# Generate ASMR audio file using WaveGAN

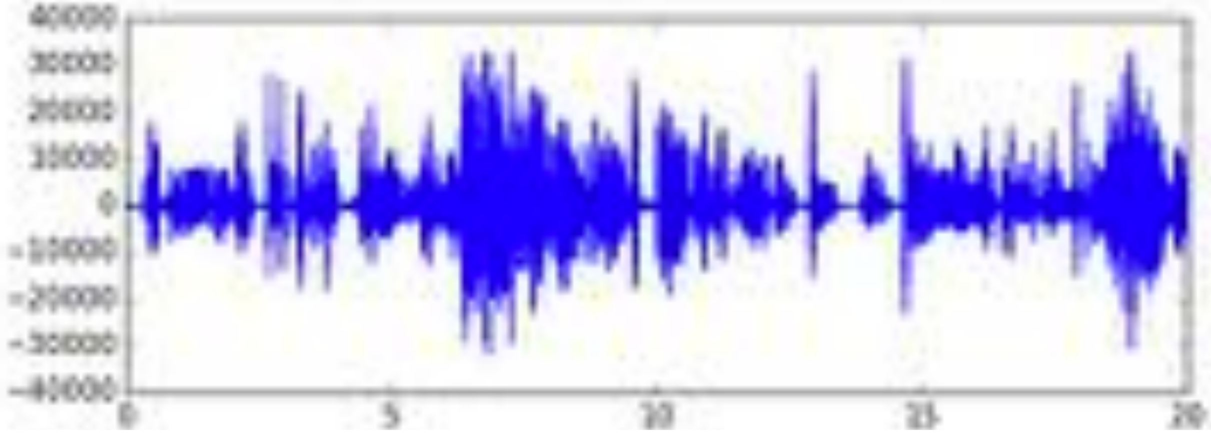
# Deep Learning Models for music composition

- 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 -32768~32767 ($2^{16}$ )



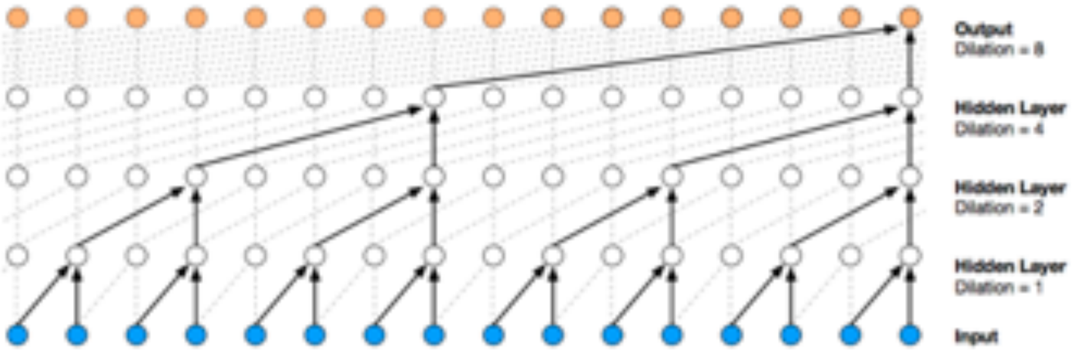
# LSTM Fails

- 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.

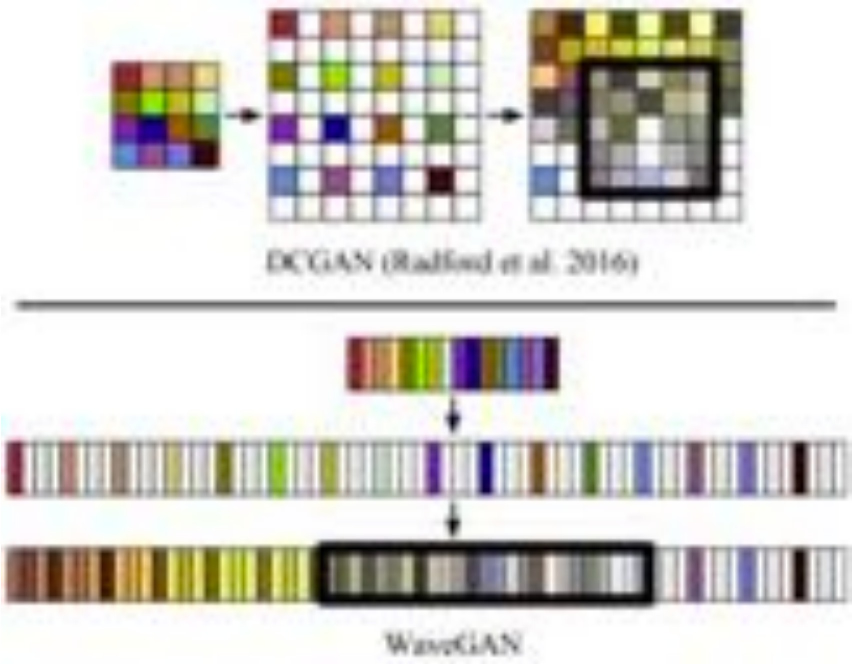
- Too large data for recurrent model
  - Hard to remember a long cycle of musical notes for a cell.
  - Adequate data for LSTM is minute to generate audio waveform.

- Need to apply other methods

- WaveNet

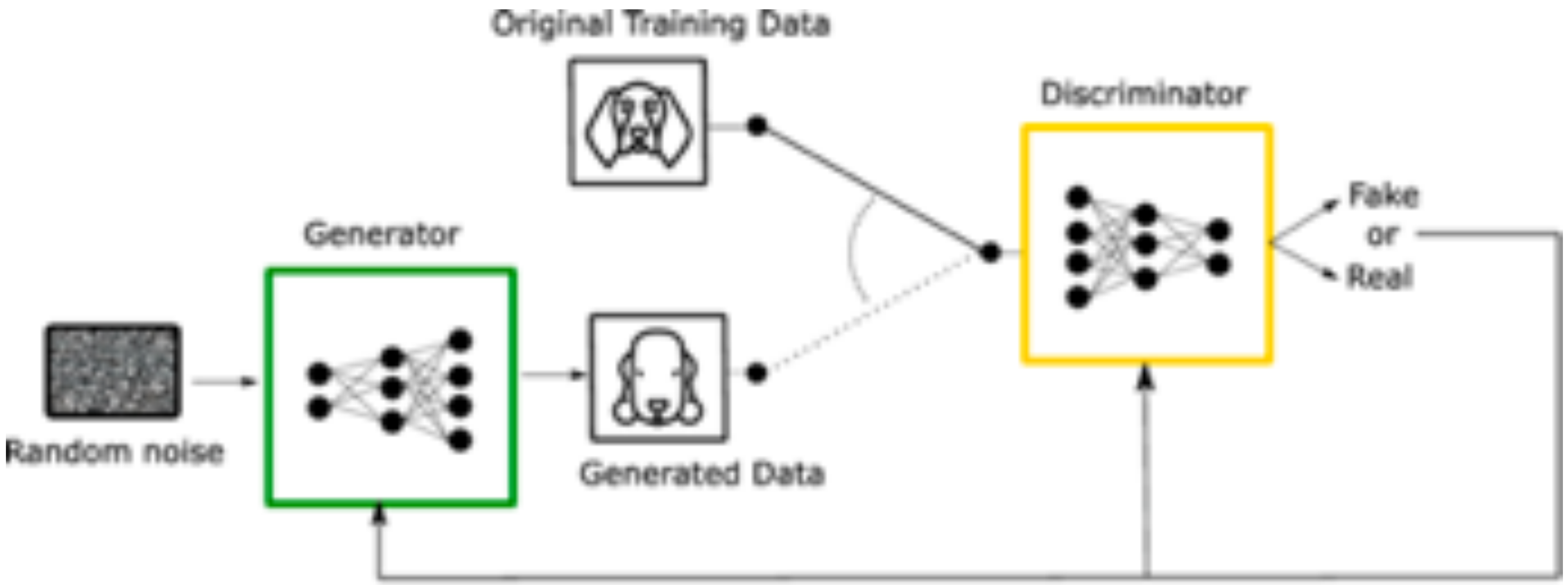


- WaveGAN



<Two examples of raising receptive fields>

# GAN(Generative Adversarial Network)



<Simple depiction of GAN>

https://www.researchgate.net/figure/Generative-Adversarial-Network-Architecture_fig3_334100947

# GAN(Generative Adversarial Network)

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

Original GAN Equation :

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

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

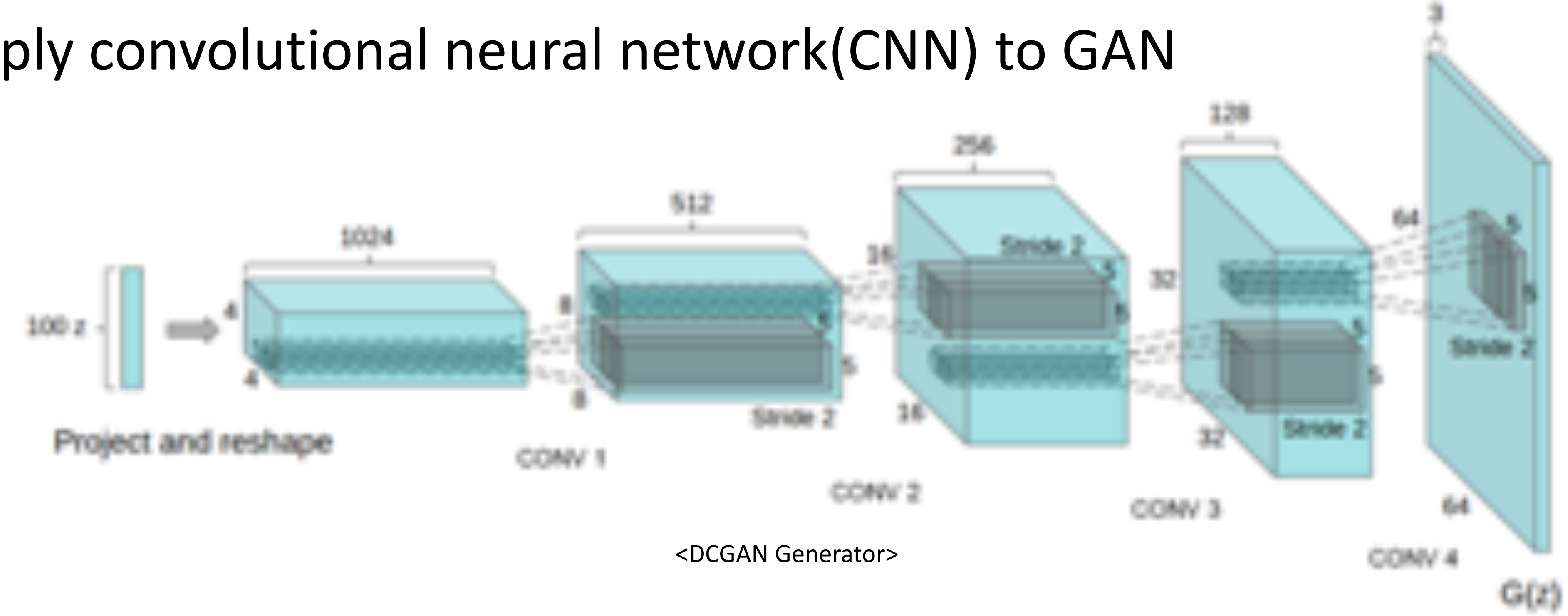
This equation minimizes the Jensen-Shannon divergence, but it’s difficult to train and prone to make failure cases.
Some solutions to improve model performance

Wasserstein-1
1-Lipshitz
Gradient penalty etc..

https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

# DCGAN (Deep Convolutional Generative Adversarial Network)

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



<DCGAN Generator>

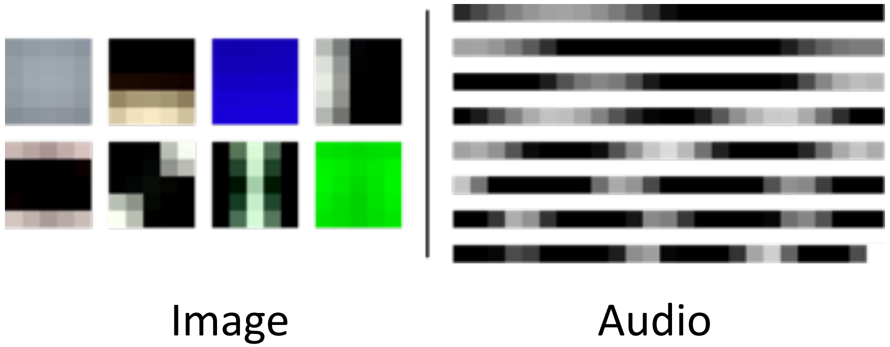
# WaveGAN

- Transformation of DCGAN
- Flatten the DCGAN architecture to operate in 1 dimeson.

- Same number of parameters and numerical operations as DCGAN.

# WaveGAN

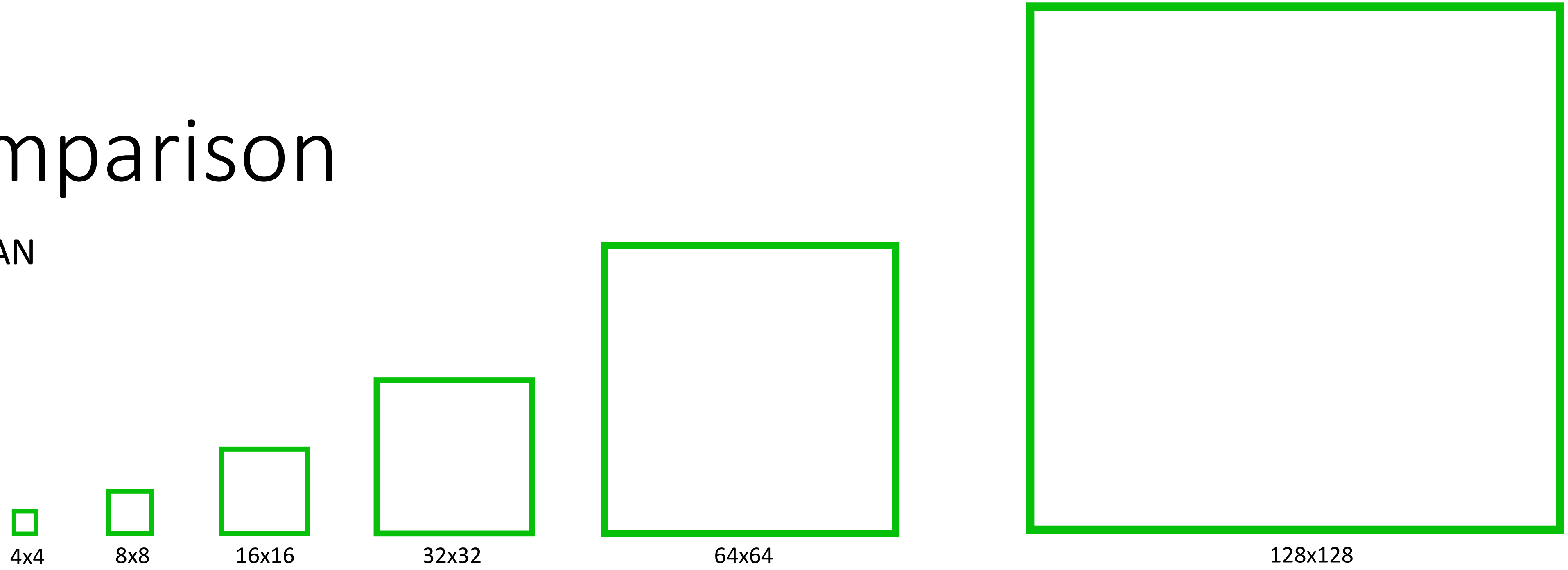
- **Periodic patterns** are unusual in natural images but a fundamental structure in audio.



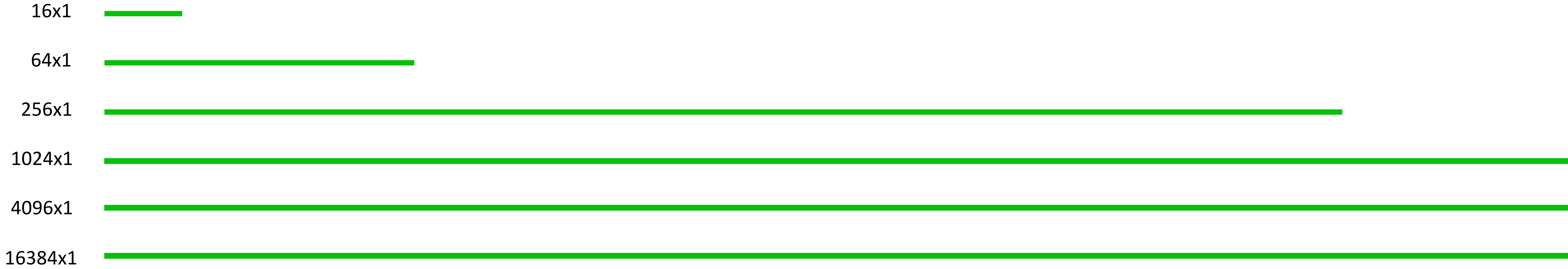
- DCGAN uses small, 2D filters while WaveGAN uses longer, 1D filters and a larger upsampling factor.

# Comparison

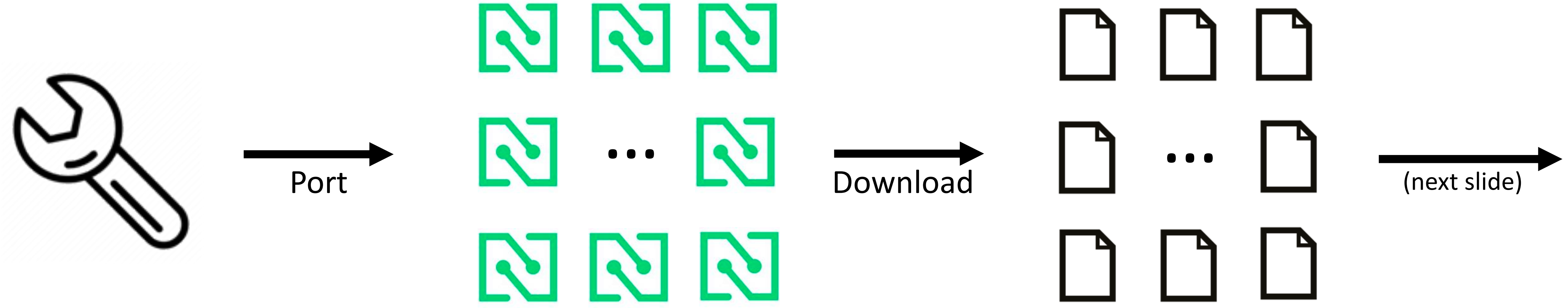
## DCGAN



## WaveGAN



# Process



Port

Download

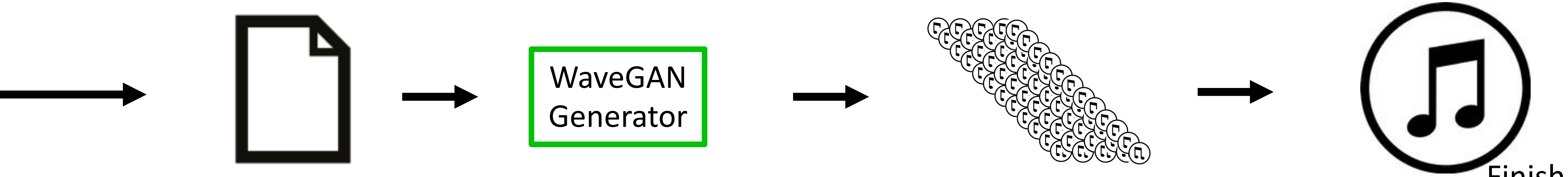
(next slide)

Adjusts code to NSML format

Run multiple cases at once on NSML

Download trained models

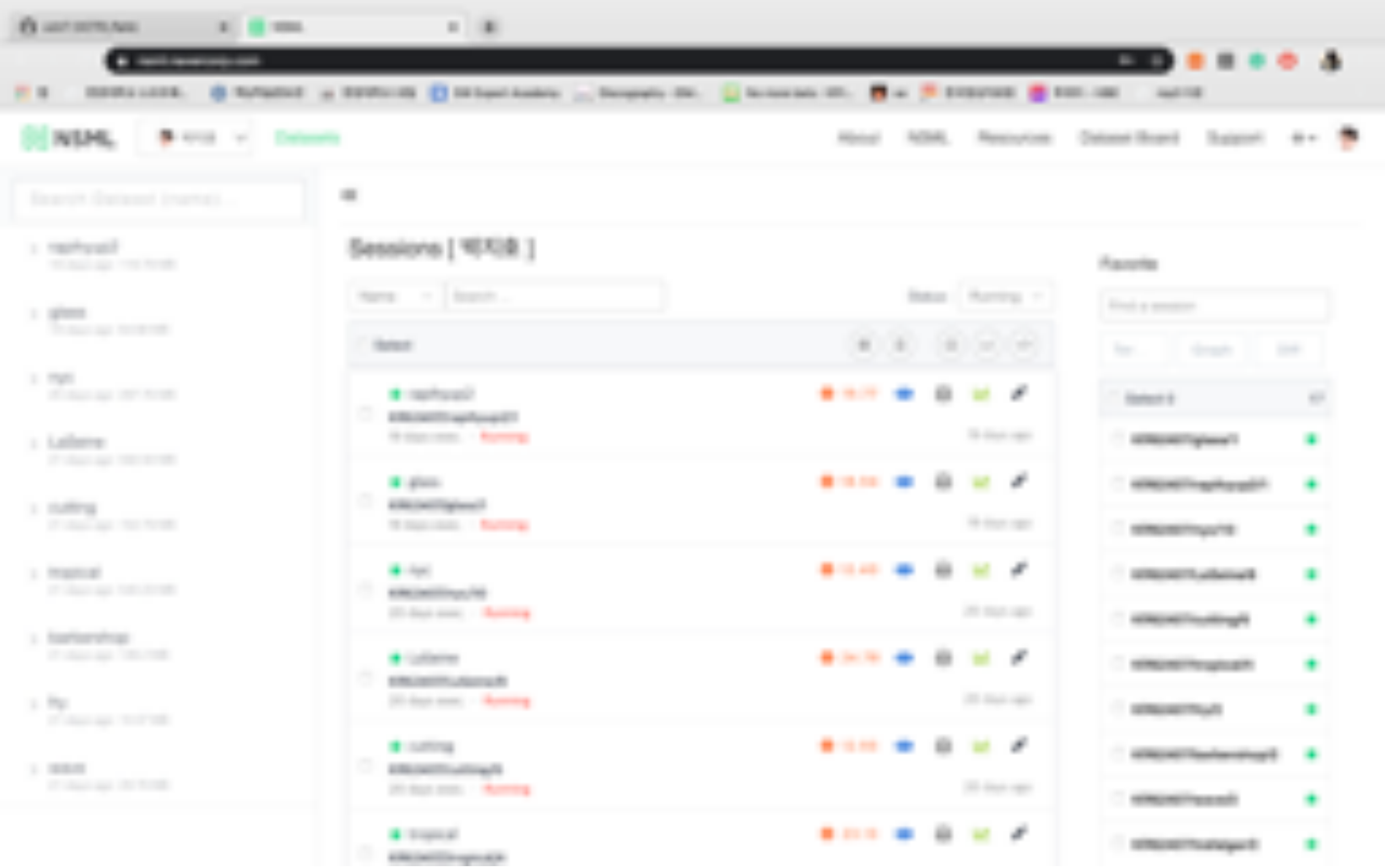
# Process



Puts each model into generator

Generates multiple audio pieces

Concatenate them via Cross fade

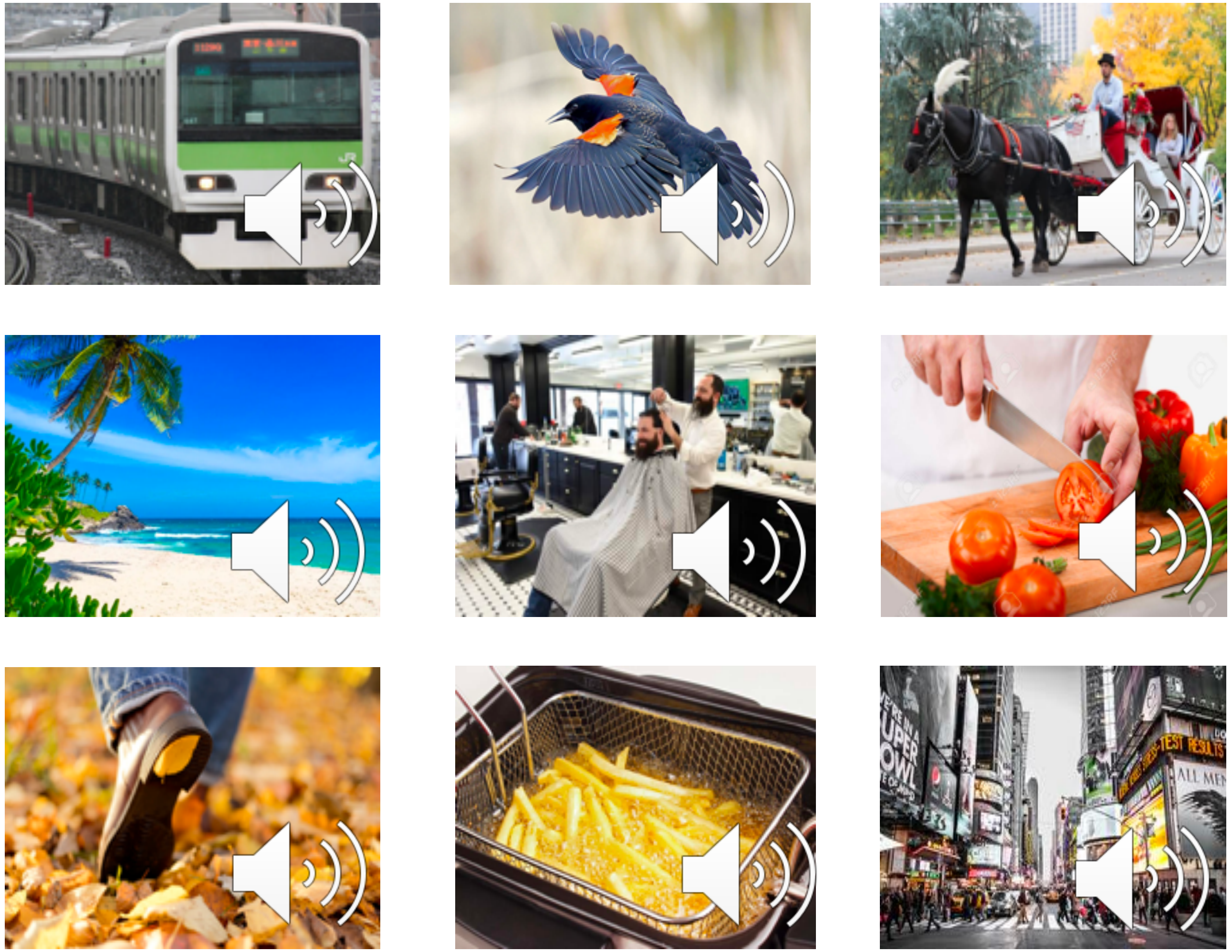


Sessions [ [illegible] ]

- rephrase3
- glass
- nyc
- LaSerne
- cutting
- tropical
- barbershop
- fly
- seaus

| Session | Status |
|---|---|
| rephrase3 | Running |
| glass | Running |
| nyc | Running |
| LaSerne | Running |
| cutting | Running |
| tropical | |

- ASMR



- Non-ASMR

https://github.com/vctr7/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.