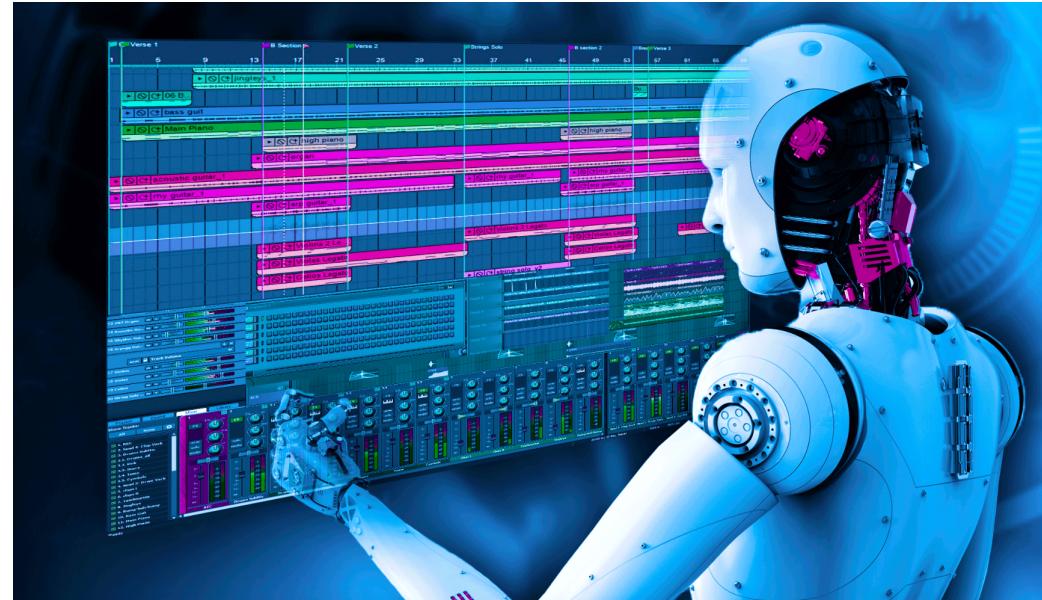


Generate ASMR audio file
using WaveGAN

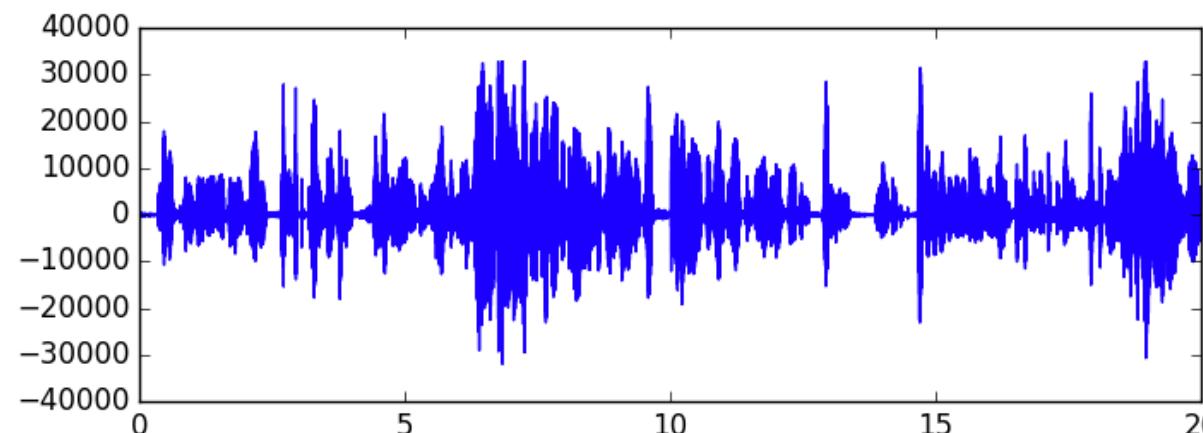
Deep Learning Models for music composition

- RNN
- CNN
- GAN
- etc



Audio File

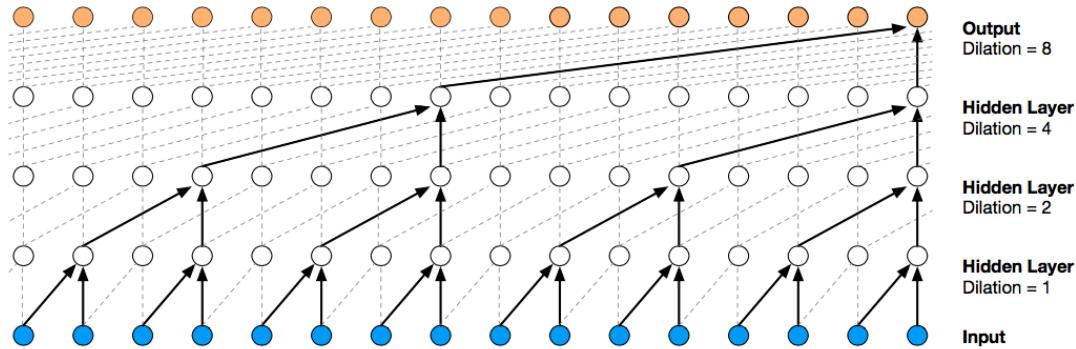
- Amplitude changes dramatically depending on time
- Consider it as time series data(stock, weather etc.)
- 44100 frames per second.
- Each frame has a range of $-32768 \sim 32767$ (2^{16})



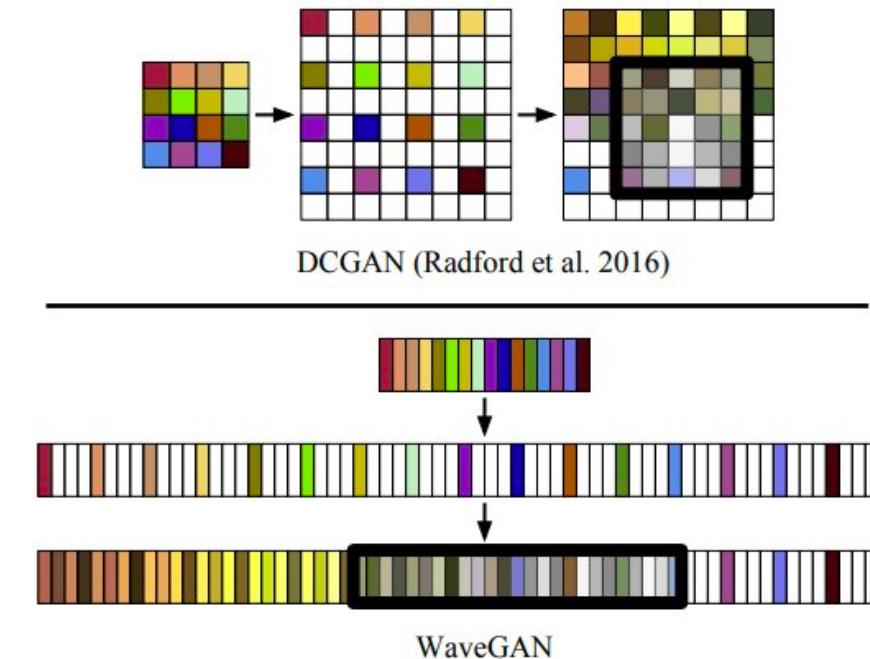
LSTM Fails

- Cannot generate properly
 - A440(pitch standard) sinusoid takes over 36 samples(waveform) to complete a **single** cycle. This suggests that filters with **larger receptive fields** are needed to process large audio.
- Too large data for recurrent model
 - Hard to remember a long cycle of musical notes for a cell.
 - Adequate data for LSTM is minute to generate audio waveform.
- Need to apply other methods

- WaveNet

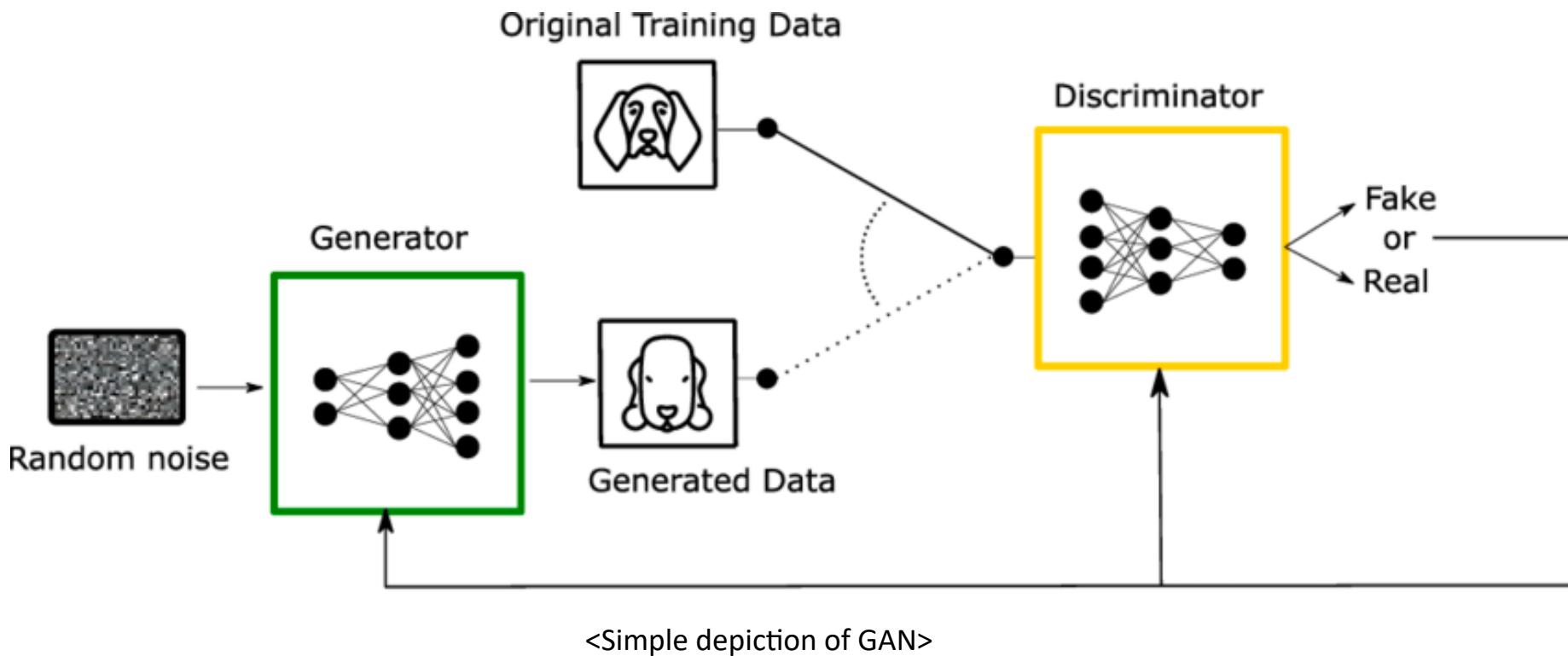


- WaveGAN



<Two examples of raising receptive fields>

GAN(Generative Adversarial Network)



GAN(Generative Adversarial Network)

GAN is unsupervised learning model.

Discriminator(D) is trained to determine if an example is real or fake, and Generator(G) is trained to fool the discriminator into thinking its output is real.

Original GAN Equation :

$$\min_G \max_D V(D, G)$$

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

This equation minimizes the Jensen-Shannon divergence, but it's difficult to train and prone to make failure cases.
Some solutions to improve model performance

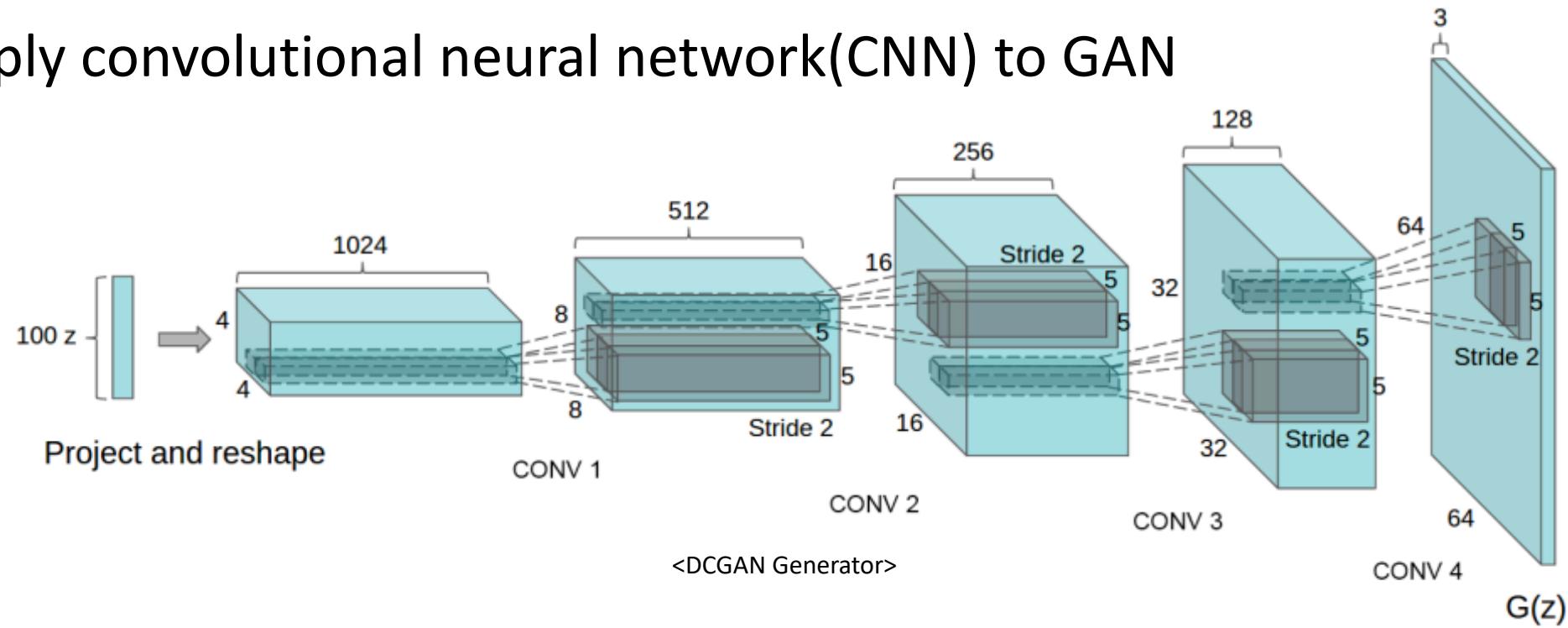
Wasserstein-1

1-Lipshitz

Gradient penalty etc..

DCGAN (Deep Convolutional Generative Adversarial Network)

- Used widely in image synthesis area.
- Apply convolutional neural network(CNN) to GAN

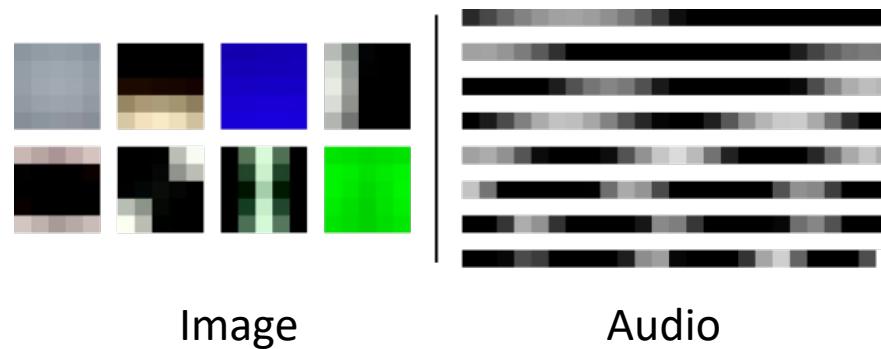


WaveGAN

- Transformation of DCGAN
- Flatten the DCGAN architecture to operate in 1 dimeson.
- Same number of parameters and numerical operations as DCGAN.

WaveGAN

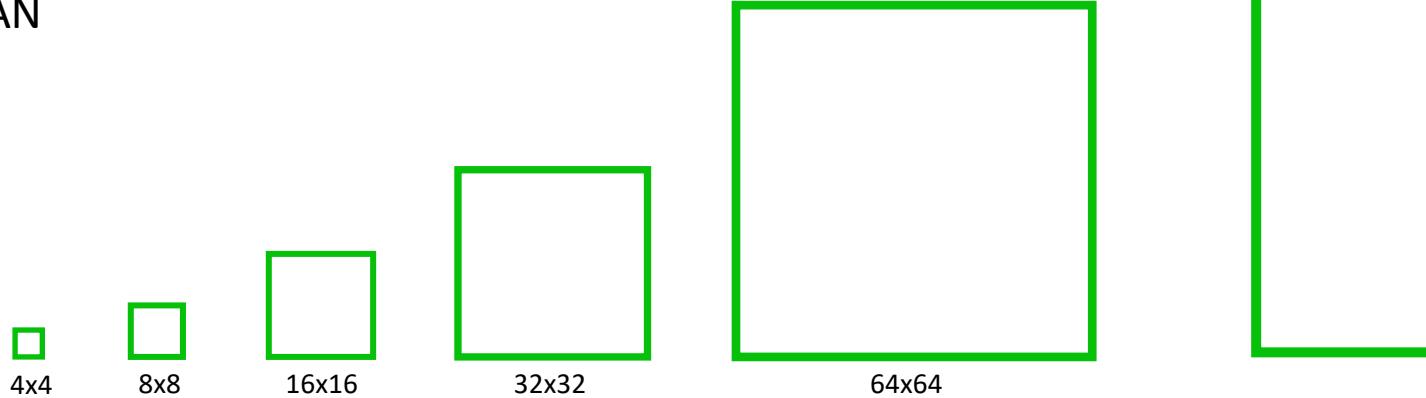
- **Periodic patterns** are unusual in natural images but a fundamental structure in audio.



- DCGAN uses small, 2D filters while WaveGAN uses longer, 1D filters and a larger upsampling factor.

Comparison

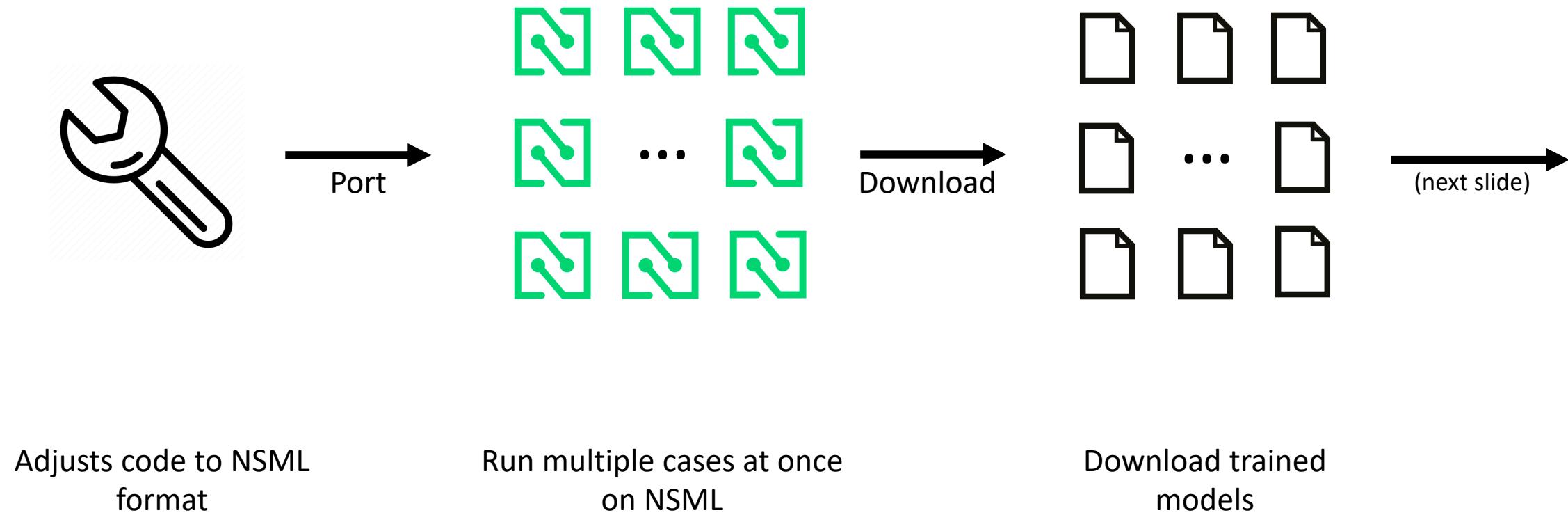
DCGAN



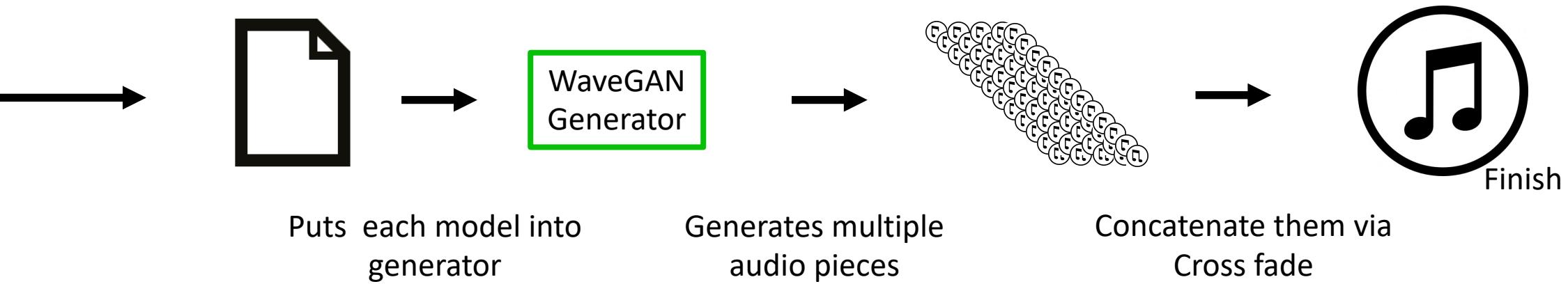
WaveGAN



Process



Process



Search Dataset (name)...

> rapihyup2
19 days ago 118.78 MB> glass
19 days ago 64.88 MB> nyc
20 days ago 297.76 MB> LaSeine
21 days ago 592.93 MB> cutting
21 days ago 152.79 MB> tropical
21 days ago 540.23 MB> barbershop
21 days ago 126.2 MB> fry
21 days ago 10.07 MB> wave
21 days ago 29.78 MB

Sessions [박지호]

Name

Search ...

Status :

Running

 Select

★ rapihyup2

:(19.77

 KR62407/rapihyup2/1

19 days exec. · Running

19 days ago

★ glass

:(18.94

 KR62407/glass/1

19 days exec. · Running

19 days ago

★ nyc

:(12.40

 KR62407/nyc/10

20 days exec. · Running

20 days ago

★ LaSeine

:(24.78

 KR62407/LaSeine/8

20 days exec. · Running

20 days ago

★ cutting

:(12.93

 KR62407/cutting/6

20 days exec. · Running

20 days ago

★ tropical

:(23.13

 KR62407/tropical/4

Favorite

Find a session

Ter...

Graph

Diff

 Select 0

17

 KR62407/glass/1

★

 KR62407/rapihyup2/1

★

 KR62407/nyc/10

★

 KR62407/LaSeine/8

★

 KR62407/cutting/6

★

 KR62407/tropical/4

★

 KR62407/fry/2

★

 KR62407/barbershop/2

★

 KR62407/wave/2

★

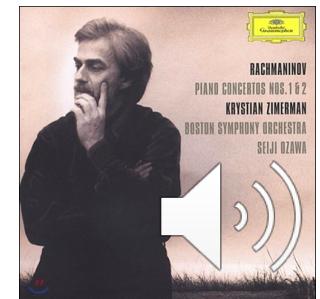
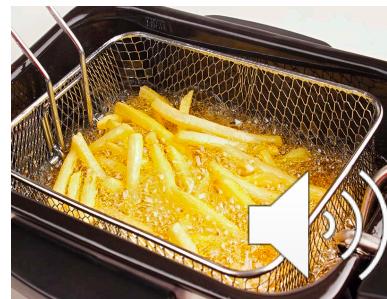
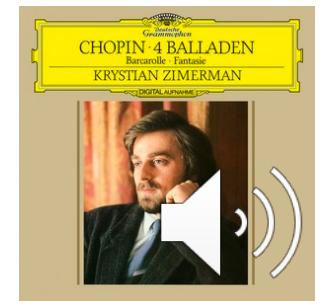
 KR62407/trafalgar/2

★

- ASMR



- Non-ASMR



Conclusion

- The outcomes(ASMR) are better than expected.
- But not flawless, somewhat incomplete.
- If enough time, reformed code and improved equipment are prepared, commercial usage of results are no longer impossible.