GAN

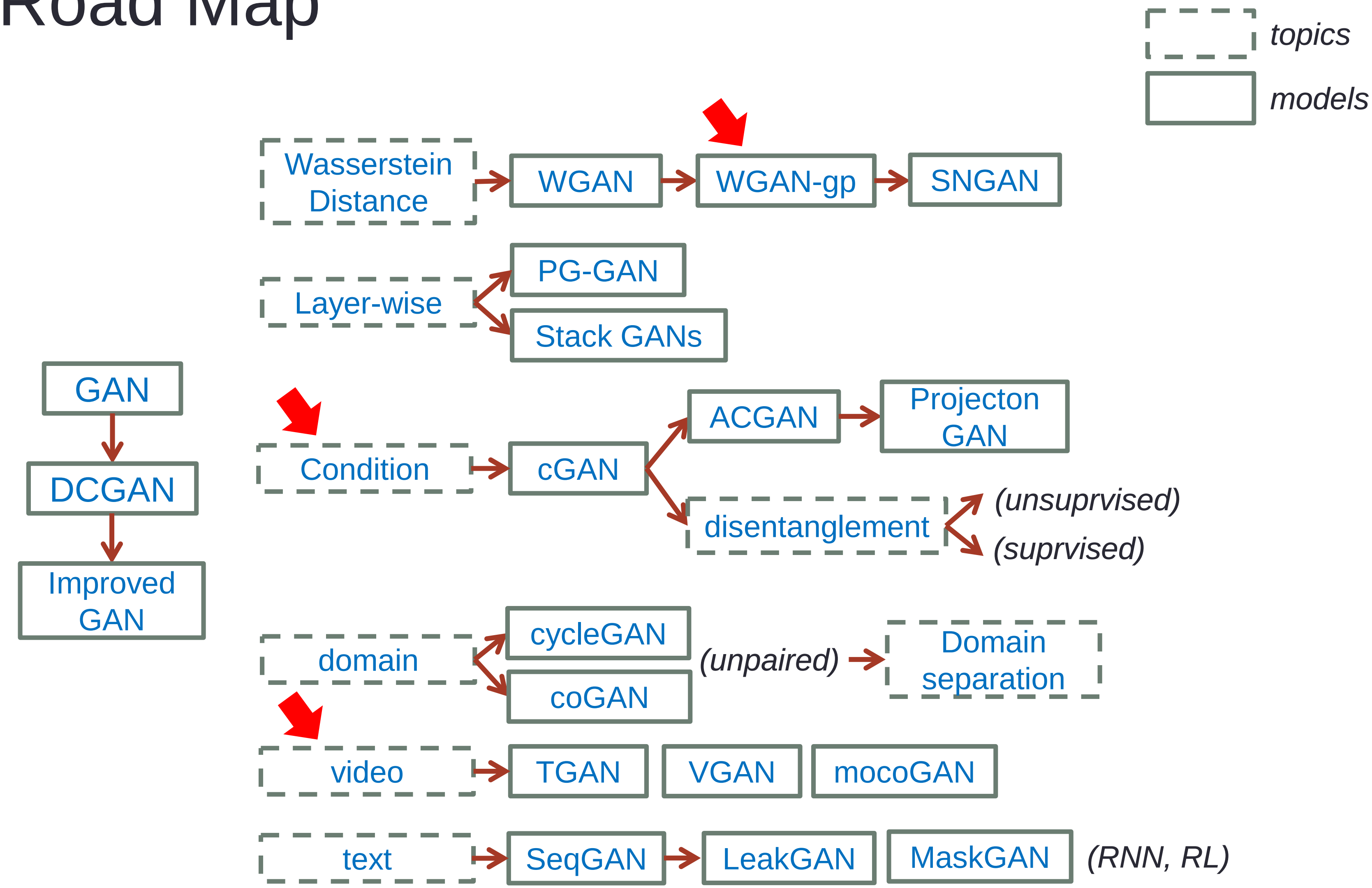
2018/7/11

Preface

- Topics
  - Experiences & tricks
  - Paper Survey
  - MuseGAN, MidiNet Review

- Outline
  - GAN Review
  - Road Map
    - Condition
    - Video

Road Map



topics
models
Wasserstein Distance
WGAN
WGAN-gp
SNGAN
Layer-wise
PG-GAN
Stack GANs
GAN
DCGAN
Improved GAN
Condition
cGAN
ACGAN
Projecton GAN
disentanglement
(unsuprvised)
(suprvised)
domain
cycleGAN
coGAN
(unpaired)
Domain separation
video
TGAN
VGAN
mocoGAN
text
SeqGAN
LeakGAN
MaskGAN
(RNN, RL)

GAN Review

- Data & Generator
- Discriminator
- Training
- Testing
- Discussion

Review:
How to start a GAN project?

- Data
- Generator
- Discriminator
- Training
- Testing
- **State-of-the-art?**

[–] **ajmooch** 6 points 3 months ago

Projection discriminator + SN-GAN + progressive growing is probably our best bet at the moment for highest-fidelity high-res images, but the resources to do an ImageNet-level variety of classes at high res will probably be pretty substantial (just to get everything tuned, let alone to train it).

- **Models:** DCGAN WGAN-gp SNGAN

* *https://www.reddit.com/r/MachineLearning/comments/890prh/r_memgen_memory_is_all_you_need_generative/*

Data & Generator

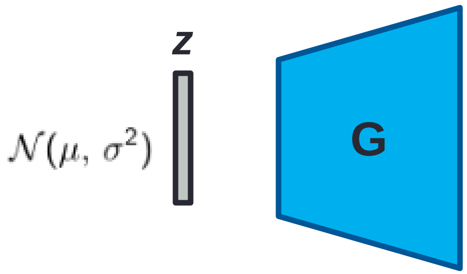
- No matter what your task is, Normalization on data is necessary.
- The activation of the output layer of the generator depends on the range of the data
  - Bounded

    [-1, 1] tanh
    [0, 1] sigmoid

  - Unbounded

    zero means unit variance
    logarithm (ex: on spectrogram)

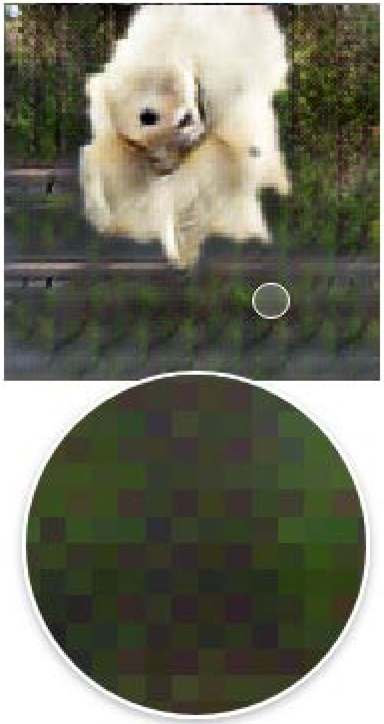
    leaky ReLU, ReLU?



- SharedArray & Shuffle data by index at runtime
- ***z***: sampled from a Gaussian distribution

Generator

- Batch Normalization is essential to the quality
- Upsampling Layers:
  - Deconvolution (transposed convolution)
    - The most common method
    - Checkerboard artifacts?
  - Resize-Convolution
    - ex: PG-GAN uses nearest neighbor upscaling
  - Pixel Shuffling
    - Super resolution



Salimans et al., 2016 [2]

- Residual Blocks can enhance the quality as well, especially when the size of the images is large and the corresponding network is deep.

* ***PG-GAN:*** *https://arxiv.org/pdf/1710.10196.pdf*
* *https://distill.pub/2016/deconv-checkerboard*

Discriminator

DCGAN

- BN applied
- Sigmoid output

WGAN-gp

- No BN
- No Sigmoid

SNGAN

- No BN
- Spectral Norm
- No Sigmoid

- Loss function:

Original GAN loss: DCGAN

$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

D Loss: - log (D(X)) - log (1 – D(G(z))
G Loss: - log (D(G(z)))

Minimize!!

In practice…

Discriminator

WGAN loss: $\boxed{\text{SNGAN}}$

D Loss: - (D(X)) + (D(G(z)))
G Loss: - (D(G(z)))

Minimize!!

remove log

Gradient penalty: $\boxed{\text{WGAN-gp}}$

D Loss: - (D(X)) + (D(G(z))) + $\underbrace{(\| \nabla_x D(x) \| - 1)^2}_{\text{penalty}}$

G Loss: - (D(G(z)))

Minimize!!

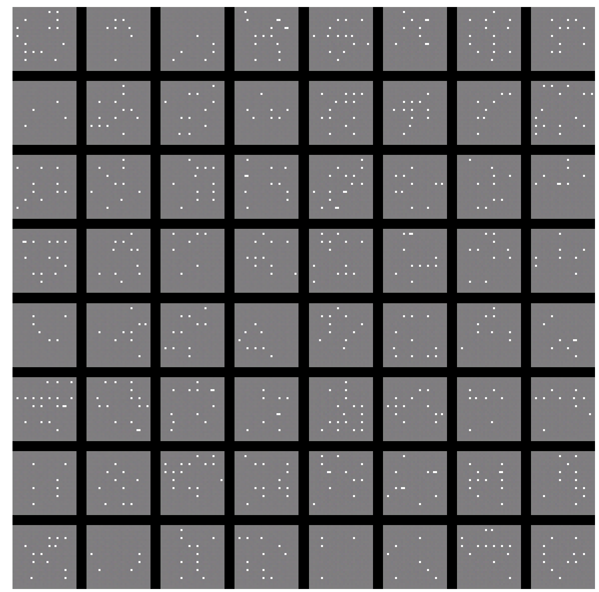
- Parameter Updating:
  - $\boxed{\text{WGAN-gp}}$ update **G** once every 5 updates of **D**

Training

- When to stop?
- How to monitor the training procedure?
- GAN Losses cannot truly reflect the quality (even for WGAN)
- Generate samples along training!

Samples/*.png:

sample_2.png	2018/7/7 下午 06:42
sample_142.png	2018/7/7 下午 06:43
sample_282.png	2018/7/7 下午 06:43
sample_422.png	2018/7/7 下午 06:43
sample_562.png	2018/7/7 下午 06:43
sample_702.png	2018/7/7 下午 06:43
sample_842.png	2018/7/7 下午 06:44
sample_982.png	2018/7/7 下午 06:44
sample_2002.png	2018/7/7 下午 06:45
sample_4002.png	2018/7/7 下午 06:48
sample_6002.png	2018/7/7 下午 06:50
sample_8002.png	2018/7/7 下午 06:53
sample_10002.png	2018/7/7 下午 06:55

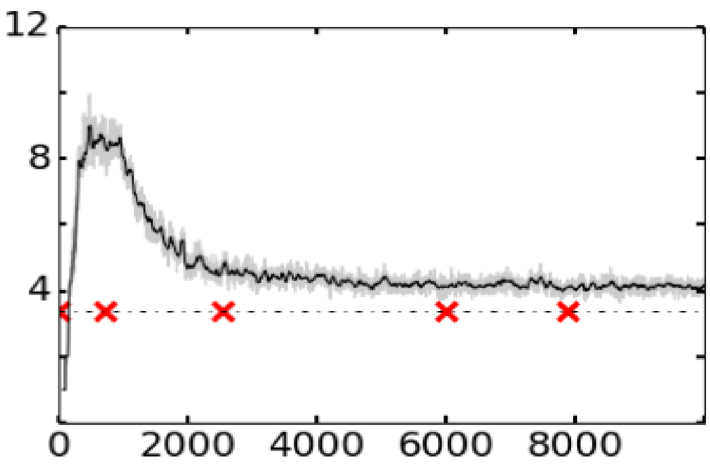


(SNGAN on CIFAR10)

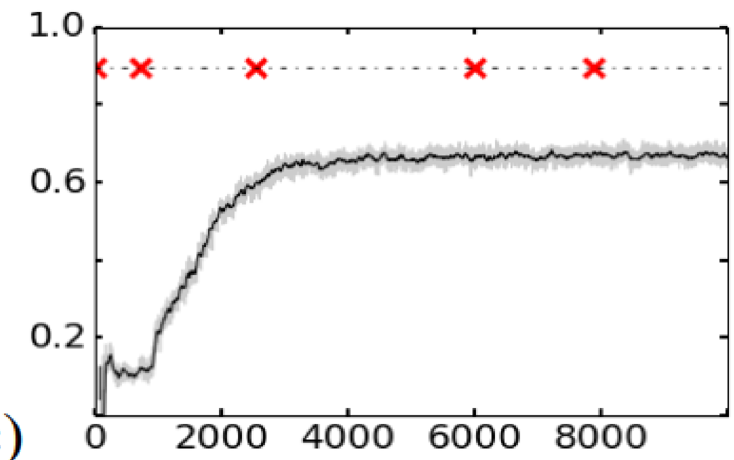
Early Stopping

- Evaluate along training!

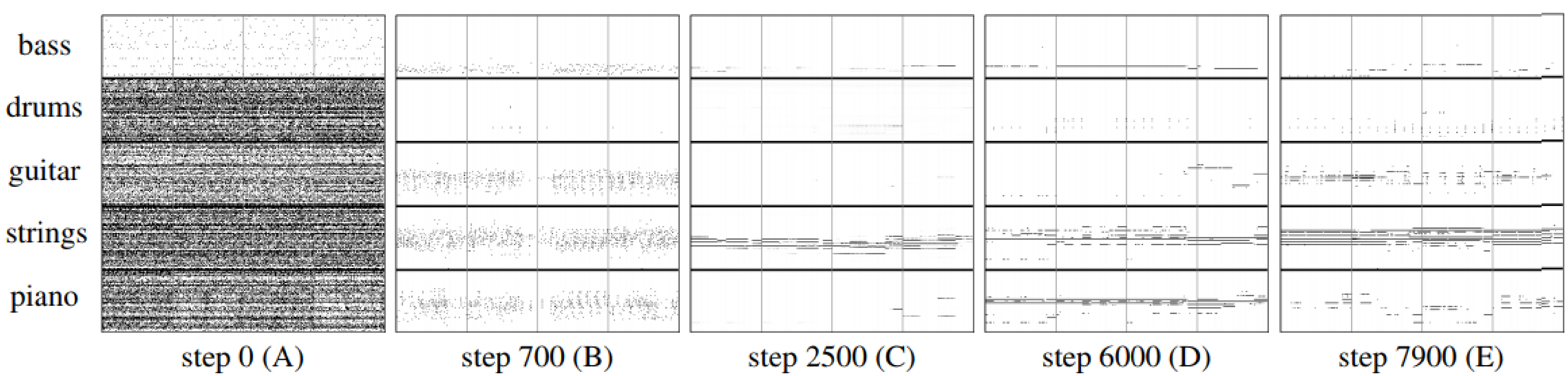
(b)



(c)



UPC **QN**



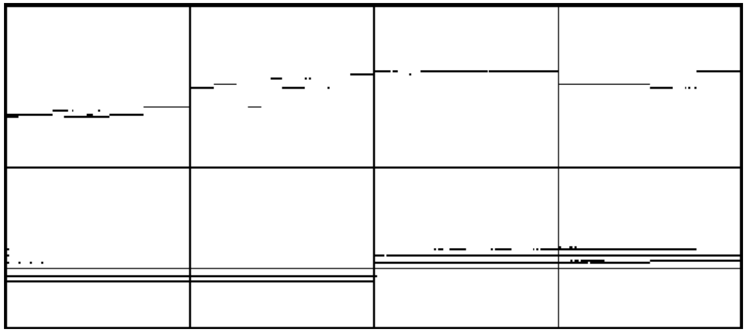
**** from MuseGAN***

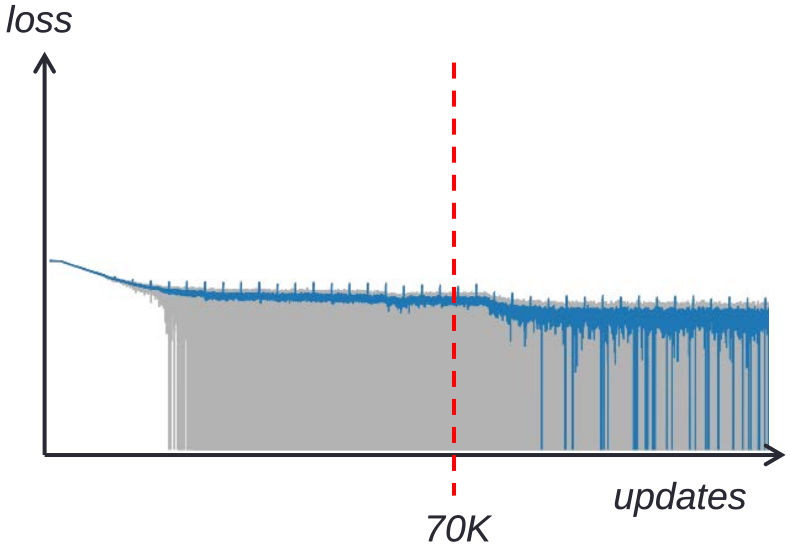
Training

- For the first time, train the model for a longer period of time and monitor the procedure.

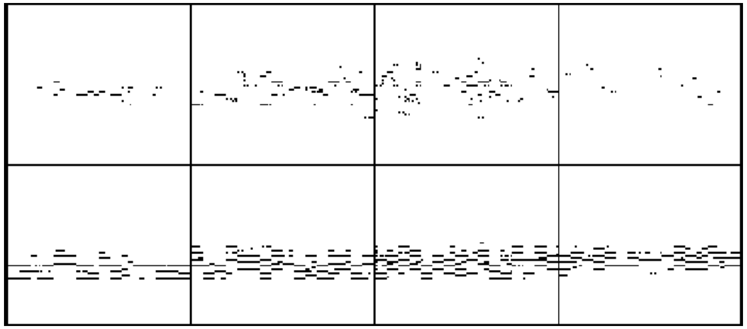
- Possible failure:
  Mode collapse
  Degeneration

(70K updates)





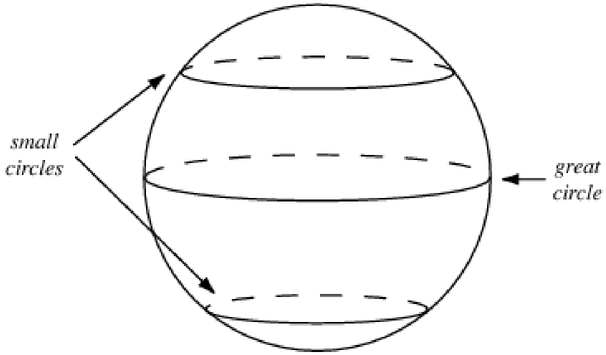
(130K updates)



**** MuseGAN (with WGAN-gp), lead sheet generation***

Testing

- Interpolation: Spherical (instead of linear)



- Batch normalization **off**

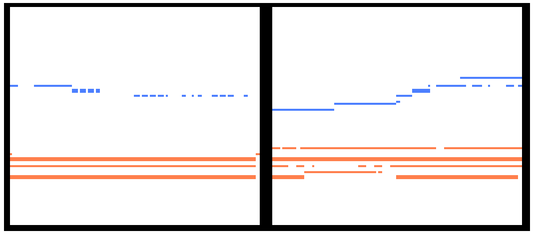
  'is_training' in TensorFlow

- Dropout **on?**
  - Image-to-Image (https://arxiv.org/pdf/1611.07004v1.pdf) turns on when testing
  - Later works seldom use it

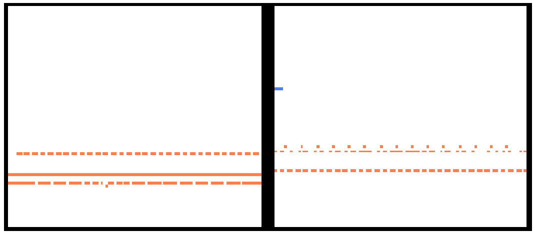
**** GAN hacks:** https://github.com/soumith/ganhacks*
**** Figure: from** https://github.com/ptrblck/prog_gans_pytorch_inference*

Discussion

- So, which model should I use?
- Loss
  - For image, SNGAN outperforms others in both of th efficiency and the quality.
  - For discrete (binary) representation, only WGAN-gp can generate results successfully.
- Network design
  - Deconvolution is simple but powerful
  - If you aim to generate images with higher qualities, try to use residual blocks and resize-upsampling layers.



(WGAN-gp on Lead Sheet)



(SNGAN on Lead Sheet)

Conditional GANs

How to apply condition

Tag? Disentanglement

Discussion

How to apply condition?

- Input Concatenation
  - cGAN: https://arxiv.org/pdf/1411.1784.pdf
- Feature Map Concatenation
  - DCGAN: https://github.com/carpedm20/DCGAN-tensorflow
- Auxiliary Classifier
  - ACGAN: https://arxiv.org/pdf/1610.09585.pdf
- Encoder
  - *S^2 GAN,* Generative Adversarial Text to Image Synthesis
  - MuseGAN/MidiNet: https://arxiv.org/pdf/1703.10847.pdf
  - FTGAN: https://arxiv.org/pdf/1711.09618.pdf
- Projection Discriminator (ICLR, 2018)
  - https://arxiv.org/pdf/1802.05637.pdf
- Principle:
  - Both of G and D need to receive the conditional information.

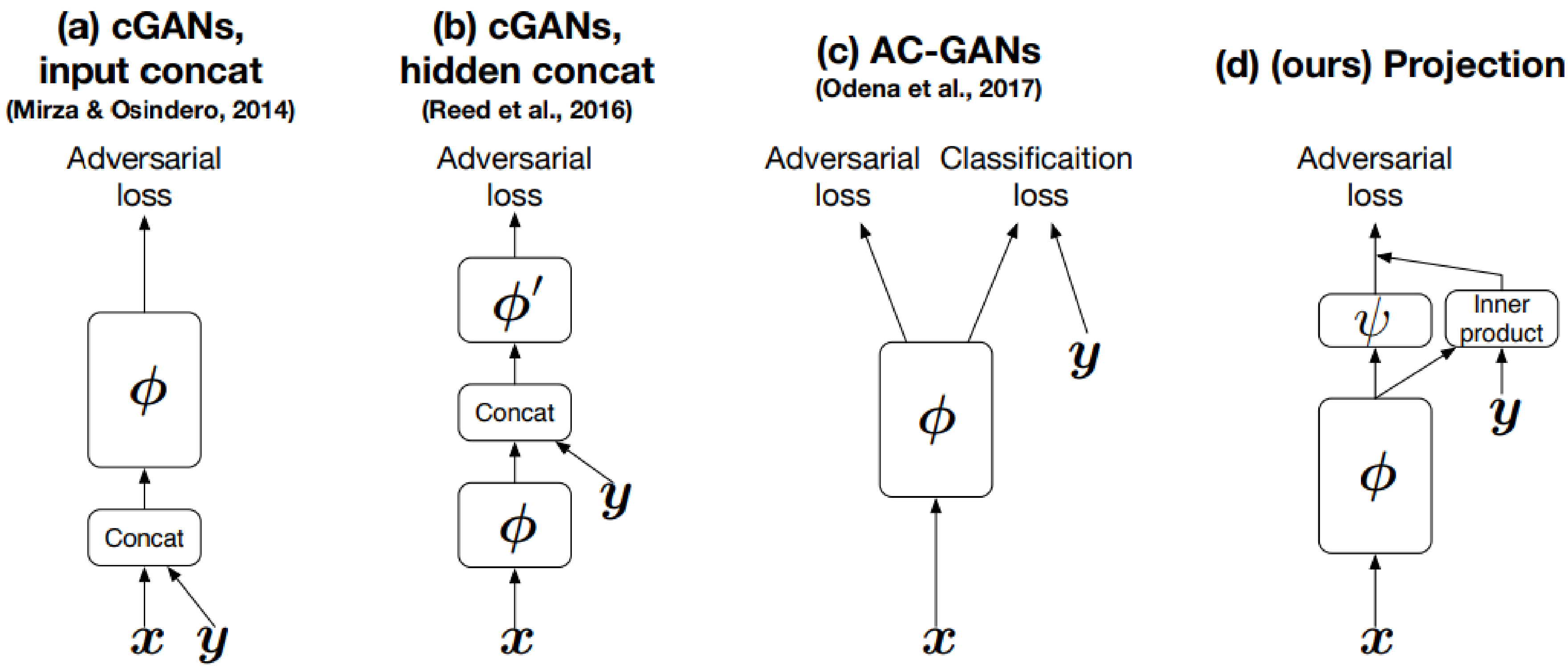


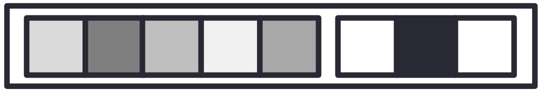
Figure 1: Discriminator models for conditional GANs

Concatenation

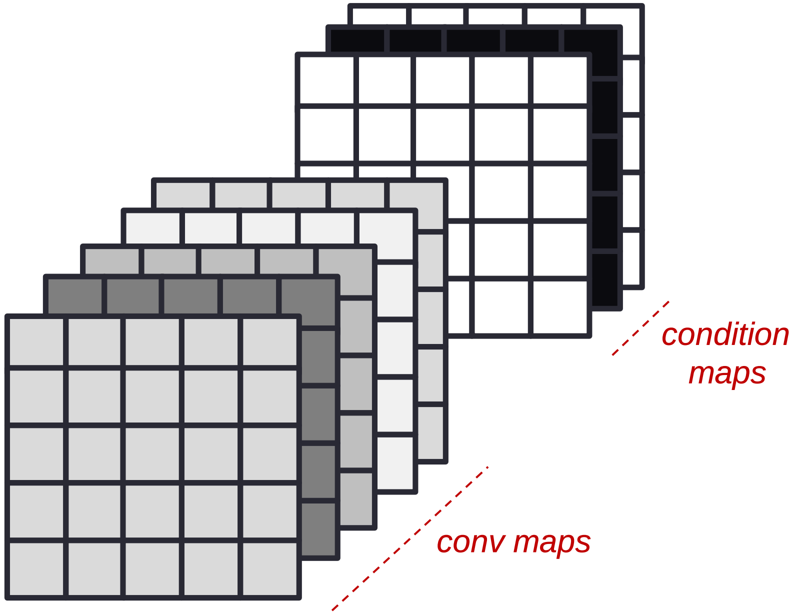
- The most common method, naive but practical.

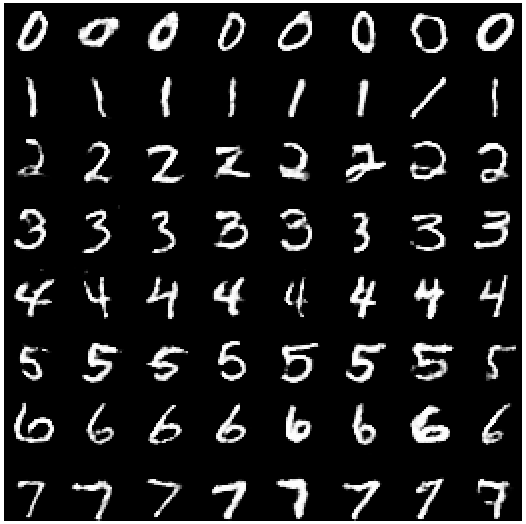
one-hot label

vector concatenation *(for linear layer, input z)*



Feature map concatenation *(for convolution layer)*

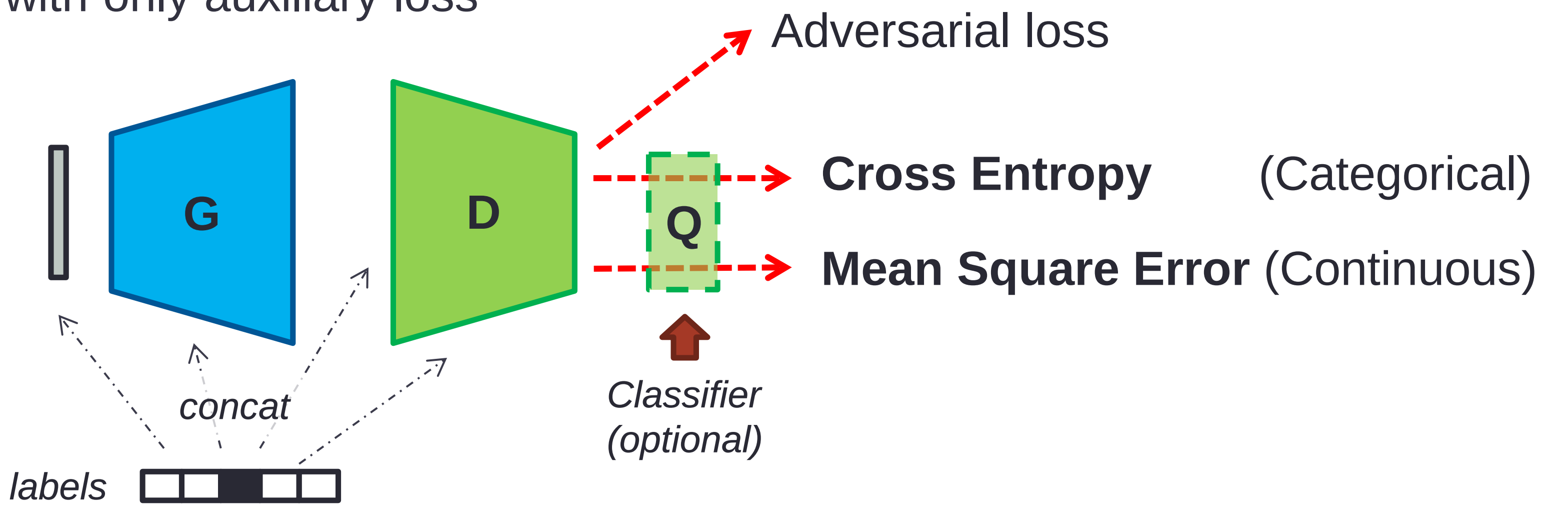




(SNGAN on MNIST)

Auxiliary Classifier

- Or with only auxiliary loss

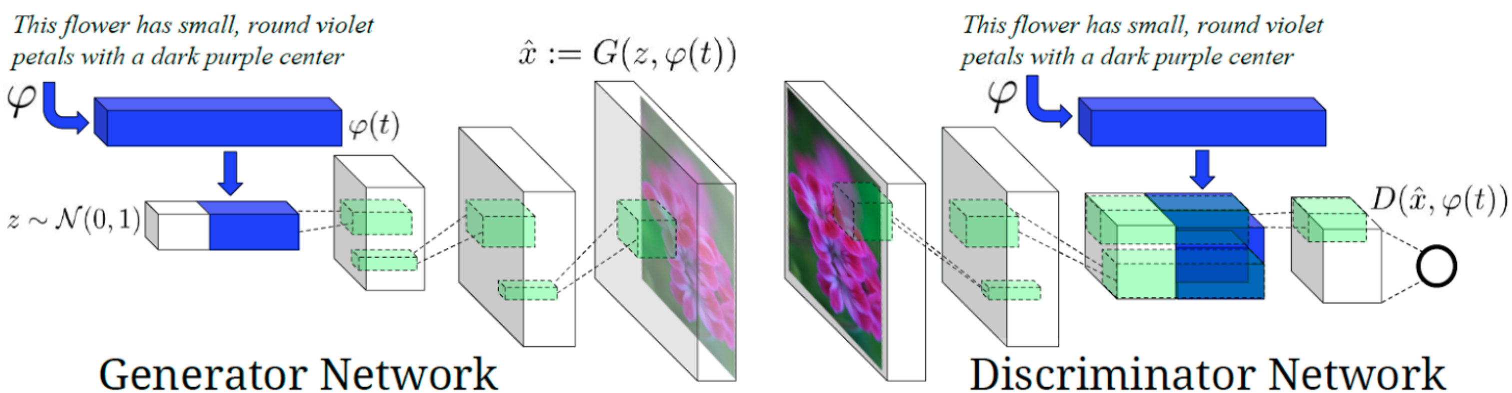


- Where should I concatenate labels? Every layer in **G** & **D**?
    - It depends. Layers which is responsible for higher level features have priorities.
- Conditional generation is still a challenging task.
- SNGAN + projection discriminator achieve promising results.

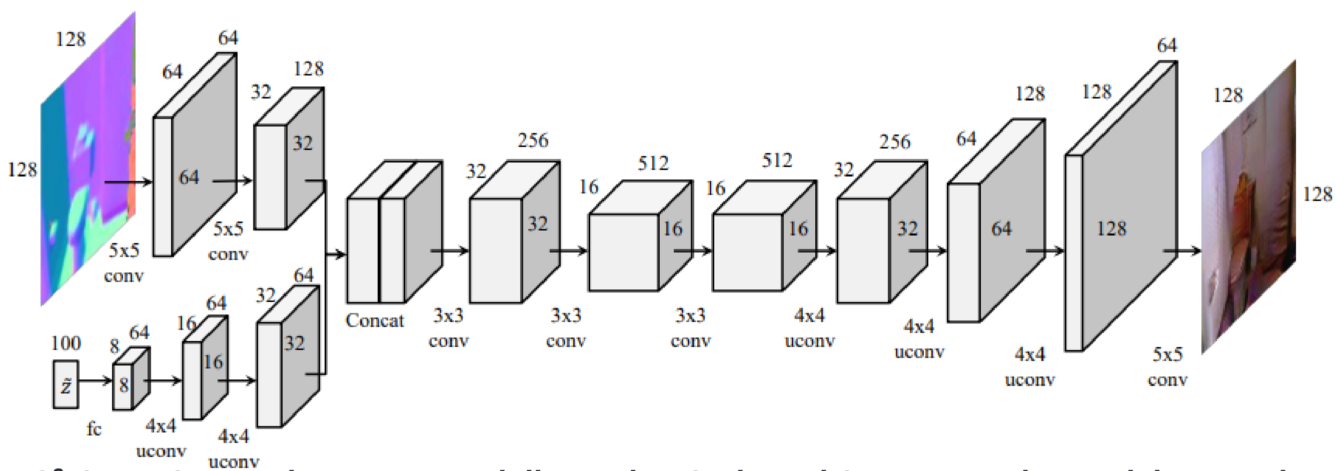
**** cods:*** *https://github.com/pfnet-research/sngan_projection*

Encoder

- For more complex conditional vectors
- Motivation: guide the generation process



Generative Adversarial Text to Image Synthesis

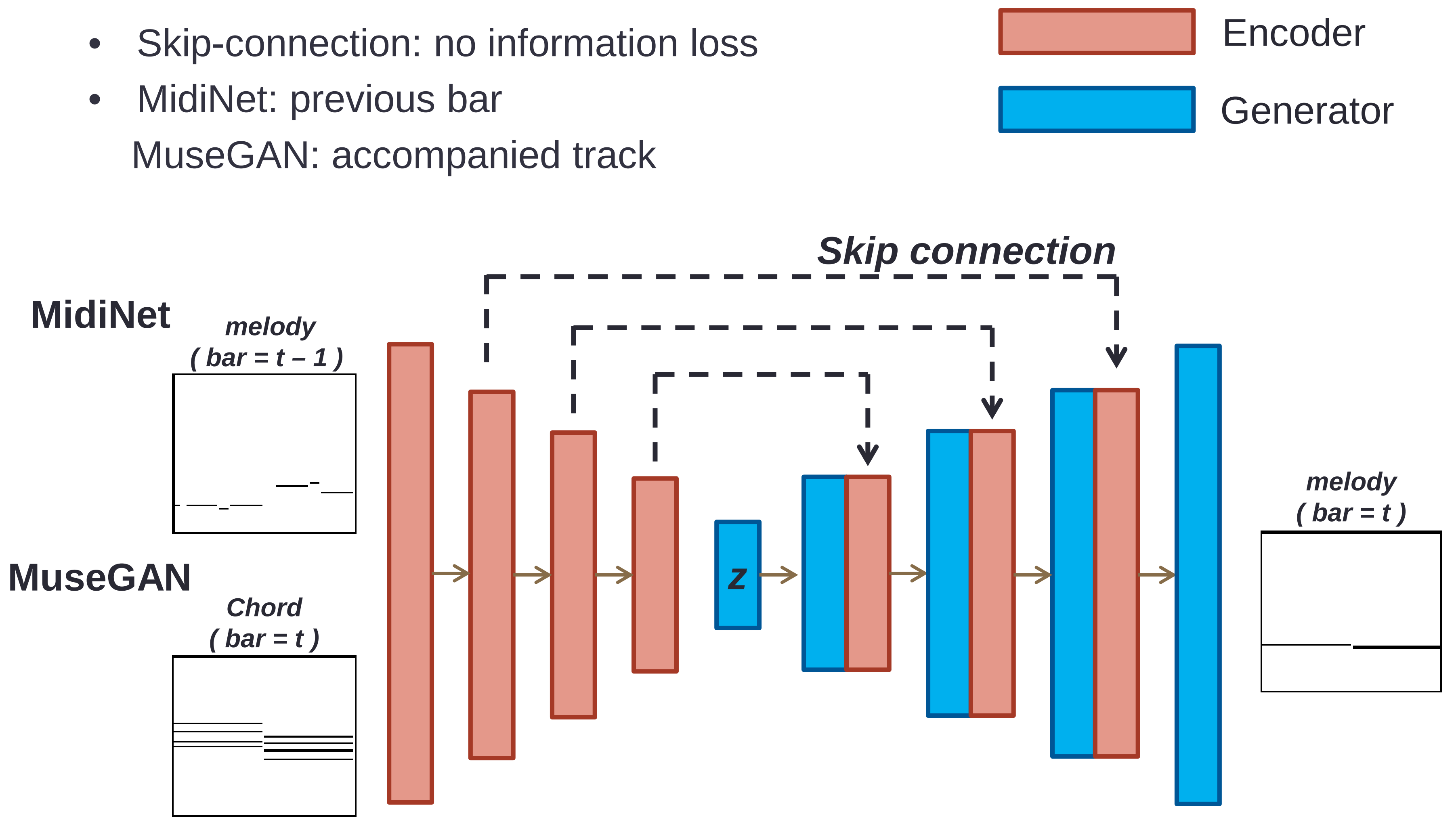


S^2 GAN: Generative Image Modeling using Style and Structure Adversarial Networks

U-Net Encoder

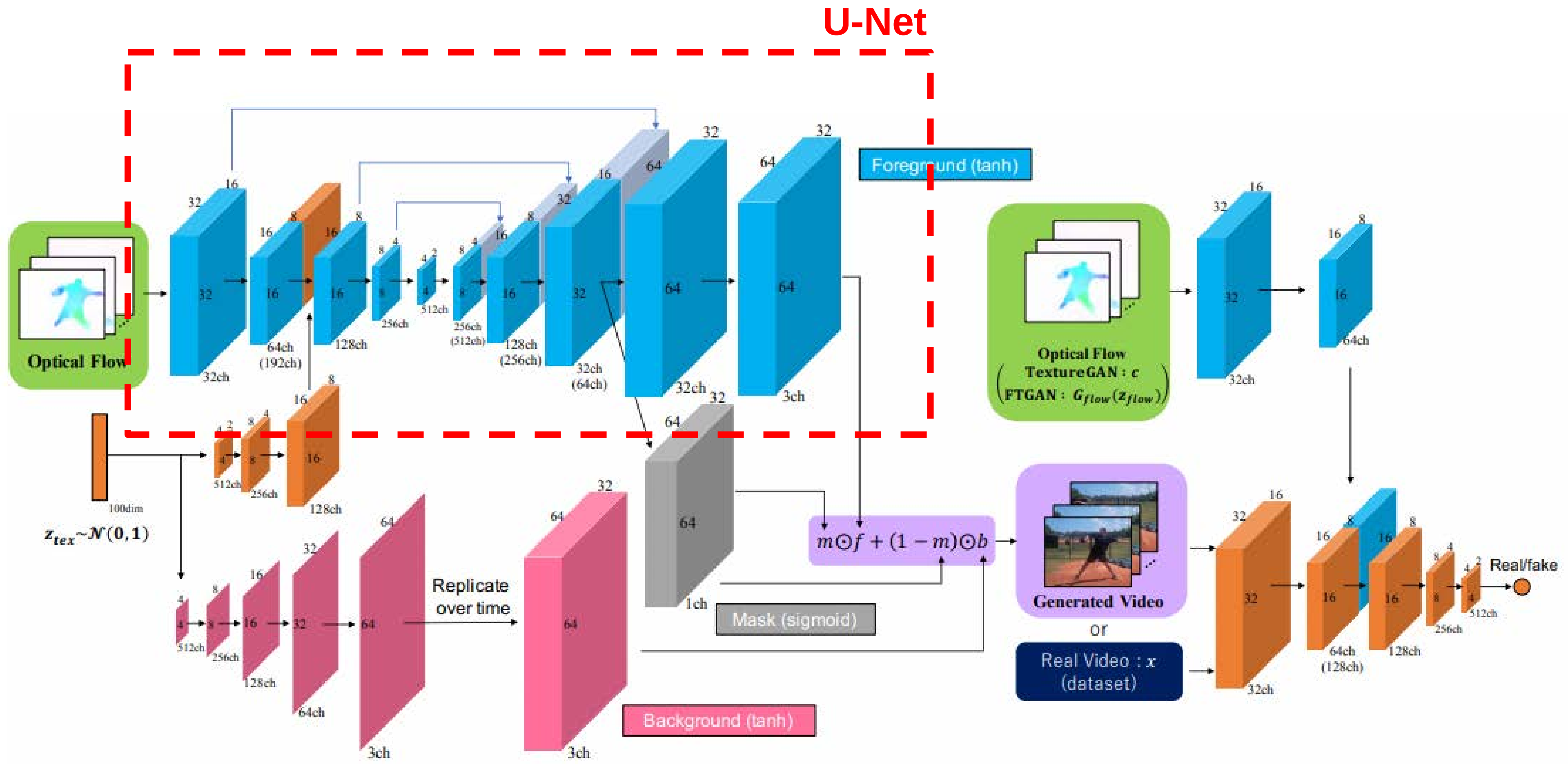
- Skip-connection: no information loss
- MidiNet: previous bar

  MuseGAN: accompanied track



U-Net Encoder

- FTGAN (video generation)
- Use conditional vector to guide the generation



** Optical flow: a feature about motion*

Tag

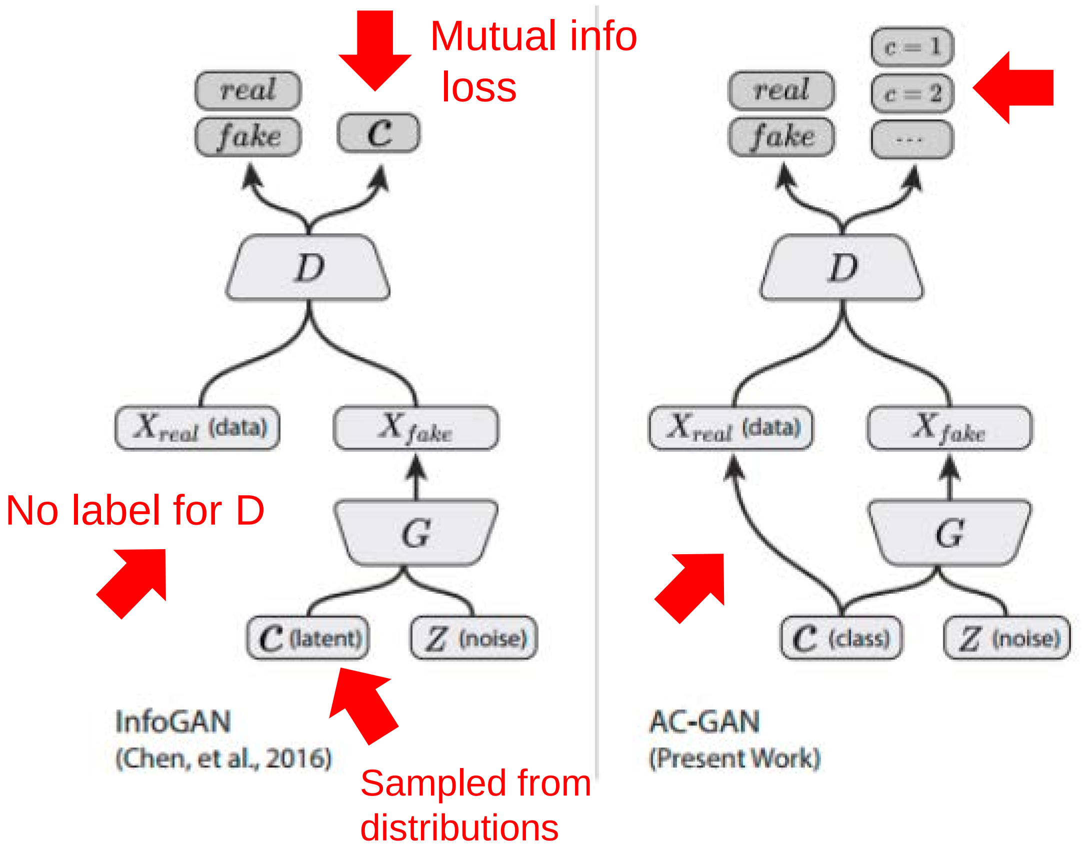
- Supervised GAN training
- **Conditional generation**
  - Given a class, generate images according to that
- **Disentanglement**
  - Try to acquire the **attribute-invariant** latent space or the controllability **with limited labels** (Ex: Gender)
  - Related Works:
    - **DR-GAN**
    - **TD-GAN**
    - **Fader Network**
    - **StarGAN** (CycleGAN)

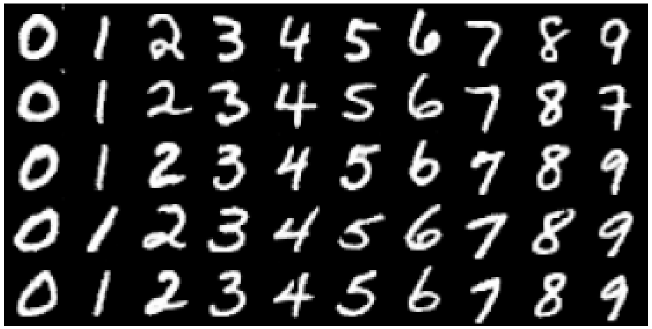
(male -> female) from Fader Network

Disentanglement

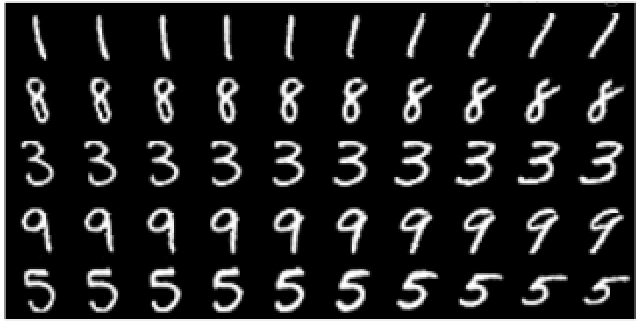
- **Unsupervised?**
  - infoGAN

    add mutual information loss to encourage latent codes learn the most obvious properties





(a) Varying c_1 on InfoGAN (Digit type)
(categorical)

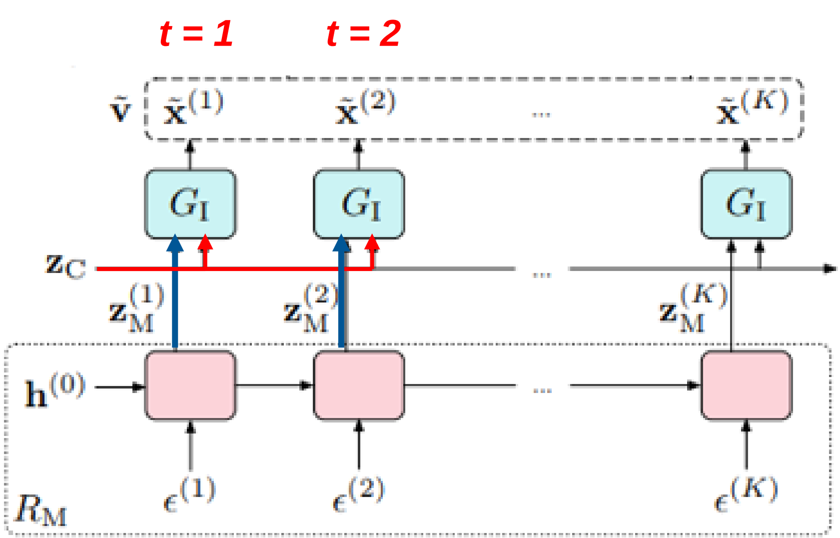


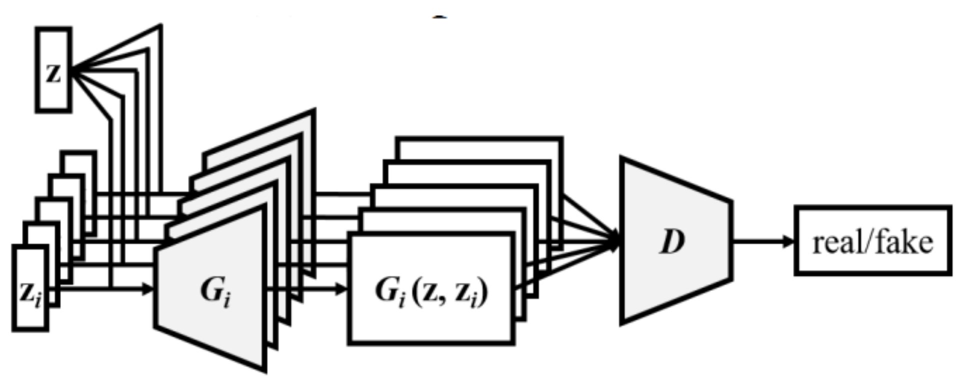
(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)
(continuous)

Disentanglement

- **Unsupervised**
  - mocoGAN & MuseGAN

    Design reasonable network architecture to encourage latent codes learn the **variant** and **invariant** properties
  - Both of works simply use 3DCNN in **D**



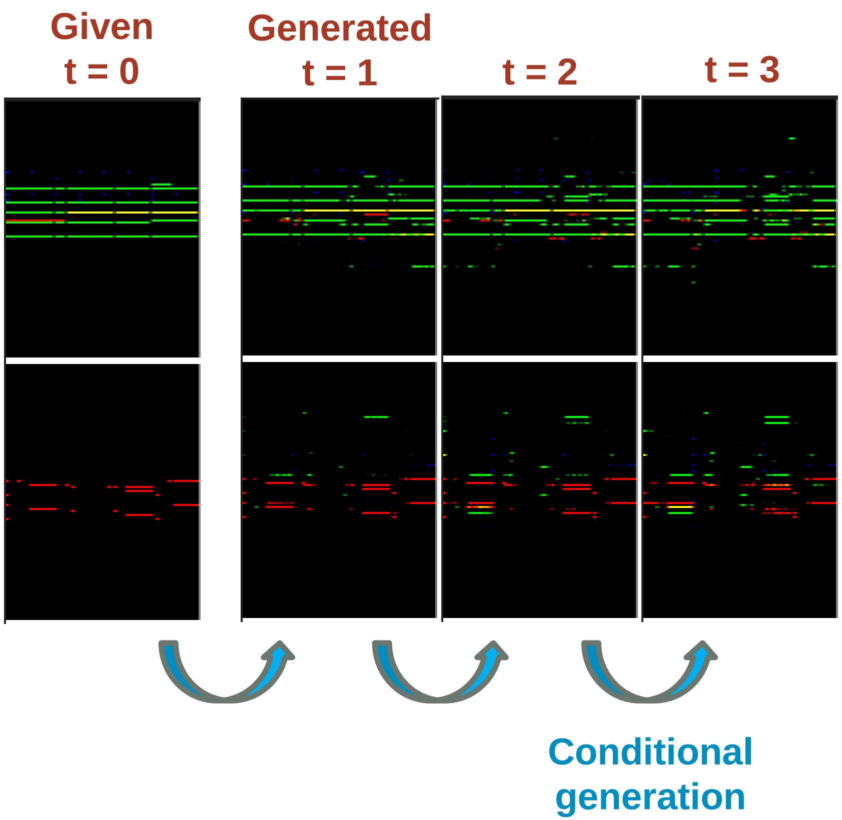


(c) Hybrid model

mocoGAN ***MuseGAN***

Discussion

- Previous bar as condition (MidiNet)
  - For melody **Good**
  - For multi-track **Failed!**



- Reasons:
  - The source and target domain are the same
  - There are too many possible results

  Finally, the network tries to copy and paste (AE, instead of U-Net)

**** MidiNet on multi-tracks***

Video GANs

Related Works

Discussion

Video Generation

- Early papers try to predict the **next frame**
  - MidiNet
- Recent works
  - VGAN
  - TGAN
  - mocoGAN
  - FTGAN
- It’s too difficult to generate a sequence of images directly.
  - Decomposition the video

    VGAN: foreground and backgrounf

    mocoGAN: content and motion

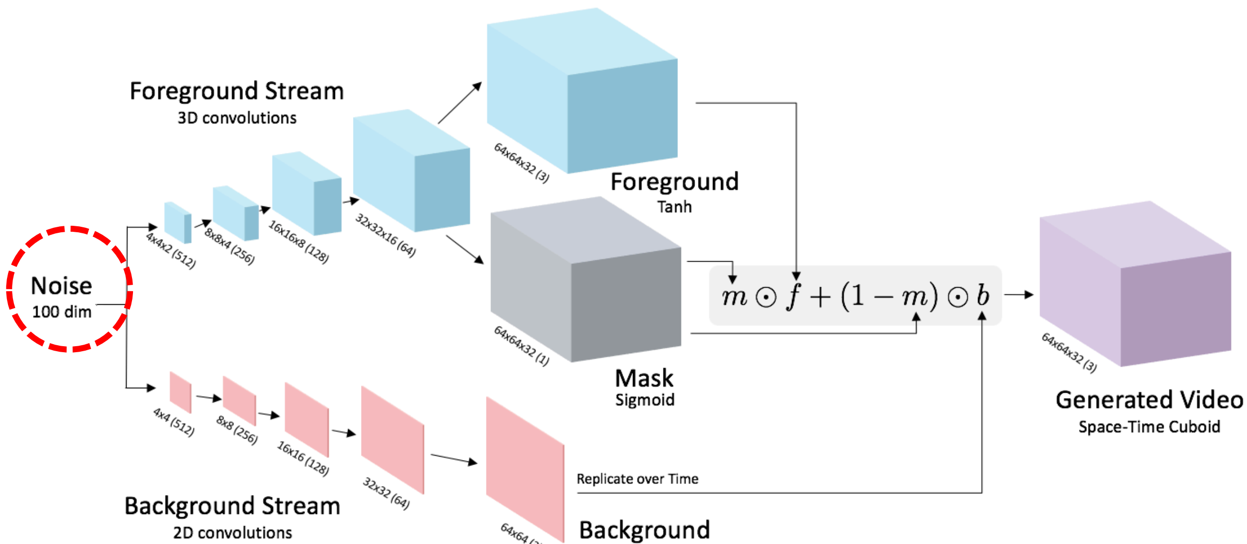
  - Using additional information to guide the generation

    FTGAN: optical flow

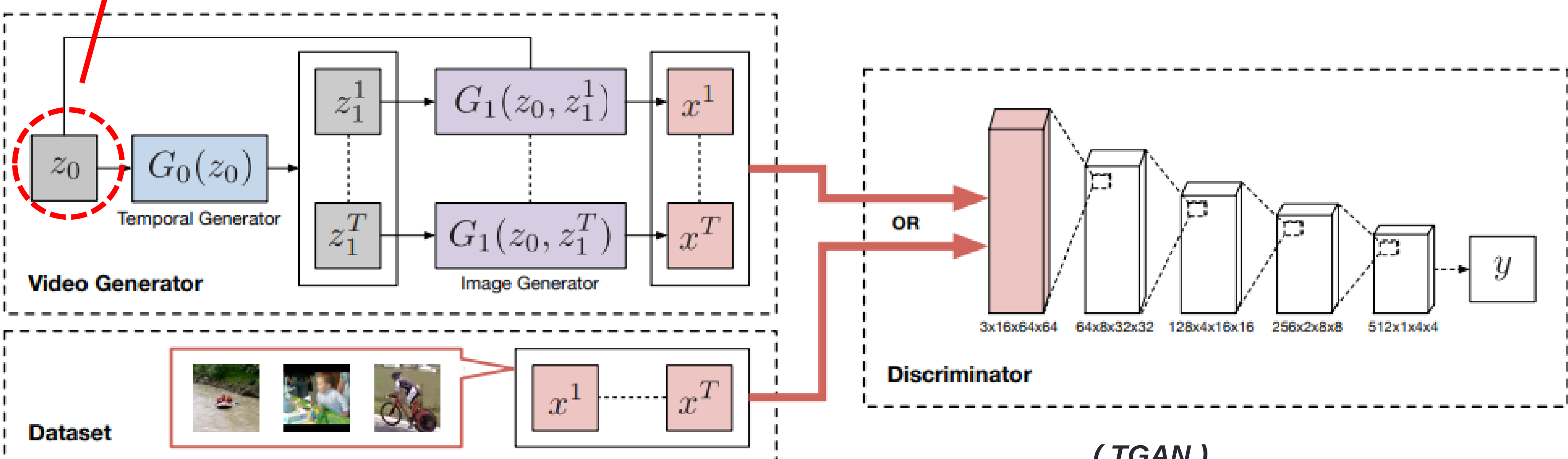
Video Generation

- Problems?

From one point at the latent space!



(VGAN)

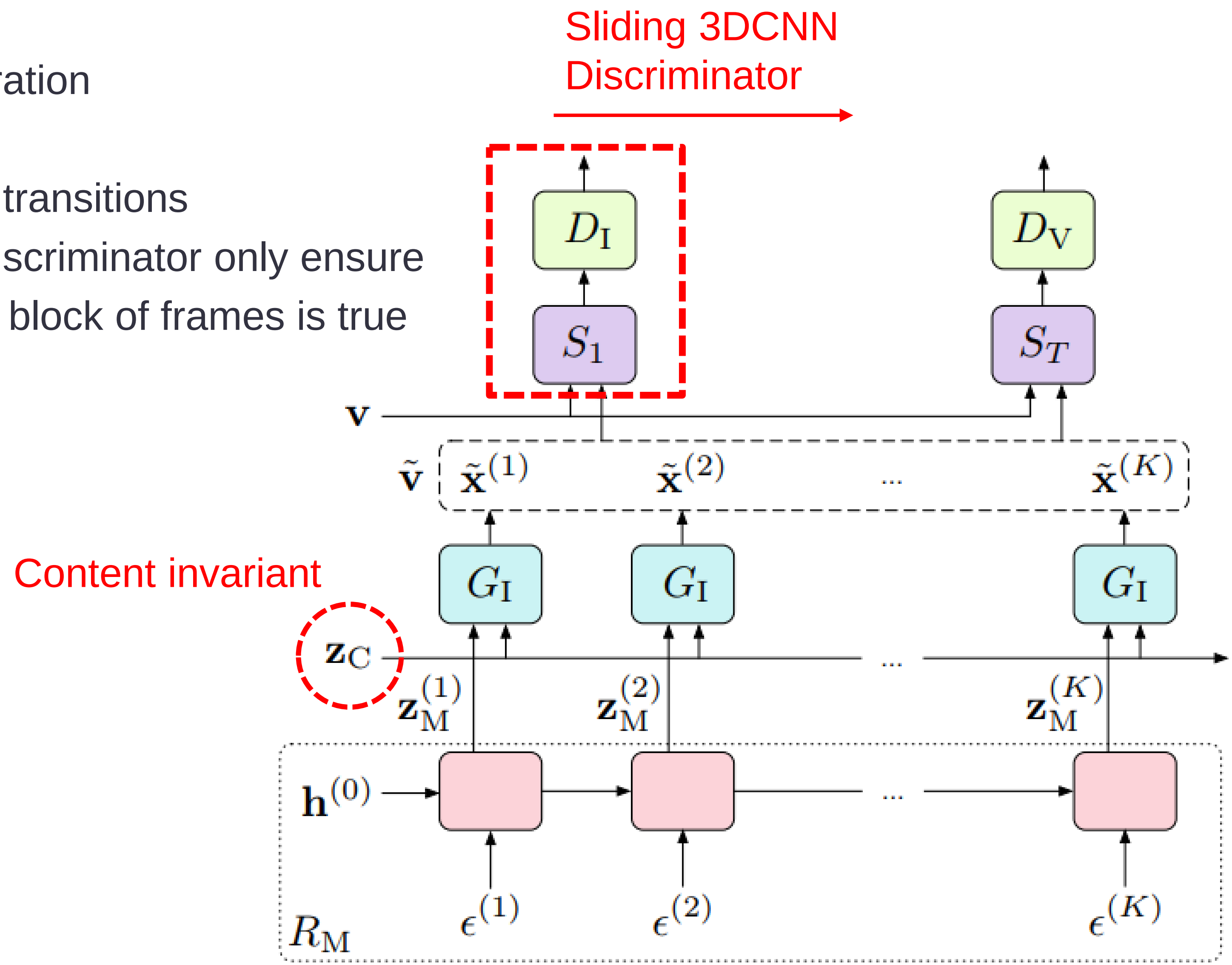


(TGAN)

Video Generation

- GIF generation

- No scene transitions
- 3DCNN discriminator only ensure every local block of frames is true



(MocoGAN)

conclusion

Video Generation

- GAN templates:
  - Chainer:
    - https://github.com/pfnet-research/chainer-gan-lib
    - https://github.com/pfnet-research/sngan_projection
  - Tensorflow:
    - https://github.com/carpedm20/DCGAN-tensorflow
    - https://github.com/wiseodd/generative-models
  - Pytorch:
    - https://github.com/eriklindernoren/PyTorch-GAN

END

Text GANs

Related Works

Discussion