CryoGAN: a new reconstruction paradigm for single-particle cryo-EM via deep adversarial learning

黃峻祿 for the paper H. Gupta, M. T. McCann, L. Donati, M. Unser, IEEE (2021) with its doi: http://dx.doi.org/10.1109/TCI.2021.3096491, Its bioRXiv version with doi: https://doi.org/10.1101/2020.03.20.001016 and Donati's PhD thesis (2020)

Group meeting,
ISS, AS
16/11/2021

Content

- Motivation
- Preliminary
- Methods
  - Distribution-Matching approach
  - CryoGAN
- Experiments and results
- Conclusion

Motivation

- Single particle cryo-EM becomes a leading tool for resolving the 3D structure of macromolecules. Current reconstruction techniques:
  - To estimate particle's orientations through a large set of **noisy tomographic projections with unknown orientations**. [1]
  - To marginalize orientations in **likelihood-based optimization procedures**. [2]
  - Both are computationally demanding processings.

→This paper proposes an approach resolving a 3D structure by using only the dataset of projections and CTF estimation.

- →This approach, CryoGAN [3], is a specialized GAN in which one entity tries to capture the distribution of real data, and another discriminates between generated samples and samples from real data.
- →CryoGAN is based on unsupervised deep adversarial learning. Its discriminator is trained to distinguish real from simulated projections.

[1]: P.A.Penczek *et al.*, "The ribosome at improved resolution", Ultramicroscopy, 1994
[2]: S.H. Scheres, "RELION", J. Struc. Biology, 2012
[3]: https://github.com/harshit-gupta-cor/CryoGAN

Introduction to deep learning (1)

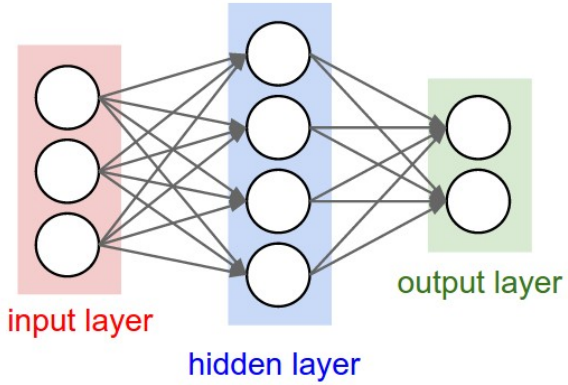
- **Deep learning is part of machine learning methods**.
  - Most of deep learning models are based on neural networks, NNs, specifically convolutional neural networks, CNNs.
  - "Deep" refers to the use of multiple layers of neurons in the network.

- **NN models are organized into layers of neurons.**
  - Within a layer, each neuron takes inputs and produces a single output which are then sent to other neurons.
  - All the inputs are summed and weighted by the weights of the connections from the inputs to the neurons.
  - The weighted sum plus a bias are passed through an activation function to produce the output.
  - There are several activation functions, such as ReLU.

Rectified linear unit (ReLU): $f(x) = \begin{cases} 0 \text{ if } x \leq 0 \\ x \text{ if } x > 0 \end{cases}$

which acts like a linear function

Visualization by [CS231n Stanford]



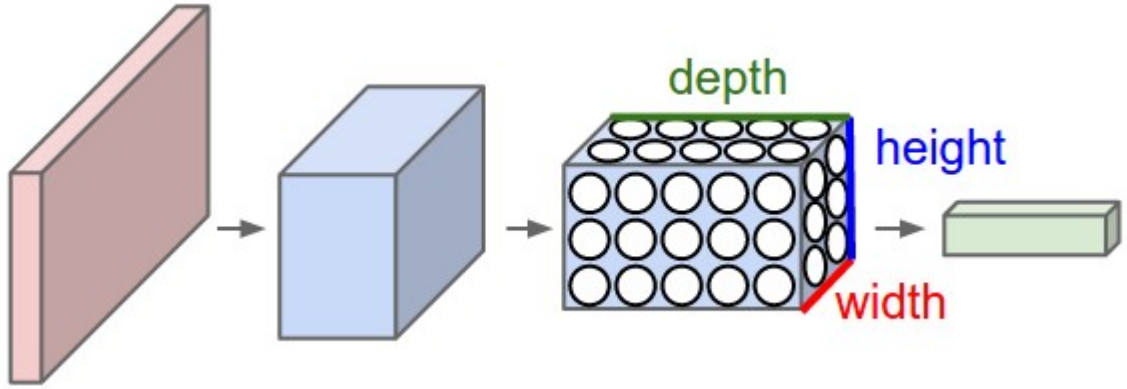
Introduction to deep learning (2)

- CNN is a sequence of layers.

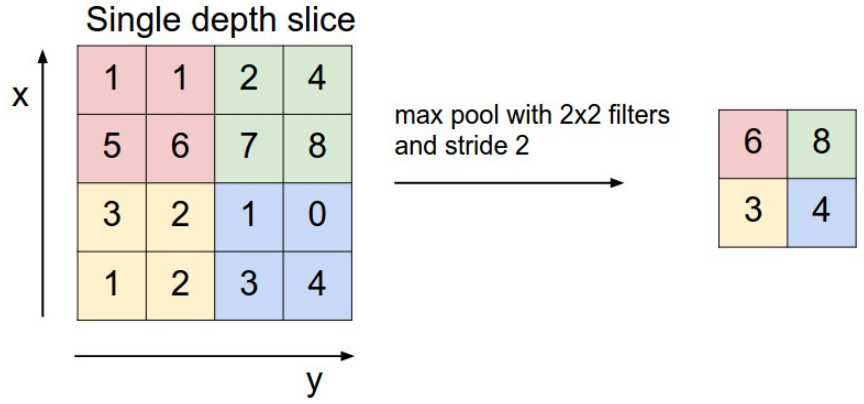
*: also known as a dense layer

  - A layer's neurons are arranged in three dimensions, width, height and depth.
  - Four popular layers, **convolutional**, **pooling**, **ReLU** and **fully-connected (FC)** layers*, build architectures.
    - **Convolutional layer** performs a convolutional operation.
    - **Pooling layer** performs a downsampling operation to reduce matrices.
    - **ReLU Layer** makes the output layer more linear.
    - **FC layer** computes the scores.

Visualization by [CS231n Stanford]



Process of max pooling. Image from [CS231n Stanford]

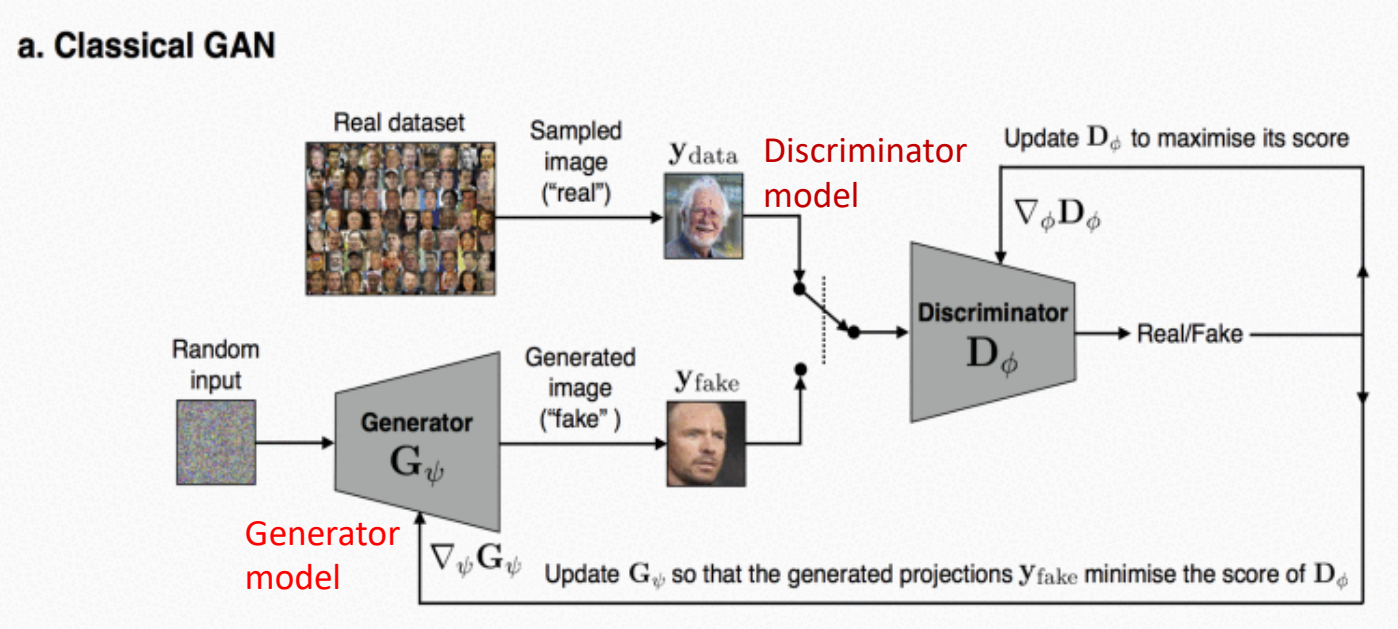


The red input layer takes an image in pixel, so its width and height are the dimensions of the image, and the depth is 3 of red, green, and blue channels.

Introduction to GAN

- **A generative adversarial network, GAN, is a class of machine learning.**
  - GAN uses two neural networks each of which competing against the other.
  - Through the competition, synthetic samples can be generated and are similar to some input data.

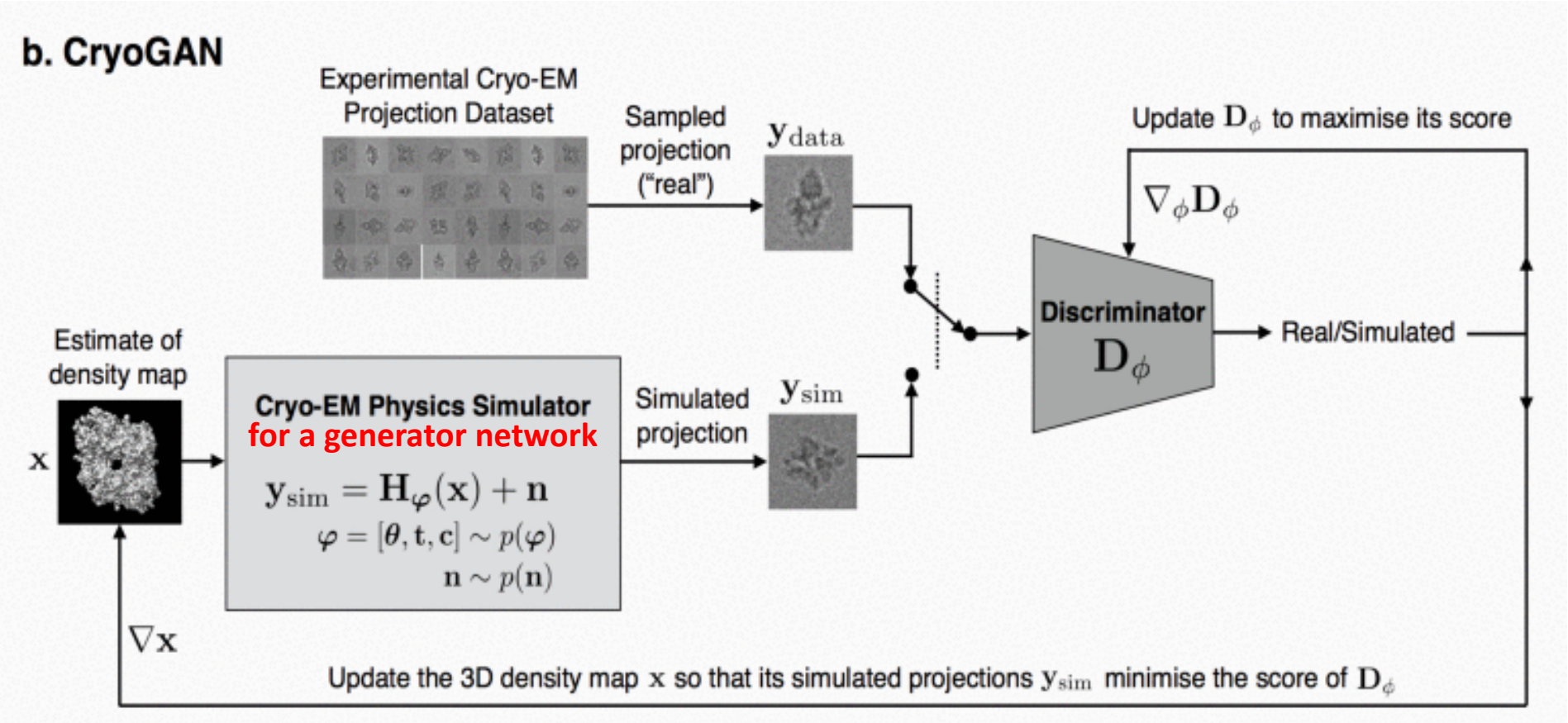
- **GAN architecture**:



- The discriminator model determines whether a given image looks like a real data image or an artificial image.
- The generator model takes random input and transforms them into images.

$\mathbf{D}_\phi$, a neural network, is called the discriminator depending on ϕ

Illustration for the CryoGAN working principle



$\mathbf{H}_{\boldsymbol{\varphi}}$: cryo-EM forward operator

$\boldsymbol{\varphi}$: imaging parameters

$\mathbf{n}$: noise

$\boldsymbol{\theta}$: Euler angles

$\mathbf{t}$: projection shifts

$\mathbf{c}$: CTF parameters

$p(\boldsymbol{\varphi})$: probability distribution of $\boldsymbol{\varphi}$

$p(\mathbf{n})$: probability distribution of $\mathbf{n}$

ϕ : parameters of discriminator

CryoGAN's learning process is unsupervised and data-driven.

Image formation model (1)

- **A cryo-EM image can be model as**

$$\boldsymbol{y}^i = \mathbf{H}_{\varphi^i}\,\mathbf{x} + \mathbf{n}^i$$

$\mathbf{y}^i$ is a 2D projection
$\mathbf{x}$ is the 3D volume
$\mathbf{H}_{\varphi}$ is the forward operator with parameters, φ^i
$\mathbf{n}$ is additive noise following distribution $p_{\mathbf{n}}$

- **The forward operator is given by**

$$\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C_c S_t P}_{\boldsymbol{\theta}}$$

$\mathbf{P}_{\boldsymbol{\theta}}$ is a projection operator at $\boldsymbol{\theta}$ (e.g. an X-ray transform)
$\mathbf{S_t}$ is shift operator
$\mathbf{C_c}$ is convolution operator

- The imaging parameters, φ, comprise the projection Euler angles, $\boldsymbol{\theta} = (\theta_1, \theta_2\theta_3)$, the projection shifts, $\mathbf{t} = (t_1, t_2)$ and the CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$.
  - φ^i are unknown and different for each projection.
  - The large number of projections, between 10^4 and 10^7 [1].
  - High level of noise in the projections

$$\varphi = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast.}})$$

[1]: A. Singer and F. J. Sigworth, “Computational methods for single-particle electron cryomicroscopy,” Annu. Rev. Biomed. Data Sci., vol. 3, 2020

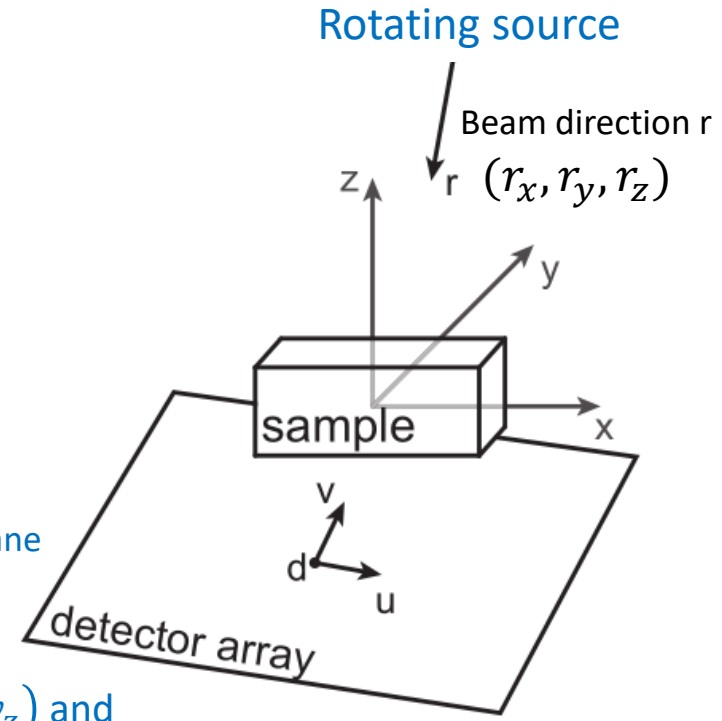
Image formation model (2)

$$\mathbf{P}_{\boldsymbol{\theta}}\{f\}(x_1, x_2) = \int_{-\infty}^{\infty} R_{\theta}\{f\}(x_1, x_2, x_3)dx_3$$

$f: \mathbb{R}^3 \to \mathbb{R}$ a density map
$R_{\theta}\{f\}(\mathbf{x}) = f\left(\mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{x}\right)$ a 3D rotation of f.
$\mathbf{R}_{\theta}$ the rotation matrix at Euler angles $\boldsymbol{\theta}$.

ASTRA toolbox [1] is used to compute the discretized version of $\mathbf{P}_{\boldsymbol{\theta}}$

- With ASTRA toolbox in cryoGAN, one can execute the projection with:
  - an ASTRA volume geometry
  - an ASTRA projection geometry (a parallel beam projection for electron tomography)
  - projection algorithm which can be executed with the above volume and projection.
  - projection geometry can be obtained



Visualization from [1]

Center of the detector plane (d_x, d_y, d_z)

Principal axes of the detector plane, described by (v_x, v_y, v_z) and (u_x, u_y, u_z)→ length of v and u correspond to the size of a pixel. Direction of v and u determine the 3D orientation of the detector

[1]: W. van Aarle *et al.*, ASTRA (2015)

Image formation model (3)

- Shift operator:

$$S_{\mathbf{t}}\{y\}(x_1, x_2) = y(x_1 - t_1, x_2 - t_2)$$

  - It models cropped subimages of a larger micrograph and can be off-center.
  - The authors implement shift in a created pad with projection and translation arrays.

- Convolution with CTF operated by $C_{\mathbf{c}}$:
  - The authors implement convolution operator with defocus parameters (from CTFFIND4), projection and translations arrays.

Distribution-matching approach (1)

- The imaging parameter distribution, $p_{\boldsymbol{\varphi}}$, acquired from a given x depending on φ.
- A conditional distribution, $p(\mathbf{y}|\mathbf{x})$, is induced by marginalizing over the imaging parameters:
  - $p(\mathbf{y}|\mathbf{x}) = \int_{\varphi\in\Phi} p_n(\mathbf{y} - \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x})p_{\varphi}(\varphi)\mathrm{d}\varphi$

Φ a set of all imaging parameters
$\mathbf{H}_{\varphi}\mathbf{x}$ noiseless projection
p_n noise distribution

- **The goal is to find a reconstruction, $\mathbf{x}_{\mathbf{rec}}$, whose projection distributions match the observed distribution, $\boldsymbol{p}(\mathbf{y}|\mathbf{x}_{\mathbf{true}})$ of the true structure, $\mathbf{x}_{\mathbf{true}}$.**
  - Distribution matching for cryo-EM reconstruction was considered by [1].
  - This paper generalizes [1] into more realistic forward model including noise and optical effects.

[1]: V. M. Panaretos, the Annals. Of Statistics, 2009

Distribution-matching approach (2)

- *Theorem* 1: if the reconstructed and true structures have the same projection distributions. Their structures are the same except for some rotation-reflection operation.

$$\mathbb{P}(y|f_{\text{true}}) = \mathbb{P}(y|f_{\text{rec.}}) \Leftrightarrow f_{\text{rec.}} = G(f_{\text{true}})$$

G is a rotation-reflection operation

- The reconstruction task can be considered as the minimization

$$\mathbf{x}_{\text{rec.}} = \arg\min_{\mathbf{x}} D(p(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\mathbf{x}_{\text{true}}))$$

D is distance between the two distributions

Distribution-matching approach (3)

- ***D* is chosen as the Wasserstein distance**:

$$D(p_1, p_2) = \inf_{\gamma \in \prod(p_1, p_2)} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma} [\|\mathbf{y}_1 - \mathbf{y}_2\|]$$

[1] demonstrates that the Wasserstein distance is more suitable to minimization than other distances (?)

$\prod(p_1, p_2)$ is the set of all the joint distributions γ $\gamma(\mathbf{y}_1, \mathbf{y}_2)$ whose marginals are p_1 and p_2

which is then applied to the minimization:

$$\mathbf{x}_{\text{rec.}} = \arg \min_{x} \inf_{\gamma \in \prod(p_{\mathbf{x}}, p_{\text{data}})} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma} [\|\mathbf{y}_1 - \mathbf{y}_2\|]$$

- **This minimization problem can be expressed its dual form [2]:**

$$\mathbf{x}_{\text{rec.}} = \arg \min_{\mathbf{x}} \max_{f: \|s\|_L \leq 1} \left\{ \mathbb{E}_{\mathbf{y} \sim p_{\text{true}}} [s(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}} [s(\mathbf{y})] \right\} \text{(duality)}$$

$$p_{\mathbf{x}} = p(\mathbf{y}|\mathbf{x})$$

$\|s\|_L$ the Lipschitz constant of the function $s: \mathbb{R}^M \to \mathbb{R}$

Lipschitz constant: $L = \sup \frac{|s(x_1) - s(x_2)|}{|x_1 - x_2|}$, the smallest upper bound of the slope

[1]: I. Gulrajani, et al., Improved Training of Wasserstein GANs, NIPS'17:Proc. (2017)

[2]: C. Villani, Optimal Transport: Old and New, Springer, Berlin, 2009

CryoGAN—connection with WGAN

- CryoGAN learns the volume whose simulated projections follow the data distribution.
  - Cryo-EM physics simulator produces projections of a volume estimate $\mathbf{x}$.
  - The simulated projections follow a distribution $p_{\mathbf{x}}$.

$$\mathbf{x}_{\text{rec.}} = \arg\min_{\mathbf{x}} \max_{\mathbf{D}_\phi : \|\mathbf{D}_\phi\|_L \leq 1} \left\{ \mathbb{E}_{\mathbf{y}\sim p_{\text{true}}}\left[\mathbf{D}_\phi(\mathbf{y})\right] - \mathbb{E}_{\mathbf{y}\sim p_{\mathbf{x}}}\left[\mathbf{D}_\phi(\mathbf{y})\right] \right\}$$

Function *s* is parameterized by $\mathbf{D}_\phi$, a neural network

- The Lipschitz constraint, $\|\mathbf{D}_\phi\|_L \leq 1$, can be enforced with gradient penalty (WGAN-GP) [1].
  - If the normal gradient is one for interpolation images, $\mathbf{y}_{\text{int.}}$, then Lipschitz constraint is satisfied.
- The formulation for the reconstruction problem becomes

$$\mathbf{x}_{\text{rec.}} = \arg\min_{\mathbf{x}} \max_{\mathbf{D}_\phi} \left\{ \mathbb{E}_{\mathbf{y}\sim p_{\text{true}}}\left[\mathbf{D}_\phi(\mathbf{y})\right] - \mathbb{E}_{\mathbf{y}\sim p_{\mathbf{x}}}\left[\mathbf{D}_\phi(\mathbf{y})\right] - \lambda \cdot \mathbb{E}_{\mathbf{y}\sim p_{\text{int.}}}\left[\left(\|\nabla_{\mathbf{y}} \mathbf{D}_\phi(\mathbf{y}_{\text{int.}})\| - 1 \right)^2 \right] \right\} \quad (1)$$

$p_{\text{int.}}$ the uniform distribution along the straight line between points sampled between p_{true} and $p_{\mathbf{x}}$. (int: Interpolations)

$\lambda \in \mathbb{R}_+$ an appropriate penalty coefficient.

[1]: I. Gulrajani, et al., Improved Training of Wasserstein GANs, NIPS'17:Proc. (2017)

CryoGAN—algorithm (1)

- By replacing the expected values with their empirical counterparts [1], the formulation is reformulated as

$$L_S\left(\mathbf{x}, \mathbf{D}_\phi\right) = \sum_{m=1}^{M} \mathbf{D}_\phi\left(\mathbf{y}_{\text{data}}^{n_m}\right) - \sum_{m=1}^{M} \mathbf{D}_\phi\left(\mathbf{y}_{\text{sim}}^{m}\right) - \lambda \sum_{m=1}^{M} \left[\left(\left\|\nabla_y \mathbf{D}_\phi\left(\mathbf{y}_{\text{int.}}^{m}\right)\right\| - 1\right)^2\right] \quad (2)$$

M number of samples in the empirical estimates
n_m random indices for measured projection images
$\mathbf{y}_{\text{sim}}^{m}$ projections from $\mathbf{x}$ generated by simulator
$\mathbf{y}_{\text{int.}}^{m} = \alpha_m \mathbf{y}_{\text{data}}^{n_m} + (1 - \alpha_m)\mathbf{y}_{\text{sim}}^{m}$, $0 < \alpha_m < 1$

The empirical estimate is identical to eq. (1) when the number of measurements becomes infinite. In real world, measurements provides very large number of particles. The empirical estimate is still reliable.

[1]: I. Gulrajani, et al., Improved Training of Wasserstein GANs, NIPS'17:Proc. (2017)

CryoGAN—algorithm (2)

- Minimize eq. (2) through stochastic gradient descent, SGD, using batches.
- Algorithm 1 alternatively updates the discriminator, $\mathbf{D}_\phi$, ($n_{\text{discr.}}$ iterations) using Adam optimizer and the volume x (one iteration) using gradients of L_S.

Algorithm 1: CryoGAN.

Parameters: number of training iterations, n_{train}; number of iterations of the discriminator per training iteration, n_{discr}; size of the batches used for SGD, M; penalty parameter, λ

1: **for** n_{train} **do**
2: **for** n_{discr} **do**
3: sample $\{\mathbf{y}^1_{\text{batch}}, \ldots, \mathbf{y}^M_{\text{batch}}\}$ from real data
4: sample $\{\mathbf{y}^1_{\text{sim}}, \ldots, \mathbf{y}^M_{\text{sim}}\}$ from current $p_{\mathbf{x}}$ ▷ (Algo. 2)
5: sample $\{\alpha_1, \ldots, \alpha_M\} \sim U[0, 1]$
6: compute $\mathbf{y}^m_{\text{int}} = \alpha_m \mathbf{y}^m_{\text{batch}} + (1 - \alpha_m)\mathbf{y}^m_{\text{sim}}$
7: update discriminator parameters ϕ using $\nabla_\phi L_S$ (14)
8: sample $\{\mathbf{y}^1_{\text{sim}}, \ldots, \mathbf{y}_{\text{sim}}{}^M\}$ from current $p_{\mathbf{x}}$ ▷ (Algo. 2)
9: update volume $\mathbf{x}$ using $\nabla_{\mathbf{x}} L_S$ (14)

See next slide

Reminder of machine learning
Batch: set of examples used in one iteration of model training

CryoGAN—Physics simulator

- The goal of the physics simulator is to sample $\mathbf{y}_{\text{sim}} \sim p_{\text{x}}(\mathbf{y})$ with three steps:
  - Sample imaging parameters, φ
  - Generate noiseless CTF-modulated and shifted projections, $\mathbf{H}_{\varphi}$, from current x.
  - Sample the noise, $\mathbf{n}$, to simulate noisy projections, $\mathbf{y} = \mathbf{H}_{\varphi}\mathbf{x} + \mathbf{n}$.

Must to have realistic noise in order to simulate projection whose distribution may closely match that of real data.

- The realistic noise is produced by extracting random background from micrographs where no particle appears.

Algorithm 2: Pseudocode for Cryo-EM Physics Simulator.

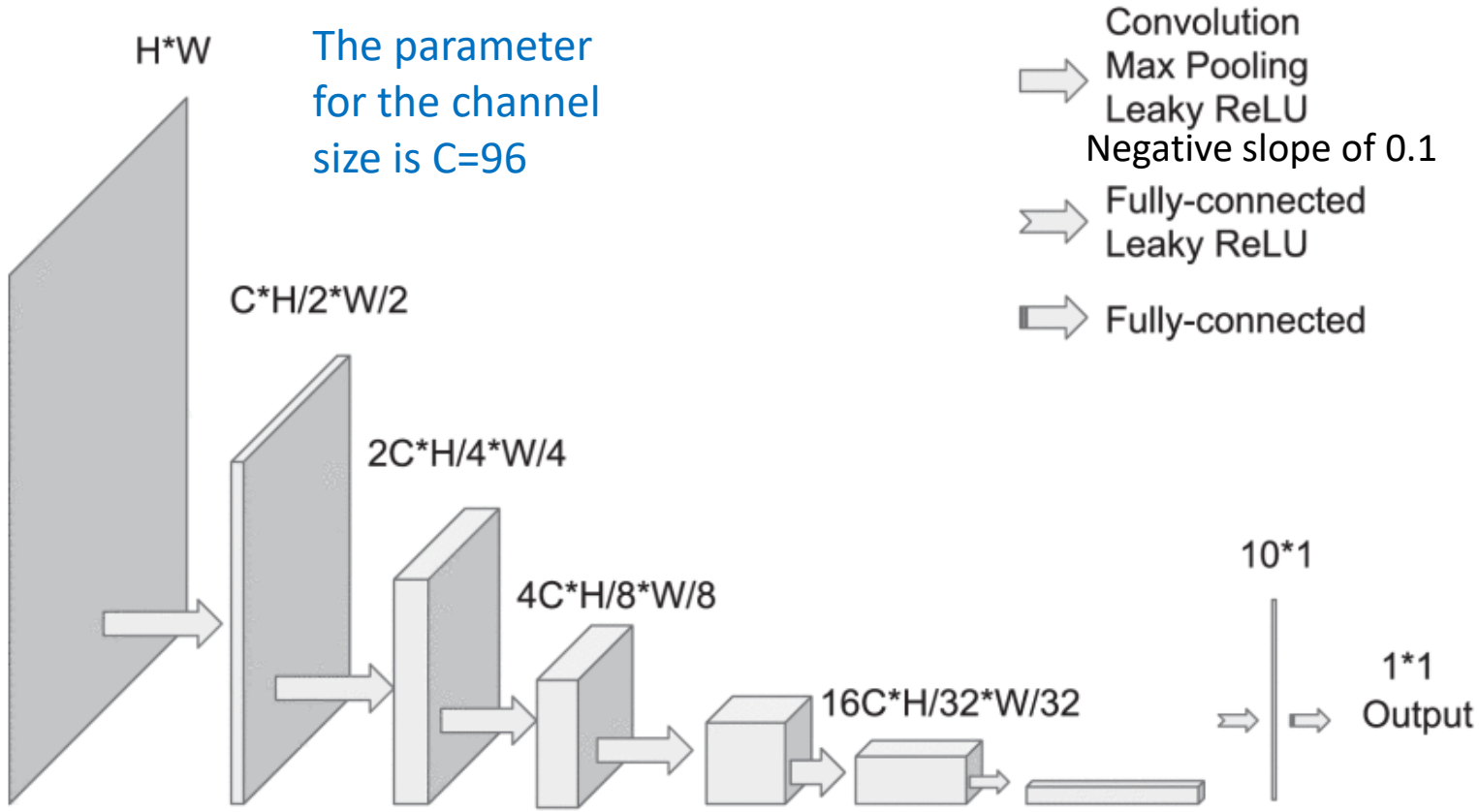
Parameters: current volume estimate $\mathbf{x}$

1: sample the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \sim p_{\boldsymbol{\theta}}$
2: sample the 2D shifts $\mathbf{t} = (t_1, t_2) \sim p_{\mathbf{t}}$
3: sample the CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}}) \sim p_{\mathbf{c}}$
4: generate a synthetic noiseless projection based on (2), with $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\varphi}\mathbf{x}$
5: sample the noise $\mathbf{n} \sim p_{\mathbf{n}}$. Add to the projection as $\mathbf{y}_{\text{sim}} = \mathbf{y}_{\text{noiseless}} + \mathbf{n}$.

CryoGAN—Discriminator Network

- Discriminator Network differentiates projections between from data and from simulation.
- The gradients from discriminator are then used by the simulator to update.
- The architecture in discriminator network in CryoGAN is similar to the one in standard GANs
  - 8 layers (6 convolutional blocks; 2 fully connected layers)
  - Each convolutional block made of a convolutional layer, max pooling and a leaky ReLU.

Input image with size H*W is downsampled to a scalar of output



Leaky ReLU: $f(x) = \begin{cases} x \text{ if } x > 0 \\ 0.01x \text{ otherwise} \end{cases}$

Experiments and results

- **For each experiment, the dataset is randomly divided into halves which are separately reconstructed.**
  - The two reconstructed structures are aligned
  - FSC is calculated between them by using the FOCUS [1,2]. The half-half resolution where the FSC crosses 0.143.
- **The two true structures are used**
  - PDB-*5ala* atomic model for the β-galactosidase (with dataset EMPIAR-10061) case
  - Preprocessed EMDB-2660 as the true structure for the 80S Ribosome case
- **Those true structures are used to created the datasets in every synthetic experiment.**

FSC: Fourier shell correlation

[1]: N. Biyani *et al.*, J.Struct. Biol. (2017)
[2]: R. D. Righetto *et al.*, Nature Commun. (2019)

Synthetic experiment and its setup (1)

- Experimental setup:

[1]: E.F. Pettersen et al., J. Comput. Chem., 2004

  - **A 2.5 Å density map** from **PDB-*5ala* atomic model** is generated by using Chimera software [1].
    - A volume of size, $302 \times 233 \times 163$, with voxel size 0.637 Å is obtained.
    - A downsampled volume of size, $180 \times 180 \times 180$, with voxel size 1.274 Å is then produced.
  - **41 000 projections are generated from the volume** according to the forward model.
    - The projection directions are sampled from a uniform distribution over 3D rotations.
    - The **CTF parameters for each projection** are extracted by using CTFFIND4 on a micrograph in EMPIAR-10061.
    - **Random noise patch** from the same micrograph is selected.

→ The **synthetic data** produced from synthetic experiment are referred as **real**.

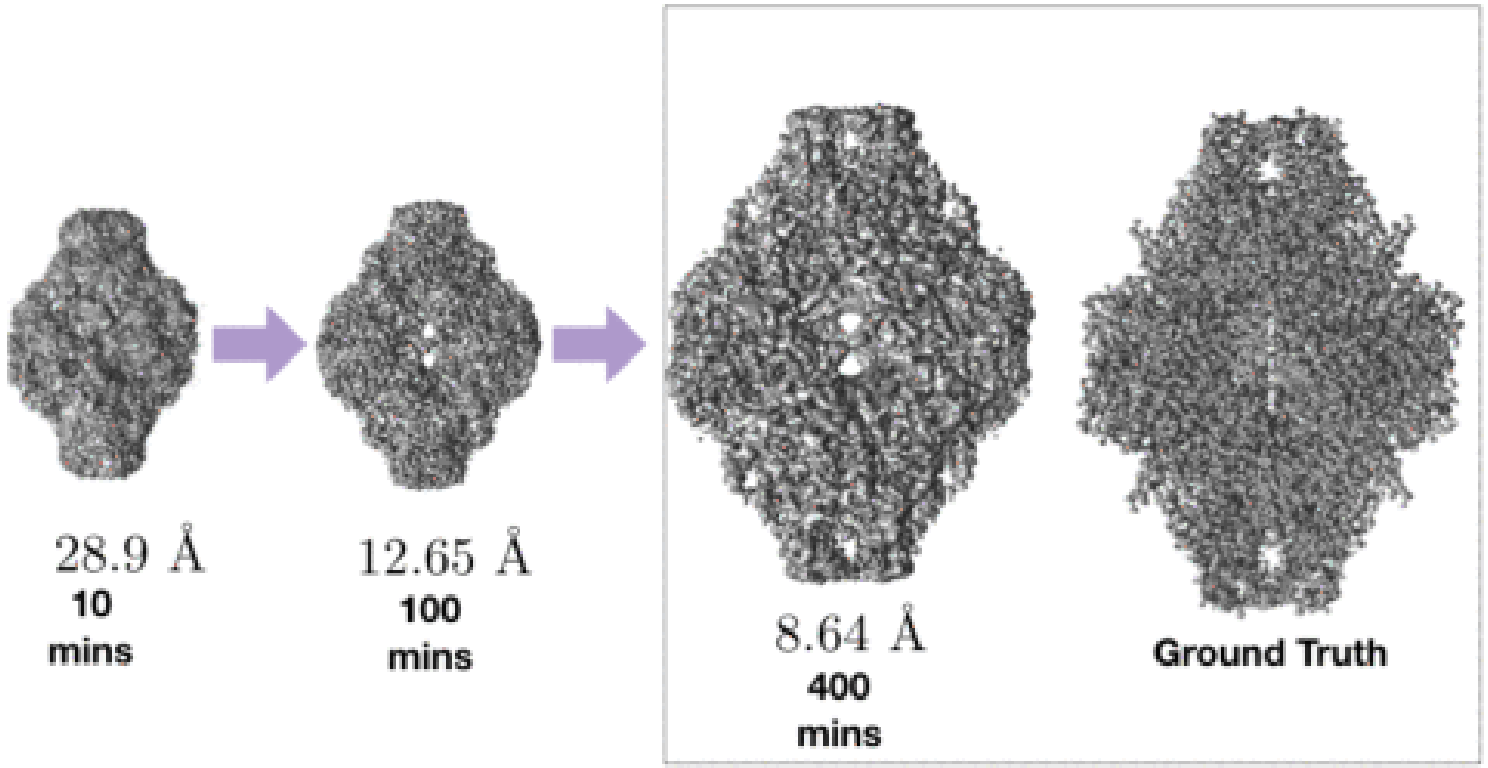
→ The production from the **physics simulator** is referred as **simulated**.

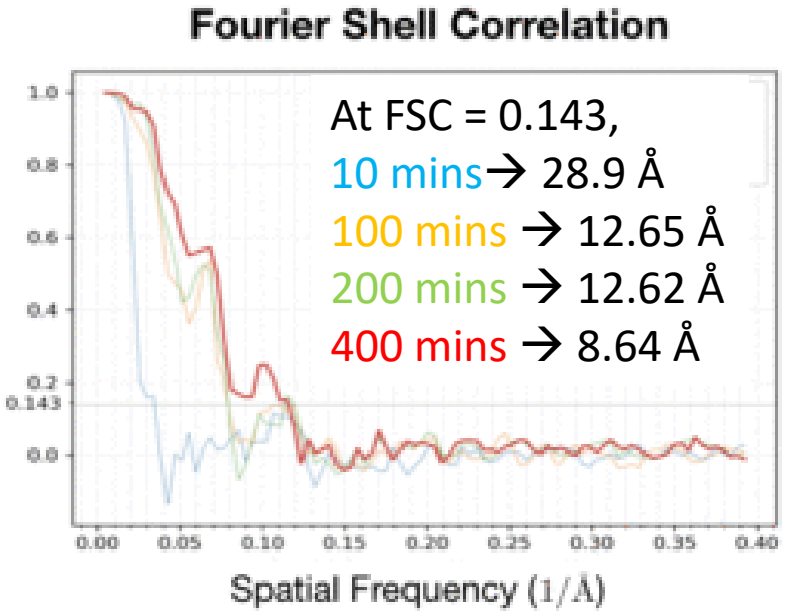
Synthetic experiment and its setup (2)

- **Generator settings:**
  - The generator was initialized with $184 \times 184 \times 184$ voxels. Pixel size: 1.274 Å
- **Training settings:**
  - Optimizer is Adam ($\beta_1 = 0.5, \beta_2 = 0.9, \epsilon = 10^{-8}$)
  - Learning rate of 10^{-3} and a batch size of 8 are set.
  - Algorithm runs for 40 epochs and learning rate decreased by 1% every epoch.
  - $\lambda = 0.001$ and $n_{\text{discr.}} = 4$.
- **Discriminator settings**
  - The weights are initialized to random values.
  - All projections are normalized to zero mean and unit standard deviation.

Synthetic experiment and its results

- CryGAN algorithm is run for 400 minutes on NVIDIA V100 GPU.
- A reconstruction with a half-map resolution of 8.64 Å is obtained.





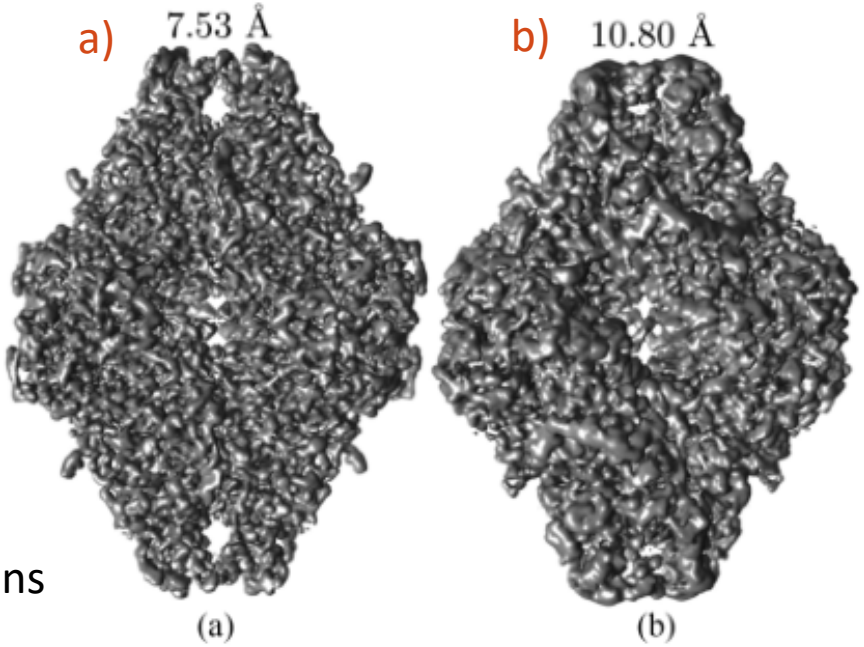
Corresponding evolution of FSC, for different time, v.s. spatial frequency

Experiment with different imaging conditions

- Different imaging conditions are given to understand its effect on CryoGAN reconstruction. Additional synthetic productions are created with the following conditions:
  a) A low noise level of -5.2 dB SNR
  b) A realistic noise level of -20 dB SNR and translations (3% of the image size)
  c) A realistic noise level of -20 dB SNR and large translations (20% of the image size)

Decreasing the noise level improves the resolution.
This case shows resolution is better for 15%.

	SNR (dB), translation (%)			
metric	-20, 0	-5.2, 0	-20, 3	-20, 20
half-map FSC = 0.143	8.6	7.5	10.8	14.3
truth FSC = 0.5	15.3	8.3	14.7	23.2
truth FSC = 0.143	11.7	6.5	11.5	19.8



$\text{SNR} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right)$ [*dB*]

-20 dB → the power of the noise is 100 times greater than the power of the signal

The presence of translations lowers resolutions.

Image SNR

Its content from [N. Marturi et al., J. of Scanning Microscopies 2014]

- Signal-to-noise ratio, SNR, in signal processing estimates the level of noise that present in a recorded signal [T. Bose *et al.*, Digital Signal and Image Processing, 2004]. SNR is given by:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{variance(f)}{variance(n)}\right)\ [\mathrm{dB}]$$

f is noiseless image

- In scanning electron microscope, SEM, image, its noises including primary, secondary emission, scintillator, phtocathode, photomultiplier follows Poisson distribution. Under the central limit theorem, the final noise can approximated as Gaussian distribution.

- The above definition can be reformulated with standard deviation, $\sigma = \sqrt{variance}$ and the SNR is then written as:

$$\mathrm{SNR} = 20\log_{10}\left(\frac{\sigma_{\hat{f}}}{\sigma_n}\right)\ [\mathrm{dB}]$$

where σ_f and σ_n are the standard deviations of noiseless image $\hat{f}$ and noise image *n*, respectively. Specifically, the standard deviation of the pixel values in image for computation.

- The observed image, $g = f + n$, and the noiseless image $\hat{f} = \mathrm{median}\{g\}$. So $n = g - \hat{f}$.

Alternative definition of SNR which is equal to $\frac{\mu}{\sigma}$ where μ is signal mean and $\sigma$is the standard deviation of noise [A. Rowlands, Physics of Digital Photography, IOP 2017]. So the definition of SNR should depend on field.

Experiment with mismatch in pose

- **The experiment is set to understand the effect from mismatch in pose distributions between real projections and simulated projections.**

- **Three synthetic productions are used with different pose distributions, p_θ^{true}**
  - Uniform distribution same as in the synthetic experiment
  - Nonuniform distributions determined by Euler angles computed from sampling 3D multivariate Gaussian. The Gaussians have zero mean and diagonal covariance $\sum \text{diag}(\sigma, 2,1)$.

- **The above pose distributions, $p_\theta^{\text{rec.}}$, are used in physics simulators.**

Experiment with mismatch in pose and its results

- The **best resolution** is achieved when the **pose distribution matches the real one**.
- An uniform $p_\theta^{\text{rec.}}$ often gives reasonable resolution → it's a good choice when the true distribution is unknown.

Reconstruction resolution (Å) for β-galactosidase

	reconstruction distribution		
true distribution	uniform	$\sigma = 2$	$\sigma = 3$
half-map FSC = 0.143			
uniform	8.6	13.2	17.2
$\sigma = 2$	9.2	9.4	14.7
$\sigma = 3$	8.6	10.6	12.7
ground truth FSC = 0.5			
uniform	15.3	16.1	18.6
$\sigma = 2$	16.4	15.0	15.5
$\sigma = 3$	16.4	16.9	16.1
ground truth FSC = 0.143			
uniform	11.7	12.6	14.5
$\sigma = 2$	11.6	10.4	10.7
$\sigma = 3$	11.5	14.0	12.3

Experiment on 80S Ribosome with various noise levels

- The ground-truth structure with dimensions (180X180x180) and voxel size of 2.68 Å is obtained by processing EMDB-2660 reconstruction from EMPIAR-10028
- Five noise levels, 20 dB, 0 dB, -5.2 dB, -14 dB and -20 dB, are given to generate synthetic productions.
  - Noise patches are extracted from the micrograph background in EMPIAR-10028
- The translations are kept at 5% of the image size for all productions.
- Other imaging conditions are the same as EMPIAR-10028.

Experiment results on 80S Ribosome with various noise levels (1)

- Algorithm is run for 20 epochs.
- The resolution results are shown in the below table

Reconstruction resolution Å for synthetic ribosome

More noise

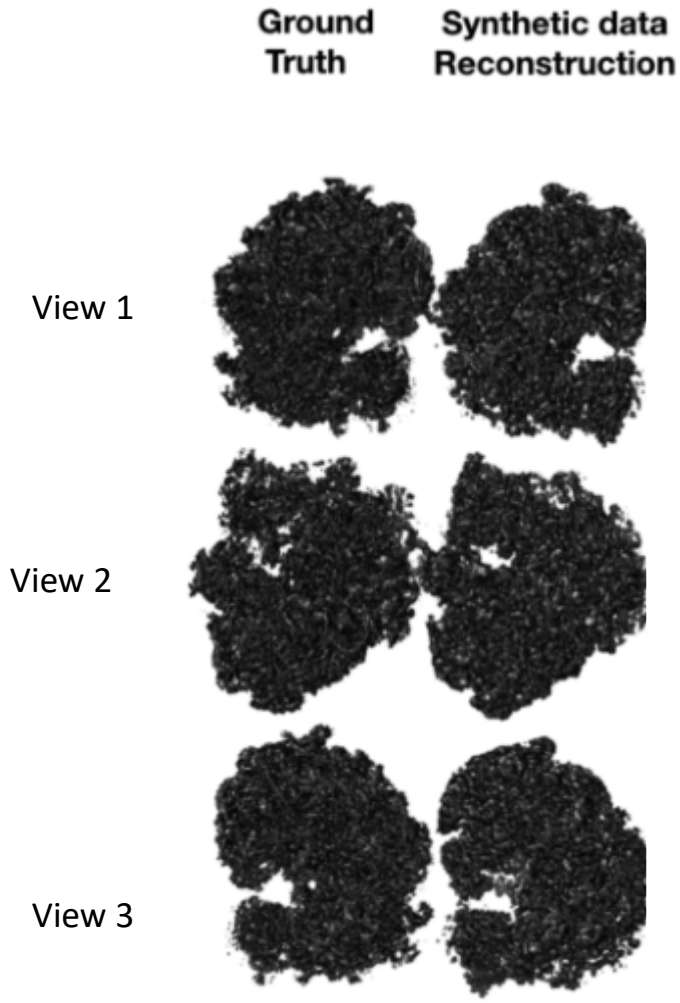
	noise level (dB)				
metric	20	0.0	-5.2	-14	-20
half-map FSC = 0.143	15.5	18.8	19.5	63.0	57.1
truth FSC = 0.5	20.9	21.9	24.1	81.0	104.3
truth FSC = 0.143	15.3	19.6	20.5	45.1	57.2

The resolution decreases with increasing noise level.

Experiment results on 80S Ribosome with various noise levels (2)

- **Ground truth**: 80S Ribosome (EMD-2660) from EMPIAR-10028
- **Synthetic data reconstruction**: CryoGAN reconstruction from synthetic production with -5.2 dB and 5% translations.
- The resolution for Ribosome is worse than the β-galactosidase by, for example, ~400% with –20 dB and 3-5% of translations → because of symmetry in Ribosome ?

Different views of 80S Ribosome and reconstruction using CryoGAN



Experimental Data—β-galactosidase (1)

- **To test the CryoGAN capacity, β-galactosidase projection data [1] are used.**
  - Data consist of 1 539 micrographs and 41 123 particle locations.
  - Data are obtained by automated particle picking with 3D classification.

[1]: A. Bartesaghi *et al.*, Science (2015)

- **Data settings**
  - Data are downsampled from (384X384) to (192X192) with a pixel size of 1.274 Å.
  - The CTF parameters for each micrograph are estimated by using CTFFIND4.
- **Generator settings**
  - A volume of size (180X180X180) is reconstructed.
  - Noise patches* are added to simulated projections during run.

*: Noise patches are randomly extracted from micrograph's background where no particle. This is done by identifying patches with the lowest variance.

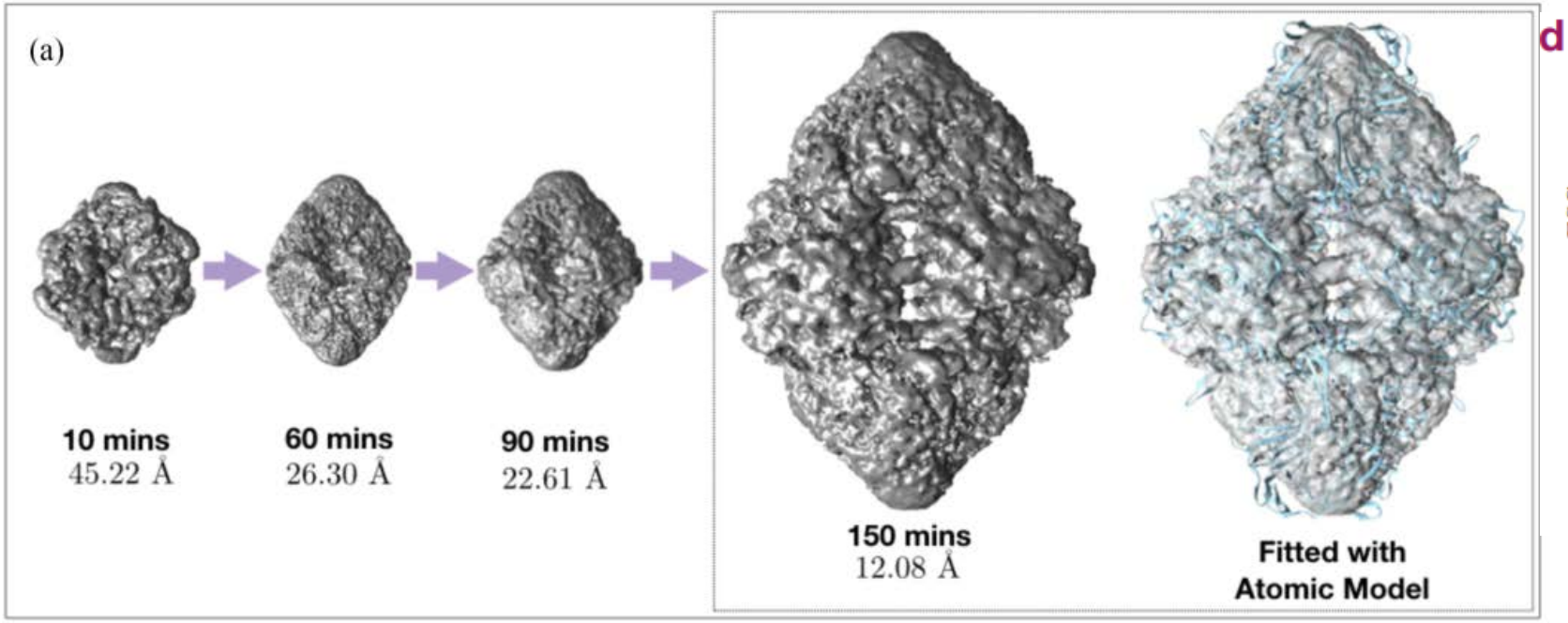
Experimental Data—β-galactosidase (2)

- **Discriminator Architecture**
  - Discriminator are initially identical for the two half datasets
  - Real and simulated projection images are smoothed with a Gaussian kernel before discriminator.
  - The standard deviation of the kernel is initially 2 and changes every iteration. It decreases by a total of 2% each epoch.
- **Training settings**
  - Algorithm runs for 15 epochs and learning rate decreases by 8% every epoch.
  - $\lambda = 1$ and $n_{\mathrm{discr.}} = 4$.

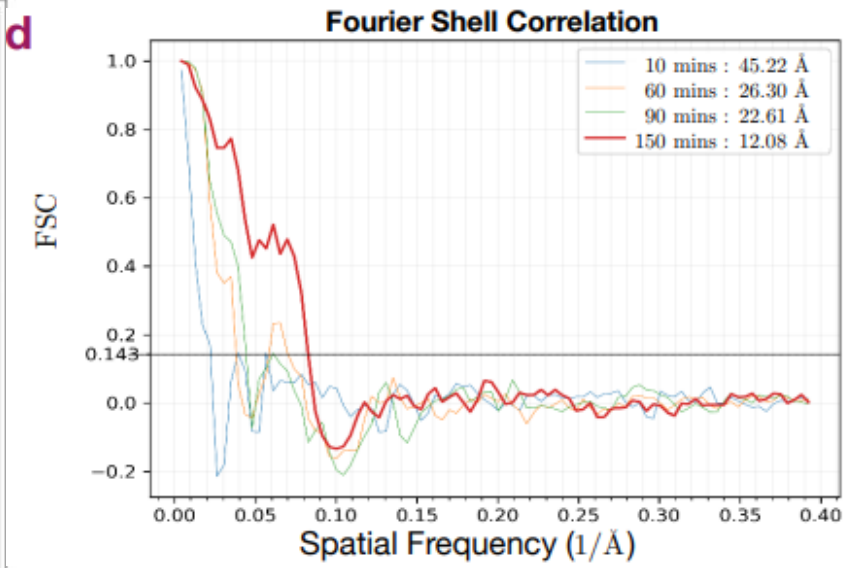
Experimental Data—β-galactosidase and its results

- CryoGAN algorithm runs for 150 minutes on NVIDIA V100 GPU.

Evolution while reconstructing the experimental β-galactosidase dataset.



d

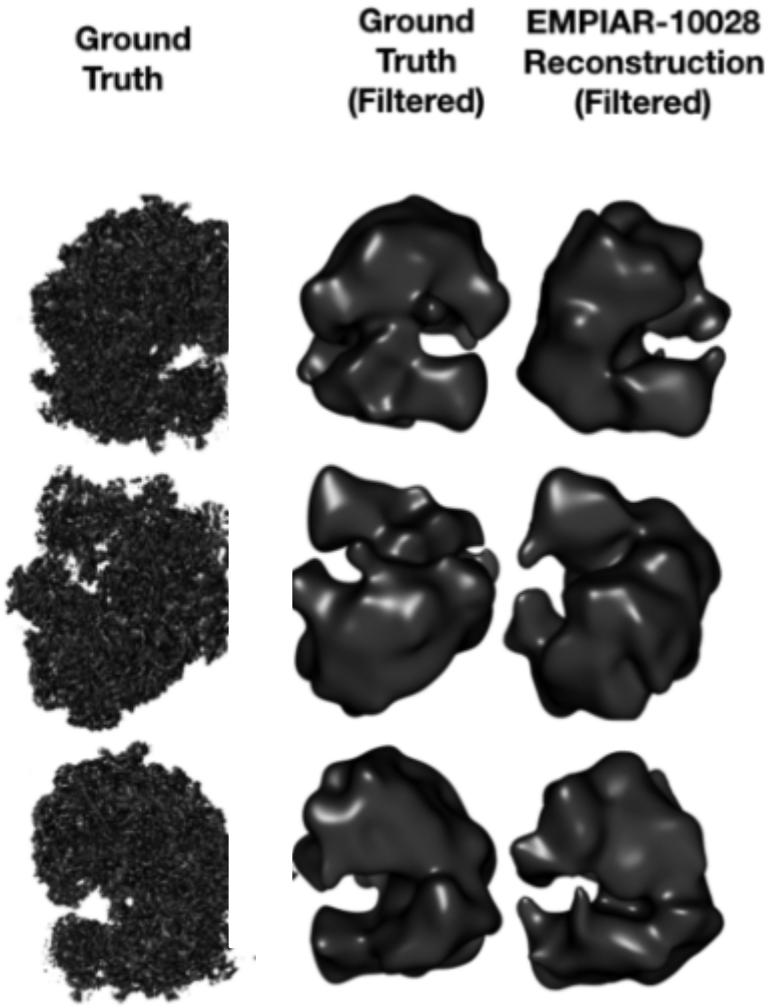


At FSC = 0.143,
10 mins → 45.22 Å
60 mins → 26.3 Å
90 mins → 22.61 Å
150 mins → 12.08 Å

Experimental Data—80S Ribosome and its results

- Experimental projection data provided by EMPIAR-10028 has 102 547 projections and is used.
- A half-half resolution of 55.5 Å is achieved.
- The resolutions w.r.t. the true structure are 100.7 Å and 54.2 Å for 0.5 and 0.143 FSC cutoffs, respectively.

- **Ground truth**: 80S Ribosome (EMD-2660) from EMPIAR-10028
- **Ground truth**: 80S Ribosome filtered with a Gaussian with standard deviation of 5 voxels
- **EMPIAR-10028 Reconstruction (Filtered)**: CryoGAN reconstruction filtered with a Gaussian.



Extended study: simple GAN for Mnist data (1)

A simple GAN [1] is constructed and trained with Mnist.

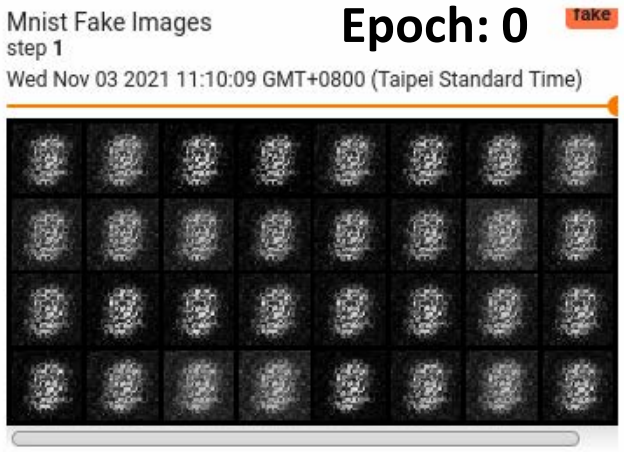
[1]: code on github is released by Aladdin Persson

- **Hyperparameters (sensitive to GAN)**:
  - Learning rate: 3×10^{-4}
  - Z =64
  - Image dim.: 28×28
  - Batch size 32
  - Number of epochs = 50
- **Discriminator**:
  - Its structure: Linear layer → LeakyReLU → Linear layer → Sigmoid
  - It goes to maximizes $\log\big(\mathrm{D}(\mathrm{real})\big) + \log\Big(1 - \mathrm{D}\big(\mathrm{G}(z)\big)\Big)$ for real and fake samples, respectively.
- **Generator:**
  - Its structure: Linear layer → LeakyReLU →Linear layer → Tanh layer (make outputs in [-1, 1])
  - It maximizes $\log\Big(\mathrm{D}\big(\mathrm{G}(z)\big)\Big)$ (minimize $\log\Big(1 - \mathrm{D}\big(\mathrm{G}(z)\big)\Big)$) for fake samples.
- Criterion measuring the Binary Cross Entropy, BCELoss, is used. The entropy loss between target, T_i, and input, x_i, can be given by:

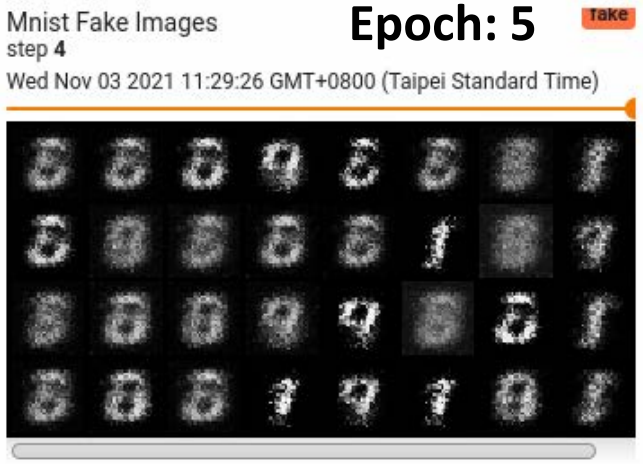
$$-\frac{1}{n}\sum_{i=1}^{n}\big[(T_i \times \log(x_i)) + \big((1 - T_i) \times (1 - \log(x_i))\big)\big]$$

Extended study: simple GAN for Mnist data (2)

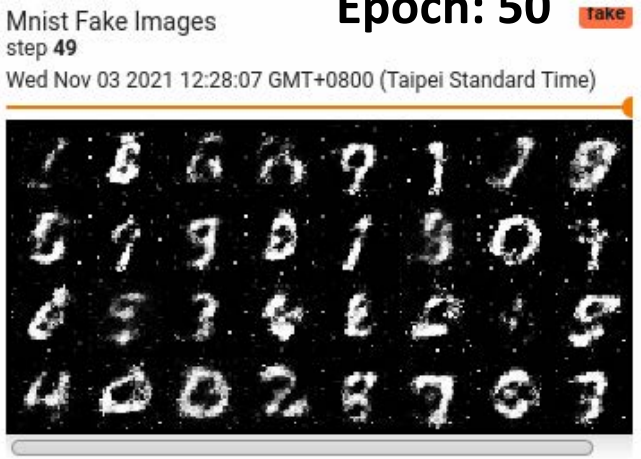
Fake



Epoch: 0



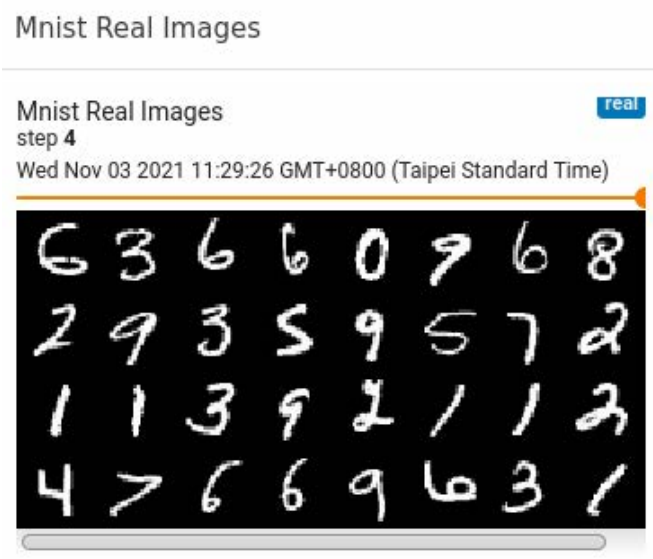
Epoch: 5



Epoch: 50

GAN is a powerful generative models but suffers from training instability.

Real



Conclusion and discussion

- CryoGAN can reconstruct reasonable the structures for β-galactosidase and 80S Ribosome, even from highly noisy experimental data.
- This paradigm is believed to open a new field with likelihood-free algorithms for cryoEM reconstruction.

- Theorem 1 does not guarantee that CryoGAN can provide a perfect reconstruction.
  - The empirical estimates in eq. 2 require infinite data to be accurate perfectly.
  - The loss function in eq. 2 is non-convex.
  - The discriminator may not span all the valid function *s* with Lipschitz constrain.
  - The algorithm approximates projections with discretization.
- The distribution-matching approach has been exploited by many algorithms [1,2]:
  - They minimize the KL-divergence to achieve distribution matching.

[1]: S.H. Scheres, "RELION", J. Struct.Biol., 2012
[2]: A. Punjani, et al., "cryoSPARC", Nature Methods, 2017

Backup

Assumptions on the image formation model

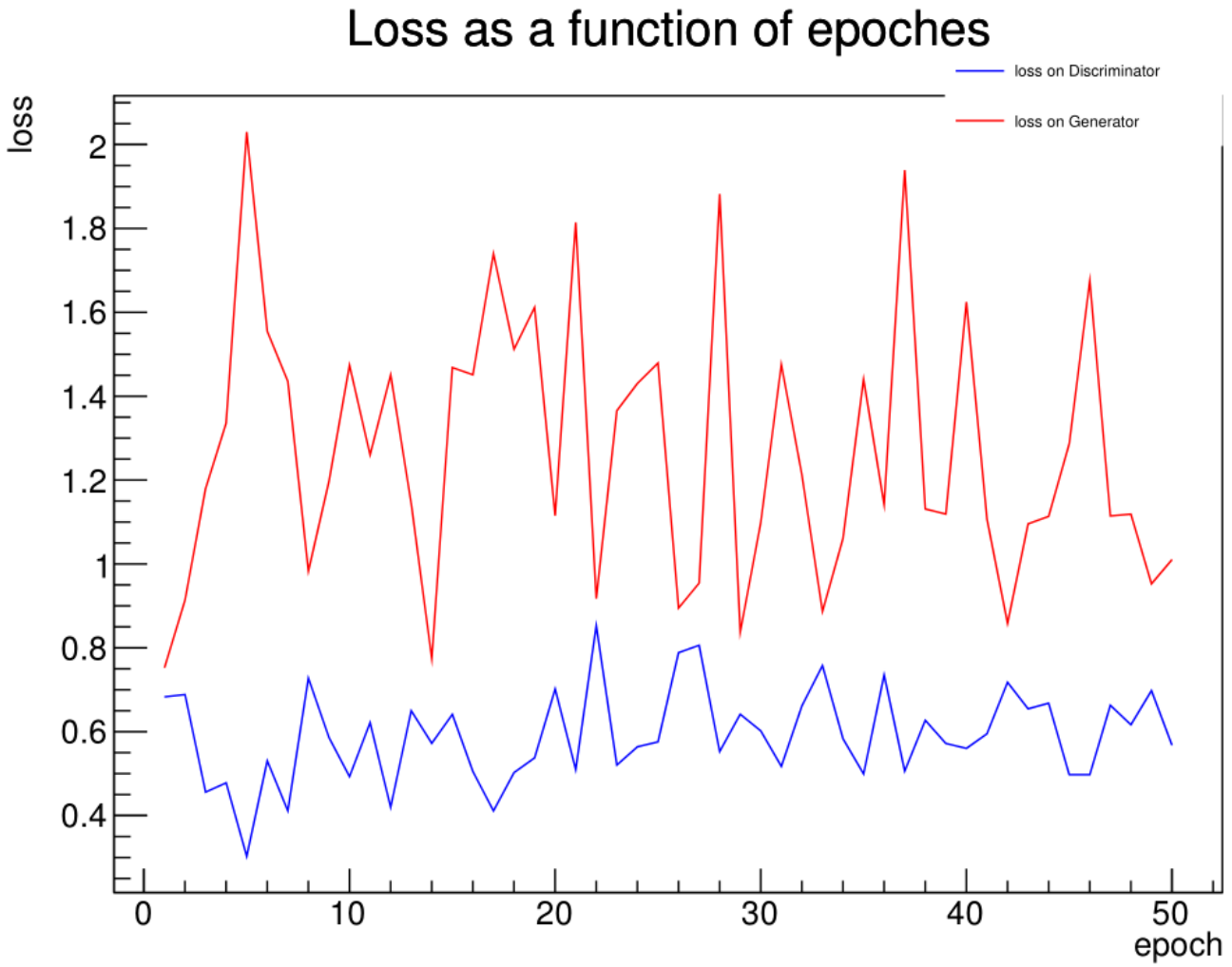
- In the continuous domain, the following **assumptions on the image formation model (forward model)**, $y = H_\varphi f + n$, are:
  - the characteristic functional $\widehat{\mathbb{P}}_n$ of the noise probability measure $\mathbb{P}_n$ is nonzero over its domain, and *n* is pointwise-defined everywhere in $\mathbb{R}^2$.
  - for any $\mathbf{c}_1, \mathbf{c}_2 \in \text{Support}\{p_\mathbf{c}\}$ and $\mathbf{c}_1 \neq \mathbf{c}_2$, $\left|\mathcal{F}\{C_{\mathbf{c}_1}\} + \mathcal{F}\{C_{\mathbf{c}_2}\}\right|$ is positive.
  - the volume, *f*, is nonnegative and has a bounded support
  - the probability distributions, $p_\theta, p_\text{c}, p_\text{t}$, are bounded.

Reconstruction resolution for Synthetic β-galactosidase

	SNR (dB), translation (%)			
metric	-20, 0	-5.2, 0	-20, 3	-20, 20
half-map FSC = 0.143	8.6	7.5	10.8	14.3
truth FSC = 0.5	15.3	8.3	14.7	23.2
truth FSC = 0.143	11.7	6.5	11.5	19.8

Loss in simple GAN

The loss in Discriminator and generator are shown:



Discriminator

Generator