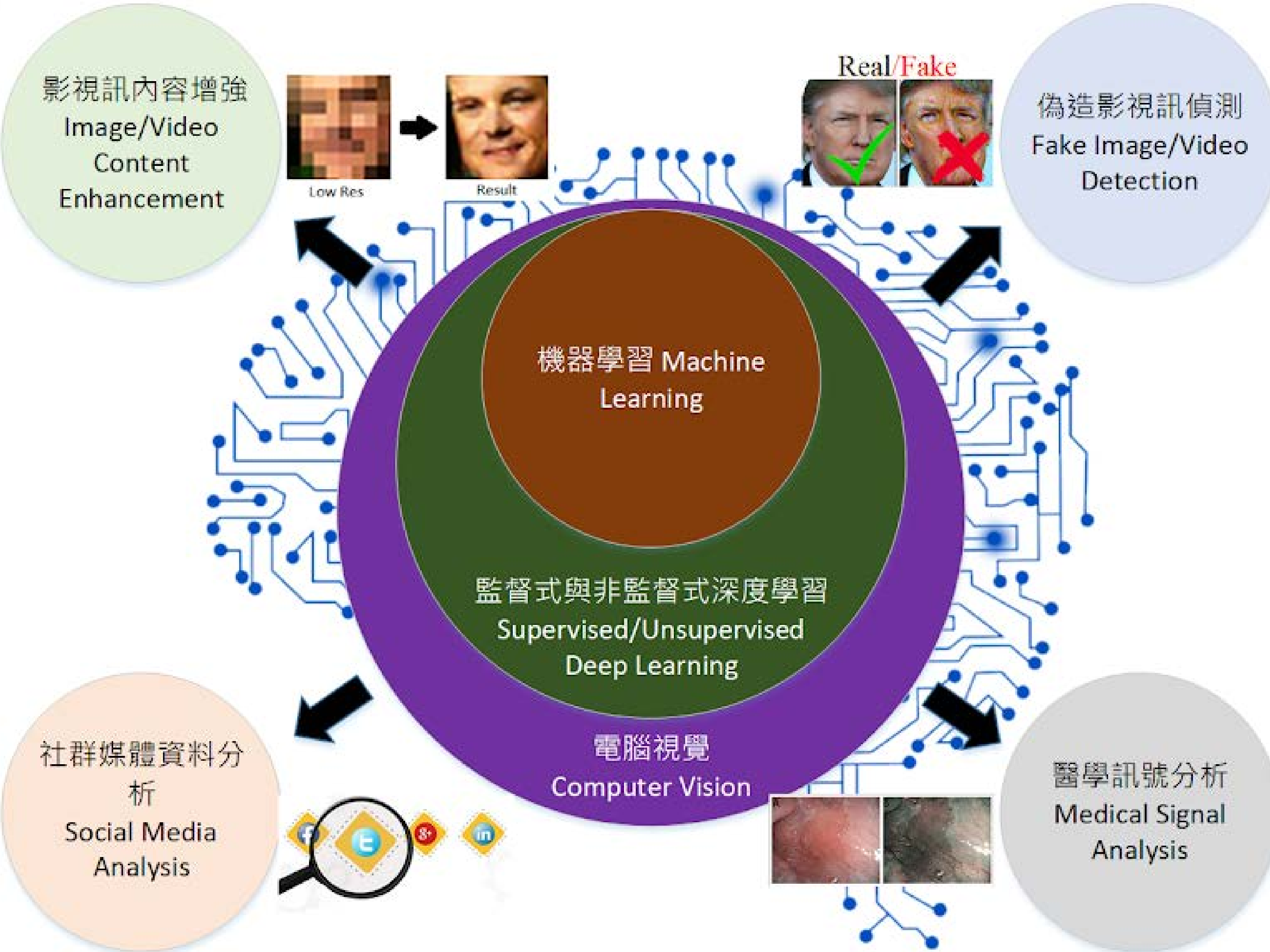
REAL OR FAKE?
GAN & ANTI-GAN

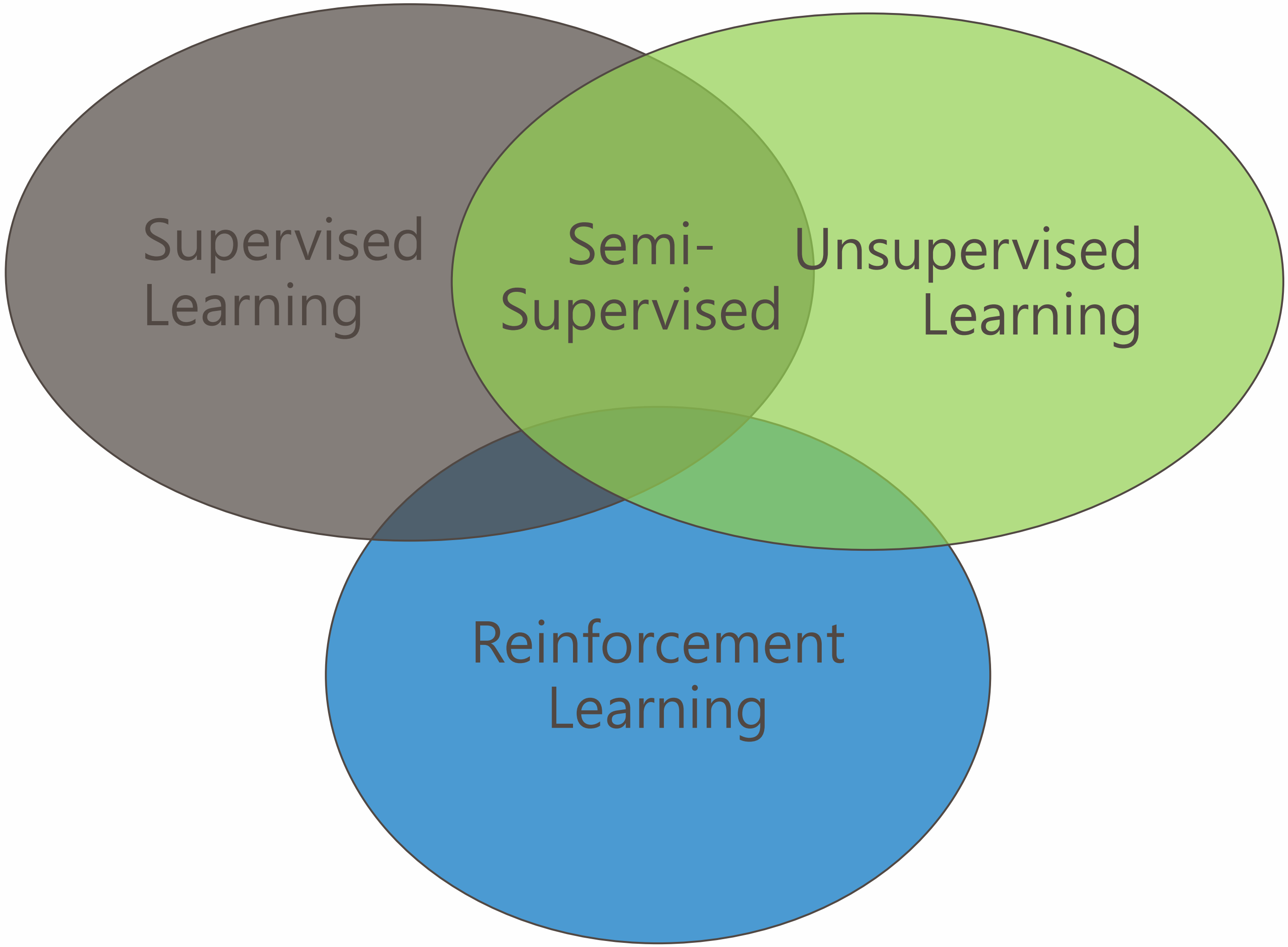
許志仲 (Chih-Chung Hsu)
Assistant Professor, cchsu@mail.npust.edu.tw
Department of Management Information Systems,
National Pingtung University of Science and Technology

Research

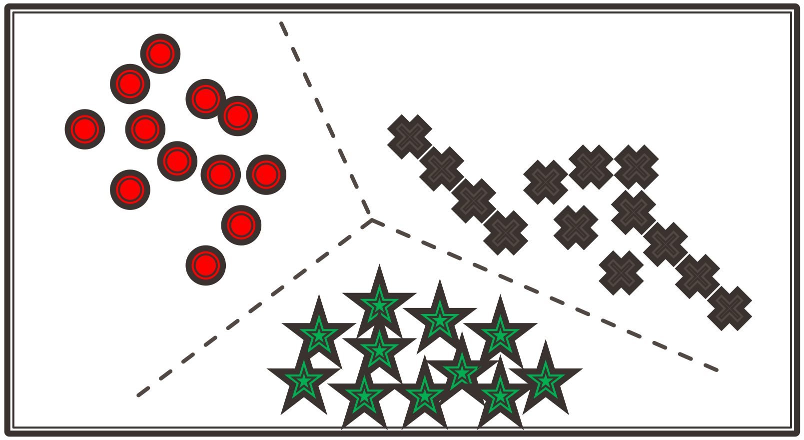
- Research Interests
  - Computer Vision, Image / Video Processing, Machine Learning, and Deep Learning
- Research Topics/Projects
  - Vision-based decision for Autonomous Car (MOST Project)
  - AI for Narrow-band imaging
  - Fake Image Detection (MOST Project)
  - Super-Resolution
  - Social Media Prediction
  - Deep Few-Shot Learning (MOST Project)
  - Other computer vision applications


影視訊內容增強
Image/Video Content Enhancement
Low Res
Result
Real/Fake
偽造影視訊偵測
Fake Image/Video Detection
機器學習 Machine Learning
監督式與非監督式深度學習
Supervised/Unsupervised Deep Learning
電腦視覺
Computer Vision
社群媒體資料分析
Social Media Analysis
醫學訊號分析
Medical Signal Analysis

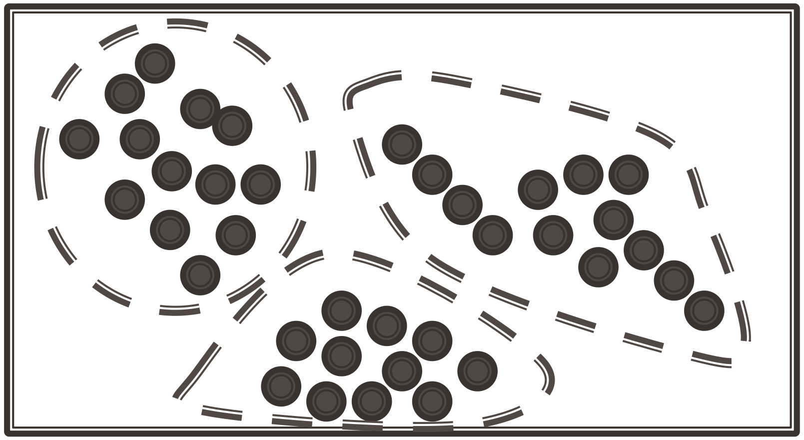
Taxonomy of Machine Learning



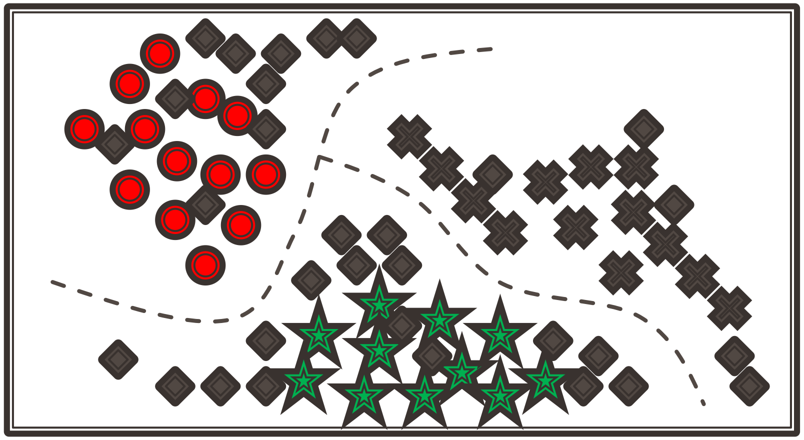
Example to Machine Learning



Supervised learning



Unsupervised learning

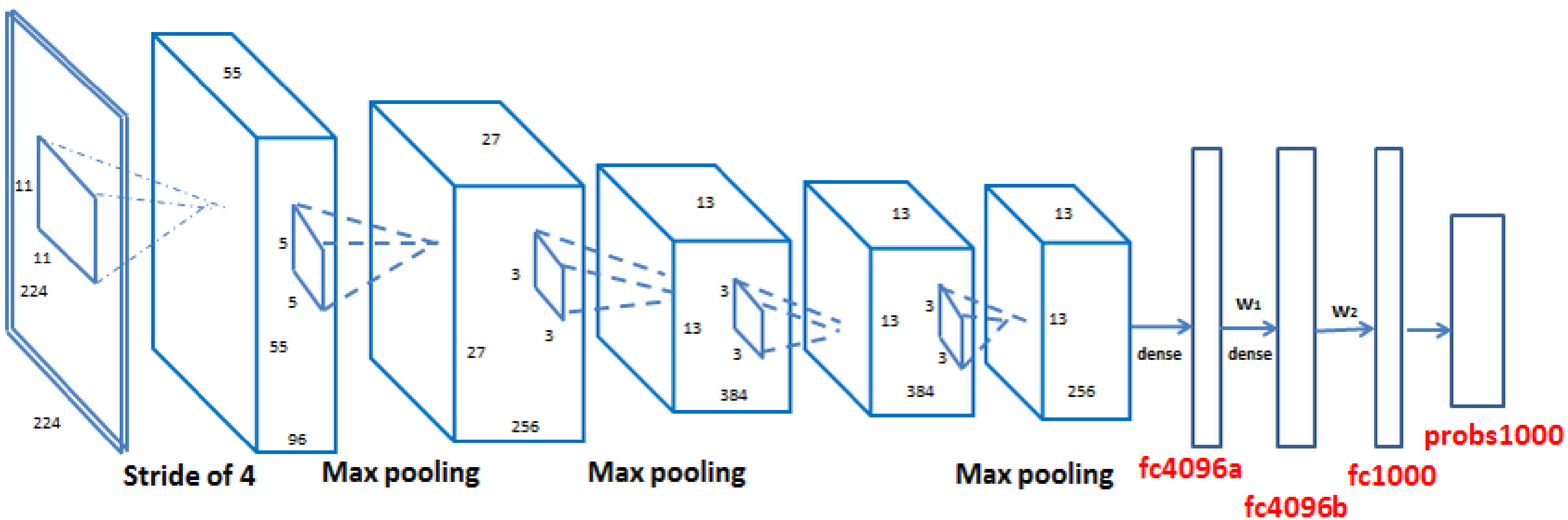


Semi-supervised learning

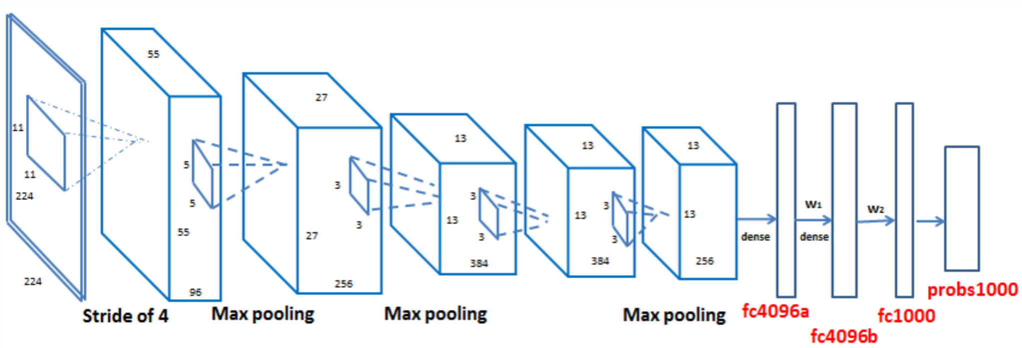
DEEP SUPERVISED LEARNING

AlexNet (2012, Hinton)

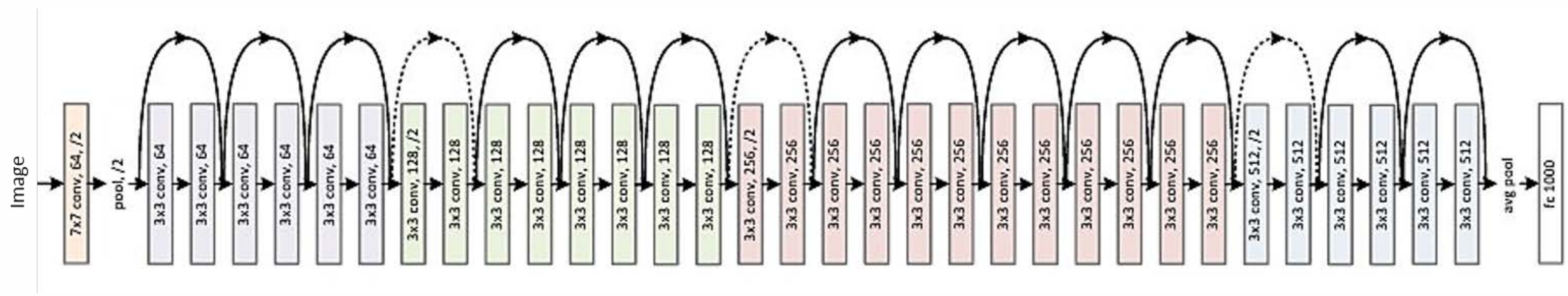
- The winner in ILSVRC Challenge based on Deep Learning in supervised way
- 9-layers
  - 5 convolution and 4 fully-connected layers



Deeper Network

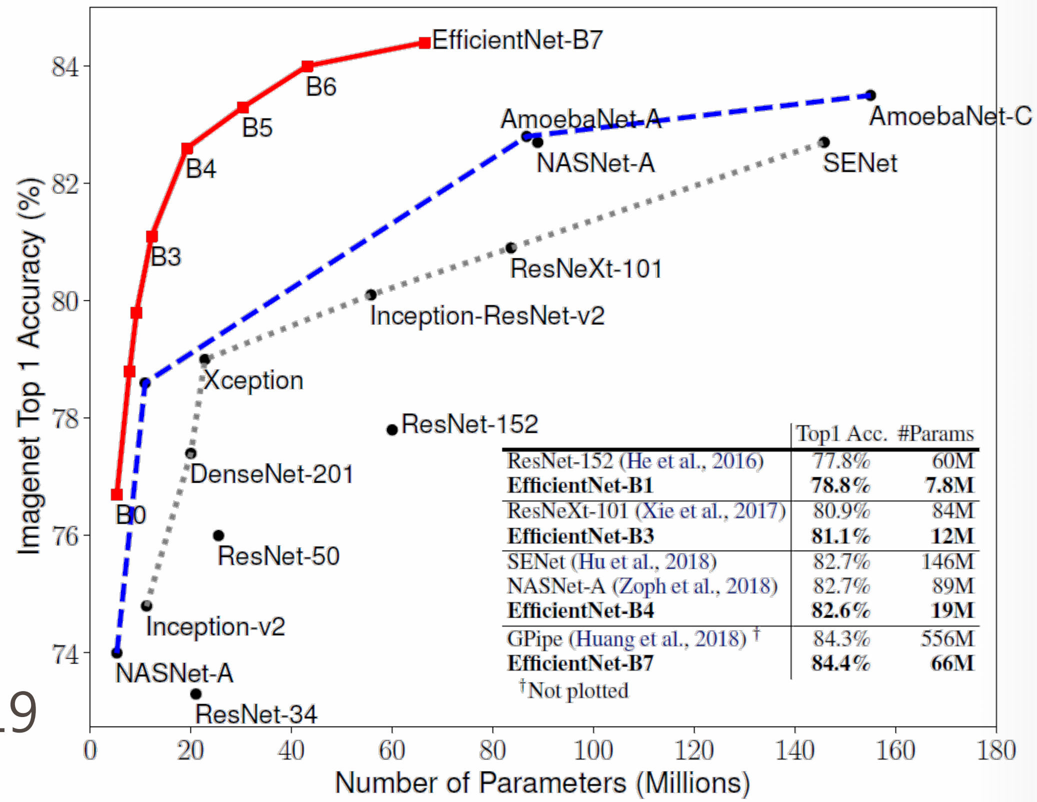


- 2013, AlexNet: 8 layers (9 layers)
- 2016, Residual Net / DenseNet: up to 152 layers…
- 2017, Stochastic depth Net: up to 1000 layers…



State-of-the-Art CNNs

- We called those CNNs trained in supervision way are "backbone " or "baseline" nets
- SOTA now
  - High-performance
    - ResNet
    - Wide-ResNet
    - ResNeXt
    - Inception v3
    - DenseNet
  - High-efficiency
    - MobileNet v3
    - EfficientNet
- Anti-aliasing CNNs ICML19



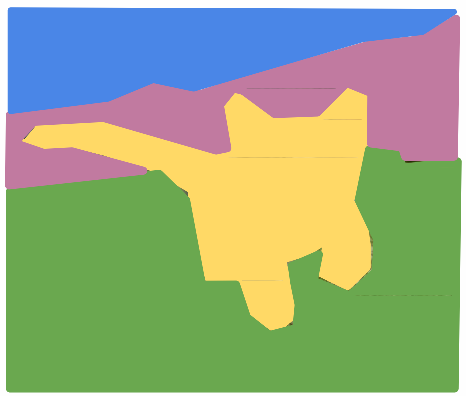
Computer Vision Applications

Classification

CAT

No spatial extent

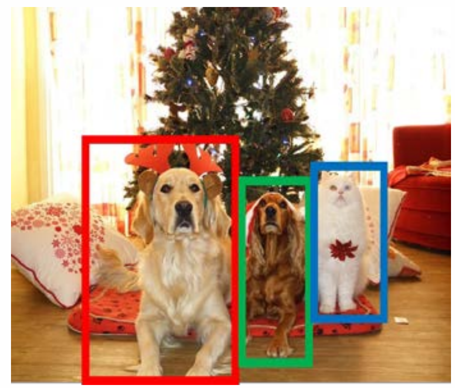
Semantic Segmentation



GRASS, CAT, TREE, SKY

No objects, just pixels

Object Detection



DOG, DOG, CAT

Instance Segmentation

DOG, DOG, CAT

Multiple Object

Slide credit: CS231n, Stanfor

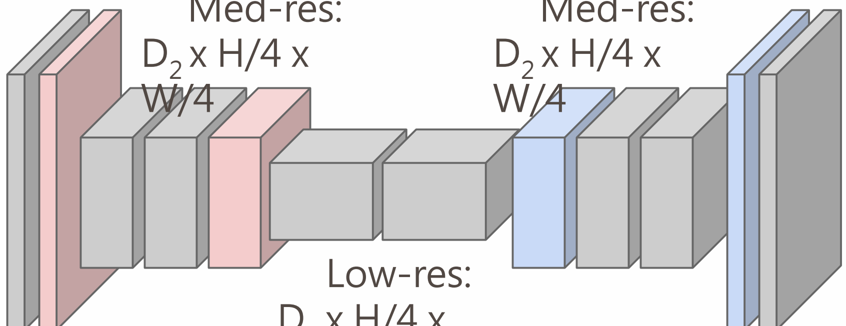
Semantic Segmentation Idea: Fully Convolutional

Design network as a bunch of convolutional layers, with
downsampling and **upsampling** inside the network!

Downsampling: Pooling, strided convolution

Upsampling: ???

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

Med-res:
D_2 x H/4 x W/4

Low-res:
D_3 x H/4 x W/4

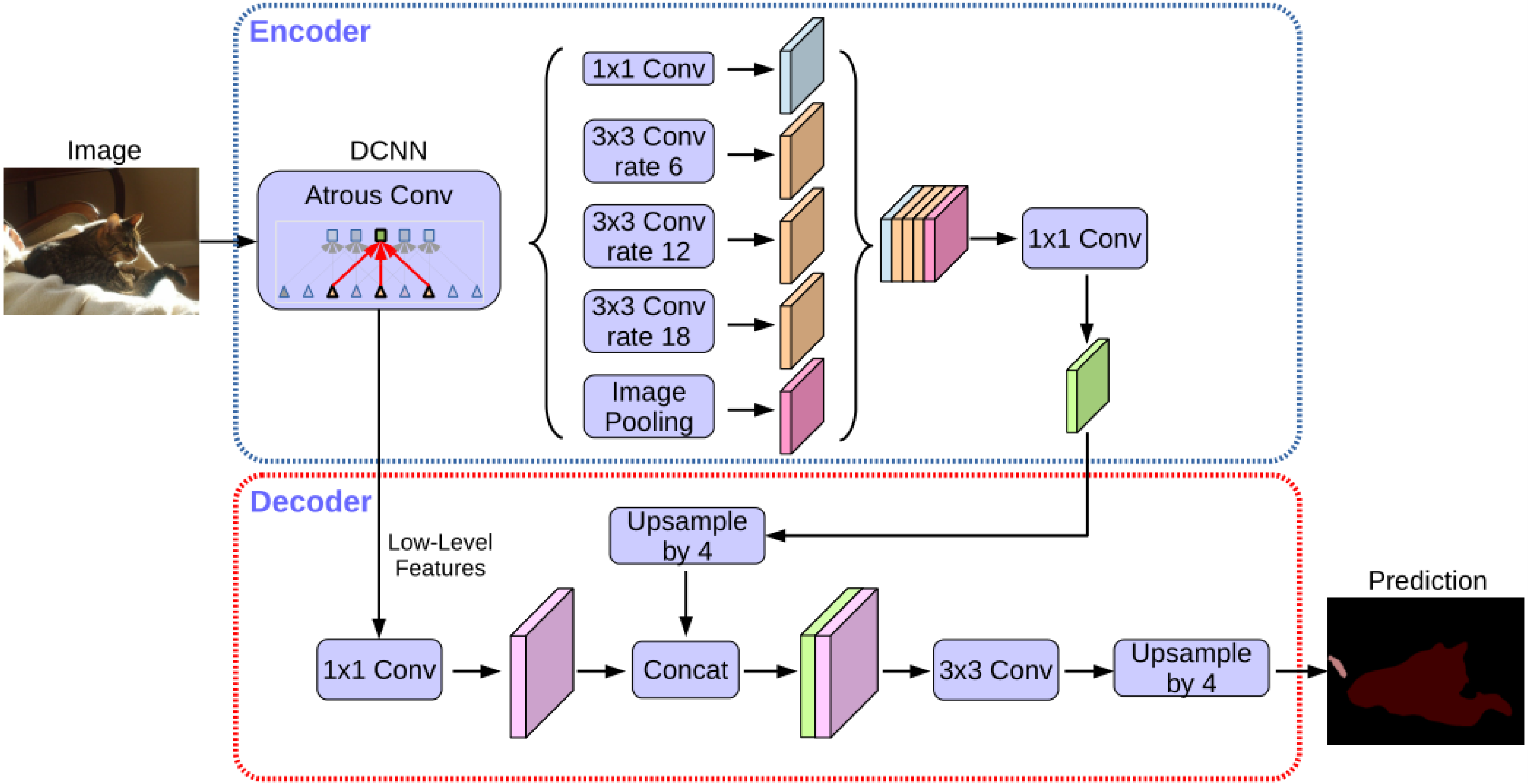
Med-res:
D_2 x H/4 x W/4

High-res:
D_1 x H/2 x W/2

Predictions: H x W

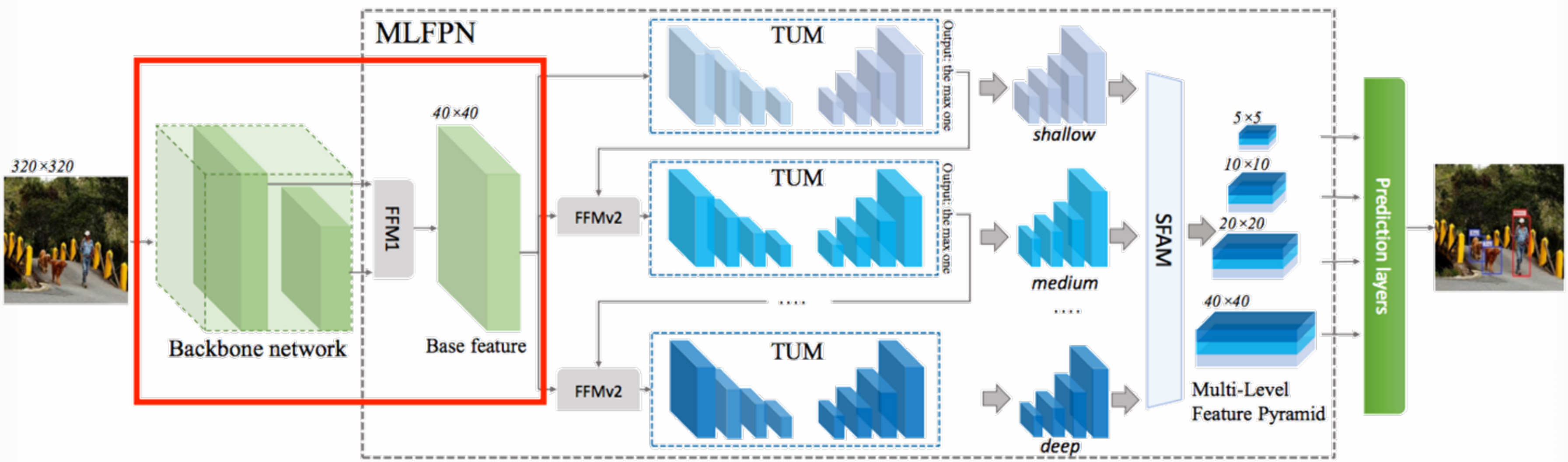
Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

DeepLab V3+ (ECCV18)



M2Det (AAAI' 19)

- FFM: Feature fusion module
- TMU: Thinned U-shape Modules
- SFAM: Scale-wise Feature Aggregation Module



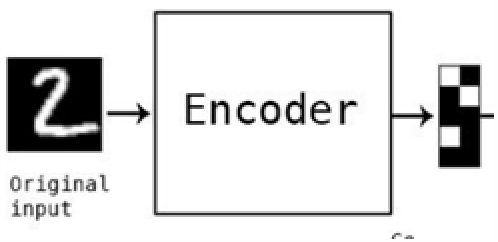
DEEP UNSUPERVISED LEARNING

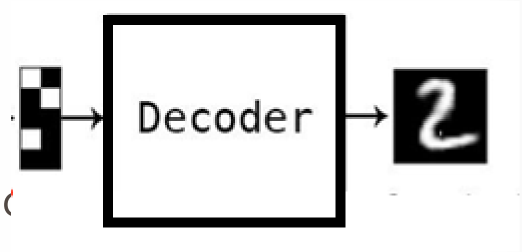
CycleGAN

UNSUPERVISED DEEP LEARNING: GENERATIVE ADVERSARIAL NETWORK

Unsupervised learning vs. Generative model

- Unsupervised learning
  - z=f(x)
- Generative model
  - x=g(z)
- It is …
  - P(z|x) vs. P(x|z)
  - An encoder vs. a decoder
    - Encoder: Feature extraction / Dimensionality reduc
    - Decoder: Generator / Upsampling
    - P(z|x) = P(x, z) / P(x) ➔ P(x) Intractable (ELBO)
    - P(x|z) = P(x, z) / P(z) ➔ P(z) is prior



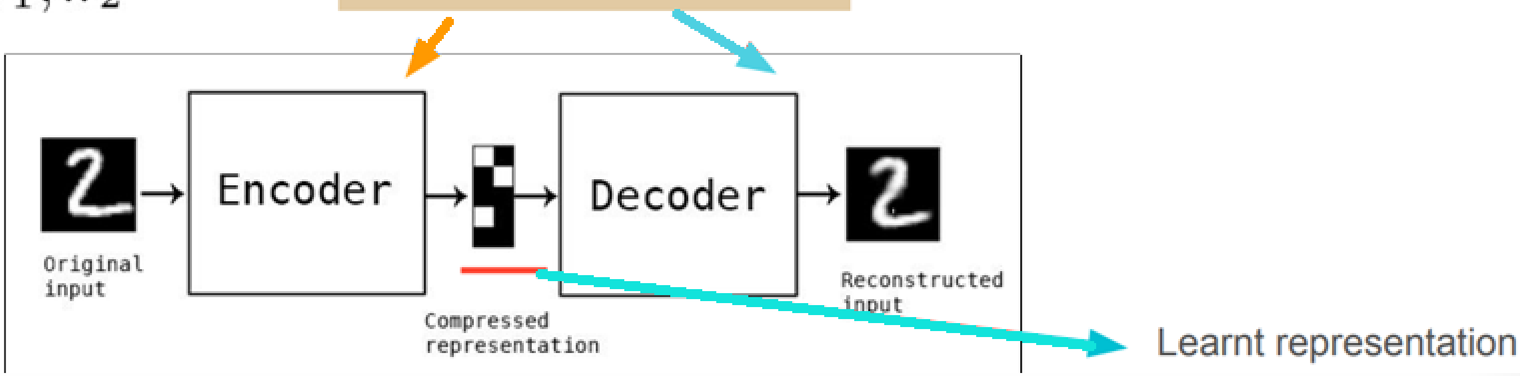


P(x,z) is necessary!!

Unsupervised Deep Learning: AutoEncoder

- With no answer "data"
  - Use "Reconstruction" to learn!!
  - A good representation should keep the information well (reconstruction error)
  - Deep + nonlinearity might help enhance the representation

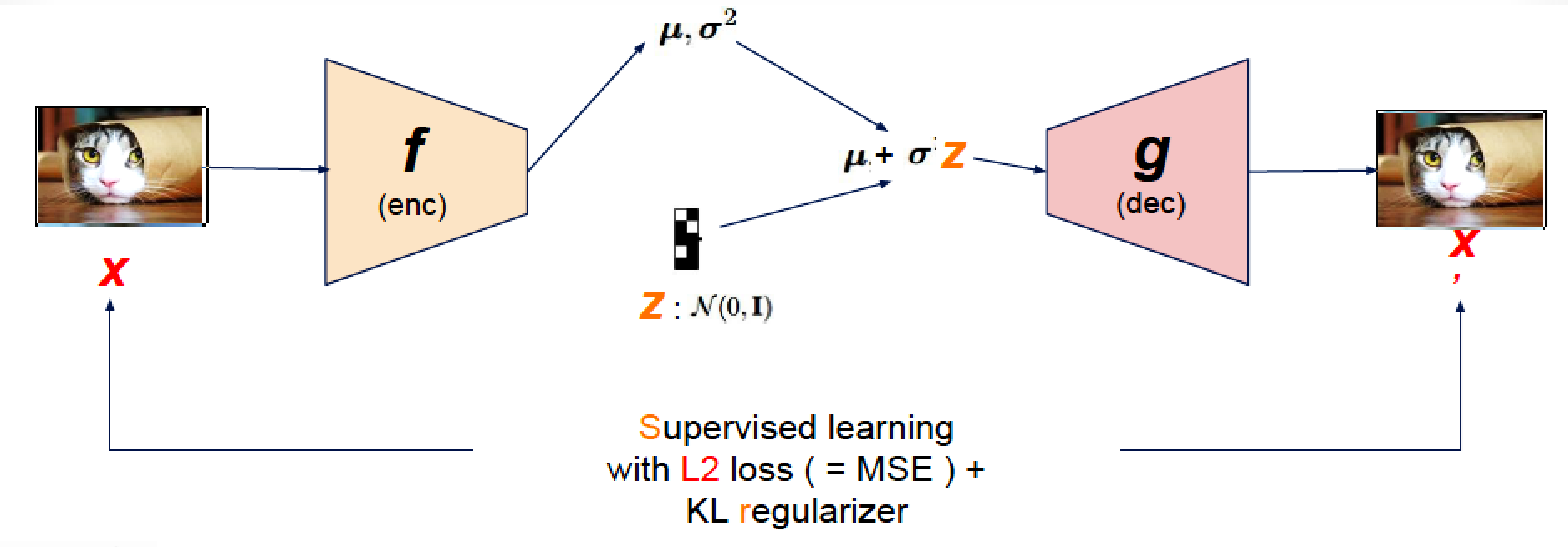
$$\min_{\mathbf{w}_1,\mathbf{w}_2} \|\mathbf{x}_i - g(f(\mathbf{x}_i;\mathbf{w}_1);\mathbf{w}_2)\|_2^2$$



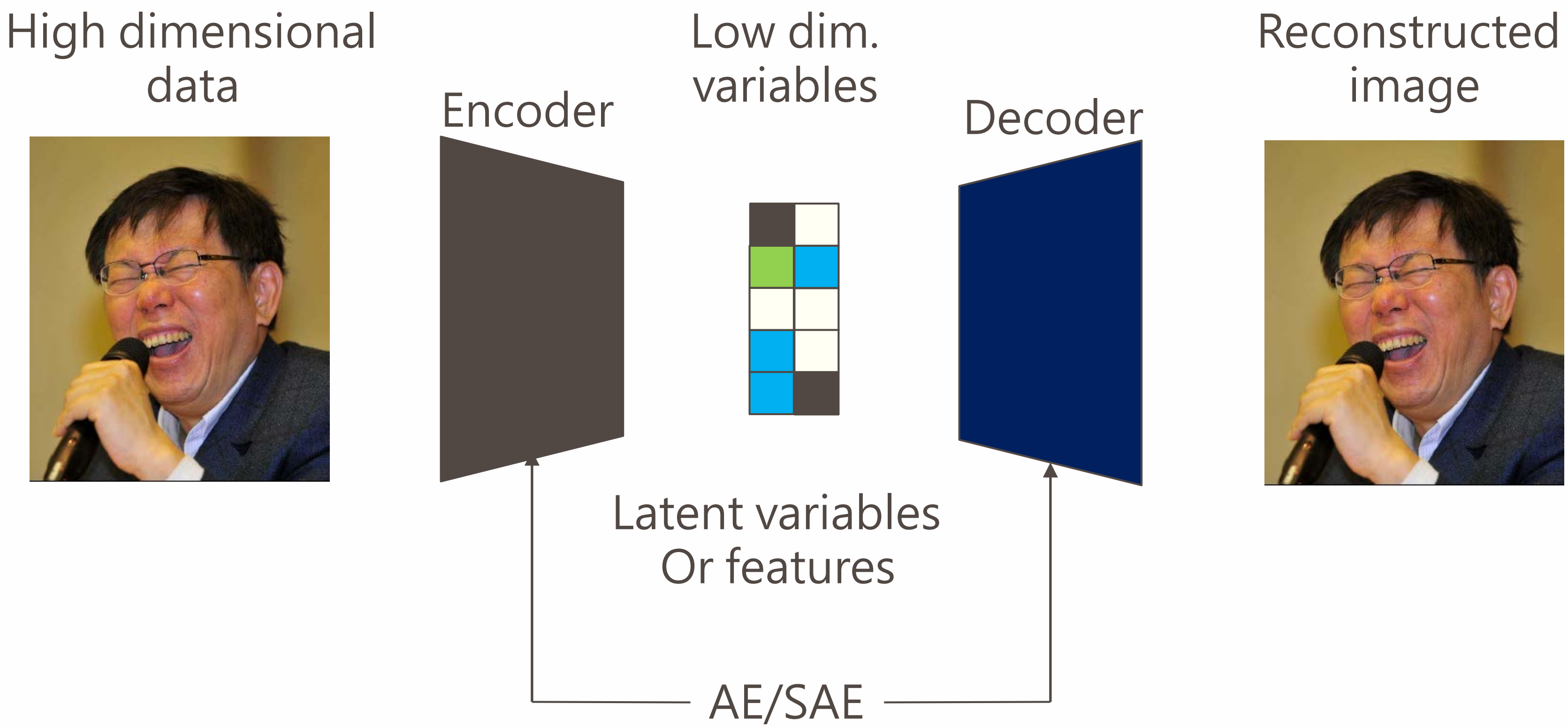
Credit: LeCun

Deep Version of AutoEncoder

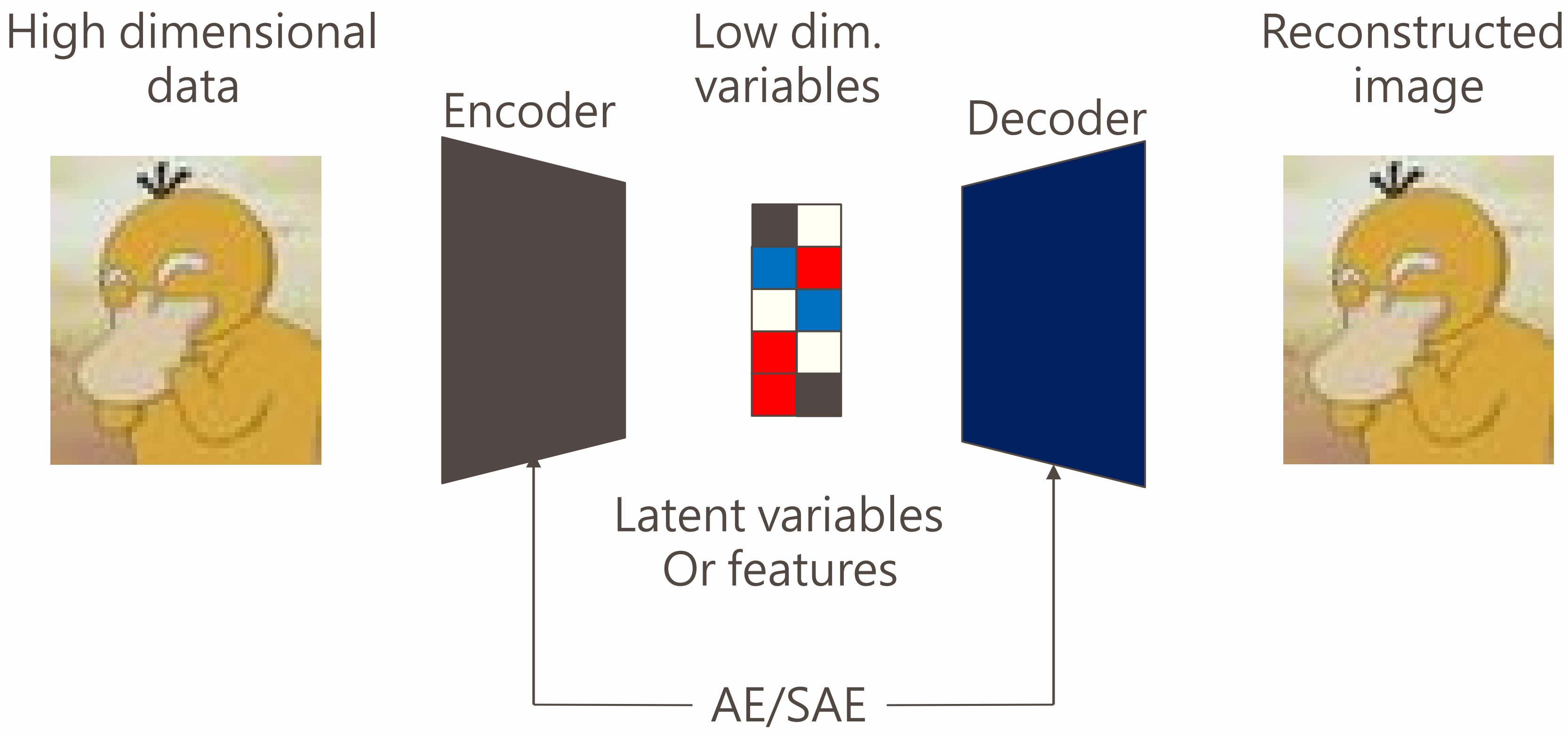
- Stacked autoencoder (SAE)
- Similar to AE but more deeper



What Exactly AE is?



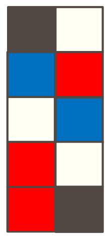
What Exactly AE is? (cont.)



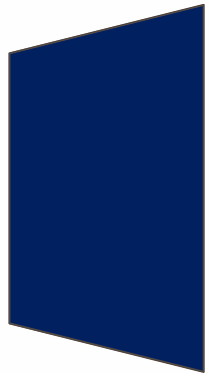
What Exactly AE is? (cont.)

For example
0.666 = 可達鴨

Low dim.
variables 1



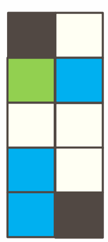
Decoder



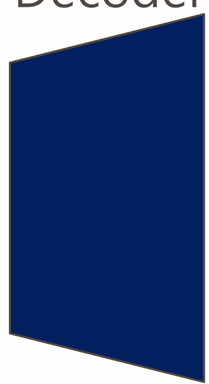
Reconstructed
image 1

For example
0.747 = 柯P

Low dim.
Variables 2



Decoder

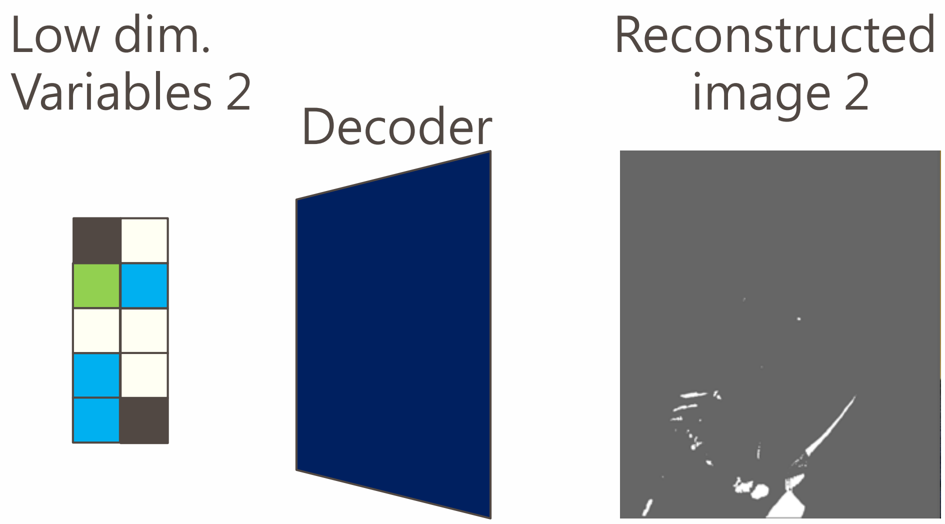


Reconstructed
image 2

Problem in SAE/AE

- One feature corresponds to one reconstructed image!
  - Feature is generated from Encoder….
    - Such AE/SAE cannot be used to generate arbitrary images

For example
0.74 = 柯P
0.76 = ?

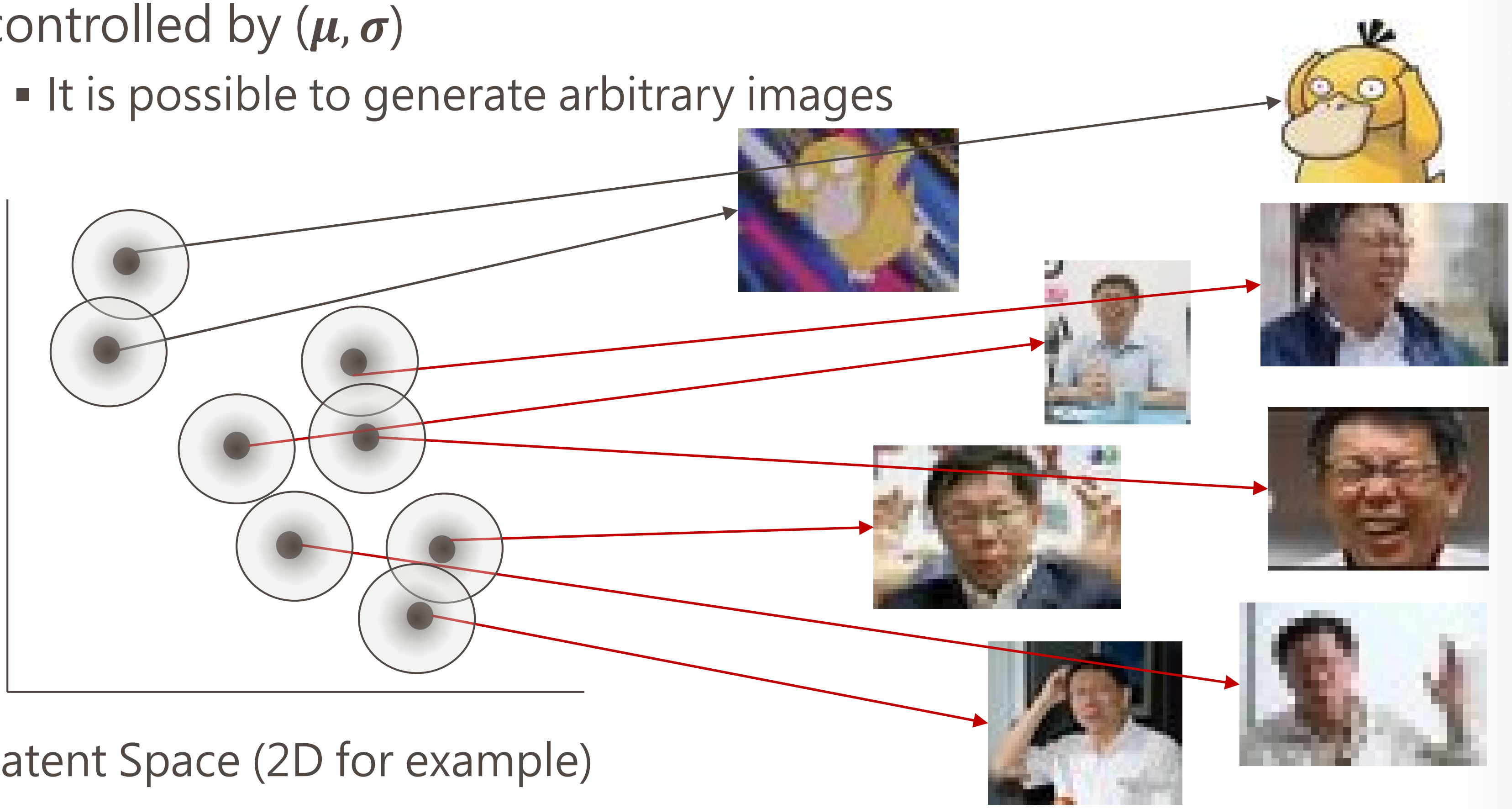


Improved AutoEncoder

- Variational autoencoder - VAE
- Kingma et al, "Auto-Encoding Variational Bayes", 2013.
  - Generative Model + Stacked Autoencoder
  - Based on Variational approximation
- From AE to VAE
  - Since the feature (latent variable) is not continuous
    - Explicit feature is required for generating an image
  - MODELING feature instead!!

From AE to VAE

- Modeling: Assume the feature is sampled from Gaussian controlled by $(\boldsymbol{\mu}, \boldsymbol{\sigma})$
  - It is possible to generate arbitrary images



Latent Space (2D for example)

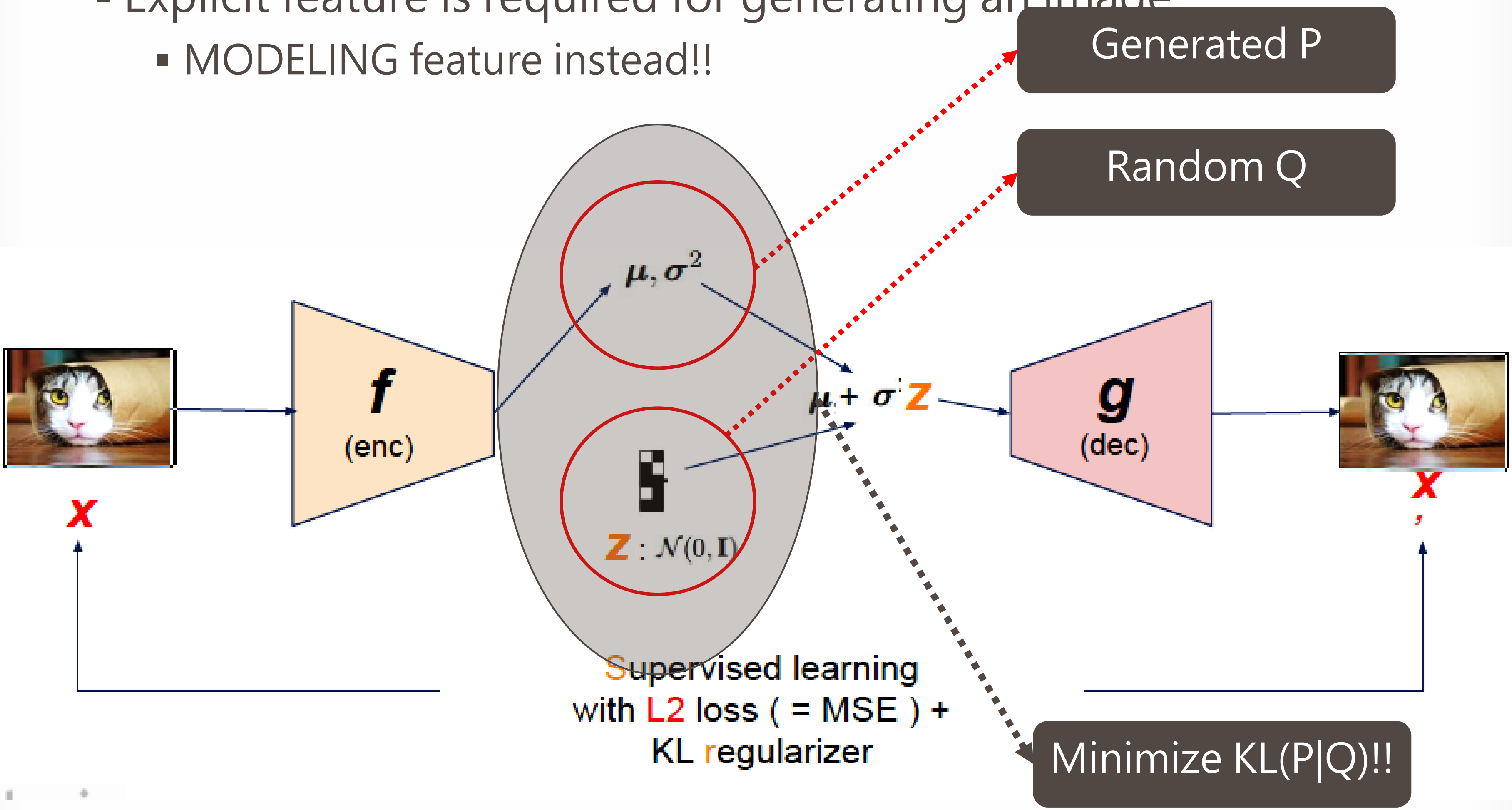
Reconstructed images

From AE to VAE

- In this way, loss function can be defined as
  - $L_{data} = ||X - \bar{X}||_2^2$, where $\bar{X}$ is the reconstructed image
  - $L_{latent} = \mathrm{KL(P|Q)}$ ➔ KL(Latent variables, Gaussian)
  - $L = L_{data} + L_{latent}$
- Difficult to optimize L
  - The distribution of latent variables is unknown & uncontrollable.
- Solution:
  - Force latent variable to be a parameters of a specified distribution: Encoder ➔ $(\boldsymbol{\mu}, \boldsymbol{\sigma})$

From AE to VAE

- Explicit feature is required for generating an image
  - MODELING feature instead!!

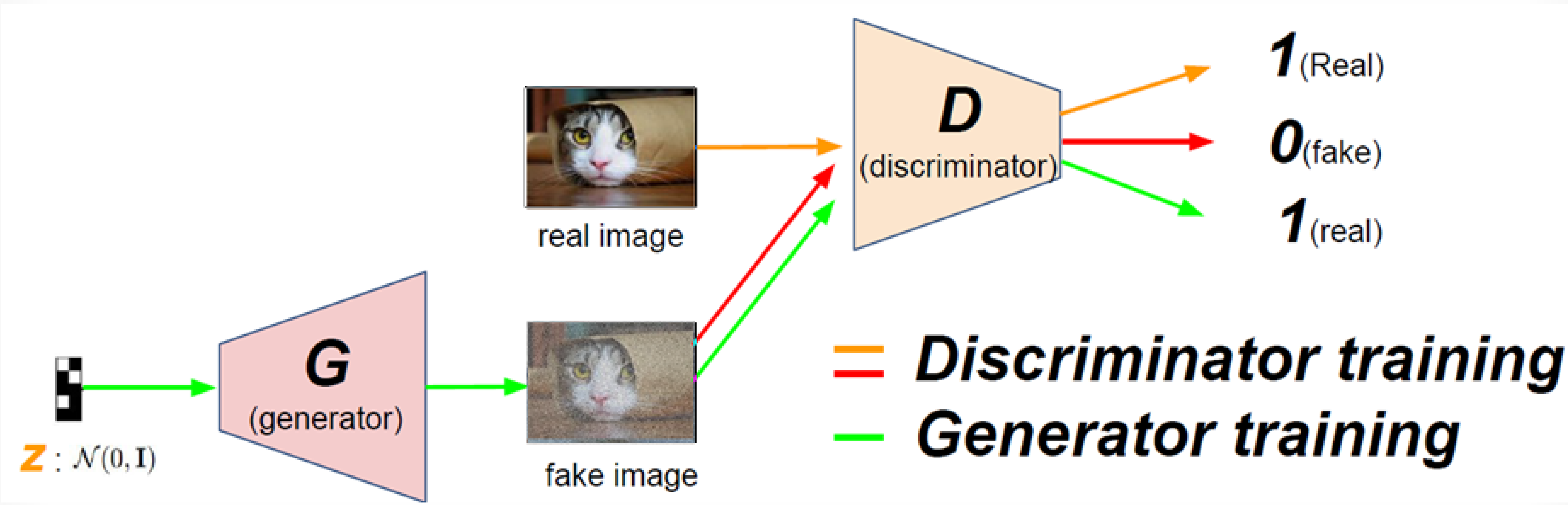


Variational AE (VAE)

- Minimize KL(P|Q)!!
  - ➔ Variational inference!!
- Recall that
  - $P(z|x) = P(x, z) / P(x)$ ➔ $P(x)$ Intractable (ELBO)
  - Approximation solution
    - Use $q(z|\theta)$ to approximate $P(z|x)$
    - Variational inference!
- Shortcoming
  - Blurred images will be generated (no guarantee its quality)

Unsupervised Deep Learning

- How to generate an image with good quality?
  - Generative adversarial network (GAN)



Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.

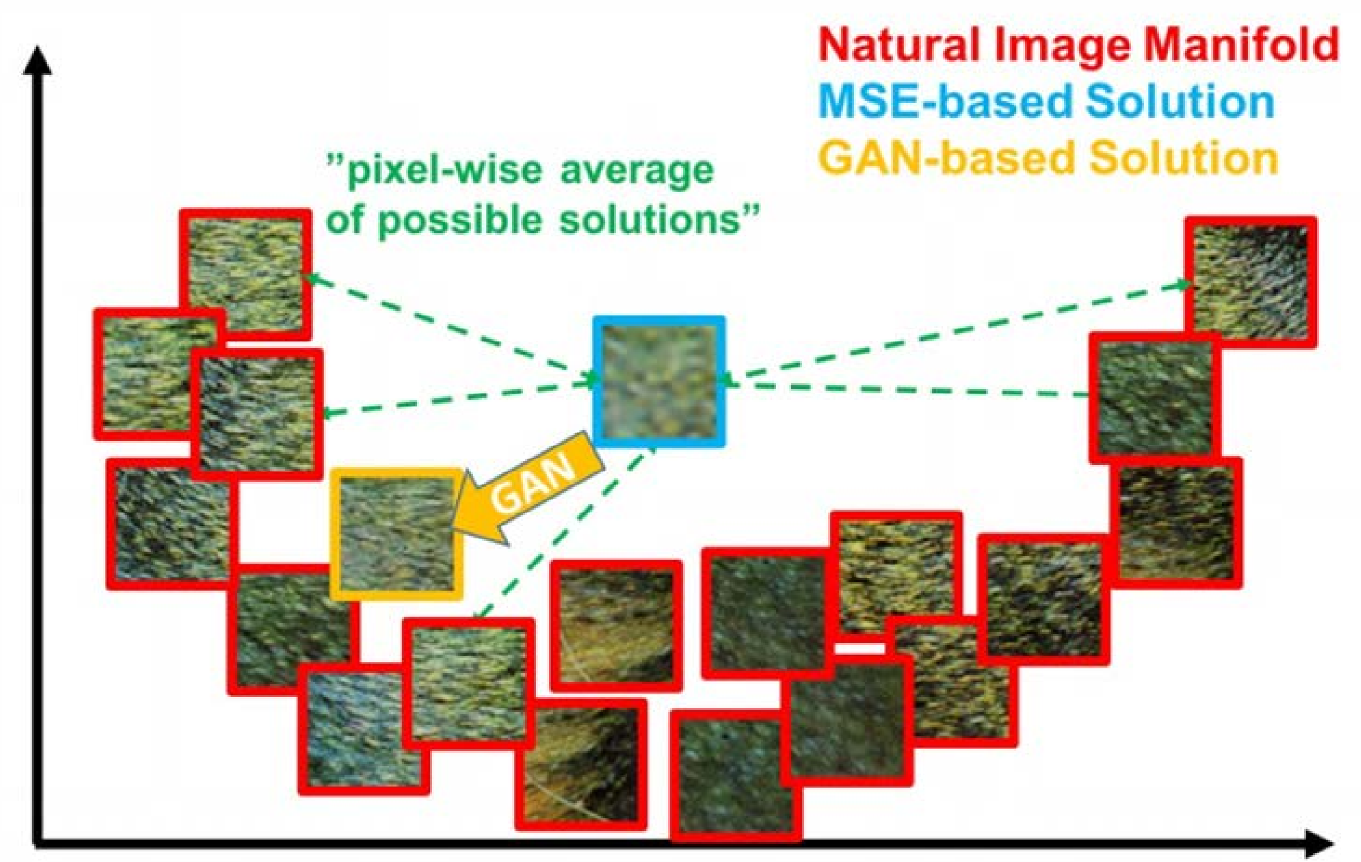
Why Generative Models?

- Excellent test of our ability to use high-dimensional, complicated probability distributions
- Simulate possible futures for planning or simulated RL
- Missing data
  - Semi-supervised learning
- Multi-modal outputs
- Realistic generation tasks

(Goodfellow 2016)

Generating an Image using GAN

- Learn and predict P(x|z)



Ledig et al. CVPR 2017

Training Procedure

- Use SGD-like algorithm of choice (Adam) on two mini-batches simultaneously:
  - A mini-batch of training examples
  - A mini-batch of generated samples
- Optional: run k steps of one player for every step of the other player.

(Goodfellow 2016)

Problems in GANs

- No guarantee to equilibrium
  - Mode collapsing
    - All smoothing results
  - Oscillation
    - May never converge
  - No indicator when to finish
- All generative models
  - Evaluation metrics (predefined)
  - Robust but difficult to train
  - Diversity testing is required

GAN’ s Ways

- Theatrical analysis of the nature of the GANs
  - WGAN
    - Wasserstein GAN (Replace KL with Wasserstein)
    - Solved the issue when there is no overlapping between distributions of generated & ground truth samples
  - BEGAN
  - WGAN-GP
  - RAGAN
  - …etc
- Applications
  - Based on a state-of-the-art GAN and fine-tune it.

Improved GAN: DCGAN

- Radford et al, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 2015
  - Tricks for gradient flow
    - Max pooling → Strided convolution or average pooling
    - Use LeakyReLU instead of ReLU
  - Other tricks
    - Use batch normal both generator and discriminator
    - Use Adam optimizer (lr = 0.0002, a = 0.9, b=0.5)

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).

Improved Versions of GAN

- There are more than 100 improved GANs/Applications since 2014!!
  - A hot topic in deep learning

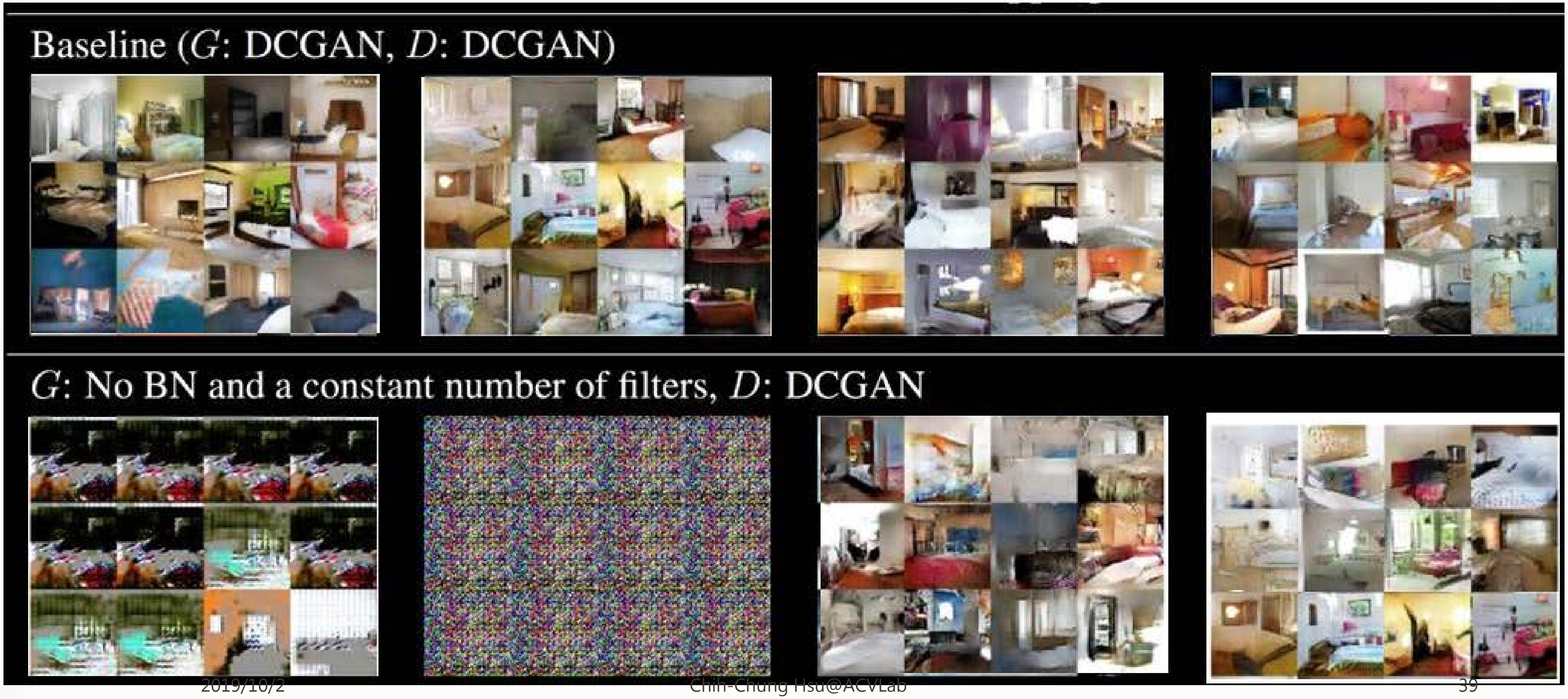
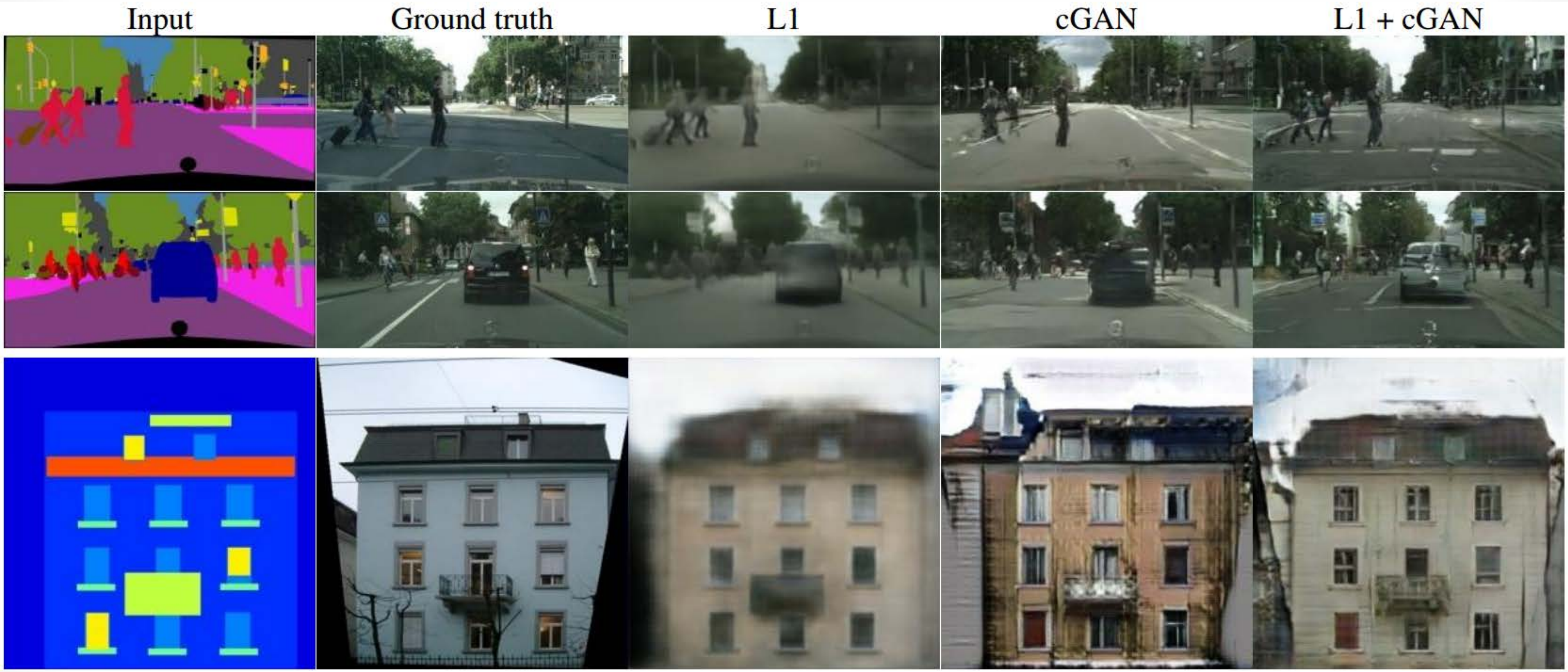


Image Super-Resolution

bicubic (21.59dB/0.6423) | SRResNet (23.53dB/0.7832) | SRGAN (21.15dB/0.6868) | original

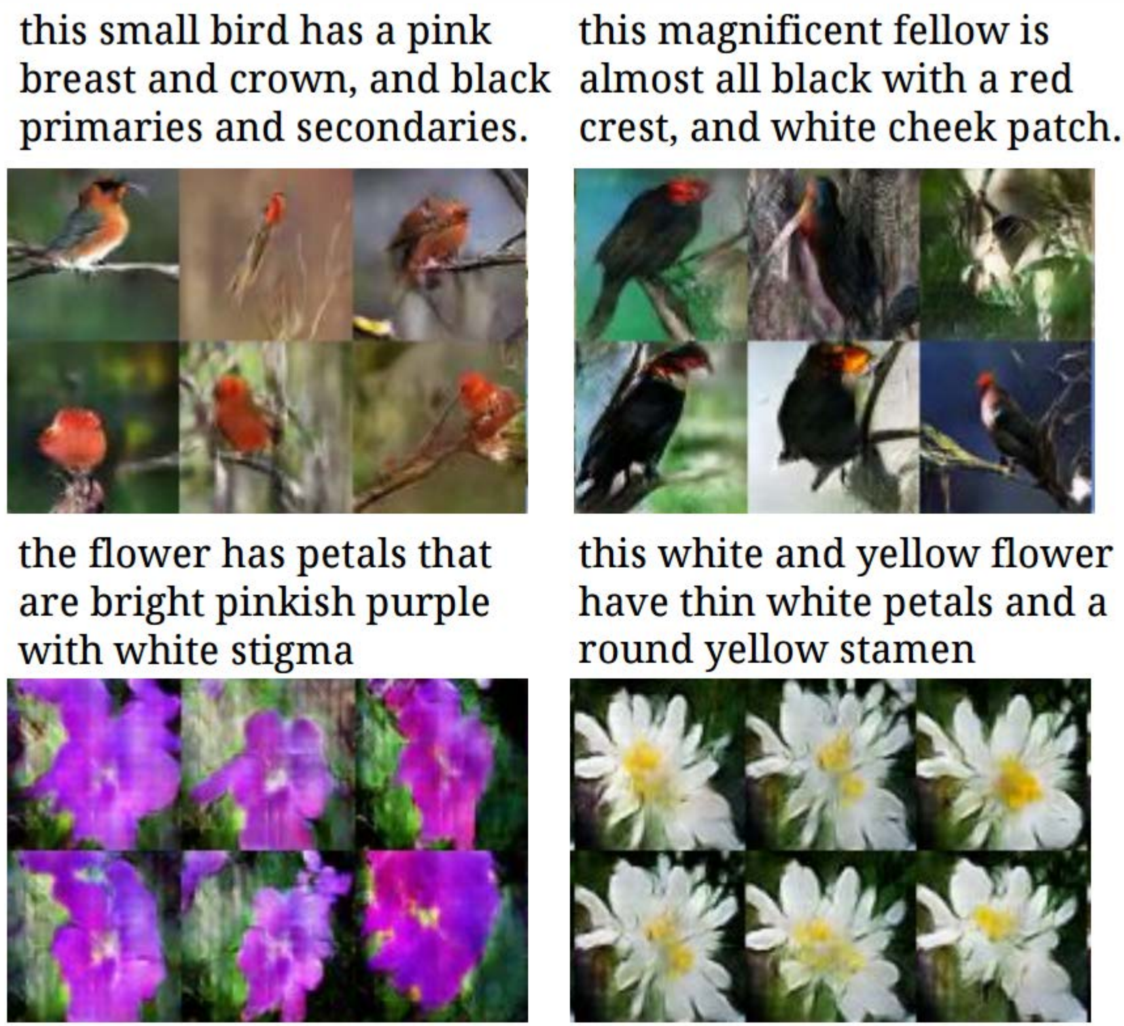
Ledig et al. CVPR 2017

Label2Image



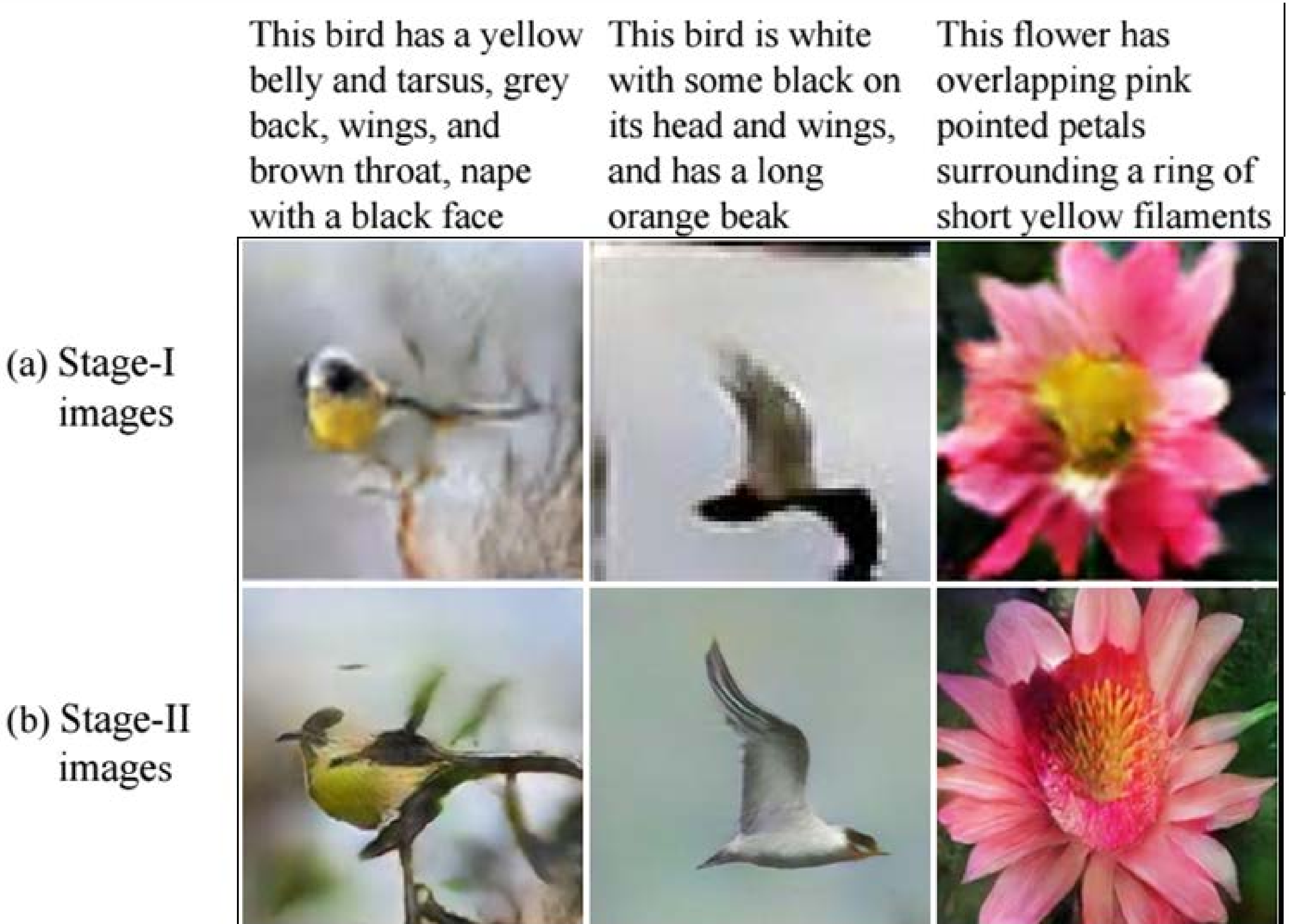
Isola et al. CVPR 2017

Text2Image



Reed et al. ICML 2016

StackGAN



Zhang et al. 2016

GAN-based Applications

- What application should GAN be used
  - Any task related to image synthesis (合成影像任務包含)
    - Image super-resolution
      - Discriminator can be used to judge its fidelity and resolution
    - Image translation
      - Discriminator can be used to identify its quality
    - Image segmentation
      - Discriminator can be used to tune generated segmentation map
    - Data argumentation
      - Discriminator can be used to check the fidelity of the simulated image
    - etc...

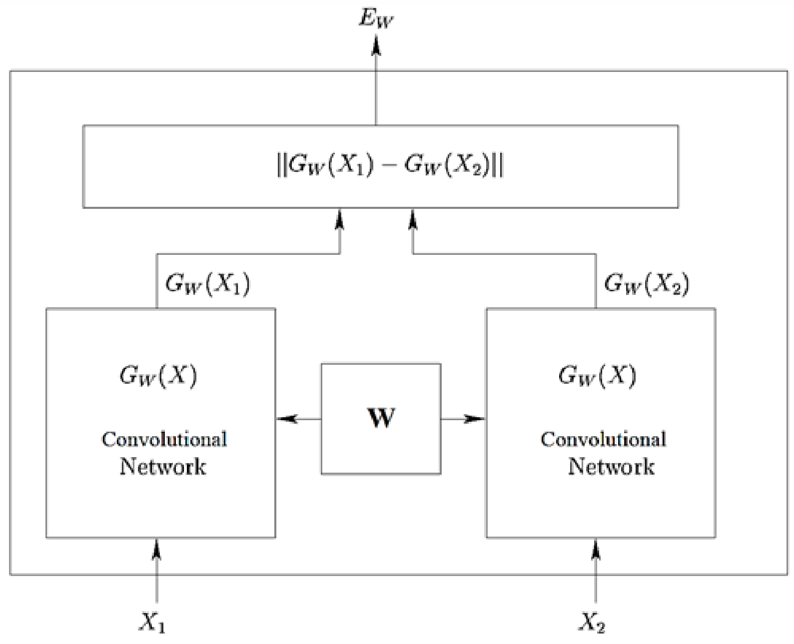
SEMI-SUPERVISED LEARNING

Deep Learning Semi-Supervised Learning

- Take some advantages form supervised learning
  - Problem:
    - How to interpret?
- Siamese Network
  - Pairwise Learning
  - Make data "Pairwise"
    - Same identity of a pair: y=1
    - Different identities of a pair: y=0
  - Usually used in "face verification" or person re-identifications

Two Phase Learning Tasks

- Siamese Network Architecture
  - Learning to capture the common features indicating unrealistic details of the fake image
  - Adopt a pairwise learning framework!!



FAKE IMAGE DETECTION: ANTI-GAN

ICIP 2019.
Best Student Paper Award

Detecting the Fake Images

- The related techniques to detect the fake images
  - Intrinsic feature based approach
    - Image forensic
    - Image forgery detection
  - Extrinsic feature based approach: Watermarking
- Intrinsic feature based approach is relatively practical
  - However, such generated images didn’t have such intrinsic features
    - Image is generated directly from noise
      - No source

Problems Caused by Fake Images

- Improper use of such fake multimedia will lead to a serious consequence

  - Police purpose, on purpose misleading, or business use

An Example of Traditional Image Forensic

(a) Original Image 1 (b) Texture replaced

An Example of Traditional Image Forensic

(a) Fake Image 1 (b) Fake Image 2

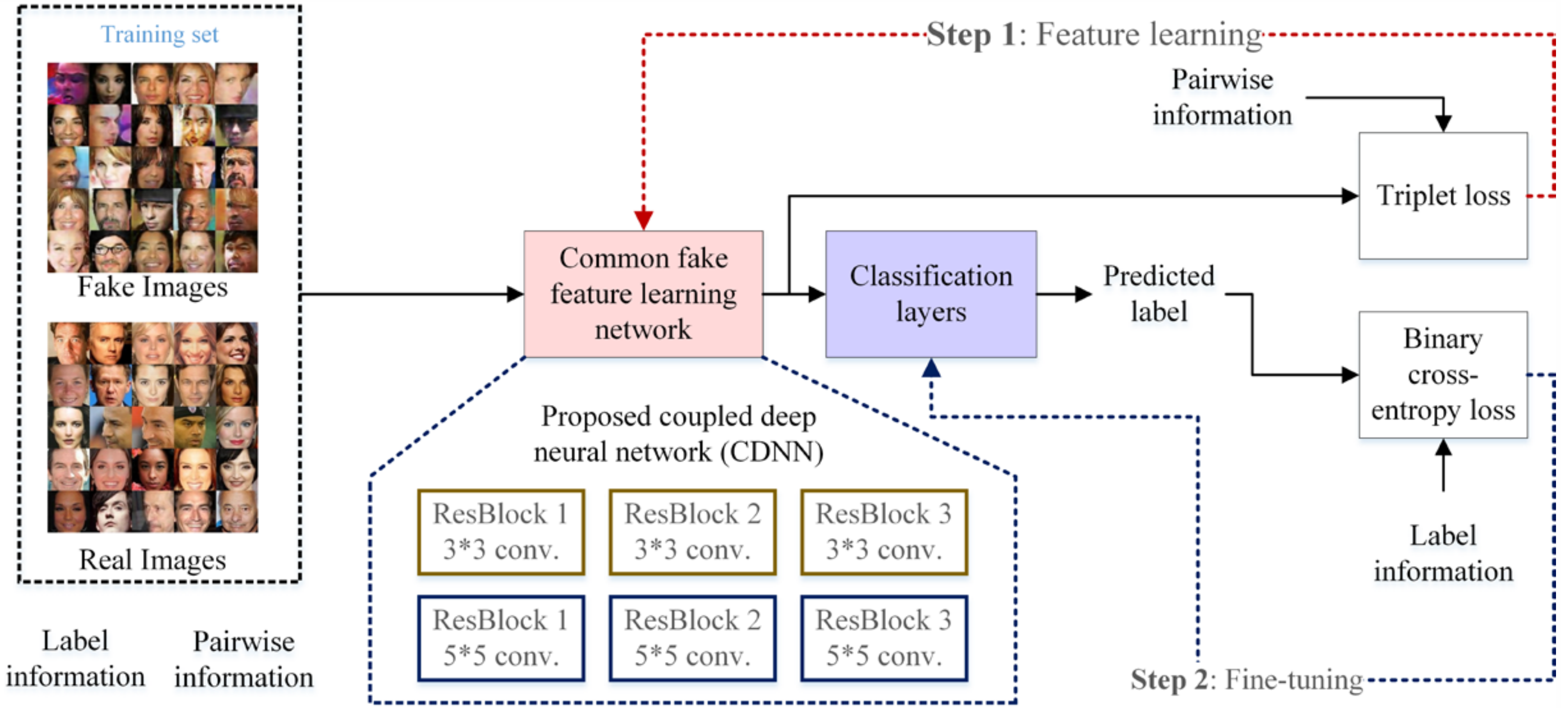
How to effectively detect such fake images remains big problem!!

We propose a novel framework to effectively address this issue!!

Fake Image Detection

- Directly learning a classifier in supervised learning manner may be ineffective.
  - It is hard to collect every GANs to learn
  - The generator can be improved
    - The fake image detector should be improved as well
    - It is too impractical
- Instead of supervised learning, we adopt pairwise learning to effectively capture the common features across different GANs
  - Two-phase learning tasks
  - Contrastive loss
  - Called deep forgery detector (DeepFD)

The Proposed Framework



Contrastive Loss

- Minimizing the feature distance between the paired inputs if they are all fake or real.

$$\mathrm{E}_W(\mathbf{x}_1, \mathbf{x}_2) = \|\mathrm{D}_1(\mathbf{x}_1) - \mathrm{D}_1(\mathbf{x}_2)\|,$$

  - Where D indicates feature representation of JDF of an image
- The contrastive loss function of the proposed JDF will be:

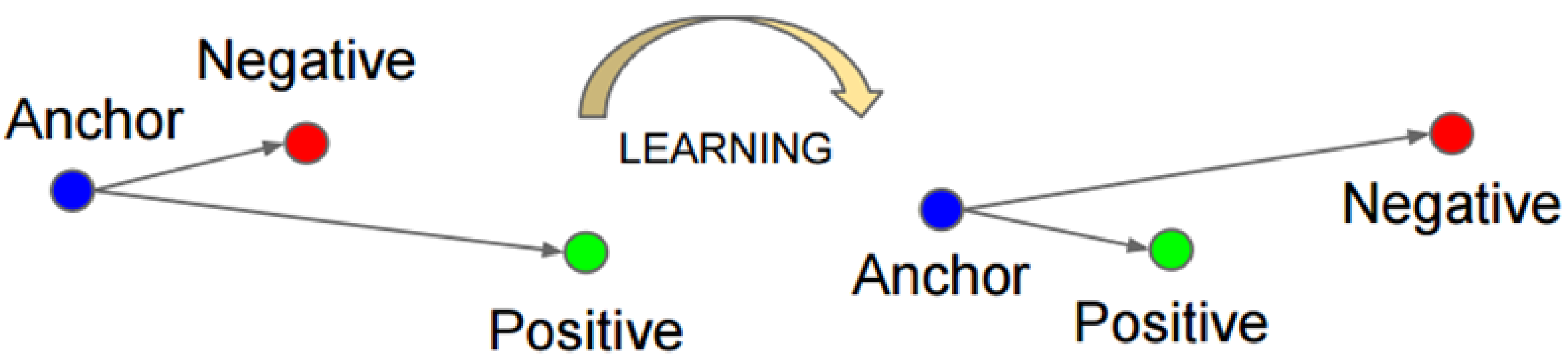
$$L(W, (P, \mathbf{x}_1, \mathbf{x}_2)) = \frac{1}{2}\left(p_{ij}(\mathrm{E}_W)^2 + (1 - p_{ij})(\max(0, m - \mathrm{E}_W)^2\right),$$

  - where p_{ij} indicates genuine ($p_{ij} = 1$) and impostor ($p_{ij} = 0$) pairs

Triplet Loss

- Calculate the distance between anchor and positive/negative samples

$$\sum_{i}^{N_r}\left[\|\mathrm{D}_1(\mathbf{x}_a)-\mathrm{D}_1(\mathbf{x}_p)\|_2^2-\|\mathrm{D}_1(\mathbf{x}_a)-\mathrm{D}_1(\mathbf{x}_n)\|_2^2+a\right]_+$$

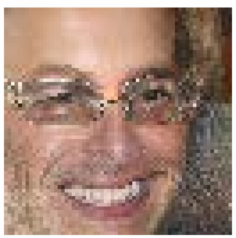


Learning Trick

- Hard mining is the most important
  - Similar to object detection nets
- Hard positive
  - Same person but different poses in two images
- Hard Negative
  - Different person but looks similar to each other in two images

Common Fake Feature Learning

GAN-1



GAN-2

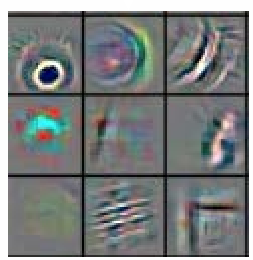
CFF Net

CFF Net

128-dim Feature

128-dim Feature

Minimizing distance



Learning to capture the features of fake images

Common Fake Feature Learning

Fake 1

Real 2

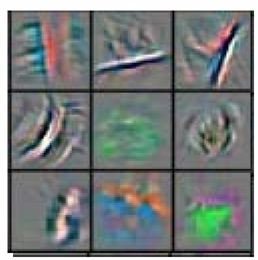
CFF Net

CFF Net

128-dim Feature

128-dim Feature

Maximizing distance



Learning to capture the features of real images

Classification Network Learning

- We can adopt existing state-of-the-art classifier to detect fake images
  - SVM, Random forest, or Bayer classifier
  - However, we don’t know what features is useful for fake image detection
- Instead, we design a network in network as the classifier
  - Learning in supervised way
  - Based on the pretrained network (CFF) learned by the proposed pairwise learning

Classification Network Learning

- The loss function of the classifier can be defined as a cross-entropy loss:

$$L_C(\mathbf{x}_i, \mathrm{y}_i) = -\sum_{i}^{N_T} \left(\mathrm{D}_2\left(\mathrm{D}_1(\mathbf{x}_i)\right) \log \ \mathrm{y}_i\right).$$

  - where N_T is the number of the training set and y_i is the label indicating 0 (fake) or 1 (real)

Network Architecture

Layers	Jointly Discriminative Feature	Discriminator
1	Conv.layer, kernel=7*7, stride=4, channel=96	Conv. layer, kernel=3*3, channel = 2
2	Residual block *2, channel=96	Global average pooling
3	Residual block *2, channel=128	Fully connected layer, neurons=2 Softmax layer
4	Residual block *2, channel=256	
5	Fully connected layer, neurons=128 Softmax layer	

Experimental Results

- Experimental settings
  - We collect 5 state-of-the-art GANs to generate fake images pool
    - 1) DCGAN (Deep convolutional GAN) [2]
    - 2) WGAP (Wasserstein GAN) [3]
    - 3) WGAN-GP (WGAN with Gradient Penalty) [4]
    - 4) LSGAN (Least Squares GAN) [5]
    - 5) PGGAN [1]
  - Each GAN generates 200,000 fake images with sized of 64x64

Experimental Results

- Experimental settings
  - We randomly pick up 202,599 fake images from the fake images pool
  - Total number of training images: 400,198
  - Total number of test images: 5,000
  - Parameter m in contrastive is 0.5
  - JDF learning in the first two epochs
  - Discriminator learning in the following epochs
- We exclude the fake images generated from one of the collected GANs to verify the proposed method is generalized

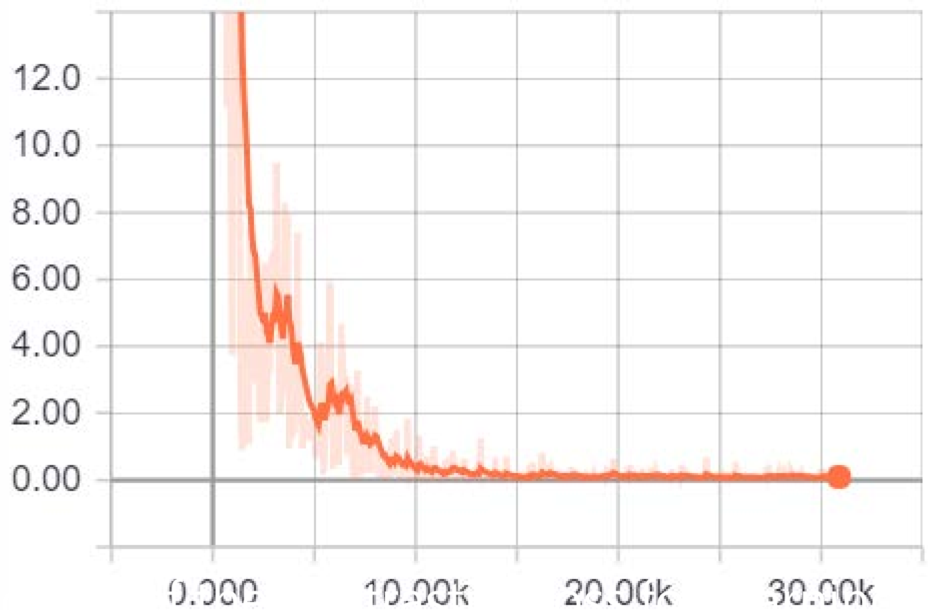
Objective Quality Comparison

The performance comparison between the proposed method and other methods

Table 1. The subjective performance comparison between the proposed fake face detector and other methods.

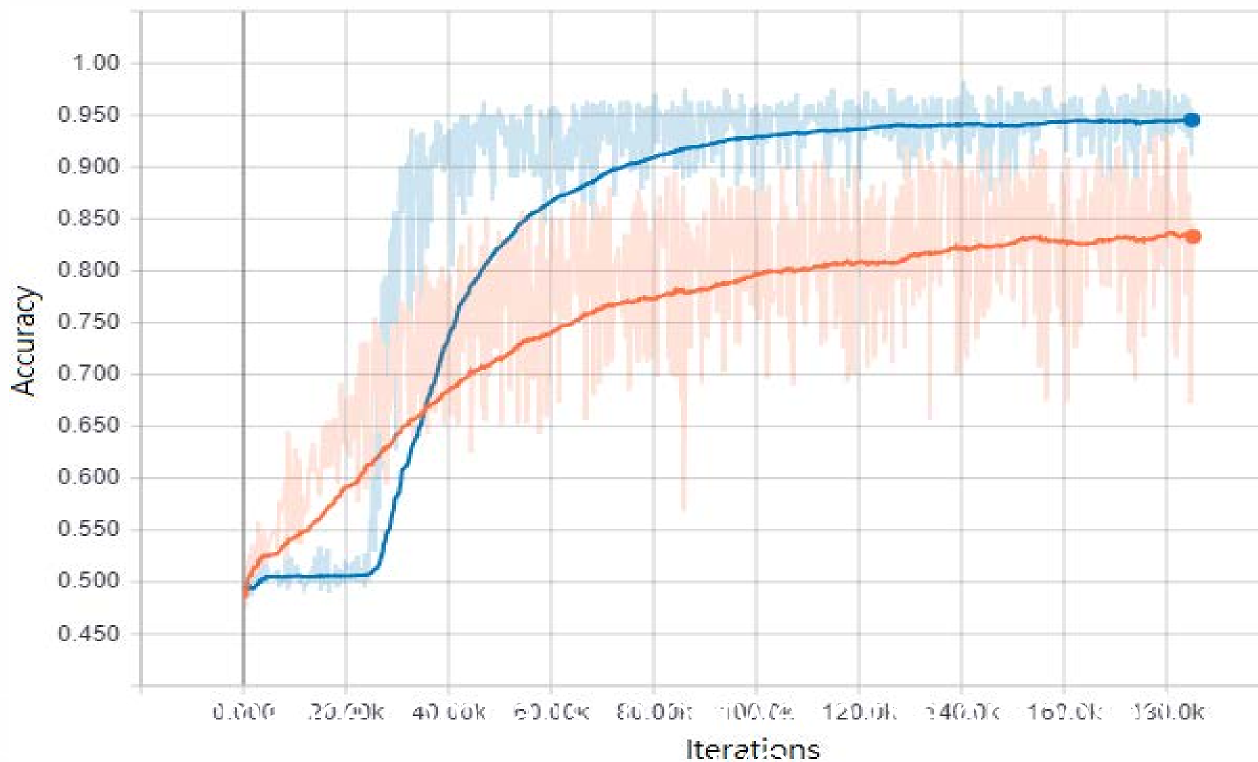
Method/Test target	LSGAN		DCGAN		WGAN		WGAN-GP		PGGAN	
	precision	recall	precision	recall	precision	recall	precision	recall	precision	recall
Method in [5]	0.205	0.580	0.253	0.774	0.235	0.673	0.242	0.604	0.222	0.862
Method in [7]	0.819	0.528	0.848	0.790	0.817	0.822	0.816	0.679	0.798	0.788
Method in [8]	0.833	0.725	0.812	0.833	0.840	0.809	0.826	0.733	0.824	0.838
Method in [15]	0.947	0.922	0.871	0.844	0.838	0.847	0.818	0.835	0.926	0.918
Baseline-I	0.921	0.915	0.887	0.831	0.860	0.855	0.822	0.837	0.919	0.898
Baseline-II	0.939	0.929	0.878	0.851	0.840	0.863	0.845	0.844	0.922	0.928
Baseline-III	0.845	0.785	0.796	0.816	0.833	0.799	0.819	0.805	0.835	0.854
The proposed	**0.981**	**0.956**	**0.986**	**0.986**	**0.895**	**0.881**	**0.876**	**0.881**	**0.951**	**0.936**

Convergence Analysis of CFF

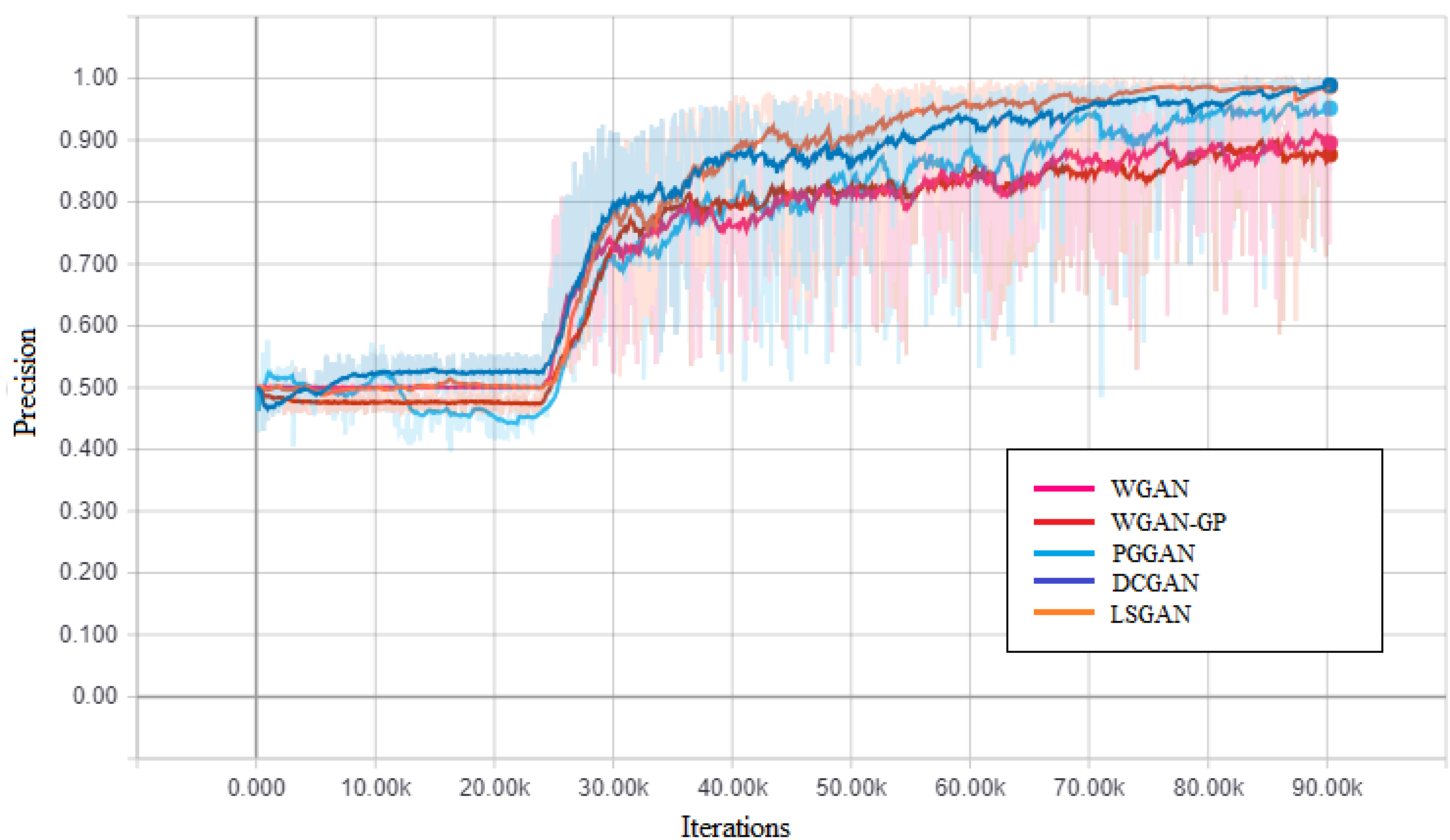


Performance Comparison

- Supervised learning vs. pairwise learning



Precision Curves for GANs Used in Our Experiments



Visualization of the Unrealistic Details in the Fake Image

- Fully convolutional network can be used to visualize the unrealistic details

Conclusion

- The proposed a novel deep forgery discriminator (DeepFD) can successfully detect the fake images
- Contribution
  - The first work to solve the problems of detecting the fake images
  - The proposed CFF can capture the common feature for fake images generated by different GANs
  - Visualization of the proposed DeepFD can be used to further improve the detector algorithm

More information can be found at
https://cchsu.info