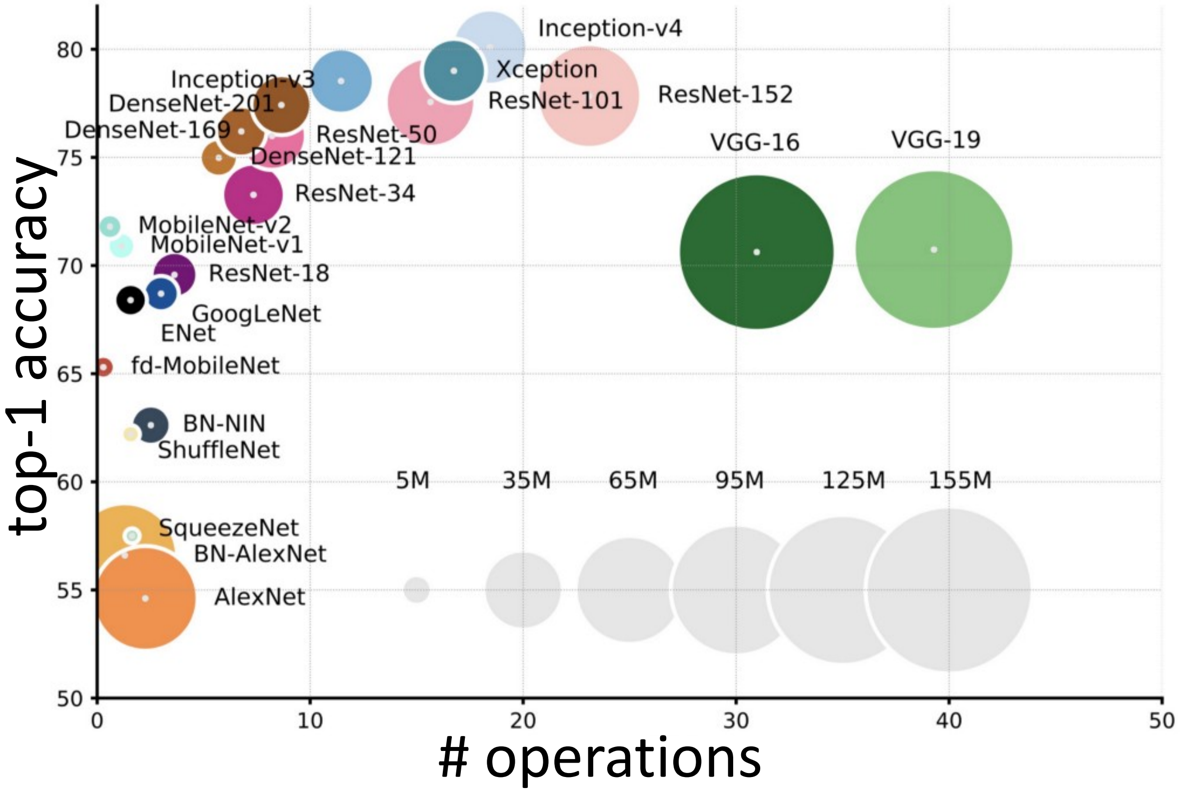
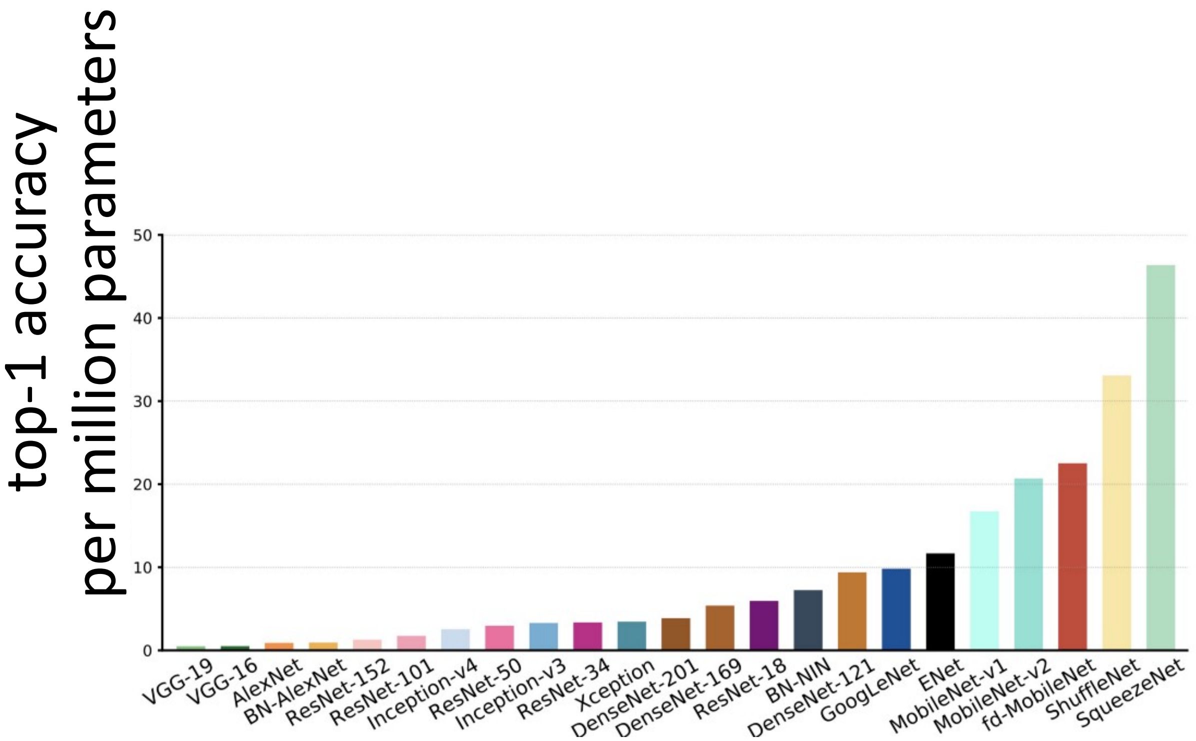
Common Architecture Elements

Classification, Segmentation, Detection

ImageNet classification performance

(for up-to-date top-performers see leaderboards of datasets like ImageNet or COCO)





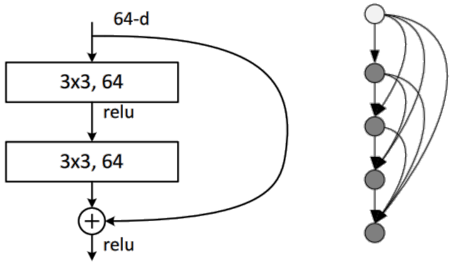
Images from: Canziani et al., *An Analysis of Deep Neural Network Models for Practical Applications*, arXiv 2017

Blog: https://towardsdatascience.com/neural-network-architectures-156e5bad51ba

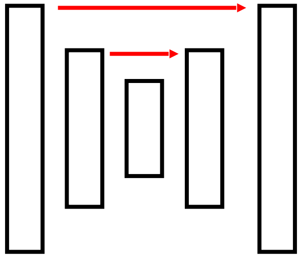
Architecture Elements

Some notable architecture elements shared by many successful architectures:

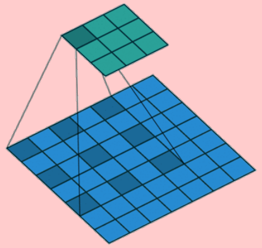
Residual Blocks
and Dense Blocks



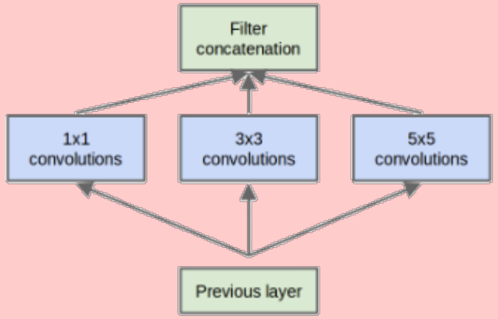
Skip Connections
(UNet)



Dilated
Convolutions



Grouped
Convolutions



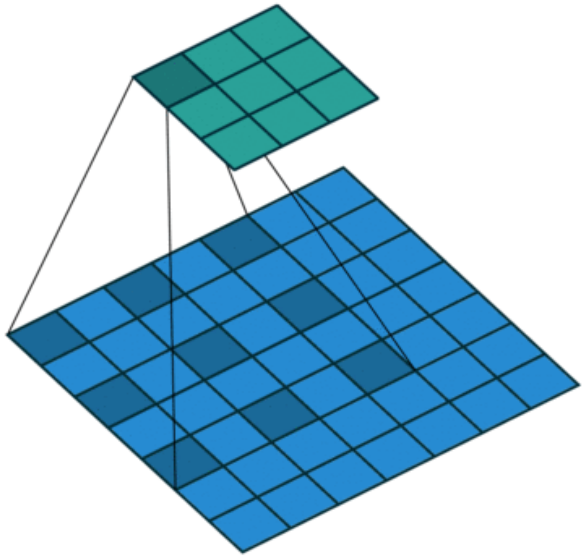
Attention
(Spatial and over Channels)

Dilated (Atrous) Convolutions

Problem: increasing the receptive field costs a lots of parameters.

Idea: spread out the samples used in each convolution.

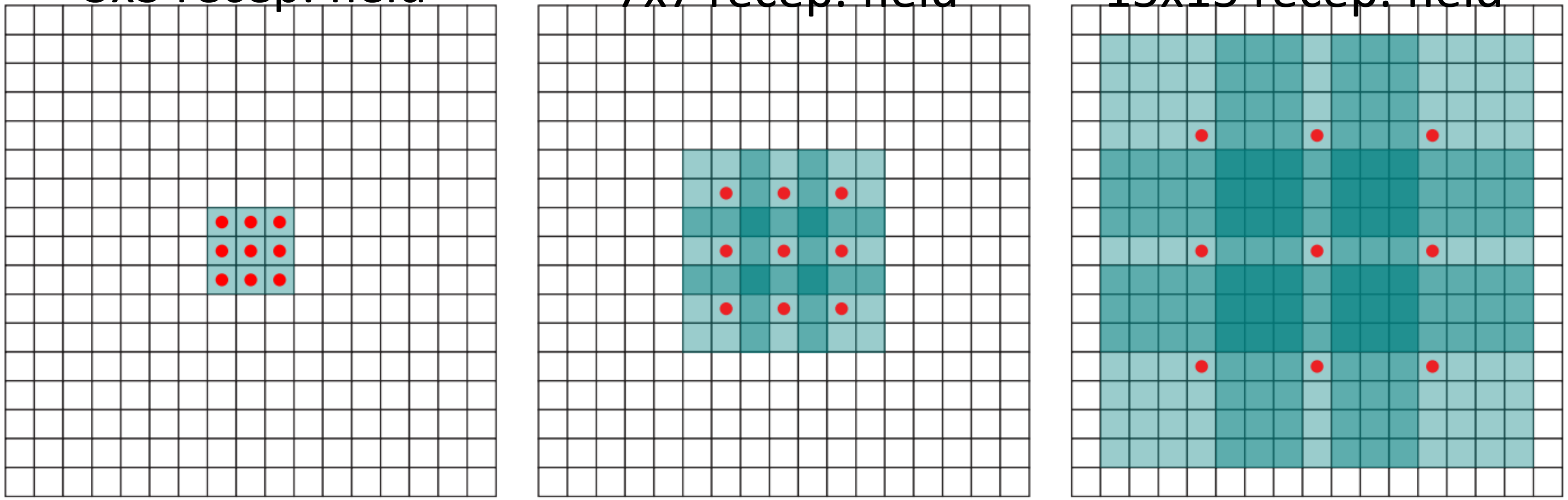
dilated convolution



1st layer: not dilated
3x3 recep. field

2nd layer: 1-dilated
7x7 recep. field

3rd layer: 2-dilated
15x15 recep. field



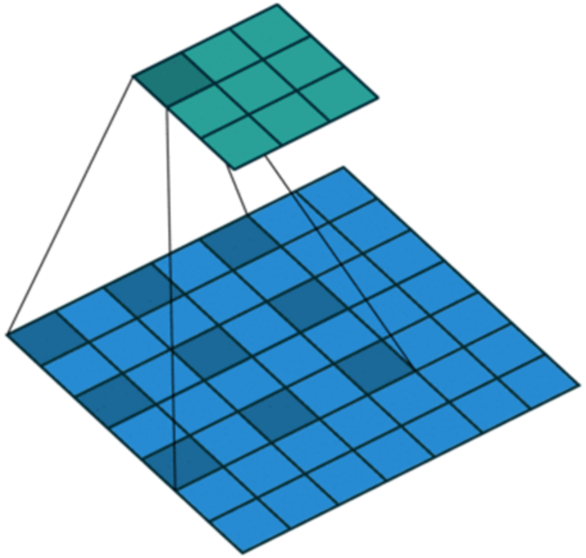
Images from: Dumoulin and Visin, *A guide to convolution arithmetic for deep learning*, arXiv 2016
Yu and Koltun, *Multi-scale Context Aggregation by Dilated Convolutions*, ICLR 2016

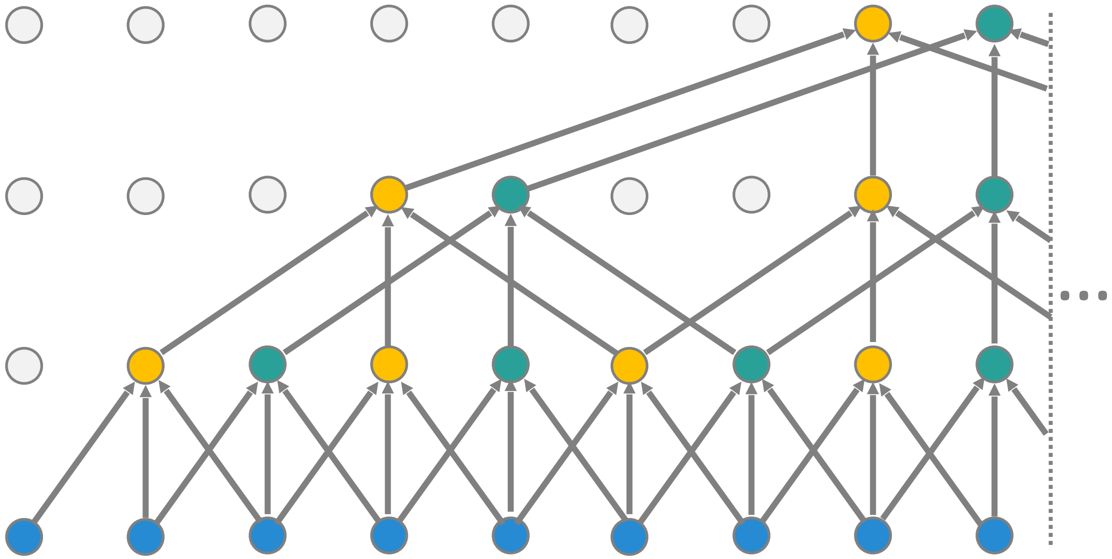
Dilated (Atrous) Convolutions

Problem: increasing the receptive field costs a lots of parameters.

Idea: spread out the samples used for a convolution.

dilated convolution





3rd layer: 2-dilated
15x15 recep. field

2nd layer: 1-dilated
7x7 recep. field

1st layer: not dilated
3x3 recep. field

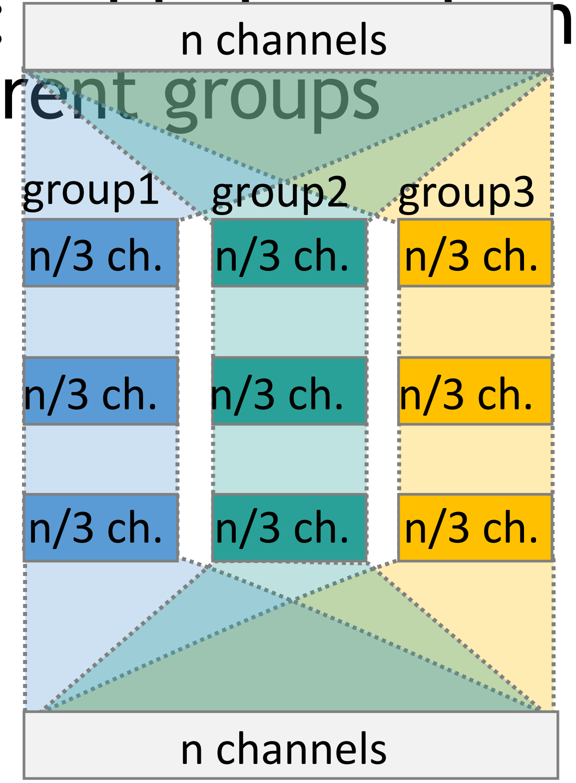
Input image

Dumoulin and Visin, *A guide to convolution arithmetic for deep learning*, arXiv 2016

Grouped Convolutions (Inception Modules)

Problem: conv. parameters grow quadratically in the number of channels

Idea: split channels into groups, remove connections between different groups



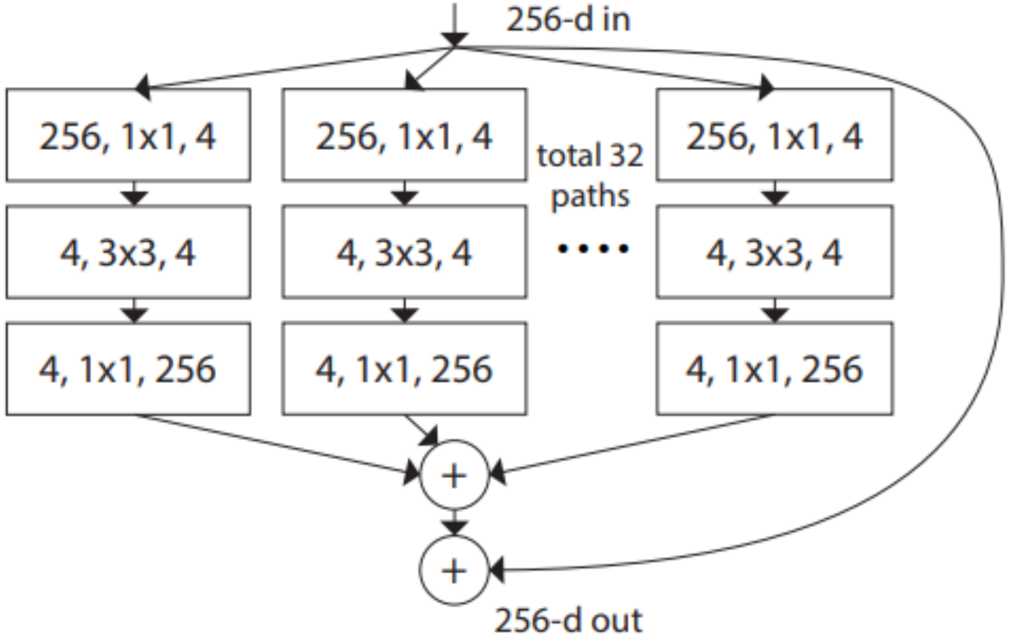
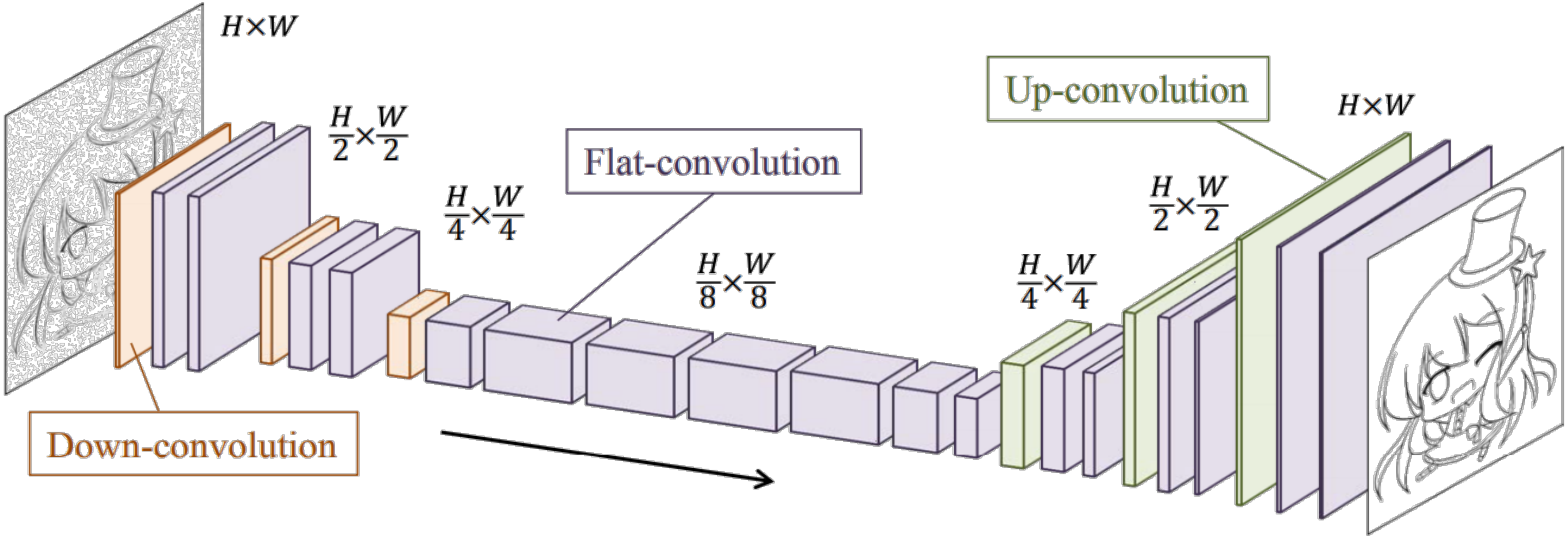


Image from: Xie et al., *Aggregated Residual Transformations for Deep Neural Networks*, CVPR 2017

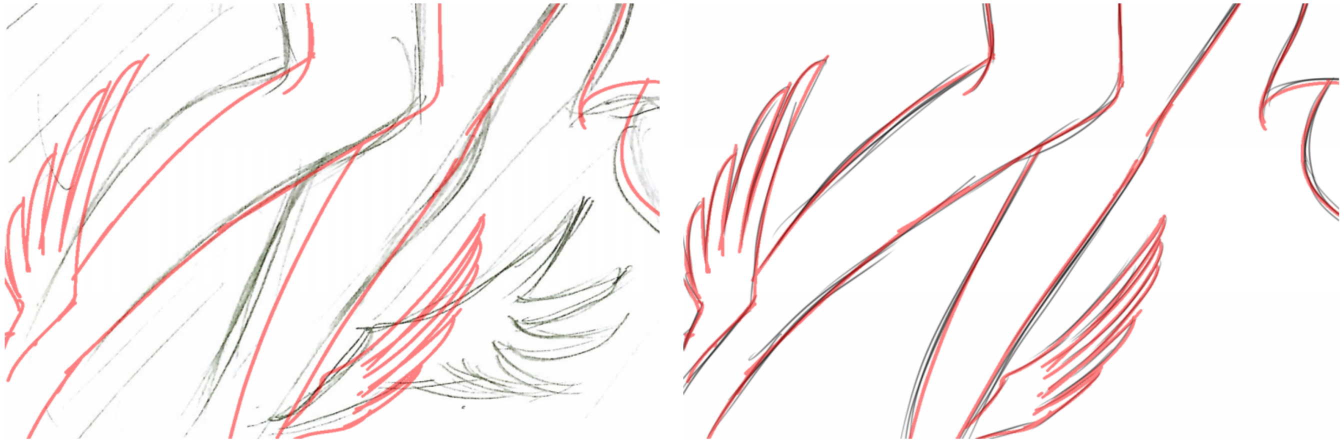
Example: Sketch Simplification



Learning to Simplify: Fully Convolutional Networks for Rough Sketch Cleanup, Simo-Serra et al.

Example: Sketch Simplification

- Loss for thin edges saturates easily
- Authors take extra steps to align input and ground truth edges



Pencil: input
Red: ground truth

Learning to Simplify: Fully Convolutional Networks for Rough Sketch Cleanup, **Simo-Serra et al.**

Image Decomposition

- A selection of methods:
- *Direct Instrinsics*, Narihira et al., 2015
- *Learning Data-driven Reflectance Priors for Intrinsic Image Decomposition,* Zhou et al., 2015
- *Decomposing Single Images for Layered Photo Retouching*, Innamorati et al. 2017

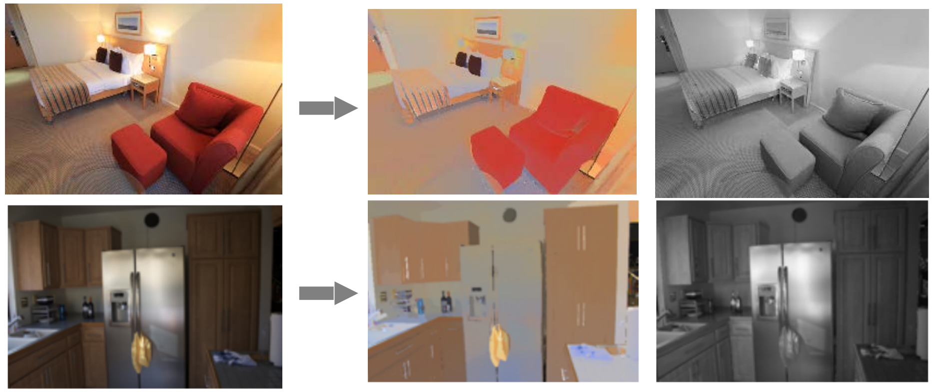
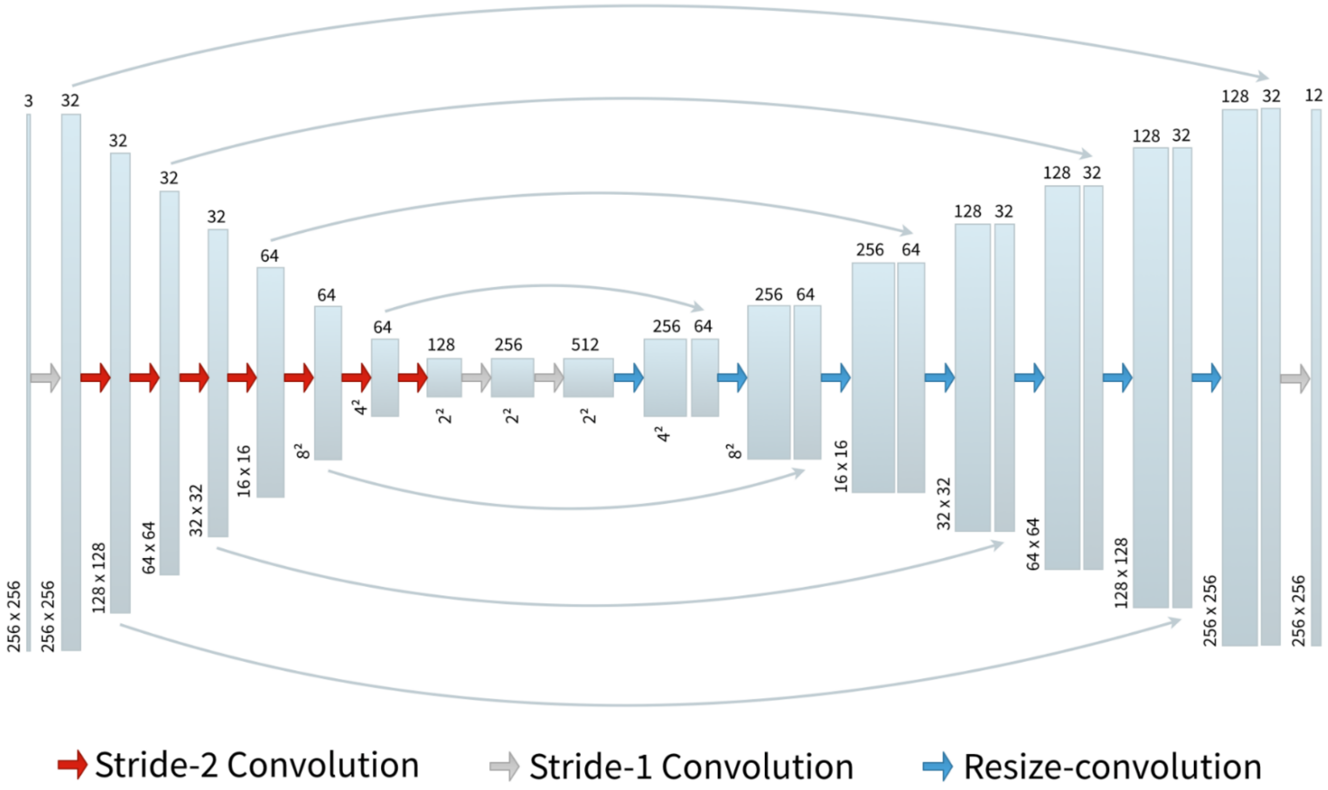
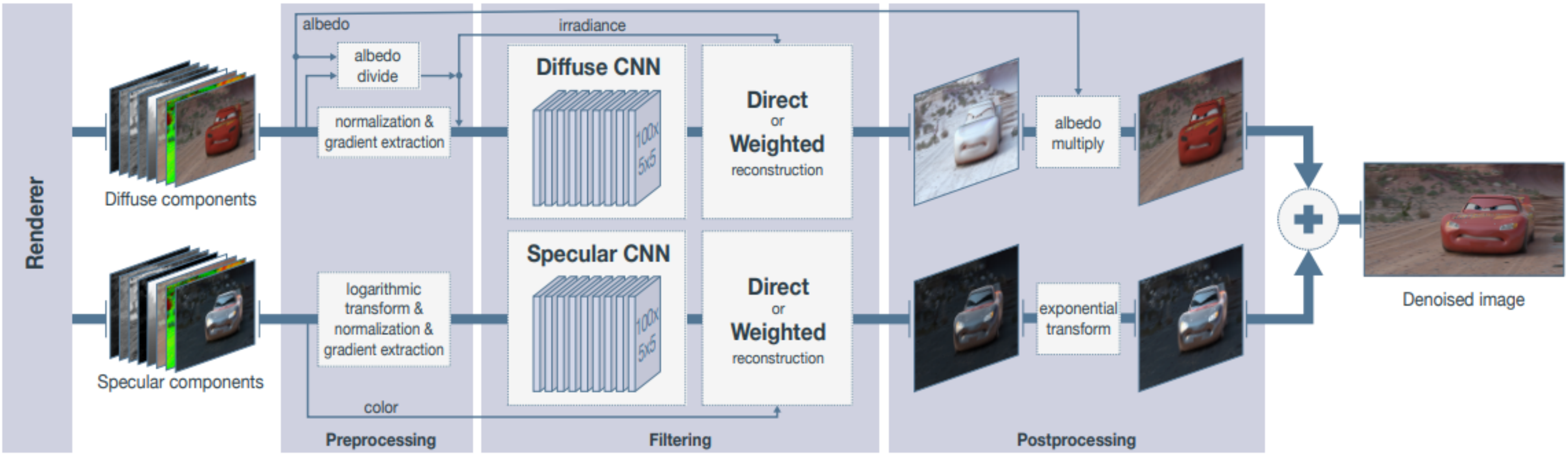


Image Decomposition: Decomposing
Single Images for Layered Photo Retouching



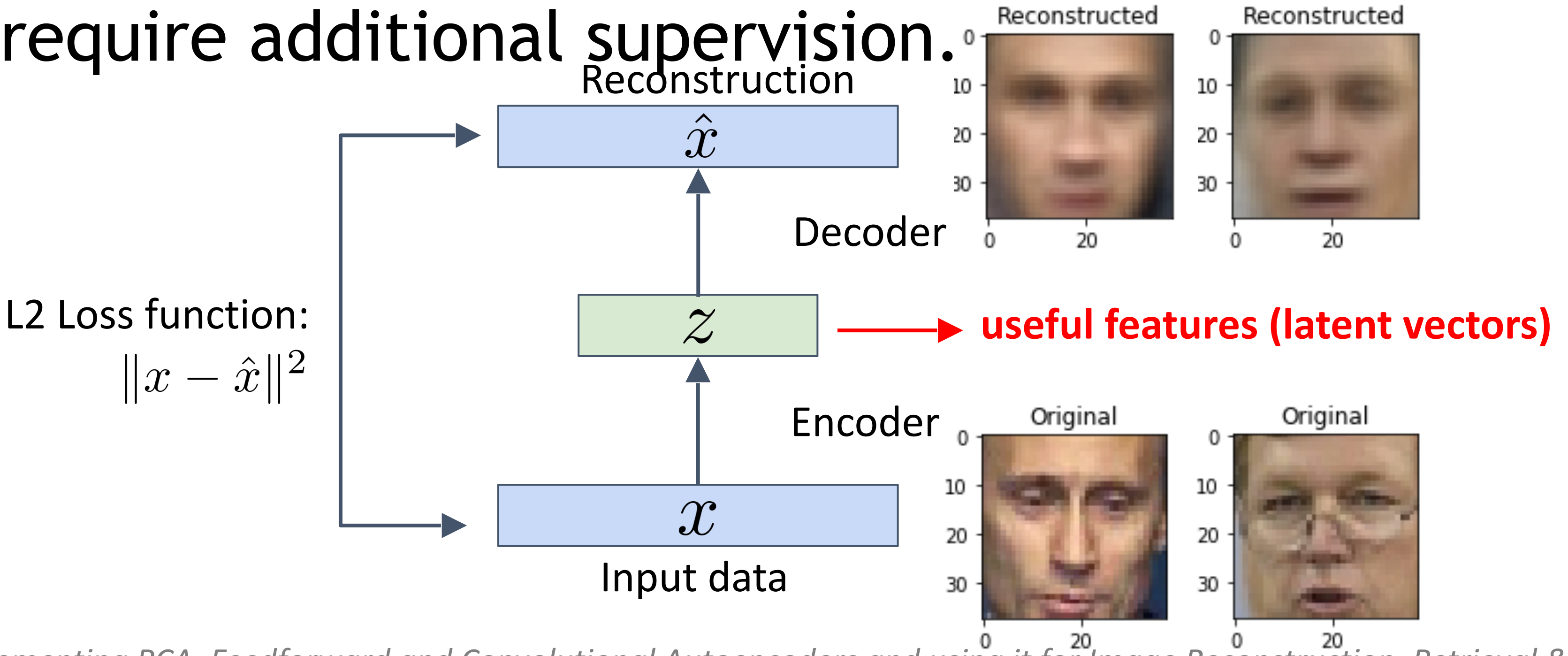
Example Application: Denoising



Deep Features

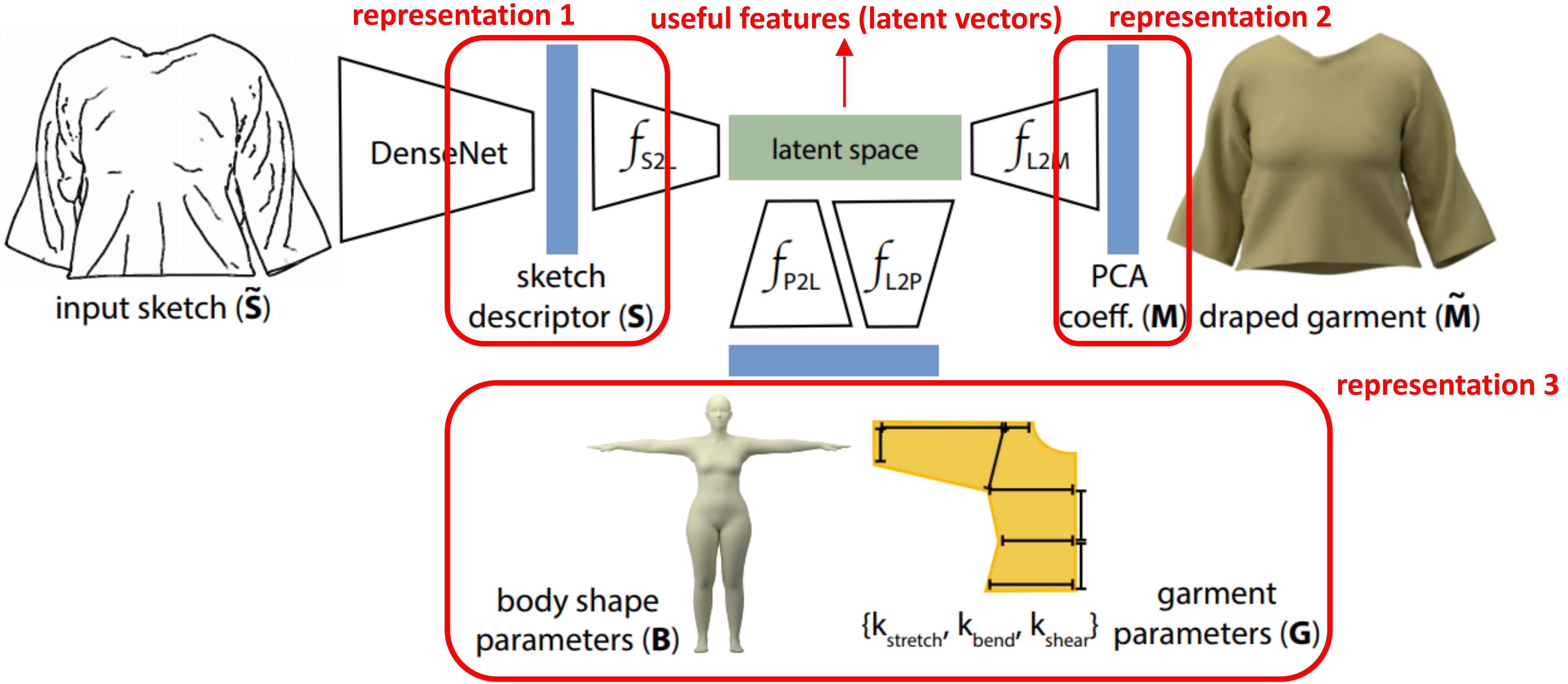
Autoencoders

- Features learned by deep networks are useful for a large range of tasks.
- An autoencoder is a simple way to obtain these features.
- Does not require additional supervision.



Manash Kumar Mandal, *Implementing PCA, Feedforward and Convolutional Autoencoders and using it for Image Reconstruction, Retrieval & Compression*, https://blog.manash.me/

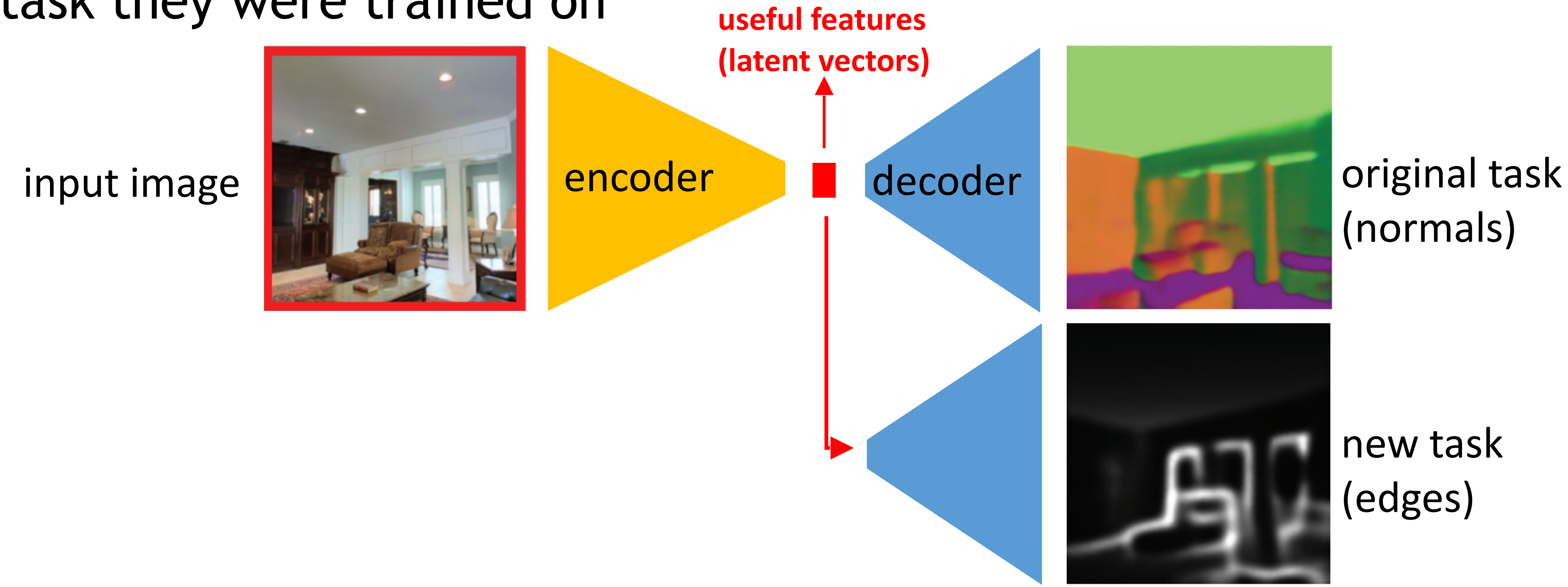
Shared Feature Space: Interactive Garments



Wang et al., *Learning a Shared Shape Space for Multimodal Garment Design*, Siggraph Asia 2018

Transfer Learning

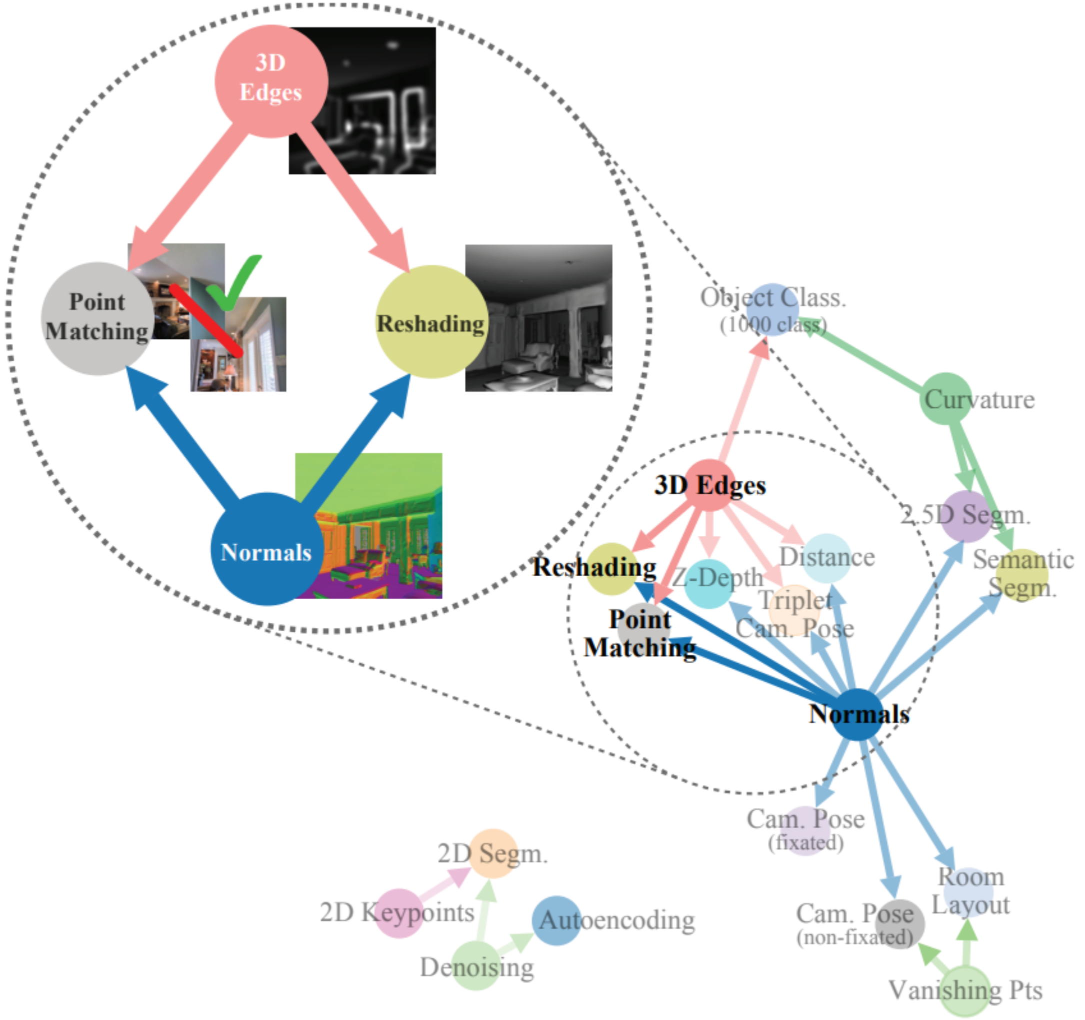
Features extracted by well-trained CNNs often generalize beyond the task they were trained on

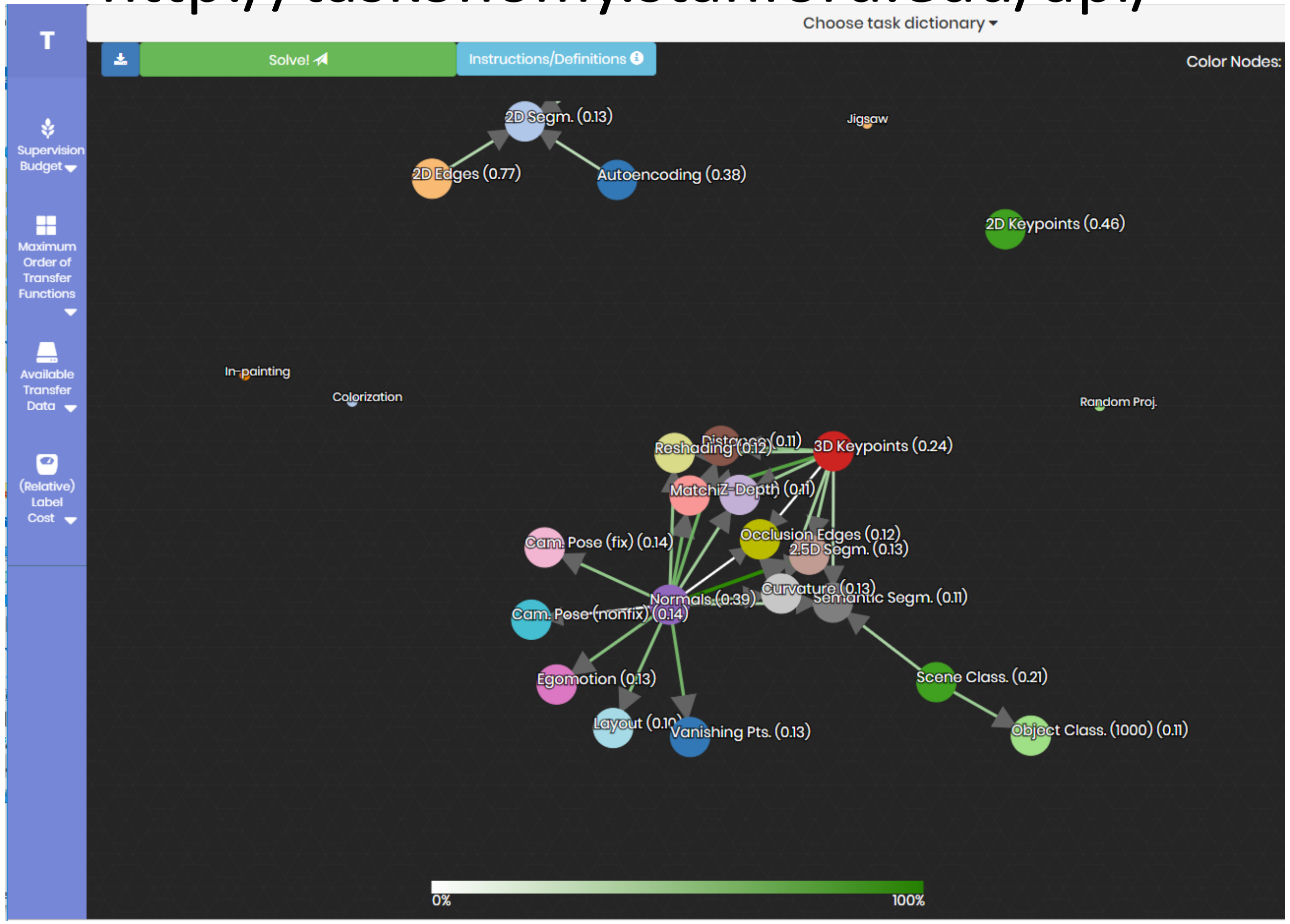


Images from: Zamir et al., *Taskonomy: Disentangling Task Transfer Learning*, CVPR 2018

Taxonomy of Tasks: Taskonomy

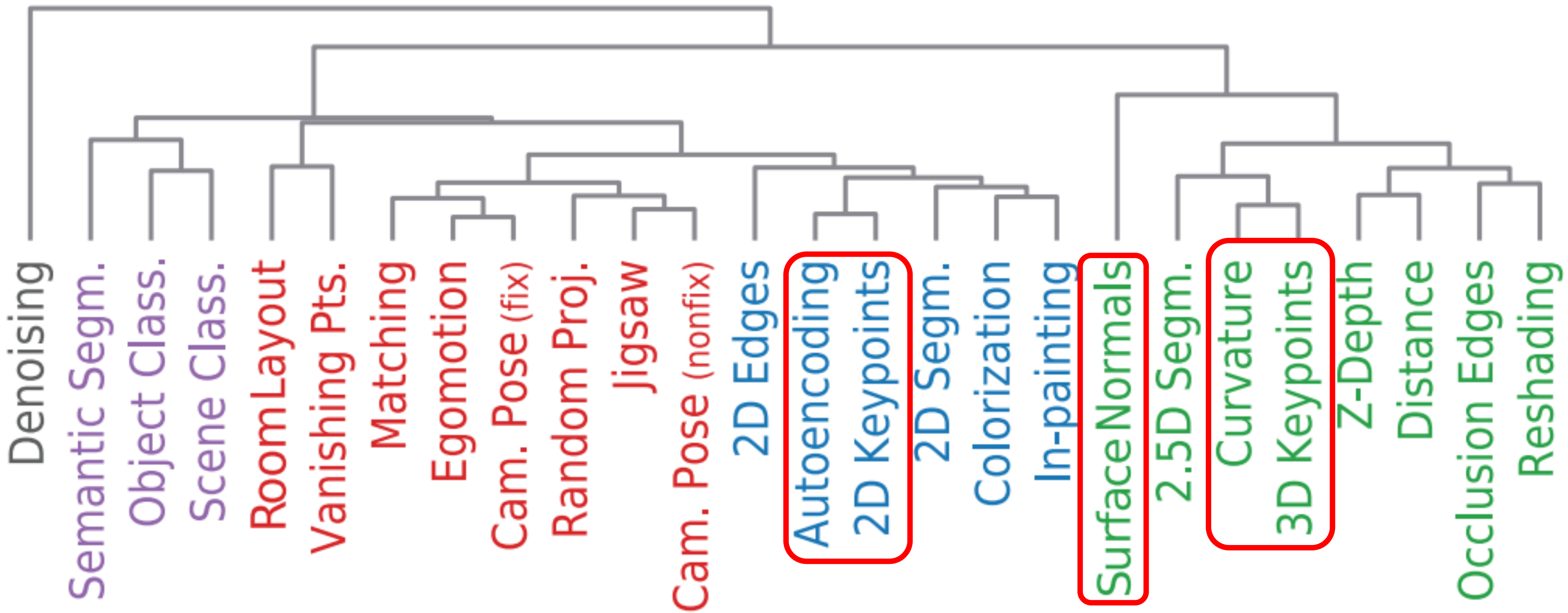
http://taskonomy.stanford.edu/api/





Images from: Zamir et al., *Taskonomy: Disentangling Task Transfer Learning*, CVPR 2018

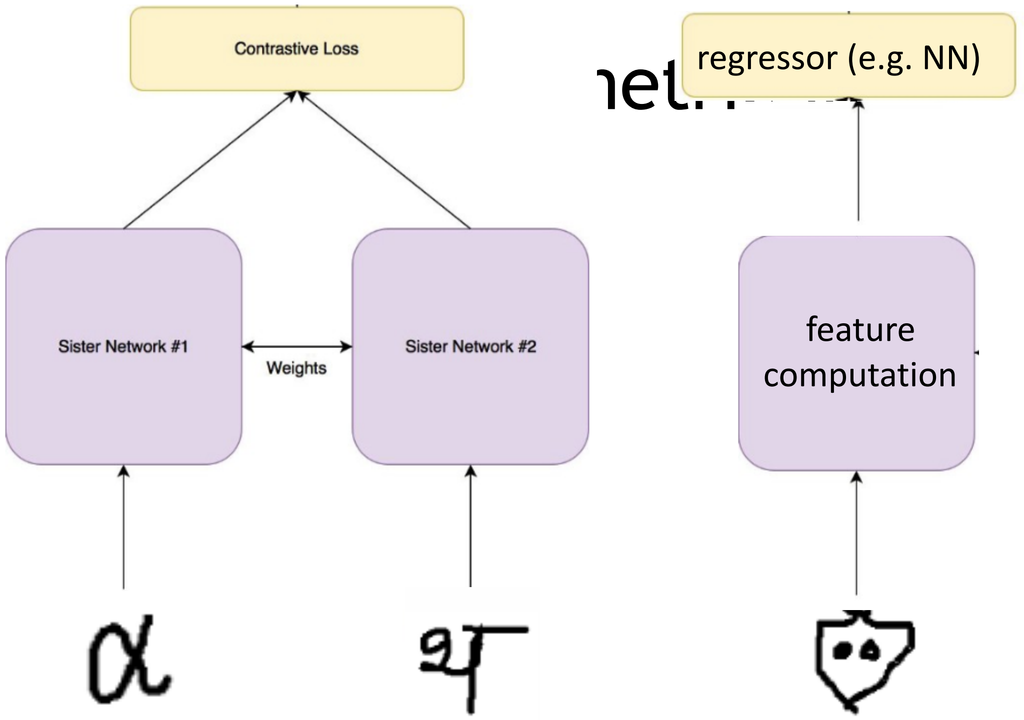
Taxonomy of Tasks: Taskonomy



Images from: Zamir et al., *Taskonomy: Disentangling Task Transfer Learning*, CVPR 2018

Few-shot, One-shot Learning

- With a good feature space, tasks become easier
- In classification, for example, nearest neighbors might already be good enough
- Often trained with a Siamese network
feature space



https://hackernoon.com/one-shot-learning-with-siamese-networks-in-pytorch-8ddaab10340e

Style Transfer

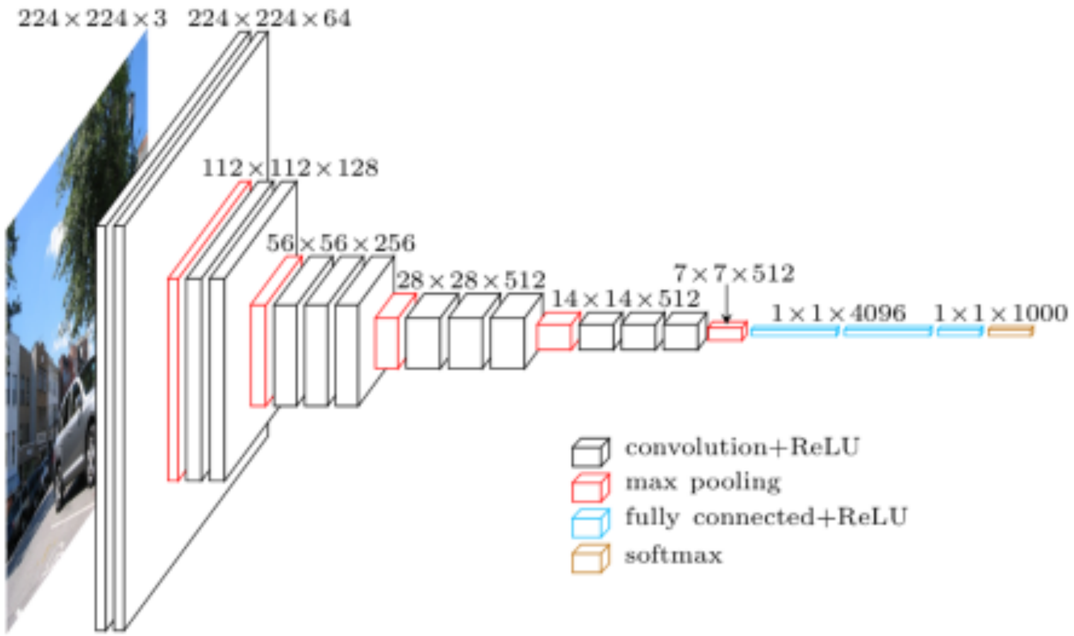
- Combine content from image A with style from image B

Images from: Gatys et al., *Image Style Transfer using Convolutional Neural Networks*, CVPR 2016

What is Style and Content?

Remember that features in a CNN often generalize well.

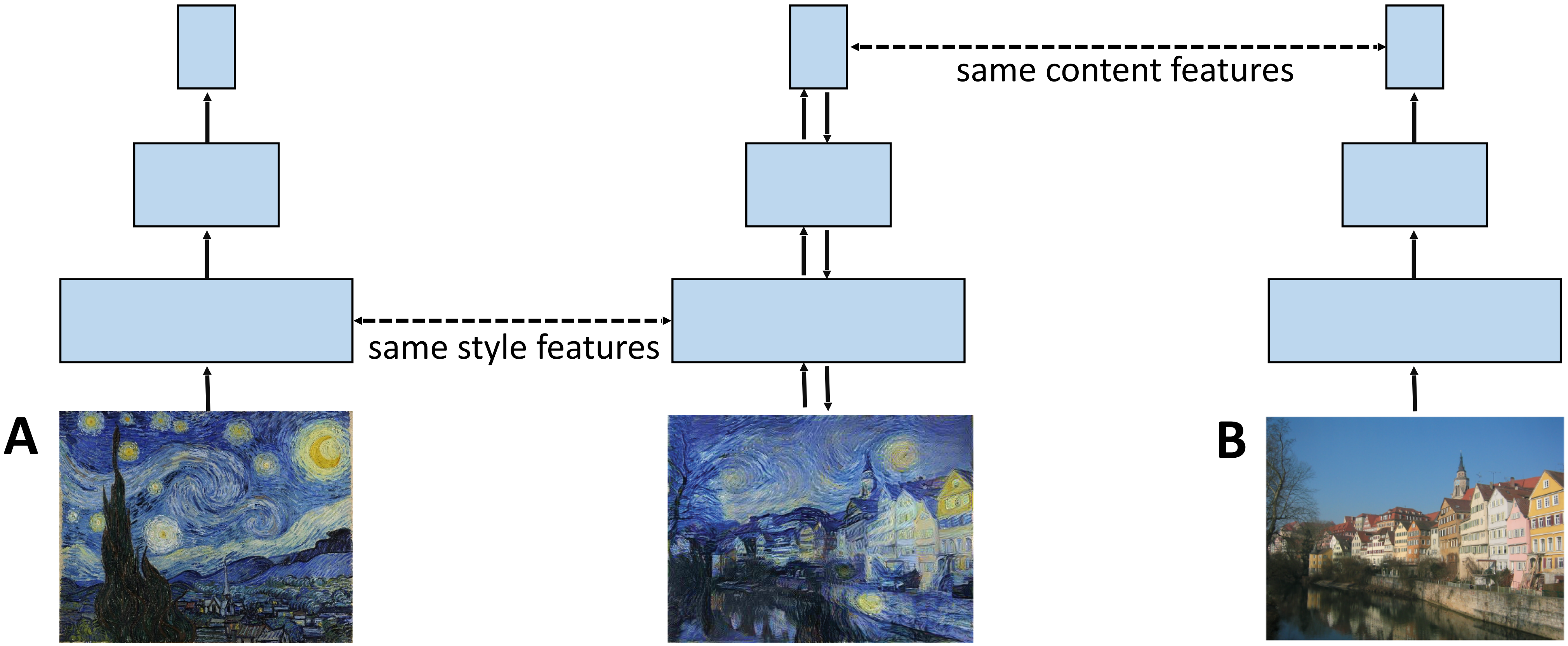
Define style and content using the layers of a CNN (VGG19 for example):



shallow layers describe style ⟷ deeper layers describe content

Optimize for Style A and Content B

same pre-trained networks, fix weights



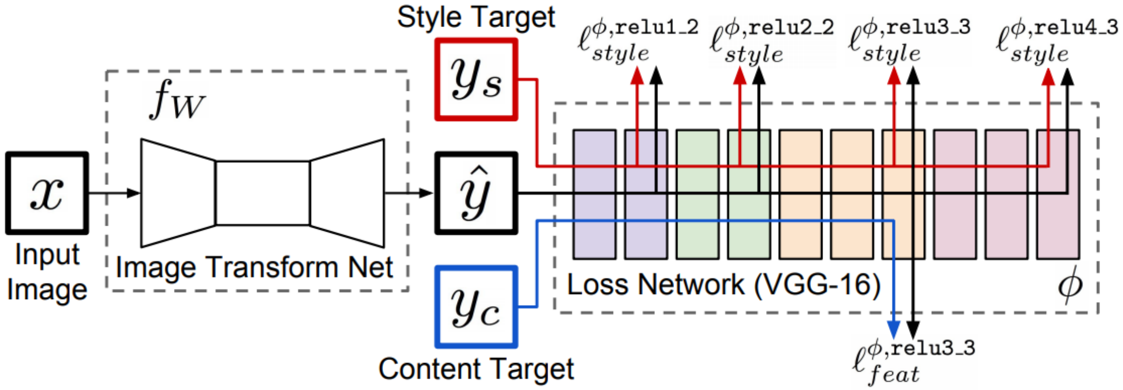
optimize to have same style/content features

Style Transfer: Follow-Ups

more control over the result

(a) Content (b) Style I (c) Style II

feed-forward networks



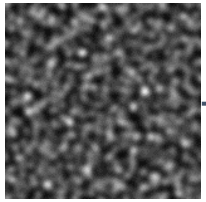
Images from: Gatys, et al., *Controlling Perceptual Factors in Neural Style Transfer*, CVPR 2017
Johnson et al., *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*, ECCV 2016

Style Transfer for Videos

Ruder et al., *Artistic Style Transfer for Videos*, German Conference on Pattern Recognition 2016

Adversarial Image Generation

Generative Adversarial Networks



Player 1: generator
Scores if discriminator can't distinguish output from real image

from dataset

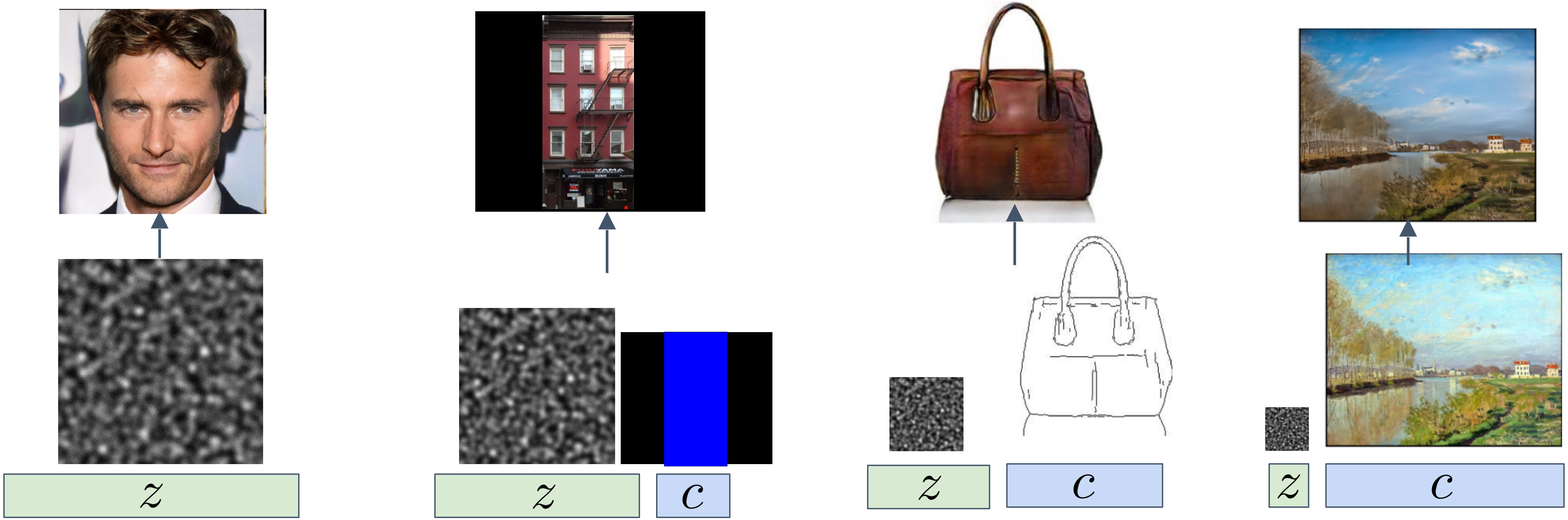
Player 2: discriminator → real/fake
Scores if it can distinguish between real and fake

GANs to CGANs (Conditional GANs)

GAN

CGAN

increasingly determined by the condition



Karras et al., *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, ICLR 2018
Kelly and Guerrero et al., *FrankenGAN: Guided Detail Synthesis for Building Mass Models using Style-Synchonized GANs*, Siggraph Asia 2018
Isola et al., *Image-to-Image Translation with Conditional Adversarial Nets*, CVPR 2017
Image Credit: Zhu et al. , *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks* , ICCV 2017

GAN

Image-to-image Translation

- ≈ learn a mapping between images from example pairs
- Approximate sampling from a conditional distribution $p_{\text{data}}(x \mid c)$

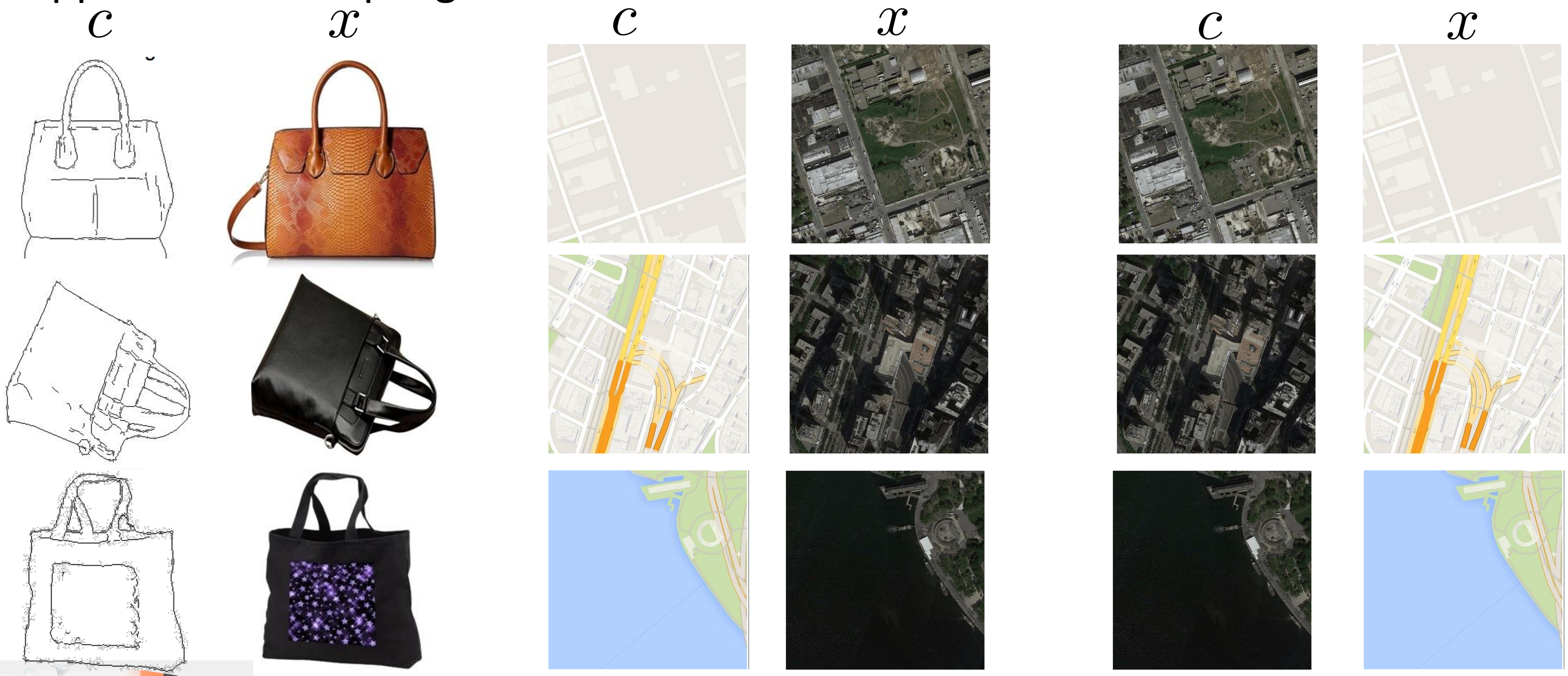
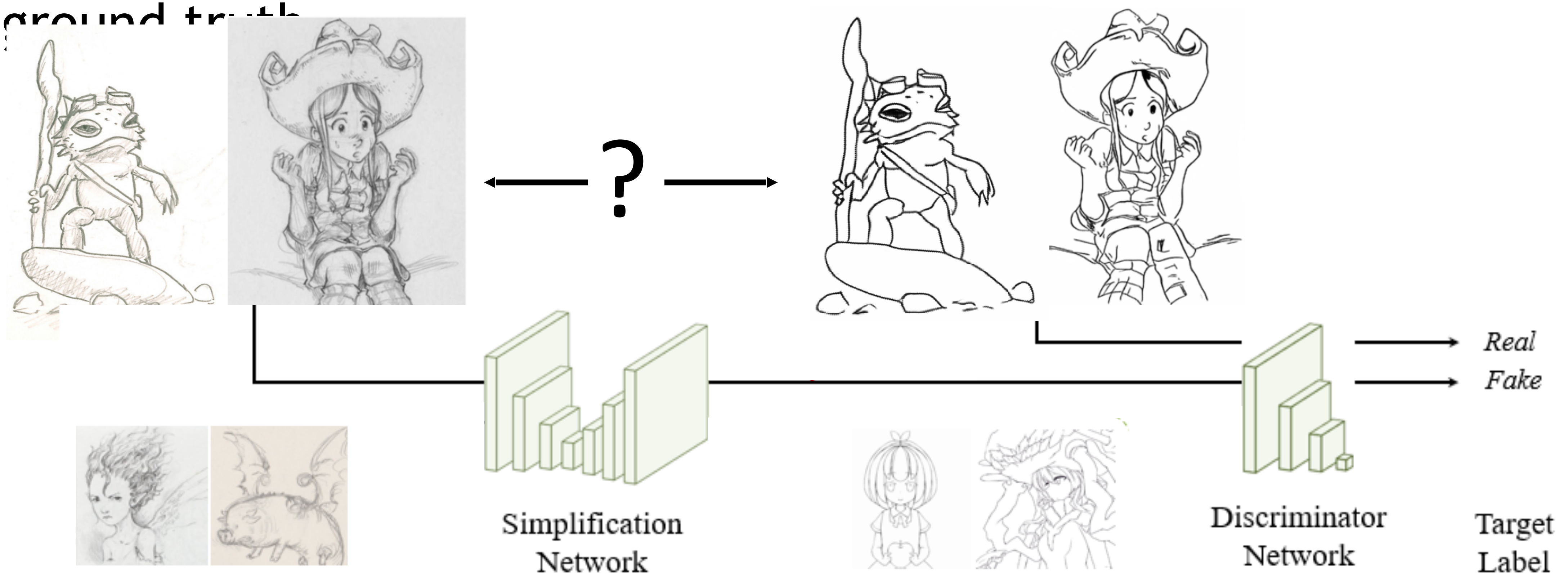


Image Credit: *Image-to-Image Translation with Conditional Adversarial Nets*, Isola et al.

Adversarial Loss vs. Manual Loss

Problem: A good loss function is often hard to find

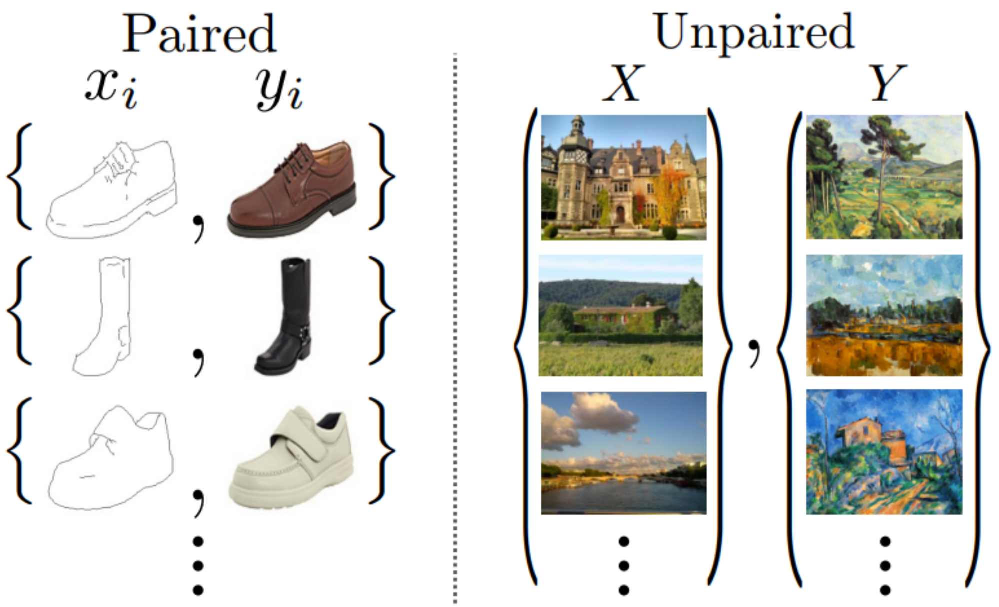
Idea: Train a network to discriminate between network output and ground truth



Images from: Simo-Serra, Iizuka and Ishikawa, *Mastering Sketching*, Siggraph 2018

CycleGANs

- Less supervision than CGANs: mapping between unpaired datasets
- Two GANs + cycle consistency



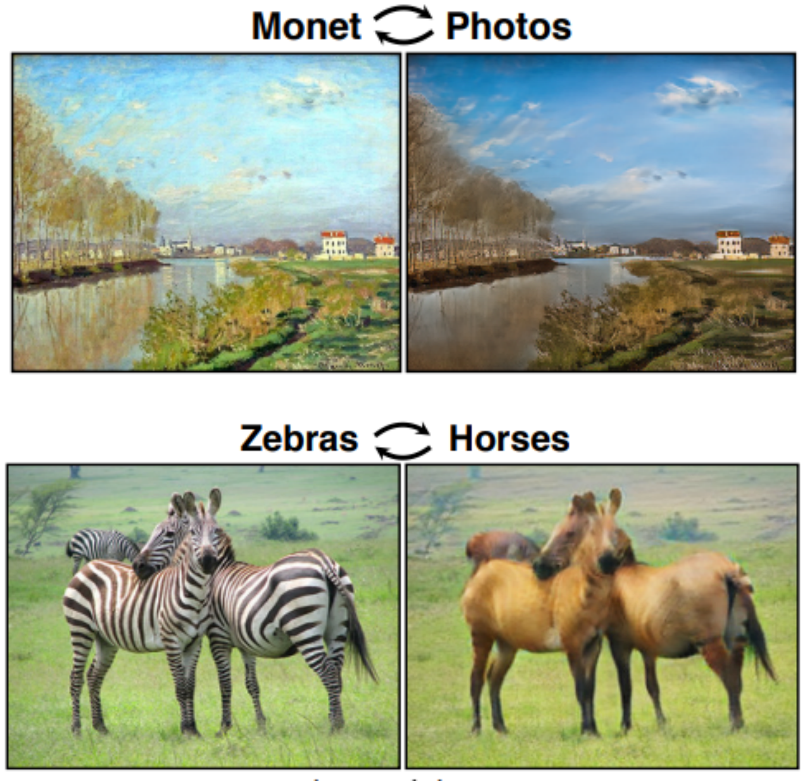


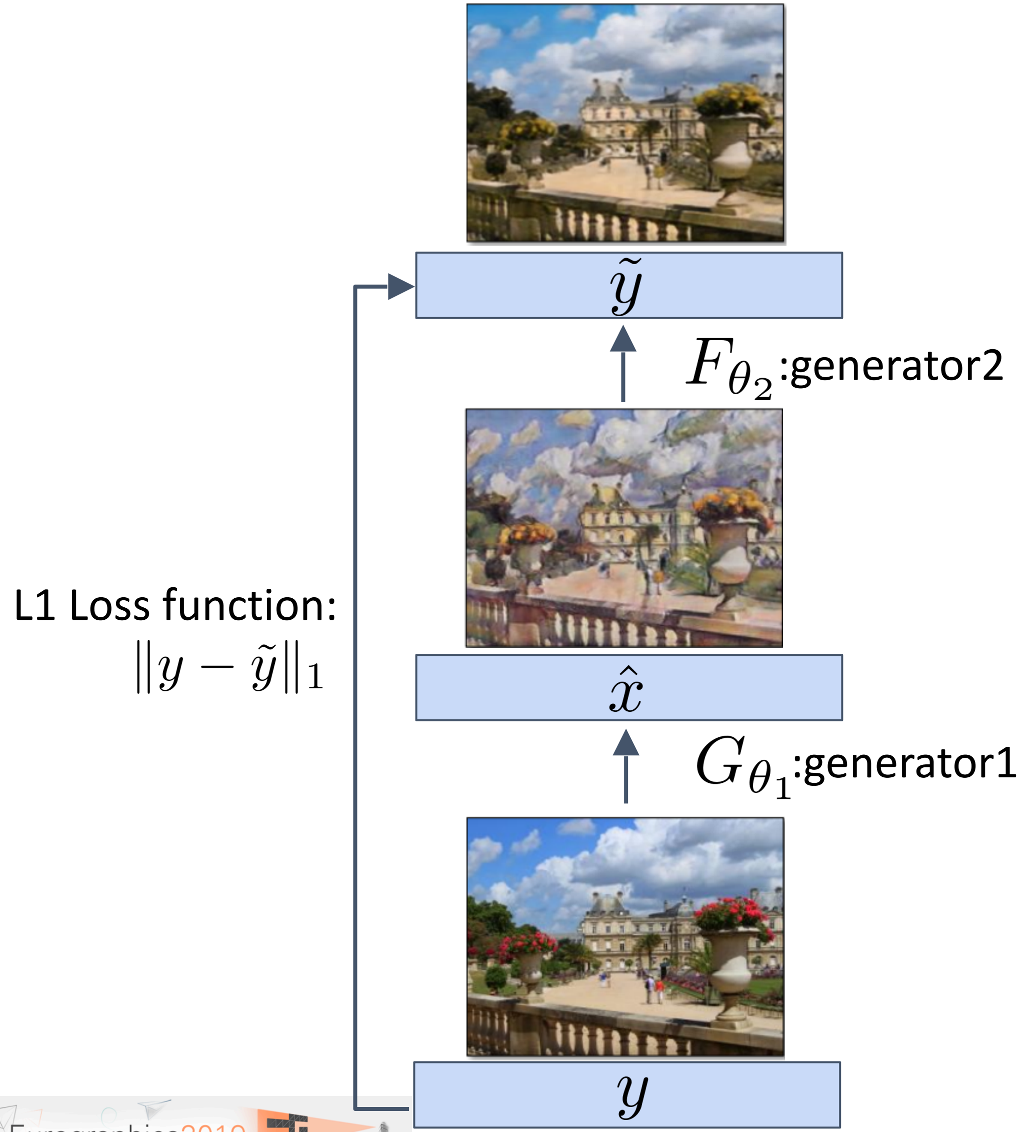
Image Credit: *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*, Zhu et al.

CycleGAN: Two GANs …

- Not conditional, so this alone does not constrain generator input and output to match

Image Credit: *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*, Zhu et al.

CycleGAN: … and Cycle Consistency



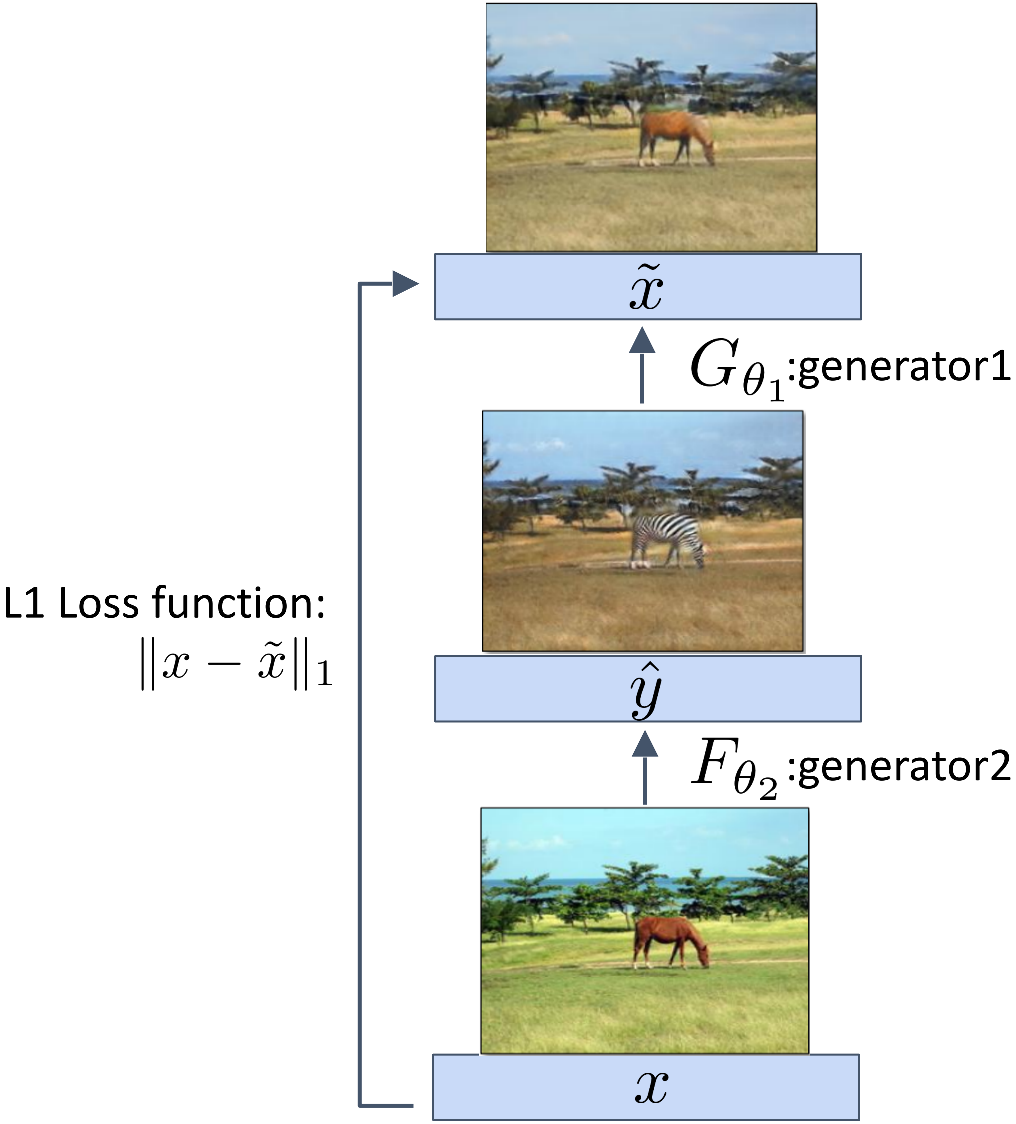


Image Credit: *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*, Zhu et al.

The Conditional Distribution in CGANs

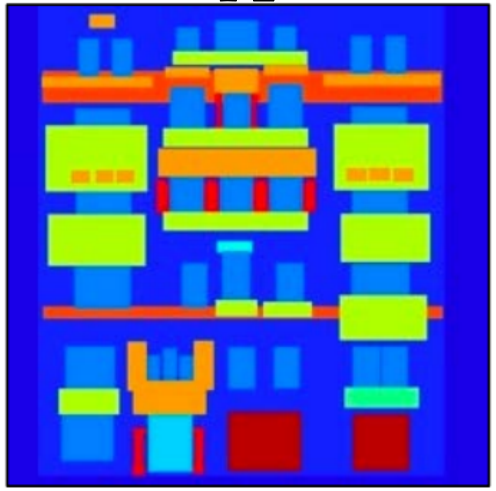
A $\qquad\qquad$ $p(B|A)$

(a) Input night image

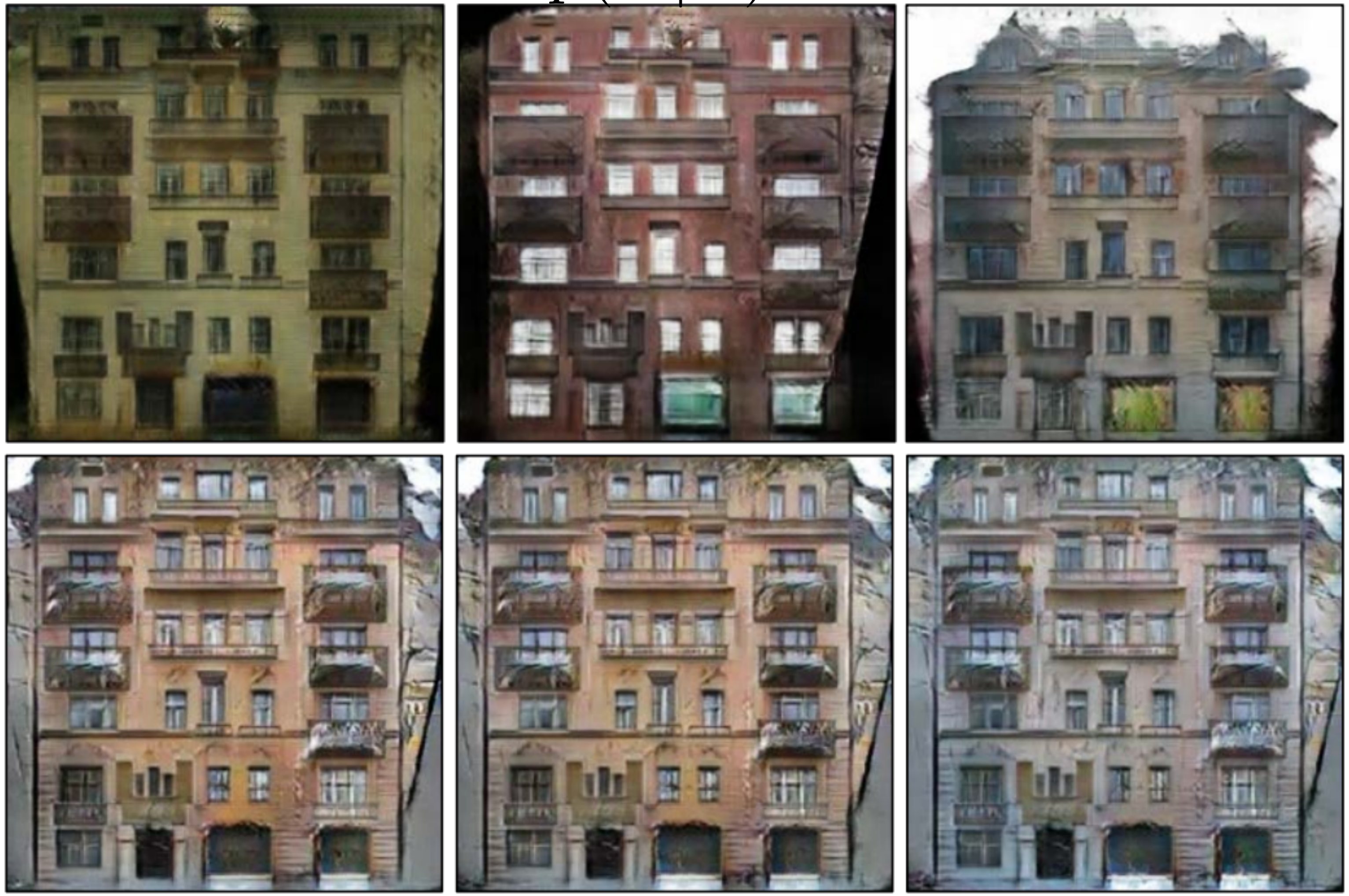
Image from: Zhu et al., *Toward Multimodal Image-to-Image Translation*, NIPS 2017

The Conditional Distribution in CGANs

A

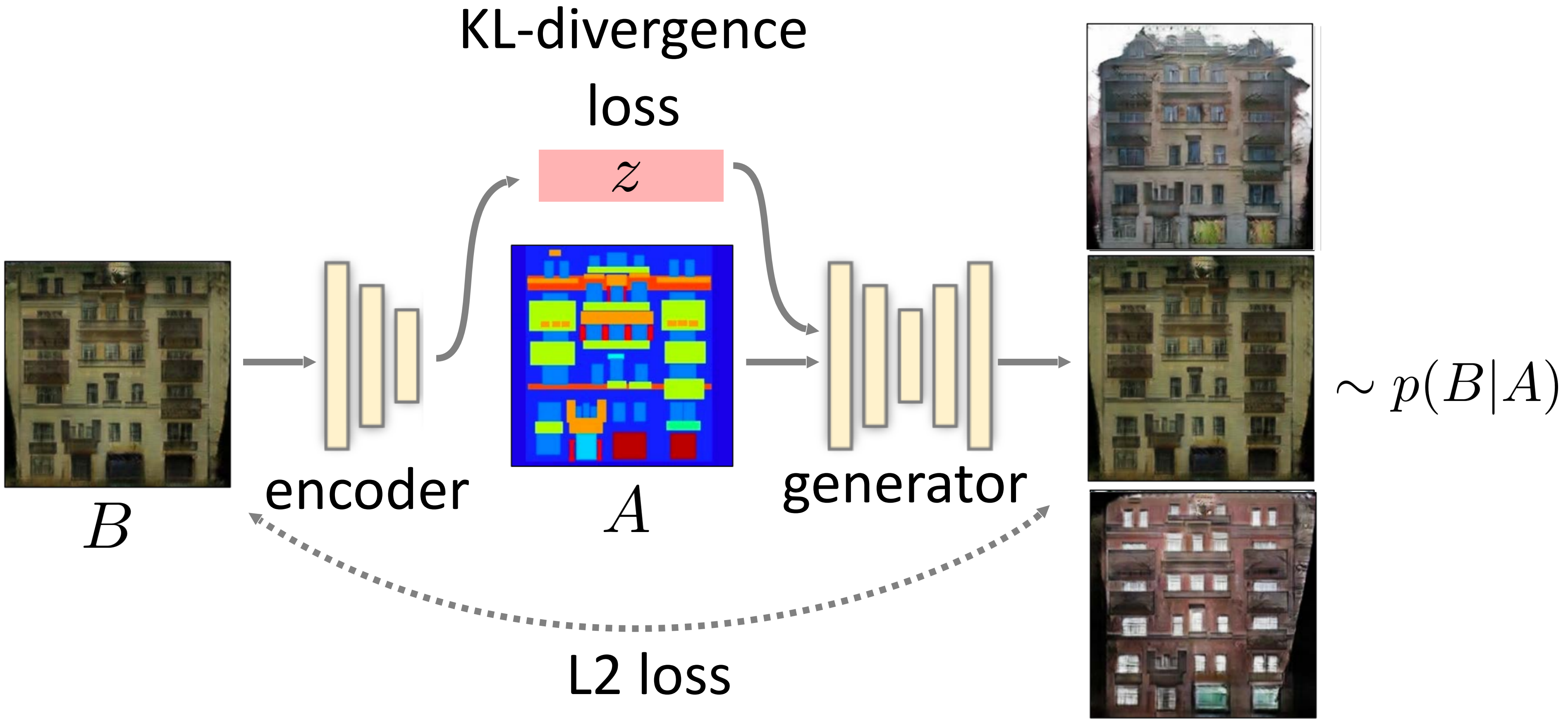


$p(B|A)$

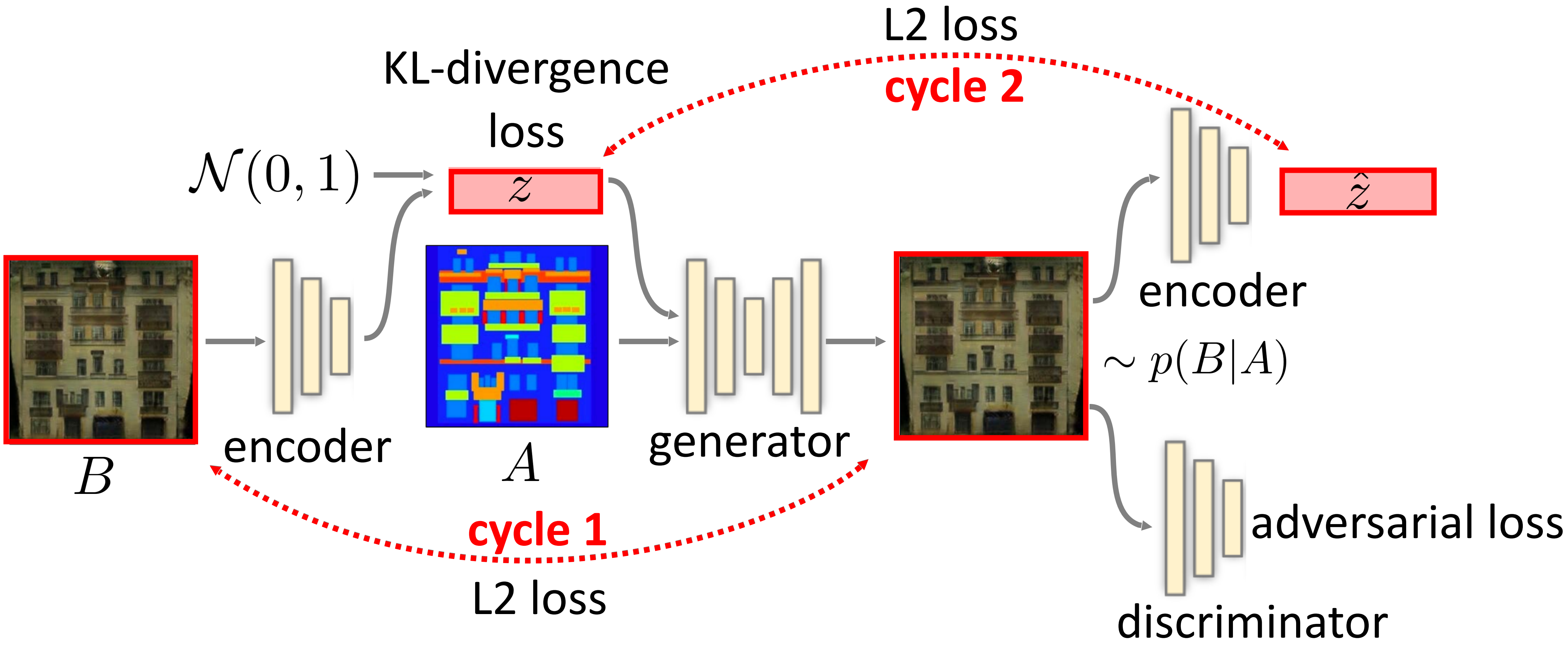


Zhu et al., *Toward Multimodal Image-to-Image Translation*, NIPS 2017

BicycleGAN

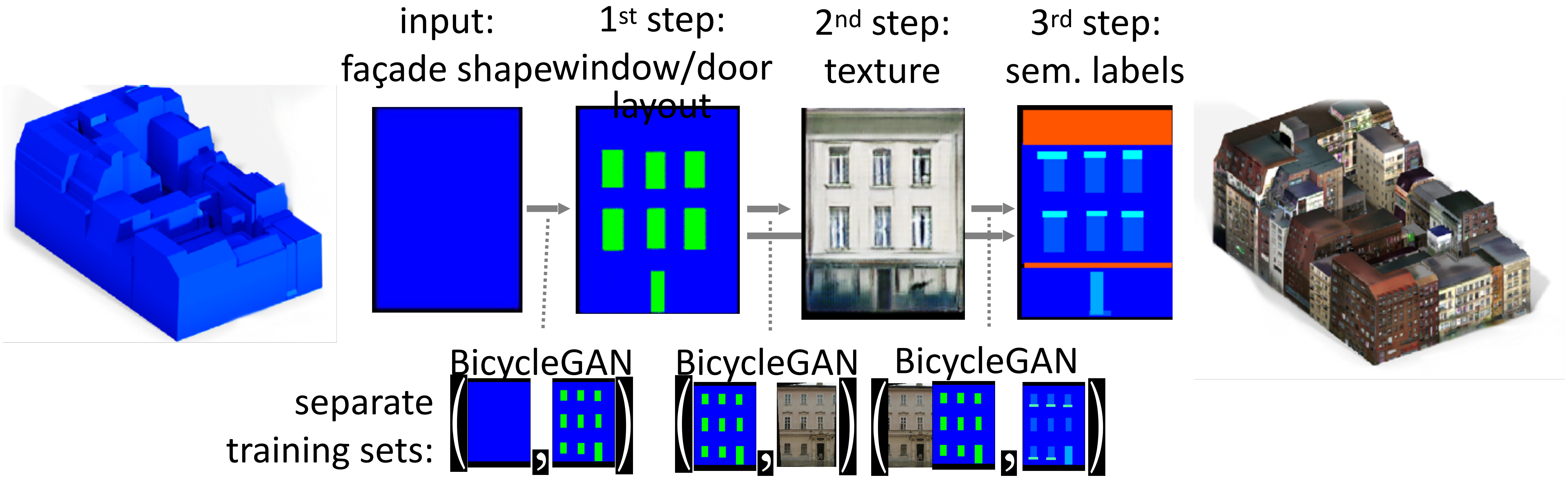


BicycleGAN



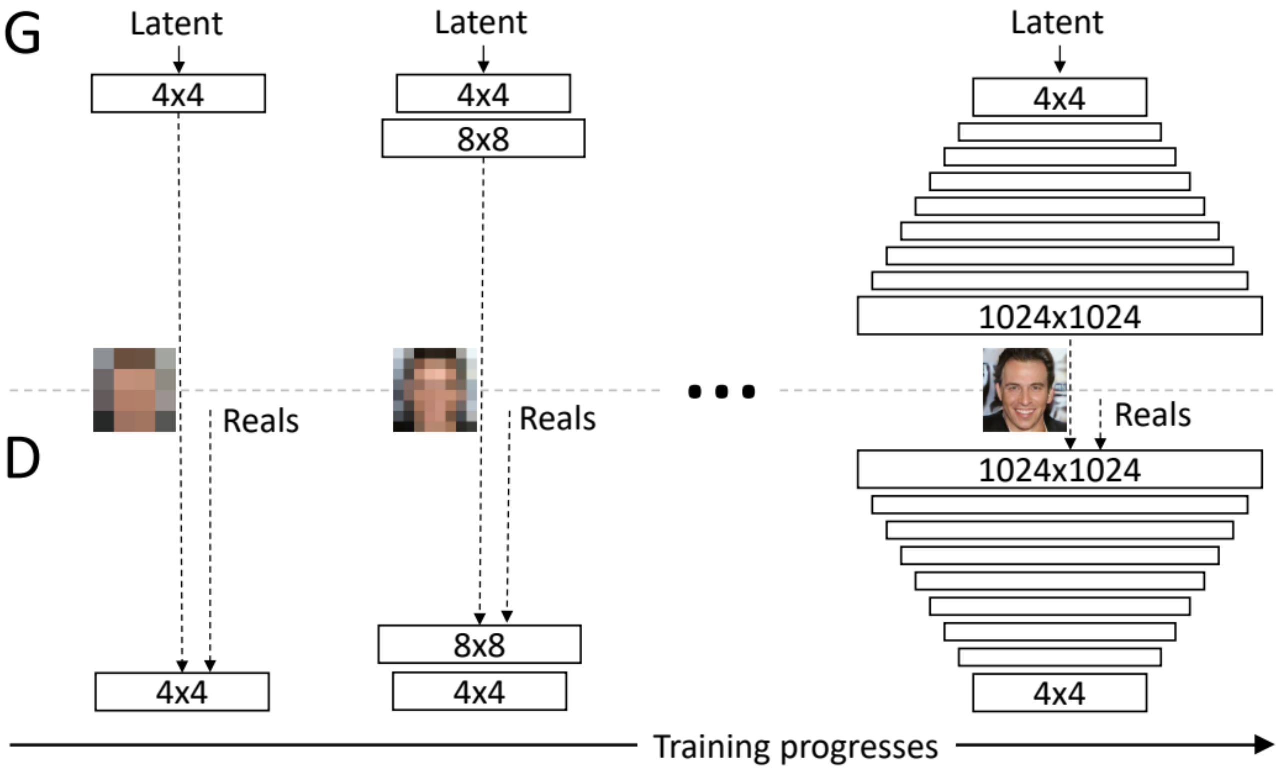
L2 loss
cycle 2
KL-divergence
loss
$\mathcal{N}(0,1)$
z
$\hat{z}$
encoder
B
encoder
A
generator
$\sim p(B|A)$
adversarial loss
discriminator
cycle 1
L2 loss

FrankenGAN


input:
façade shape
1st step:
window/door layout
2nd step:
texture
3rd step:
sem. labels
BicycleGAN
BicycleGAN
BicycleGAN
separate
training sets:

Progressive GAN

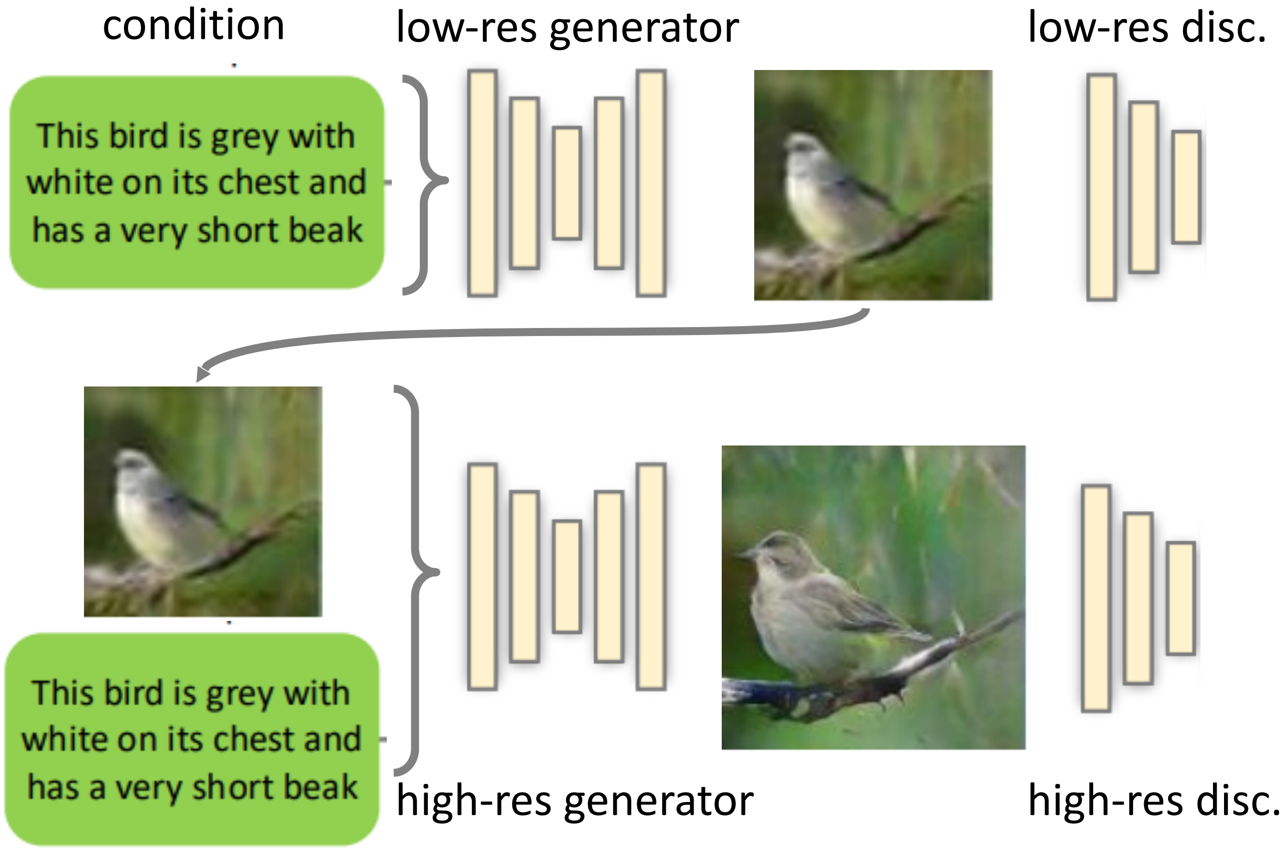
- Resolution is increased progressively during training
- Also other tricks like using minibatch statistics and normalizing feature vectors



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

StackGAN

Condition does not have to be an image



This flower has white petals with a yellow tip and a yellow pistil

A large bird has large thighs and large wings that have white wingbars

Zhang et al., *StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks*, ICCV 2017

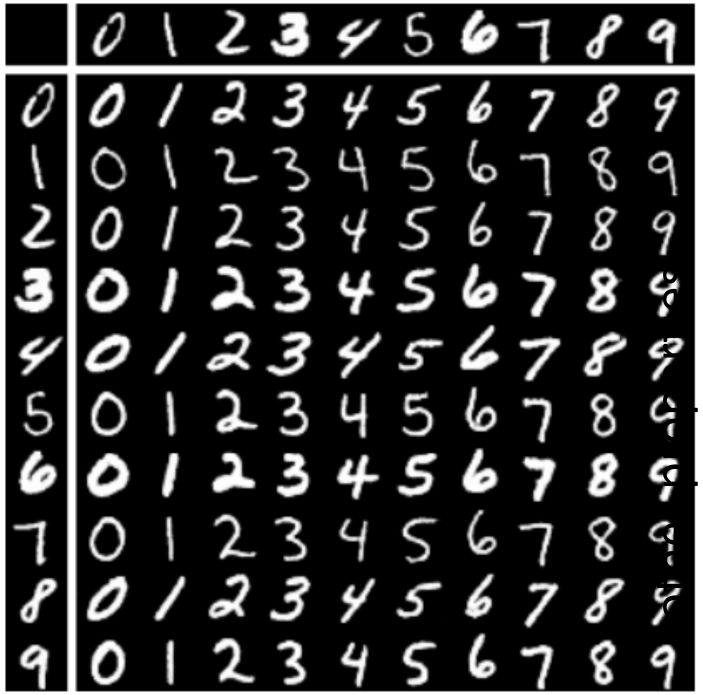
Disentanglement

z — Entangled: different properties may be mixed up over all dimensions

$z_a \; z_b \; \cdots$ — Disentangled: different properties are in different dimensions

specified property: number

other properties



specified property: character

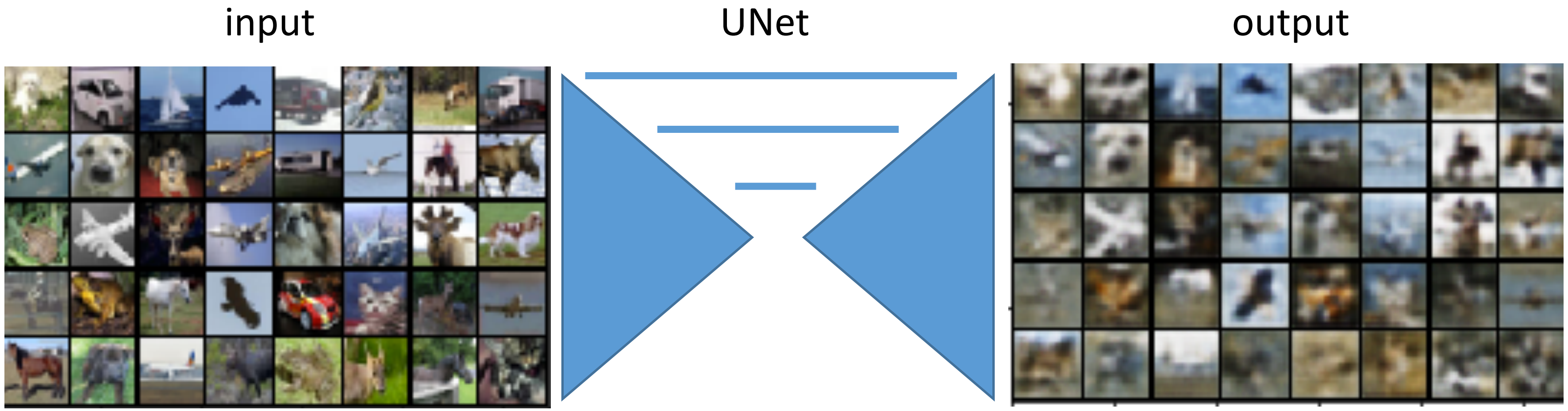
other properties

Mathieu et al., *Disentangling factors of variation in deep representations using adversarial training*, NIPS 2016

Attention and Gray Box Learning

Attention in Deep Learning

target: horizontal mirroring

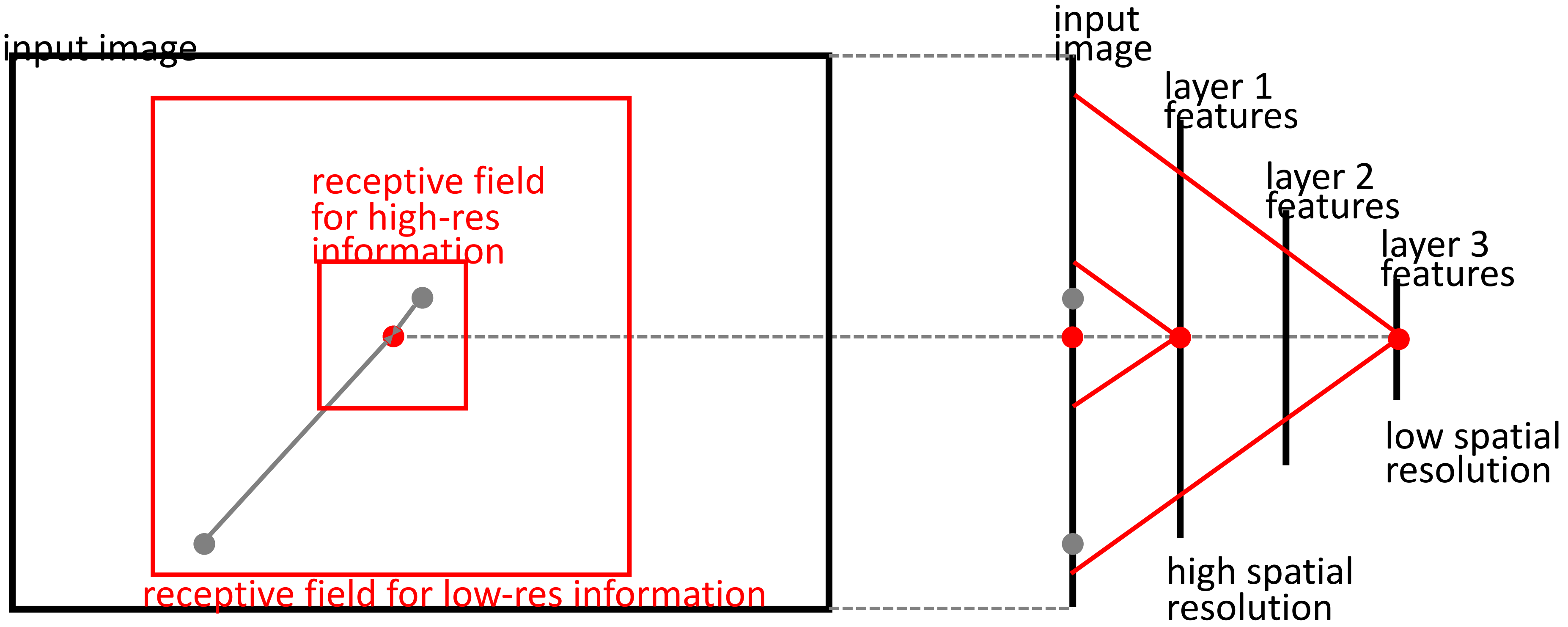


Why is this hard for the network?

1) Locality of convolutions
2) Driven only by data from shallower layers (no semantics)

Attention in Deep Learning

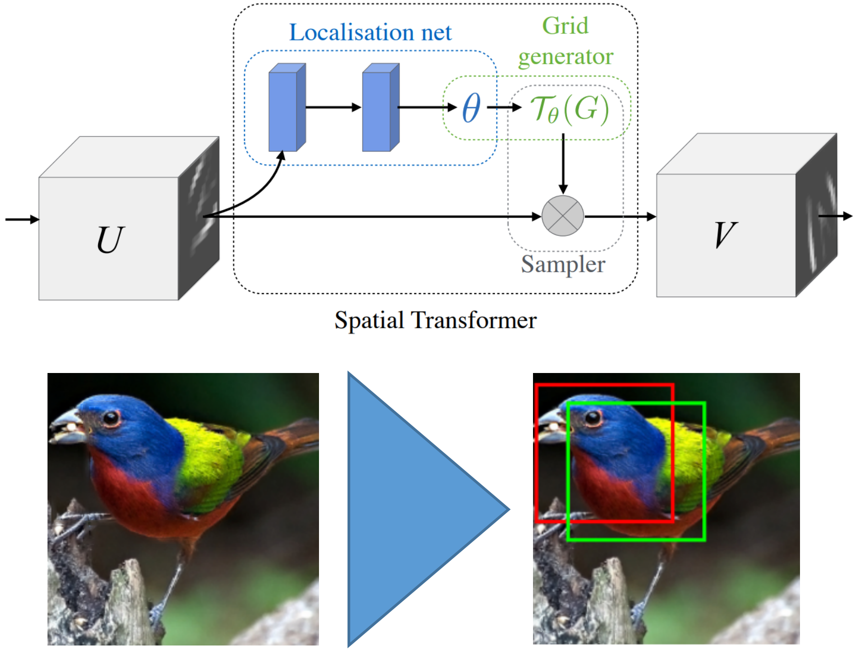
Problem: architecture constrains information flow. For example, in a typical CNN, at a given image location (red), information about other image locations (grey) is available in a resolution that depends on the spatial distance.



Attention Based on Semantics

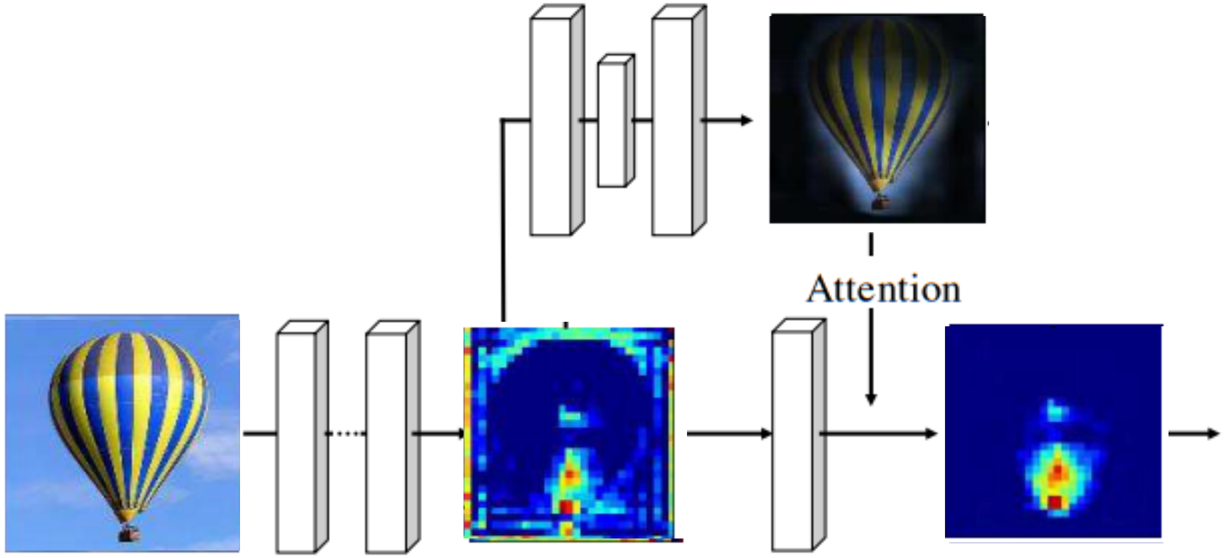
Idea: use higher-level semantics to select relevant information

Spatial Transformer Networks



Jaderberg et al., *Spatial Transformer Networks*, NIPS 2015

Residual Attention Network for Image Classification



Wang et al., *Residual Attention Network for Image Classification*, CVPR 2017

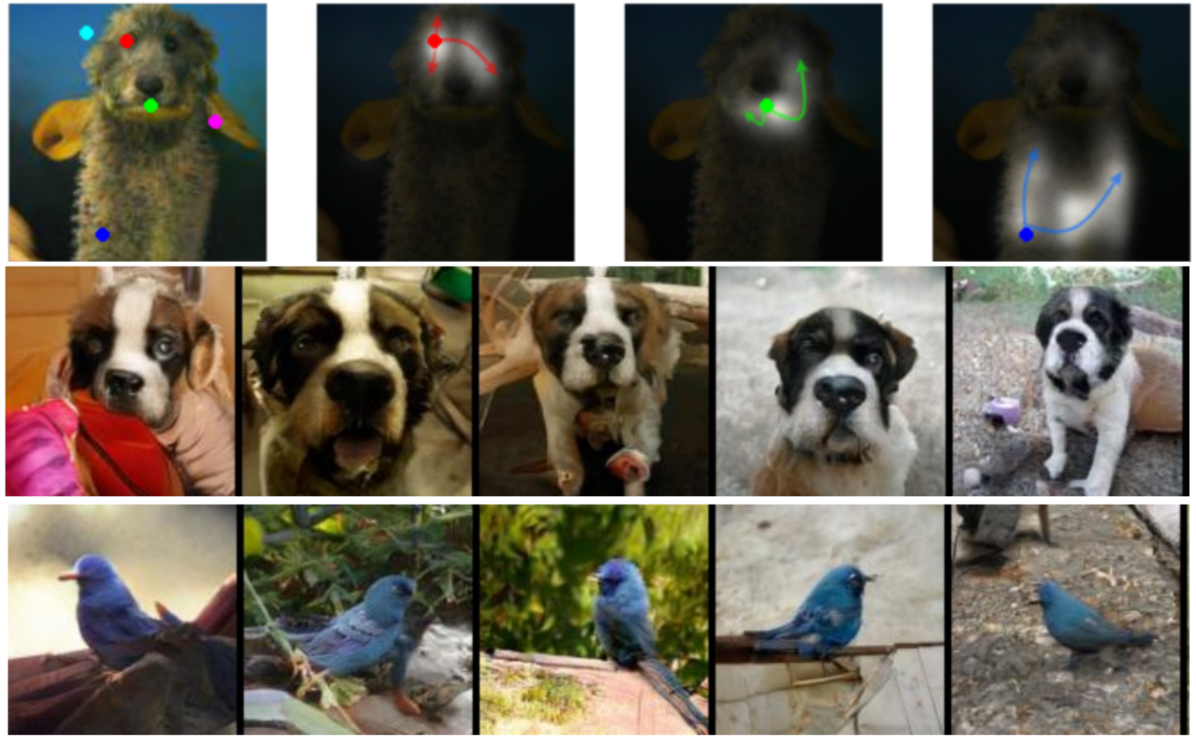
Attention to Distant Details

Idea: gather information from distant details based on their features

Non-local Neural Networks

Wang et al., *Non-local Neural Networks*, CVPR 2018

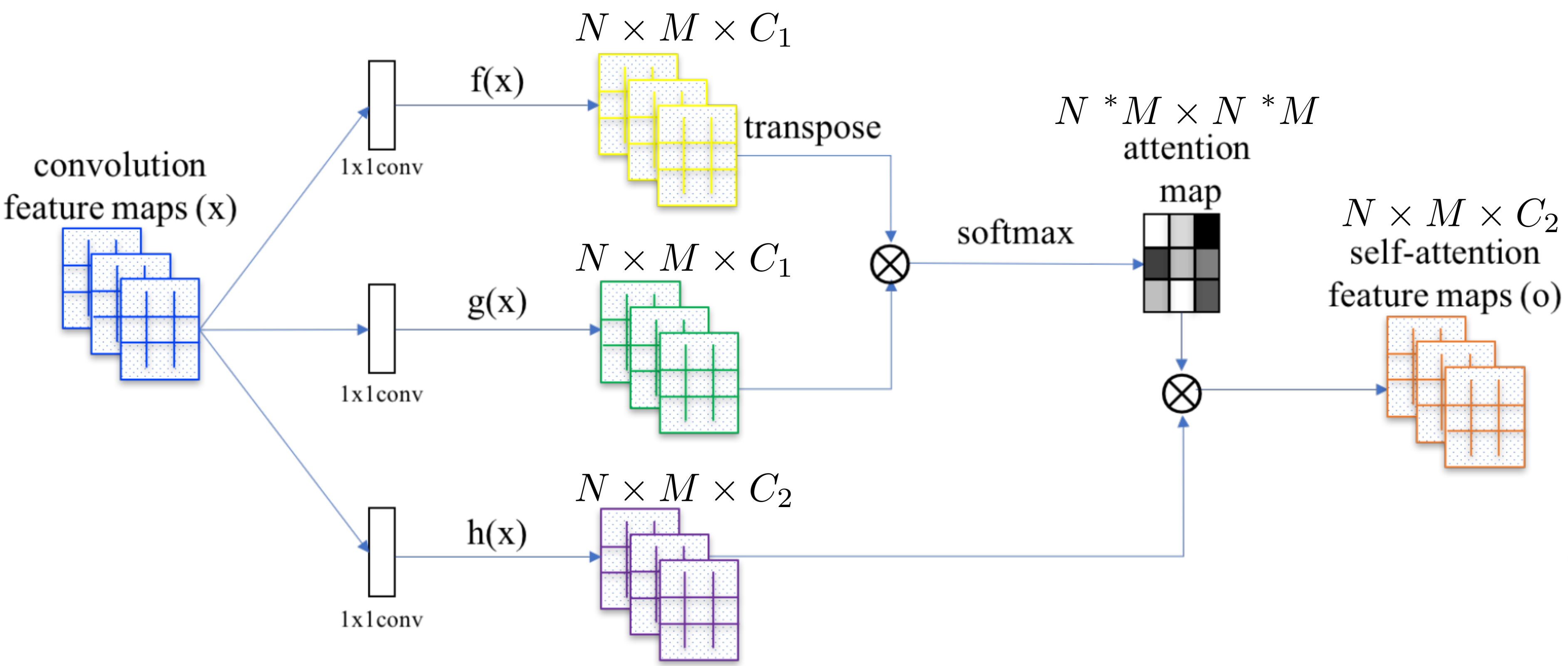
Attention GAN



Zhang et al., *Self-Attention Generative Adversarial Networks*, CVPR 2018

Attention to Distant Details

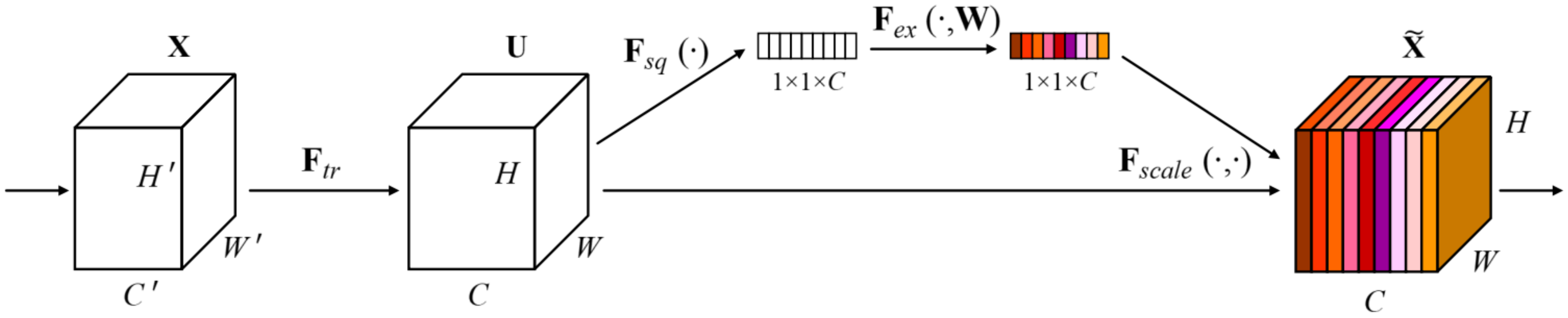
Idea: gather information from distant details based on their features



Zhang et al., *Self-Attention Generative Adversarial Networks*, CVPR 2018

Squeeze and Excitation: Attention over Channels

Idea: weigh (emphasize and suppress) channels based on global information

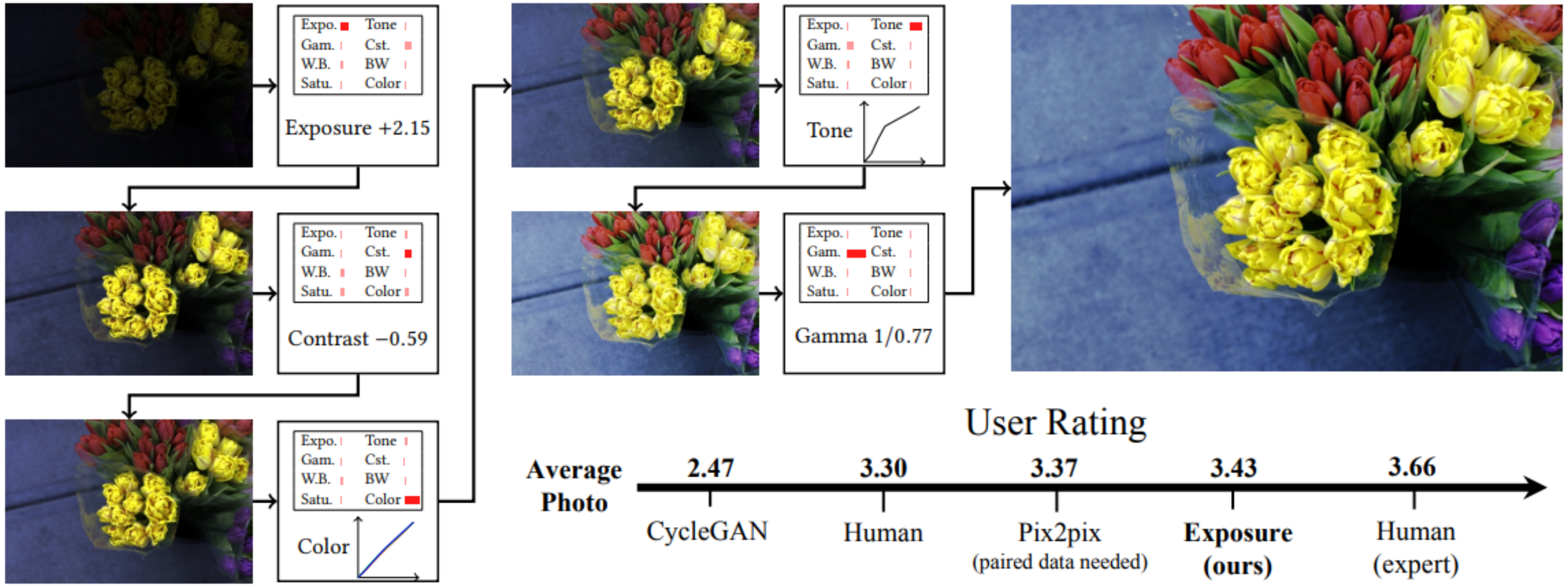


Hu et al., *Squeeze-and-Excitation Networks*, CVPR 2018

Gray Box Learning

Problem: Most networks are black boxes.

Idea: Regress parameters for a small set of well-known operations.



Hu et al., *Exposure: A White-Box Photo Post-Processing Framework*, Siggraph 2018

Summary

- Common Architecture Elements
  (Dilated Convolution, Grouped Convolutions)
- Deep Features
  (Autoencoders, Transfer Learning, One-shot Learning, Style Transfer)
- Adversarial Image Generation
  (GANs, CGANs)
- Interesting Trends
  (Attention, “Gray Box” Learning)