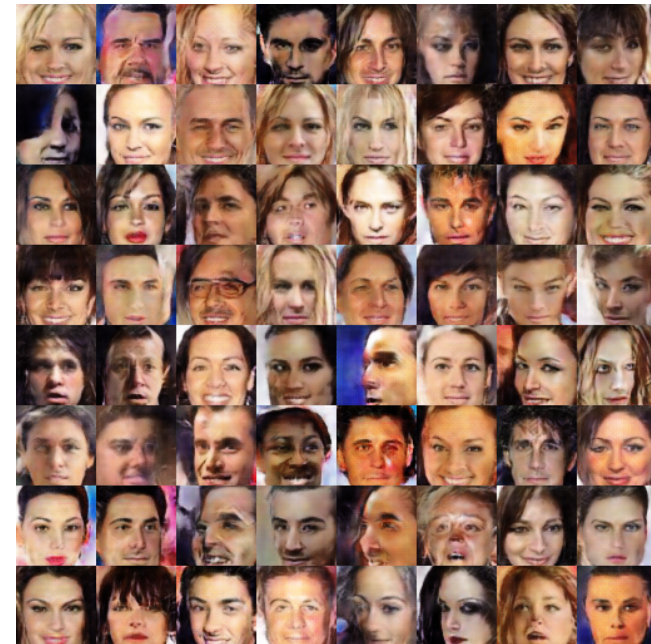


Large Scale (e.g. Resolution)
(e.g., image, video, ...)

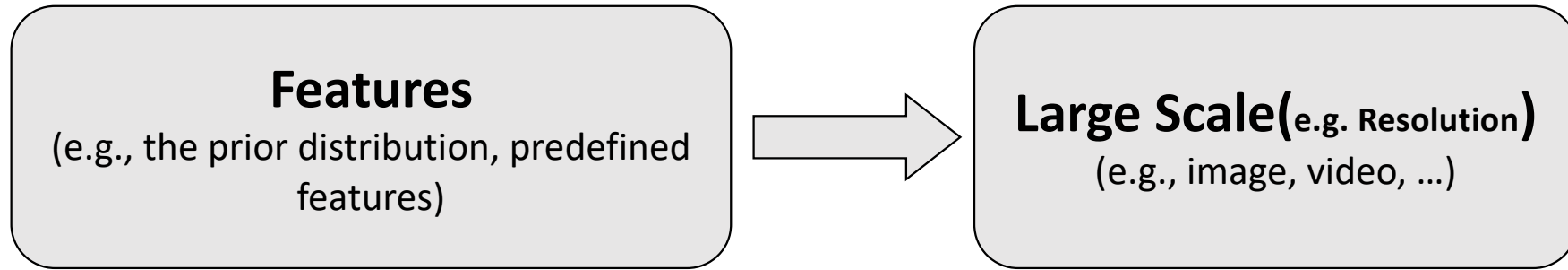
$$p(x) = p(x|z)p(z)$$



(Prior) Normal Distribution
z = 100 values

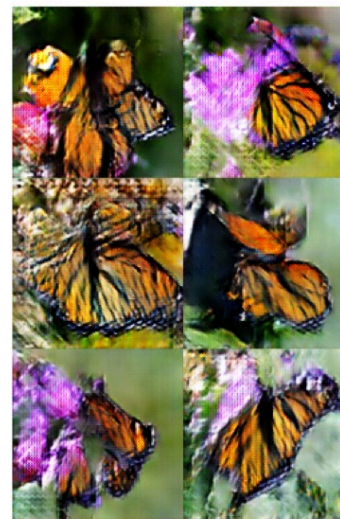
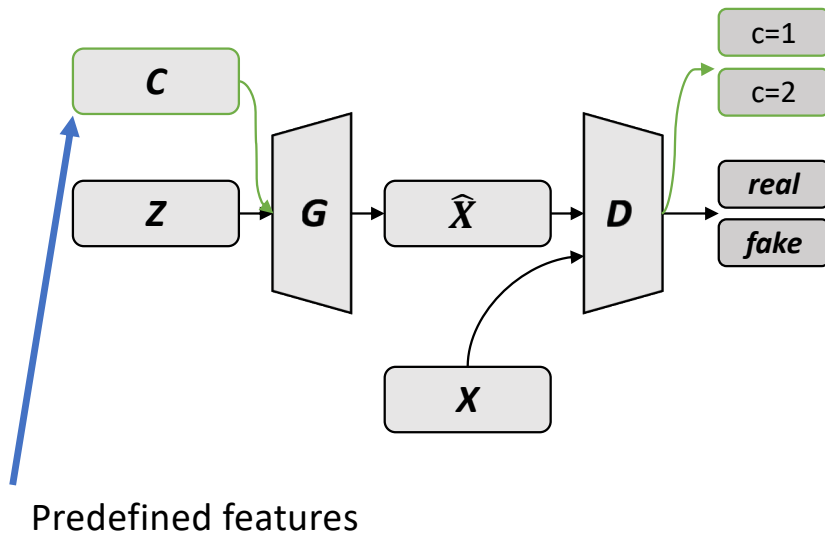


Formulation



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

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



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

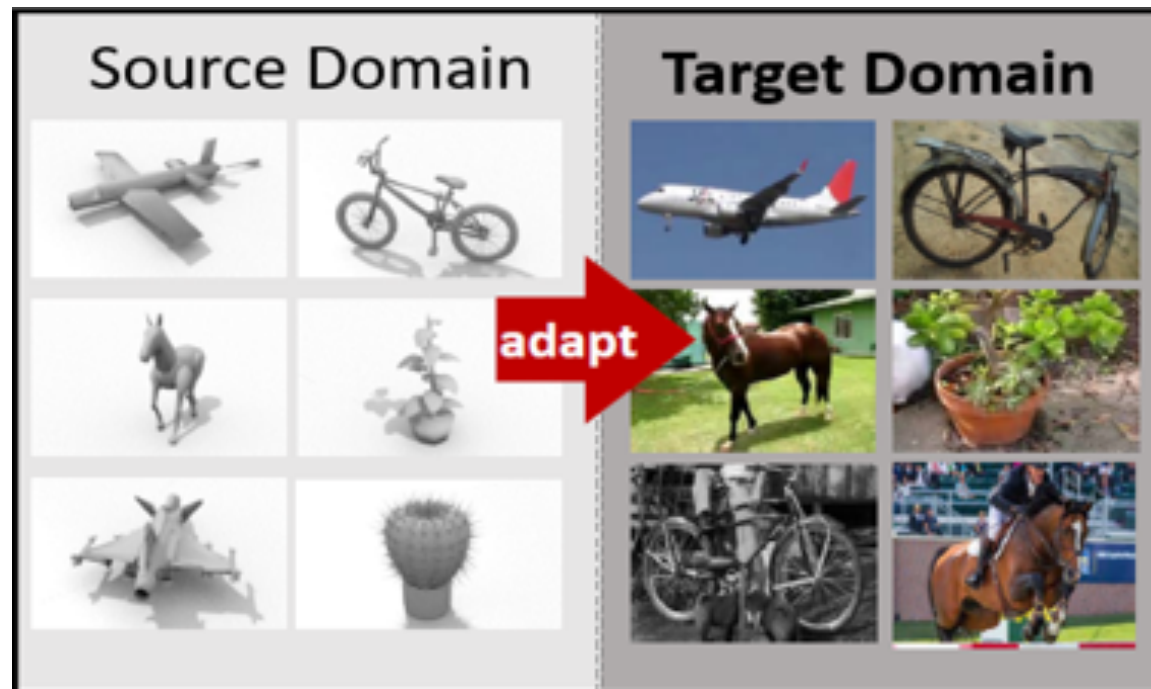


GTA 5

- **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



CG-based Methods

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



Input GTA5 CG

<https://blog.csdn.net/gdymind>



Output image with German street view style

blog.csdn.net/gdymind

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

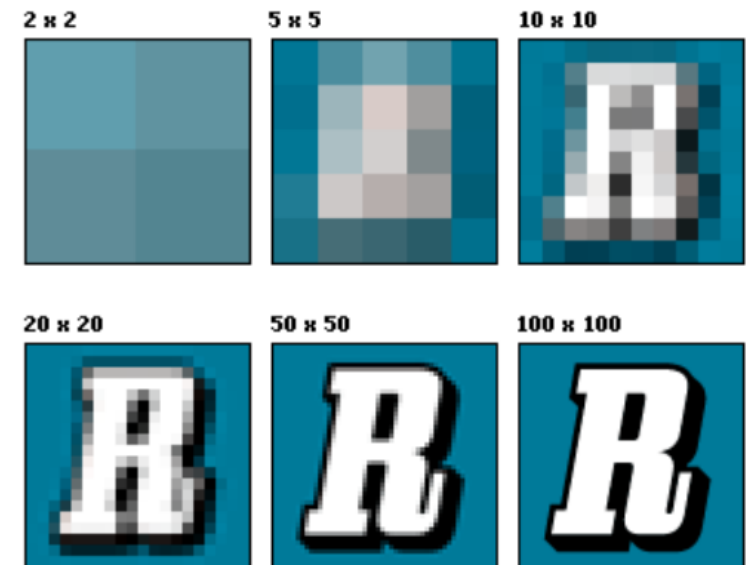
- Fully CG-based
- Hybrids
- Limitation
 - Need prior knowledge
 - Intensive Engineering...
 - Limited Generalization
 - Artificially designed generation rule can only capture limited latent structure of domain
 - Improvement need more prior ...
 - Anyway, automatically learning prior knowledge is necessary.

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 modeling 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**

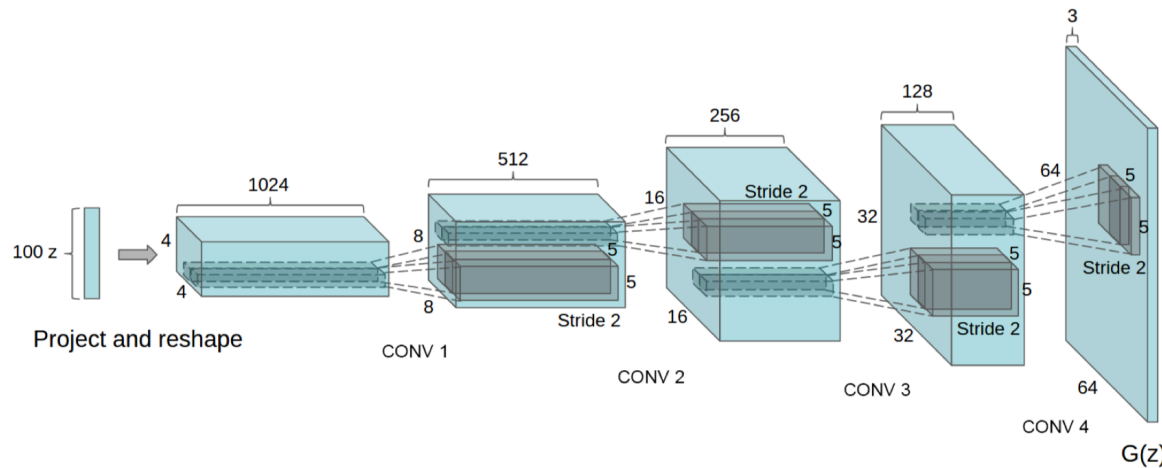


Challenge: High-dimensional data generation

- 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:

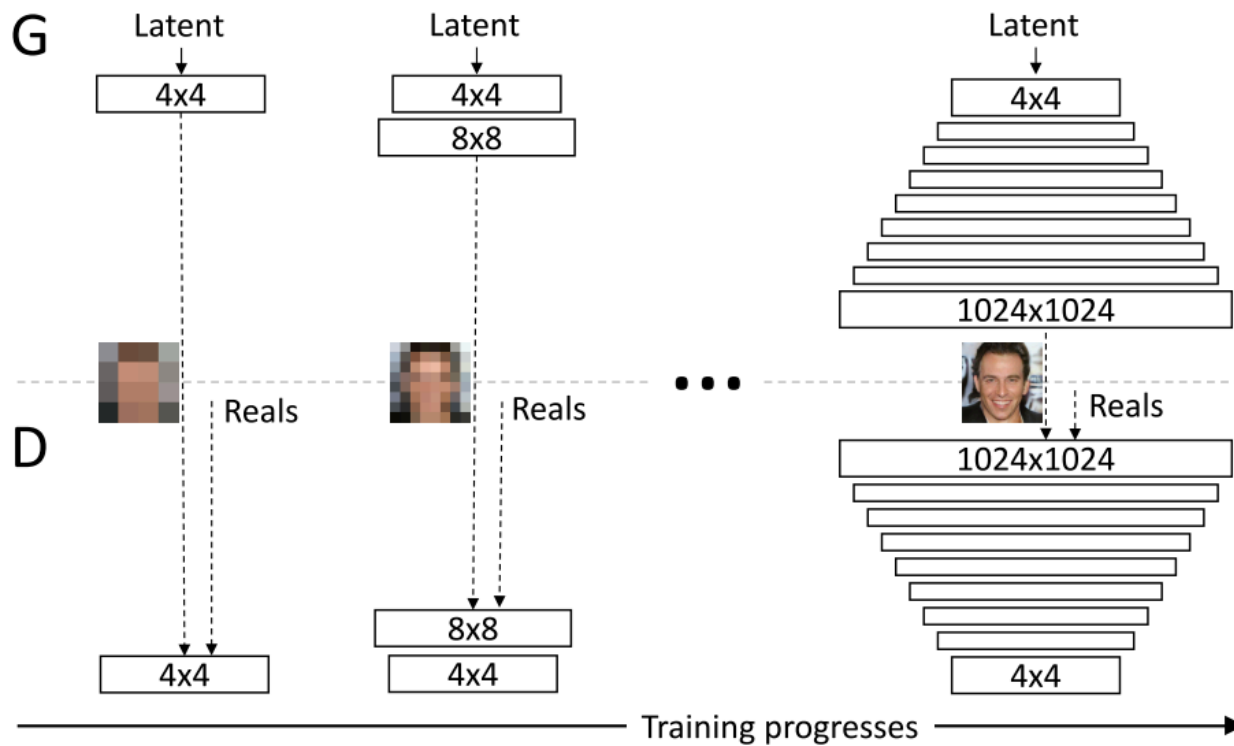
- Unstable training
- Computer memory constraints
- High resolution images make the discriminator easier to discriminate the fake and real images, amplifying the gradient problem.

64x64 work !

1024x1024 fail

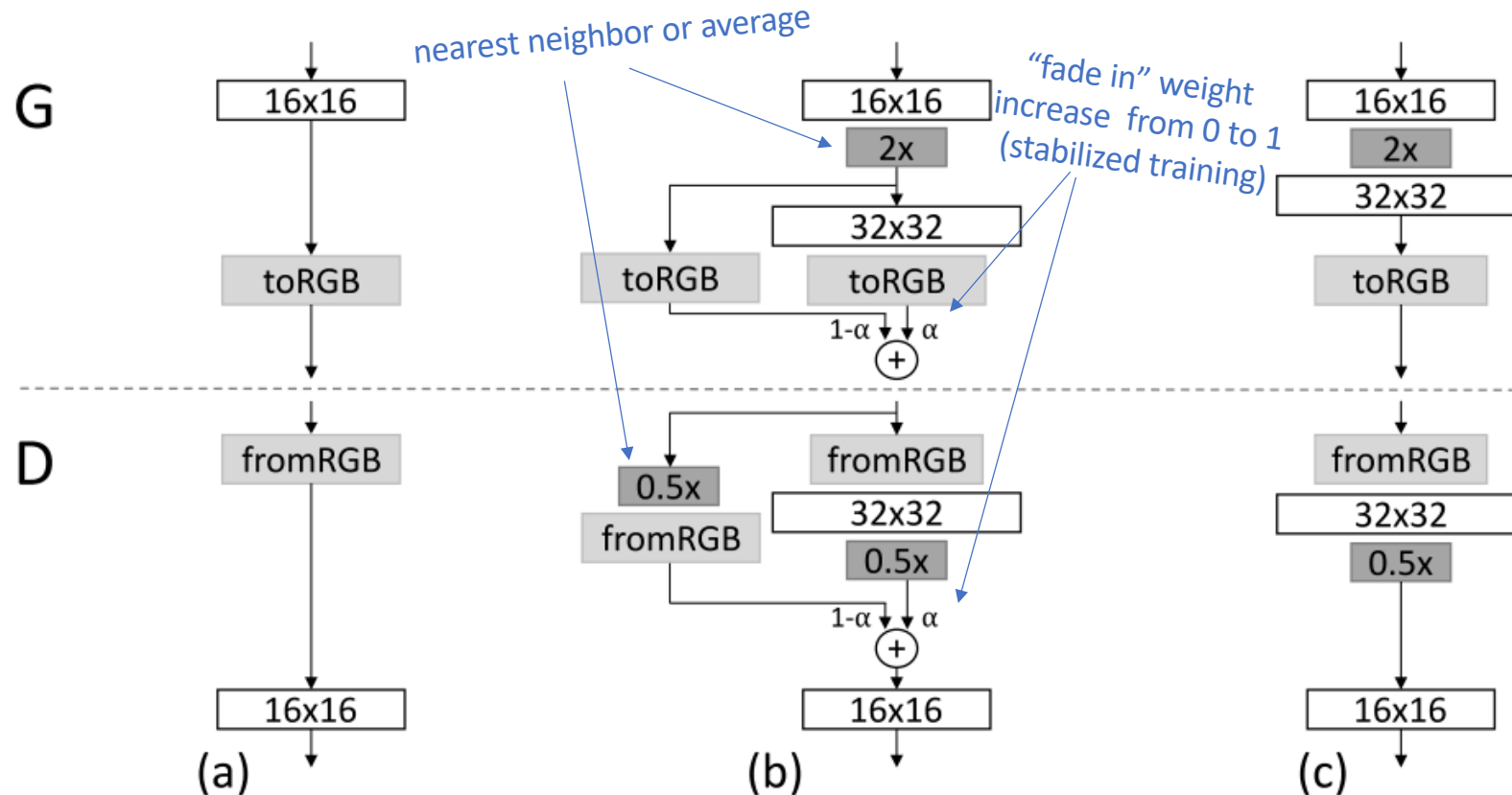
Progressive GAN

- From Coarse to Fine : $4 \times 4 \rightarrow 8 \times 8 \rightarrow 16 \times 16 \rightarrow 32 \times 32 \dots \rightarrow 1024 \times 1024$



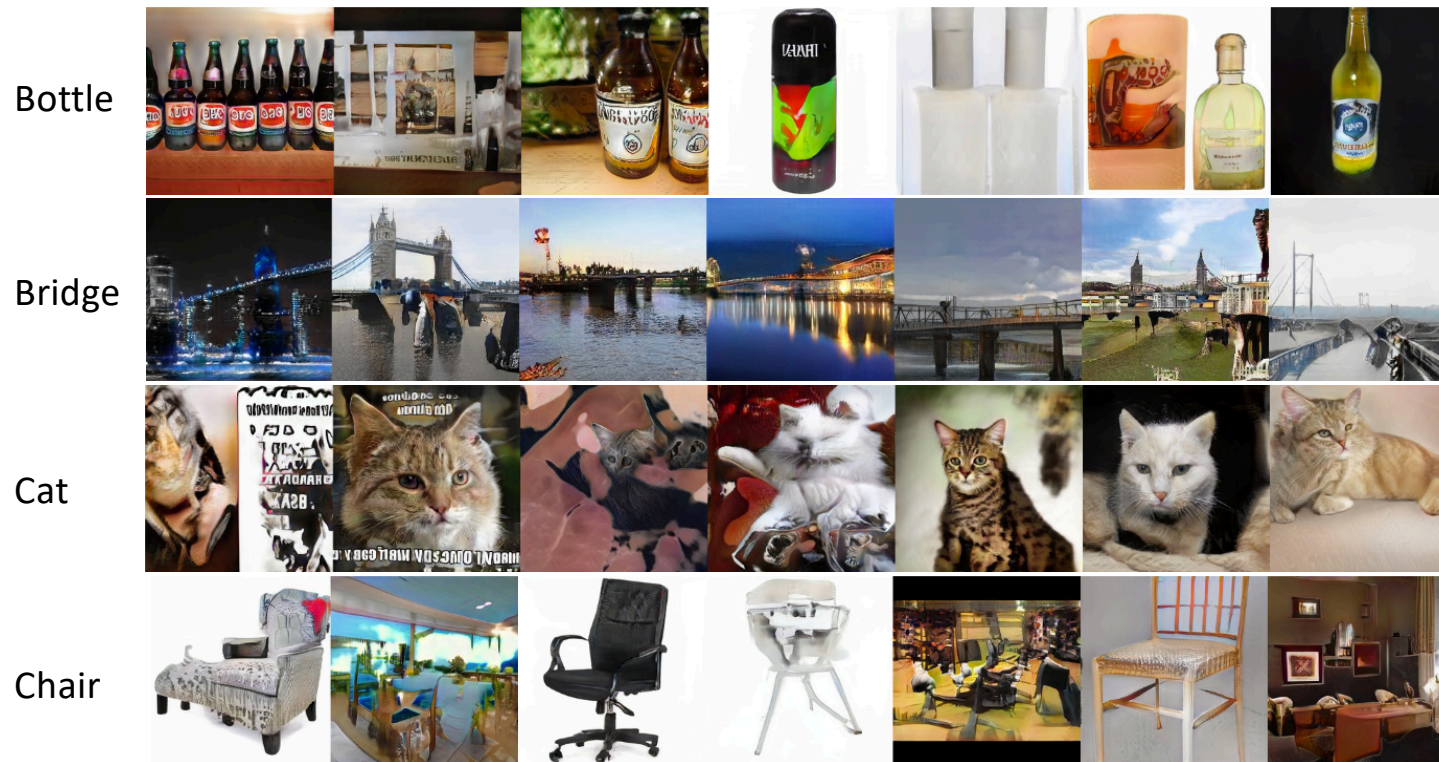
Progressive GAN

- From Coarse to Fine : $4 \times 4 \rightarrow 8 \times 8 \rightarrow 16 \times 16 \rightarrow 32 \times 32 \dots \rightarrow 1024 \times 1024$




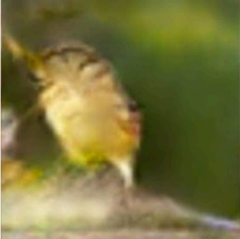

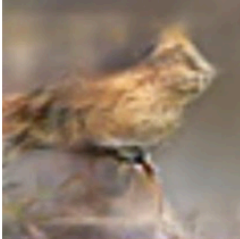
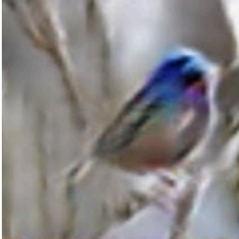
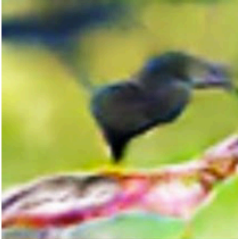
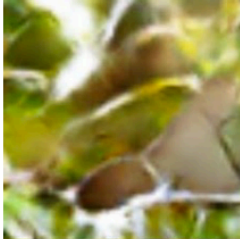


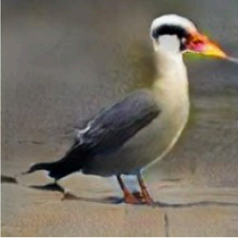
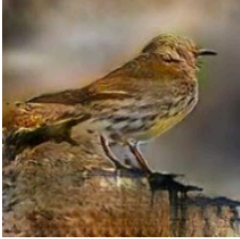

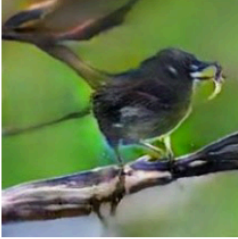
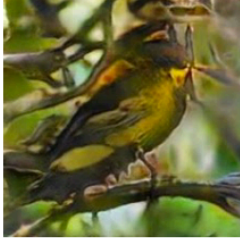
Progressive GAN

- From Coarse to Fine with Condition



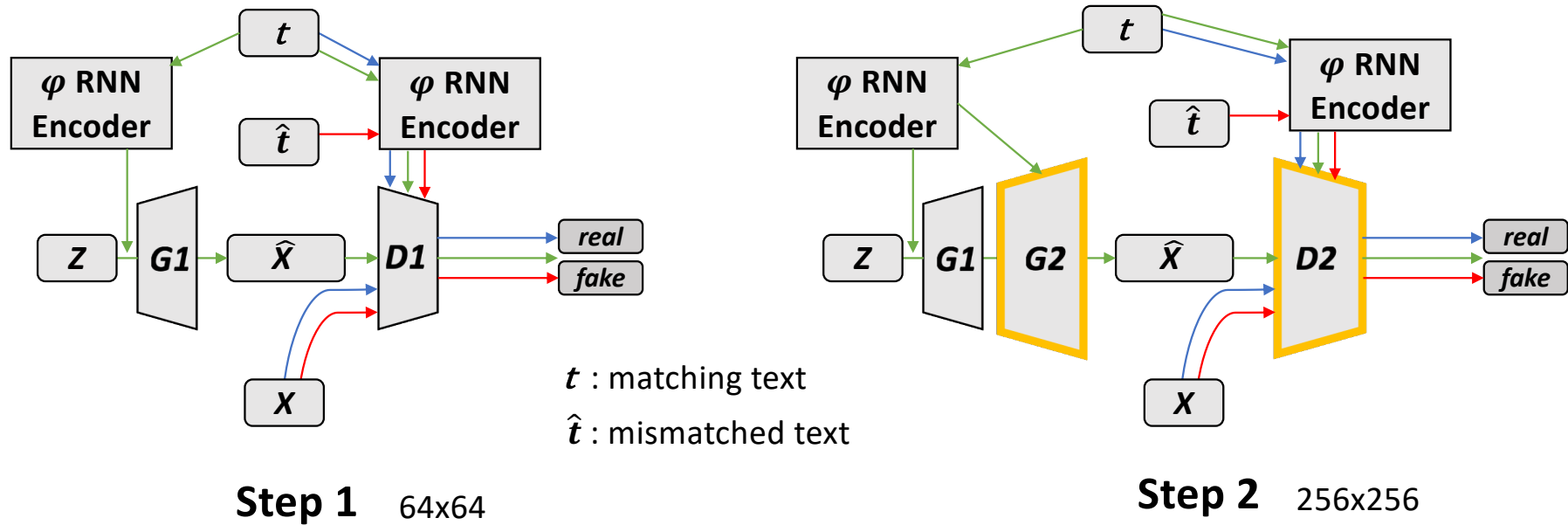
Progressive GAN → StackGAN

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

Text description	This bird is blue with white and has a very short beak	This bird has wings that are brown and has a yellow belly	A white bird with a black crown and yellow beak	This bird is white, black, and brown in color, with a brown beak	The bird has small beak, with reddish brown crown and gray belly	This is a small, black bird with a white breast and white on the wingbars.	This bird is white black and yellow in color, with a short black beak
Stage-I images							
Stage-II images							

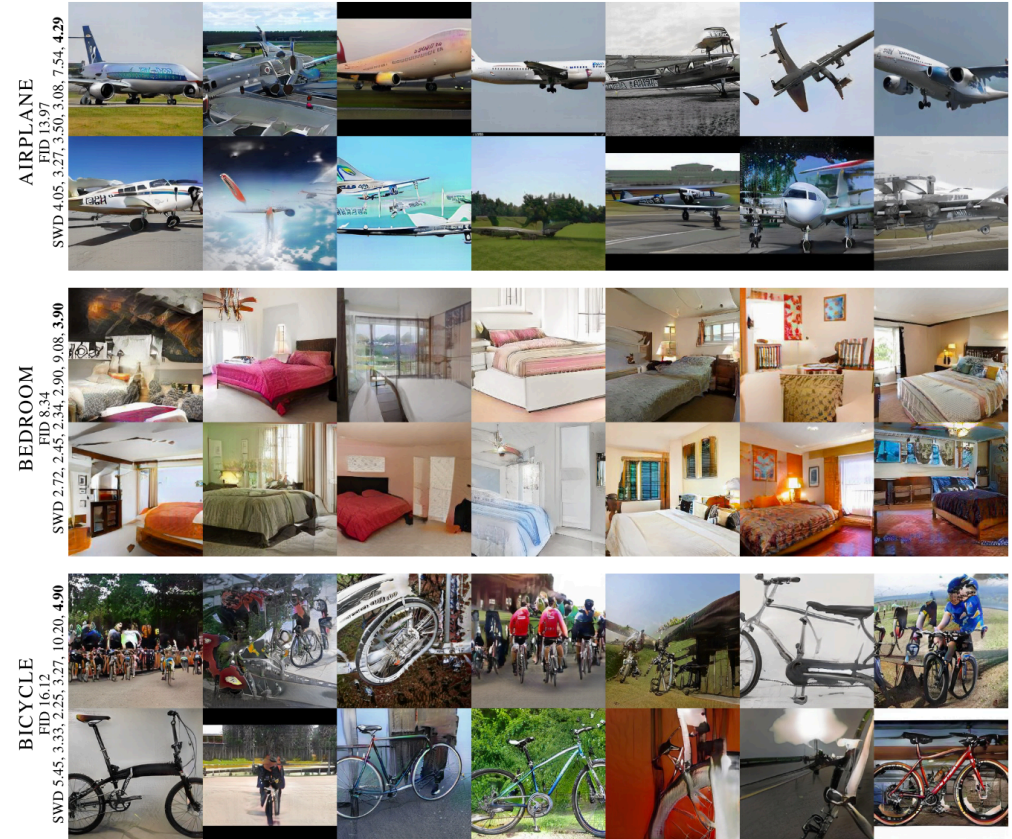
Progressive GAN \rightarrow StackGAN

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



Progressive GAN

- Question: Can Computer Graphic Generates This?



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

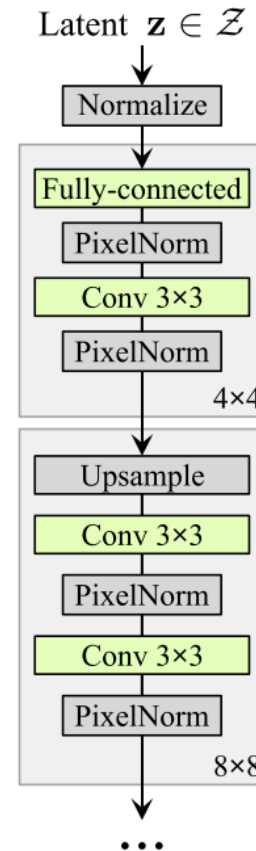
StyleGAN

- Insert Feature as Style Transfer

Adaptive Normalization:

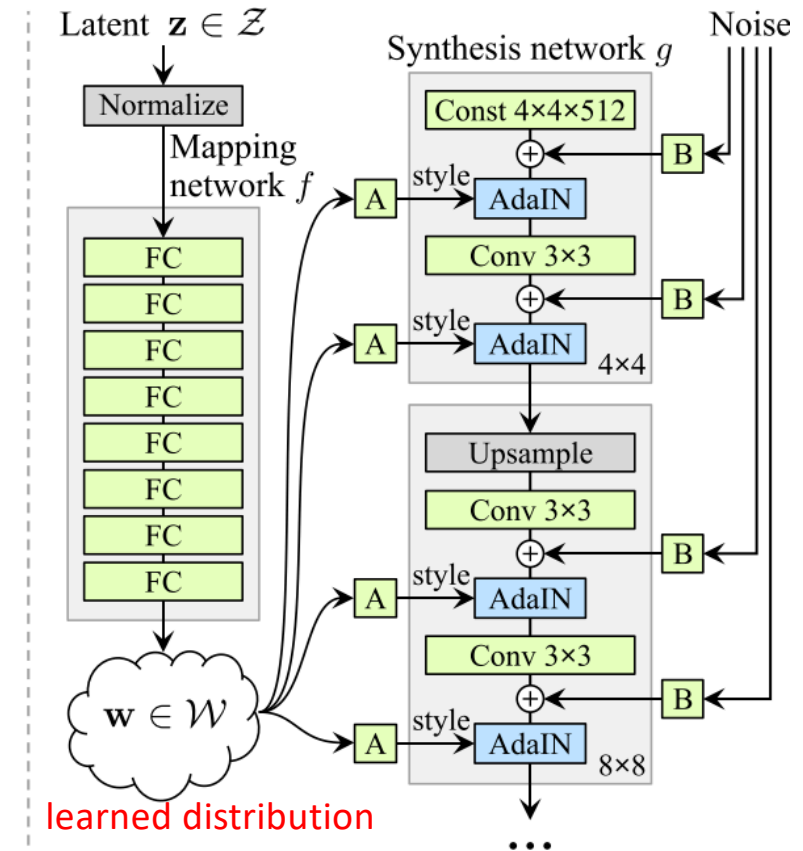
$$\text{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i}$$

where each feature map \mathbf{x}_i is normalized separately, and then scaled and biased using the corresponding scalar components from style \mathbf{y} .



(a) Traditional

prior distribution

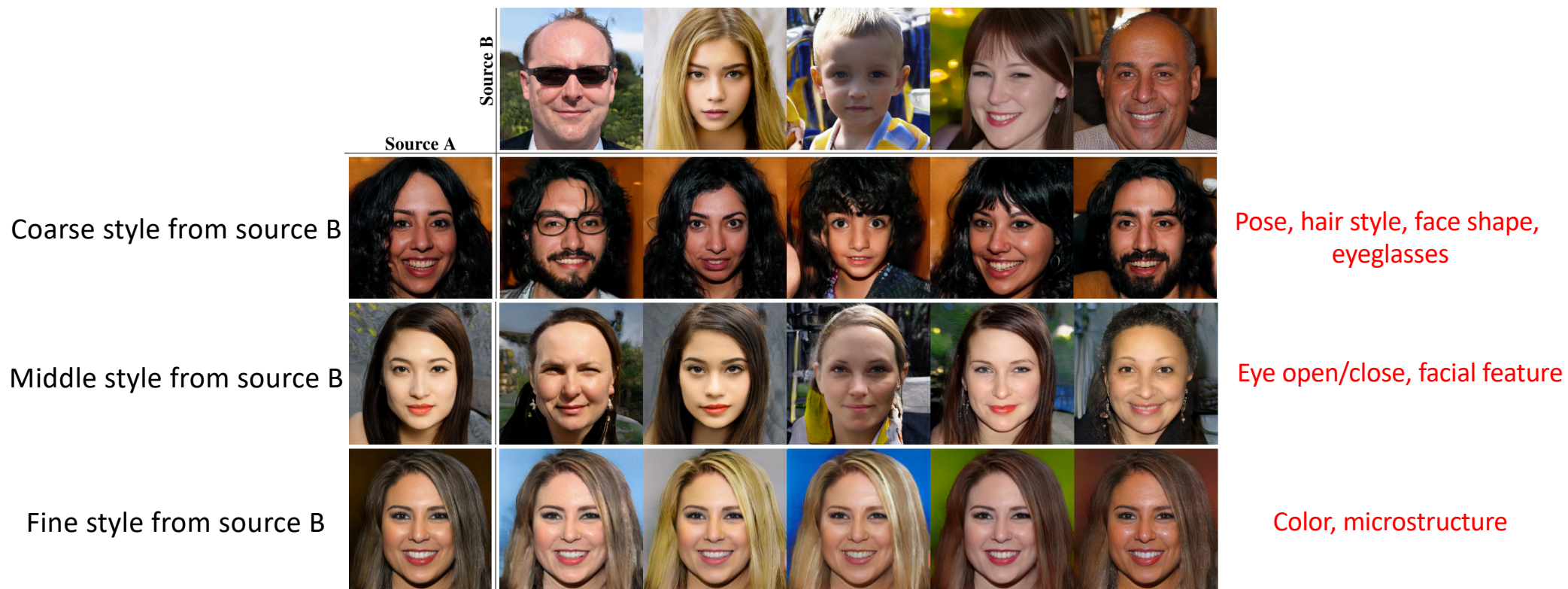


learned distribution

(b) Style-based generator

StyleGAN

- Hierarchical Latent Code



StyleGAN

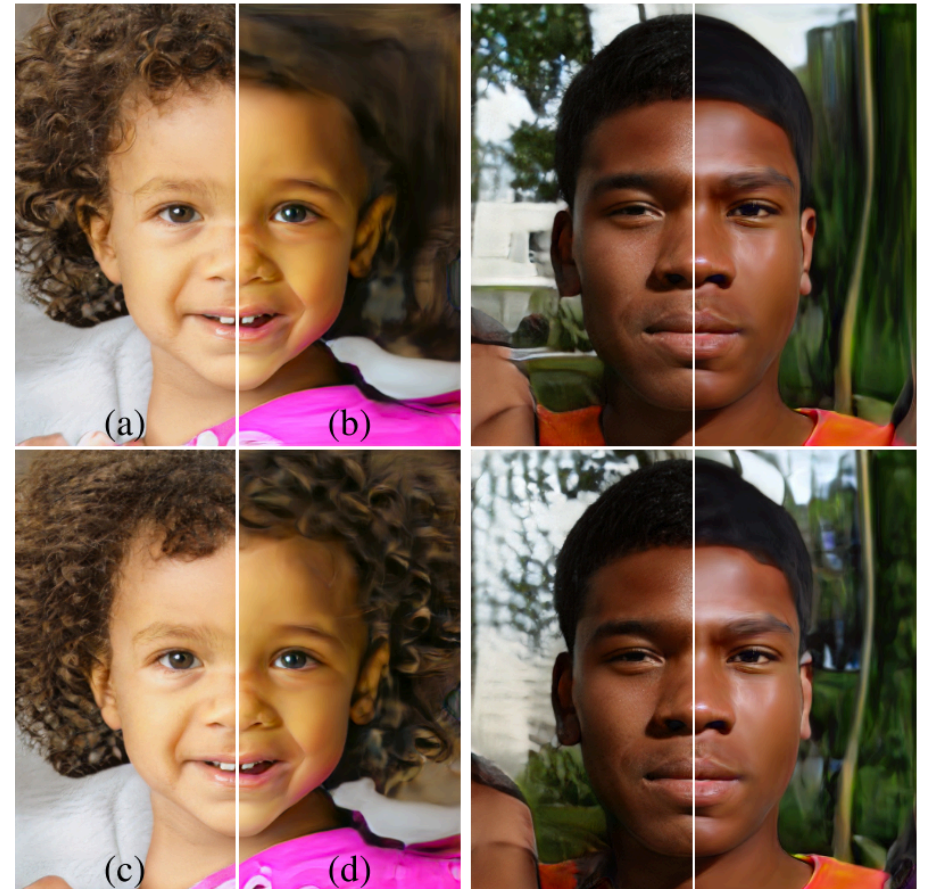
- Hierarchical Noise

(a) Noise is applied to all layers.

(b) No noise. look “smooth”

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

(d) Noise in coarse layers only ($4^2 - 32^2$). coarse details

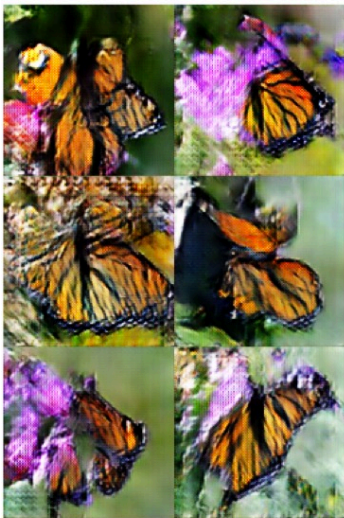


Challenge: High-dimensional data generation

- 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

- Recap: Shape Deformation When Directly Scaling Up DCGAN
 - And recall that deep model's challenges lie on
 - Intensively modeling details
 - Long range dependency
- CNN is a strong inductive bias to model natural details, but fails when modeling long range dependency.



monarch butterfly



goldfinch



daisy



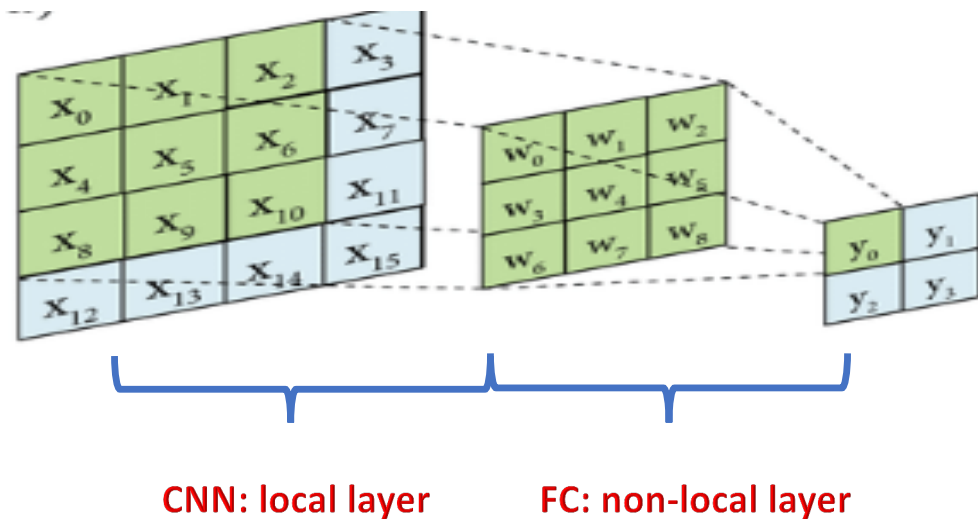
redshank



grey whale

SAGAN

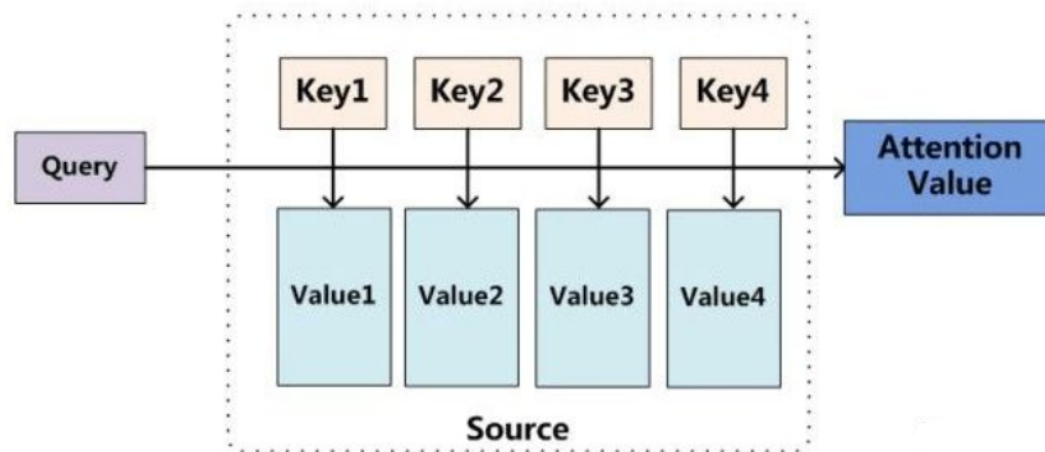
- Non-local layer vs Local layer
 - CNN is “local layer”, a neuron only observes part elements of the previous layer.



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

SAGAN

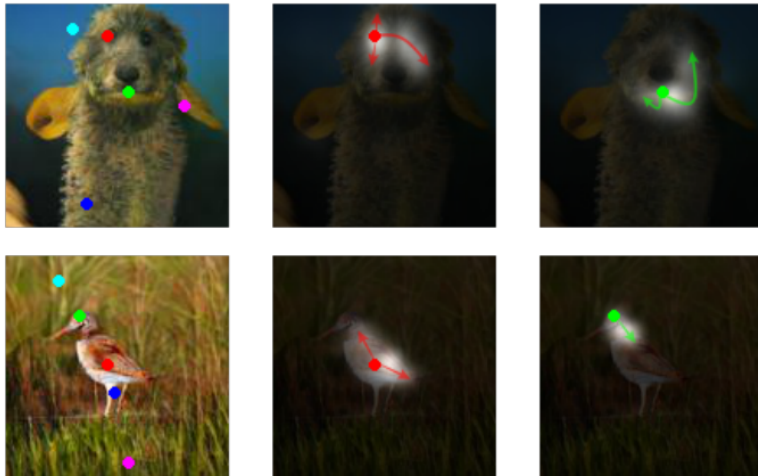
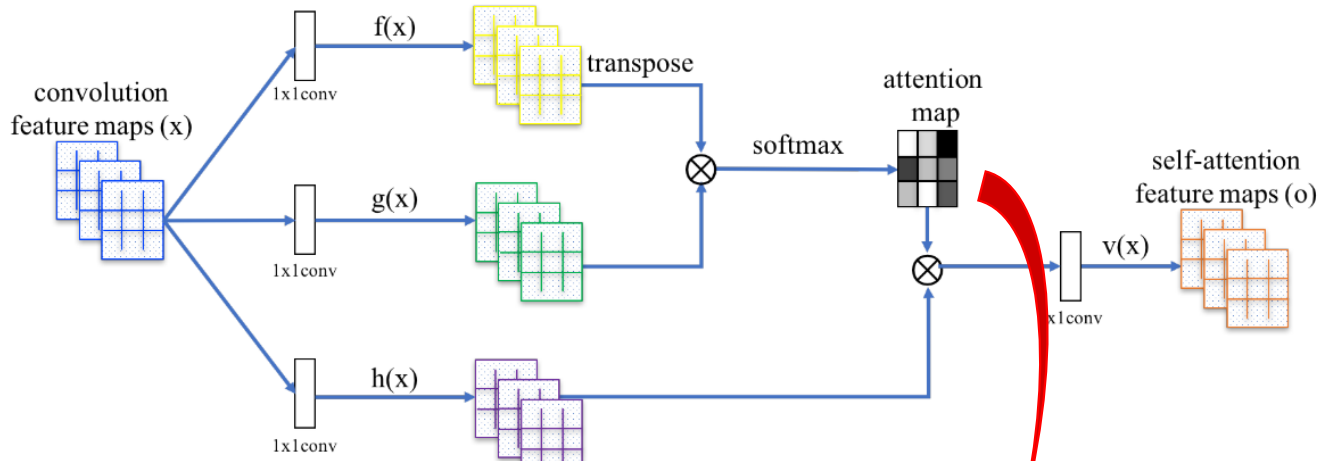
- SAGAN: Introduce attention layer into DCGAN backbone
 - Attention: have become an integral part of models that must **capture global dependencies**
 - Illustration of attention:



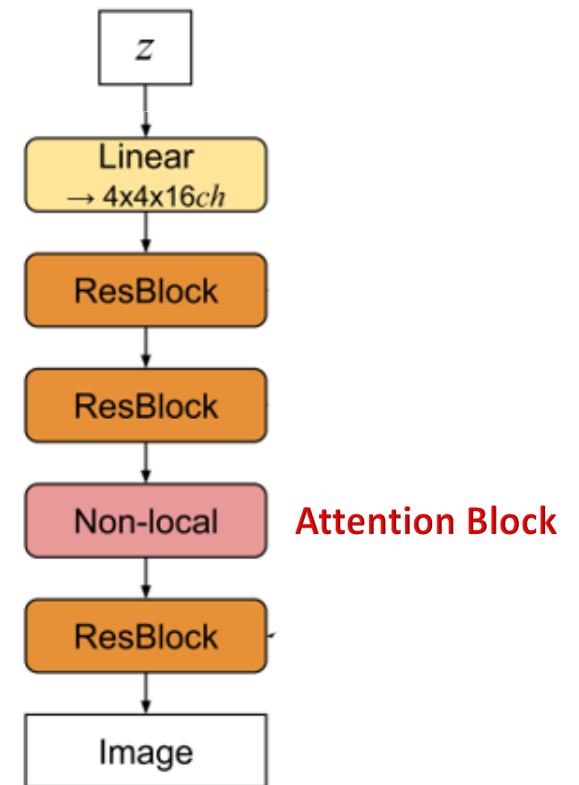
- $Attention\ Value = \sum_{i=1}^4 value_i * coefficient_i$ **Convex combination of value w.r.t. coefficients**
- $coefficient_i = \frac{e^{-Key_i \circ Query}}{\sum_{j=1}^4 e^{-Key_j \circ Query}}$ **Correlation coefficient**

SAGAN

- SAGAN: Introduce attention layer into DCGAN backbone



Visualization of attention Coefficients (attention map)



SAGAN

- SAGAN: Introduce attention layer into DCGAN backbone



DCGAN



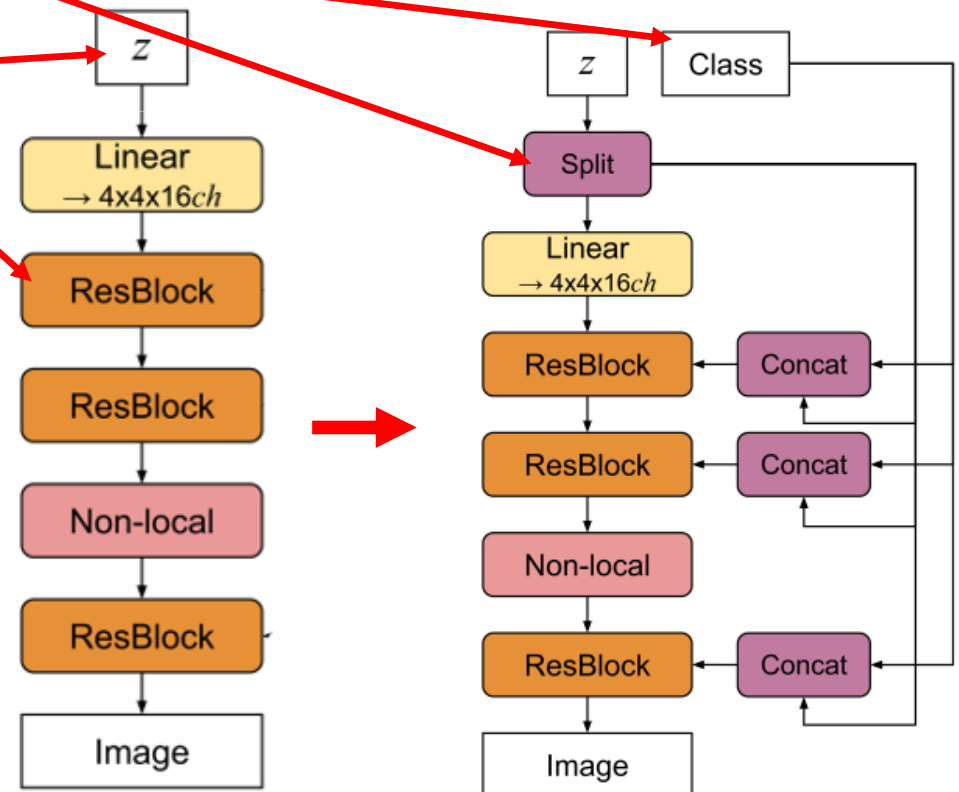
SAGAN

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

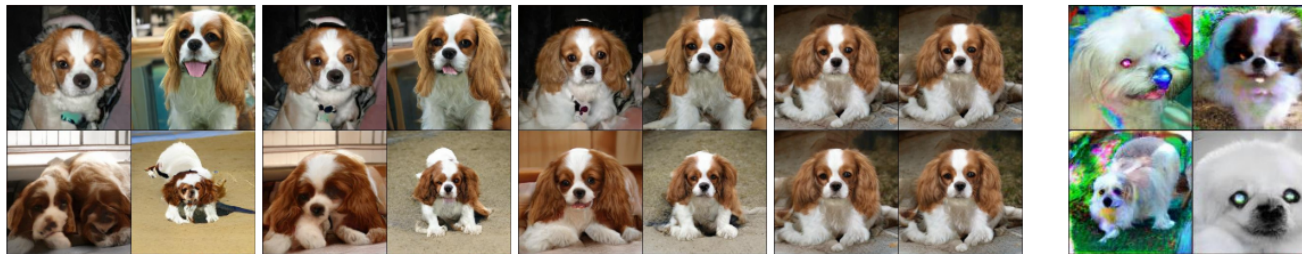
Big-GAN

- 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 regularization
 - Enforce Generator to be more amenable to truncation
 - Orthogonal regularization can make G smoother

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

Big-GAN

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

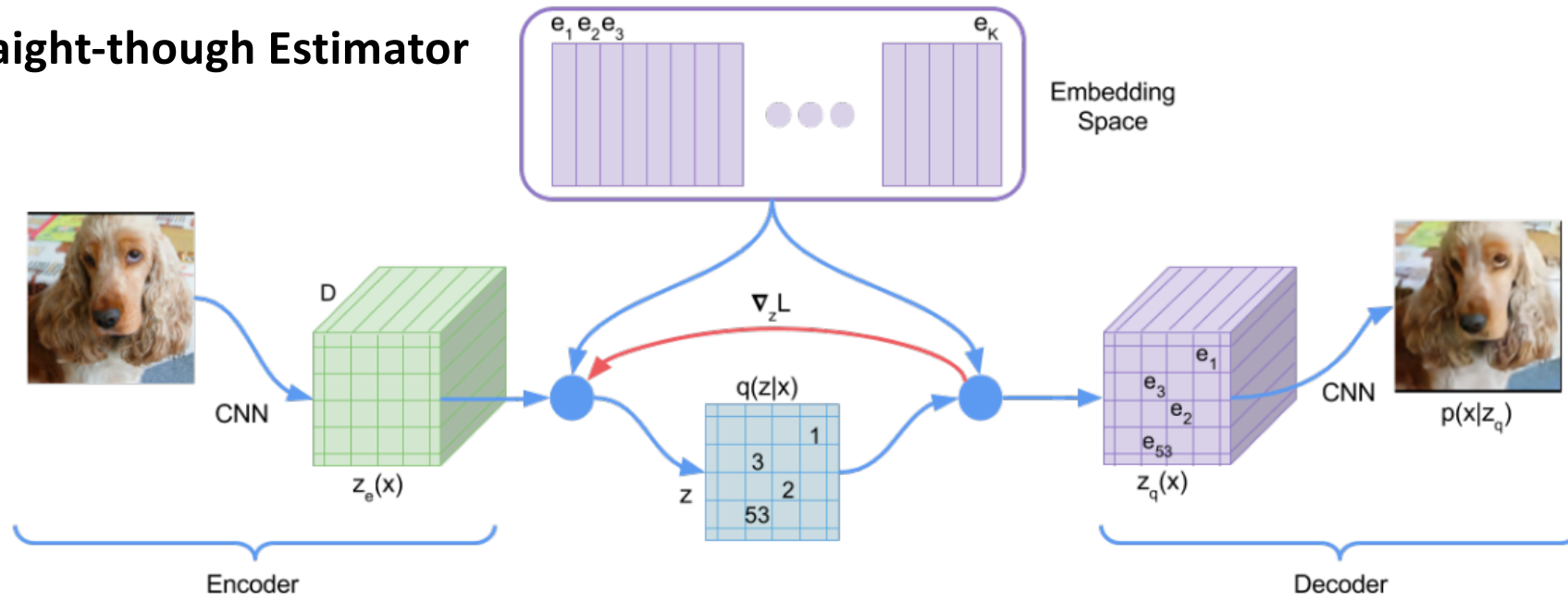


Challenge: High-dimensional data generation

- 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

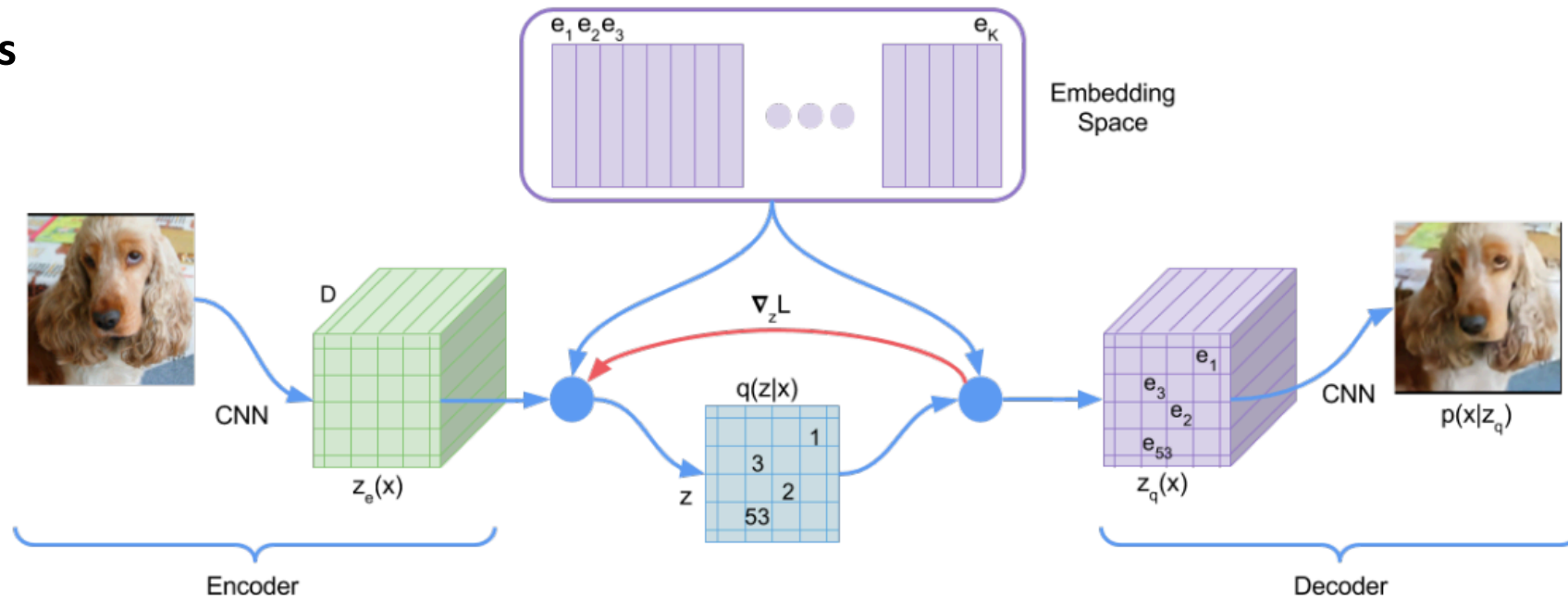
VQ-VAE

- Straight-through Estimator



VQ-VAE

- Loss



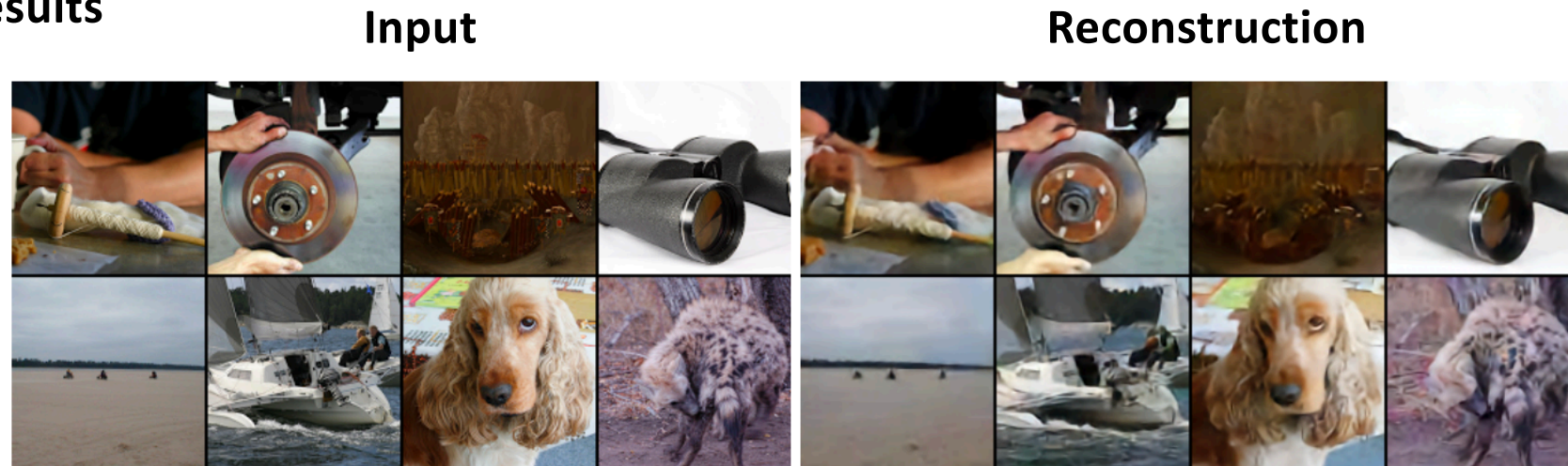
$$L = \underbrace{\log p(x|z_q(x))}_{\text{Image reconstruction loss}} + \underbrace{\| \text{sg}[z_e(x)] - e \|_2^2 + \beta \| z_e(x) - \text{sg}[e] \|_2^2}_{\text{Make the quantized vector as close as the original vector}}$$

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

Image reconstruction loss **Make the quantized vector as close as the original vector**

VQ-VAE

- Results

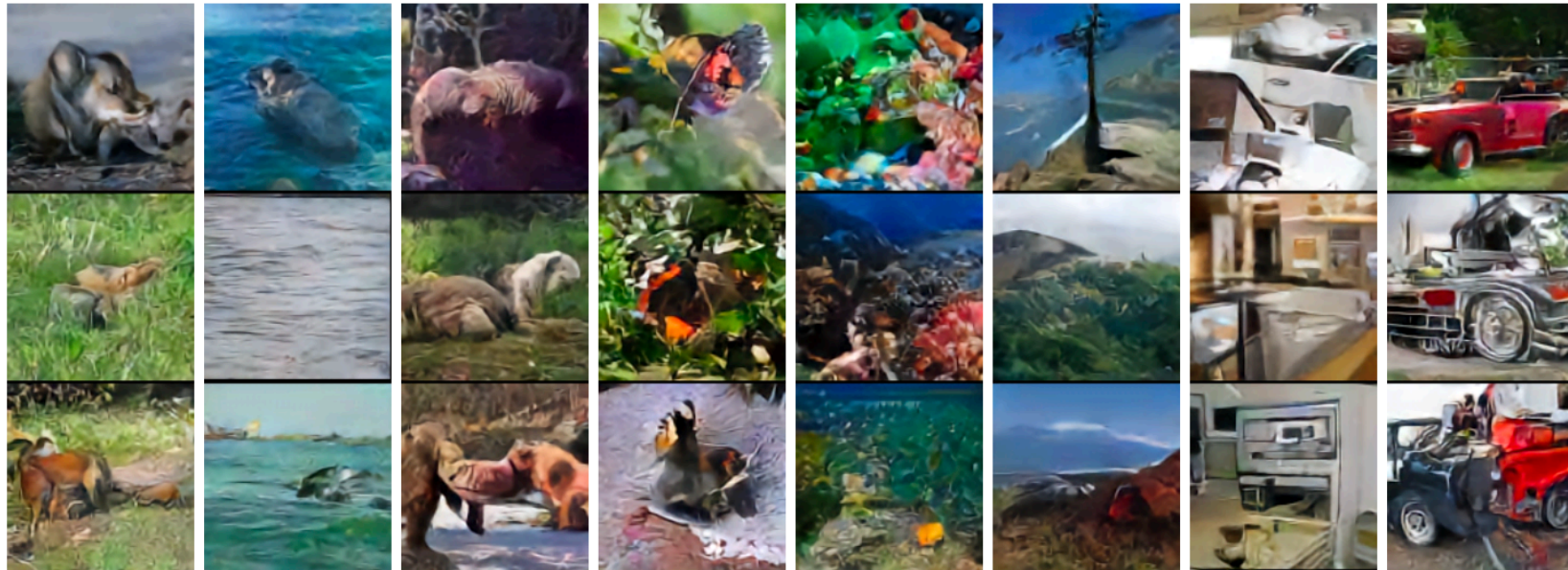


Images contain a lot of redundant information as most of the pixels are correlated and noisy, therefore learning models at the pixel level could be wasteful.

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 $\frac{128 \times 128 \times 3 \times 8}{32 \times 32 \times 9} \approx 42.6$ 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

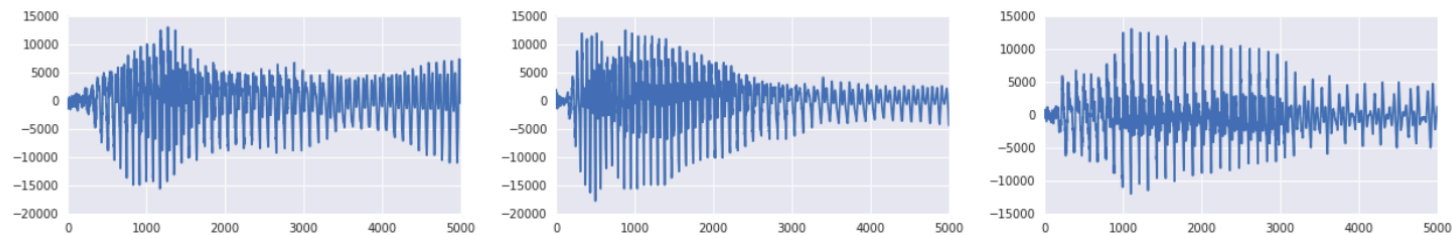
- Results – Random Sampling



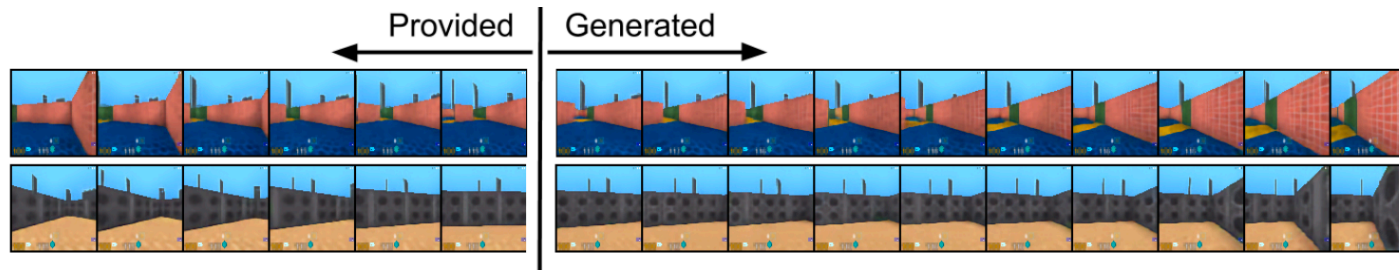
Samples (128x128) from a VQ-VAE with a PixelCNN prior trained on ImageNet images. From left to right: kit fox, gray whale, brown bear, admiral (butterfly), coral reef, alp, microwave, pickup.

VQ-VAE

- 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-2

- An intuitive interpretation of different level's information



VQ-VAE-2



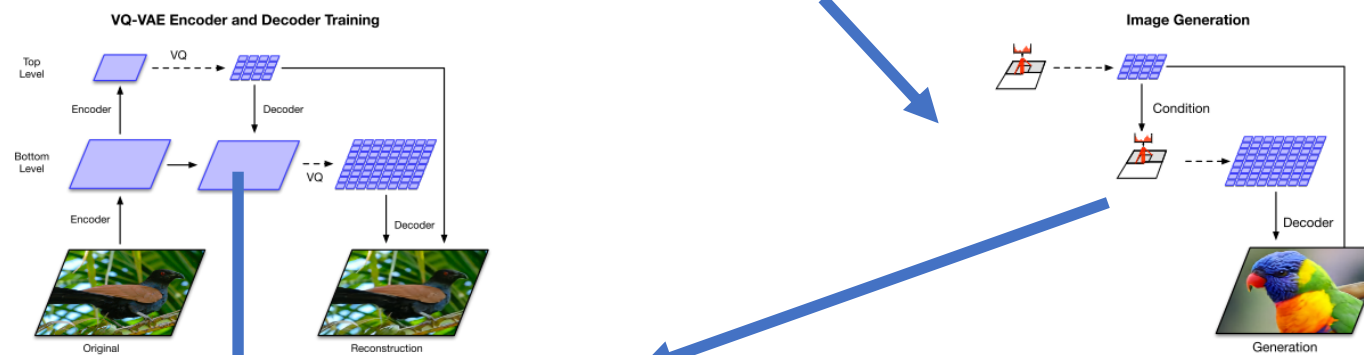
VQ-VAE-2
(more diverse)

BigGAN

Generating Diverse High-Fidelity Images with VQ-VAE-2. Razavi, Ali Oord, Aaron van den Vynals, 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”

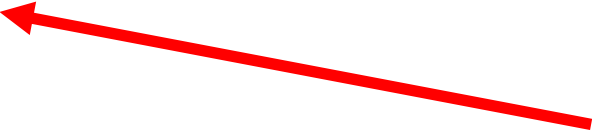


- For VQ-VAE-2, the hierarchical representations are not independent, we cannot change the hierarchical feature individually.
- For both VQ-VAE and VQ-VAE-2, the spatial representations (the features within a same latent map) are not independent, we cannot change the spatial feature individually.

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

- High-dimensional data generation
 - Data encoding, implicit inverse $x \rightarrow z$
 - More
 - Interpolation in latent space
 - Multi-modality
 - Mode collapse
 - Fast training
 - Disentanglement
 - Hierarchical representation with independent property
 - Spatial representation with independent property
- Next Lecture**
- 

Thanks