TUM·adl4cv | Advanced Deep Learning for Computer Vision(2020)

adl4cv (2020)· 课程资料包 @ShowMeAI

视频	课件	笔记	代码
中英双语字幕	一键打包下载	官方笔记翻译	作业项目解析

视频 ·B 站 [扫码或点击链接]
https://www.bilibili.com/video/BV1Tf4y1L7wg

课件 & 代码 · 博客 [扫码或点击链接]
http://blog.showmeai.tech/adl4cv

神经网络可解释性 GNN transformer 相似度与度量学习 GAN Transfer learning 迁移学习 自注意力 model interpretability 图神经网络 Siamese neural network

Awesome AI Courses Notes Cheatsheets 是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 **TOP20+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets· 持续更新中			
知识图谱	**图机器学习**	**深度强化学习**	**自动驾驶**
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094

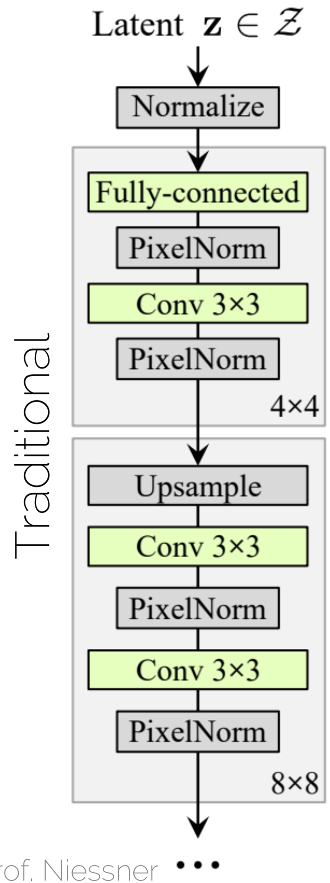
微信公众号

资料下载方式 2：扫码点击底部菜单栏

称为 **AI 内容创作者？** 回复 [添砖加瓦]

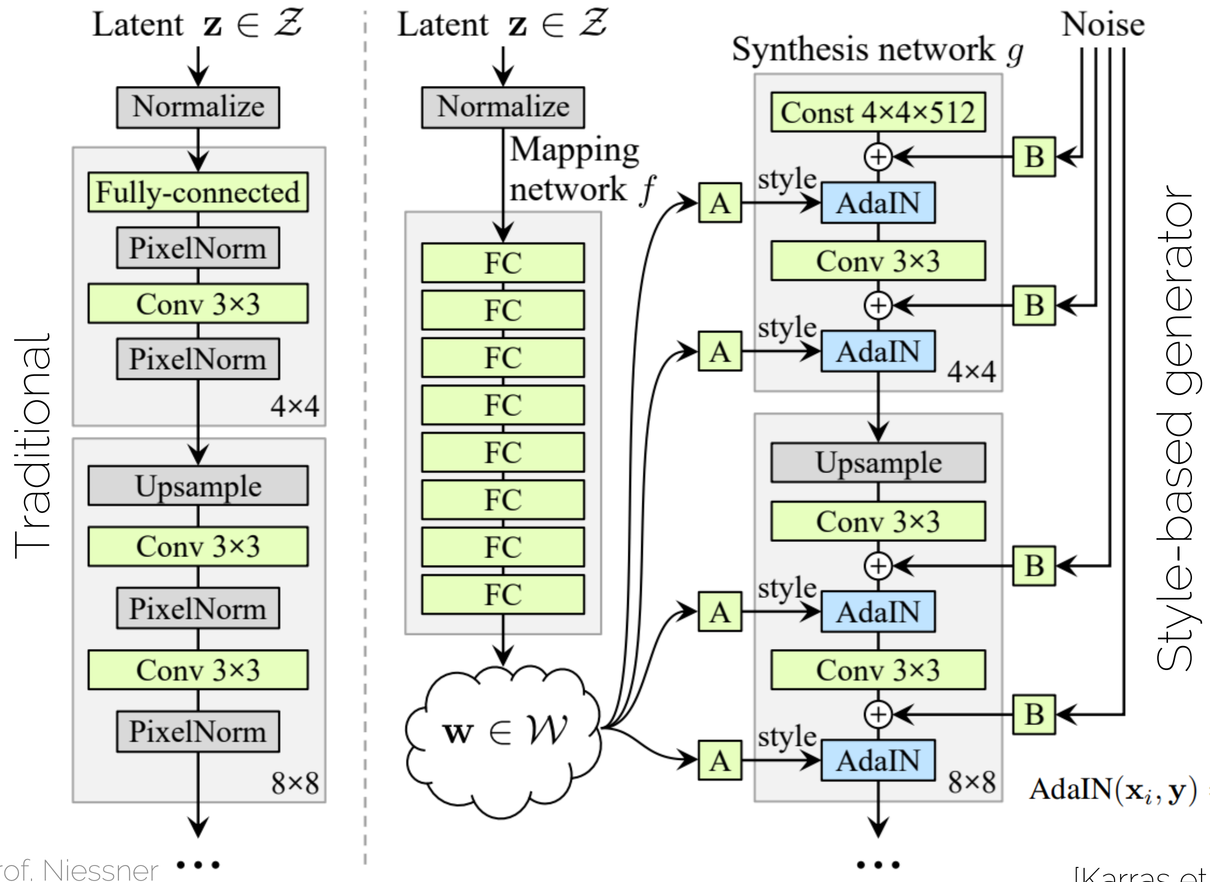
Style GAN

StyleGAN



[Karras et al. 19] StyleGAN

StyleGAN



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

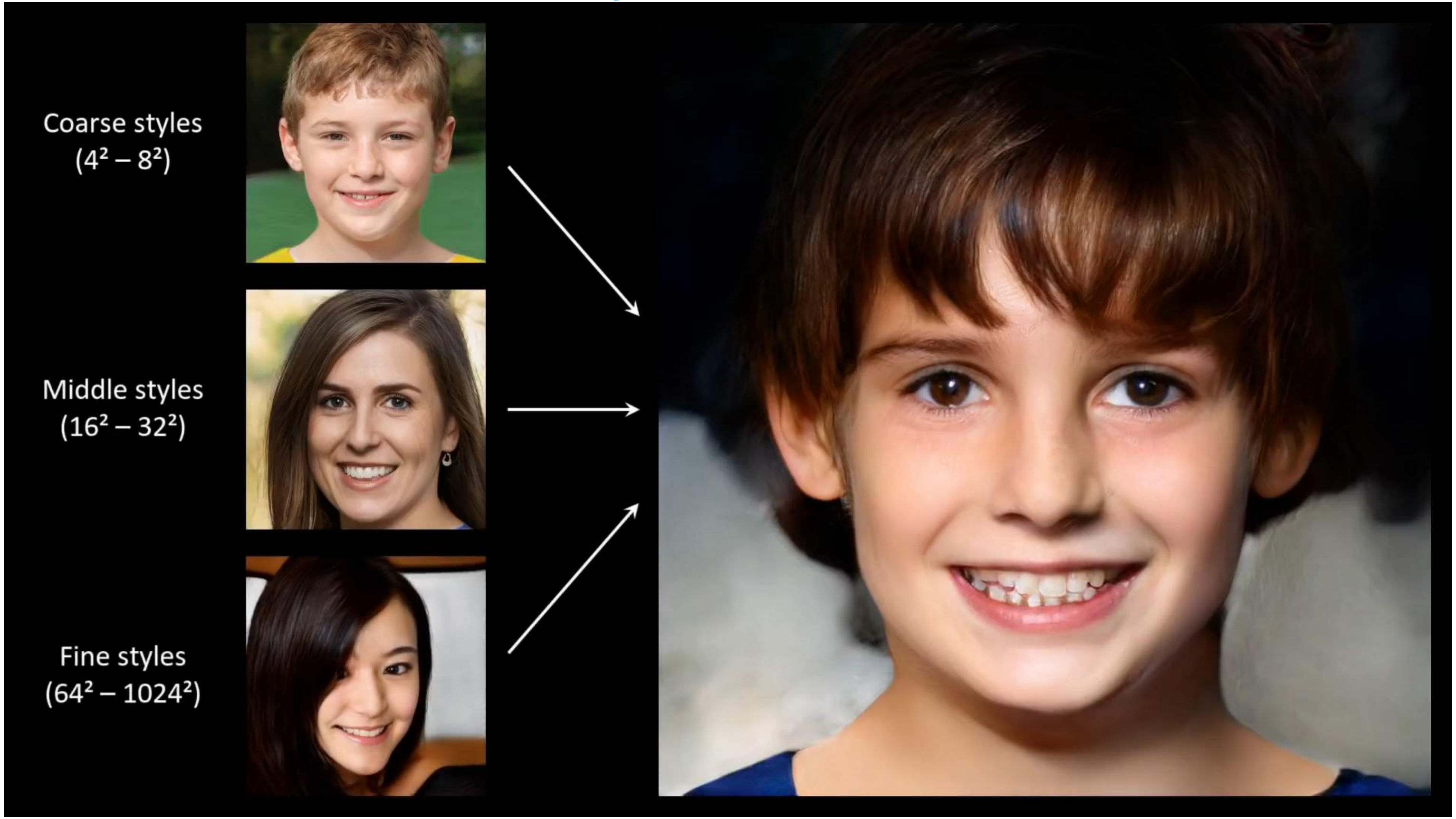
[Karras et al. 19] StyleGAN

StyleGAN

Method	CelebA-HQ	FFHQ
A Baseline Progressive GAN [30]	7.79	8.04
B + Tuning (incl. bilinear up/down)	6.11	5.25
C + Add mapping and styles	5.34	4.85
D + Remove traditional input	5.07	4.88
E + Add noise inputs	**5.06**	4.42
F + Mixing regularization	5.17	**4.40**

FID (Frechet inception distance) on 50k gen. images
-> Architecture is similar to Progressive Growing GAN

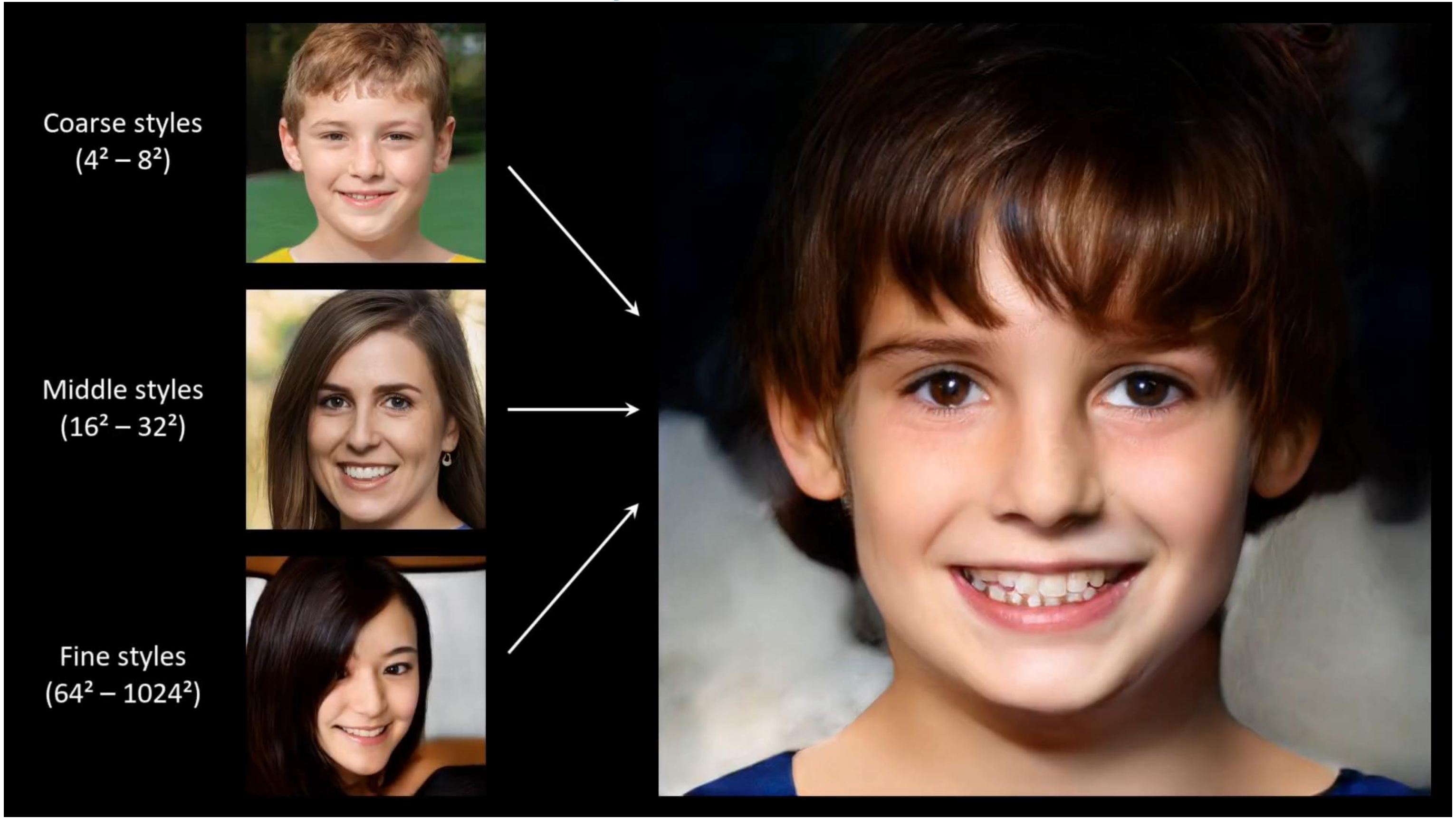
StyleGAN



https://youtu.be/kSLJriaOumA

[Karras et al. 19] StyleGAN

StyleGAN

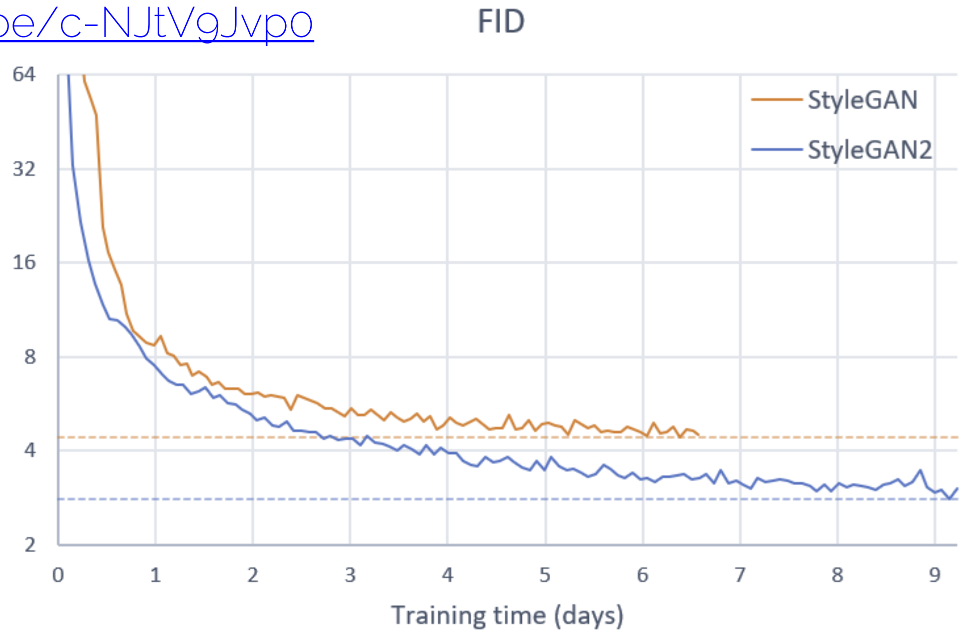


https://youtu.be/kSLJriaOumA
[Karras et al. 19] StyleGAN

StyleGAN2

Interesting analysis about design choices!

- https://arxiv.org/pdf/1912.04958.pdf
- https://github.com/NVlabs/stylegan2
- https://youtu.be/c-NJtV9Jvp0



Autoregressive Models

Autoregressive Models vs GANs

- GANs learn implicit data distribution
  - i.e., output are samples (distribution is in model)

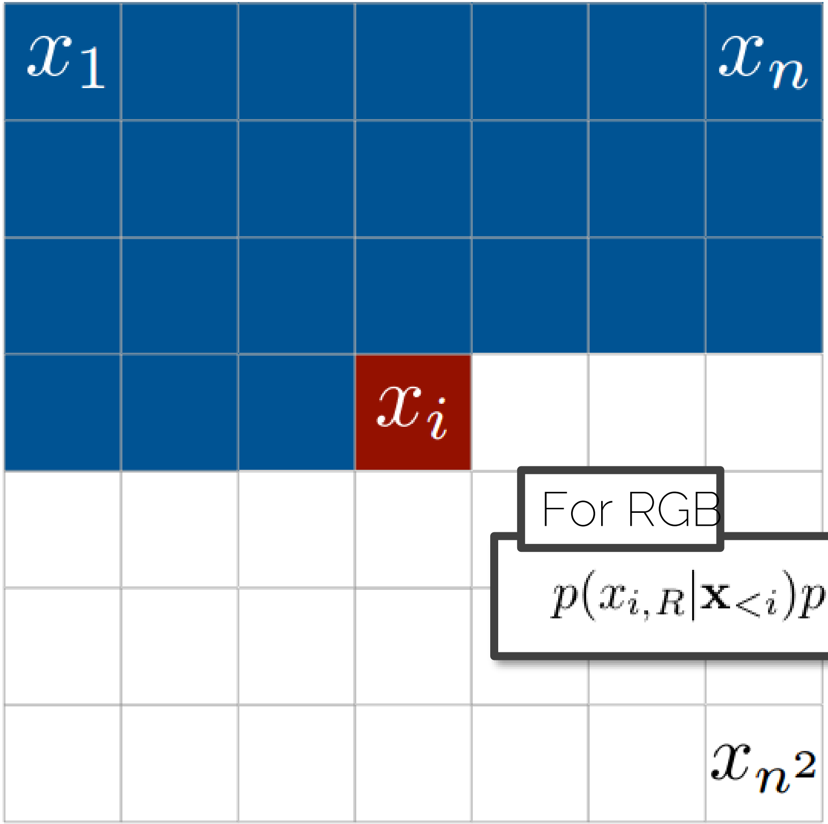
- Autoregressive models learn an explicit distribution governed by a prior imposed by model structure
  - i.e., outputs are probabilities (e.g., softmax)

PixelRNN

- Goal: model distribution of natural images
- Interpret pixels of an image as product of conditional distributions
  - Modeling an image → sequence problem
  - Predict one pixel at a time
  - Next pixel determined by all previously predicted pixels
  - ➢ Use a Recurrent Neural Network

[Van den Oord et al 2016]

PixelRNN



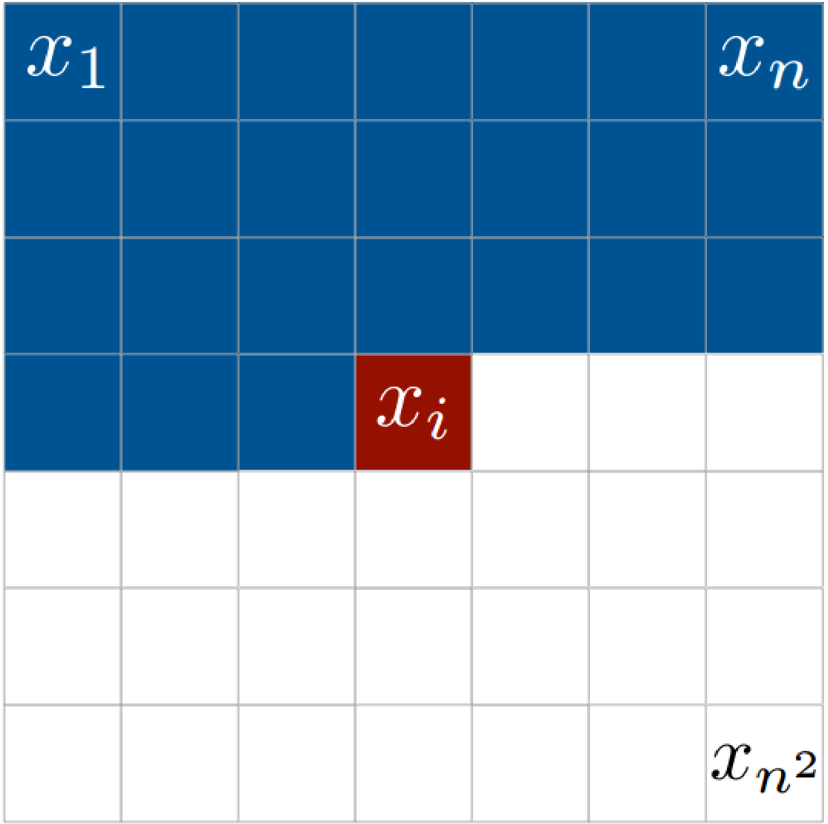
$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i | x_1, ..., x_{i-1})$$

For RGB

$$p(x_{i,R}|\mathbf{x}_{<i})p(x_{i,G}|\mathbf{x}_{<i}, x_{i,R})p(x_{i,B}|\mathbf{x}_{<i}, x_{i,R}, x_{i,G})$$

[Van den Oord et al 2016]

PixelRNN



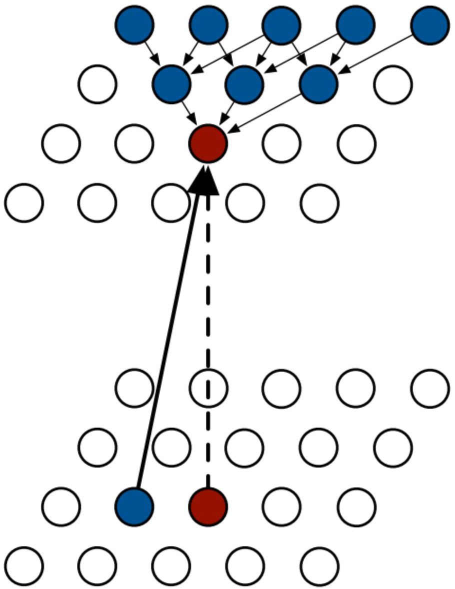
$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i | x_1, \ldots, x_{i-1})$$

$x_i \in [0,255]$
→ 256-way softmax

[Van den Oord et al 2016]

PixelRNN

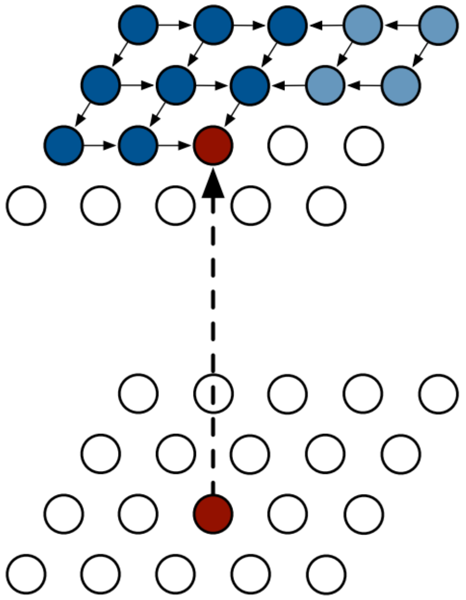
- Row LSTM model architecture
- Image processed row by row
- Hidden state of pixel depends on the 3 pixels above it
  - Can compute pixels in row in parallel
- Incomplete context for each pixel



[Van den Oord et al 2016]

PixelRNN

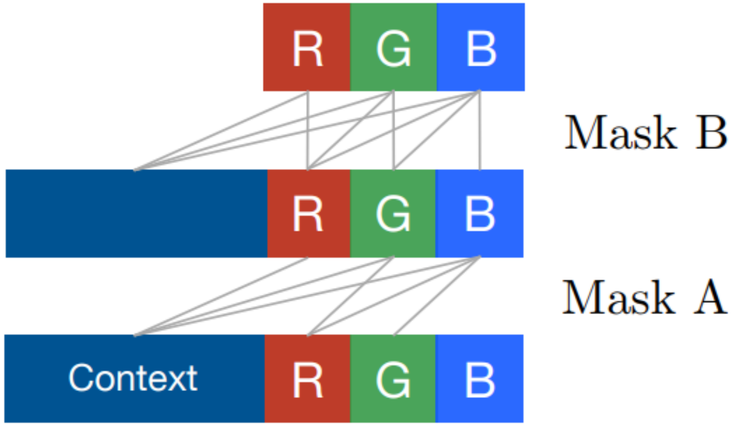
- Diagonal BiLSTM model architecture
- Solve incomplete context problem
- Hidden state of pixel $p_{i,j}$ depends on $p_{i,j-1}$ and $p_{i-1,j}$
- Image processed by diagonals



[Van den Oord et al 2016]

PixelRNN

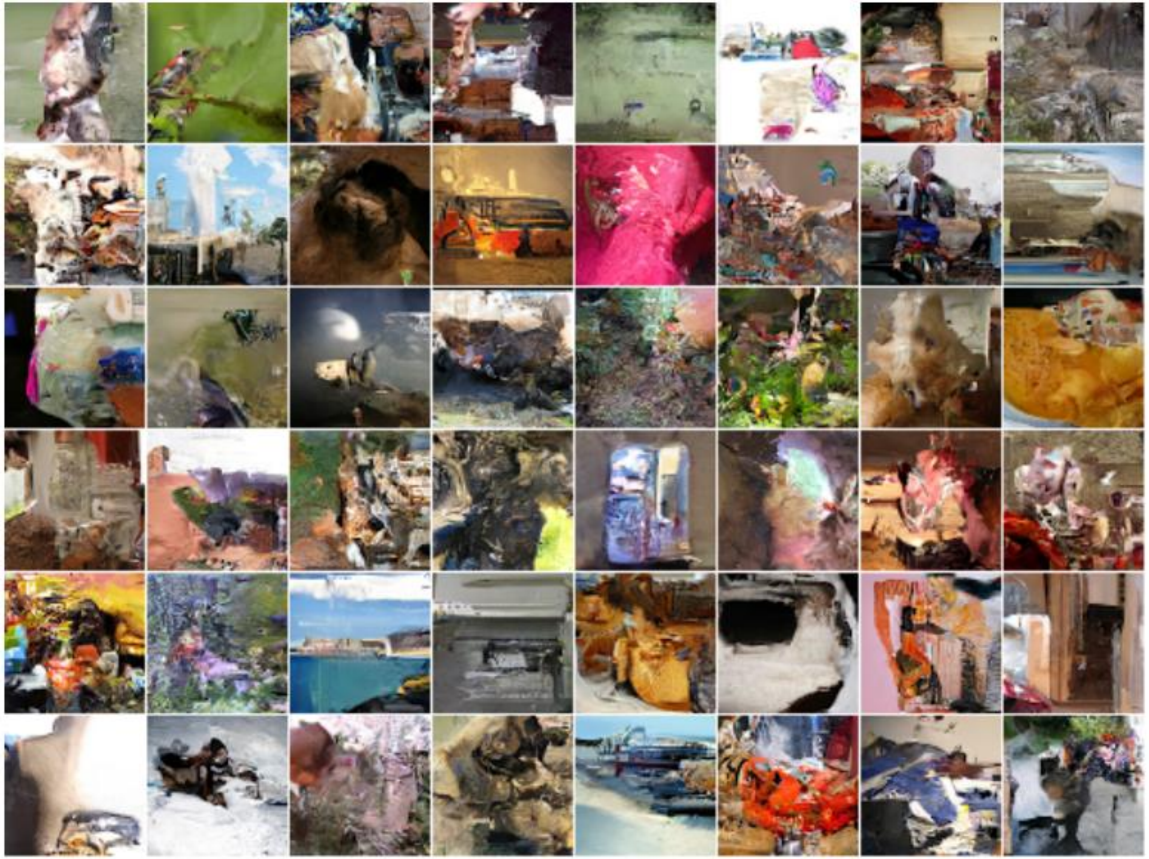
- Masked Convolutions
- Only previously predicted values can be used as context
- Mask A: restrict context during 1st conv
- Mask B: subsequent convs
- Masking by zeroing out values



[Van den Oord et al 2016]

PixelRNN

- Generated 64x64 images, trained on ImageNet



[Van den Oord et al 2016]

PixelCNN

- Row and Diagonal LSTM layers have potentially unbounded dependency range within the receptive field
  - Can be very computationally costly
- ➢ PixelCNN:
  - standard convs capture a bounded receptive field
  - All pixel features can be computed at once (during training)

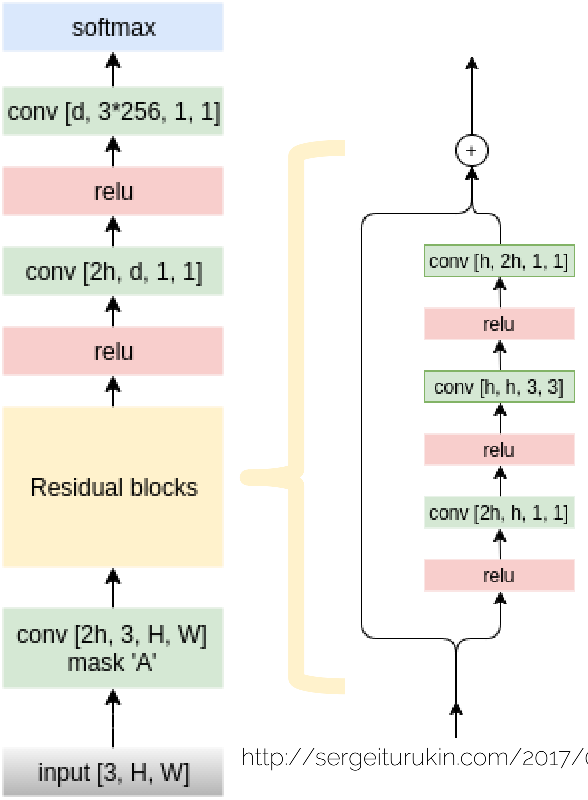
[Van den Oord et al 2016]

PixelCNN

- Model preserves spatial dimensions
- Masked convolutions to avoid seeing future context

1	1	1	1	1
1	1	1	1	1
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0

Mask A



http://sergeiturukin.com/2017/02/22/pixelcnn.h

[Van den Oord et al 2016]

Gated PixelCNN

- Gated blocks
- Imitate multiplicative complexity of PixelRNNs to reduce performance gap between PixelCNN and PixelRNN
- Replace ReLU with gated block of sigmoid, tanh

k^th layer sigmoid

$$y = \tanh\left(W_{k,f} * x\right) \odot \sigma\left(W_{k,g} * x\right)$$

element-wise product convolution

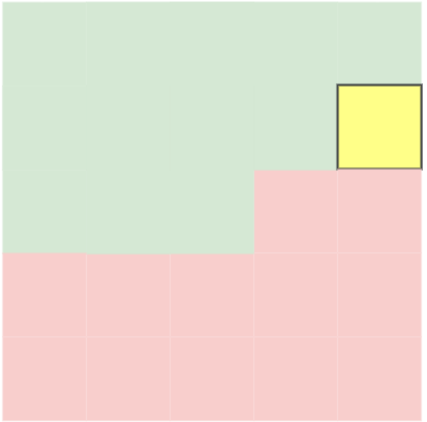
[Van den Oord et al 2016]

PixelCNN Blind Spot

1	1	1	1	1
1	1	1	1	1
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0

5x5 image / 3x3 conv

Receptive Field



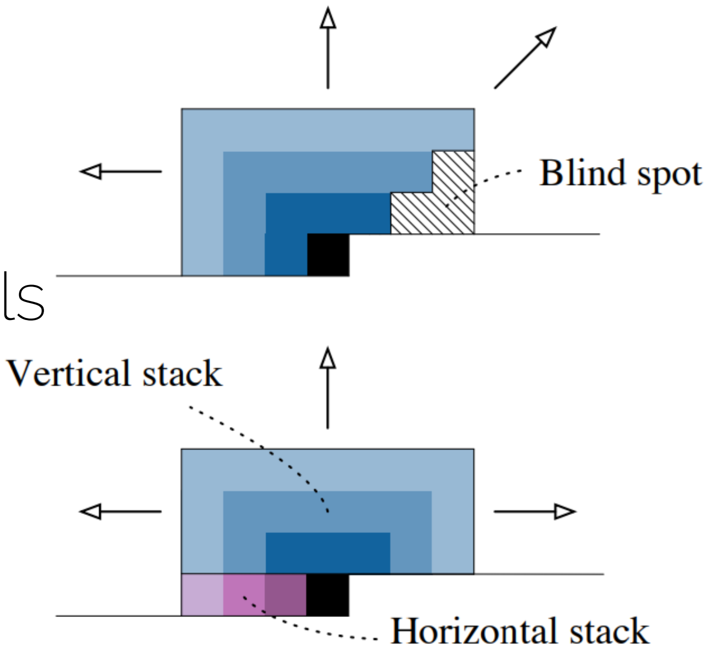
Unseen context

http://sergeiturukin.com/2017/02/24/gated-pixelcnn.html

[Van den Oord et al 2016]

PixelCNN: Eliminating Blind Spot

- Split convolution to two stacks
- Horizontal stack conditions on current row
- Vertical stack conditions on pixels above



Conditional PixelCNN

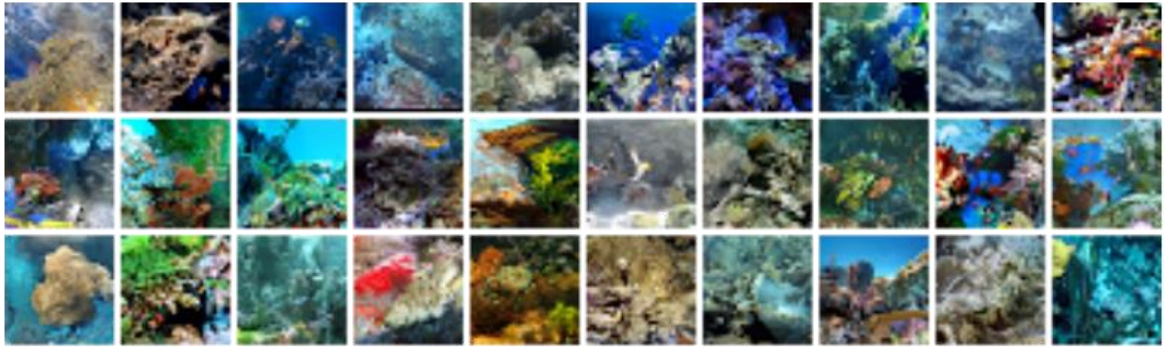
- Conditional image generation
- E.g., condition on semantic class, text description

latent vector to be conditioned on

$$y = \tanh\left(W_{k,f} * x + V_{k,f}^T h\right) \odot \sigma(W_{k,g} * x + V_{k,g}^T h)$$

[Van den Oord et al 2016]

Conditional PixelCNN



Coral Reef

Sorrel horse

[Van den Oord et al 2016]

Autoregressive Models vs GANs

- Advantages of autoregressive:
  - Explicitly model probability densities
  - More stable training
  - Can be applied to both discrete and continuous data
- Advantages of GANs:
  - Have been empirically demonstrated to produce higher quality images
  - Faster to train

Autoregressive Models

- State of the art is pretty impressive ☺

Vector Quantized Variational AutoEncoder

Generating Diverse High-Fidelity Images with VQ-VAE-2
https://arxiv.org/pdf/1906.00446.pdf [Razavi et al. 19]

Generative Models on Videos

GANs on Videos

Two options

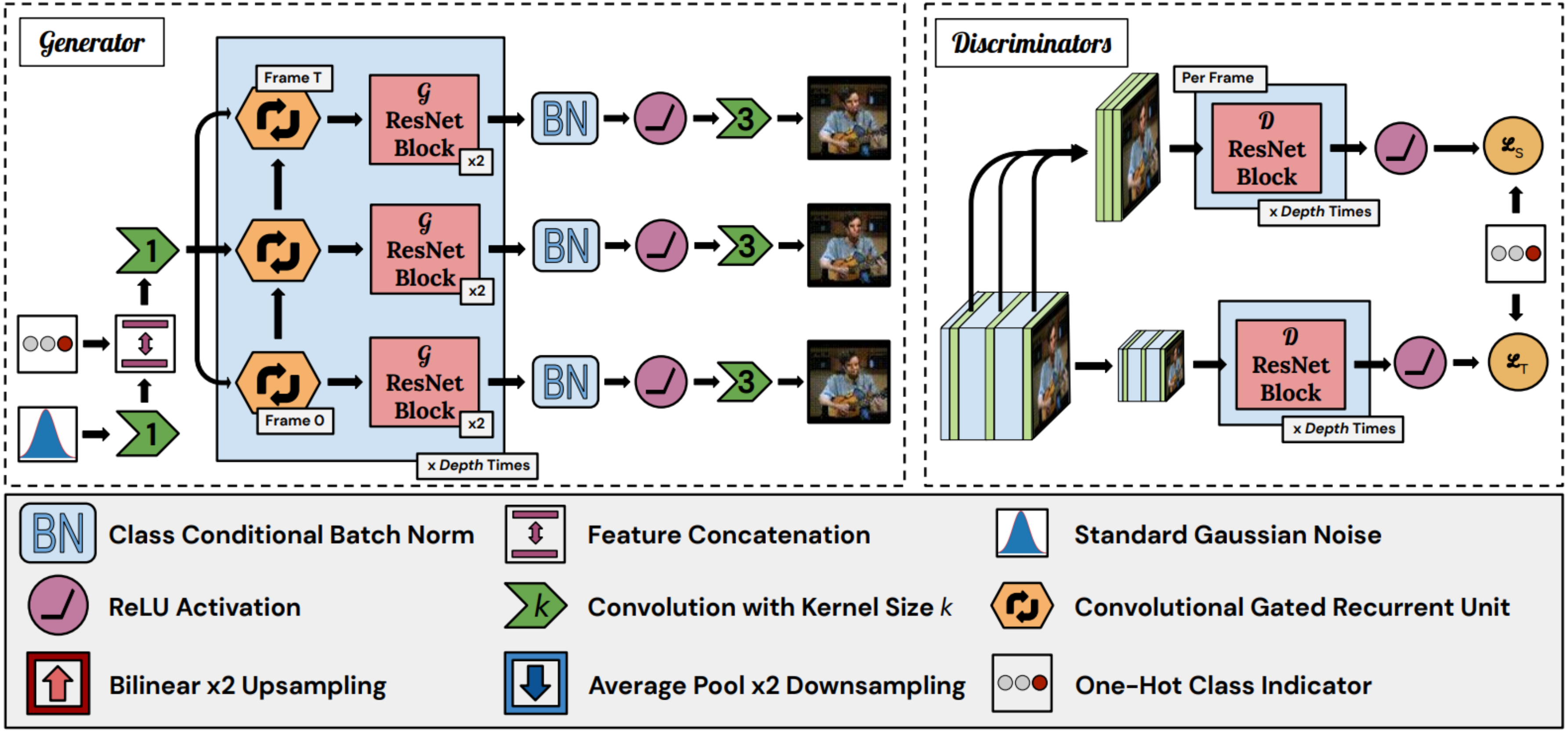
- Single random variable z seeds entire video (all frames)
  - Very high dimensional output
  - How to do for variable length?
  - Future frames deterministic given past

- Random variable z for each frame of the video
  - Need conditioning for future from the past
  - How to get combination of past frames + random vectors during training

General issues

- Temporal coherency
- Drift over time (many models collapse to mean image)

GANs on Videos: DVD-GAN



[Clark et al. 2019] Adversarial Video Generation on Complex Datasets

GANs on Videos: DVD-GAN

[Clark et al. 2019] Adversarial Video Generation on Complex Datasets

GANs on Videos: DVD-GAN

- Trained on Kinetics-600 dataset
  - 256 x 256, 128 x 128, and 64 x 64
  - Lengths of up 48 frames

-> This is state of the art!

-> Videos from scratch still incredibly challenging

[Clark et al. 2019] Adversarial Video Generation on Complex Datasets

Conditional GANs on Videos

- Challenge:
  - Each frame is high quality, but temporally inconsistent

Video-to-Video Synthesis

- Sequential Generator:

$$p(\tilde{\mathbf{x}}_1^T | \mathbf{s}_1^T) = \prod_{t=1}^{T} p(\tilde{\mathbf{x}}_t | \tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^{t}).$$

past L generated frames   past L source frames

(set L = 2)

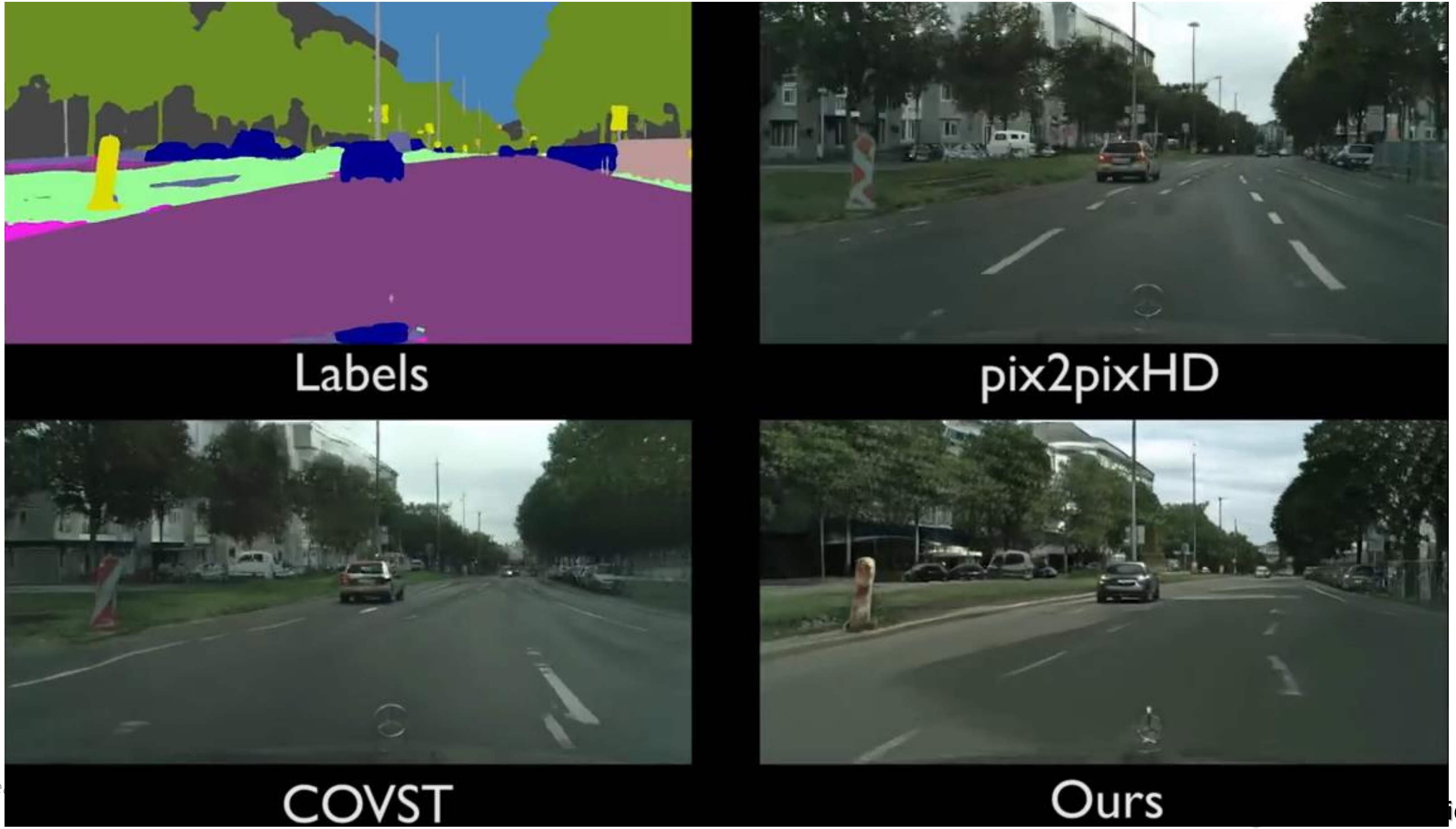
- Conditional Image Discriminator $\boldsymbol{D_i}$ (is it real image)
- Conditional Video Discriminator $\boldsymbol{D_v}$ (temp. consistency via flow)

Full Learning Objective:

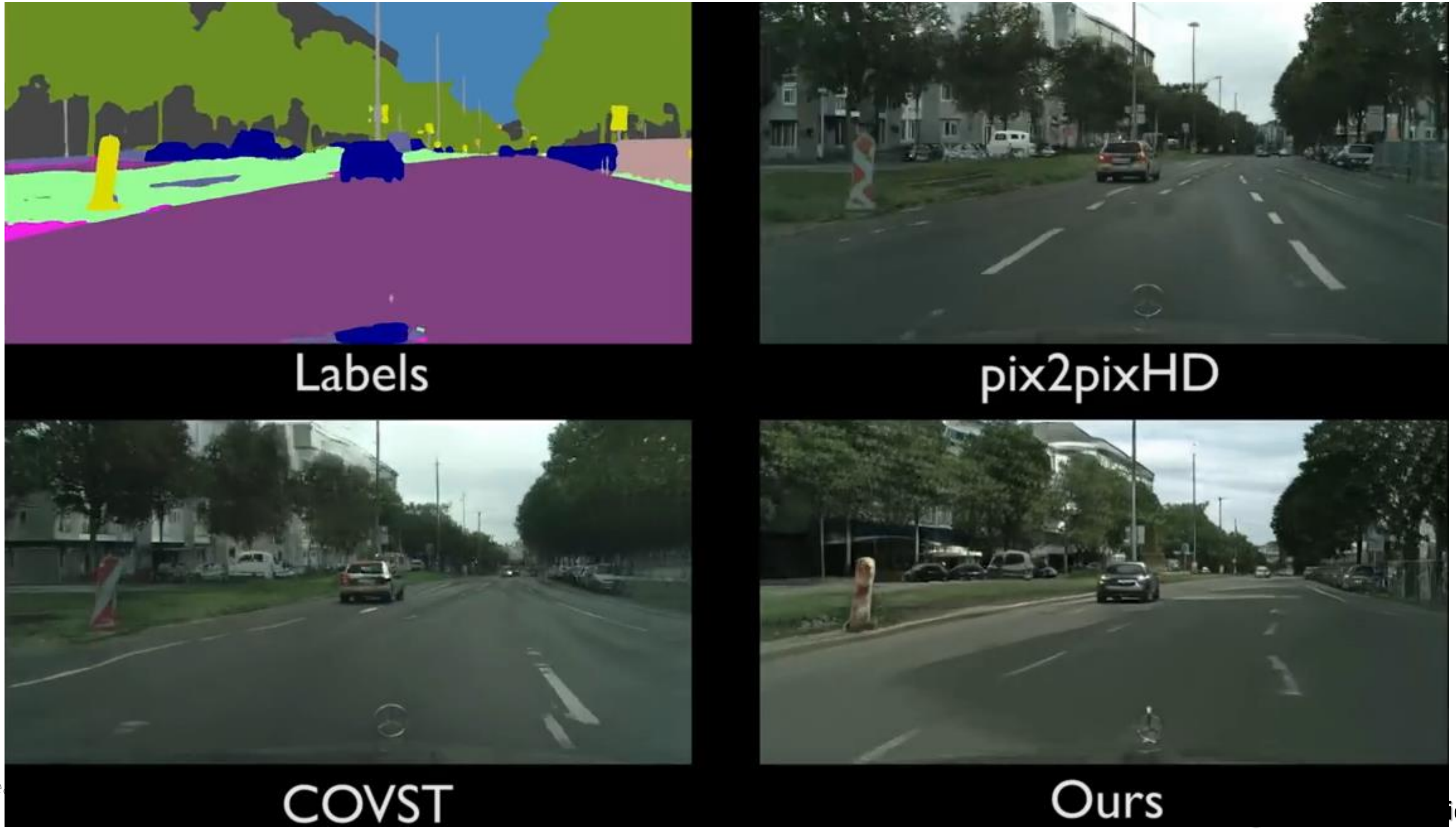
$$\min_F \left(\max_{D_I} \mathcal{L}_I(F, D_I) + \max_{D_V} \mathcal{L}_V(F, D_V) \right) + \lambda_W \mathcal{L}_W(F),$$

Wang et al. 18: Vid2Vid

Video-to-Video Synthesis



Video-to-Video Synthesis



Video-to-Video Synthesis

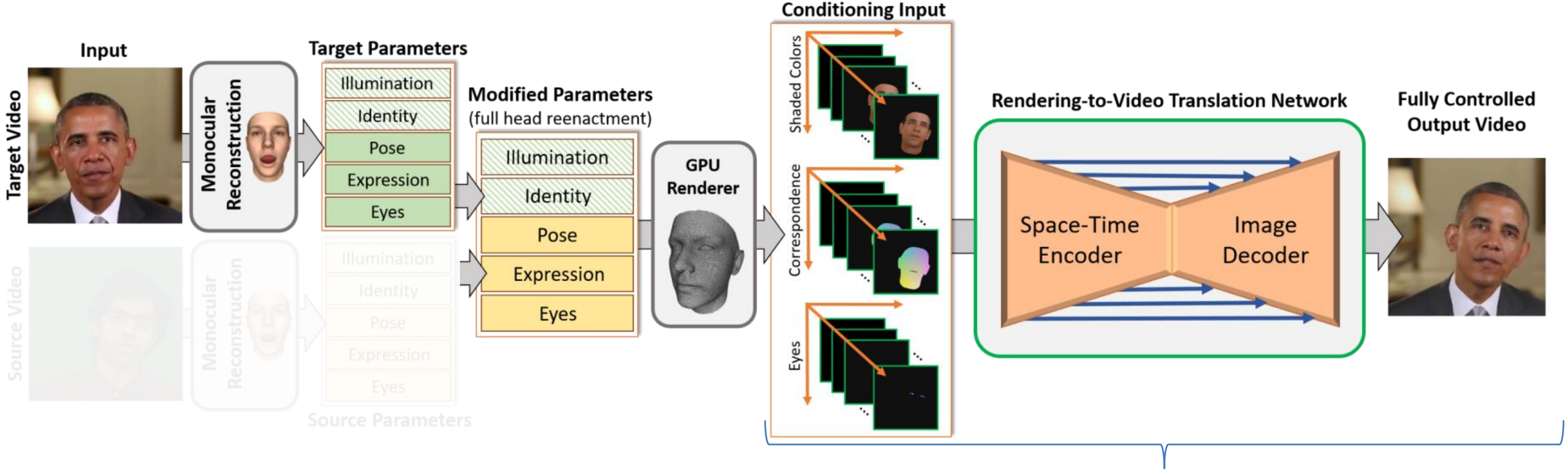
- Key ideas:
  - Separate discriminator for temporal parts
    - In this case based on optical flow
  - Consider recent history of prev. frames
  - Train all of it jointly

Wang et al. 18: Vid2Vid

Deep Video Portraits

Siggraph'18 [Kim et al 18]: Deep Portraits

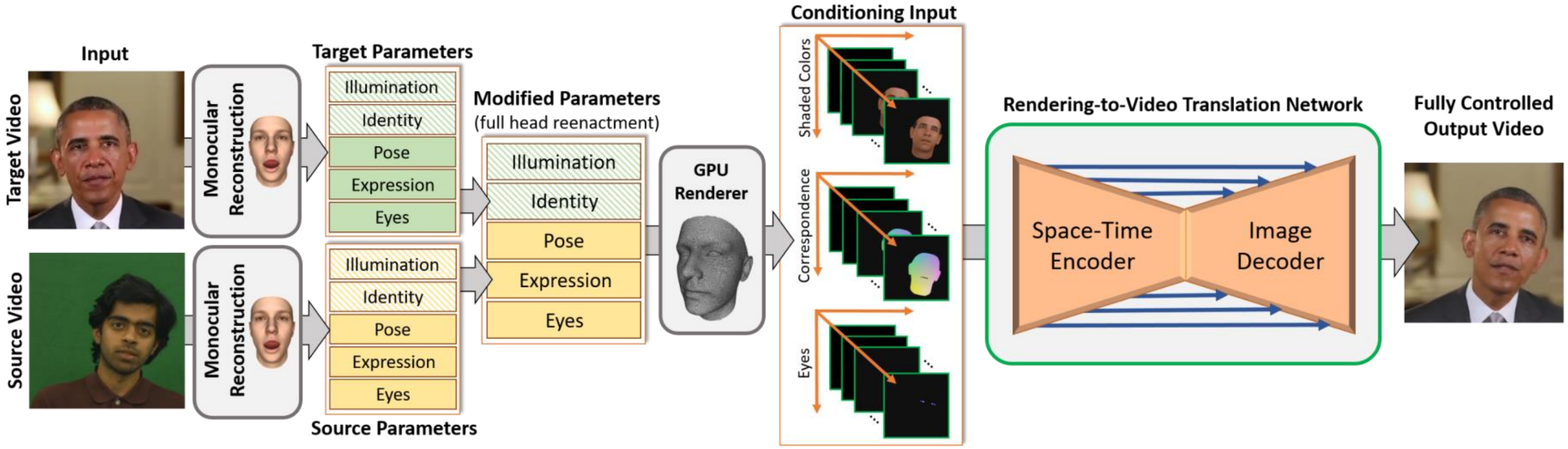
Deep Video Portraits



Similar to "Image-to-Image Translation" (Pix2Pix) [Isola et al.]

Siggraph'18 [Kim et al 18]: Deep Portraits

Deep Video Portraits



Siggraph'18 [Kim et al 18]: Deep Portraits

Deep Video Portraits

Source Sequence Conditioning Images Result

Neural Network converts synthetic data to realistic video

Siggraph'18 [Kim et al 18]: Deep Portraits

Deep Video Portraits

Source Sequence — Conditioning Images — Result

Siggraph'18 [Kim et al 18]: Deep Portraits

Deep Video Portraits

Source Sequence — Conditioning Images — Result

Deep Video Portraits

Siggraph'18 [Kim et al 18]: Deep Portraits

Deep Video Portraits

Interactive Video Editing

2x speed

Siggraph'18 [Kim et al 18]: Deep Portraits

Deep Video Portraits: Insights

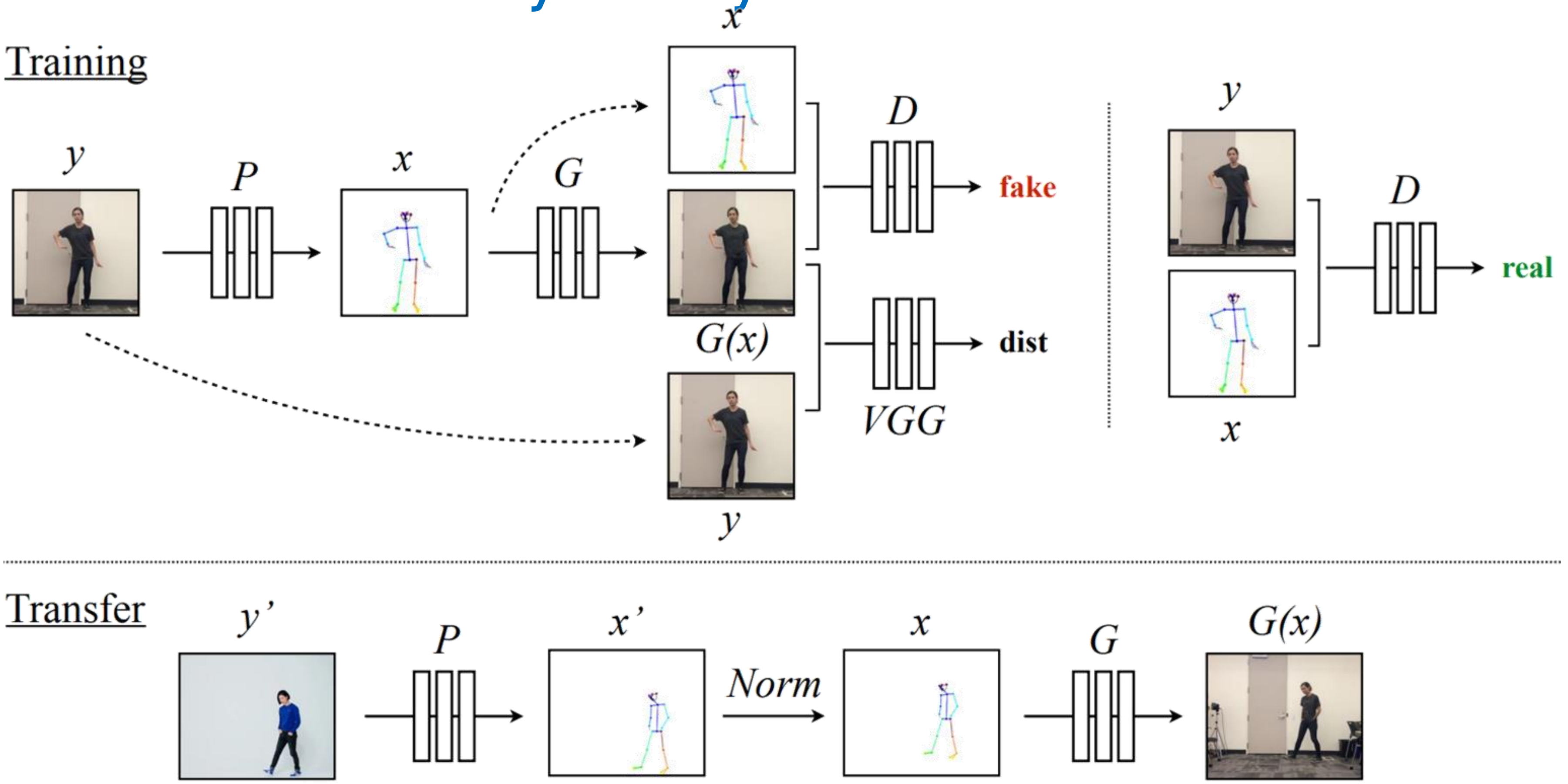
- Synthetic data for tracking is great anchor / stabilizer
- Overfitting on small datasets works pretty well
- Need to stay within training set w.r.t. motions
- No real learning; essentially, optimizing the problem with SGD

  -> should be pretty interesting for future directions

Everybody Dance Now

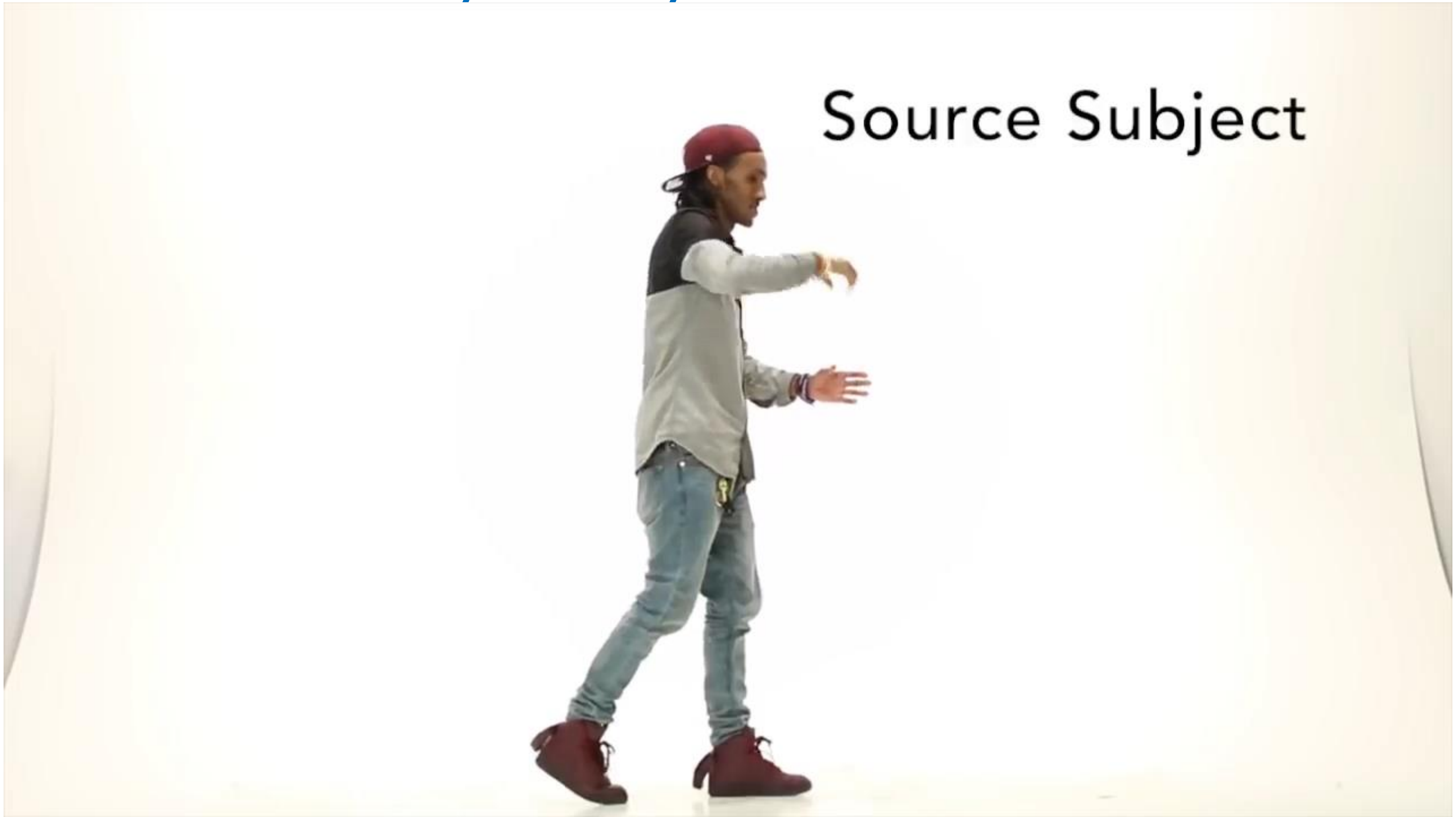
[Chan et al. '18] Everybody Dance Now

Everybody Dance Now



[Chan et al. '18] Everybody Dance Now

Everybody Dance Now



[Chan et al. '18] Everybody Dance Now

Everybody Dance Now

- cGANs work with different input

- Requires consistent input i.e., accurate tracking

- Network has no explicit 3D notion

[Chan et al. '18] Everybody Dance Now

Everybody Dance Now: Insights

- Conditioning via tracking seems promising!
  - Tracking quality translates to resulting image quality
  - Tracking human skeletons is less developed than faces
    - Temporally it's not stable… (e.g., OpenPose etc.)
  - Fun fact, there were like 4 papers with a similar idea that appeared around the same time…

[Chan et al. '18] Everybody Dance Now

Next Lectures

- Next Lectures:
  - Neural Rendering
  - 3D Deep Learning

- Keep working on the projects!

See you next week ☺

TUM·adl4cv | Advanced Deep Learning for Computer Vision(2020)

adl4cv (2020)· 课程资料包 @ShowMeAI

视频	课件	笔记	代码
中英双语字幕	一键打包下载	官方笔记翻译	作业项目解析

视频 ·B 站 [扫码或点击链接]
https://www.bilibili.com/video/BV1Tf4y1L7wg

课件 & 代码 · 博客 [扫码或点击链接]
http://blog.showmeai.tech/adl4cv

神经网络可解释性 GNN transformer 相似度与度量学习 GAN Transfer learning 迁移学习 自注意力 model interpretability 图神经网络 Siamese neural network

Awesome AI Courses Notes Cheatsheets 是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 **TOP20+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets· 持续更新中			
知识图谱	**图机器学习**	**深度强化学习**	**自动驾驶**
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094

微信公众号

资料下载方式 2：扫码点击底部菜单栏

称为 **AI 内容创作者？** 回复 [添砖加瓦]