Learning and forgetting in image classification and generation

Luis Herranz
Learning and machine perception (LAMP) group
Computer Vision Center (Barcelona)

Universitat Autònoma de Barcelona

MARIE CURIE ACTIONS

Horizon 2020

About me

ETS Ing. Telecomunicación
Universidad Politécnica de Madrid (Spain)

(Ph.D) Escuela Politécnica Superior
Universidad Autónoma de Madrid (Spain)

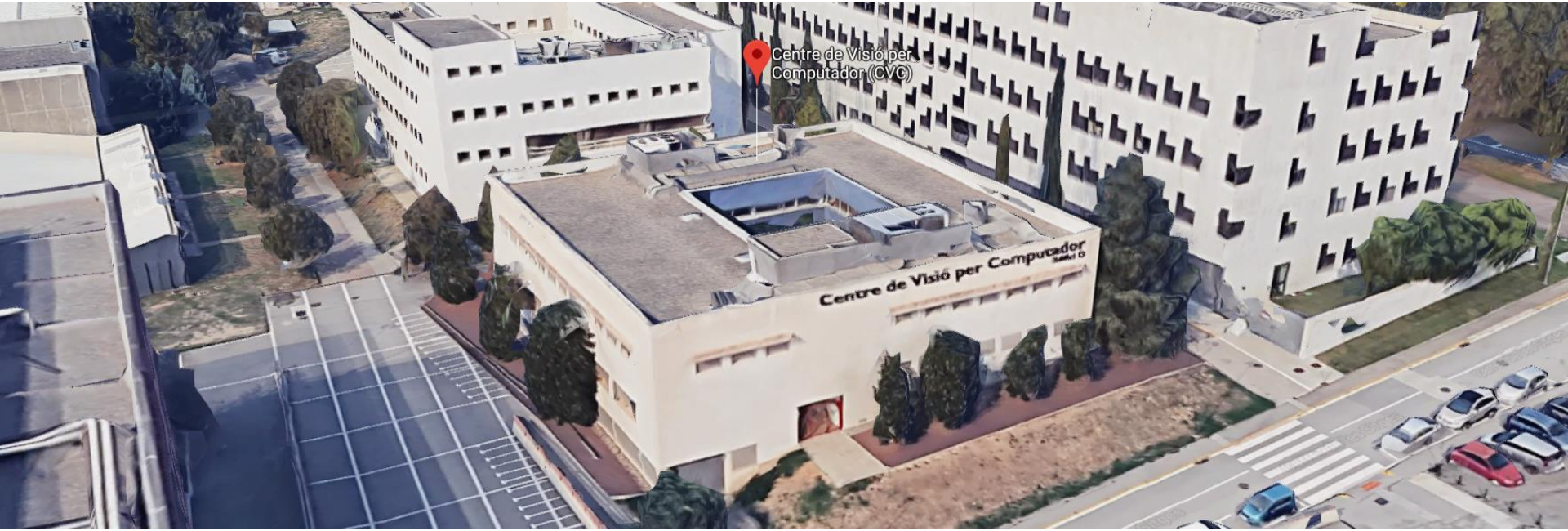
Misubishi Electric R&D (UK)

Institute of Computing Technology,
Chinese Academy of Sciences (China)

Computer Vision Center,
Universitat Autònoma de Barcelona (Spain)

Research interests:
Multimedia
Computer vision
Deep learning
Multimodal representations
Lifelong learning

Computer Vision Center (UAB campus)



Only Center in Europe fully devoted to Computer Vision

23 Years

+130 Staff

+20 Nationalities

M€2,3 Income / year

8 PhD thesis /year

+100 Intl publications / year

Learning and Machine Perception (LAMP) group

Joost van de Weijer
Leader

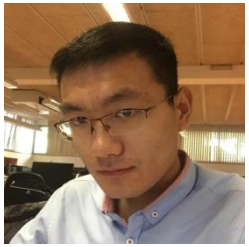
Senior PhDs



Postdocs


PhD students











Outline

- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)

Outline

- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)

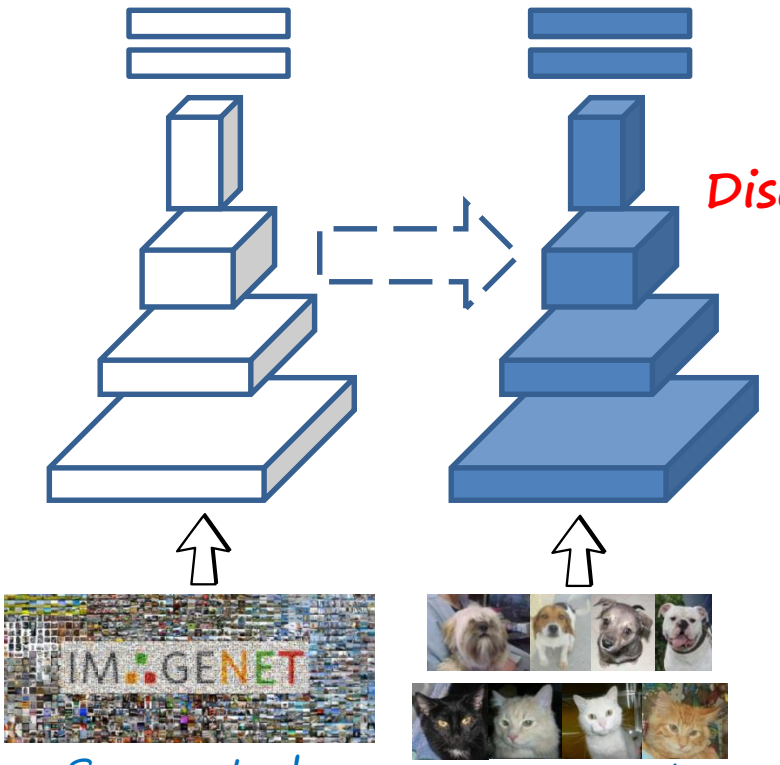
Transfer learning and lifelong learning

Forgets source task, i.e. catastrophic forgetting **(who cares?)**

Forgets task 1 **(big deal!!)**

Transfer+adaptation

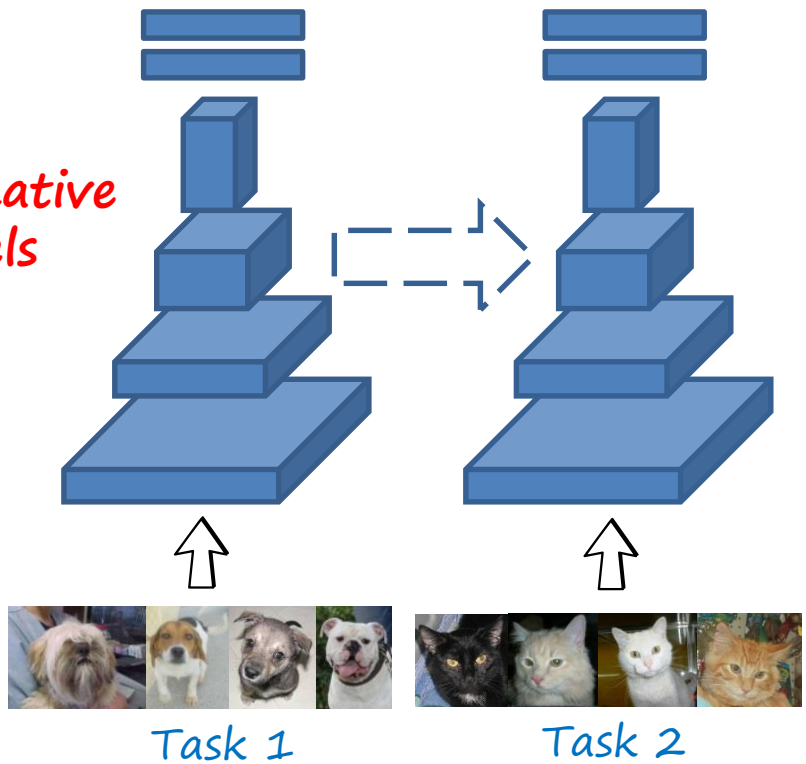
Lifelong learning



Source task

Target task

Discriminative models



Task 1

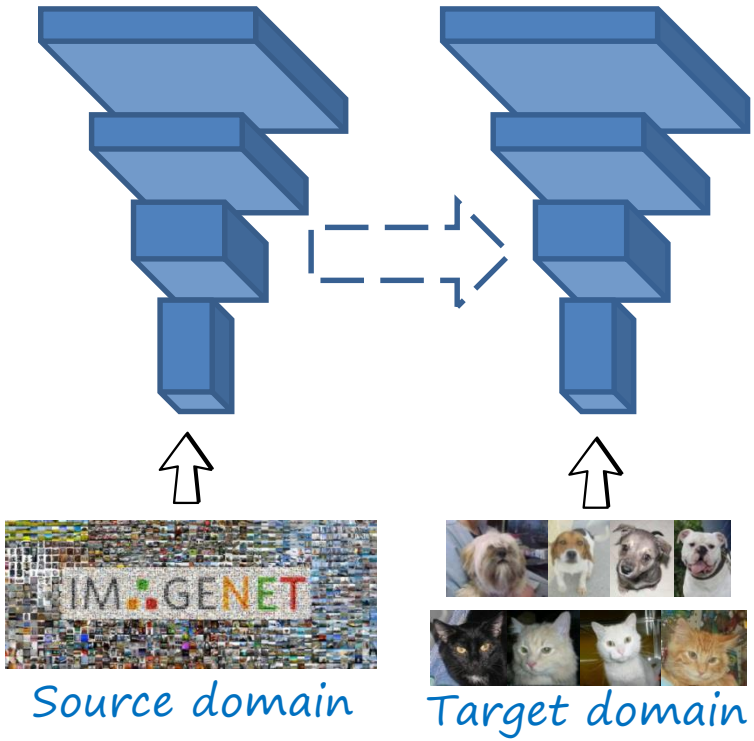
Task 2

Rotated elastic weight consolidation (ICPR 2018)

Liu et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

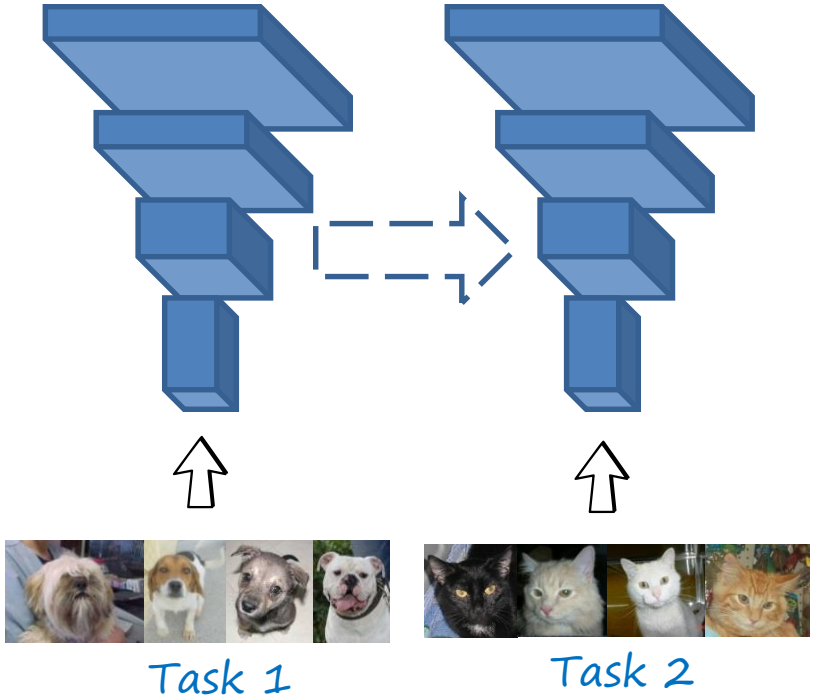
Transfer learning and lifelong learning (now with GANs for image generation)

Transfer+adaptation (generative)



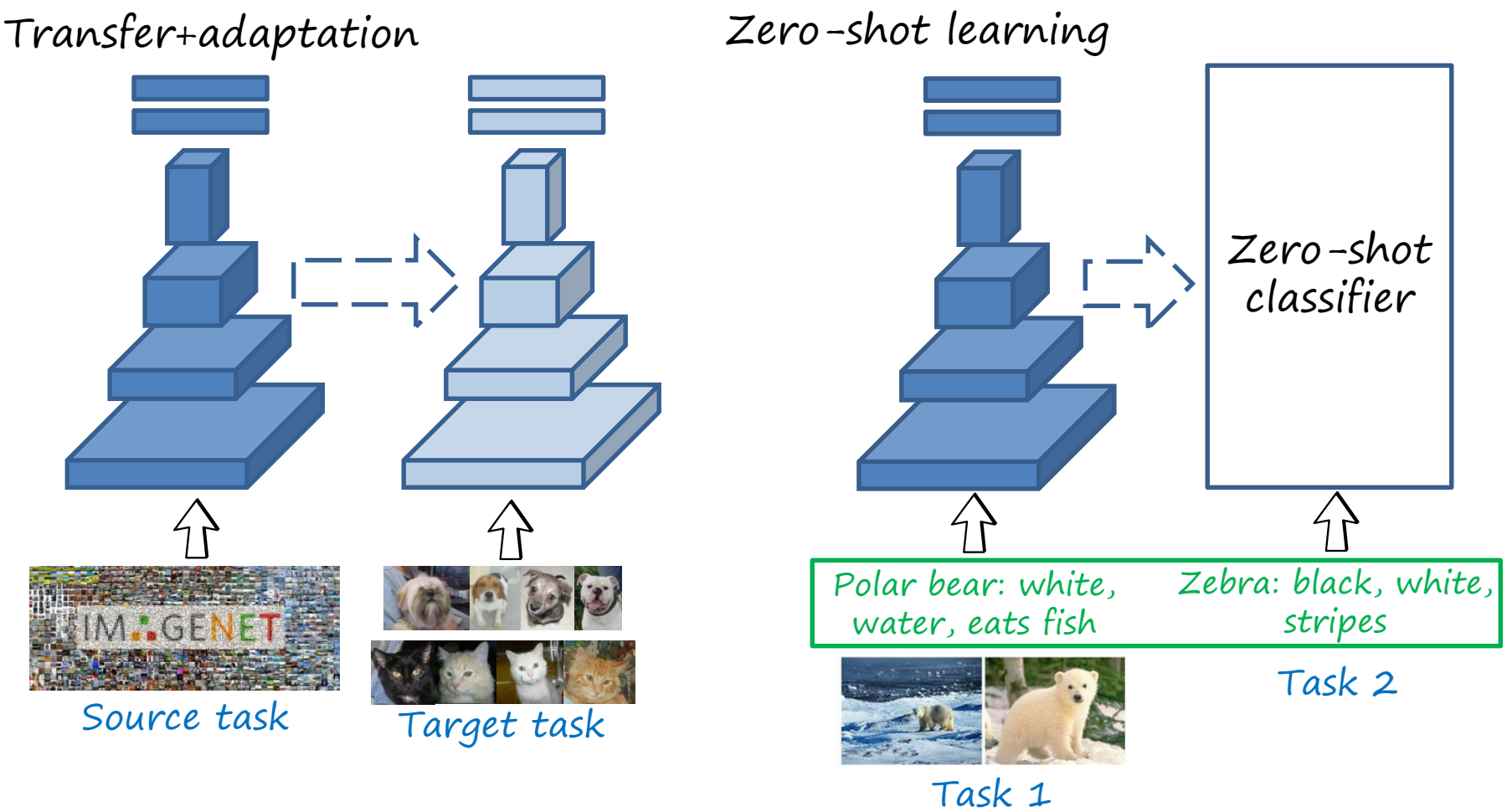
Transferring GANs (ECCV 2018)

Lifelong learning (generative)



Memory Replay GANs (NIPS 2018)

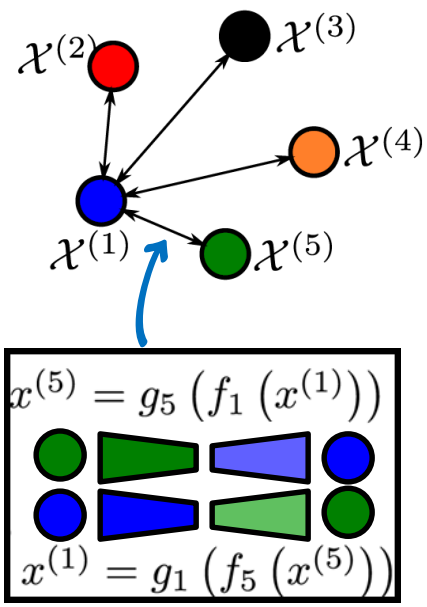
Transfer learning and zero-shot learning



Zero-pair image-to-image translations

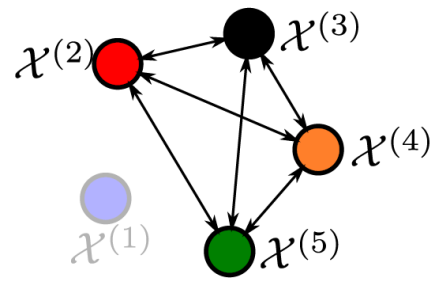
*Only these translations are trained (**seen**)*

Train



*Evaluate on these **unseen** translations (no training pairs)*

Test



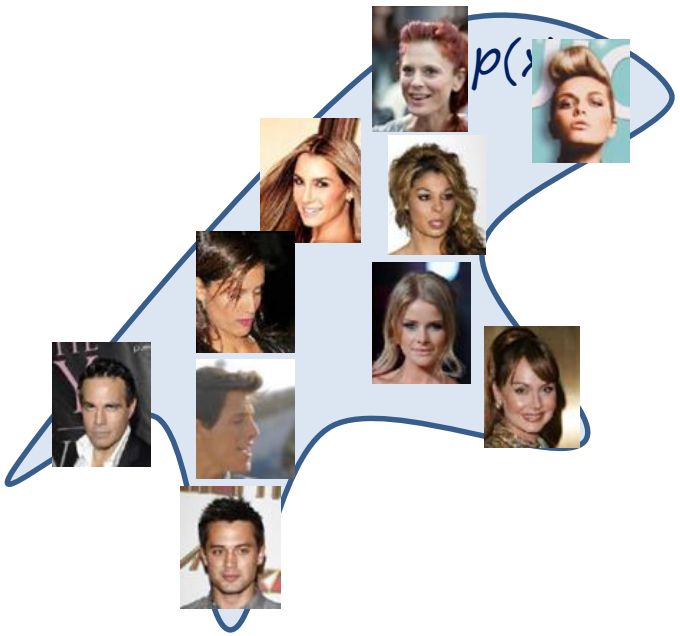
Mix and match networks (CVPR 2018)

Outline

- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)

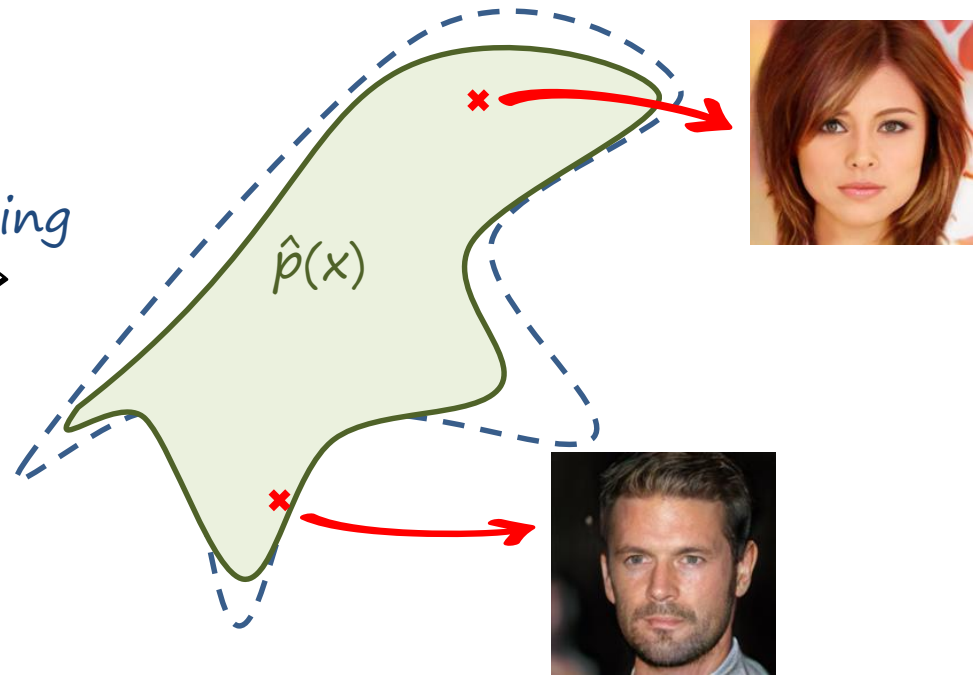
Generative models: networks that imagine

Training data
(e.g. 64x64x3≈12K dims)



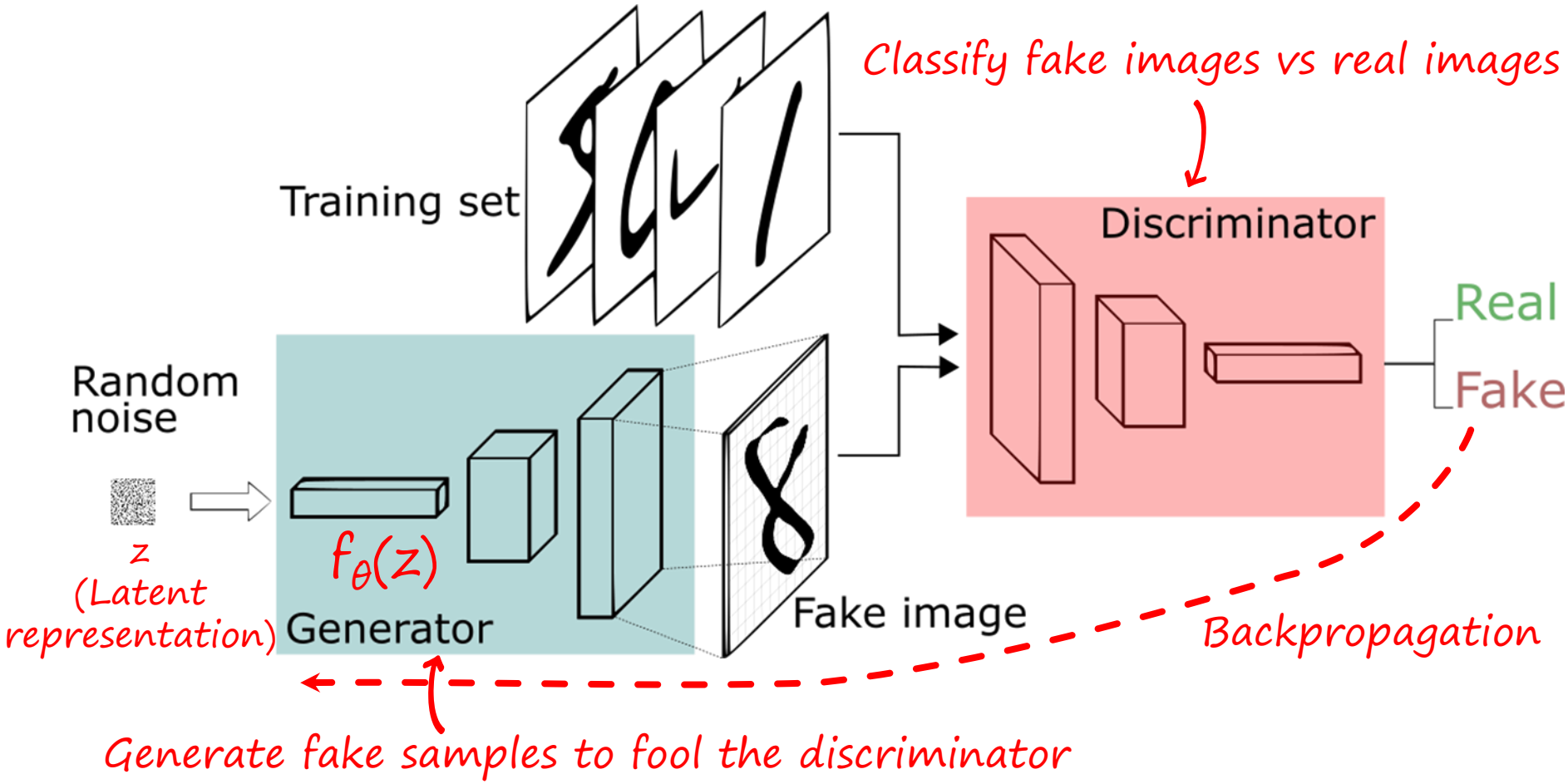
Learning

Sampling



Different approaches
- Density estimation
- Variational autoencoders
- Autoregressive models
- Generative adversarial networks (GANs)

Generative Adversarial Networks (GANs)

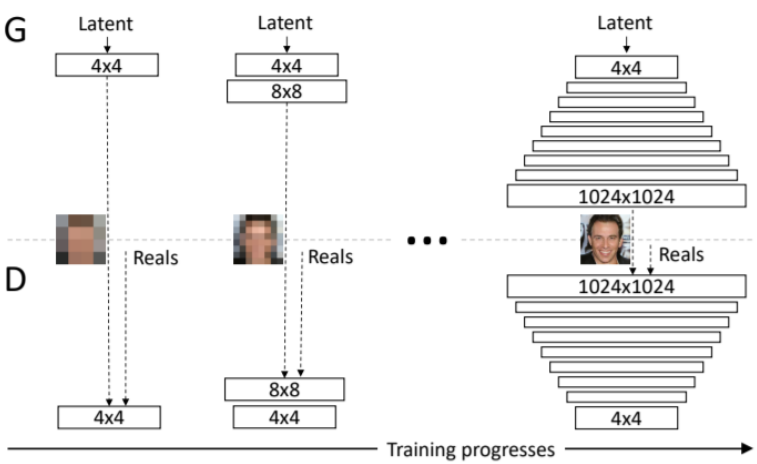


Goodfellow et al., "Generative Adversarial Networks", NIPS 2014
Figure from https://deeplearning4j.org/generative-adversarial-network

Generative Adversarial Networks

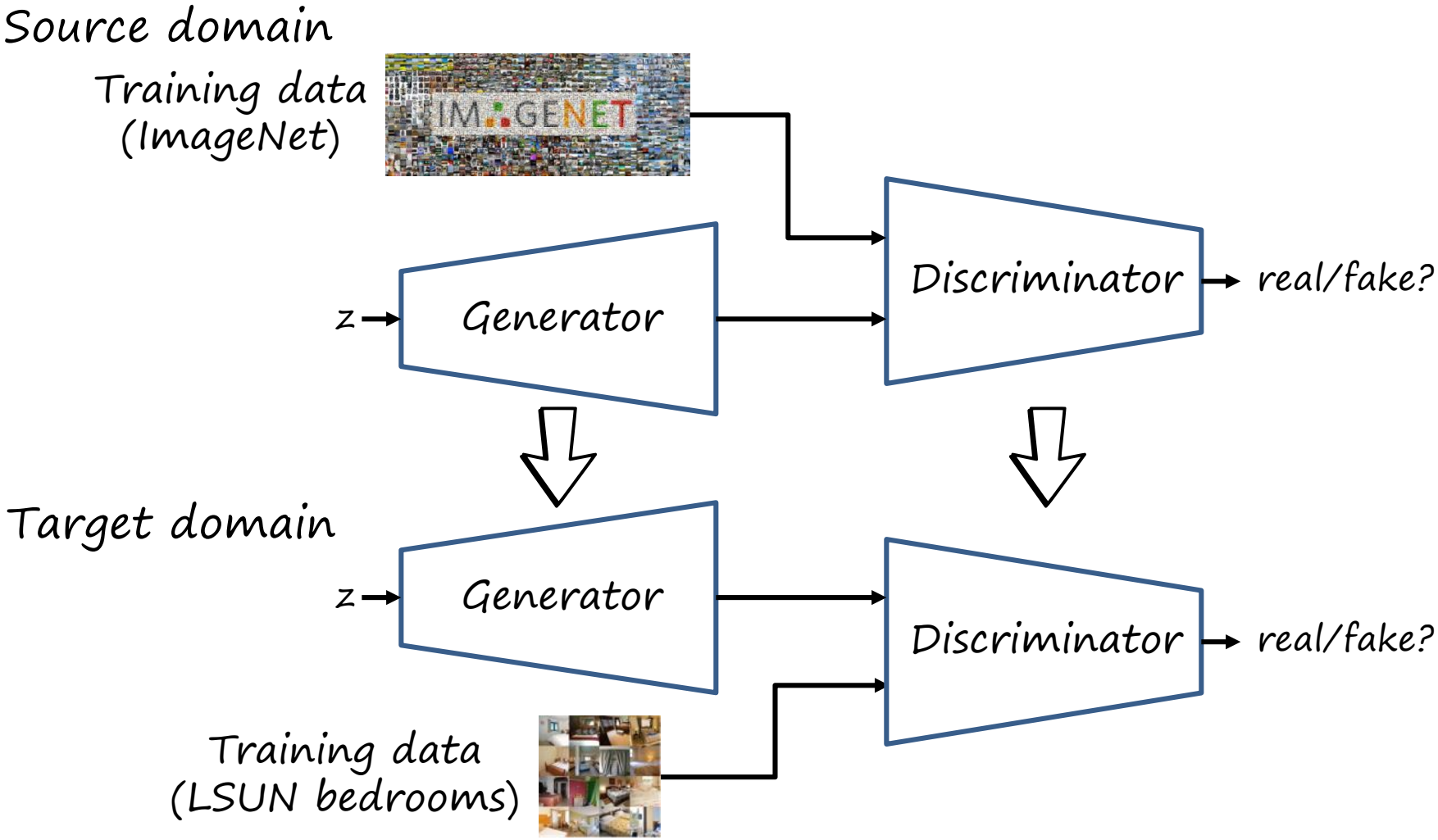
Wasserstein GAN (WGAN-GP)

Progressive growing of GANs



and many more…

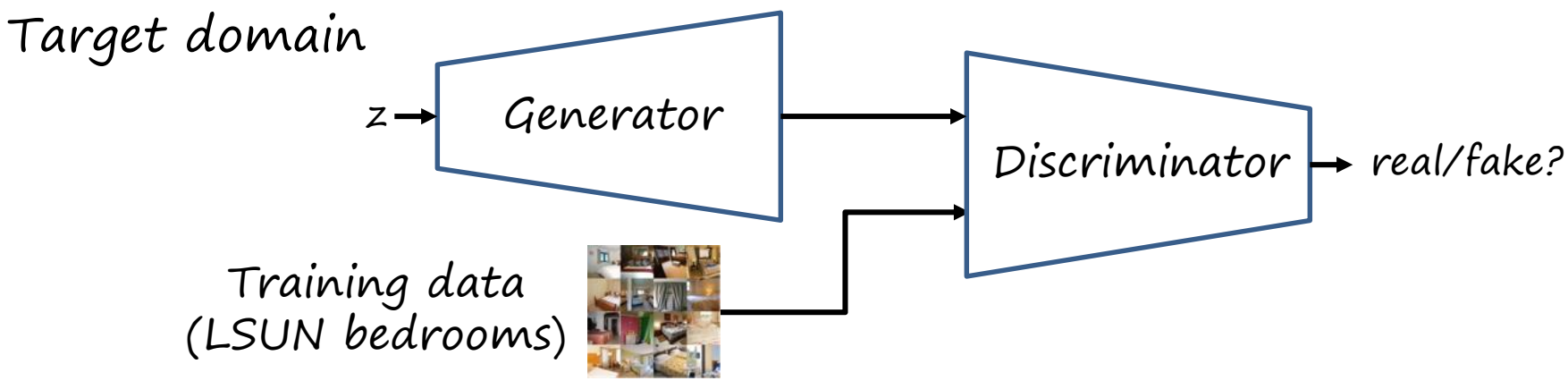
Transferring GAN representations



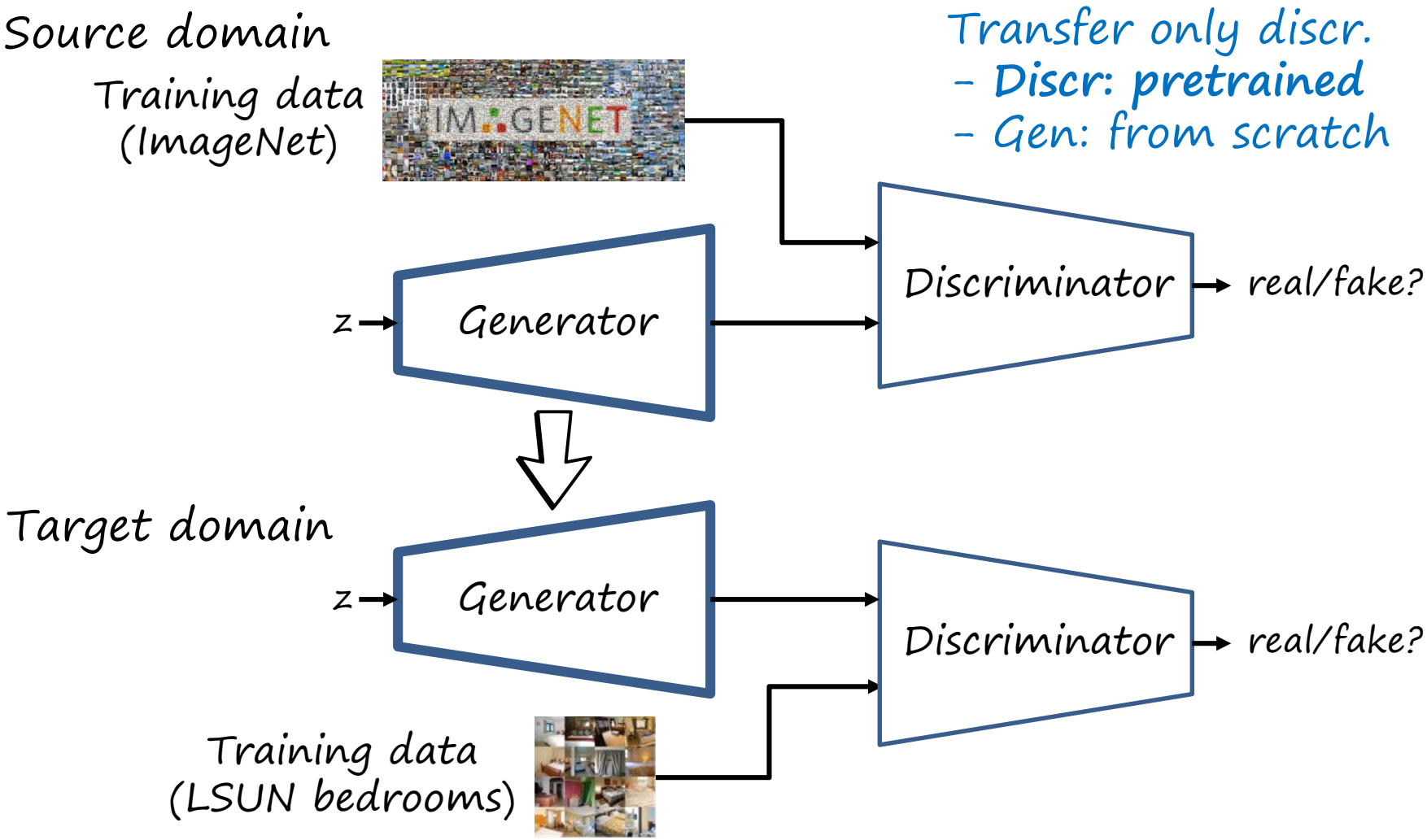
Wang *et al.*, "Transferring GANs: generating images from limited data", ECCV 2018

Transfer configuration

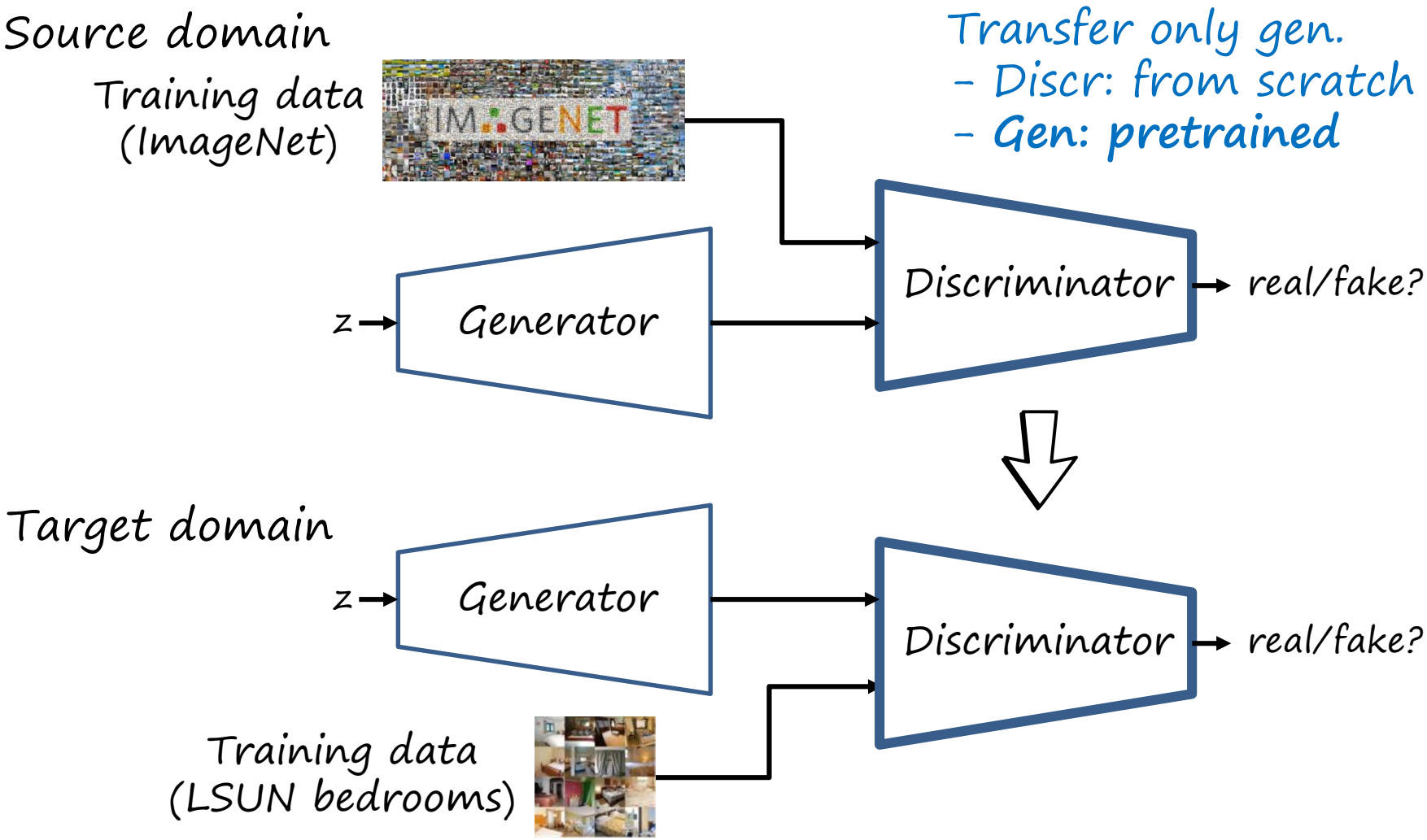
Training from scratch
- Discr: from scratch
- Gen: from scratch



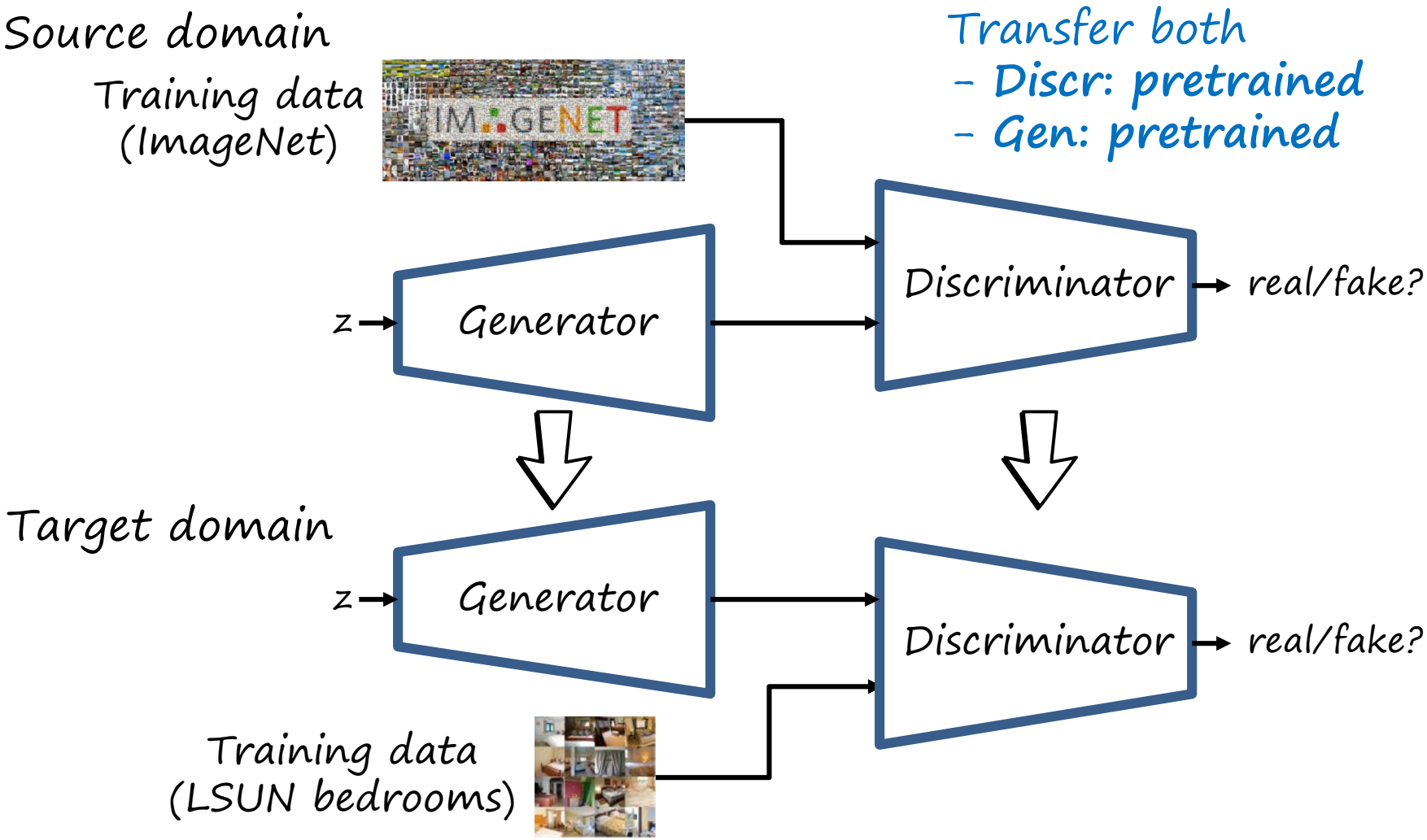
Transfer configuration



Transfer configuration



Transfer configuration

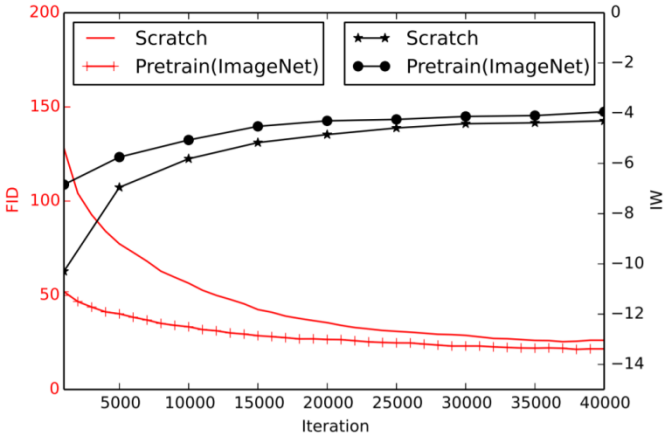


Transfer configuration

- What should I transfer?
  - Experiment: ImageNet to Bedrooms (100K images)

	Generator	Scratch		Pretrained	
	Discriminator	Scratch	Pretrained	Scratch	Pretrained
Lower better	FID $(\mathcal{X}^{tgt}_{data}, \mathcal{X}^{tgt}_{gen})$	32.87	30.57	56.16	**24.35**
Higher better	IW $(\mathcal{X}^{tgt}_{val}, \mathcal{X}^{tgt}_{gen})$	-4.27	-4.02	-6.35	**-3.88**
		None	Only discr.	Only gen.	Both

- Training is faster and images have better quality
  - Especially when data is limited



Learning with limited data

Target dataset:
Bedroom

From scratch

Pretrained (ImageNet)

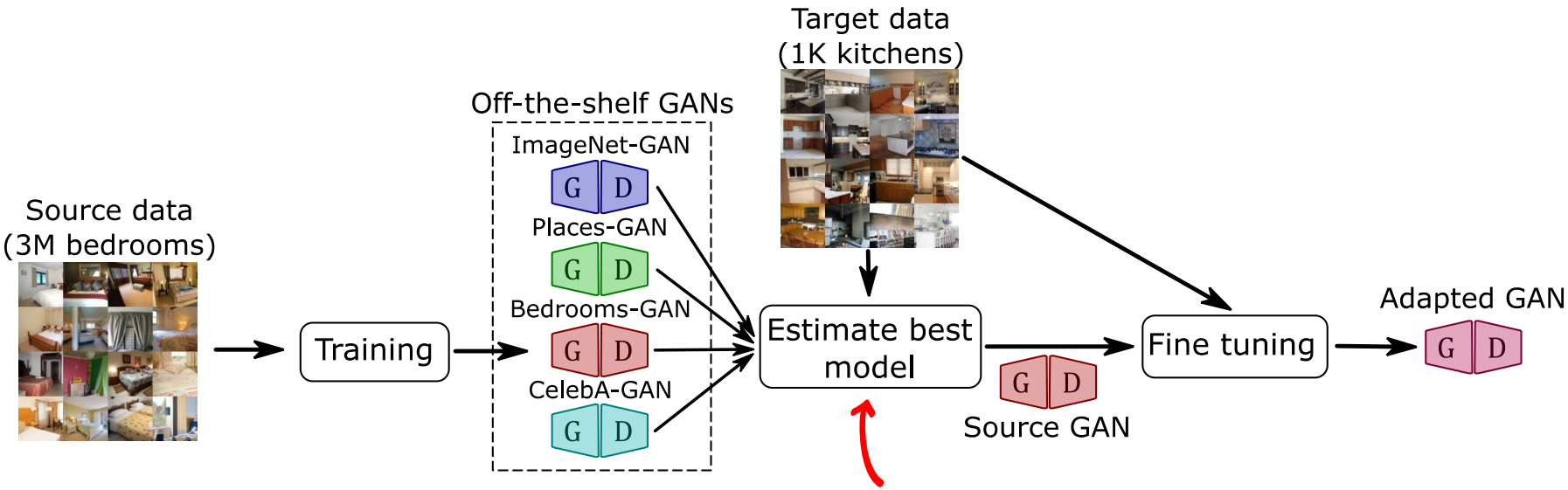
1000 images

10000 images

100000 images

Wang *et al.*, "Transferring GANs: generating images from limited data", ECCV 2018

Pretrain model selection



Estimated as
$\min_k$ FID(generated by GAN_k, target real)

Wang *et al.*, "Transferring GANs: generating images from limited data", ECCV 2018

Good/bad source datasets

Source

Places
(205 classes
~10K img/class)

Target

Kitchen
(50K)

Bedrooms
(1 class
3M images)

Kitchen
(50K)

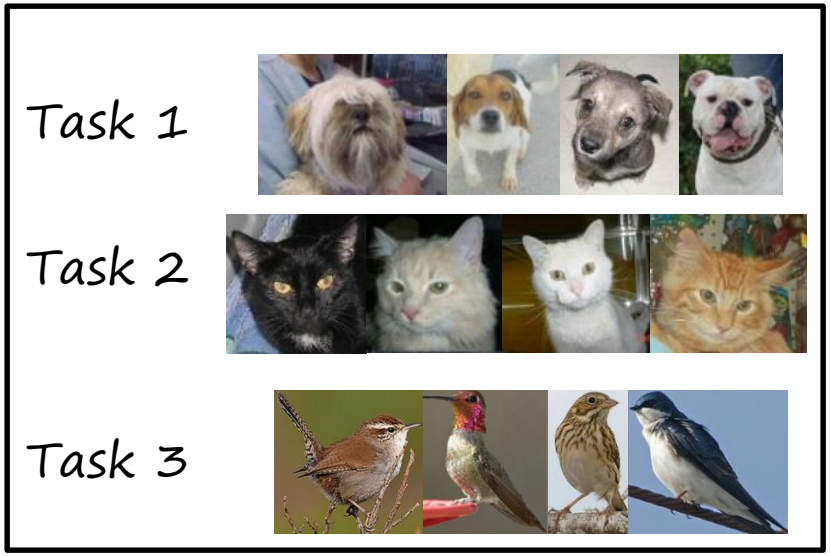
Generative: very **dense**, diversity not so important (e.g. LSUN Bedrooms)

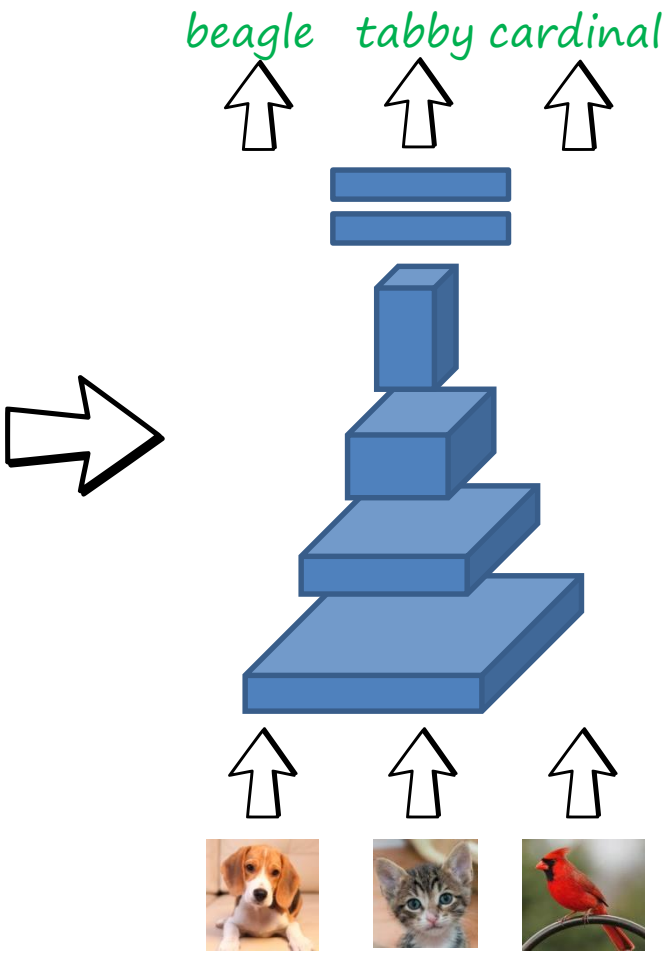
Discriminative: very **diverse**, medium density (e.g. ImageNet, Places)

Outline

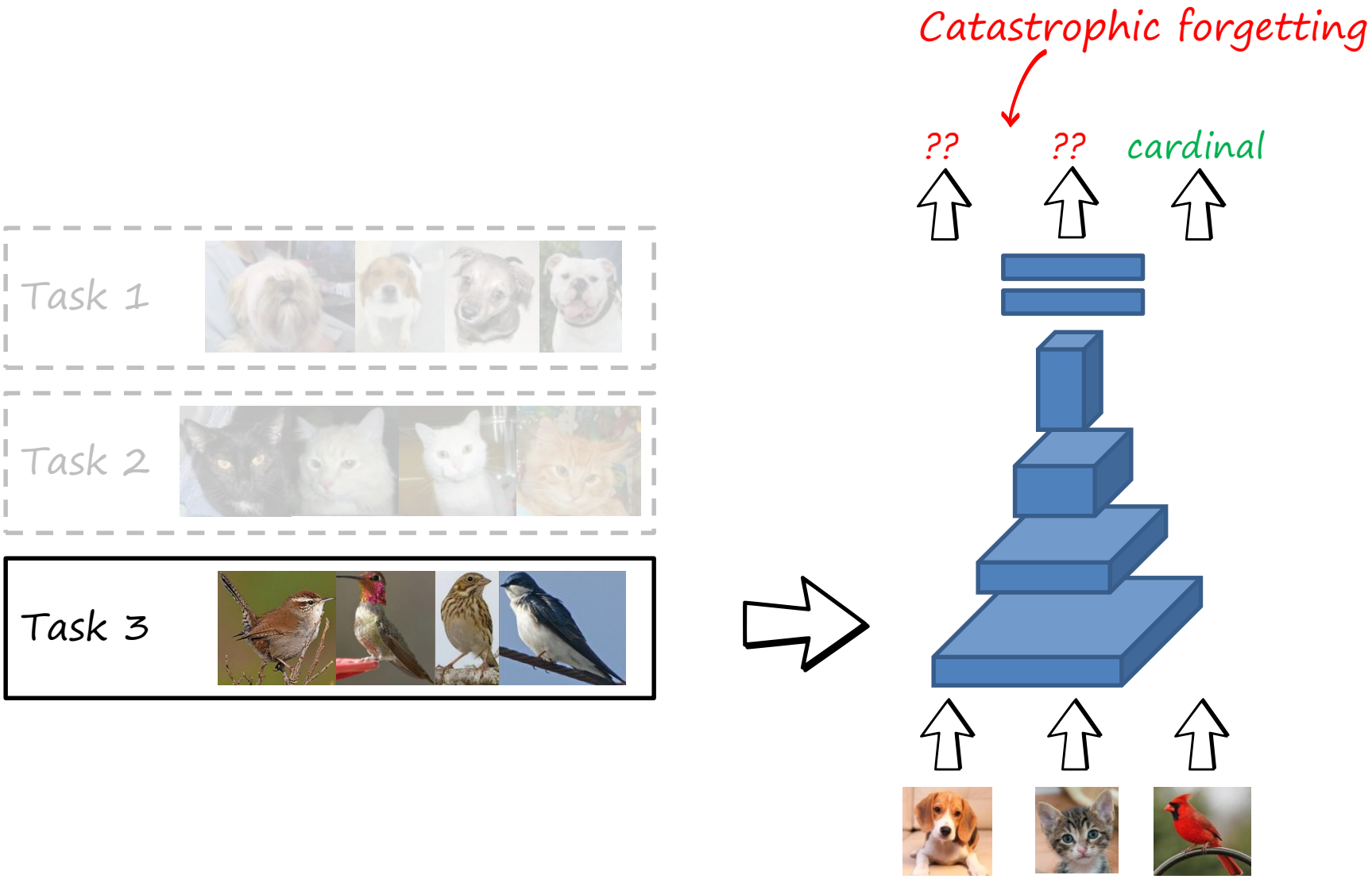
- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)

When are neural networks good?

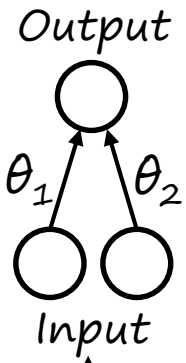




Sequential learning

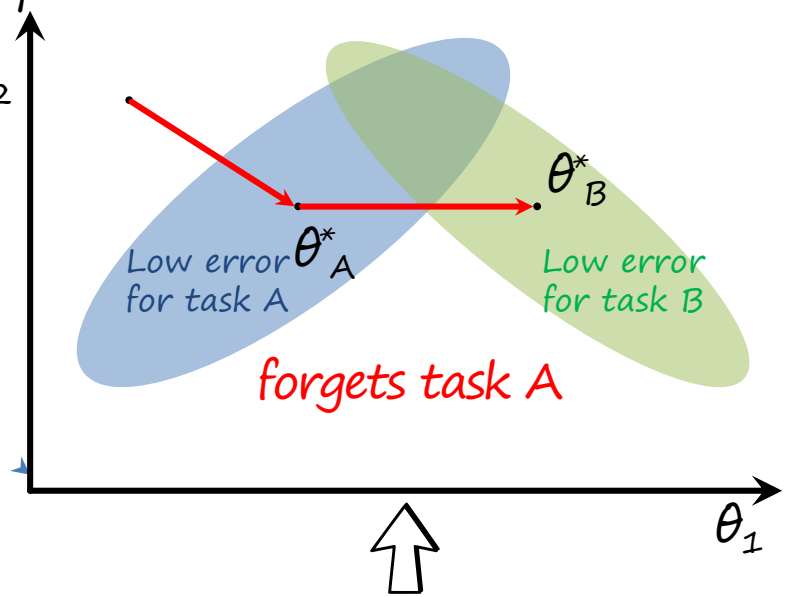


Catastrophic interference and forgetting



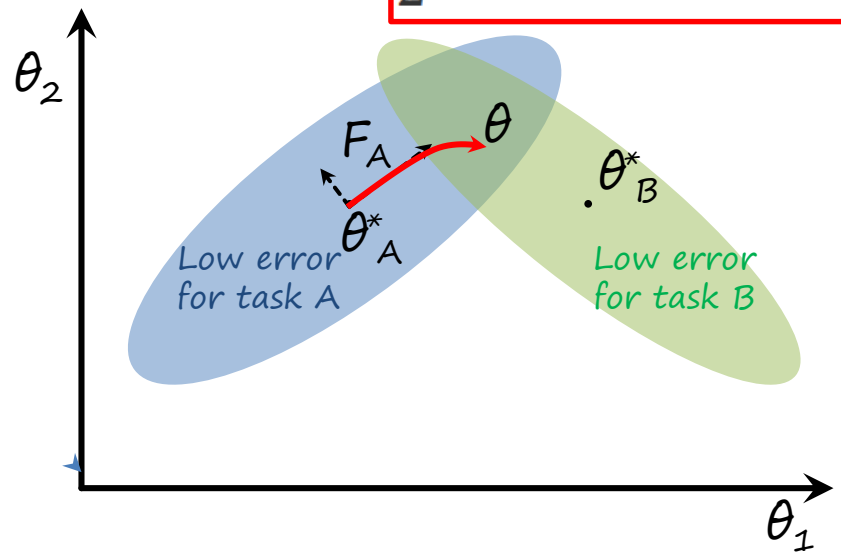
Training/fine tuning

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta)$$



Elastic weight consolidation (EWC)

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \boxed{\frac{\lambda}{2}(\theta - \theta^*_A)^\intercal F_A (\theta - \theta^*_A)}$$



Problem: size of F_A is $(\#\theta)^2$!!!

Task B

Kirkpatric et al., Overcoming catastrophic forgetting in neural networks, PNAS, 2017

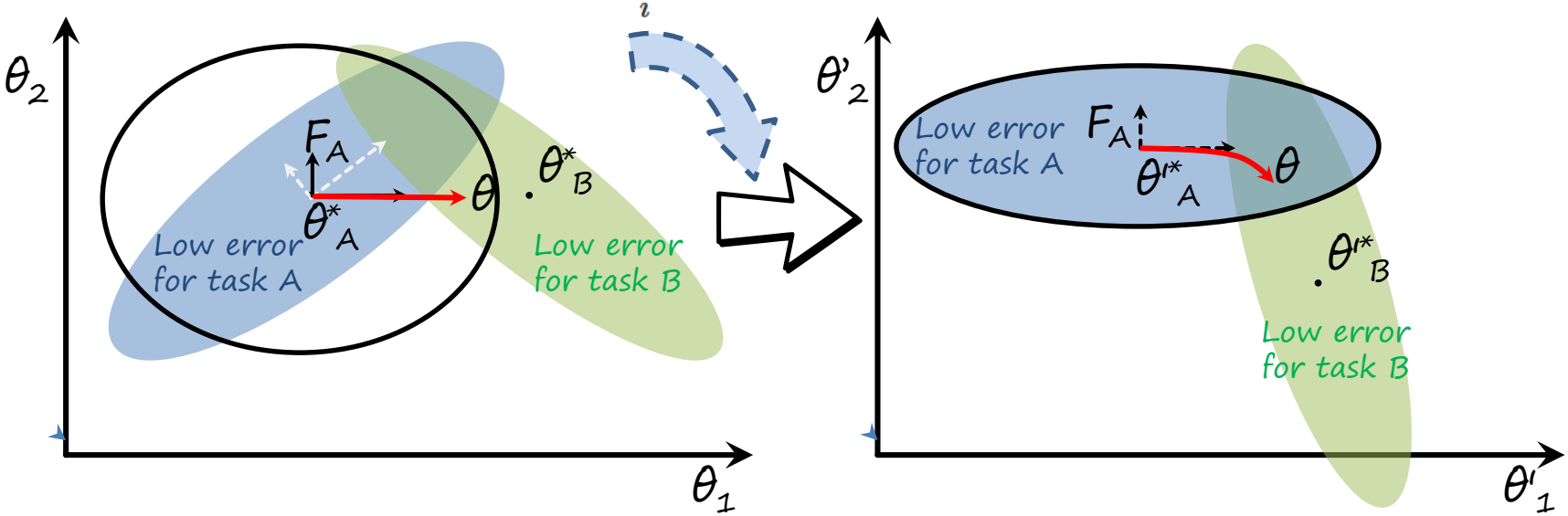
Rotated elastic weight consolidation

Size of the diagonal of F_A is $\#\theta$

Elastic weight consolidation in practice
(with diagonal approx. of F_A)

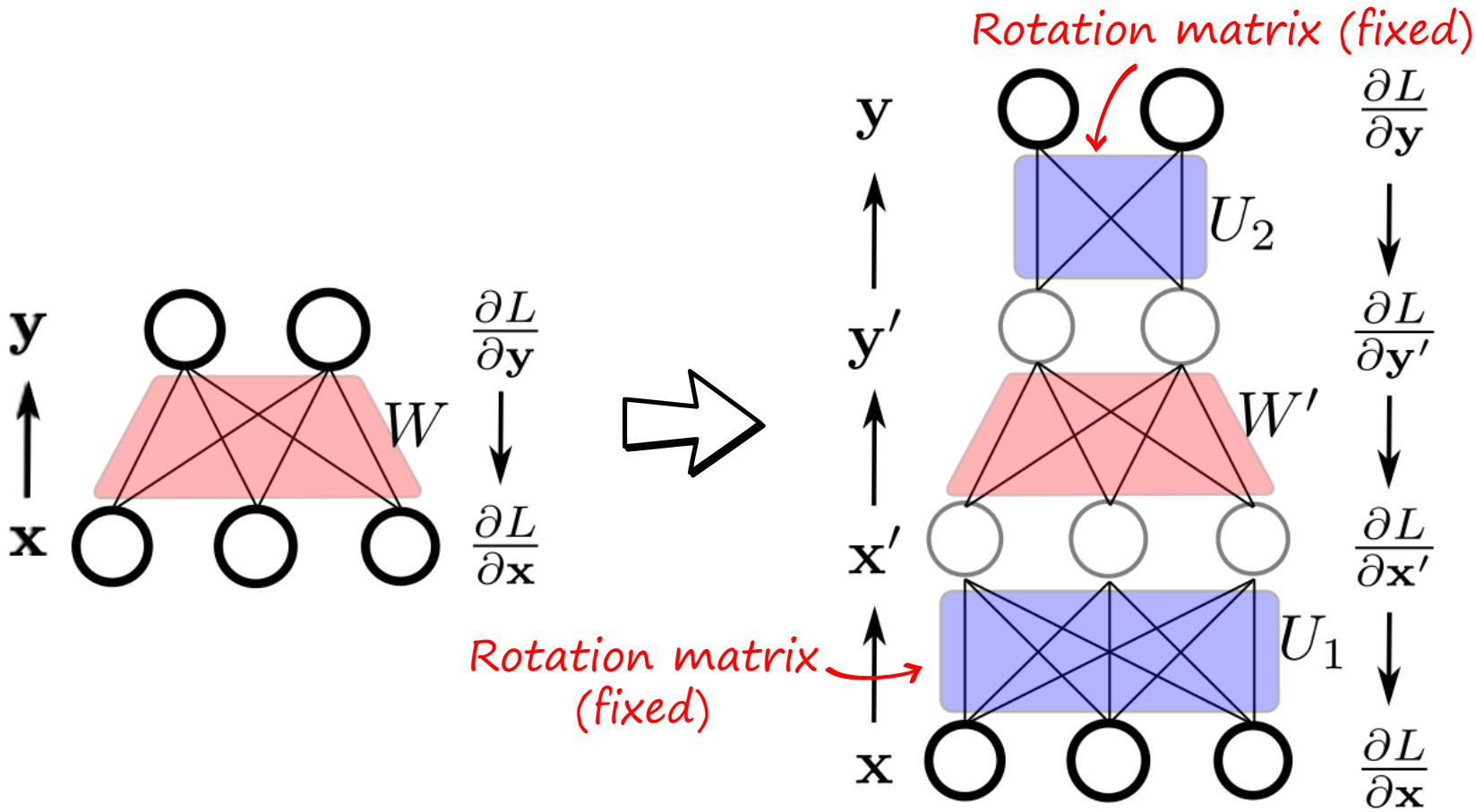
Rotated elastic weight consolidation (R-EWC)

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \frac{\lambda}{2}\sum_i (F_A)_i \left(\theta_i - (\theta^*_A)_i\right)^2$$



Liu, Masana et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

Rotating fully connected layers



Liu , Masana et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

Computing the rotations

$$F_W = \mathbb{E}_{p\sim\pi}\left[\left(\frac{\partial L}{\partial \mathbf{y}}\right)\mathbf{x}\mathbf{x}^\intercal\left(\frac{\partial L}{\partial \mathbf{y}}\right)^\intercal\right]$$

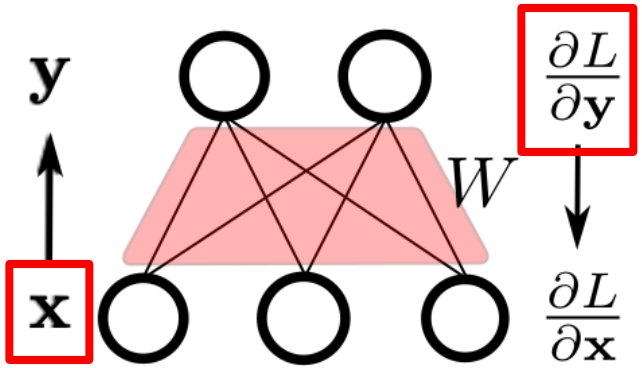
Assuming x and δL/δy independent

$$F_W \approx \mathbb{E}_{\substack{x\sim\pi\\ y\sim p}}\left[\left(\frac{\partial L}{\partial \mathbf{y}}\right)\left(\frac{\partial L}{\partial \mathbf{y}}\right)^\intercal\right]\mathbb{E}_{x\sim\pi}\left[\mathbf{x}\mathbf{x}^\intercal\right]$$

$$\mathbb{E}_{x\sim\pi}\left[\mathbf{x}\mathbf{x}^\intercal\right] = \boxed{U_1}S_1V_1^\intercal$$

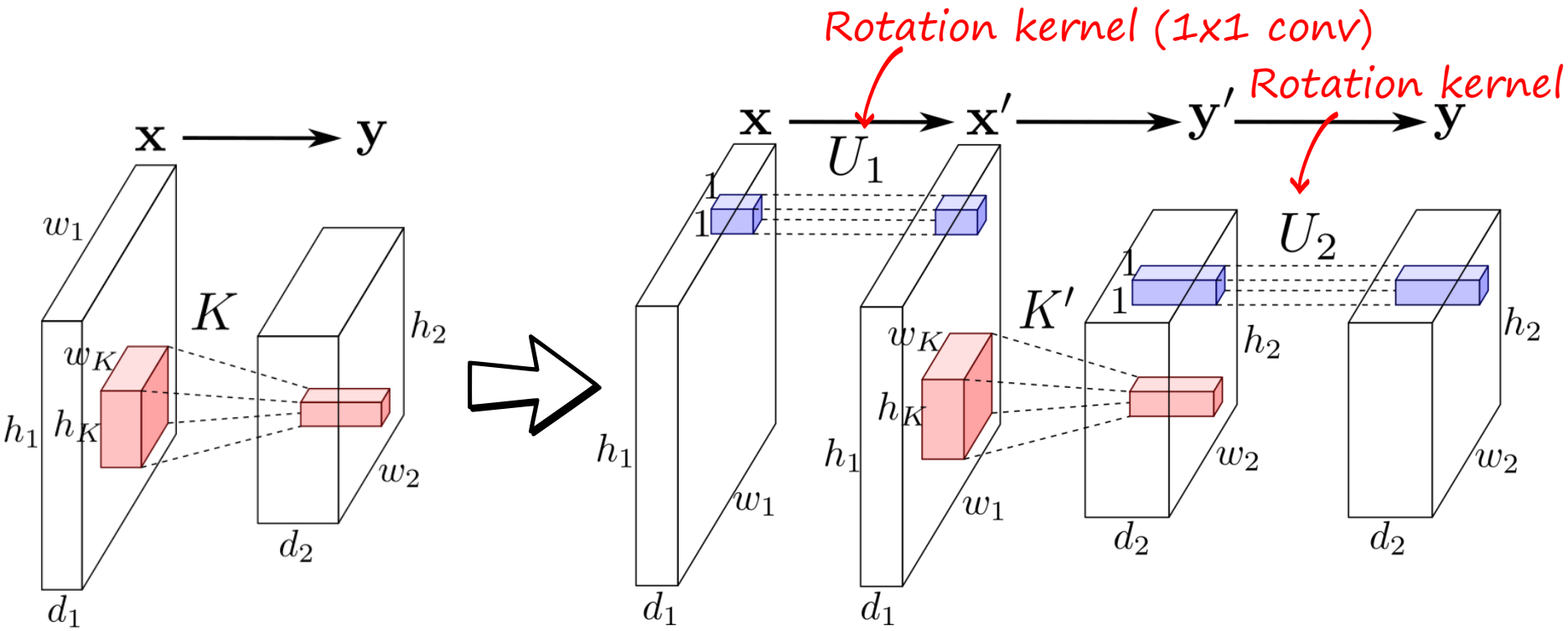
$$\mathbb{E}_{\substack{x\sim\pi\\ y\sim p}}\left[\left(\frac{\partial L}{\partial \mathbf{y}}\right)\left(\frac{\partial L}{\partial \mathbf{y}}\right)^\intercal\right] = \boxed{U_2}S_2V_2^\intercal$$

Using SVD



Liu , Masana et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

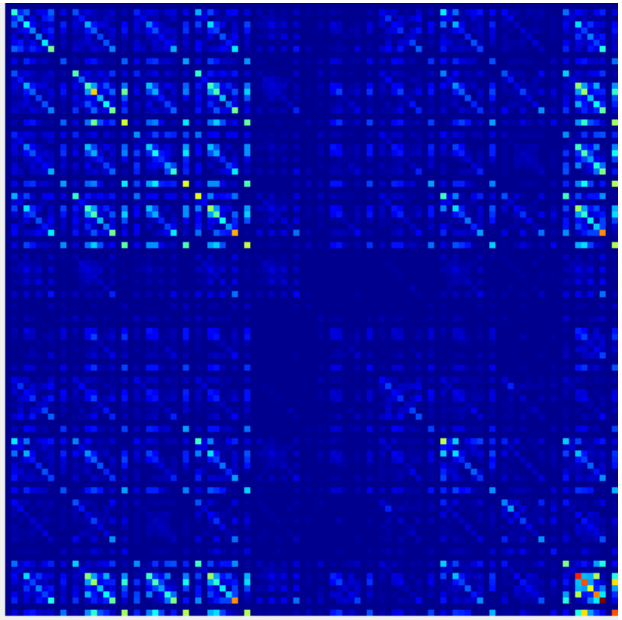
Rotating convolutional layers



Liu et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

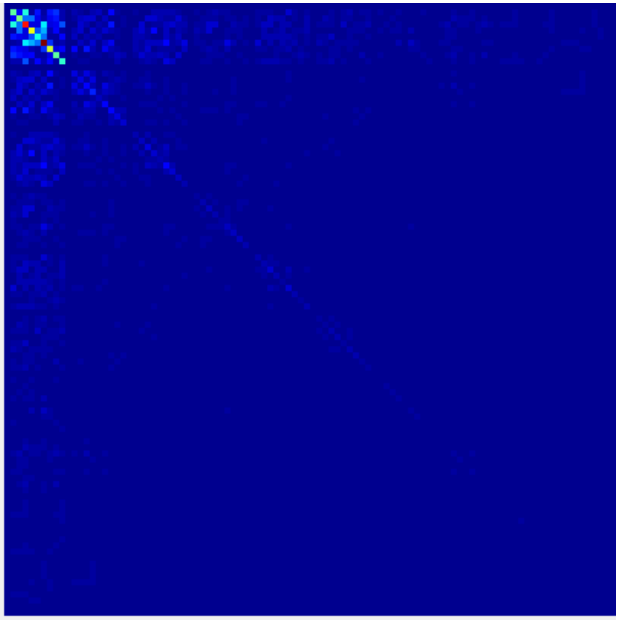
Fisher Information matrix

No rotation (i.e. EWC)



Energy in the diagonal: 40%

After rotation (i.e. R-EWC)



Energy in the diagonal: 74%

Liu , Masana et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

Experimental results (2 tasks)

- MNIST dataset. Two tasks: 0-4 and 5-9

	$\lambda = 1$		$\lambda = 10$		$\lambda = 100$		$\lambda = 1000$		$\lambda = 10000$	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
FT	6.1	97.6	6.1	97.6	6.1	97.6	6.1	97.6	6.1	97.6
EWC [5]	66.8	90.9	75.3	95.6	**85.8**	**92.8**	78.4	93.7	81.0	88.8
R-EWC - conv only	62.7	89.2	67.5	96.1	80.4	91.4	**84.7**	**93.1**	75.5	93.7
R-EWC - fc only	78.9	95.3	79.0	95.8	87.4	93.5	93.0	82.3	**94.3**	**88.0**
R-EWC - all	77.2	96.7	**91.7**	**91.2**	86.9	95.9	96.3	81.1	92.1	86.0
R-EWC - all no last	71.5	91.8	84.9	97.0	**91.6**	**94.5**	94.6	88.4	97.9	79.4

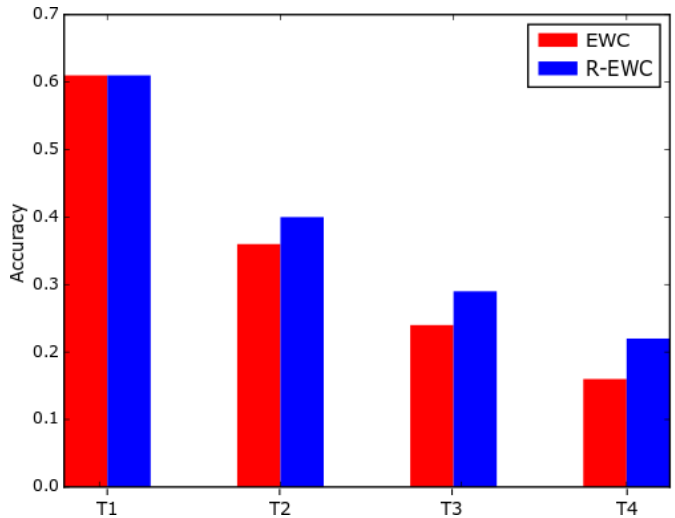
- Several datasets

	EWC [5] (T1 / T2)	R-EWC (T1 / T2)
MNIST	89.3 (85.8 / 92.8)	**93.1** (91.6 / 94.5)
CIFAR-100	37.5 (23.5 / 51.5)	**42.5** (30.2 / 54.7)
CUB-200 Birds	45.3 (42.3 / 48.6)	**48.4** (53.3 / 45.2)
Stanford-40 Actions	50.4 (44.3 / 58.4)	**52.5** (52.3 / 52.6)

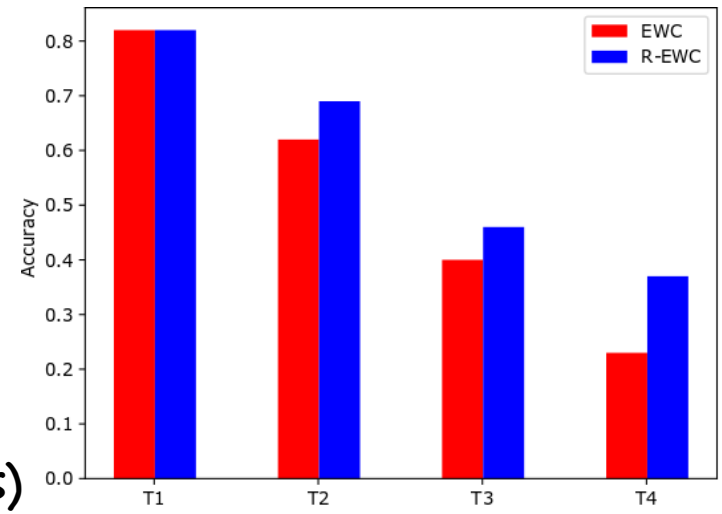
Liu , Masana et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

Experimental results (4 tasks)

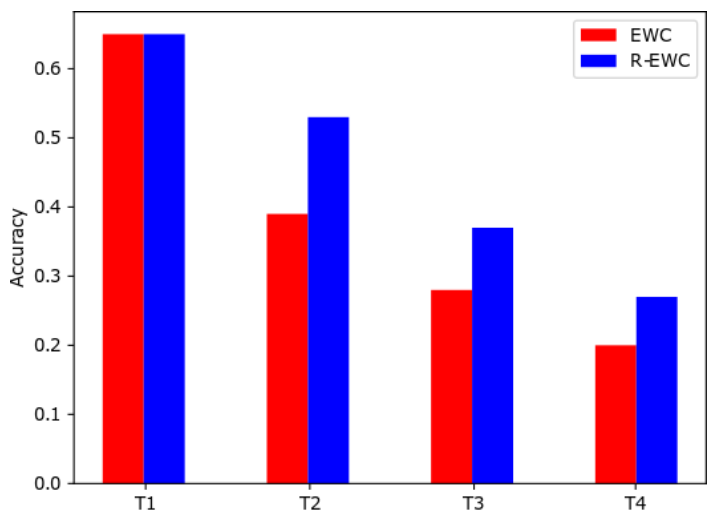
CIFAR 100



Stanford Actions



CUB 200 (birds)

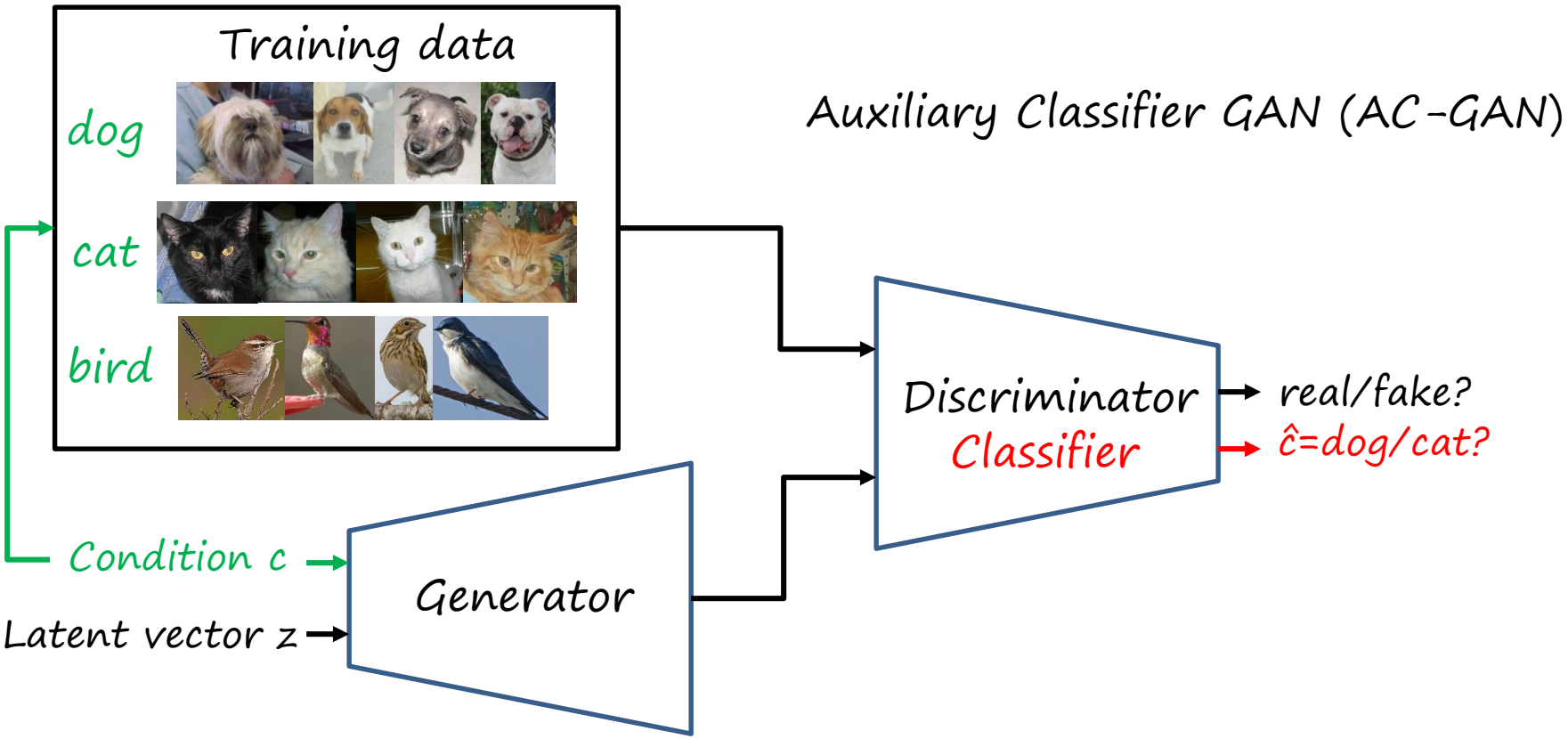


Liu , Masana et al., Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting, ICPR 2018

Outline

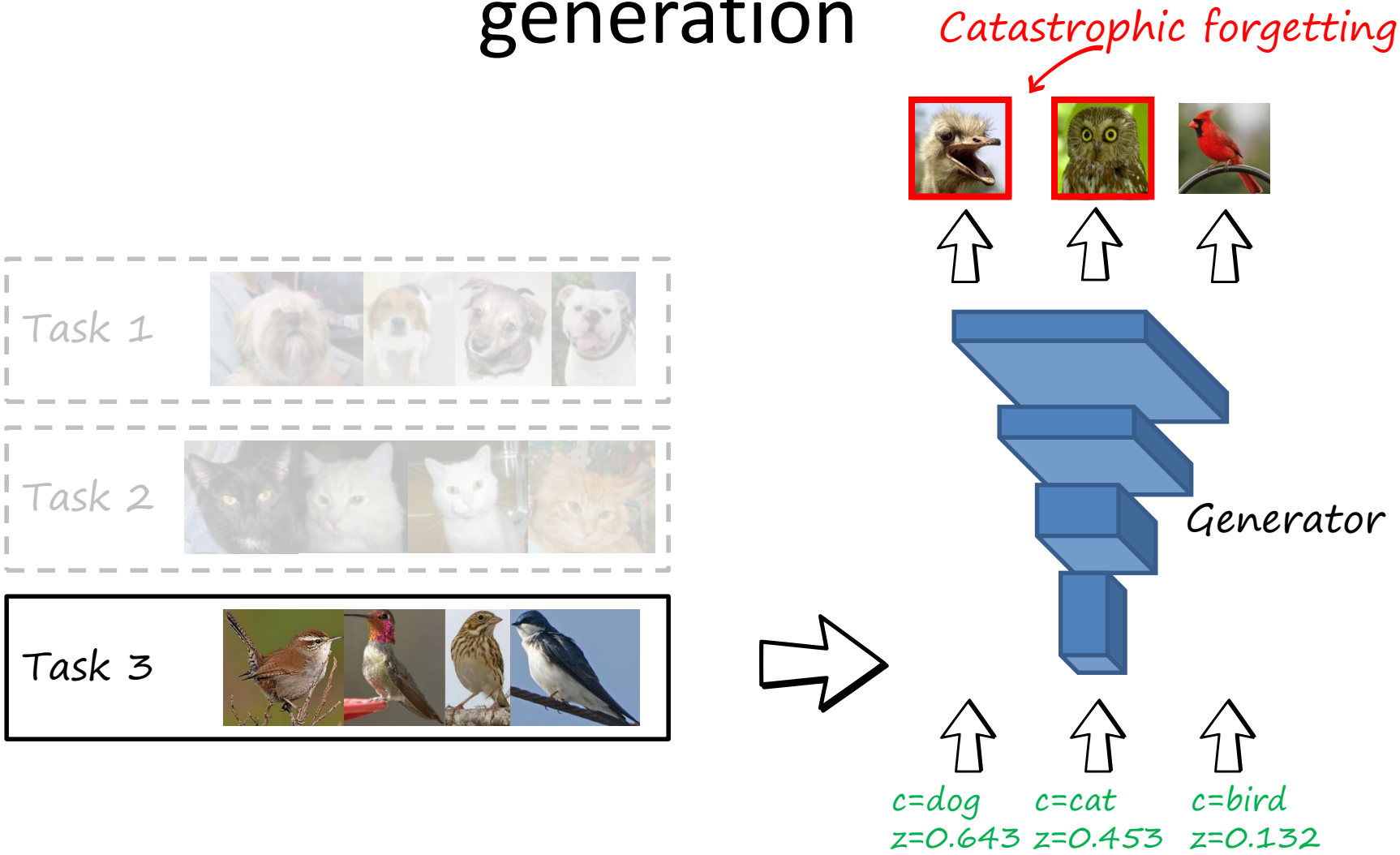
- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)

Joint training (non-sequential)



Odena et al., Conditional Image Synthesis With Auxiliary Classifier GANs, ICML 2017

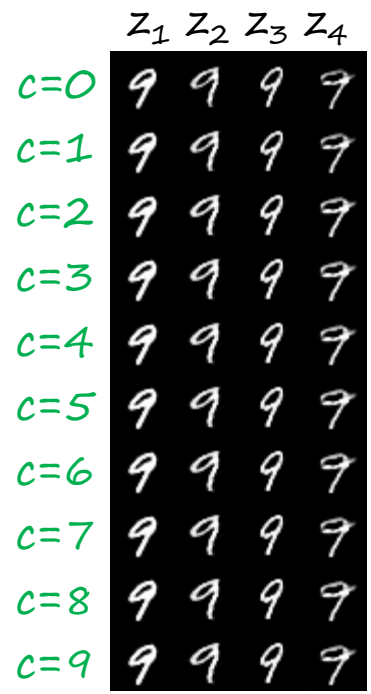
Sequential learning for image generation



Sequential fine tuning and forgetting

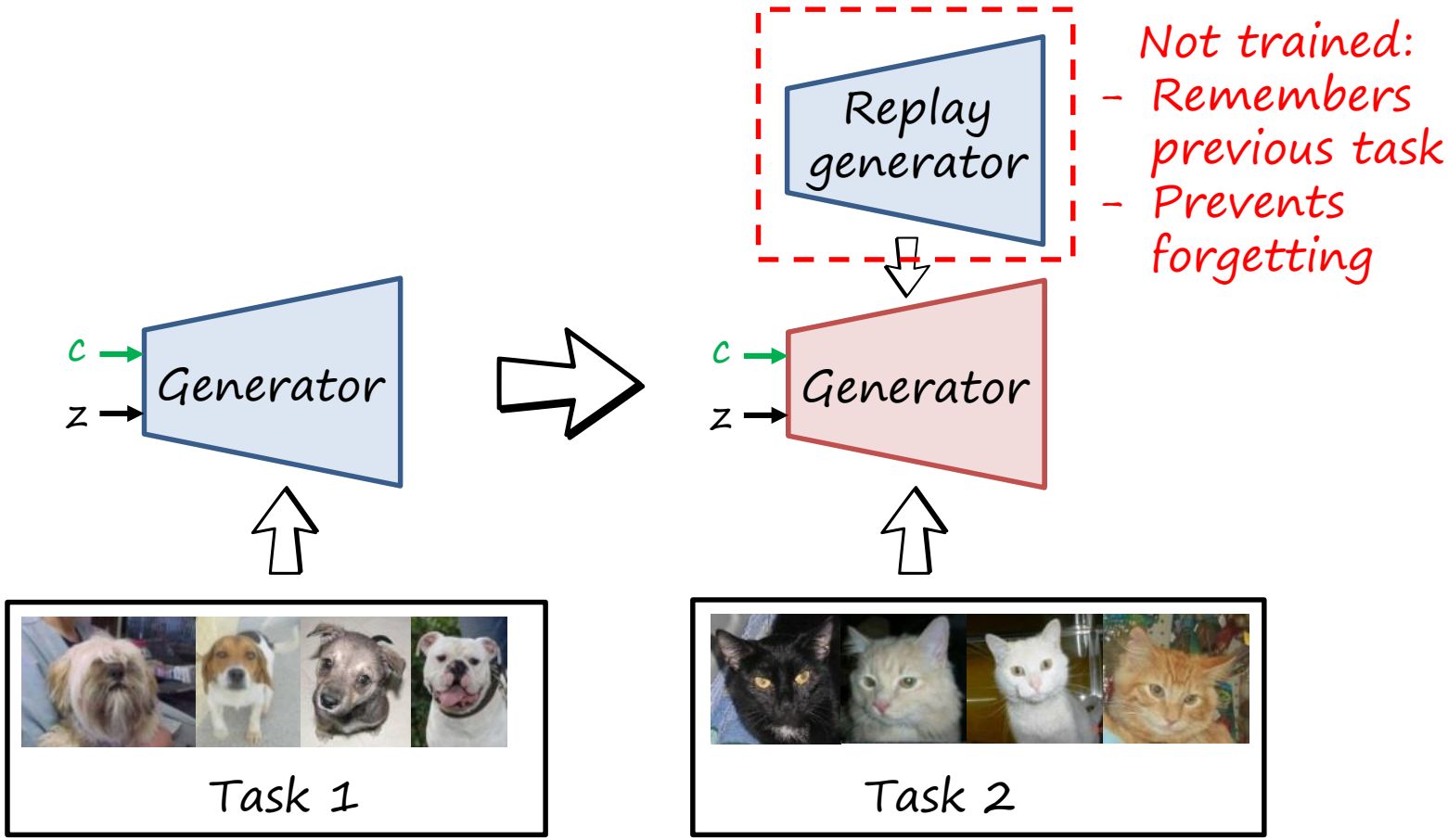
LSUN 4 categories (4 tasks)

MNIST 10 categories (10 tasks)

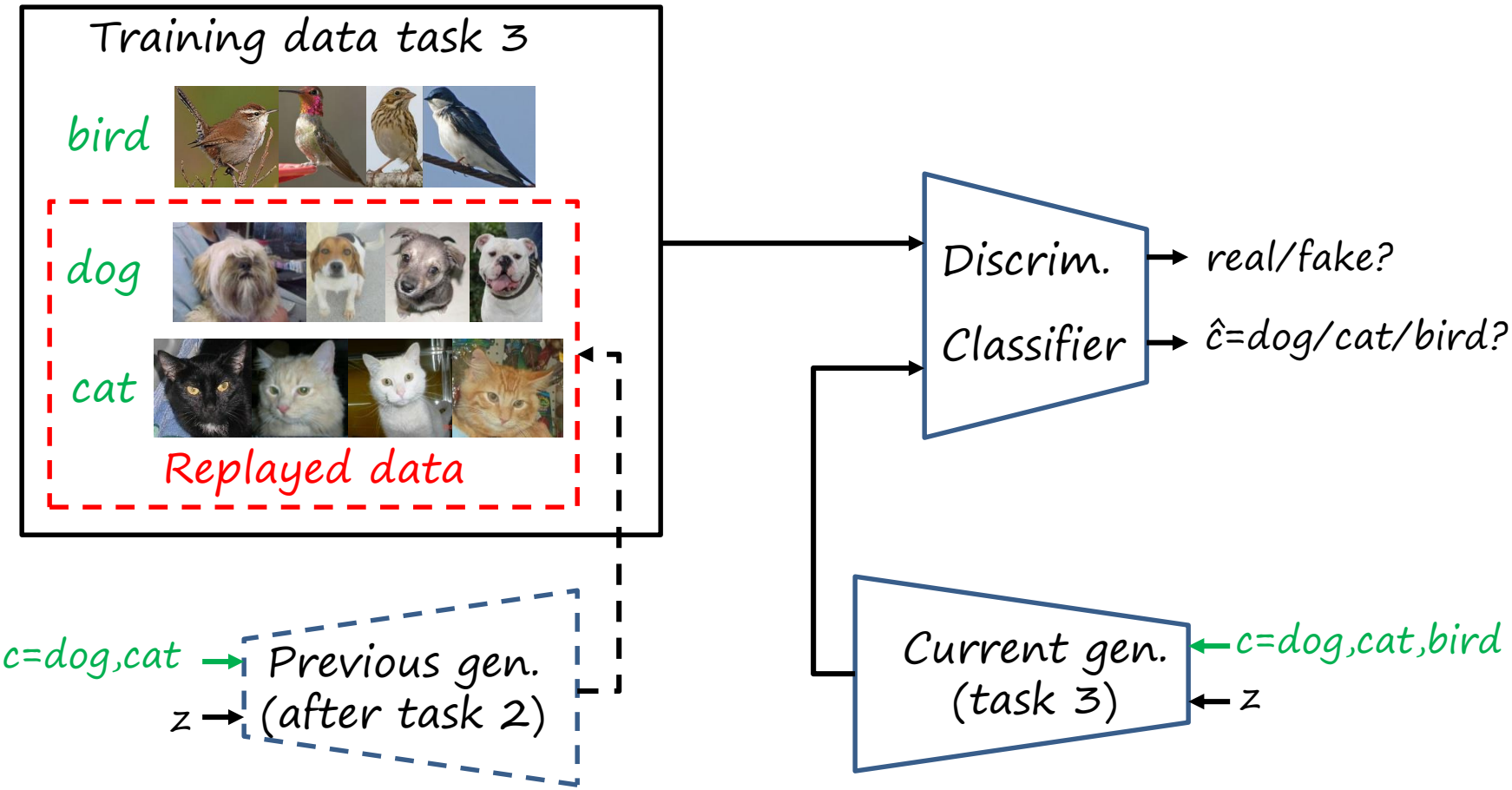


After task 10

Memory Replay GANs (MeRGANs)



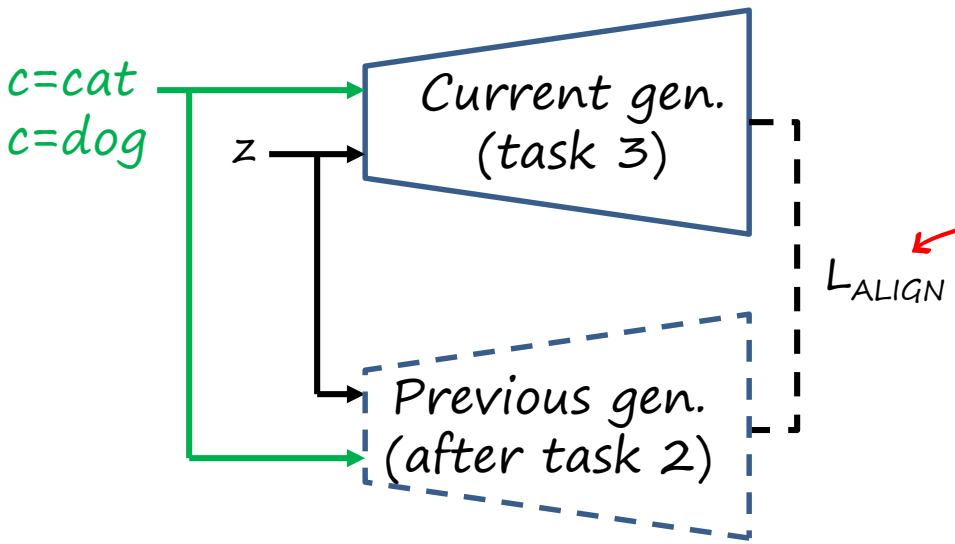
MeRGAN-JTR: joint training w/ replay



Step 1: replay previous tasks

Step 2: joint training

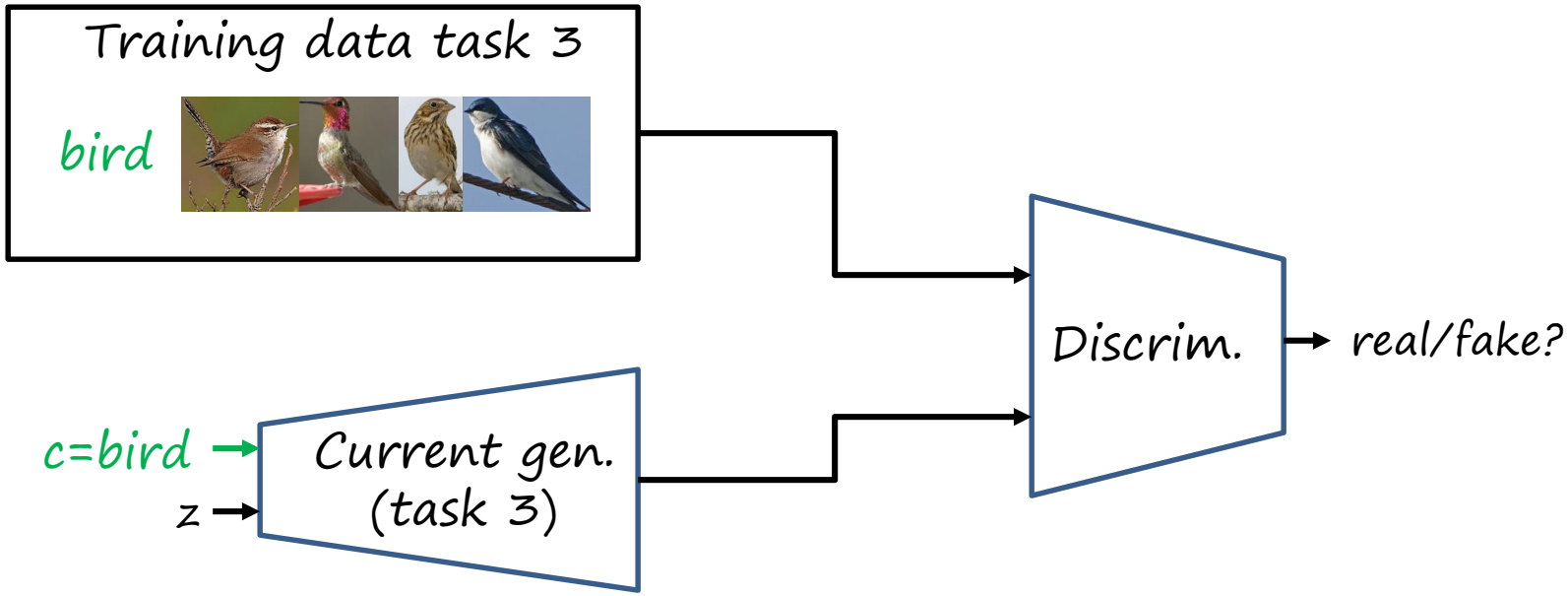
MeRGAN-RA: replay alignment



We can do pixelwise alignment because for given z and c output is deterministic (thanks conditional GAN)

Step 1: replay previous tasks and align

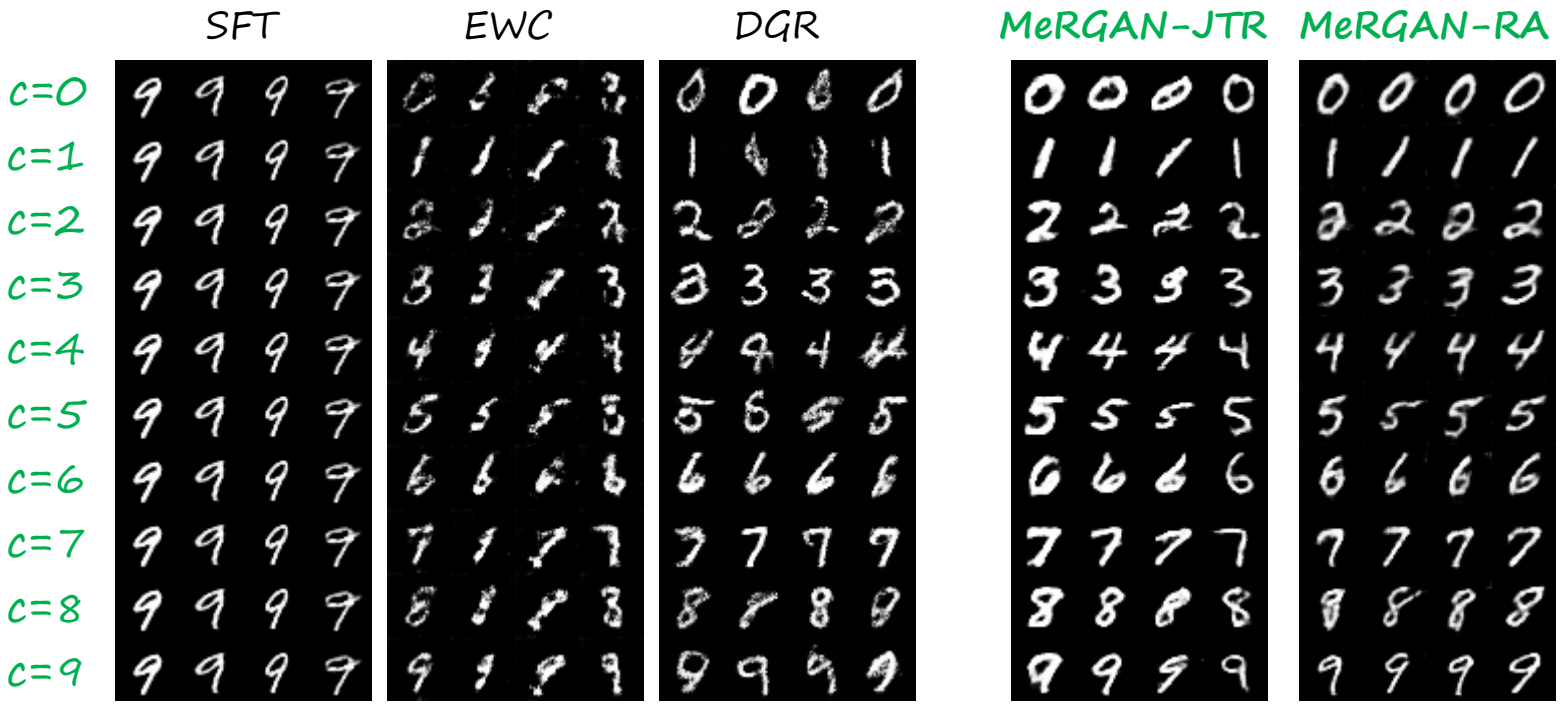
MeRGAN-RA: replay alignment



Step 2: learning new task

Digit generation (MNIST)

10 tasks (10 categories) 3 layers generator



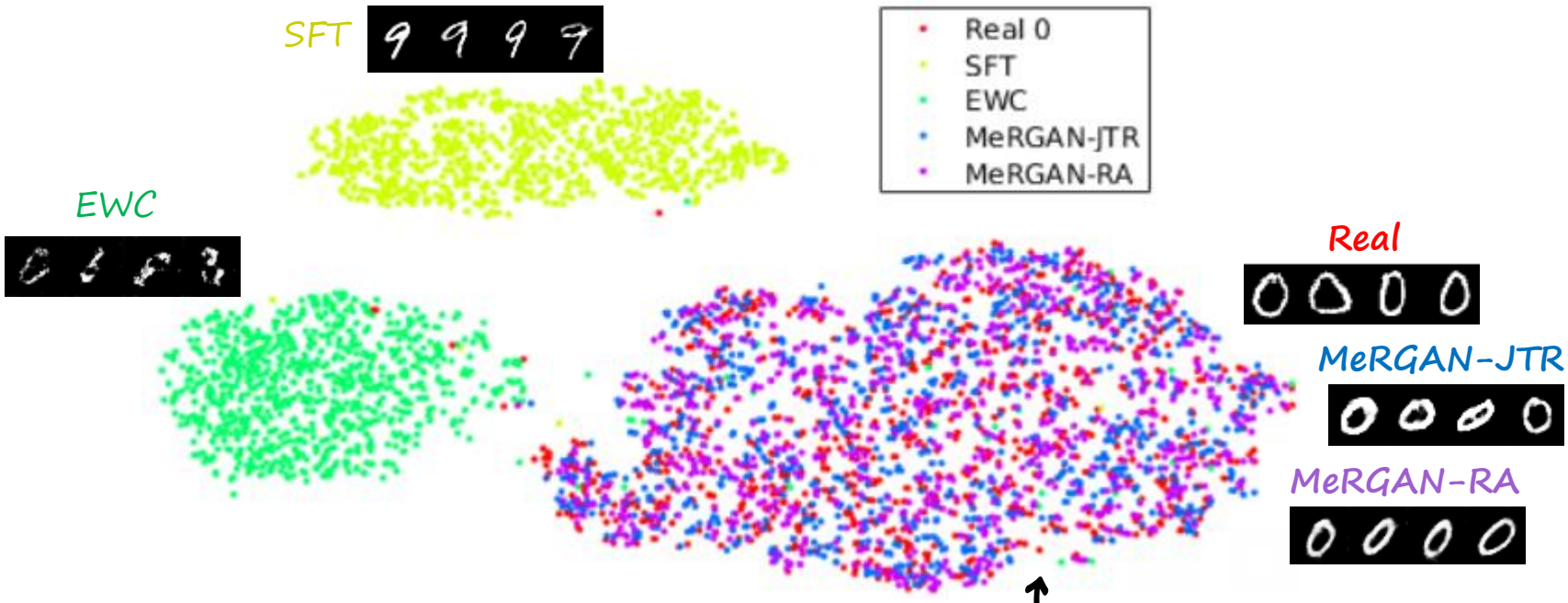
SFT: Sequential fine tuning **EWC**: GAN with EWC [arxiv17] **DGR**: Deep generative replay [NIPS17]

Scene generation (LSUN)

4 tasks (4 categories) 18-layer ResNet generator

t-SNE visualizations (MNIST)

Generating digit 0 (i.e. first task) after learning 10 tasks

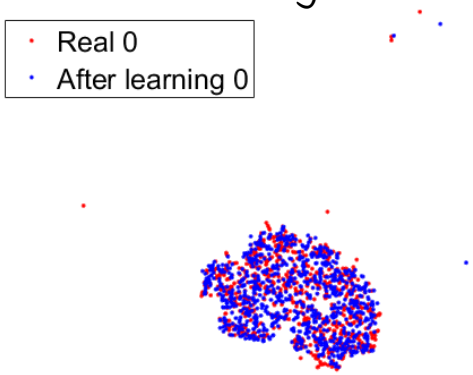


Samples from MeRGANs overlap with real data

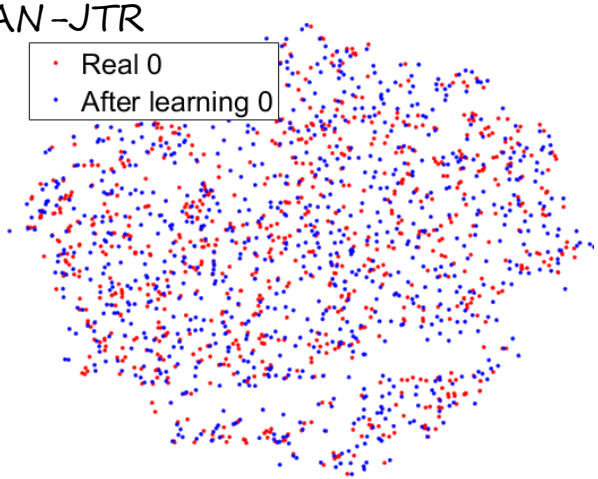
Learning and forgetting in t-SNE (MNIST)

Generating digit 0 (i.e. task 1)

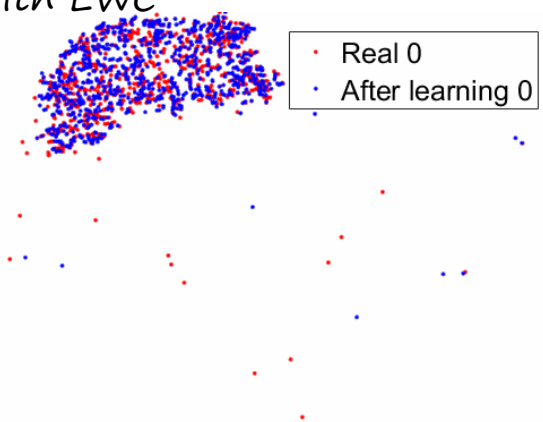
Sequential fine tuning



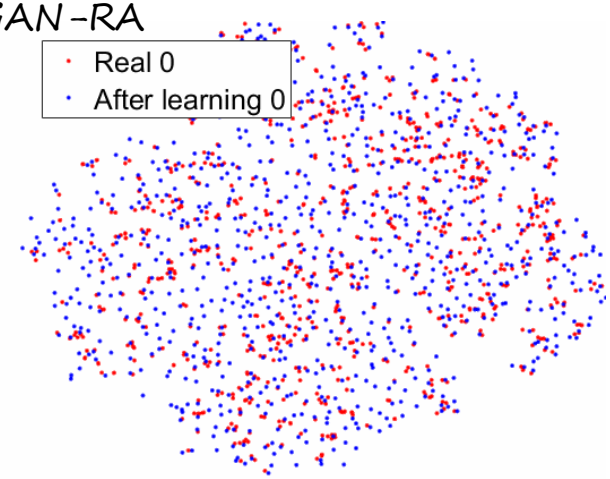
MeRGAN-JTR



GAN with EWC



MeRGAN-RA



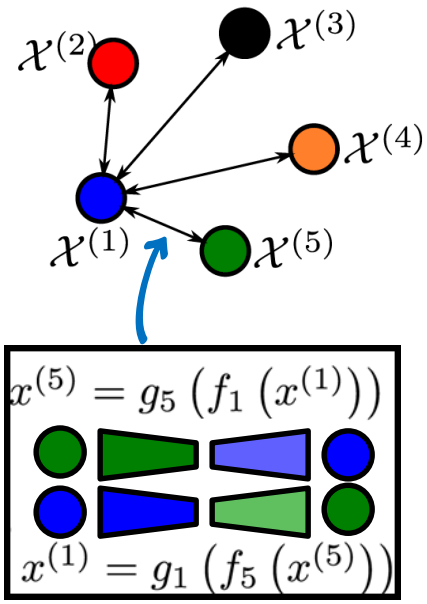
Outline

- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)

Unseen image-to-image translations

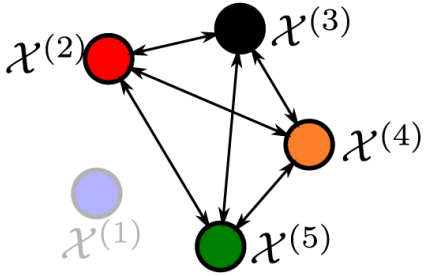
Only these translations are trained (seen)

Train



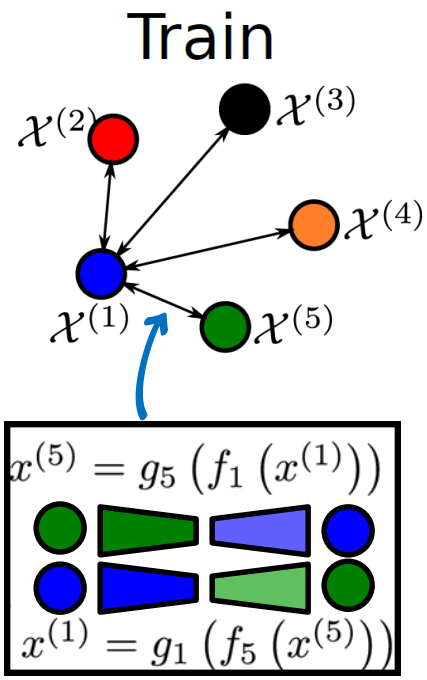
*Evaluate on these **unseen** translations*

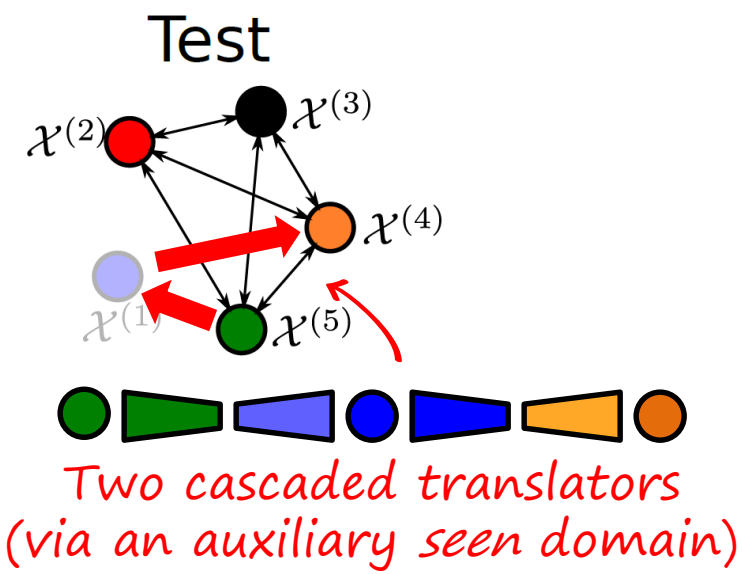
Test



Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

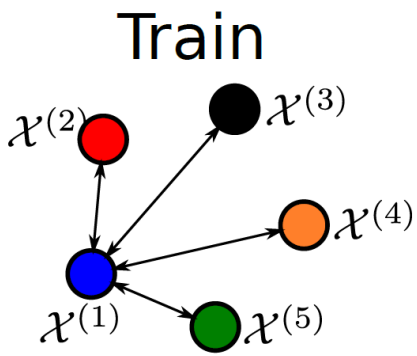
Cascading image-to-image translators

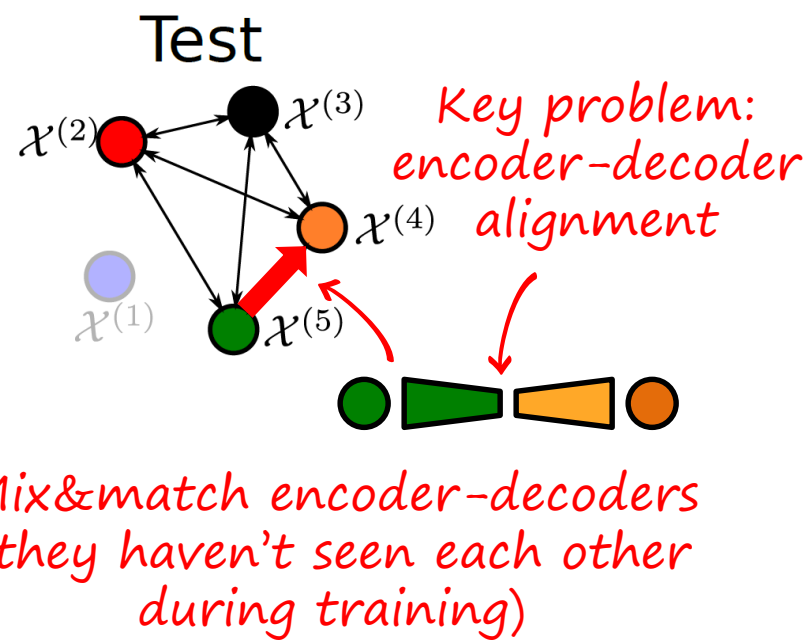




Wang, van de Weijer, Herranz, “Encoder-decoder alignment for zero-pair image-to-image translation”, CVPR 2018

Mix and match networks

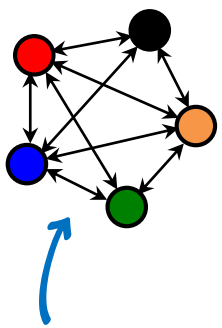


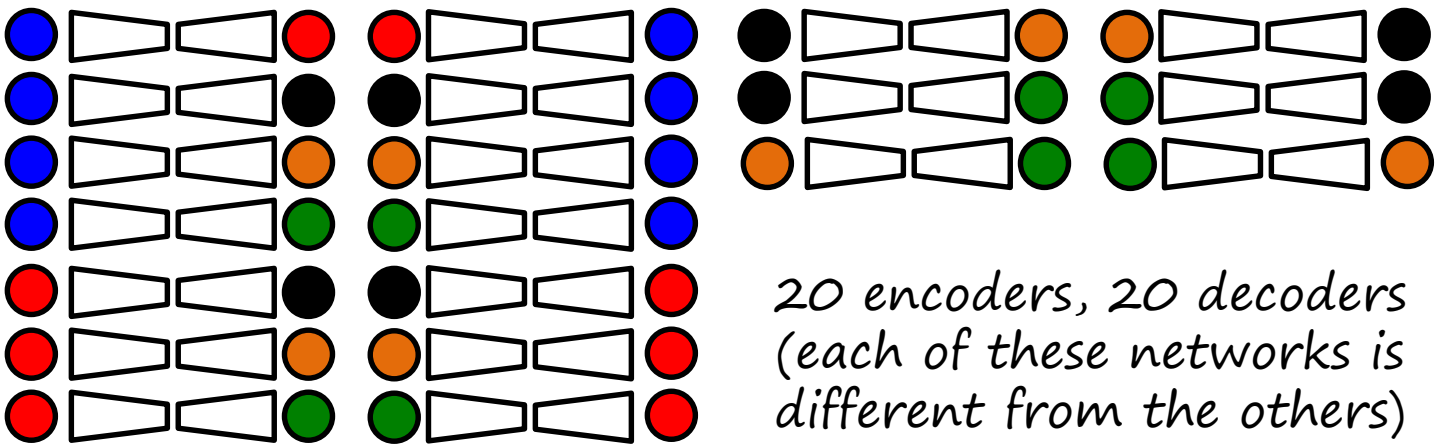


Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

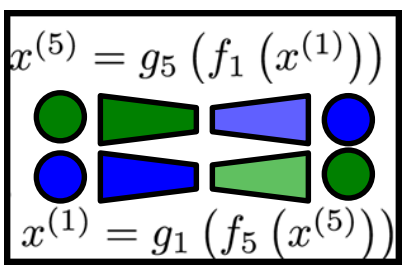
Application: many-to-many translations

We could train all possible translators
Problems:
- No sharing
- Poor scalability: number of networks $O(N^2)$





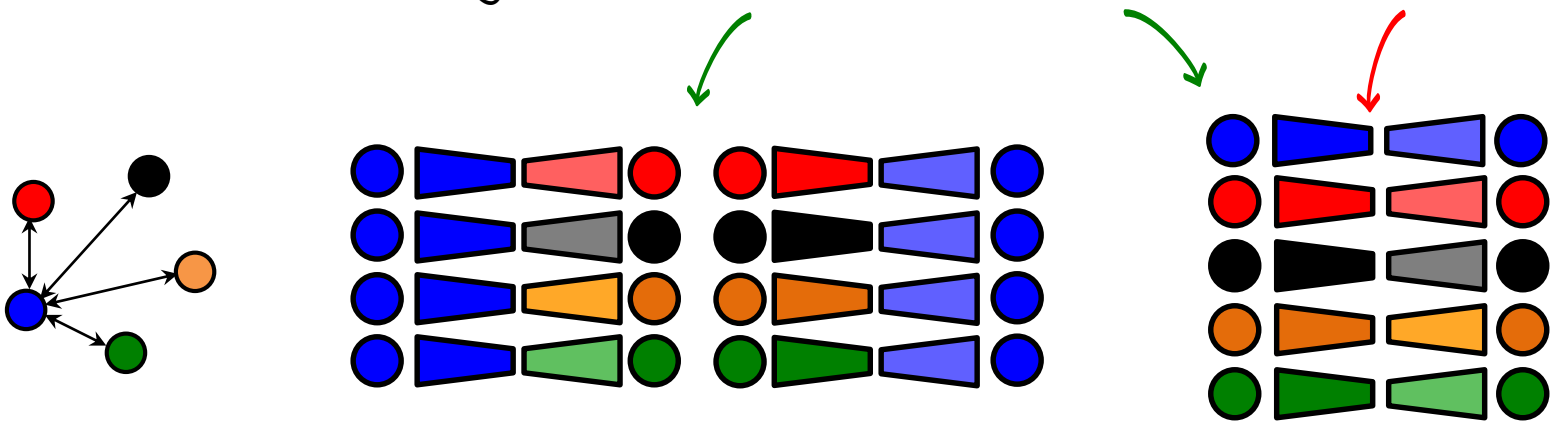
20 encoders, 20 decoders
(each of these networks is different from the others)



Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Mix and match networks

Unseen encoder-decoder alignment
- Latent representation should be domain-independent
- Achieved using shared encoder/decoders and autoencoders



5 encoders, 5 decoders

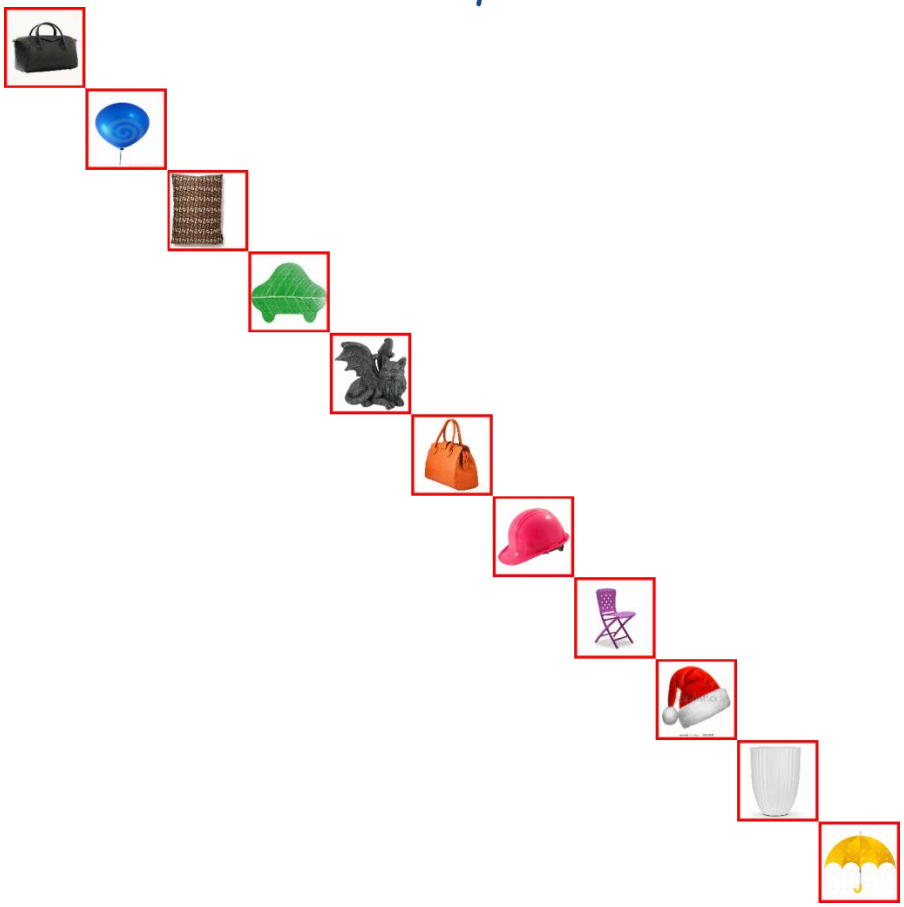
- Scalable: number of networks O(N)

Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Example: scalable recolorization

Input

Unpaired translation
11 colors (i.e. 11 domains)

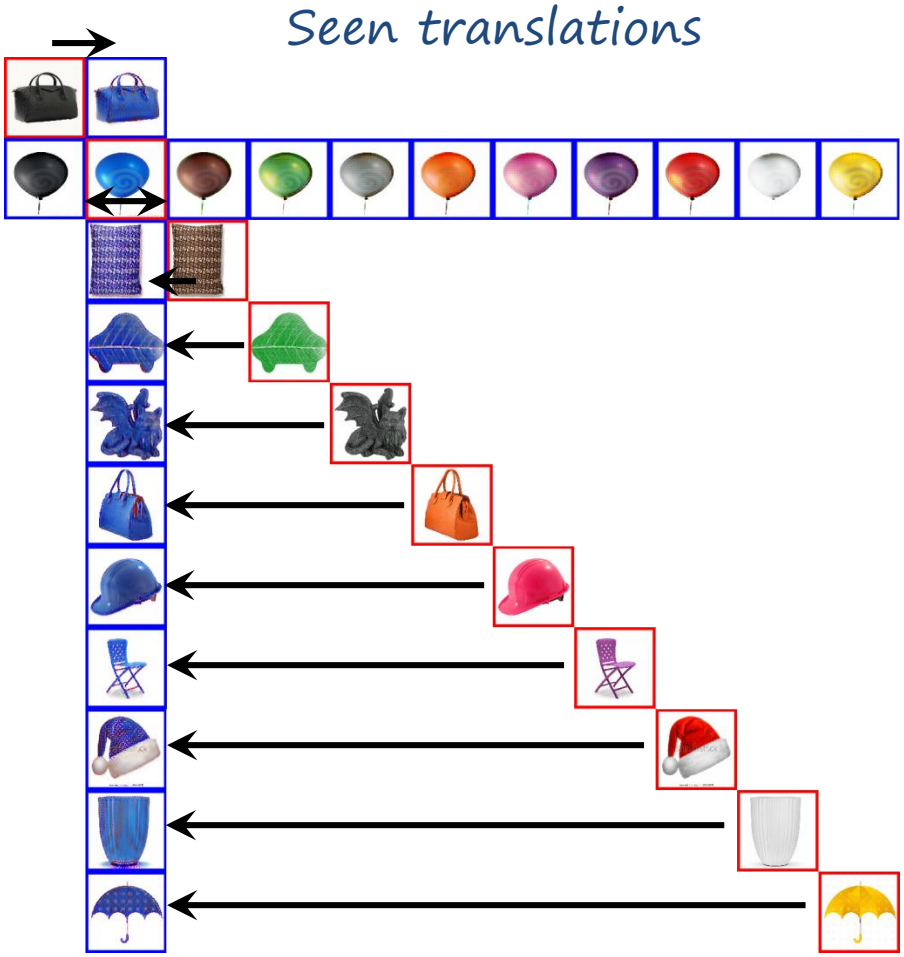


Wang, van de Weijer, Herranz, “Encoder-decoder alignment for zero-pair image-to-image translation”, CVPR 2018

Example: scalable recolorization

Unpaired translation
11 colors (i.e. 11 domains)

Requires training 11 encoders and 11 decoders



Wang, van de Weijer, Herranz, “Encoder-decoder alignment for zero-pair image-to-image translation”, CVPR 2018

Example: scalable recolorization

Unpaired translation
11 colors (i.e. 11 domains)

Requires training 11 encoders and 11 decoders

CycleGANs for all combinations would require 55 encoders and 55 decoders

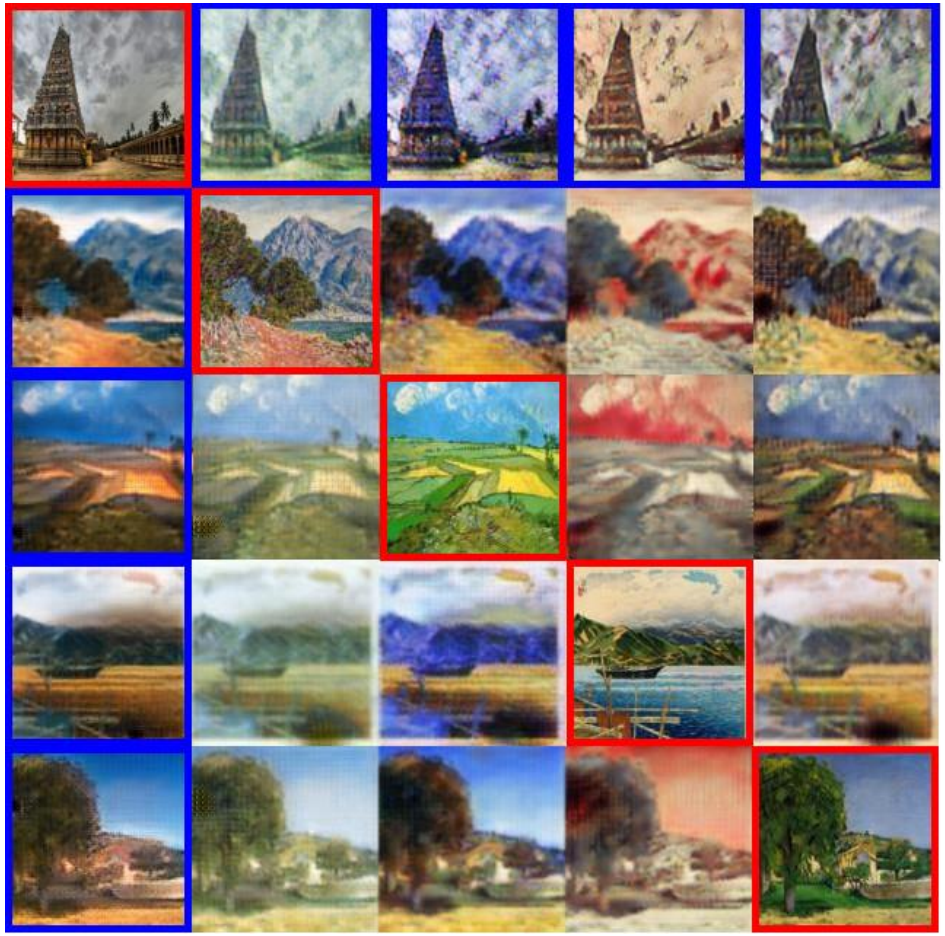
Seen+unseen translations

Wang, van de Weijer, Herranz, “Encoder-decoder alignment for zero-pair image-to-image translation”, CVPR 2018

Example: scalable style transfer

Unpaired translation
Five domains
(photo, Monet, van Gogh, Ukiyo-e, Cezanne)

(5 encoders and 5 decoders)



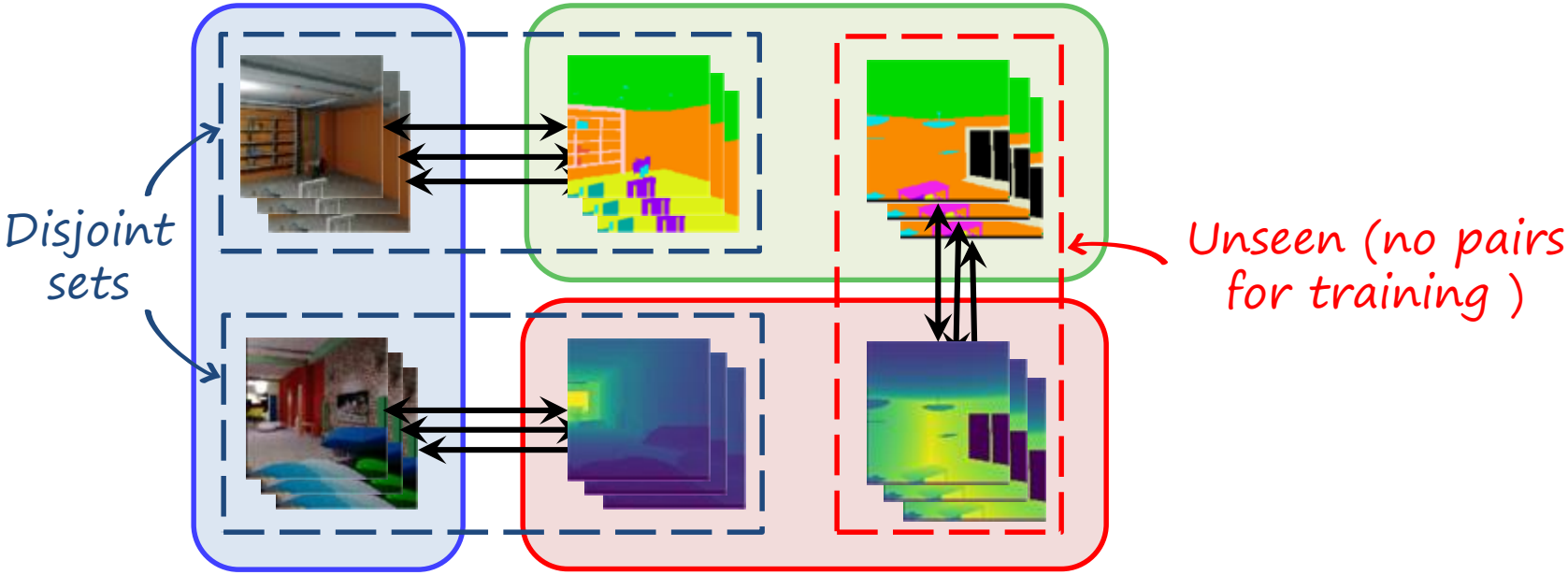
Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Application: zero-pair translation

Cross-modal translation setting (RGB, segmentation and depth)

Paired data available for (RGB, segm.) and (RGB, depth)

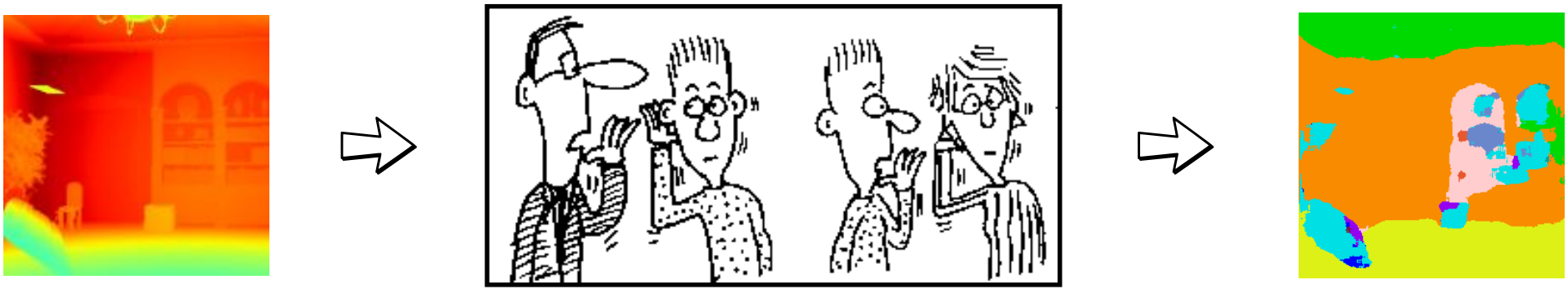
Evaluate on the **unseen zero-pair translations** (depth, segm.)



Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Zero-pair translation with two cascaded pix2pix (paired translations)

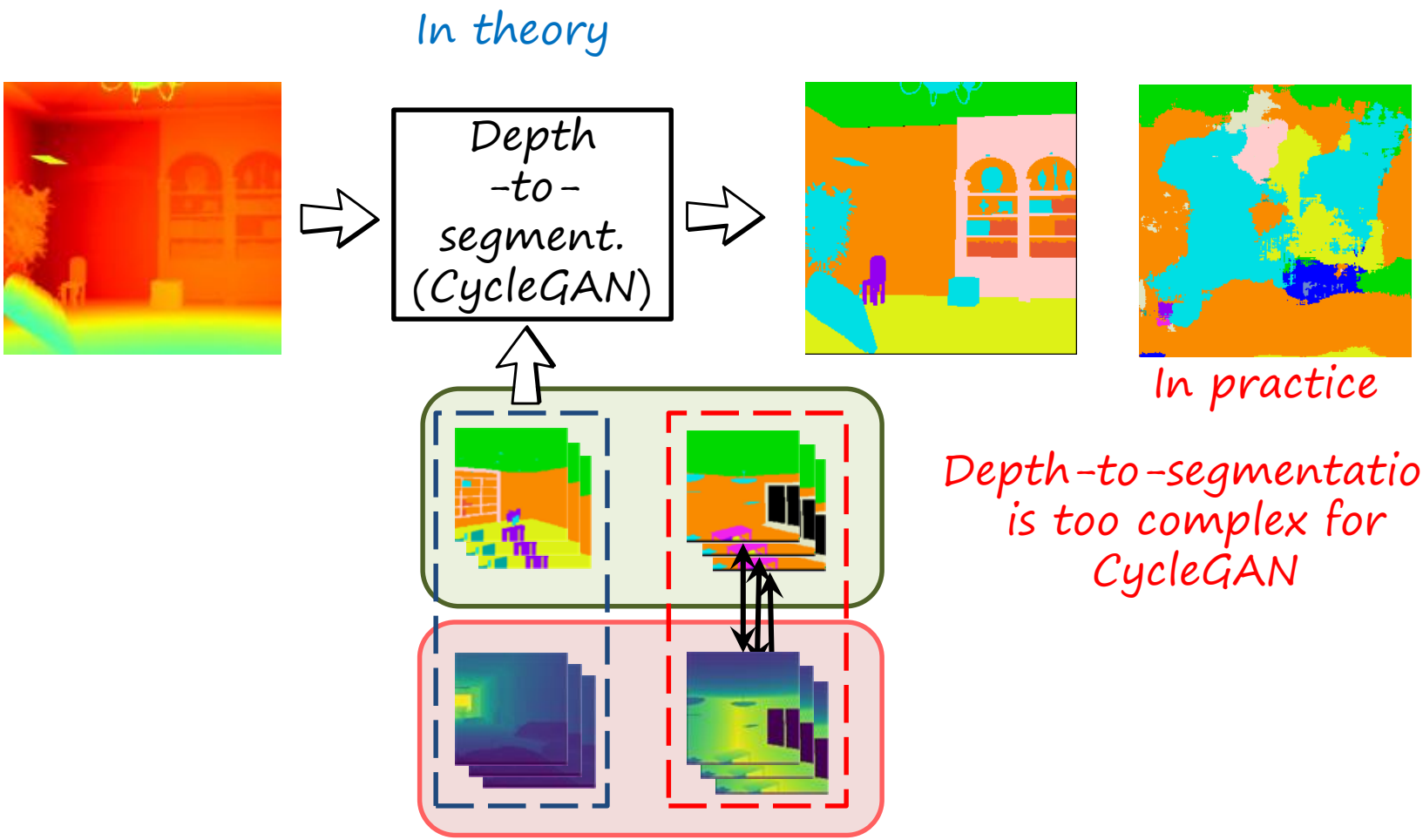
In theory



In practice

Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

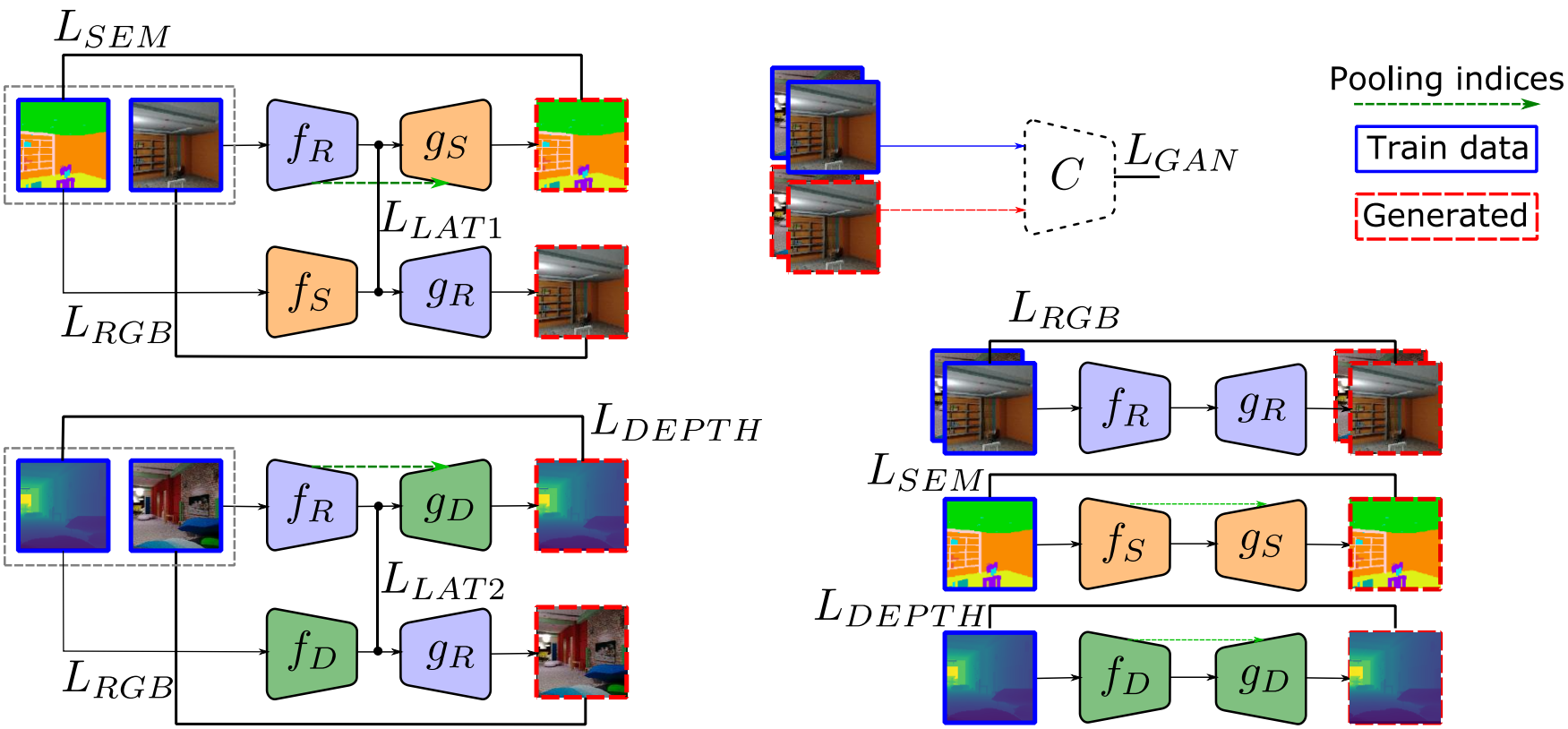
Zero-pair translation with CycleGAN (unpaired translation)



Depth-to-segmentation is too complex for CycleGAN

Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Zero-pair translation with mix and match networks

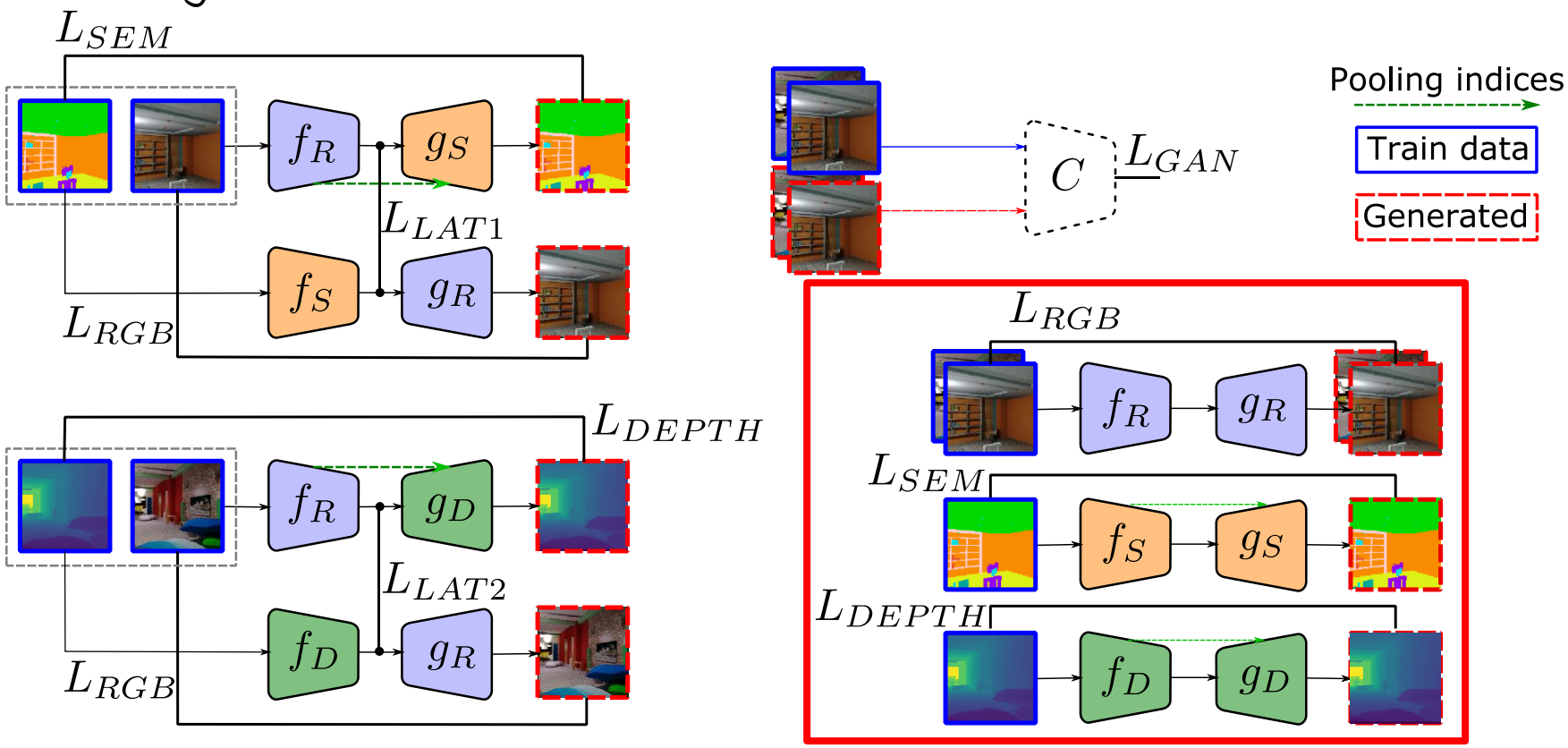


Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Zero-pair translation with mix and match networks

Training for encoder-decoder alignment:

Autoencoders



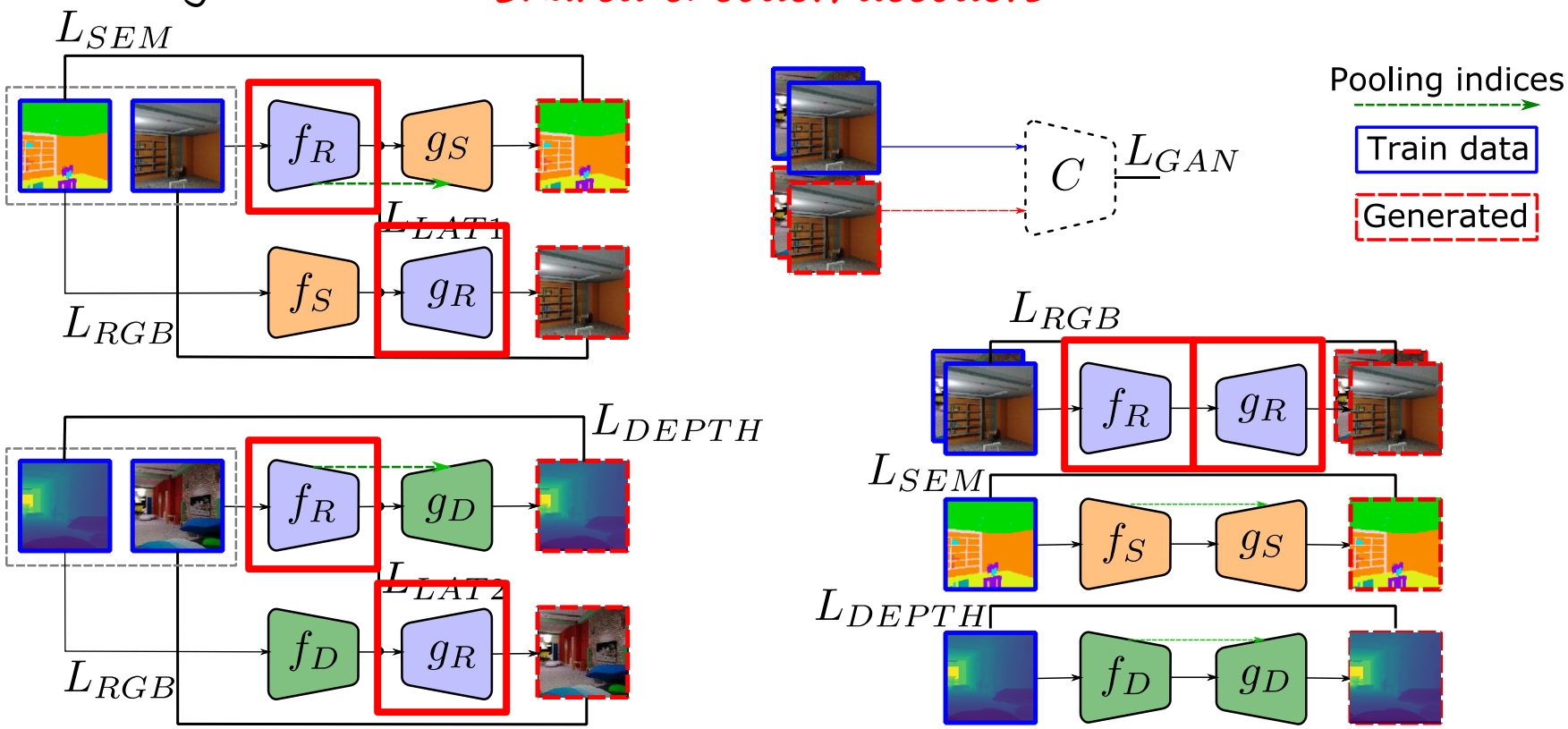
Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Zero-pair translation with mix and match networks

Training for encoder-decoder alignment:

Autoencoders

Shared encoder/decoders



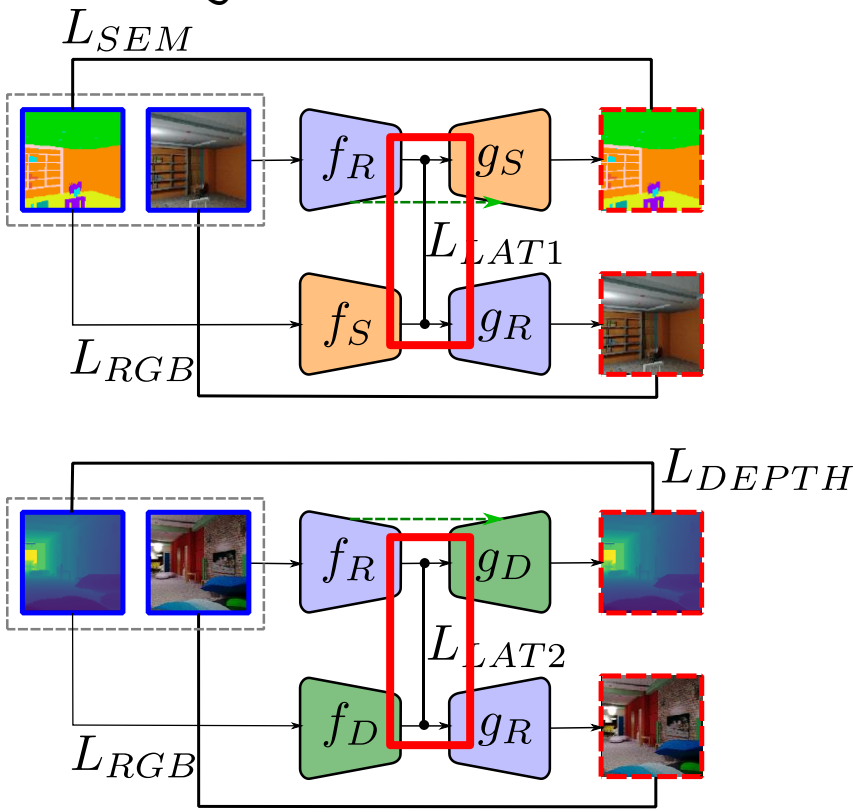
Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

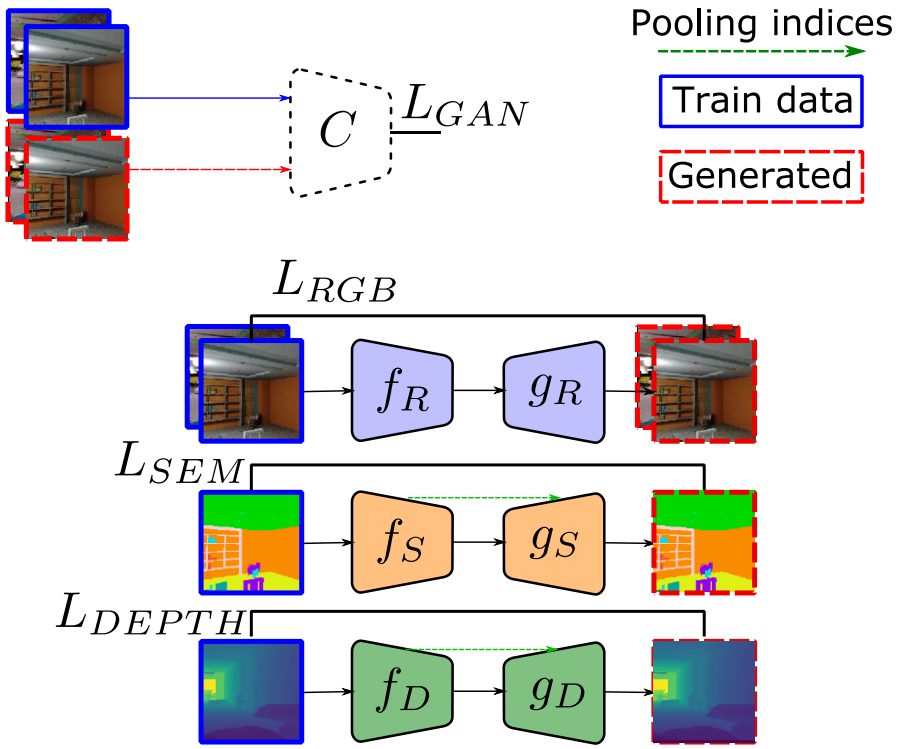
Zero-pair translation with mix and match networks

Training for encoder-decoder alignment:

Autoencoders
Shared encoder/decoders

Latent losses





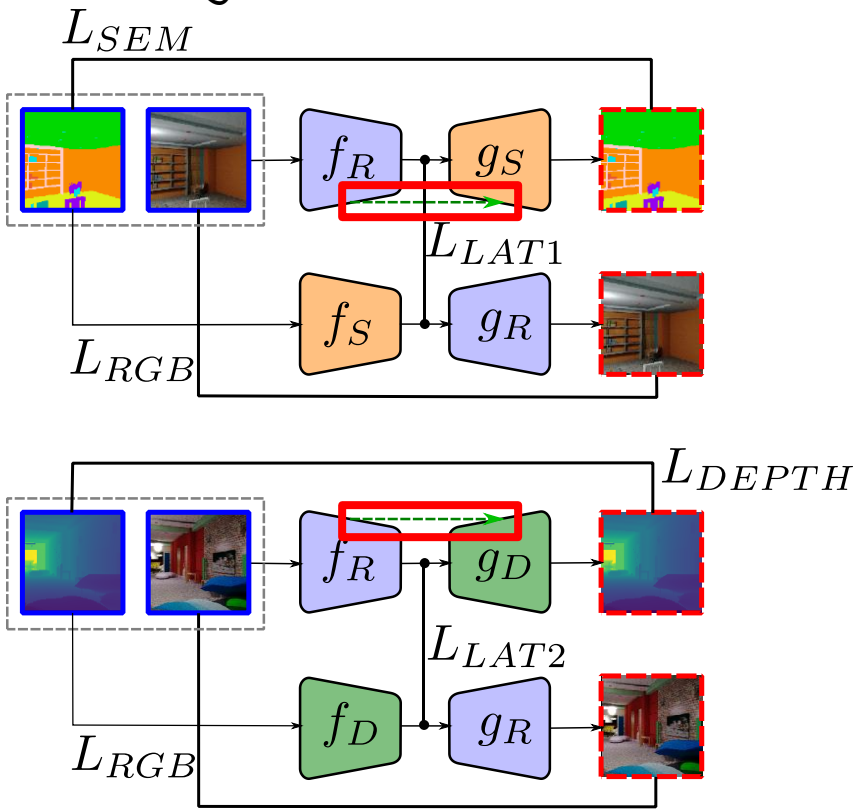
Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

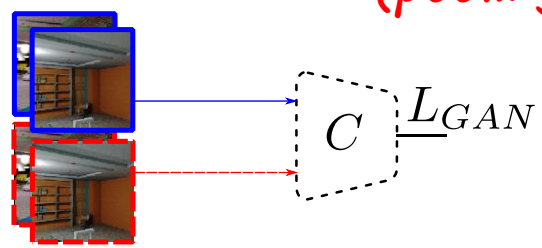
Zero-pair translation with mix and match networks

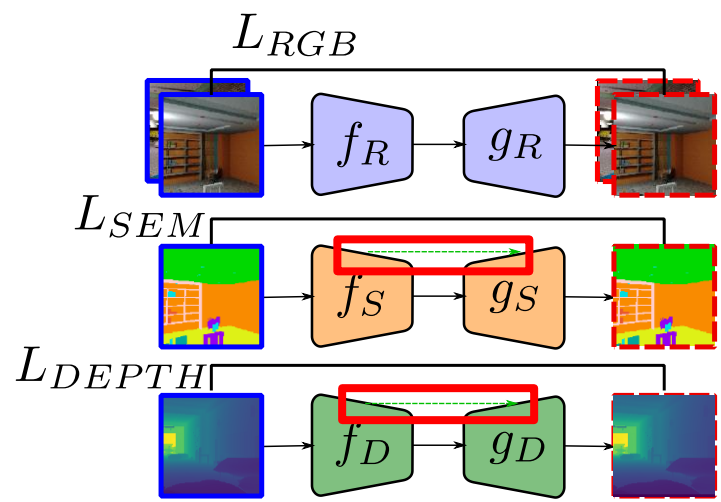
Training for encoder-decoder alignment:

Autoencoders
Shared encoder/decoders

Latent losses
Robust side information (pooling indices)







Pooling indices

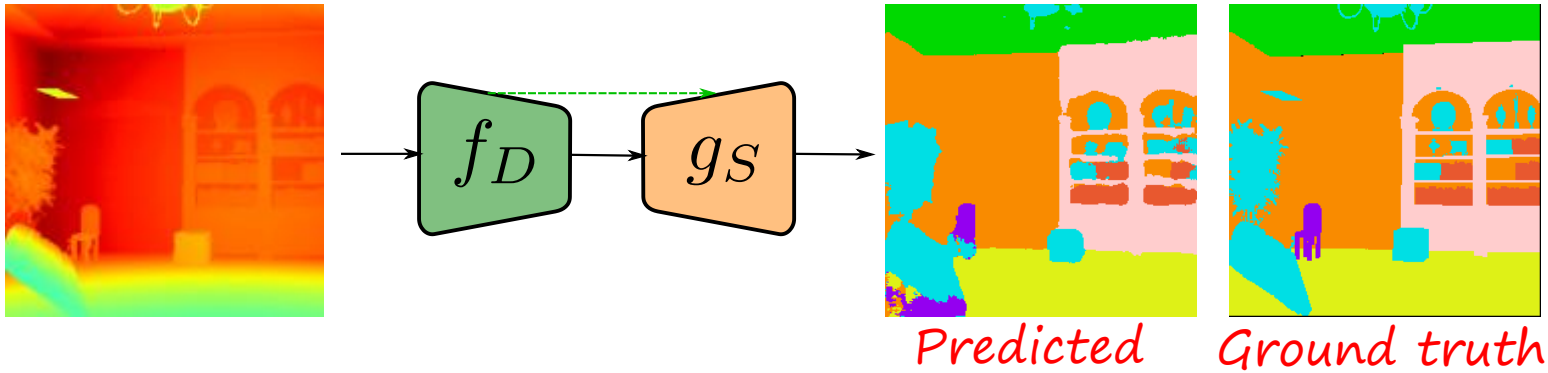
Train data

Generated

Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

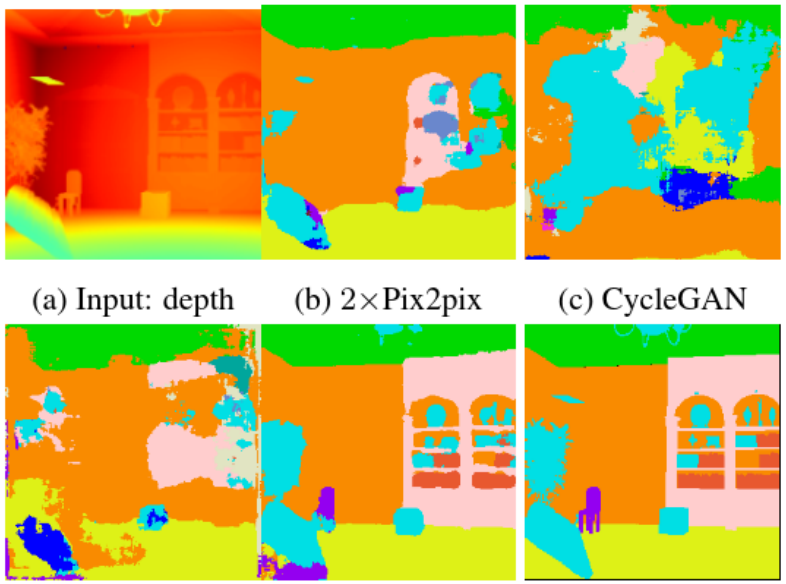
Zero-pair translation with mix and match networks

Test on zero-pair translation depth-to-segmentation



Wang, van de Weijer, Herranz, “Encoder-decoder alignment for zero-pair image-to-image translation”, CVPR 2018

Comparison: depth-to-segmentation



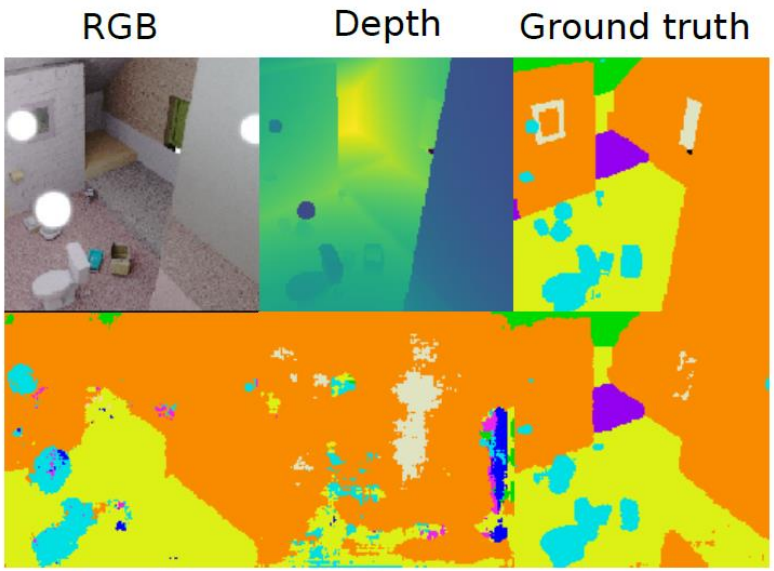
(a) Input: depth (b) 2×Pix2pix (c) CycleGAN

(d) D→R→S (e) Proposed (f) Ground truth

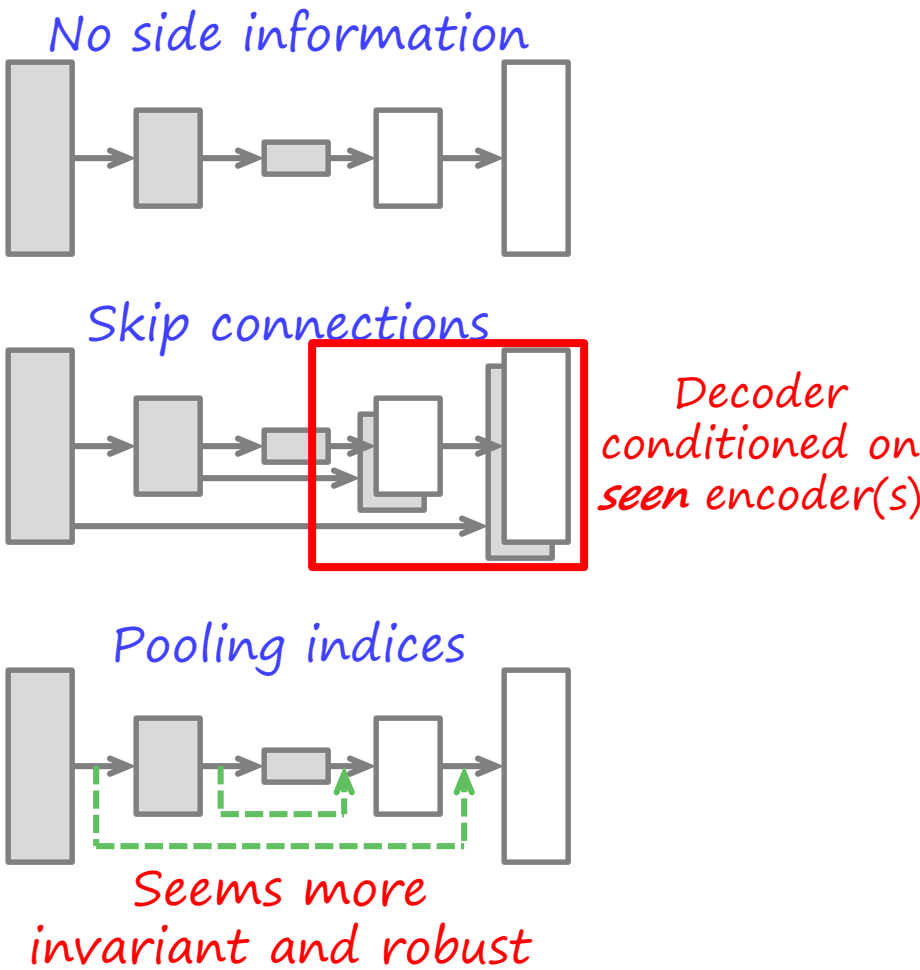
Figure 1: Zero-pair depth→segmentation, trained on (depth,RGB) and (RGB,segmentation).

Wang, van de Weijer, Herranz, “Encoder-decoder alignment for zero-pair image-to-image translation”, CVPR 2018

Side information in mix and match networks



Side information	Pretrained	mIoU	Global
-	N	32.2%	63.5%
Skip connections	N	14.1%	52.6%
Pooling indices	N	45.6%	73.4%
Pooling indices	Y	49.5%	80.0%



Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

Quantitative evaluation

Method	Conn.	L_{SEM}	Bed	Book	Ceiling	Chair	Floor	Furniture	Object	Picture	Sofa	Table	TV	Wall	Window	mIoU	Global
Baselines																	
CycleGAN [34]	SC	CE	2.79	0.00	16.9	6.81	4.48	0.92	7.43	0.57	9.48	0.92	0.31	17.4	15.1	6.34	14.2
2×pix2pix [10]	SC	CE	34.6	1.88	70.9	20.9	63.6	17.6	14.1	0.03	38.4	10.0	4.33	67.7	20.5	25.4	57.6
M&MNet $D \to R \to S$	PI	CE	0.02	0.00	8.76	0.10	2.91	2.06	1.65	0.19	0.02	0.28	0.02	58.2	3.3	5.96	32.3
M&MNet $D \to R \to S$	SC	CE	25.4	0.26	82.7	0.44	56.6	6.30	23.6	5.42	0.54	21.9	10.0	68.6	19.6	24.7	59.7
Zero-pair																	
M&MNet $D \to S$	PI	CE	50.8	18.9	89.8	31.6	88.7	48.3	44.9	62.1	17.8	49.9	51.9	86.2	79.2	55.4	80.4
Multi-modal																	
M&MNet $(R, D) \to S$	PI	CE	49.9	25.5	88.2	31.8	86.8	56.0	45.4	70.5	17.4	46.2	57.3	87.9	79.8	57.1	81.2

Table 3: Zero-pair depth-to-semantic segmentation. **SC**: skip connections, **PI**: pooling indexes, **CE**: cross-entropy

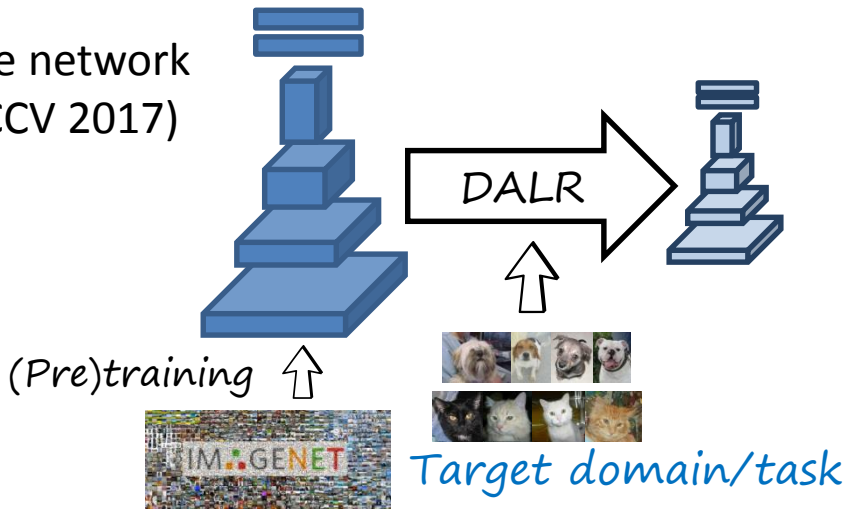
Wang, van de Weijer, Herranz, "Encoder-decoder alignment for zero-pair image-to-image translation", CVPR 2018

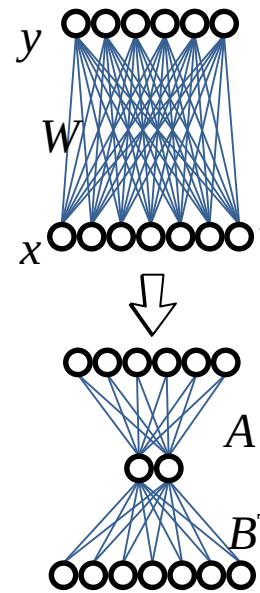
Outline

- Introduction
- Transferring GANs (ECCV 2018)
- Rotated elastic weight consolidation (ICPR 2018)
- Memory Replay GANs (NIPS 2018)
- Mix and match networks (CVPR 2018)
- Other works

Other works

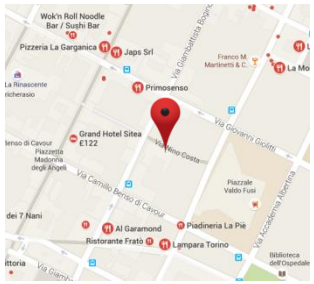
Domain-adaptive network compression (ICCV 2017)



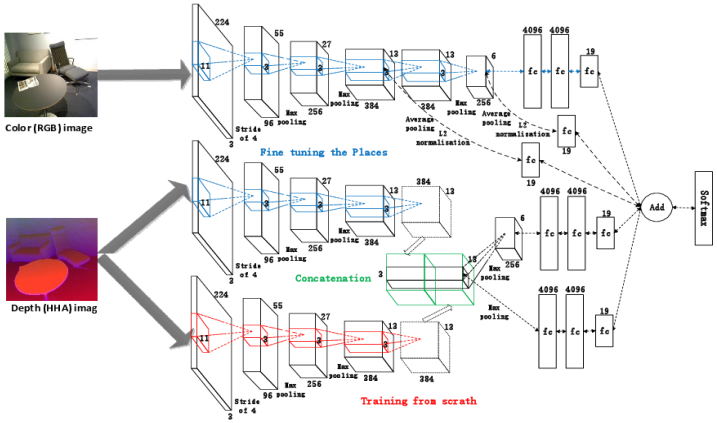


Domain-adaptive low rank (DALR) matrix decomposition

Contextual food recognition and analysis (TMM17,TMM18)



RGB-D deep representations (AAAI17,IJCAI17,TIP18)



THANK YOU!

lherranz@cvc.uab.es
www.lherranz.org