Aplikasi Visi Komputer

Hari 2 – Pesantren Sains Data 2023

12 April 2023/ 22 Ramadhan 1444

Arrie Kurniawardhani

arrie.kurniawardhani@uii.ac.id

Chat GPT

Can GPT get questions from image?

GPT-4 Multimodal

- OpenAI’s recent update does include new image processing capabilities.
- The latest version of Chat GPT, powered by the new multimodal model (GPT-4) can now understand and respond to images you feed it.
- https://youtu.be/KzSpKf4l6k4?t=58

GPT-4 Multimodal

Input

What can I make with these ingredients?

Output

There are many options for what you can make with these ingredients! Some possibilities include:

- Pancakes or waffles
- Crepes
- French toast
- Omelette or frittata
- Quiche
- Custard or pudding
- Cake or cupcakes
- Muffins or bread- Cookies or biscuits

These are just a few examples, but the possibilities are endless!

GPT-4 Multimodal

OpenAI

Visual inputs: VGA charger
Sample 1 of 7

User: What is funny about this image? Describe it panel by panel.

Source: hmmm (Reddit)

Visual inputs: VGA charger
Sample 1 of 7

Next sample

GPT-4

The image shows a package for a "Lightning Cable" adapter with three panels.

Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

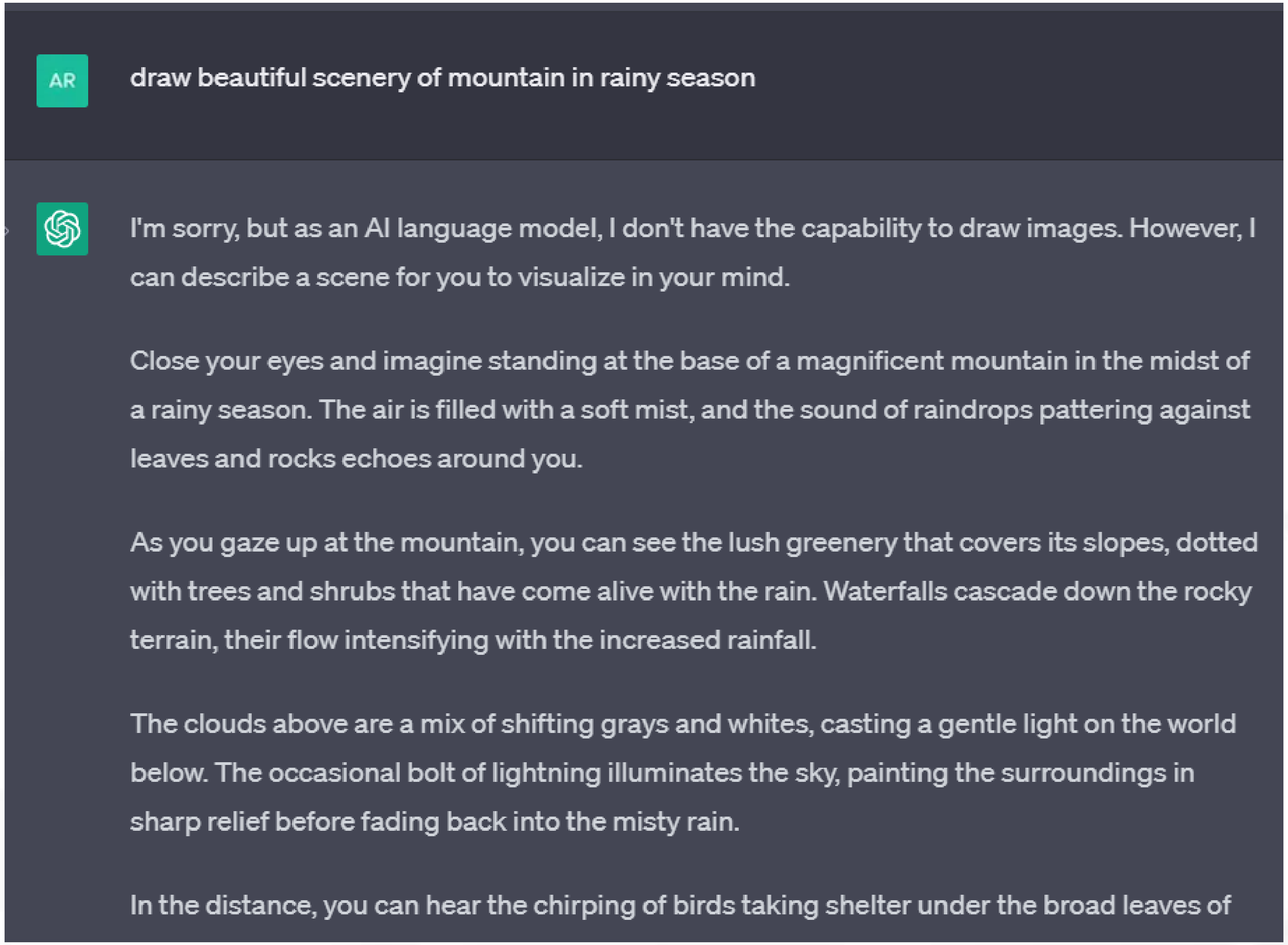
Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

Chat GPT

Can Chat GPT draw image ?



Chat GPT

- Chat GPT (Generative Pre-training Transformer) does not have the ability to generate images or draw pictures.
- The model is a conversational bot built to output text responses and not pixel images.

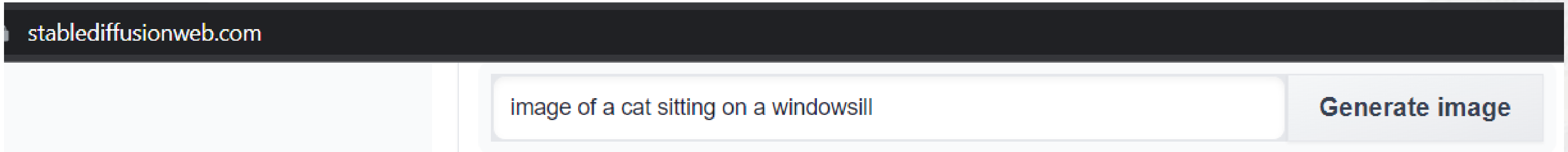
Chat GPT
Text description for generate image

» Chat GPT can create a text description which could then be used for image generation using text-to-image synthesis.

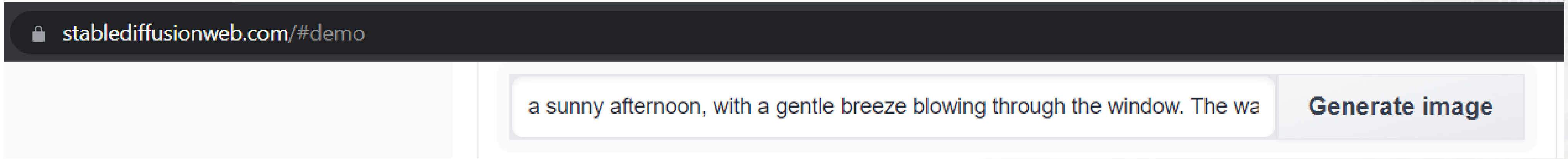
» Text-to-image synthesis:
1. Training a model to create images based on language descriptions.
2. The model would receive a text input and use it to map out visual content and generate image pixels based on the themes of the description

» Stable Diffusion https://stablediffusionweb.com/

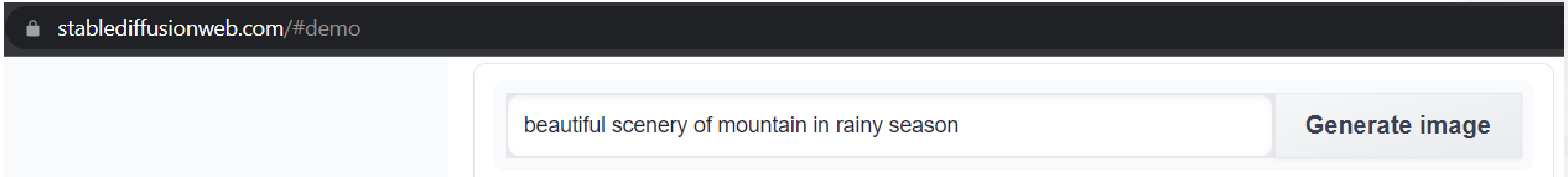
Stable Diffusion



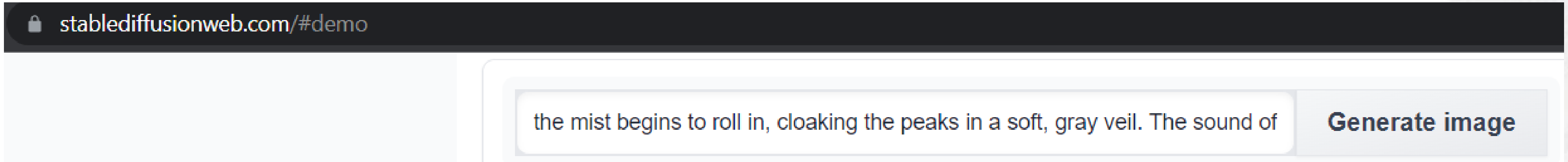
Stable Diffusion



Stable Diffusion



Stable Diffusion



Limitation GPT and Text-to-Image Synthesis

- The quality of the generated images.
  - The accuracy and detail of the generated images will depend on the quality of the natural language descriptions generated by GPT and the capabilities of the text-to-image synthesis model.
  - There may be some errors or inconsistencies in the generated images, especially if the natural language descriptions are ambiguous or incomplete.
- The computational cost of using GPT for image generation.
  - Generating natural language descriptions using GPT is a resource-intensive process, and training a text-to-image synthesis model also requires significant computational resources.
  - This may limit the scalability and practicality of using GPT for image generation in some contexts.

History of Computer Vision

- <2012: Traditional Machine Learning
- 2012: ImageNet, Convolutional Neural Network (CNN)
- 2014: Variational Autoencoder (VAE).
- 2014: Generative Adversarial Network (GAN) - Ian Goodfellow.
- 2015: CNN variation surpass human ability
- 2017: Year of GAN
- 2017: Transformer (Vaswani et al, NIPS)
- 2021: Vision Transformer (Dosovitskiy et al, ICLR)
- 2021: DALL-E
- 2022: DALL-E2

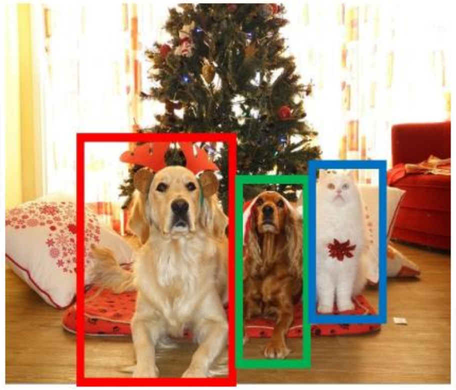
Generative Model

Supervised Learning

- Data: (x,y) -> x is feature, y is label
- Build a function to map x -> y

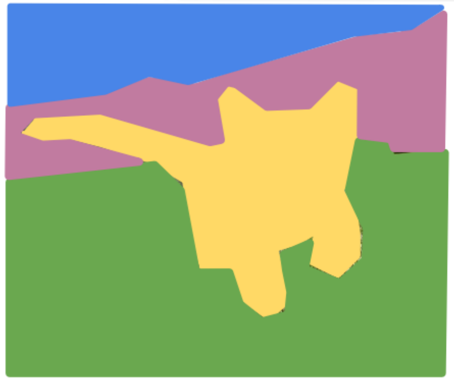
→ Cat

Classification



DOG, DOG, CAT

Object Detection



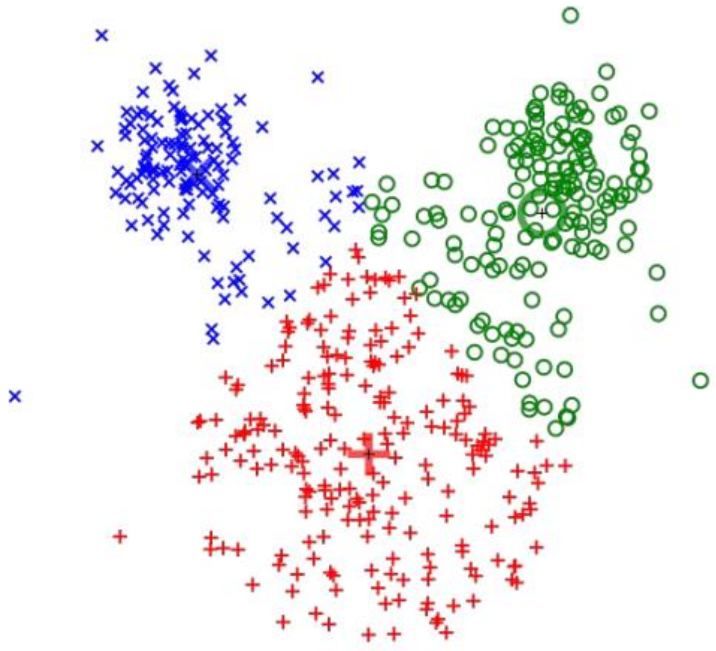
GRASS, CAT, TREE, SKY

Semantic Segmentation

A cat sitting on a suitcase on the floor

Image captioning

Unsupervised Learning



K-means clustering

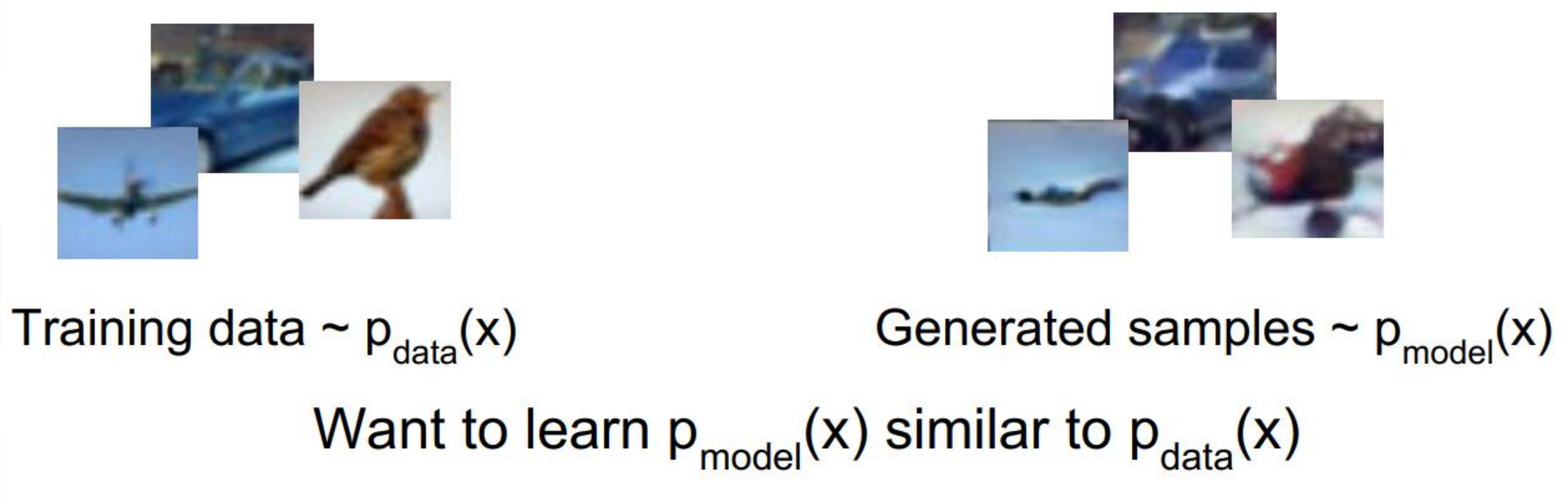
- Data: only x, there is no label.
- Learn some underlying hidden structure of the data.

- Examples: Clustering, dimensionality reduction, feature learning, density estimation

- Implementation: Image super-resolution, colorization, artwork generation, Simulation and planning

Generative Model

- Given training data, generate new samples from same distribution.
- Addresses density estimation, a core problem in unsupervised learning



Taxonomy of Generative Models

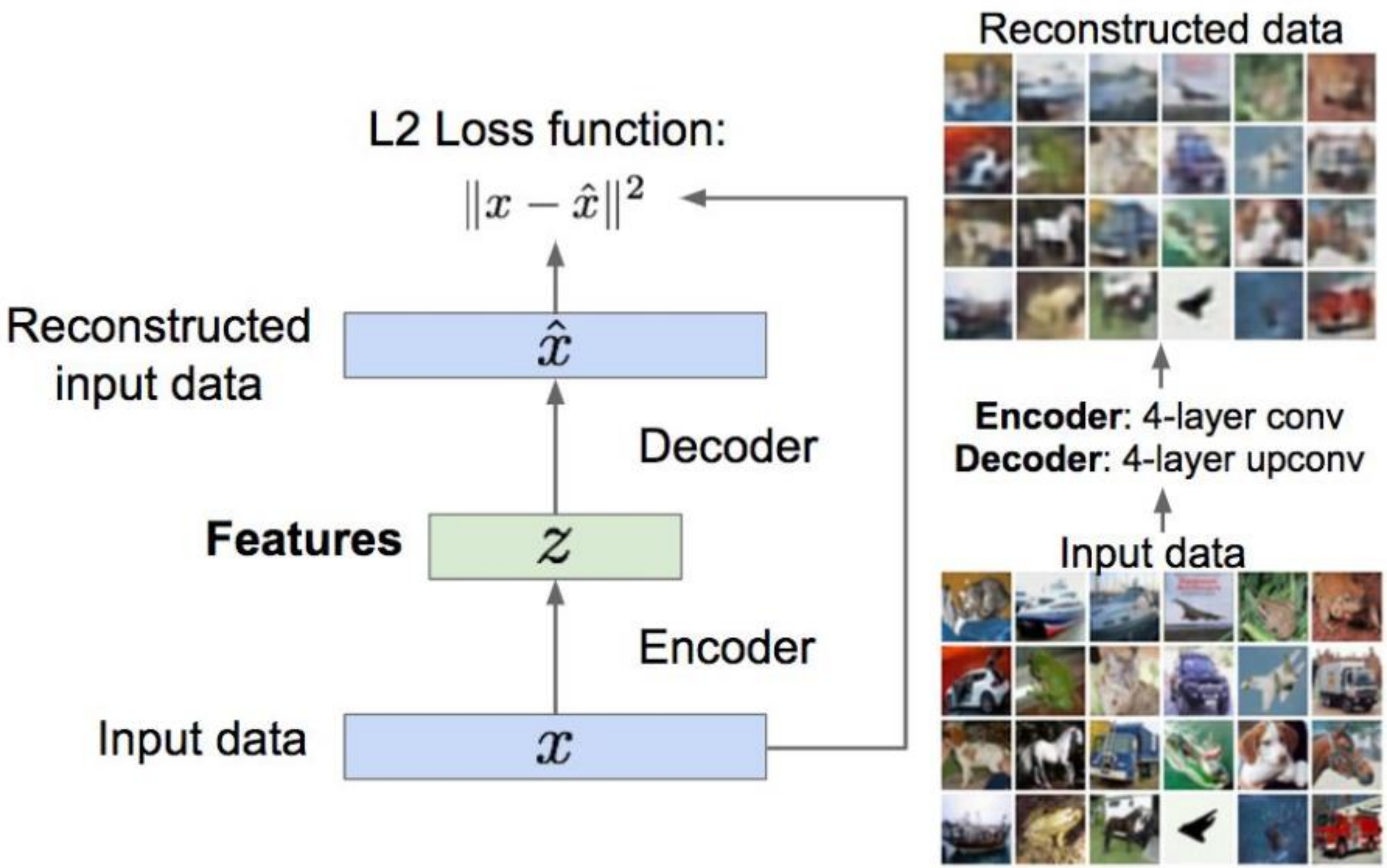
- **Generative models**
  - Explicit density
    - Tractable density
      - Fully Visible Belief Nets
        - NADE
        - MADE
        - PixelRNN/CNN
      - Change of variables models (nonlinear ICA)
    - Approximate density
      - Variational
        - Variational Autoencoder
      - Markov Chain
        - Boltzmann Machine
  - Implicit density
    - Direct
      - GAN
    - Markov Chain
      - GSN

Figure copyright and adapted from Ian Goodfellow, Tutorial on Generative Adversarial Networks, 2017.

Autoencoder

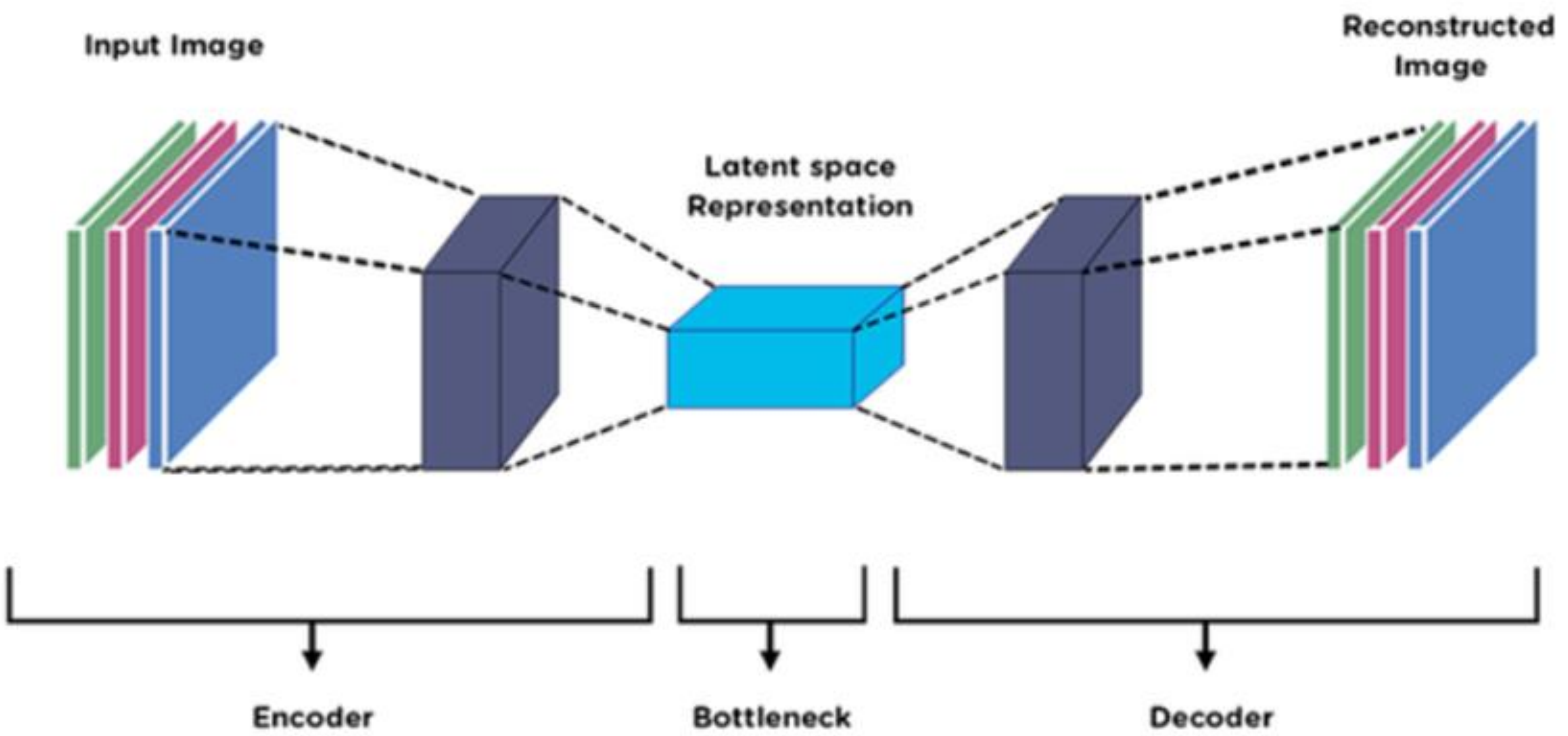
Autoencoder (AE)

- Autoencoding- encoding itself
- Membangkitkan suatu data:
  1. Memetakan antara input dengan output spasial yang lebih kecil dengan encoder
  2. Features capture factors of variation in training data.
  3. Lalu memetakan kembali ke output spasial yang lebih besar dengan decoder



Autoencoders (Feature learning)

Autoencoder (AE)



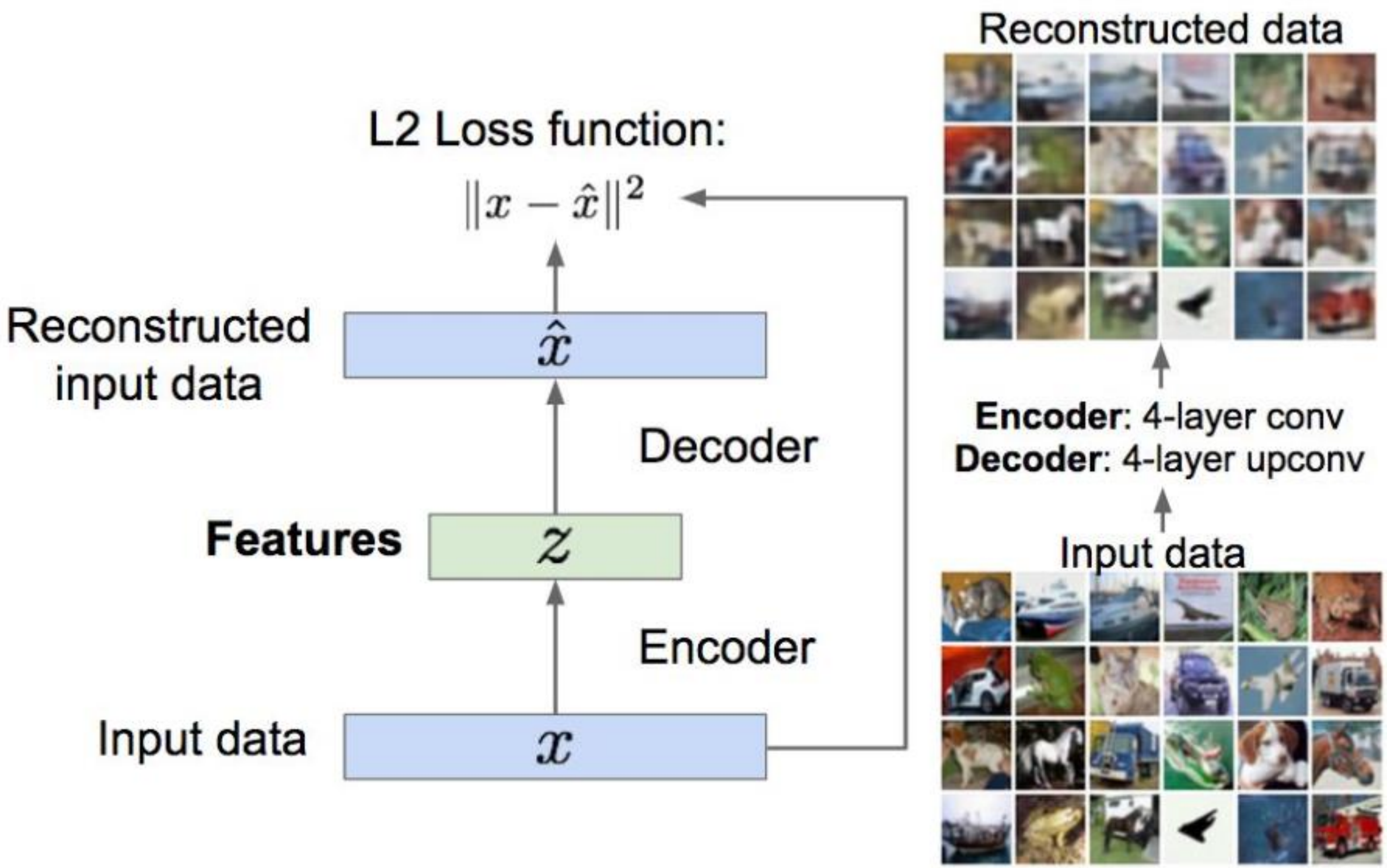
Autoencoder (AE)

Encoder:

» Unsupervised approach for learning a lower-dimensional feature representation from unlabeled training data

» Dimensionality Reduction: Generate features to capture meaningful factors of variation in data

» Linear +nonlinearity (sigmoid), Deep fully-connected, ReLU CNN



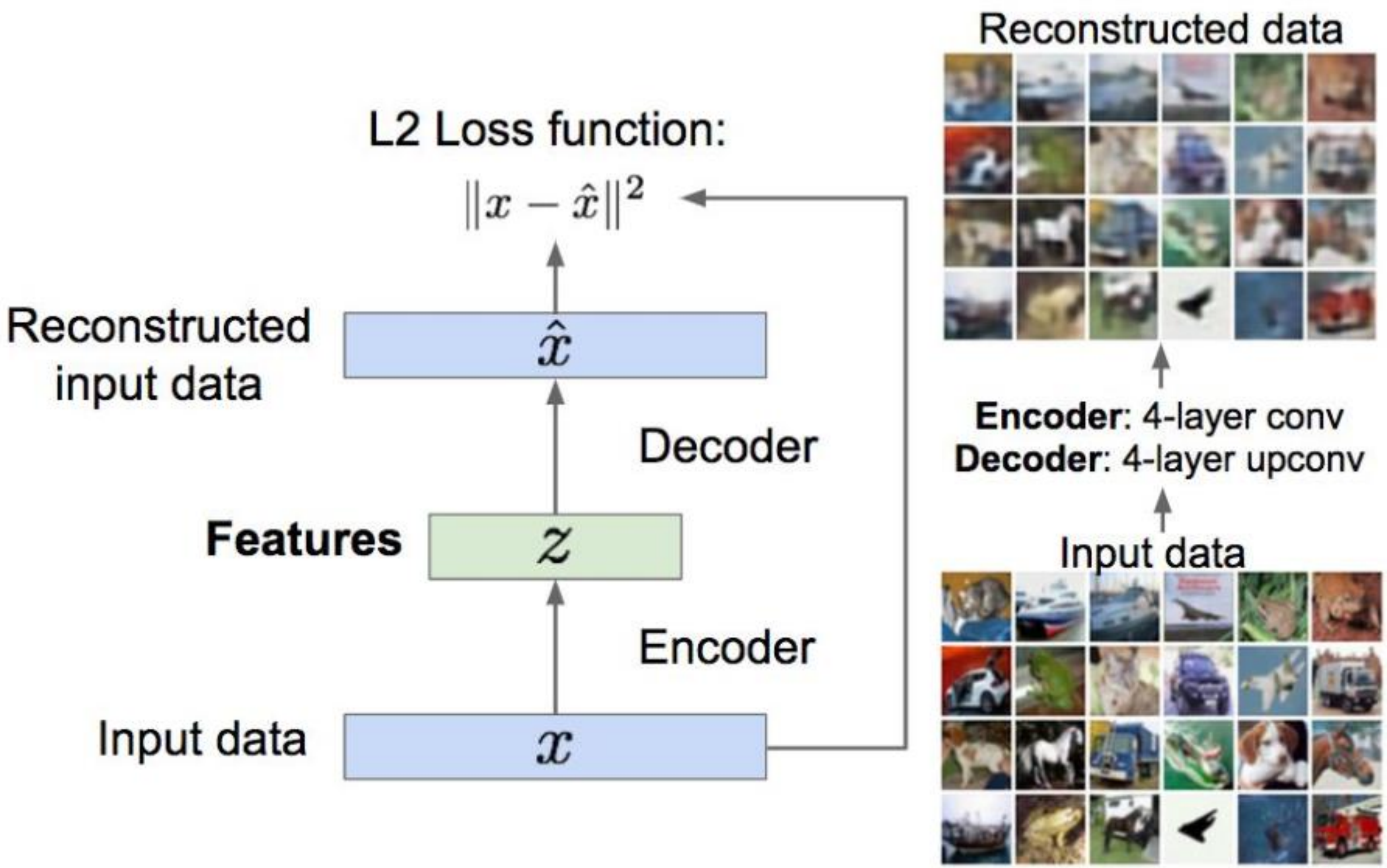
Autoencoders (Feature learning)

Autoencoder (AE)

Decoder:

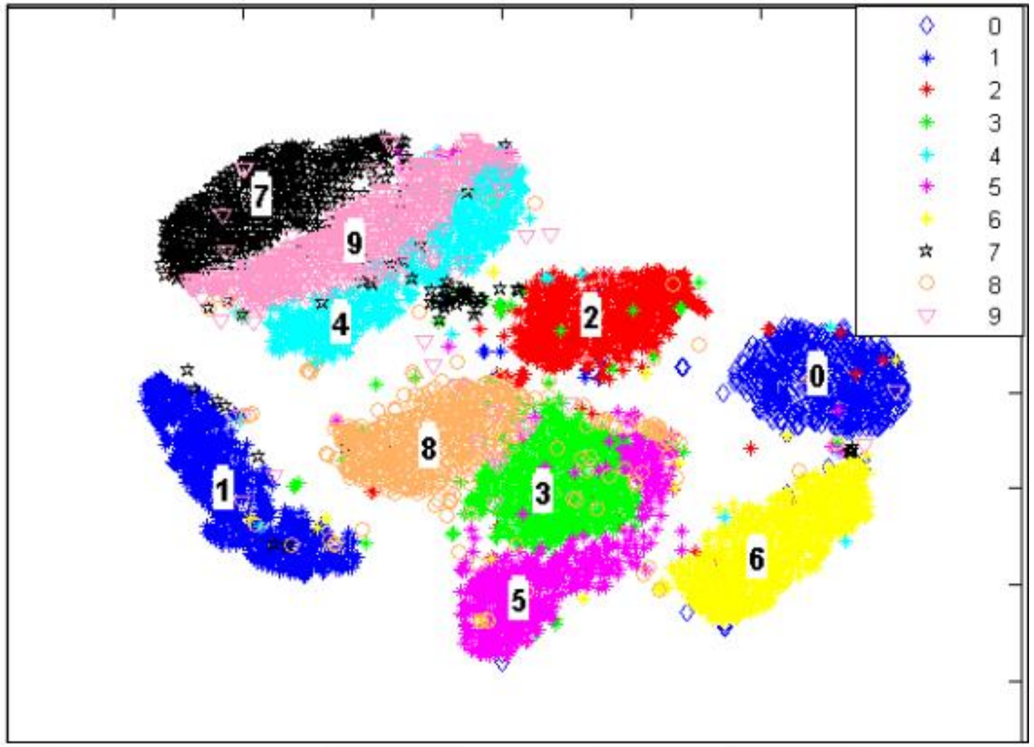
» How to learn this feature representation? Train such that features can be used to reconstruct original data

» Linear +nonlinearity (sigmoid), Deep fully-connected, ReLU CNN



Autoencoders
(Feature learning)

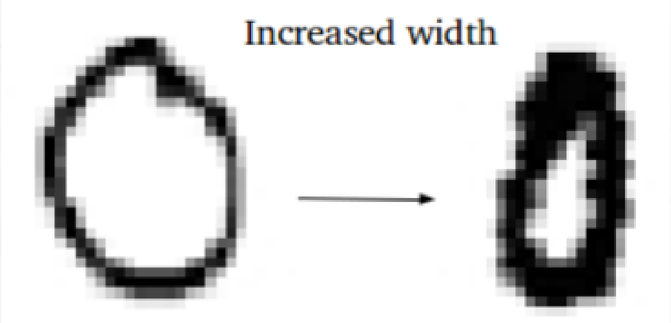
Autoencoder Problem



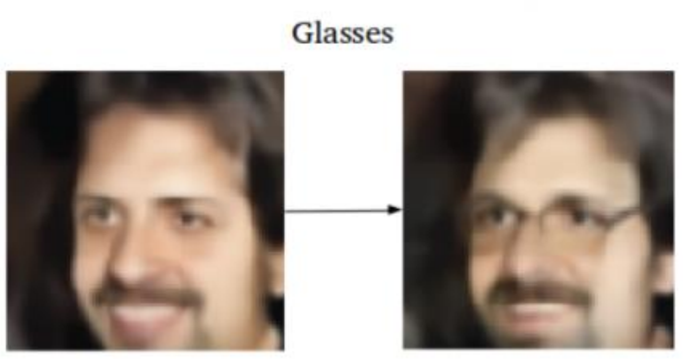
- Hasil vektor encoding terkluster dan terpisah berdasarkan data.
- Laten space tempat mengonversi input menjadi bentuk vektor encoding, hasilnya tidak terdistribusi secara kontinyu.
- Latent space yang diskontinuiti (jarak antar cluster), hasil gambar yang dibangkitkan menjadi aneh dan tidak realistis

Autoencoder Problem

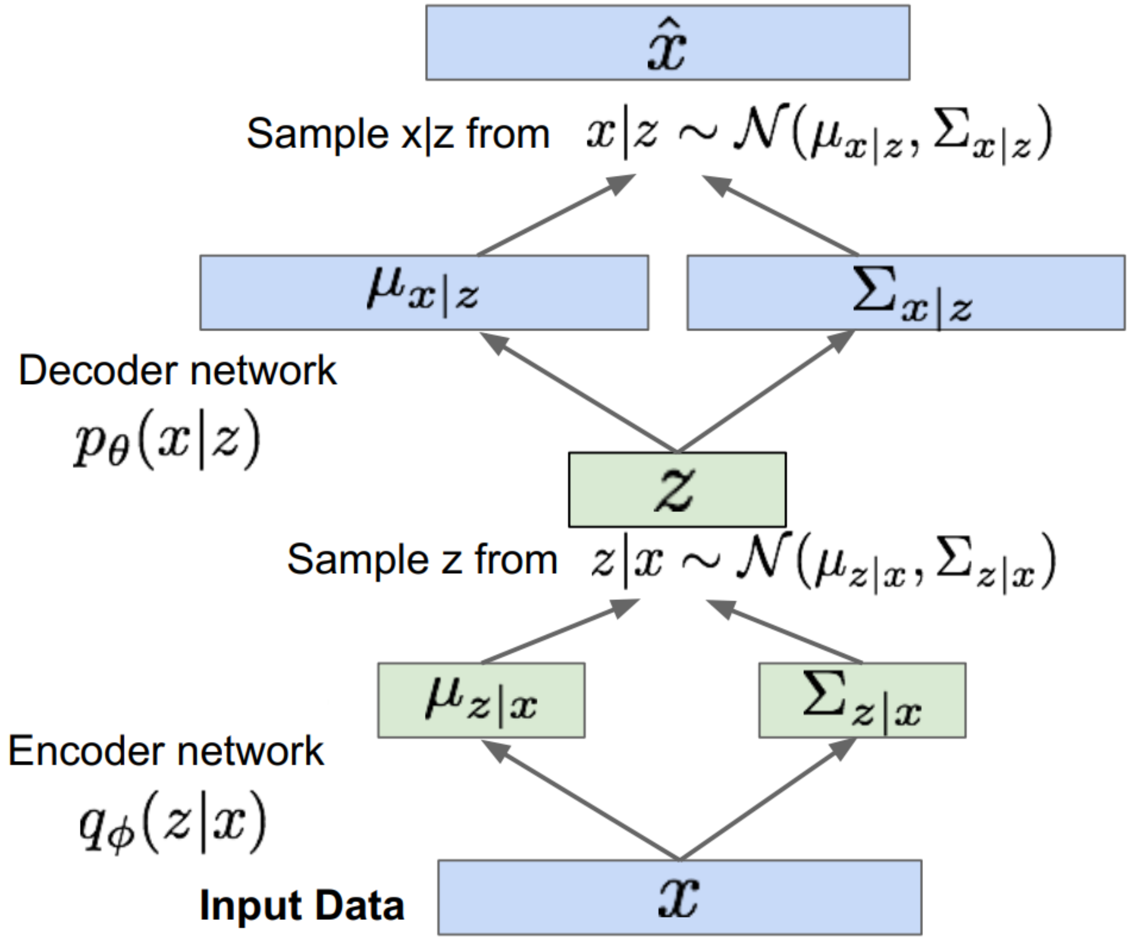
» Tepat jika kita ingin mereplikasi data yang sama.
» Contoh: Image coloring, struktur output image tidak boleh berubah dari input



» AutoEncoder bukanlah (belum) Generative Model.
» Karena AutoEncoder melakukan pemetaan, bukan melakukan pembangkitan data baru.
» Agar dapat membangkitkan data baru:
  » Distribusi latent space harus kontinyu.
  » Tersebar secara merata untuk seluruh encoding data.



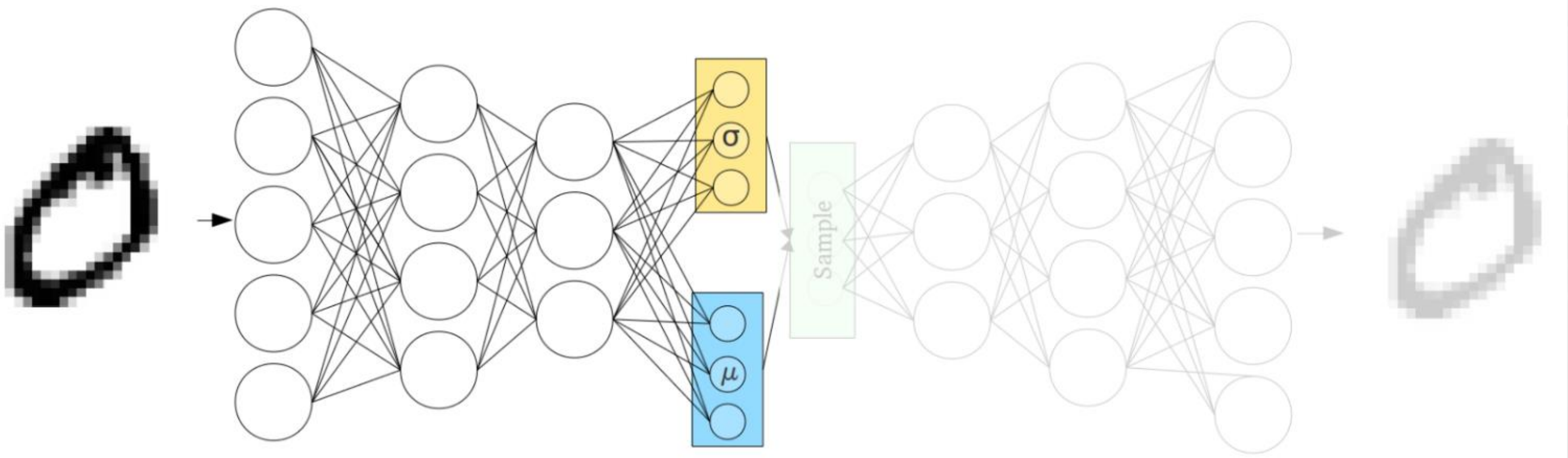
Variational Autoencoder (VAE)



- Perbedaan dengan AE standar: saat proses pelatihan, output encoding VAE (di tengah) bukanlah sekedar pemetaan input menjadi output.
- Input yang sama, menghasilkan mean dan standar deviasi yang sama
- Namun, hasil encoding akan berbeda setiap kali dicoba karena dibangkitkan secara random (bukan sekedar pemetaan)

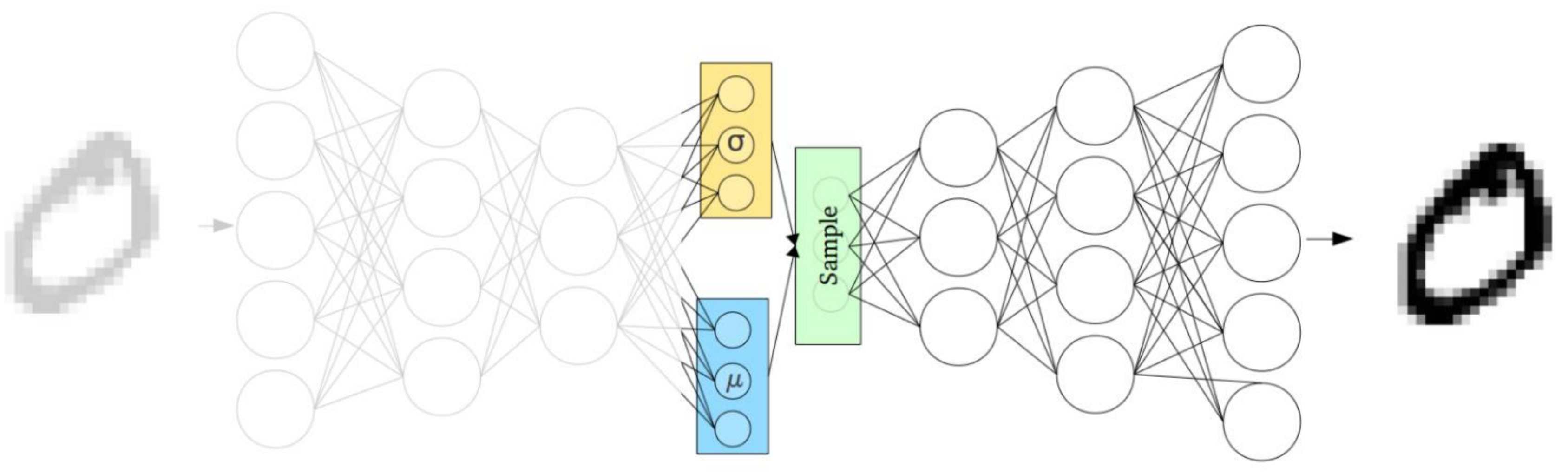
Variational Autoencoder (VAE)

- VAE memetakan input terhadap output distribusi data berisi prior mean dan kovarian (standar deviasi) dari keseluruhan dataset.
- Encoder tidak mempelajari untuk mereduksi data, namun mempelajari untuk merangkum dataset, bagaimana persebaran rata-rata atribut terhadap kovarian nya (hubungan antar atribut)

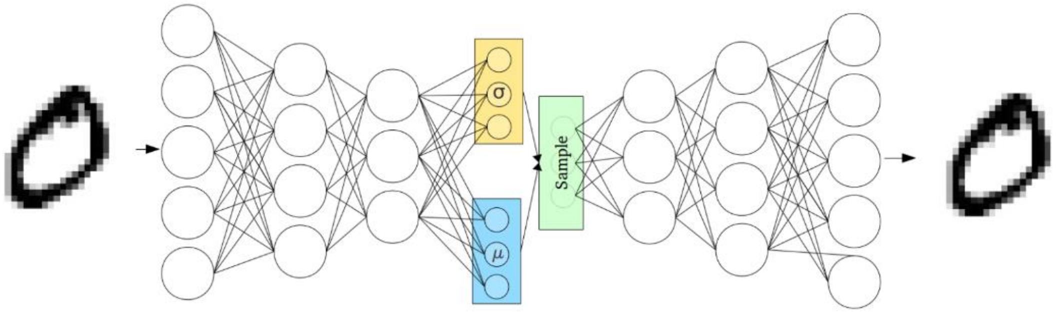


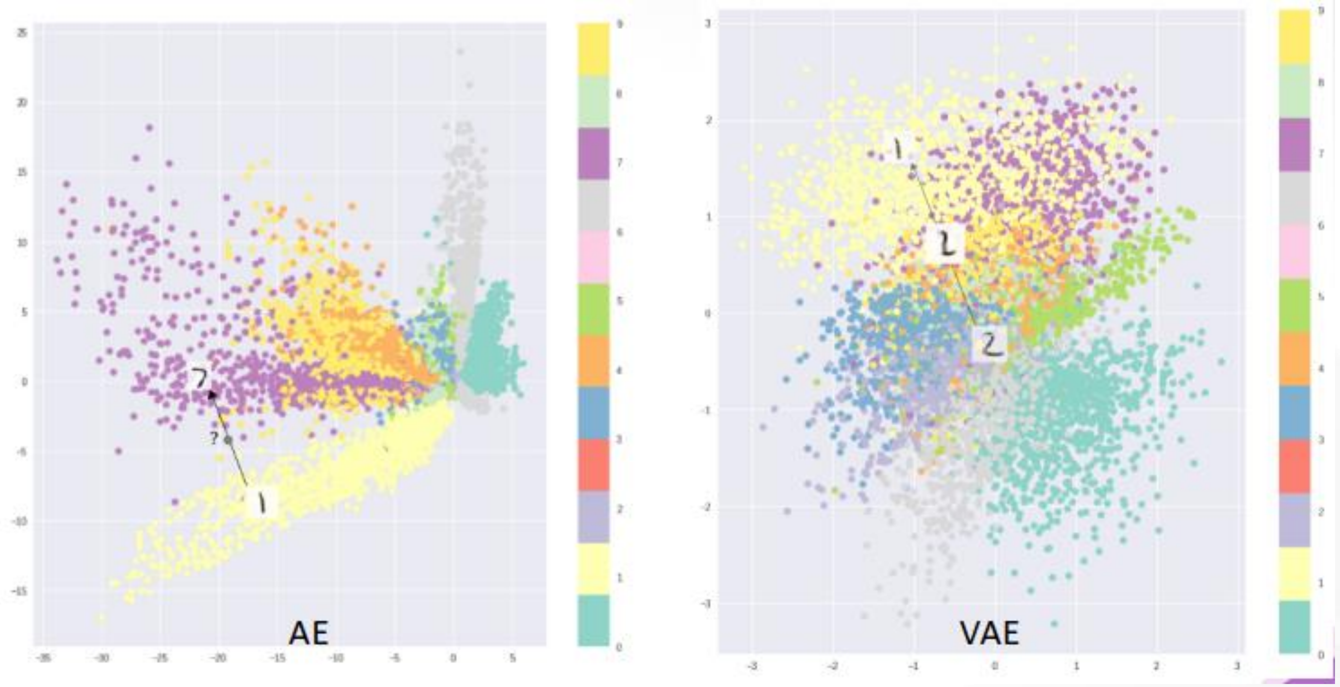
Variational Autoencoder (VAE)

- Pada Decoder, data disample (bangkitkan latent space secara random) berdasarkan mean dan kovarian (standar deviasi) yang didapat.
- Hasil latent space random tsb kemudian diberikan pada decoder untuk membangkitkan gambar.
- Input yang sama, menghasilkan mean dan standar deviasi yang sama, namun hasil encoding akan berbeda setiap kali dicoba karena dibangkitkan secara random (bukan sekedar pemetaan)



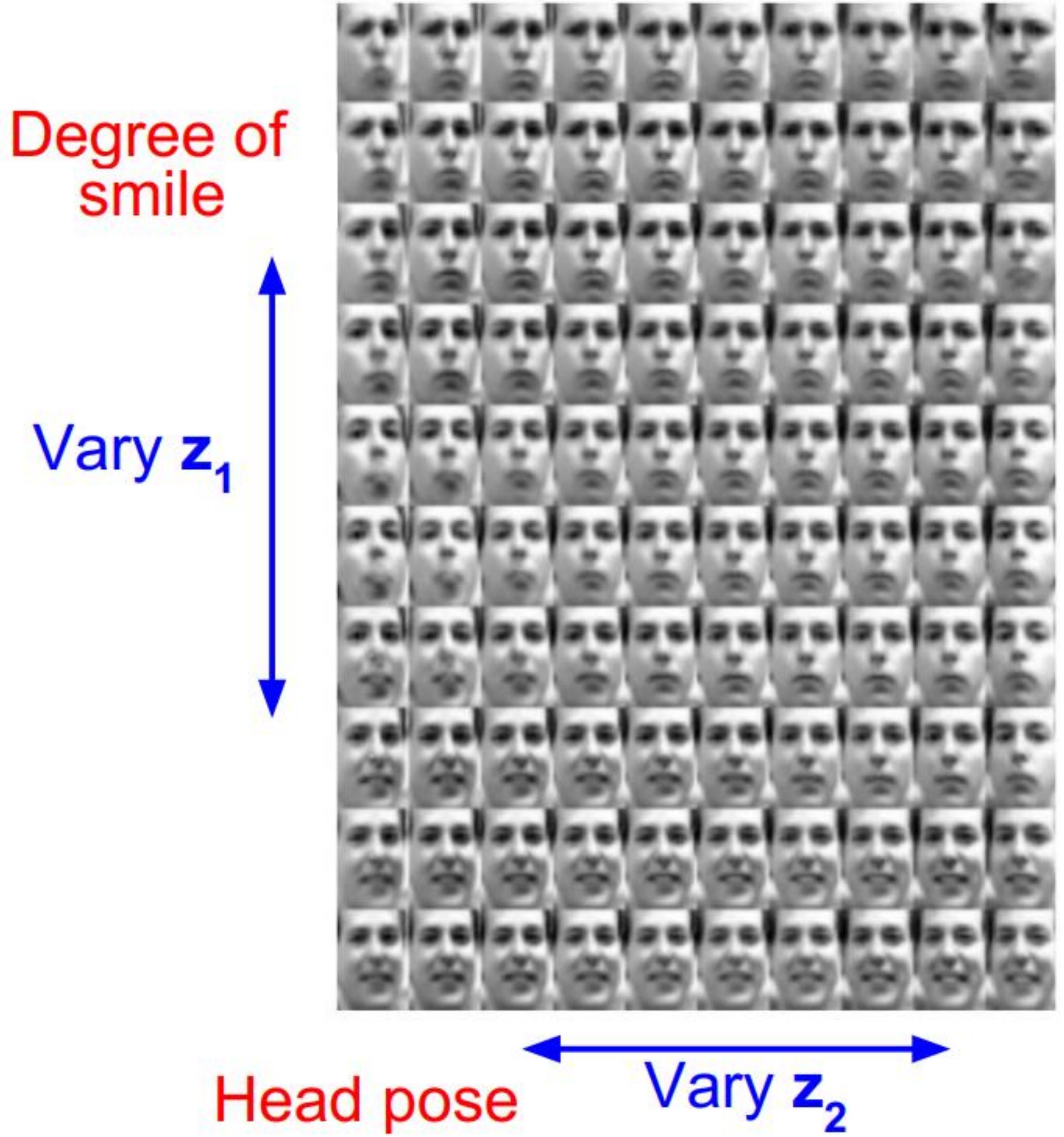
Variational Autoencoder (VAE)





- Dari encoding yang berbeda tersebut, maka bisa membangkitkan gambar yang berbeda-beda.
- Proses tersebut menghasilkan vektor encoding (Latent Space) dari VAE yang terdistribusi secara kontinu.

Variational Autoencoder (VAE)



- VAE dapat menghasilkan data yang baru yang sangat mirip dengan data input, namun bukanlah duplikat dari data asli
- Bisa digunakan untuk membantu (boosting) proses Supervised Learning: Klasifikasi, Dataset kurang, latih VAE, generate data baru, lanjutkan pelatihan.

UNIVERSITAS ISLAM INDONESIA

Center of Data Science Universitas Islam Indonesia

Generative Adversarial Network

Generative Adversarial Network (GAN)

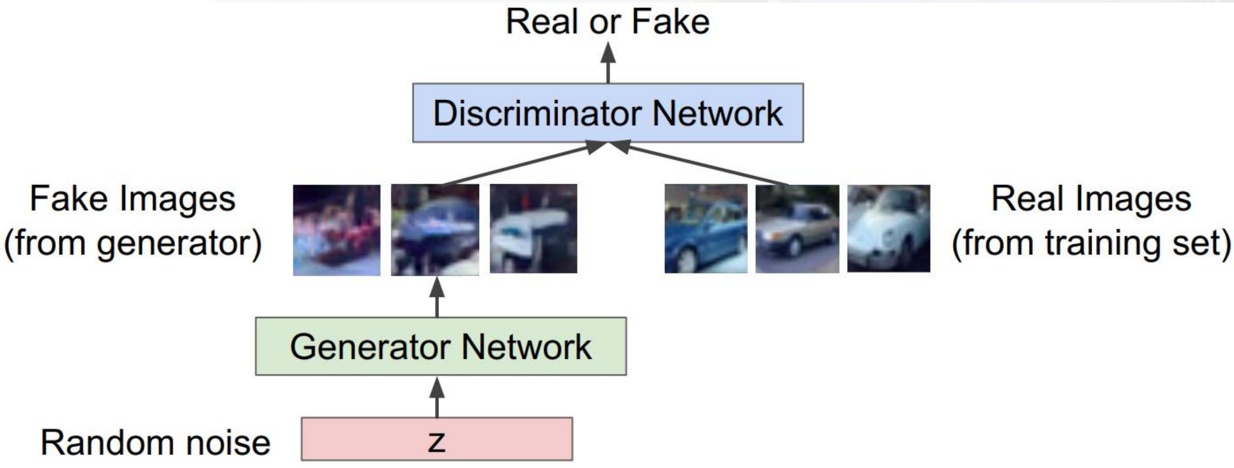
- Untuk melakukan sampling data dari data yang memiliki distribusi kompleks dan berdimensi tinggi sangat susah dan tidak ada cara langsung.
- Menggunakan pendekatan Game Theory: data dibangkitkan dari distribusi data latih menggunakan sistem bernama “2-player Game”

- Terdapat 2 jaringan di dalam GAN:
  - Generator Model
  - Discriminator Model
- Berbeda dari AutoEncoder, kedua jaringan tidak bekerja sama, justru kedua jaringan dilatih untuk saling bertentangan

2-player Game

1. Generator Network
   - » Dilatih untuk membangkitkan gambar baru dari input “seed” latent space random
   - » Dilatih untuk bisa mengelabui Discriminator agar mengira gambar baru adalah gambar asli.

2. Discriminator Network
   - » Dilatih untuk membedakan input gambar baru (fake examples) dari generator dengan gambar asli dari data latih (real examples).
   - » Terus dilatih untuk selalu bisa membedakan



2-player Game - exemple

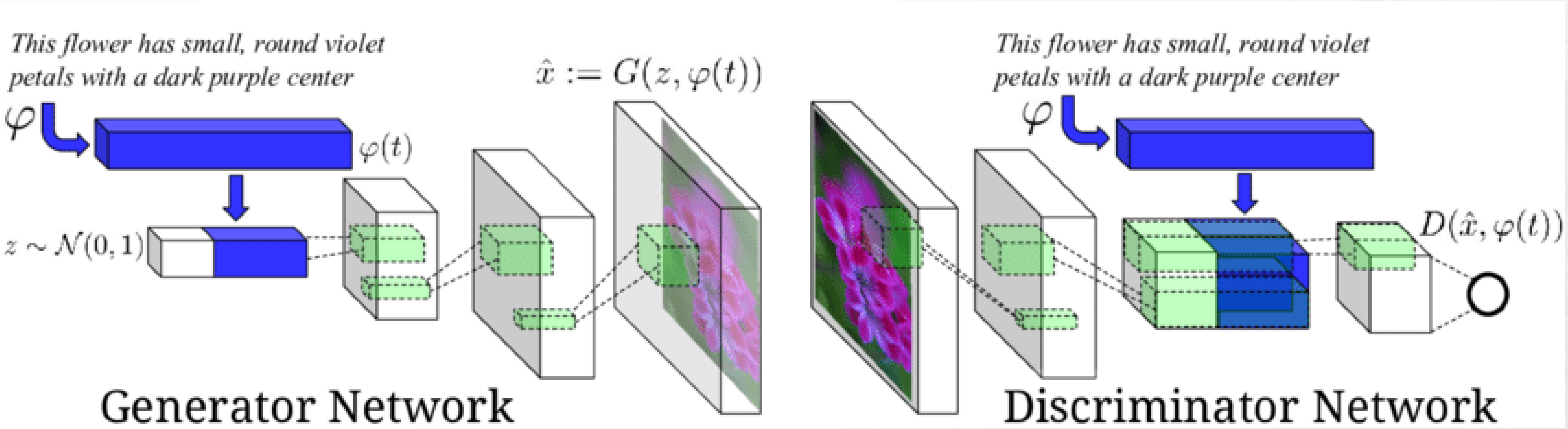
1. Generator Network
   - » Jaringan pemalsu uang,
   - » Selalu berusaha membuat uang palsu yang lolos dari pengecekan polisi

2. Discriminator Network
   - » Jaringan polisi/pengecek uang palsu
   - » Selalu berusaha untuk mengidentifikasi mana yang uang palsu

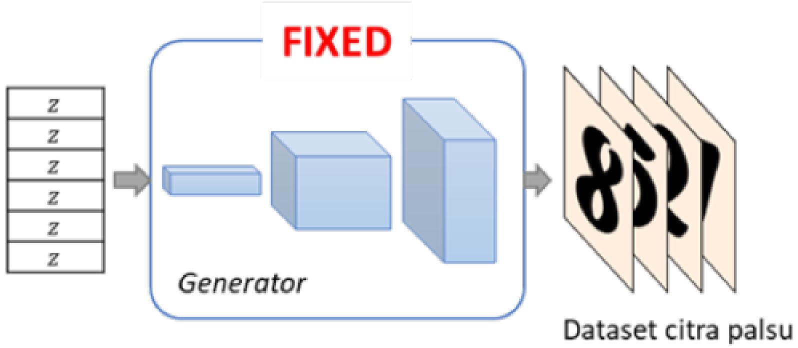
2-player Game

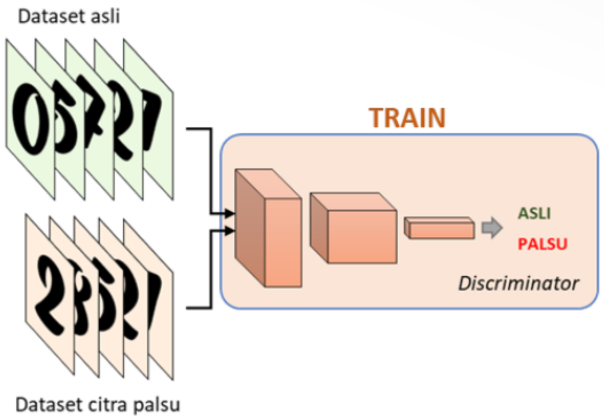
1. Generator Network
   » Arsitektur seperti Decoder Convolutional
   » Input vector, output image volume

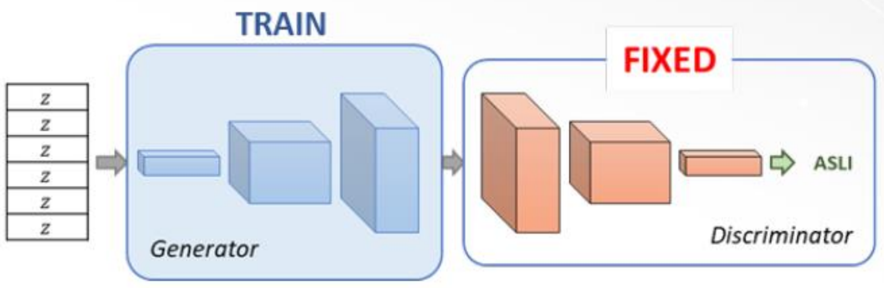
2. Discriminator Network
   » Arsitektur klasifikasi biner
   » Input image volume, output satu neuron sigmoid (biner)



GAN Training in 1 epoch







1. Generator membangkitkan sejumlah n data baru dari input matrix random z berukuran $n \times v$.
   v= panjang vektor seed

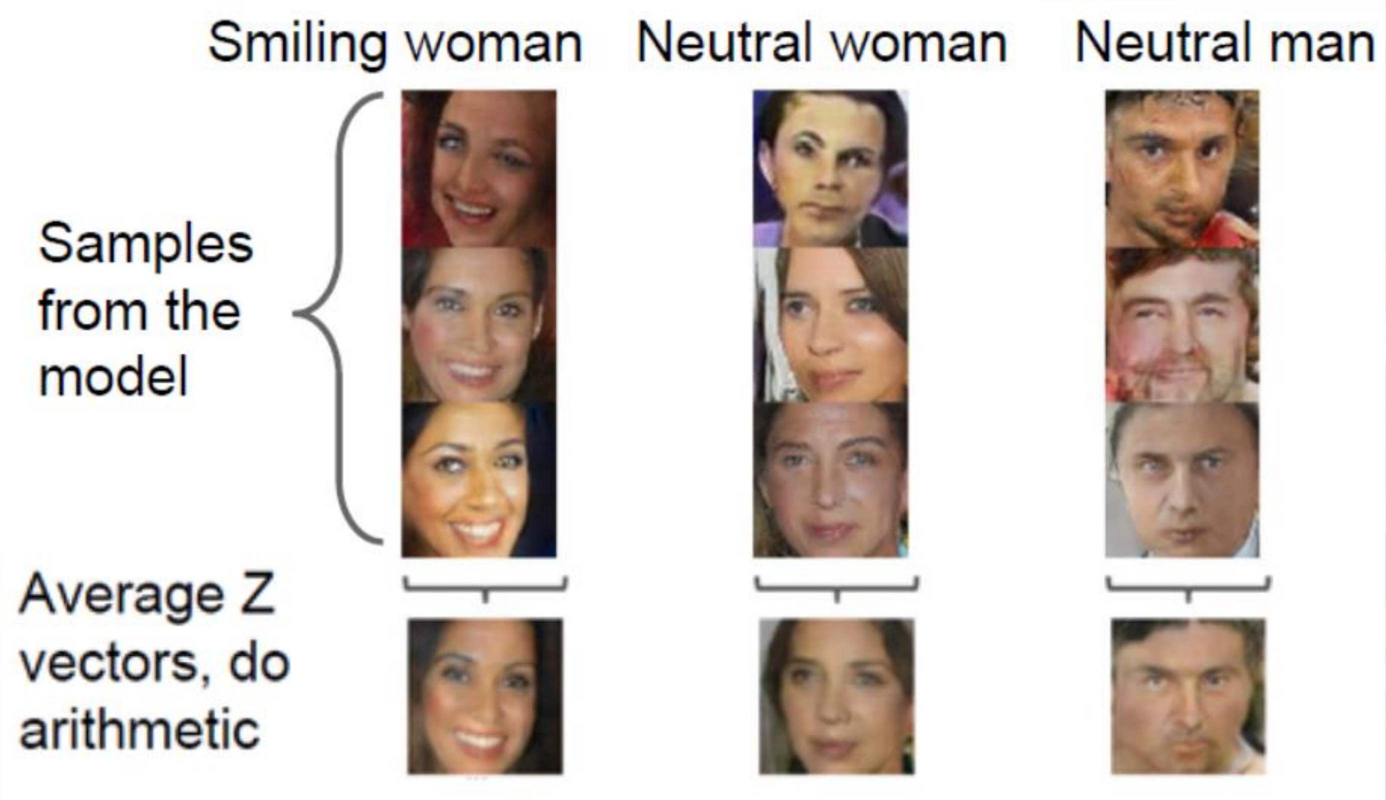
2. Latih Discriminator:
   a. Ambil sejumlah n data asli
   b. Labeli data asli dengan label y = 1
   c. Labeli data palsu dengan y = 0
   d. Jalankan pelatihan diskriminator terhadap dataset tersebut.
   e. Hingga diskriminator konvergen (mampu mengenali asli dan palsu)

3. Latih Generator,
   a. Bangkitkan matrix random z baru
   b. Bangkitkan n gambar baru
   c. Cek gambar baru ke diskriminator
   d. Hitung error: berapa banyak gambar yang gagal mengelabui diskriminator
   e. Update bobot generator agar bisa membuat gambar lebih bagus

GAN - bedroom

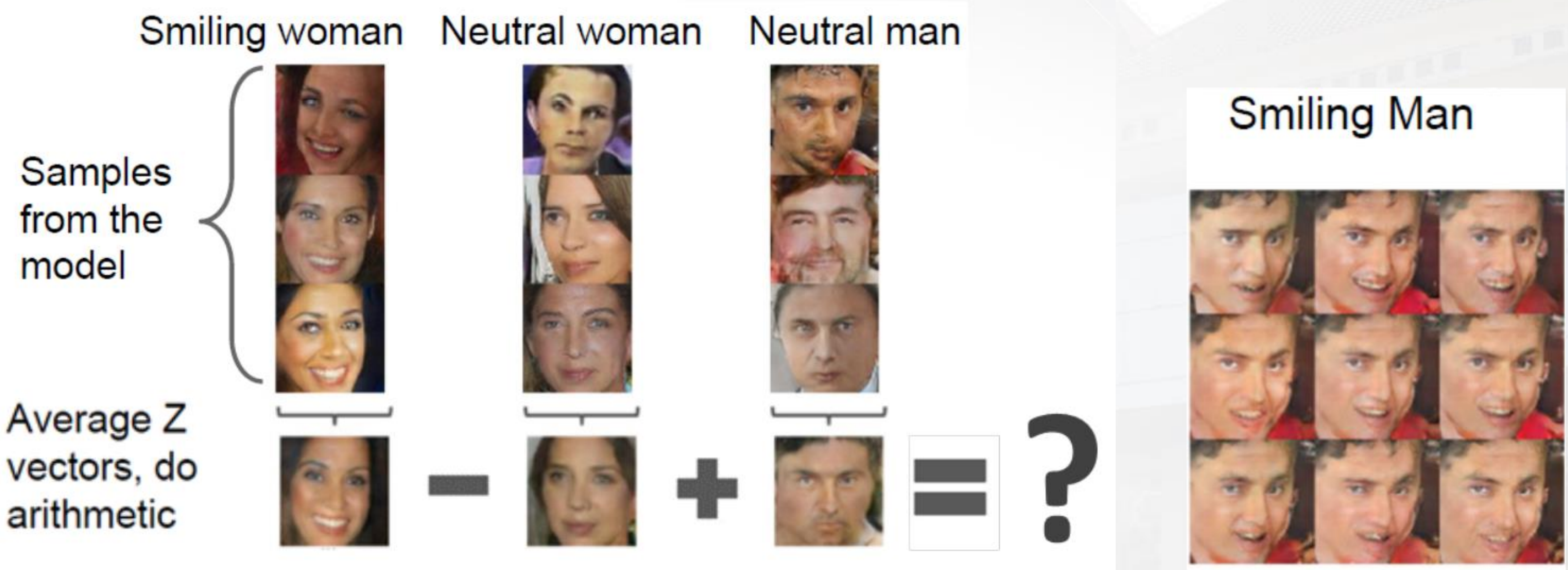
Radford et al, ICLR, 2016

GAN - face



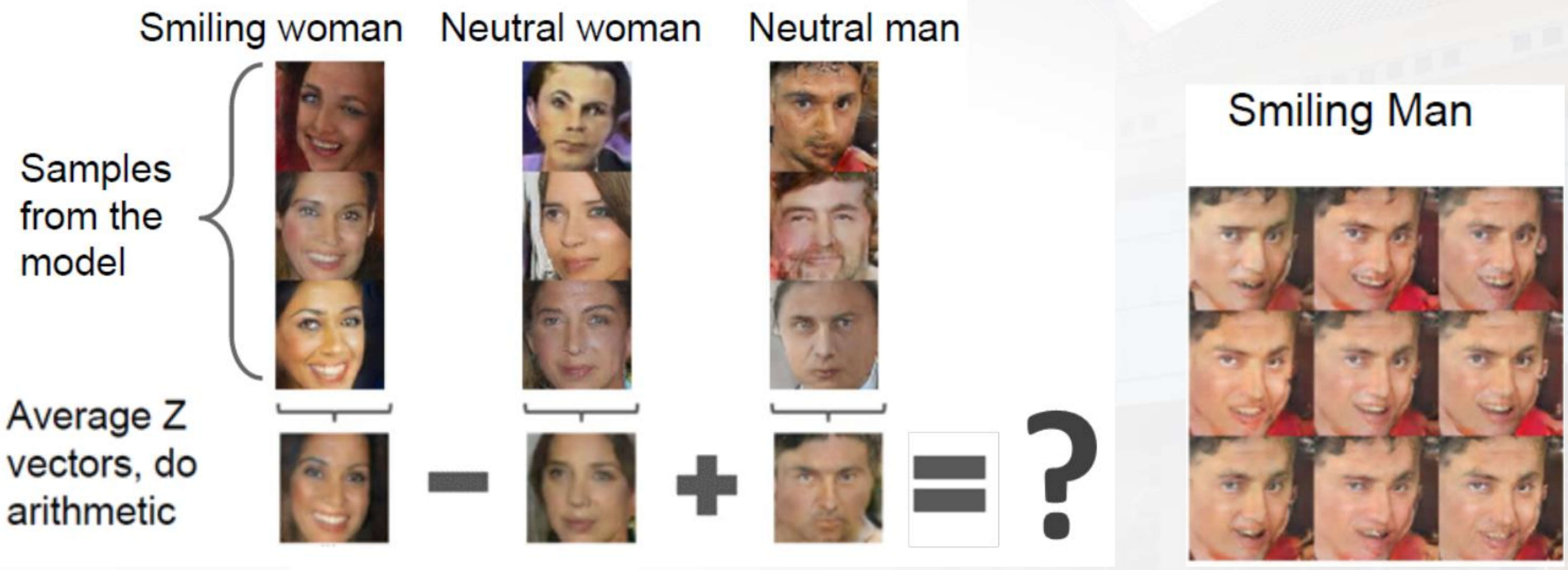
Radford et al, ICLR, 2016

GAN - face



Radford et al, ICLR, 2016

GAN - face



Smiling woman
Neutral woman
Neutral man
Samples from the model
Average Z vectors, do arithmetic
Smiling Man

GAN – Face

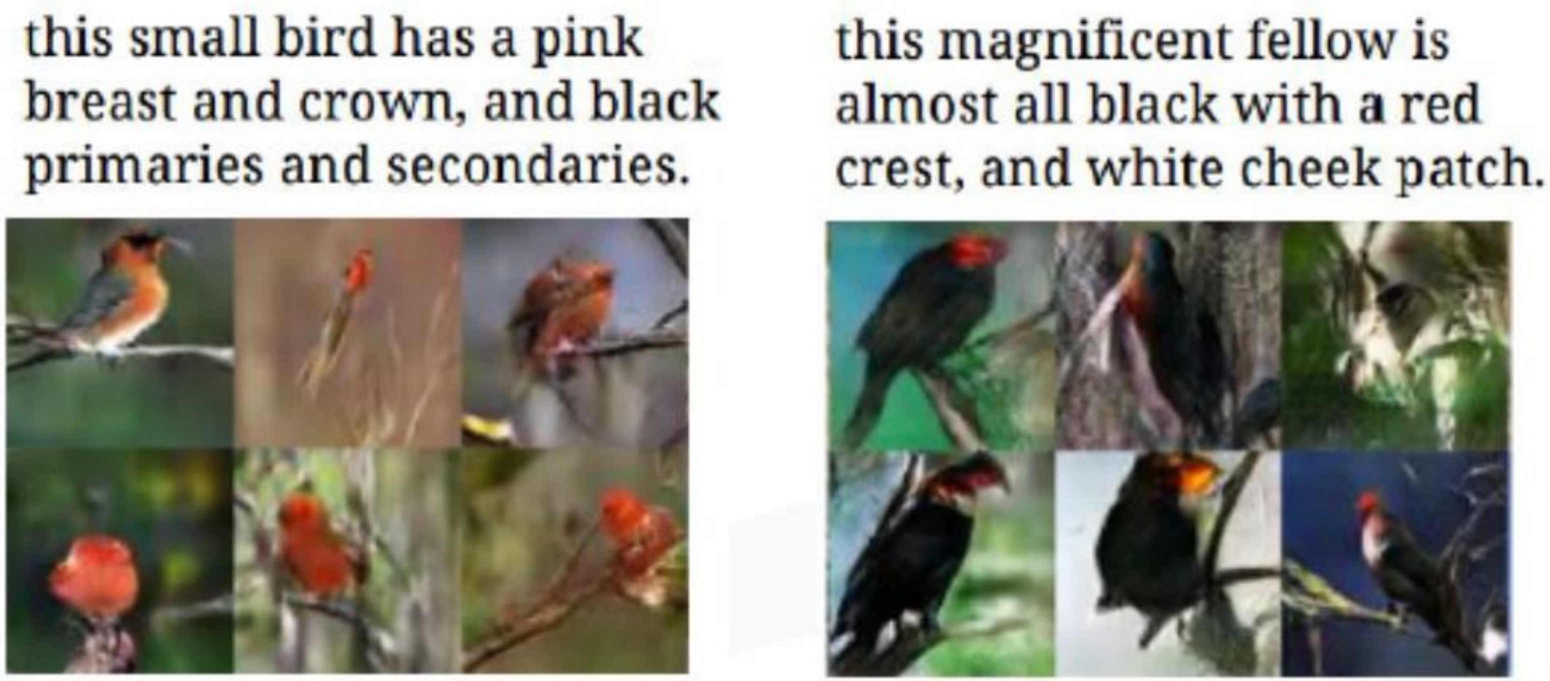
DCGAN 11/2015

EBGAN-PT 9/2016

BEGAN. Bertholet et al. 2

Progressive GAN, Karras 2018.

GAN – Text-image Synthesis



Akata et al, 2017

GAN – Super Image Resolution

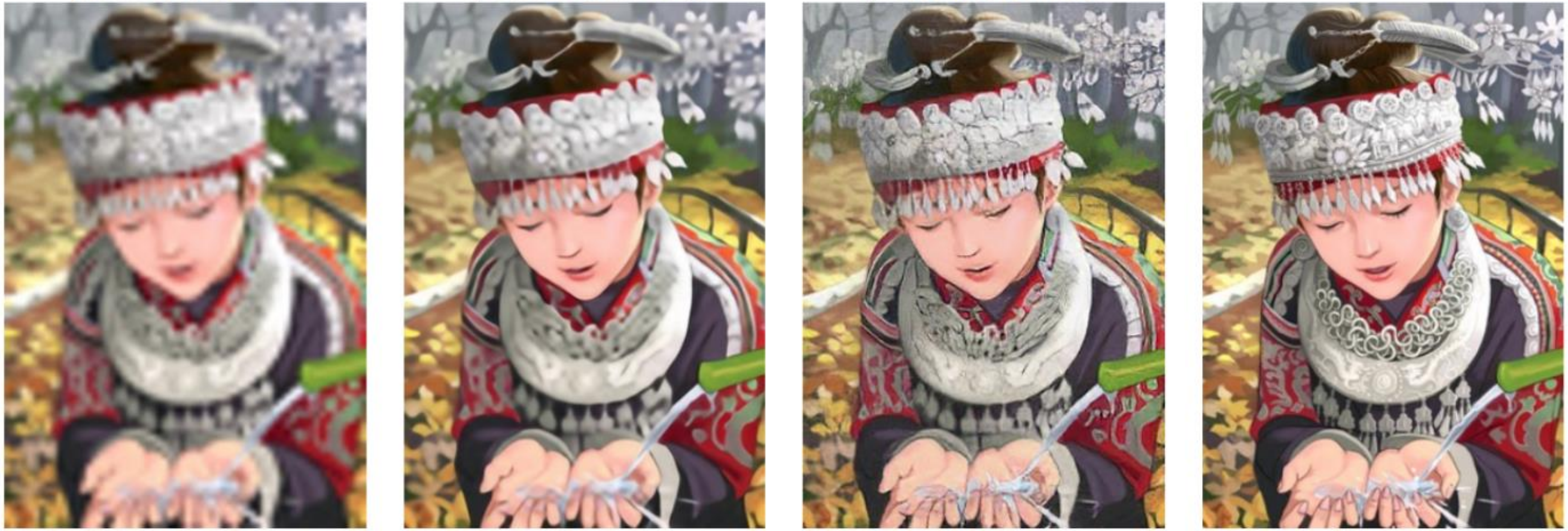
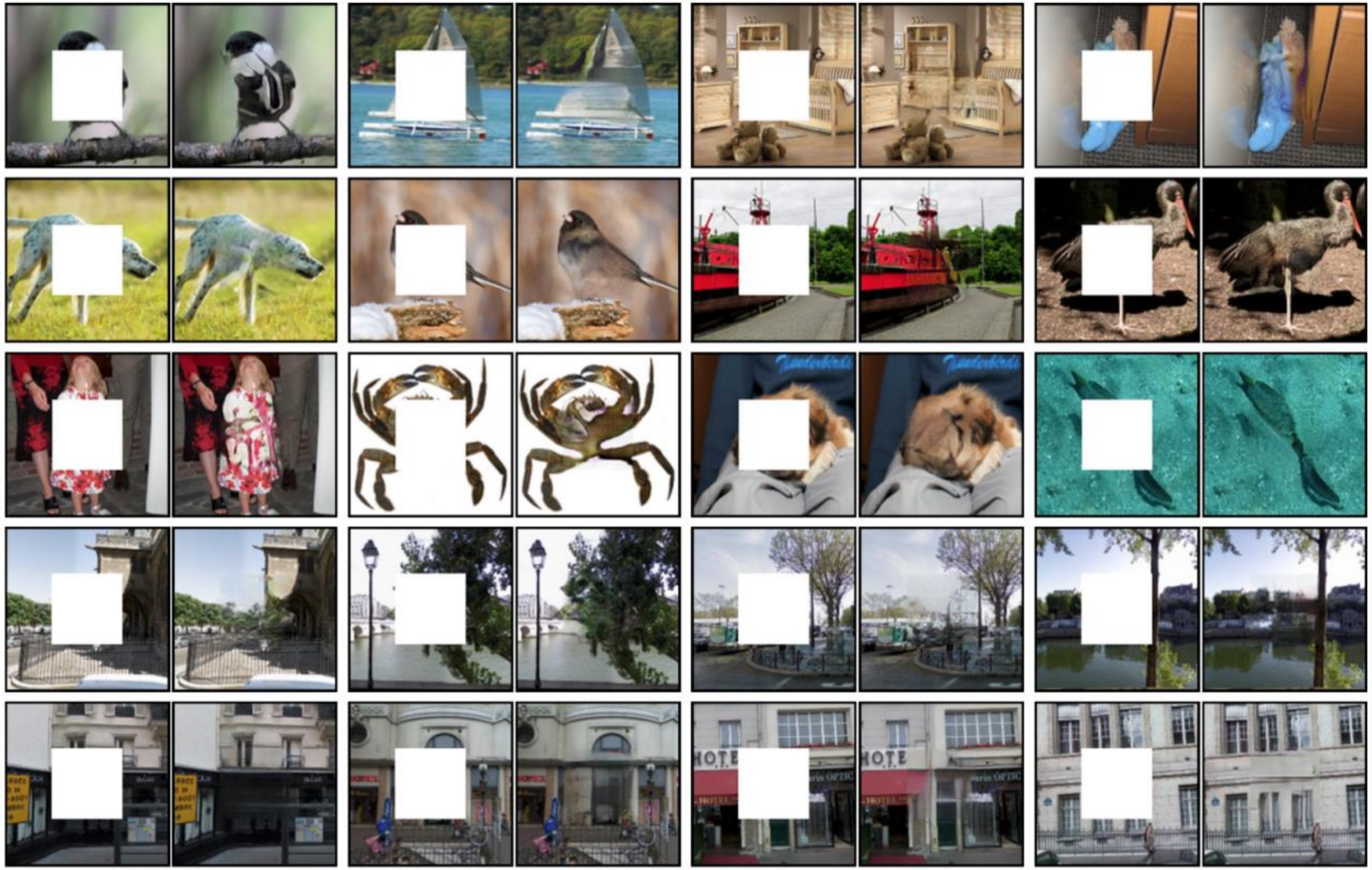


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

GAN – Image Stitching and Inpainting



AI art generator

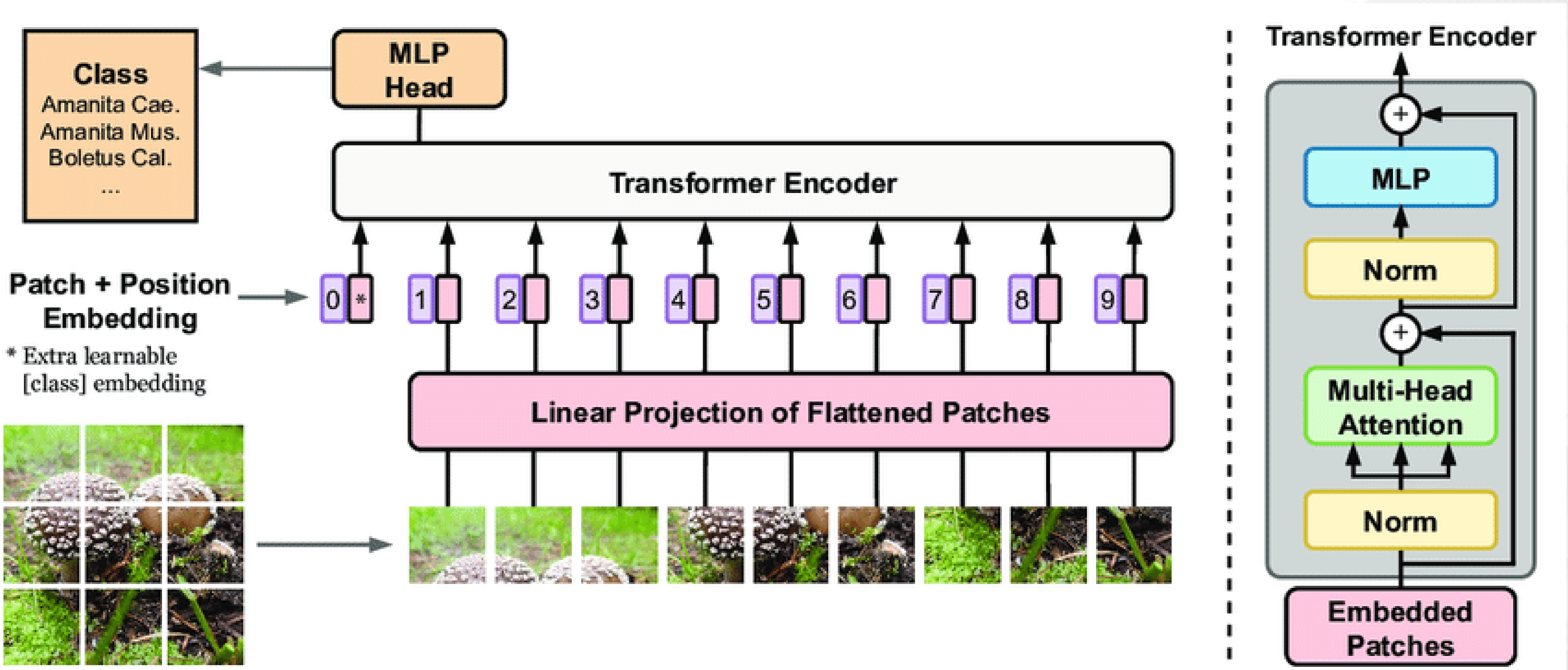
- DALL-E2 – OpenAI's image-generating tool, known for creating astonishingly realistic images. https://openai.com/product/dall-e-2
- Jasper Art – A popular AI tool that converts text inputs to images. https://art.jasper.ai/
- Night cafe – Another popular tool, known for its wider range of features compared to its rivals. https://nightcafe.studio/
- Photosonic – An art generator with the option of image generation or image modification. https://photosonic.writesonic.com/

iGPT

iGPT

- https://openai.com/research/image-gpt
- Just as a large transformer model trained on language can generate coherent text, the same exact model trained on pixel sequences can generate coherent image completions and samples.
- iGPT trained GPT-2 on images unrolled into long sequences of pixels. The result showed that the model appears to understand 2-D image characteristics such as object appearance and category.

Vision Transformer (VIT)



https://www.researchgate.net/publication/357885173_Automatic_Fungi_Recognition_Deep_Learning_Meets_Mycology/figures?lo=1

Terima Kasih

arrie.kurniawardhani@uii.ac.id

Source

» Generative Models, Program Fresh Graduate Academy Digital Talent Scholarship 2019 | Machine Learning.

» Lecture 13: Generative Models, Fei-Fei Li & Justin Johnson & Serena Yeung.

» Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020 Oct 22.

» The GAN Zoo https://github.com/hindupuravinash/the-gan-zoo

» For tips and tricks for trainings GANs https://github.com/soumith/ganhacks