

# Denoising cryo-EM data with conditional generative adversarial networks

---

Michael Cianfrocco  
Life Sciences Institute  
Department of Biological Chemistry  
University of Michigan

# Motivation

---

*How can we use previously determined structures to aid in particle picking and micrograph assessment?*

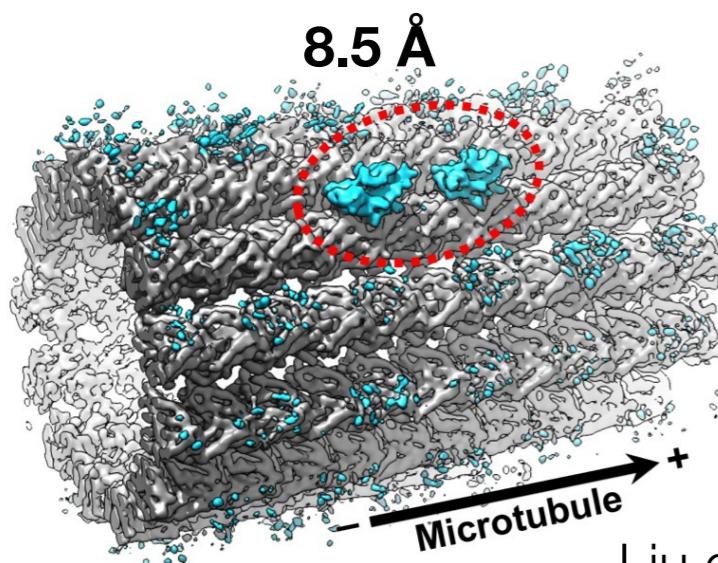
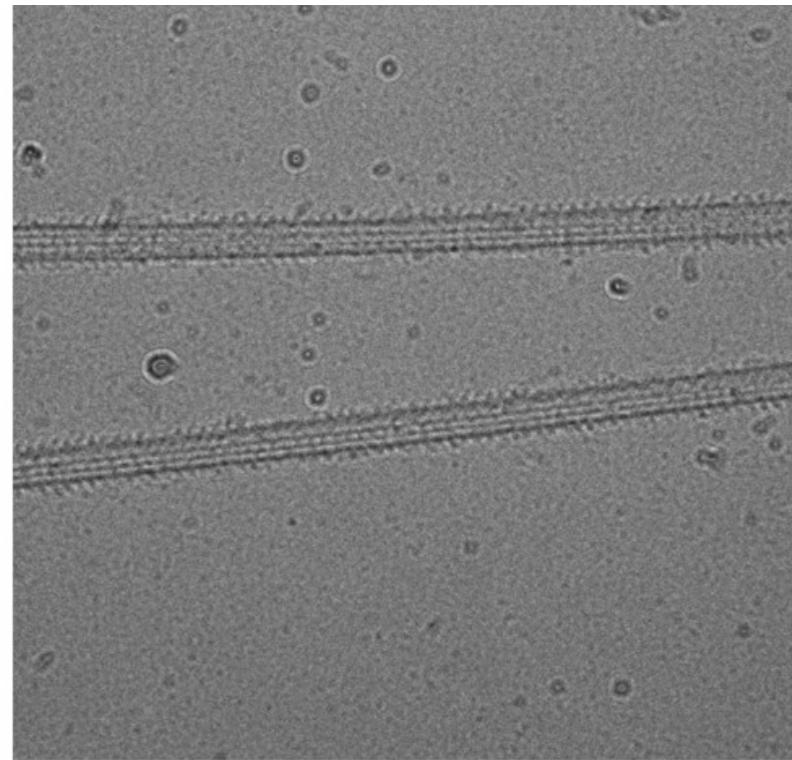
Two examples:

- 1) Small protein bound to large complex
  - ▶ Microtubule bound to a kinesin dimer
- 2) Previously known structures studied as a co-complex
  - ▶ Nucleosomes bound to chromatin remodeling factor

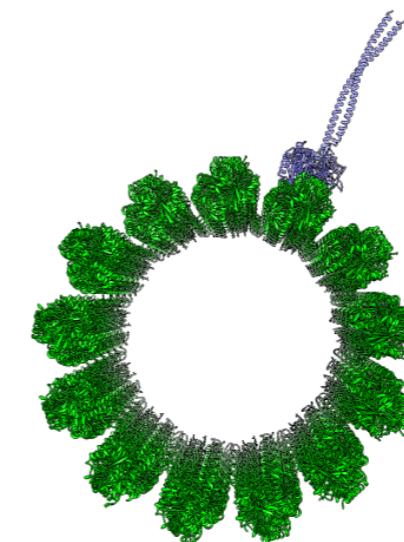
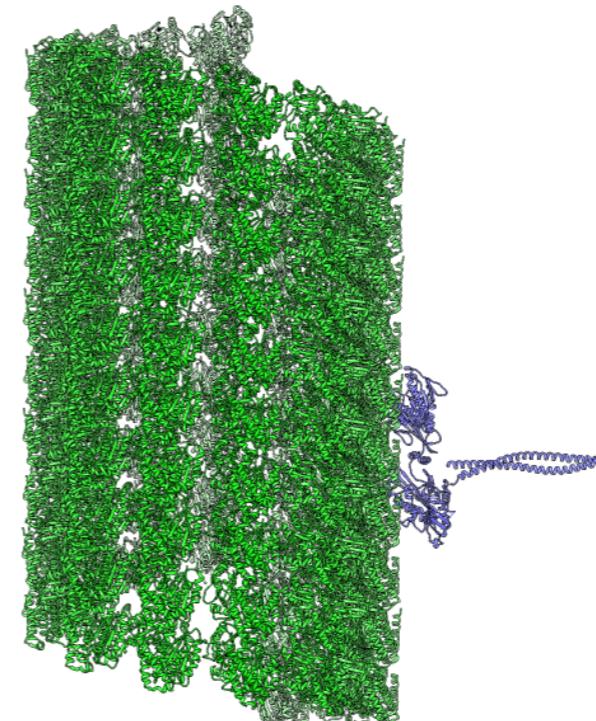


# Example #1: Small protein bound to larger complex

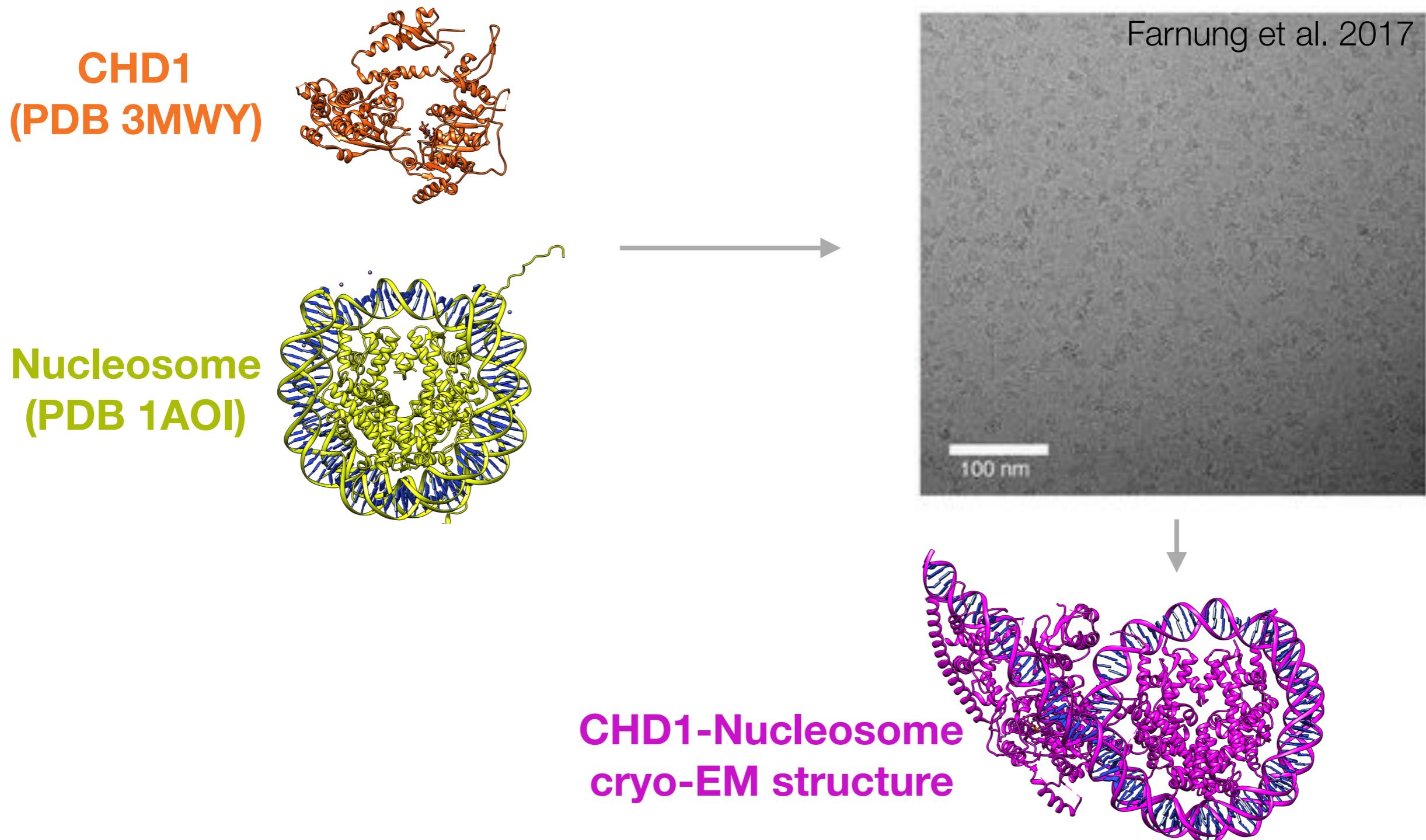
Kinesin bound to microtubule



Liu et al. 2017



## Example #2: Known structures in complex together



# Motivation

---

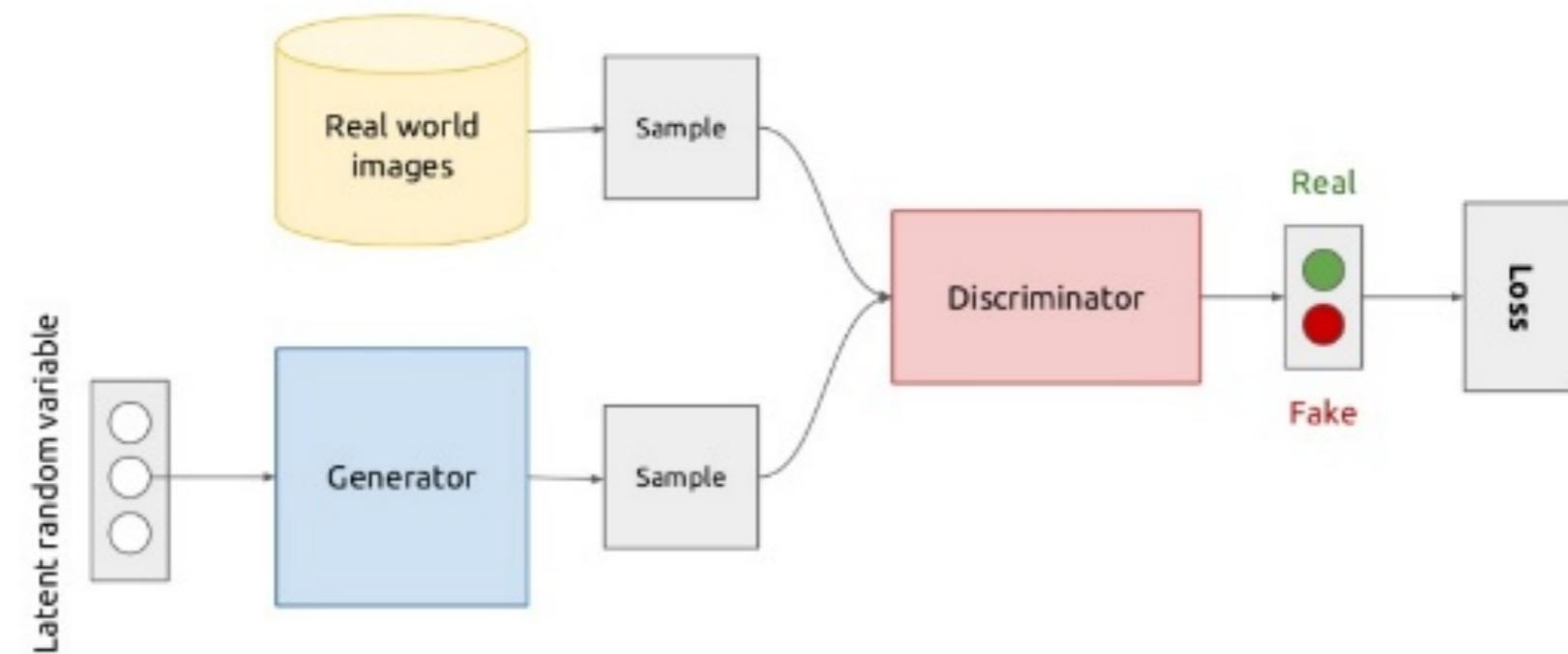
*How can we use previously determined structures to aid in particle picking and micrograph assessment?*

‘Super-resolution’ or ‘denoising’ approaches using neural networks

1. Generative adversarial networks
2. Conditional adversarial networks
3. Application to single particle cryo-EM

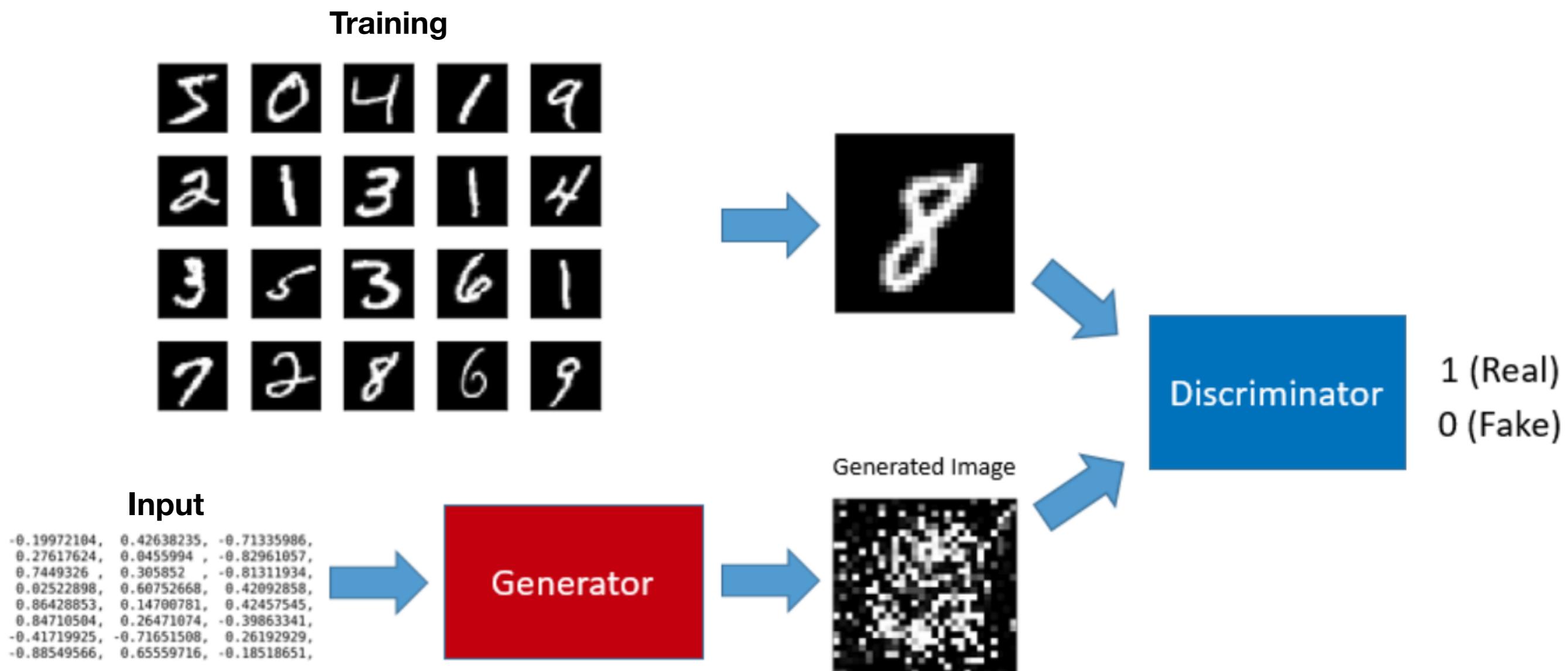


# Generative adversarial networks ('GANs')

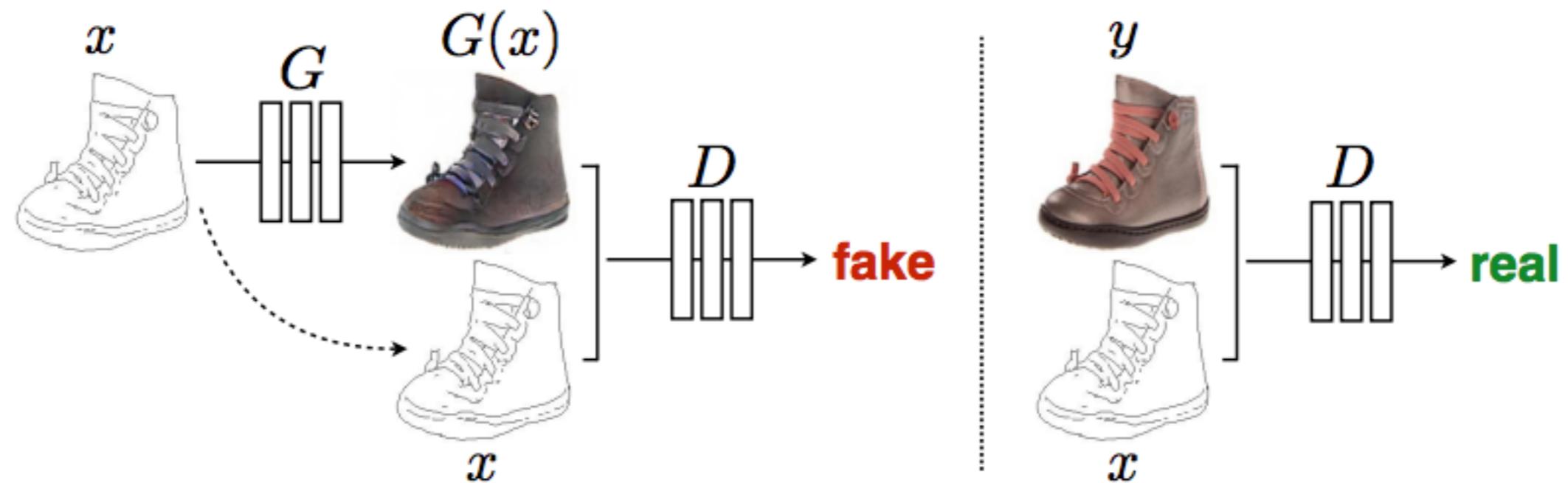


<http://www.slideshare.net/xavigiro/deep-learning-for-computer-vision-generative-models-and-adversarial-training-upc-2016>

# Generative adversarial networks ('GANs')



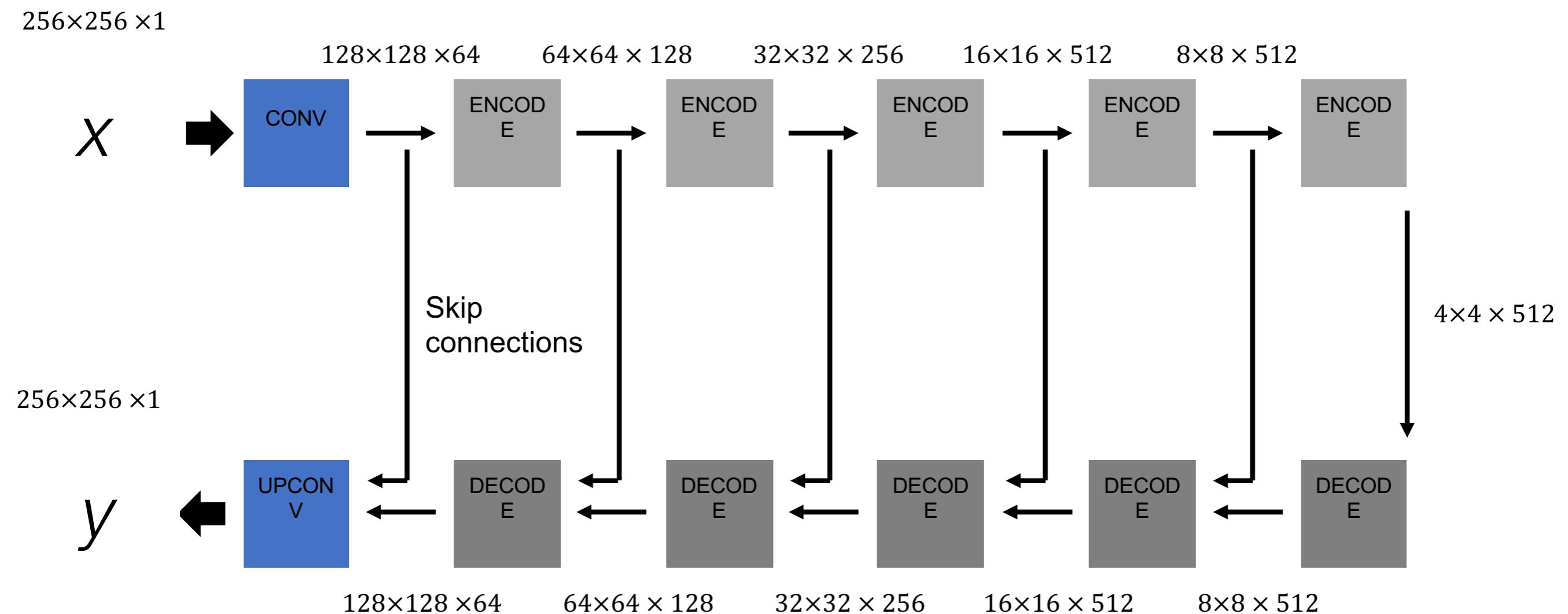
# Conditional adversarial networks ('cGANs')



Learn a mapping from observed image  $x$  and random noise vector  $z$ , to  $y$ ,  $G : \{x, z\} \rightarrow y$

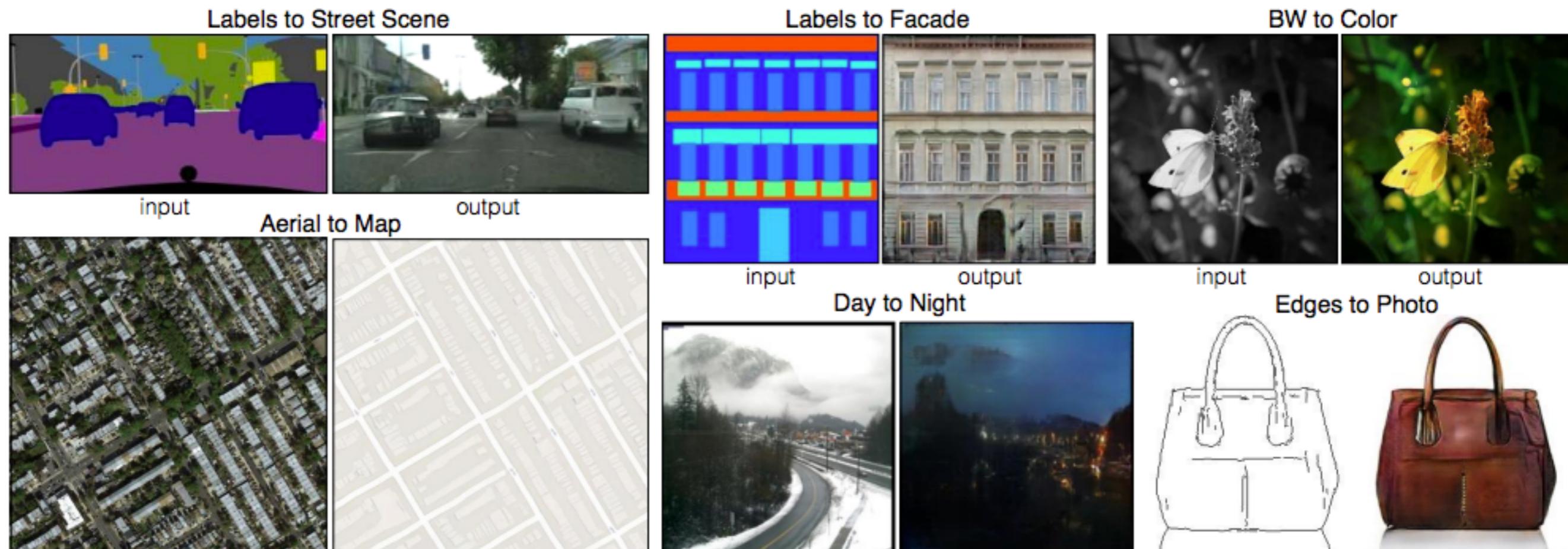
# Conditional adversarial networks ('cGANs')

*U-net architecture for generator*



# Conditional adversarial networks ('cGANs')

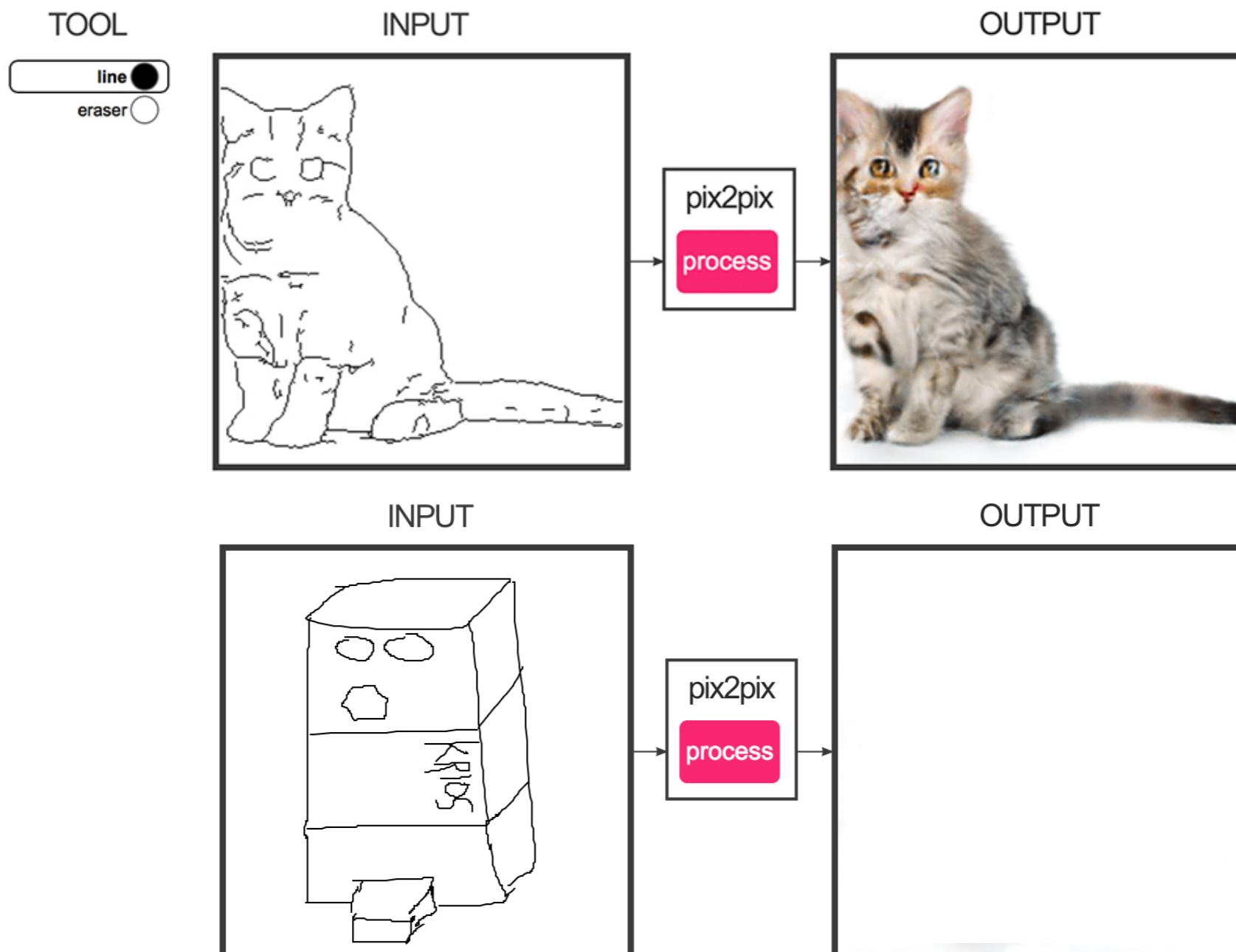
Application of algorithm ('pix2pix') to different image-to-image translations



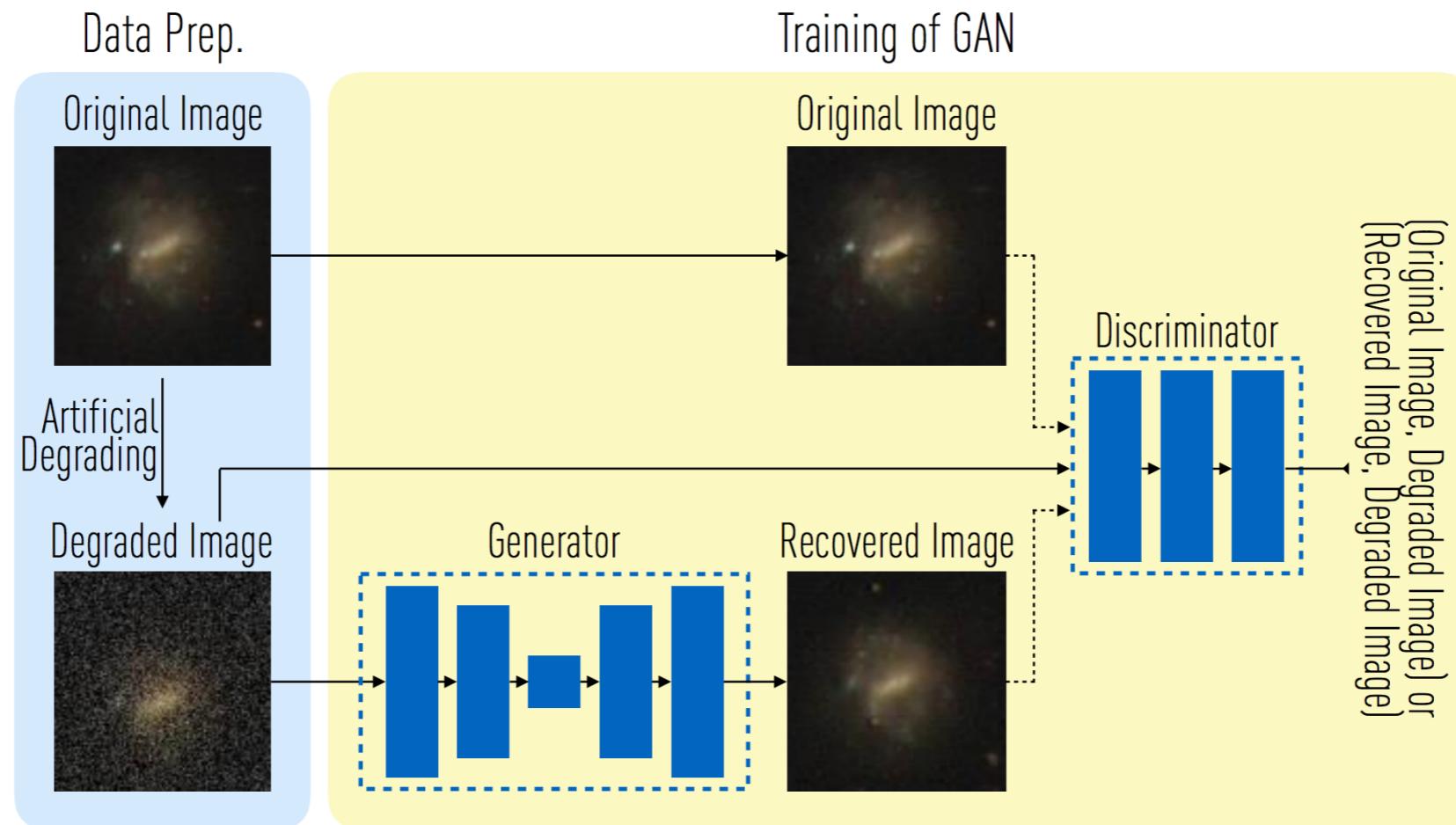
# Conditional adversarial networks ('cGANs')

Application of algorithm ('pix2pix') to different image-to-image translations

edges2cats



# Conditional adversarial networks: galaxyGAN



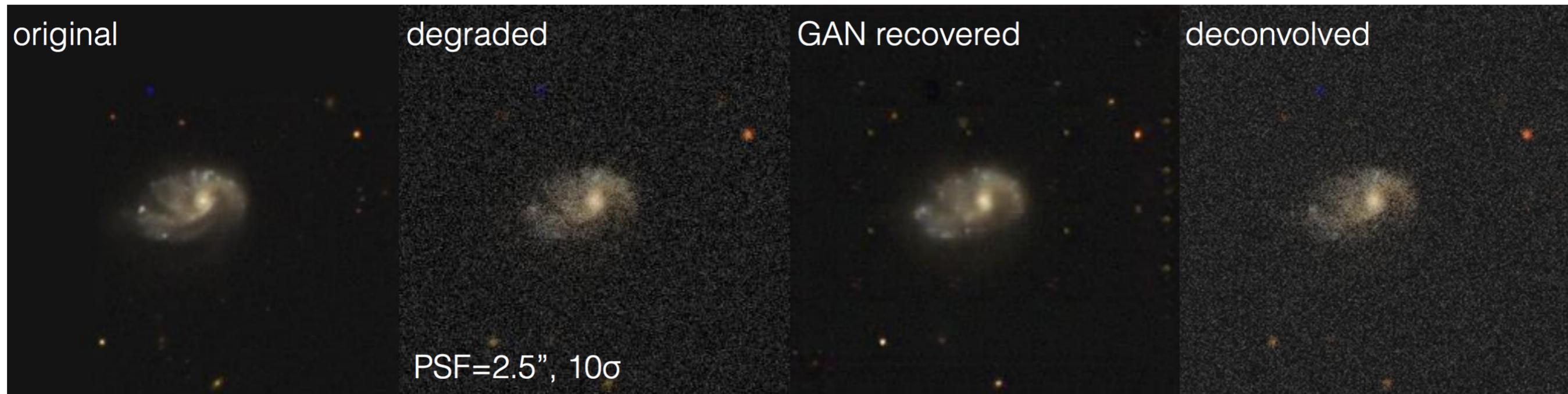
*Training:*

-4,550 galaxy images with  
noise added

*Testing:*

-Noise added images that had not  
been seen previously by cGAN

# Conditional adversarial networks: galaxyGAN



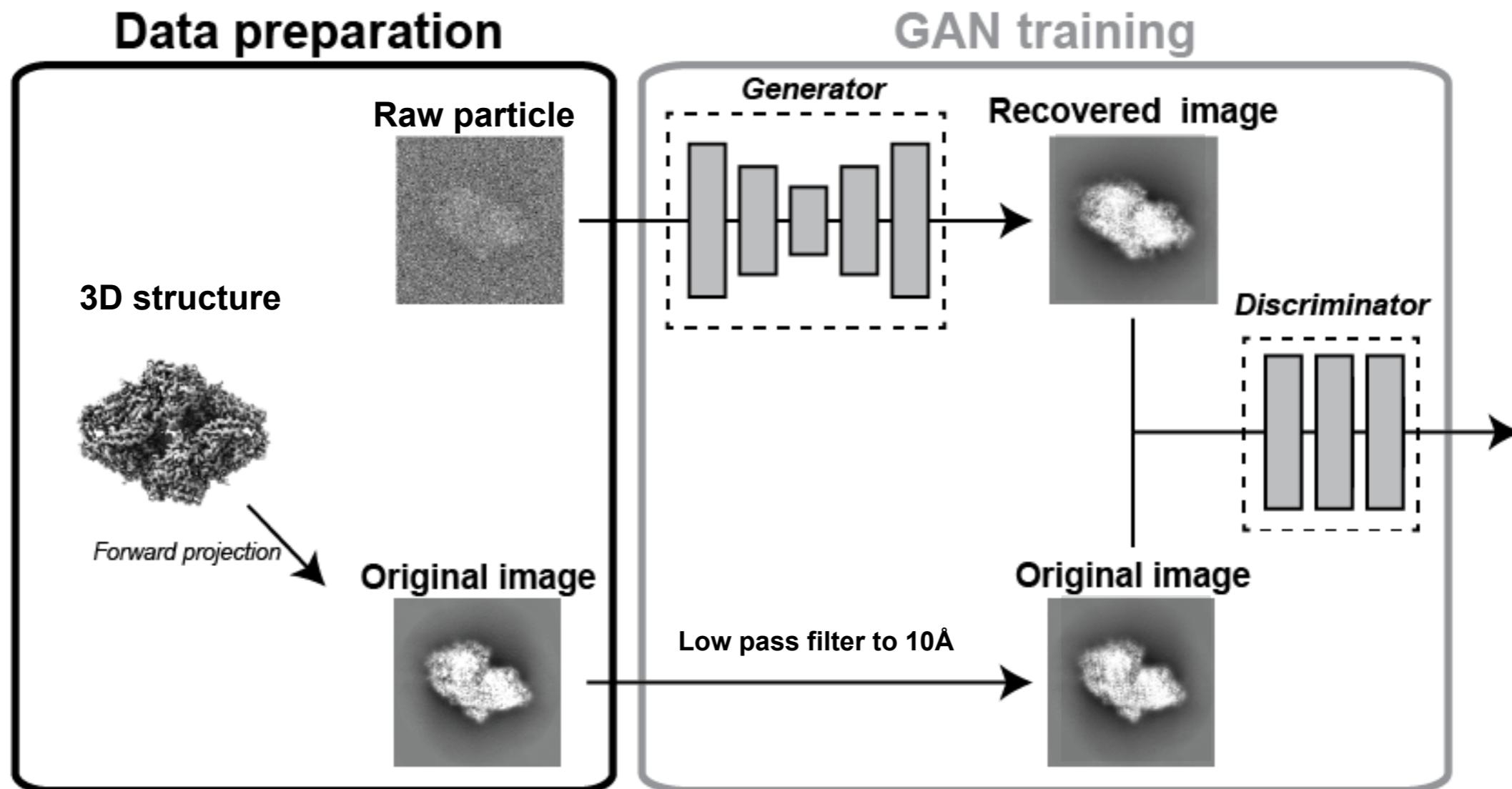
# *Training:*

-4,550 galaxy images with  
noise added

## *Testing:*

-Noise added images that had not been seen previously by cGAN

# Implementing conditional adversarial networks for cryo-EM



*Adaptation notes:*

*Tensorflow*

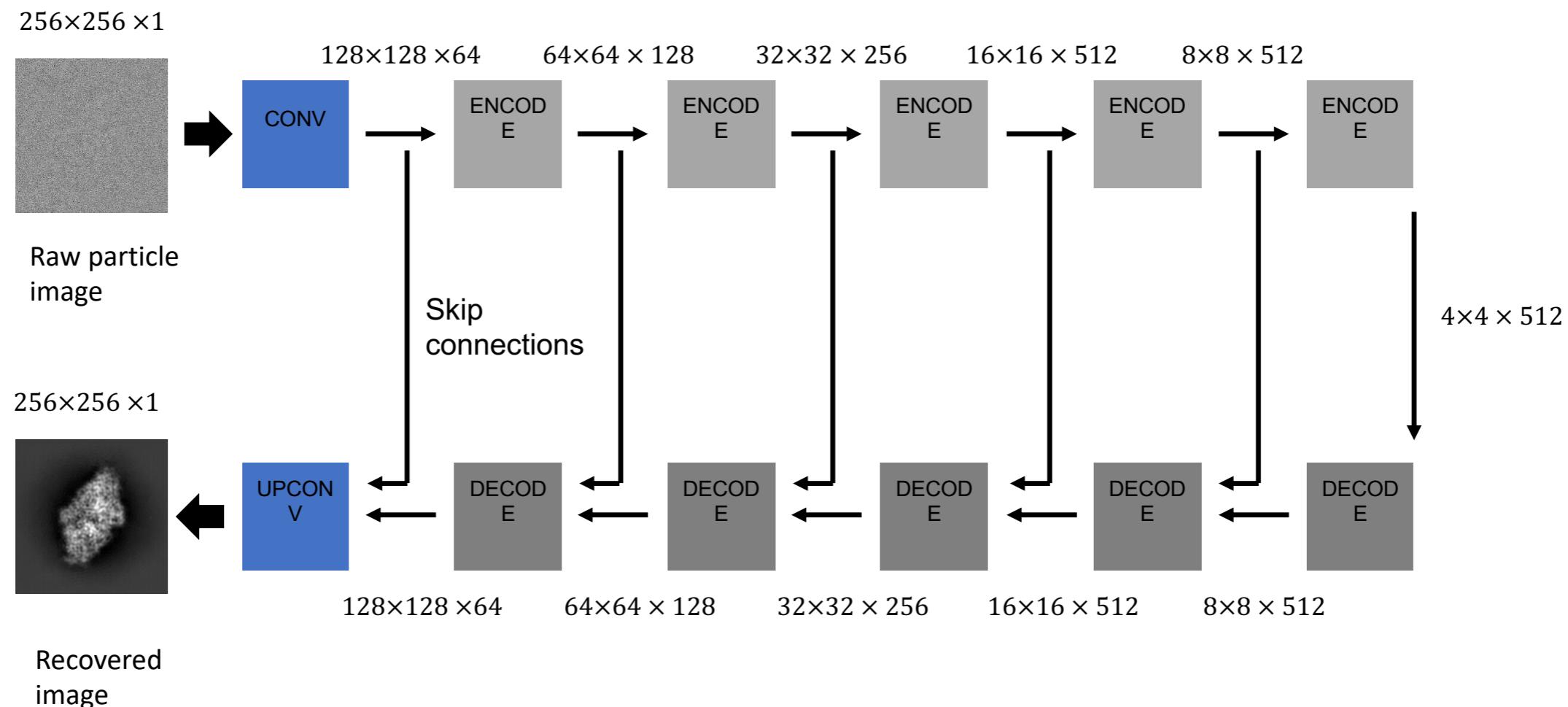
*Residual blocks*

*Exponential loss function*

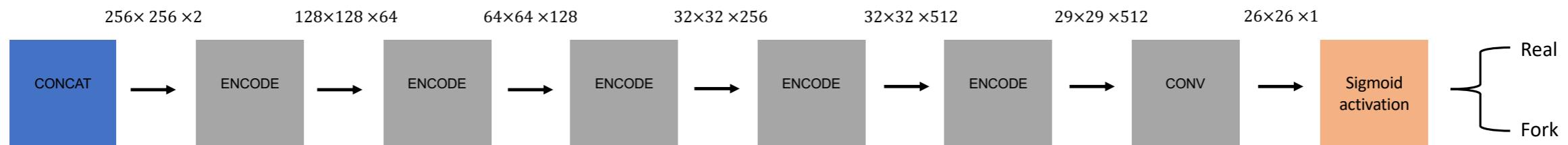


# Implementing conditional adversarial networks for cryo-EM

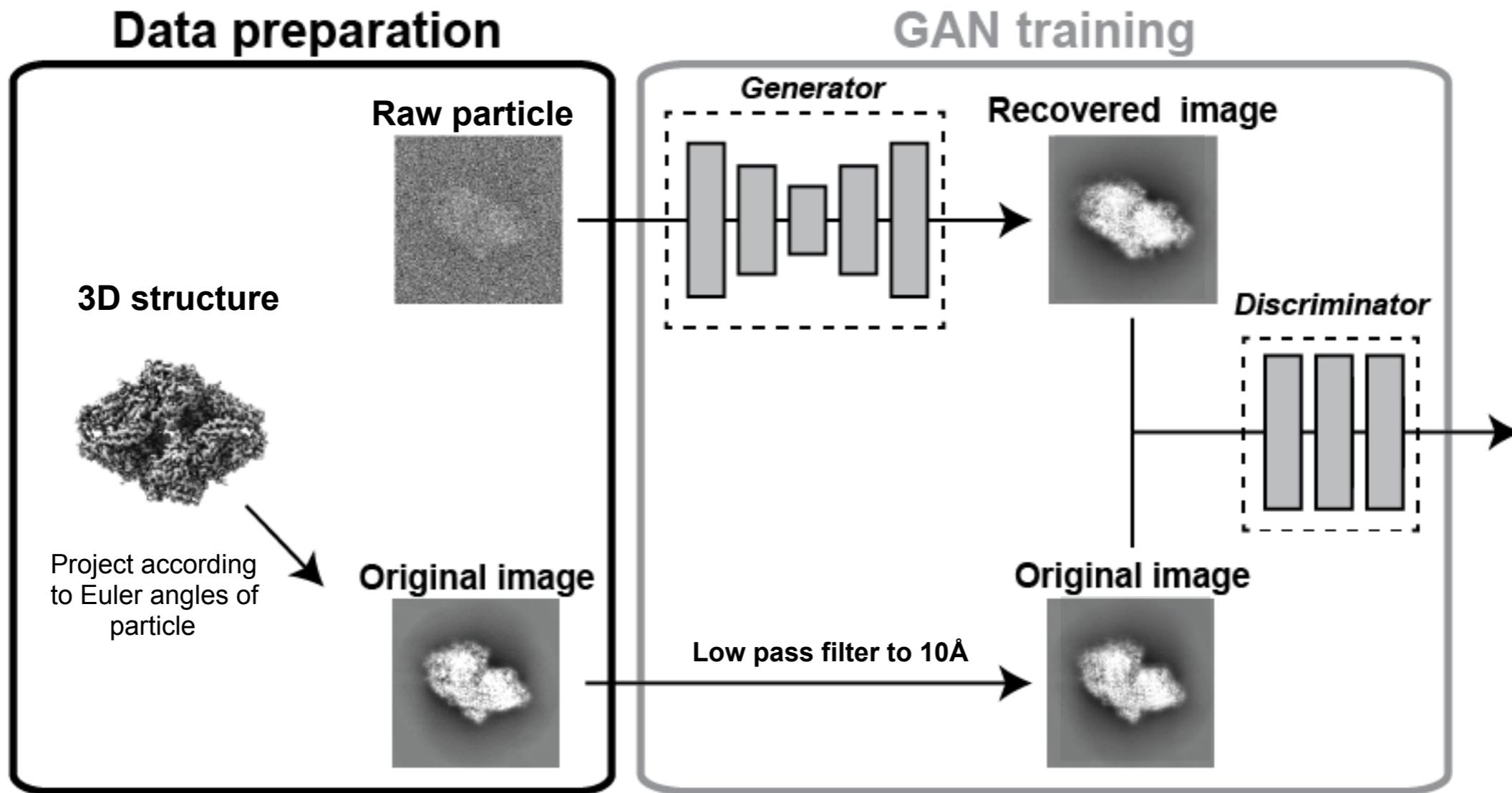
*Generator (U-net architecture):*



*Discriminator:*



# Test #1: Beta-galactosidase (EMPIAR10061)

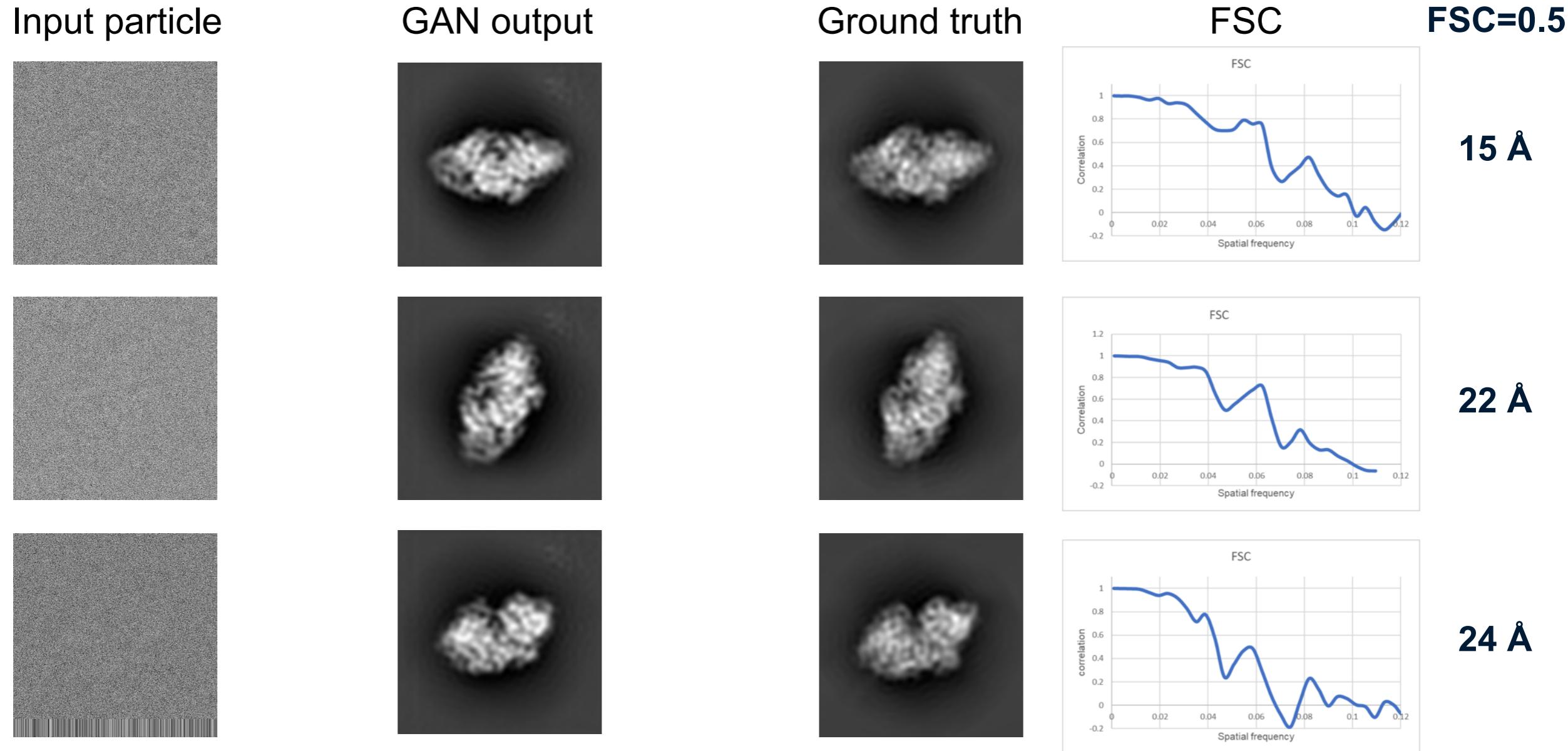


Training notes:

10,000 particle / projection pairs (randomly selected)  
256 x 256 (0.997 Å/pix)  
~10 hours on 1xGTX1080Ti GPU



# Test #1: Beta-galactosidase (EMPIAR10061)



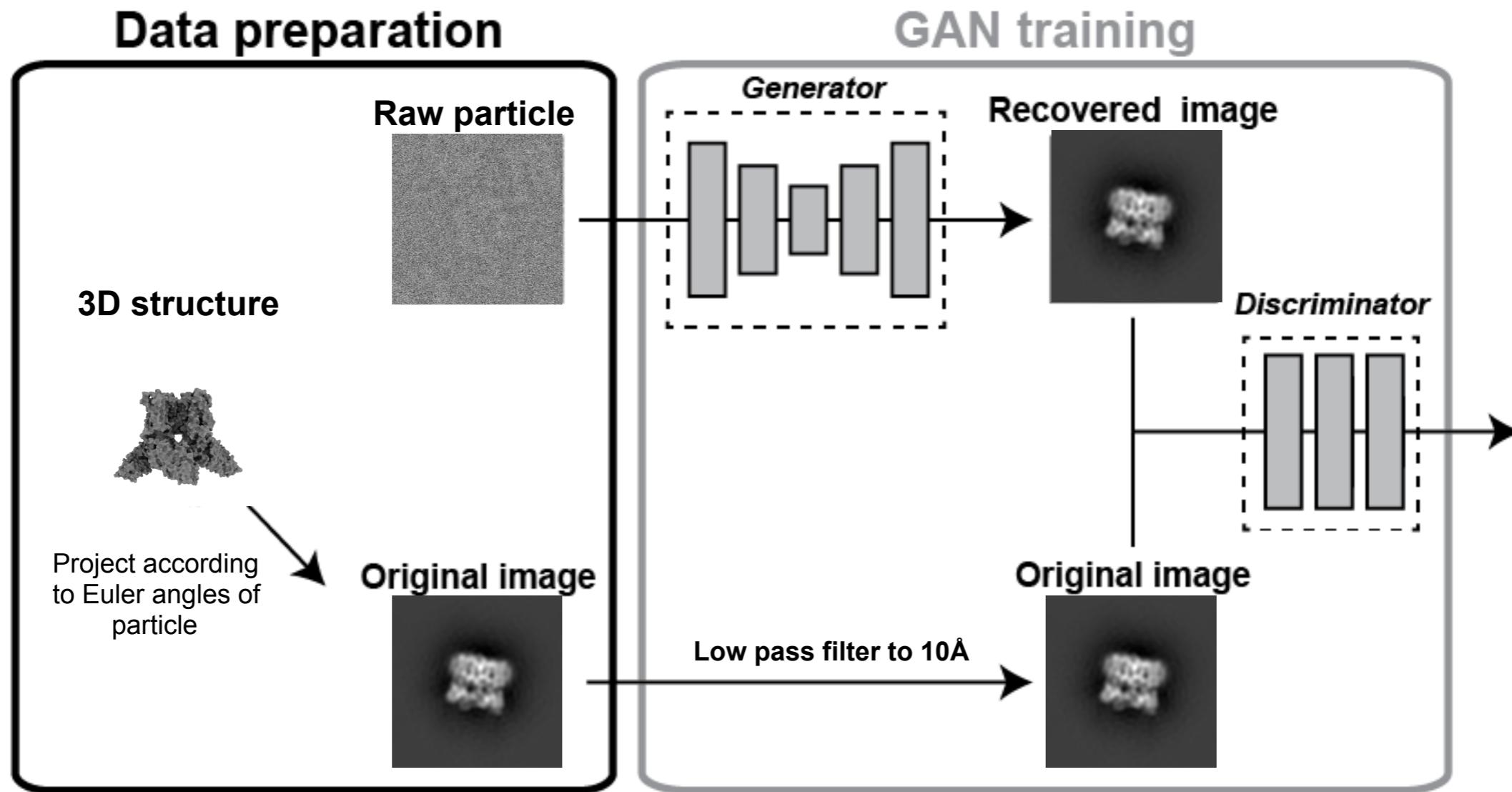
Testing notes:

Different particles than training set

<1 sec/particle to generate with GTX1080Ti GPU



# Test #2: TRPV1 (EMPIAR10005)



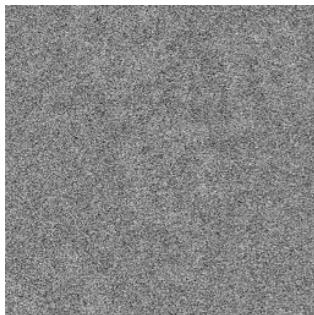
Training notes:

10,000 particle / projection pairs (randomly selected)  
256 x 256 (1.22 Å/pix)  
~10 hours on 1xGTX1080Ti GPU

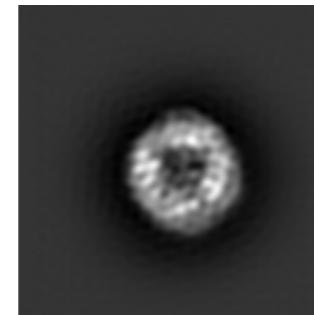


# Test #2: TRPV1 (EMPIAR10005)

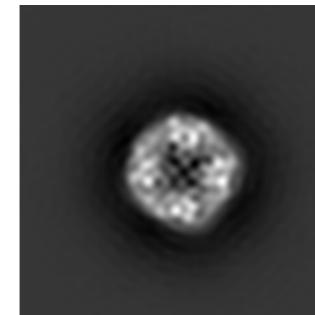
Input particle



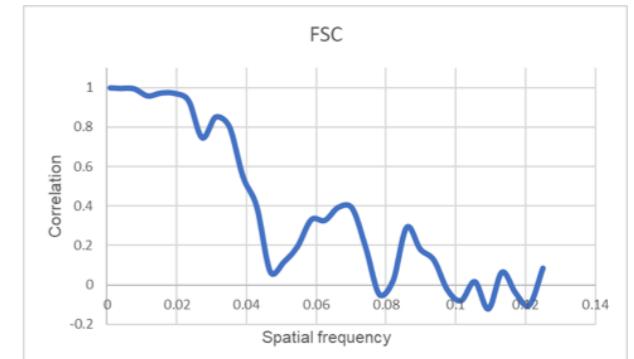
GAN output



Ground truth

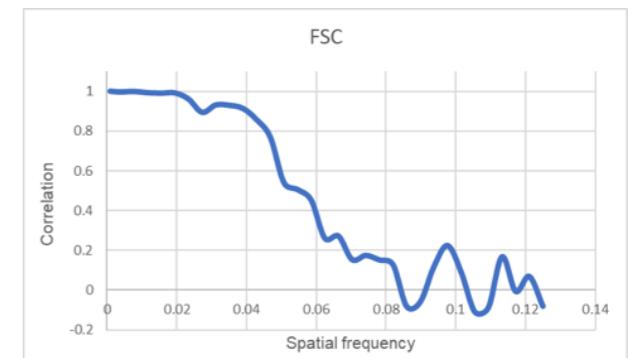
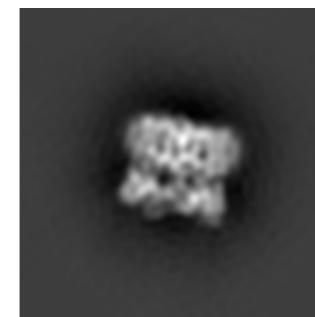
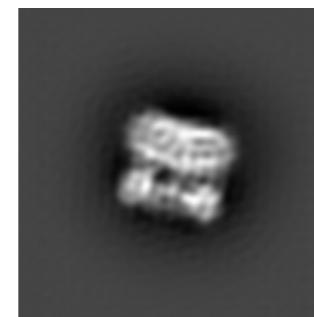
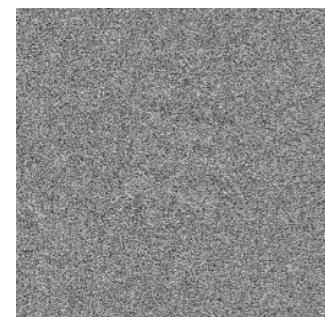


FSC

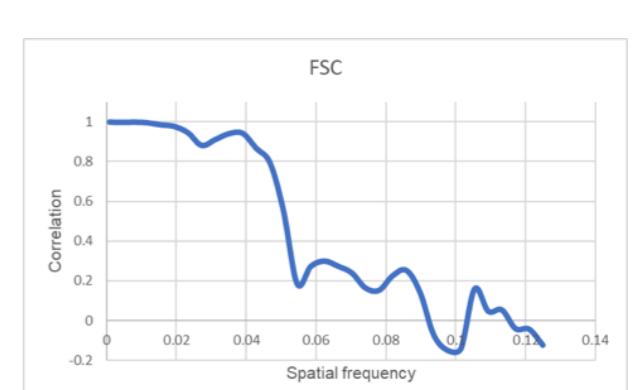
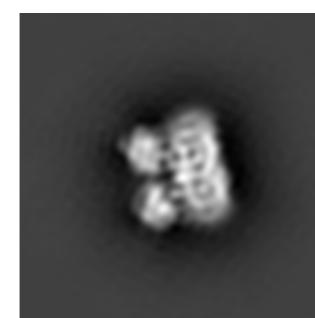
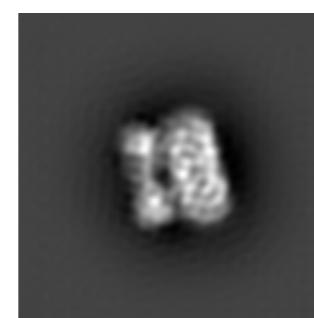
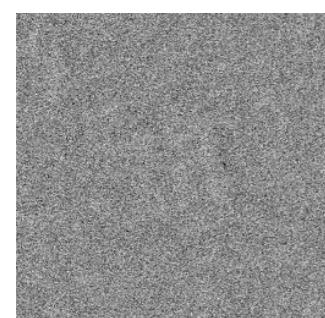


**FSC=0.5**

**31 Å**



**23 Å**

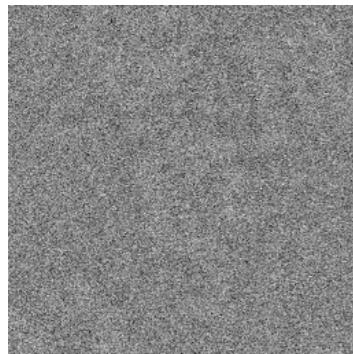


**23 Å**

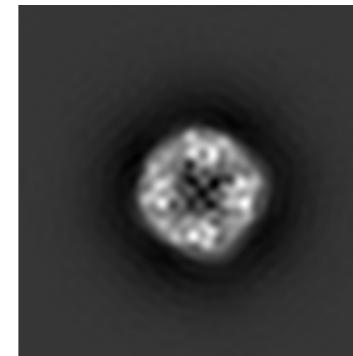


## Test #2: TRPV1 (EMPIAR10005)

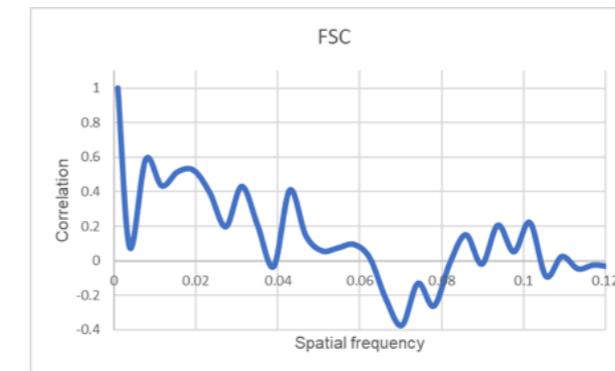
Input particle



Ground truth

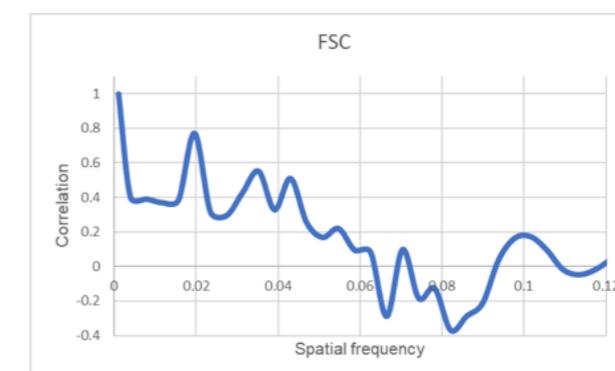
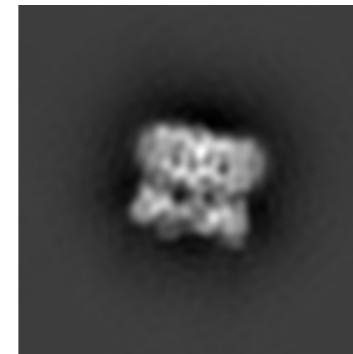
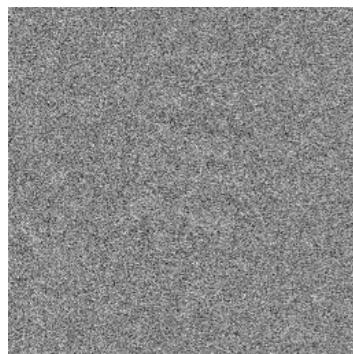


FSC b/w raw particle  
& ground truth

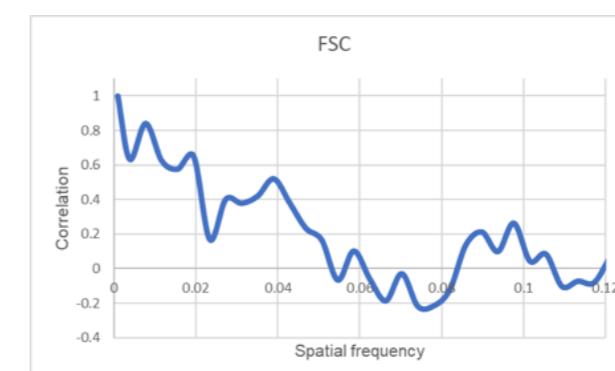
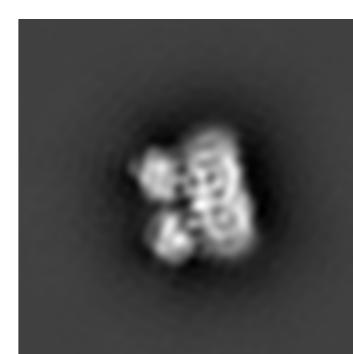
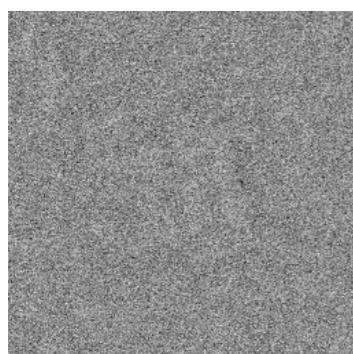


**FSC=0.5**

**> 200 Å**



**> 200 Å**

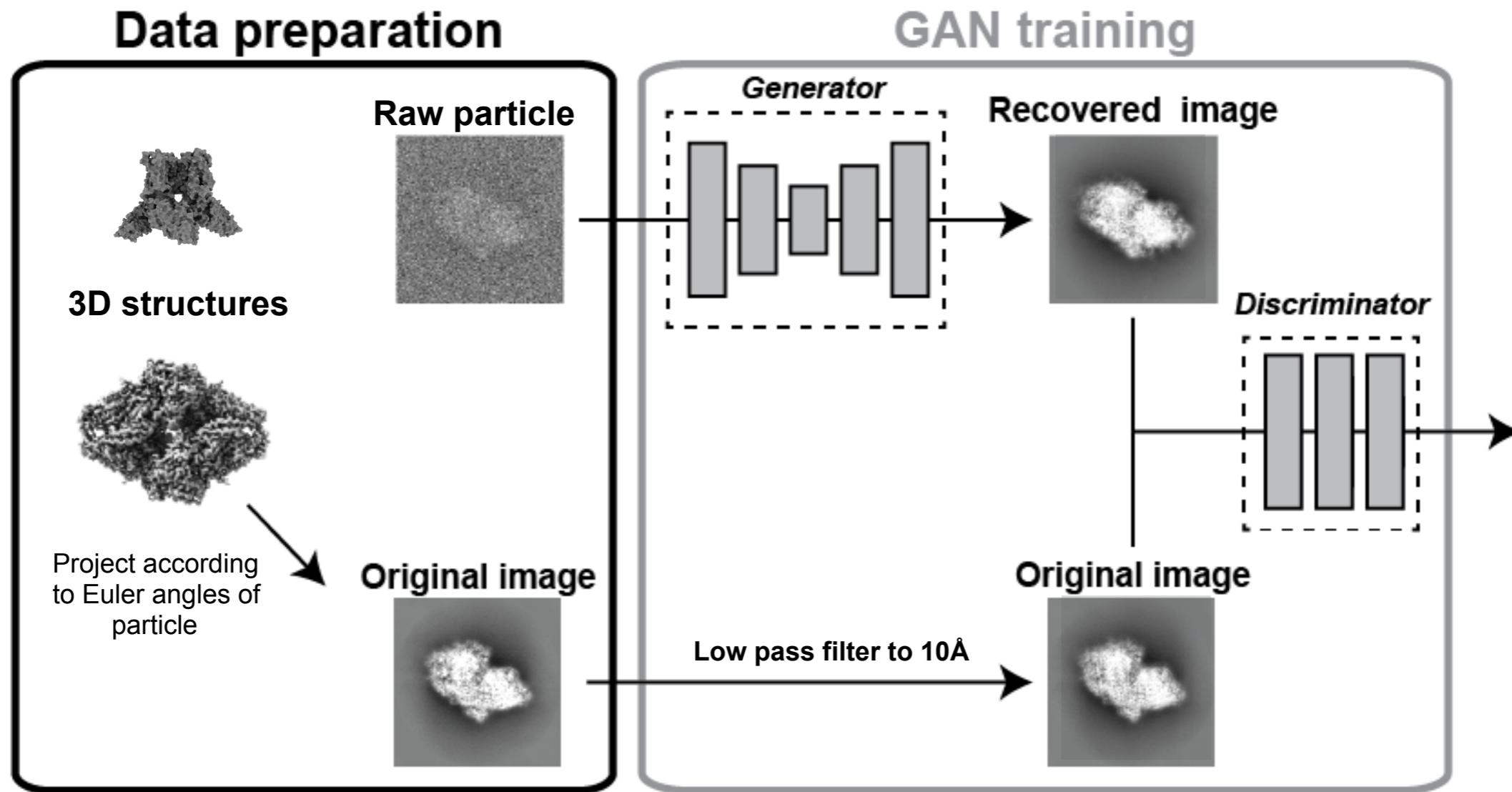


**61 Å**

*Determining FSC curve between raw particles and ground truth*



# Test #3: Training on Beta-Gal & TRPV1 combined



*Training notes:*

20,000 particle / projection pairs (randomly selected)

(10,000 from each TRPV1 & Beta-Gal)

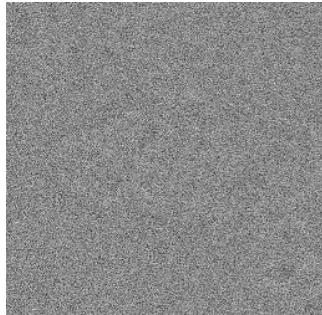
256 x 256 (0.997 Å/pix)

~20 hours on 1xGTX1080Ti GPU

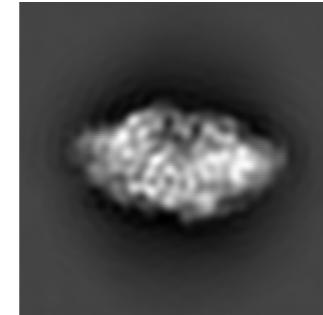


# Test #3: Training on Beta-Gal & TRPV1 combined

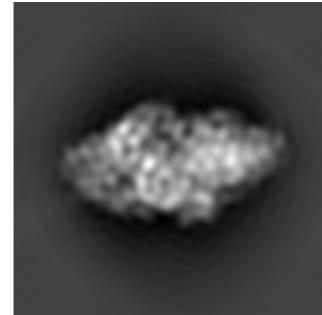
Input particle



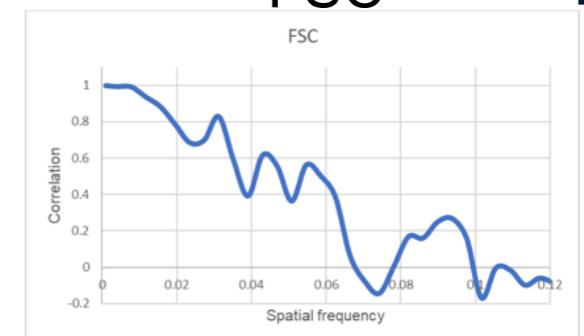
GAN output



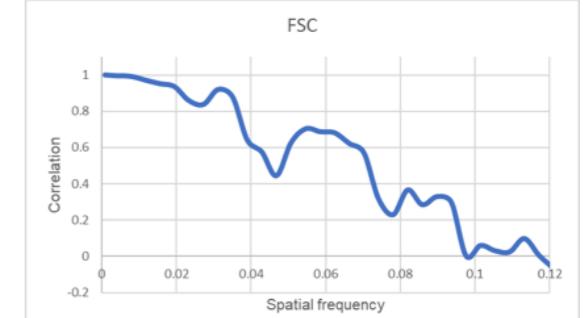
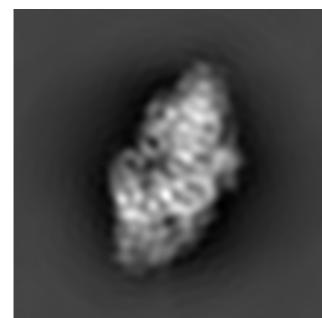
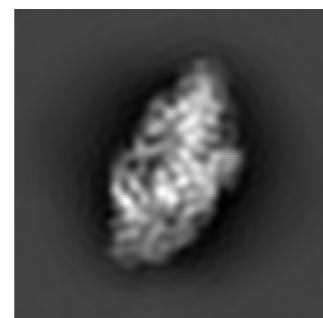
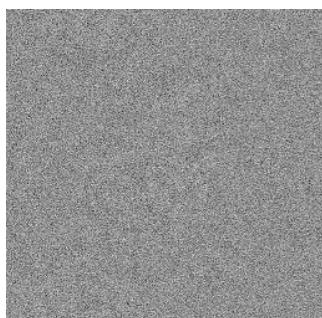
Ground truth



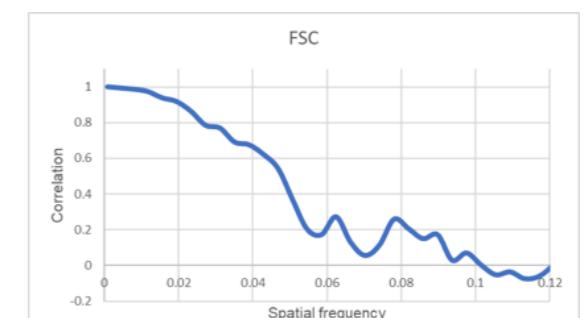
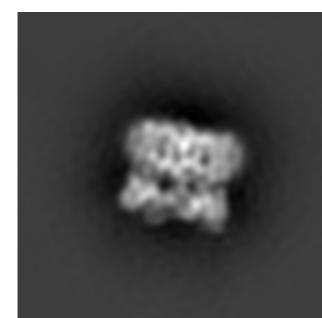
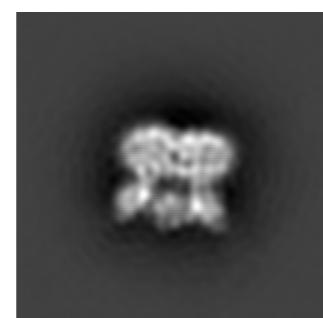
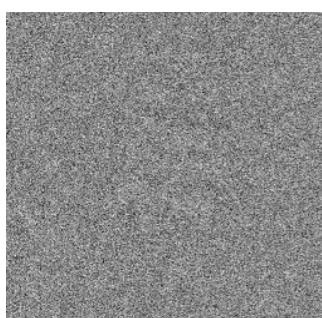
FSC



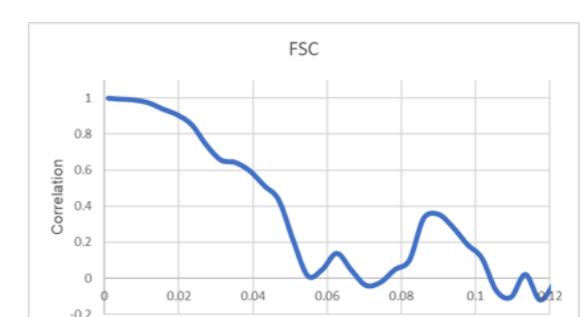
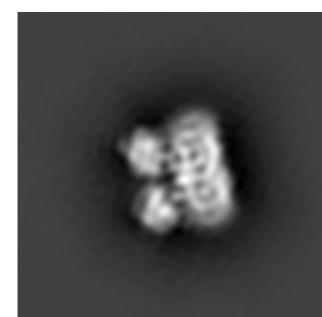
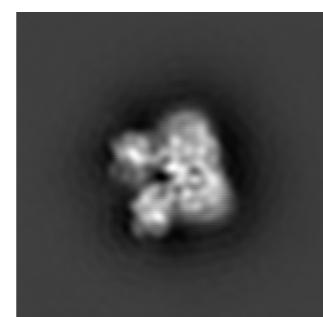
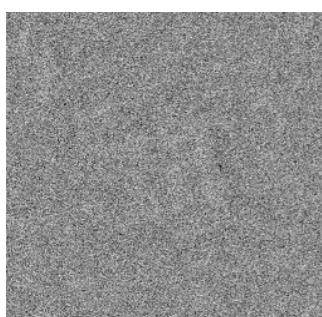
28 Å



22 Å



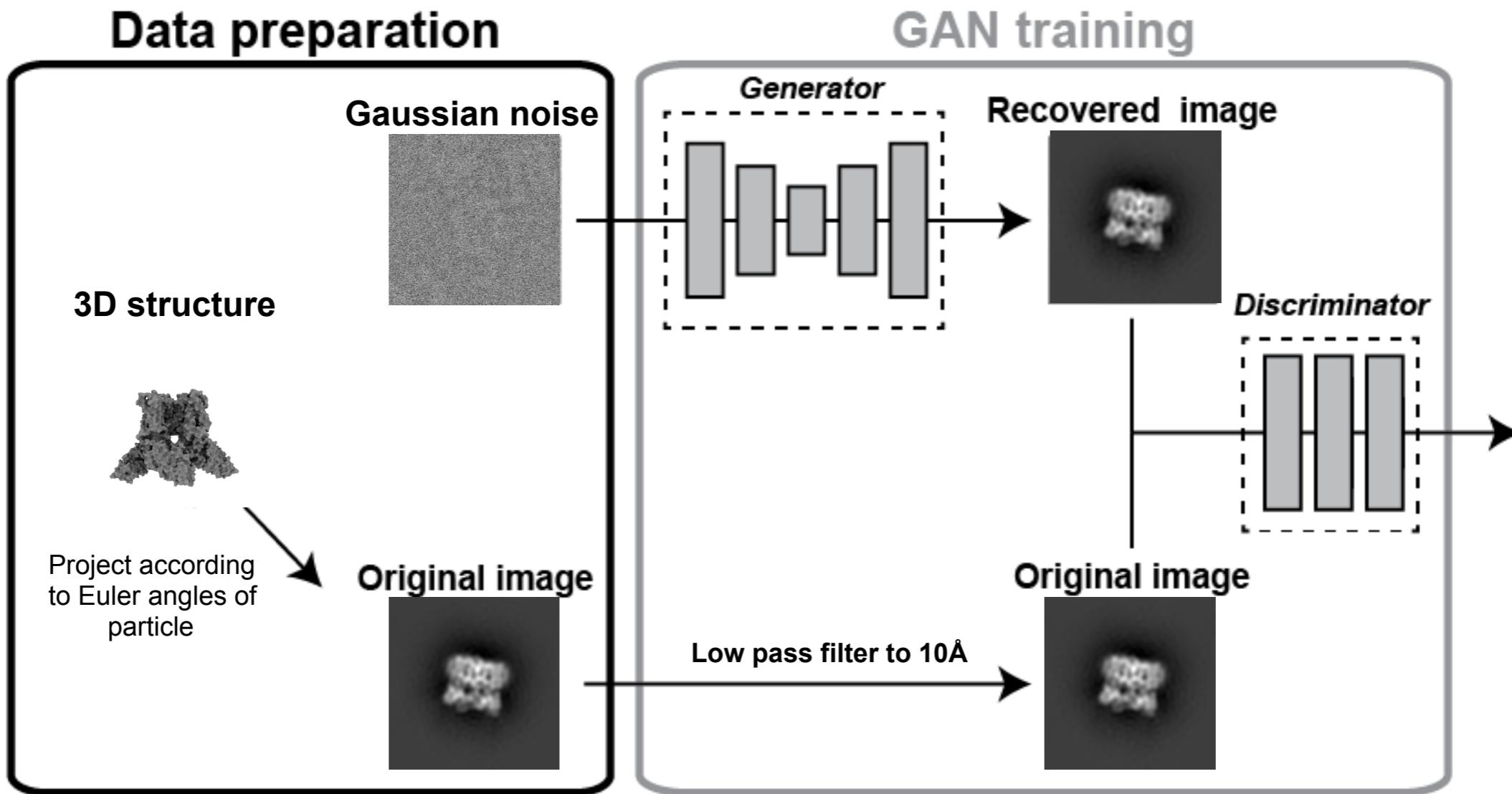
25 Å



27 Å



# Test #4: Training on noise with TRPV1



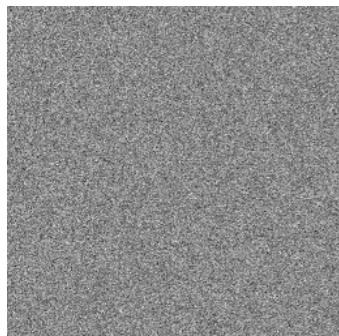
Training notes:

1,000 particle / projection pairs (randomly selected)  
256 x 256 (1.22 Å/pix)  
~1 hours on 1xGTX1080Ti GPU

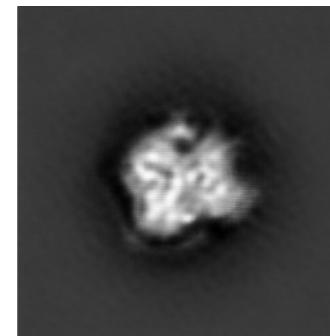


# Test #4: Training on noise with TRPV1

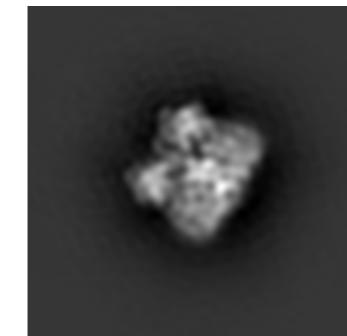
Input particle



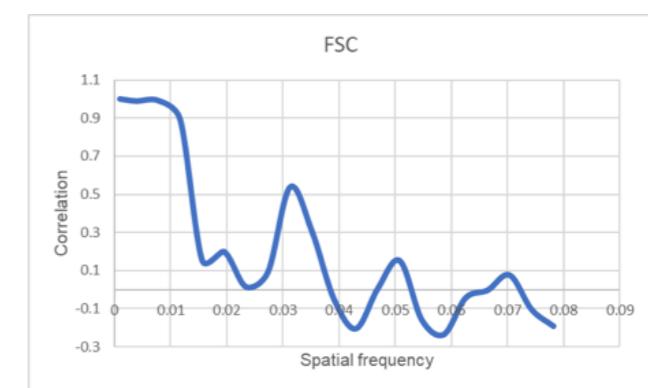
GAN output



Ground truth

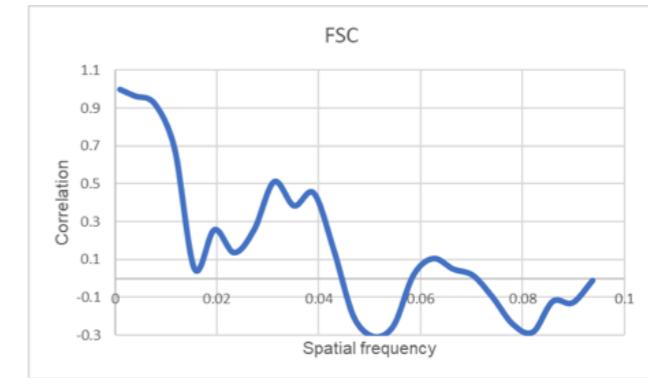
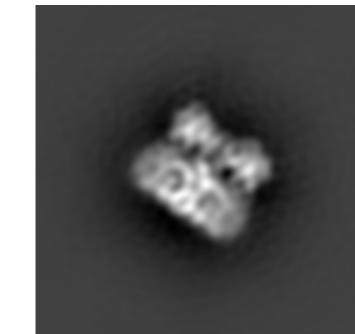
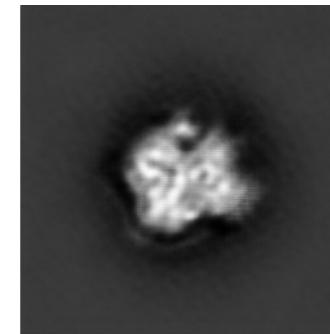
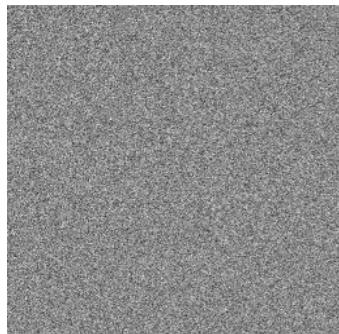


FSC



**FSC=0.5**

**87 Å**

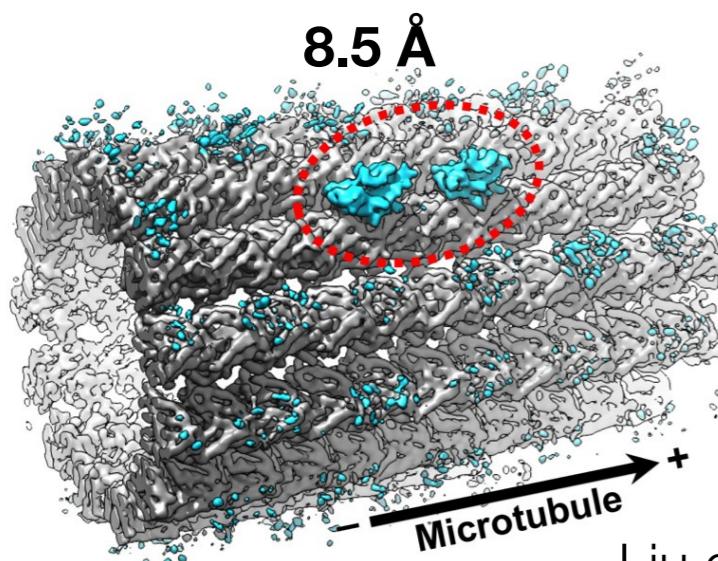
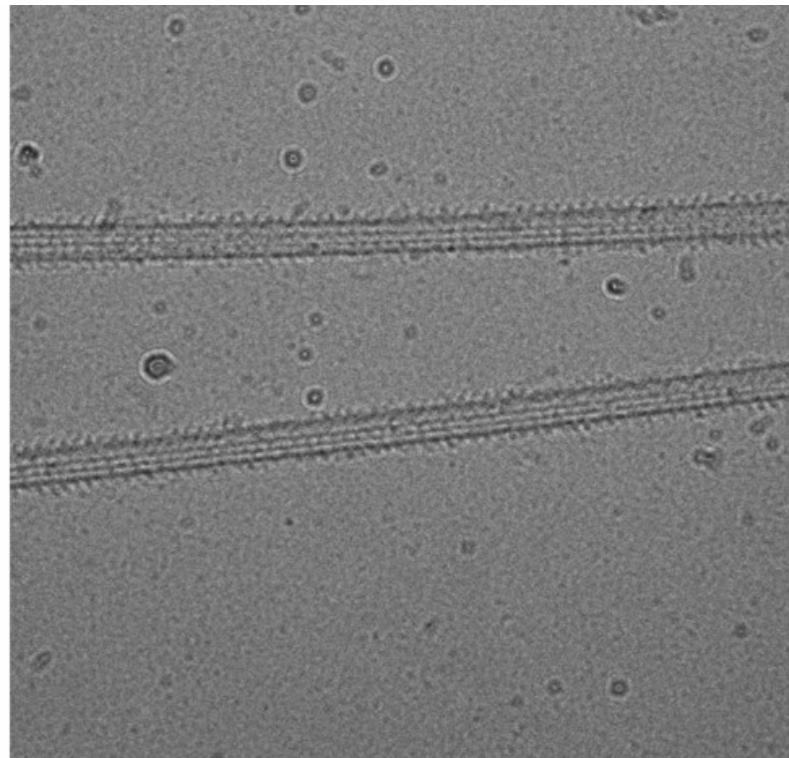


**81 Å**

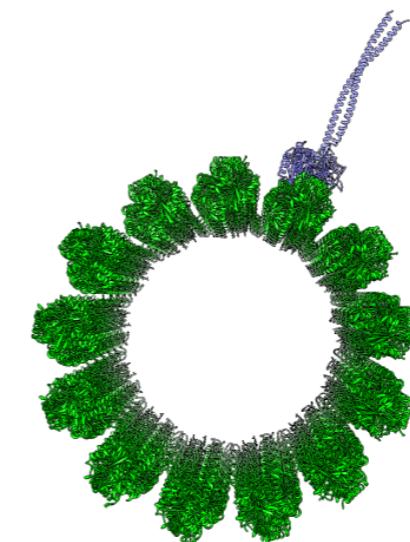
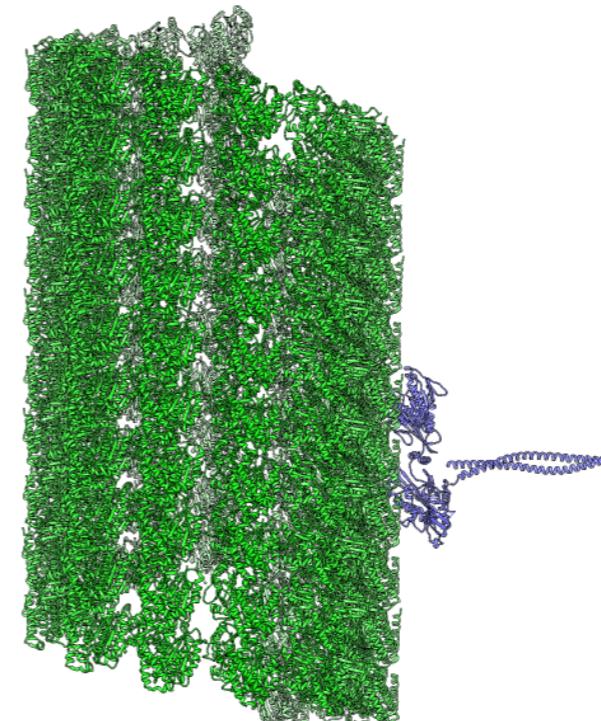


# Example #1: Small protein bound to larger complex

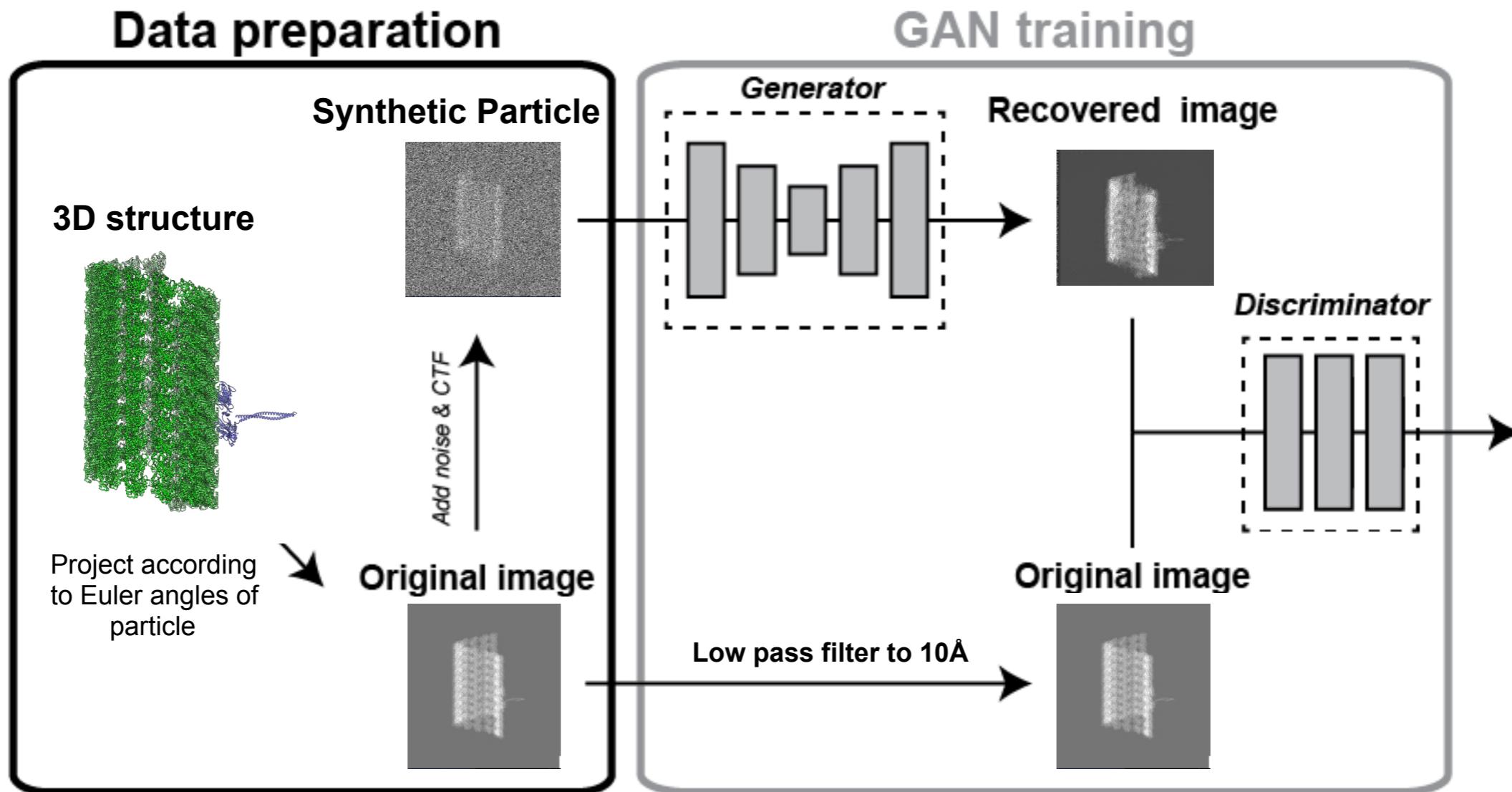
Kinesin bound to microtubule



Liu et al. 2017



# Example #1: Small protein bound to larger complex



*Training notes:*

## Synthetic data

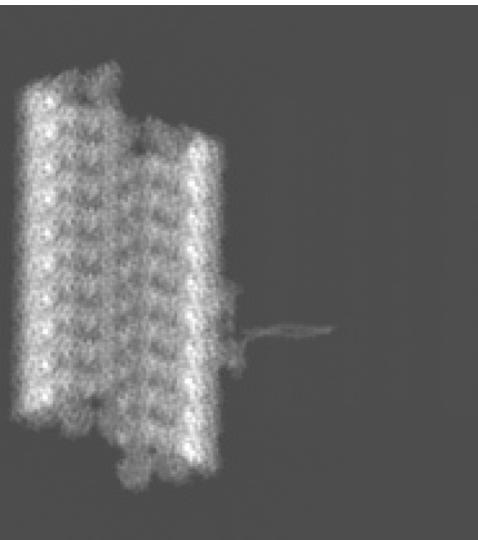
- 10,000 particle / projection pairs (random Euler angles)
- 256 x 256 (4.56 Å/pix)
- ~10 hours on 1xGTX1080Ti GPU



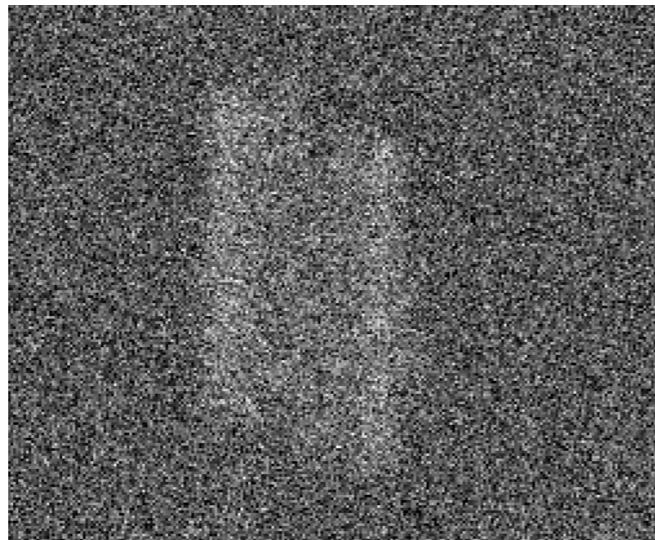
# Example #1: Small protein bound to larger complex

---

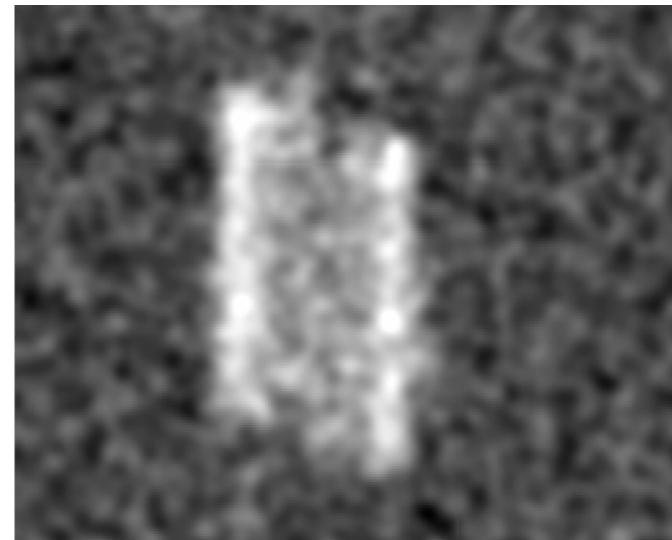
Ground truth



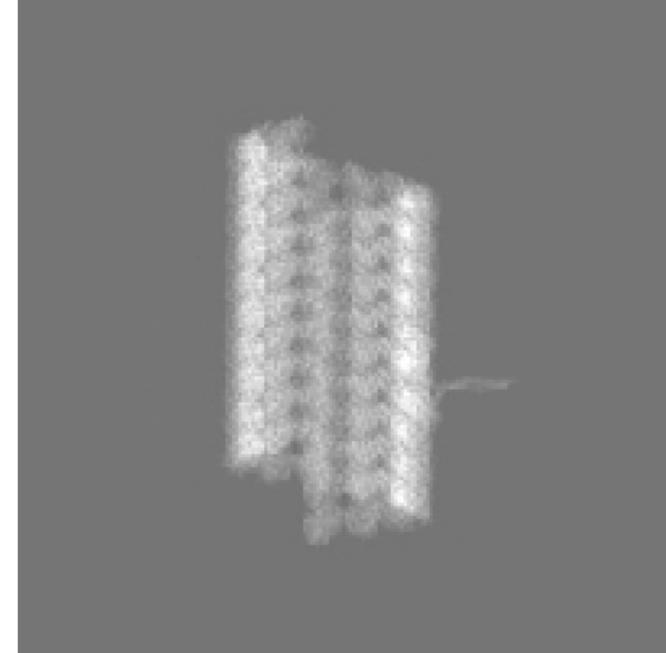
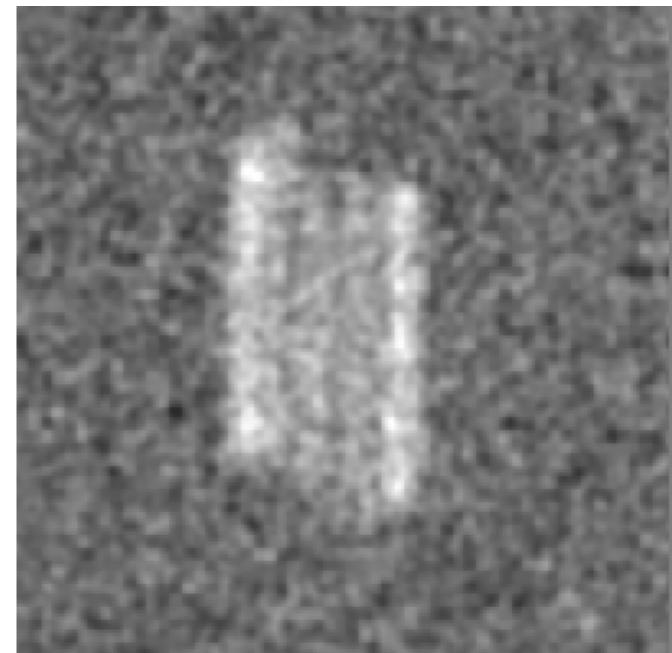
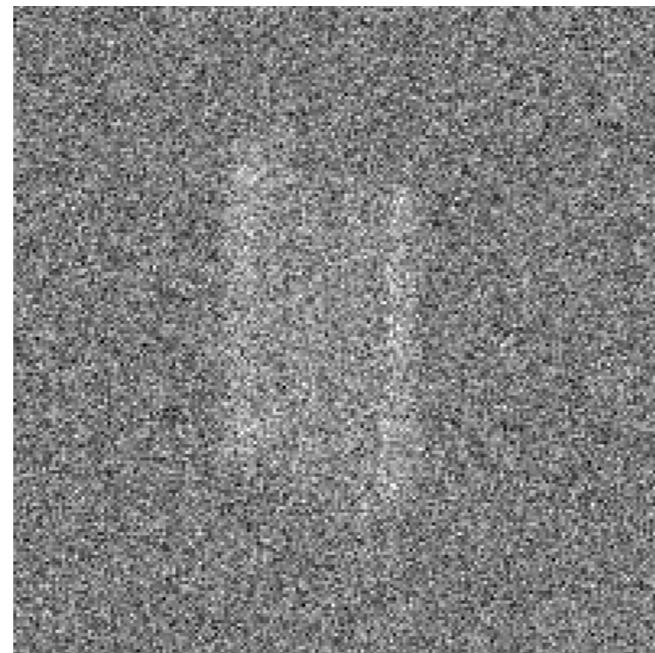
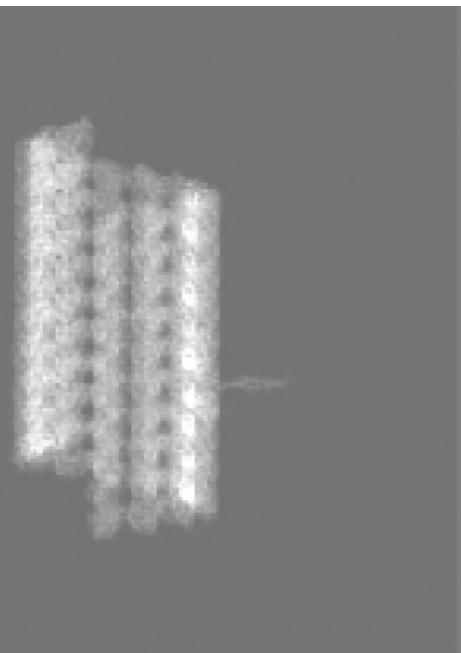
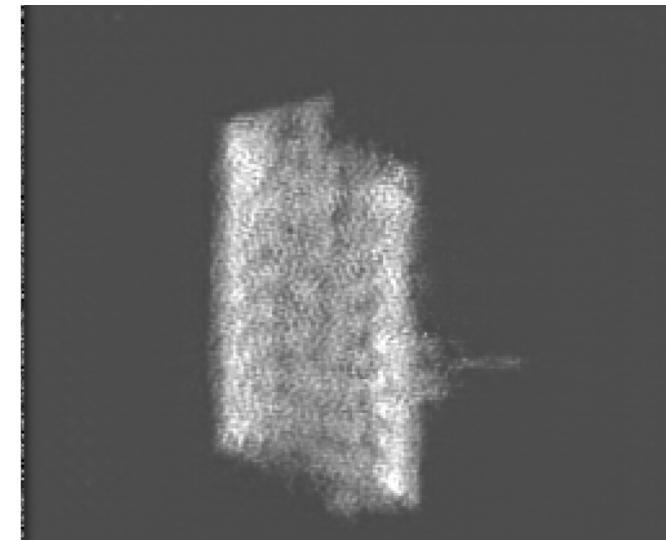
Simulated particle  
(sigma=15)



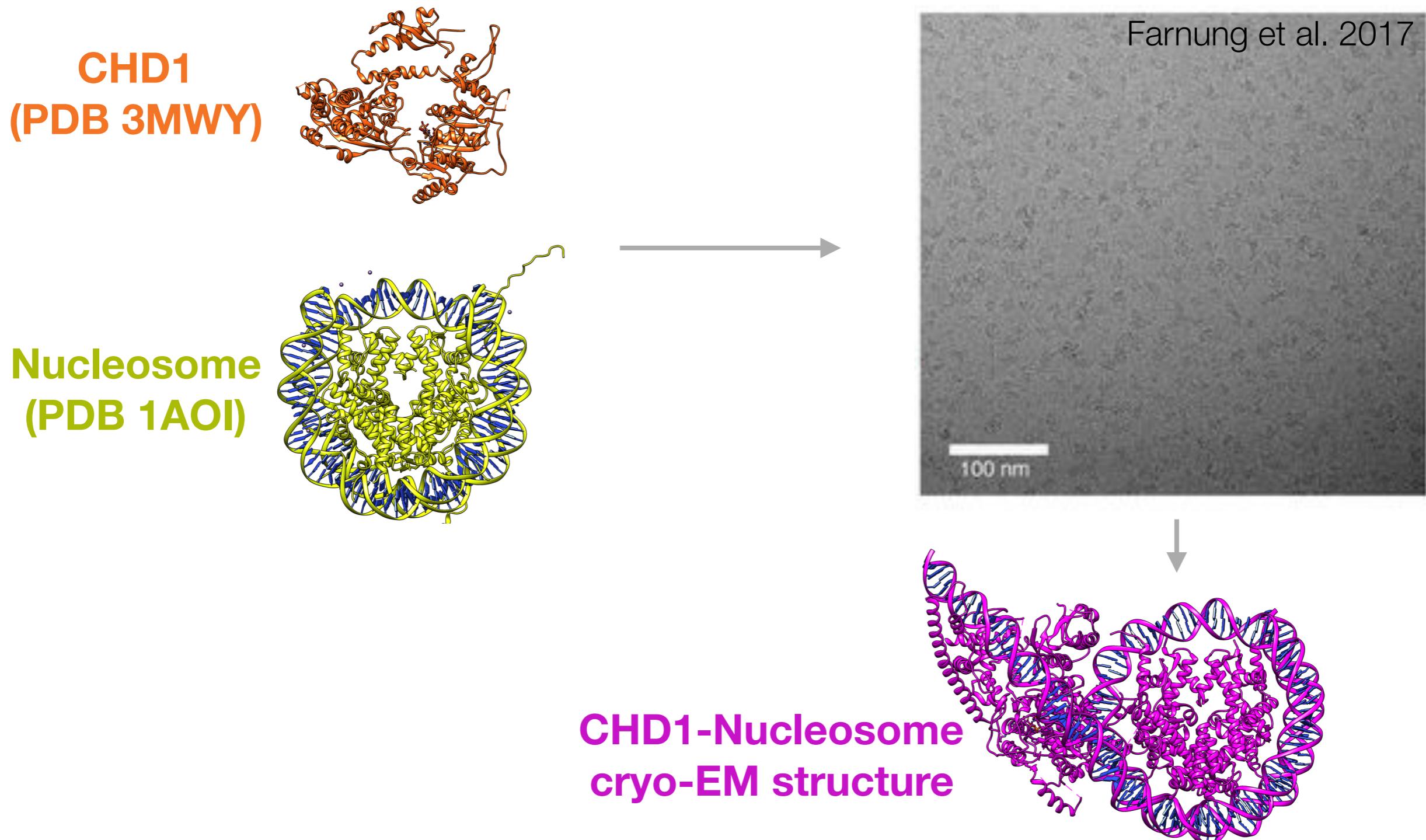
Filtered particle at 40 Å



GAN output

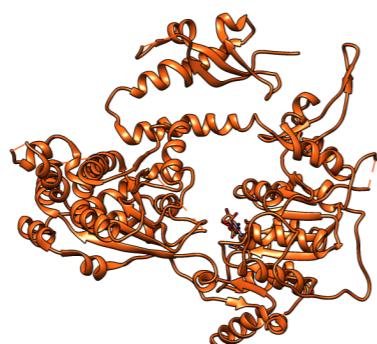


## Example #2: Known structures in complex together

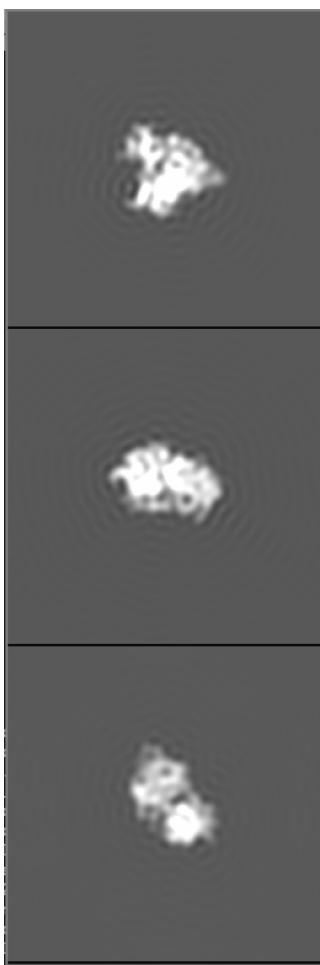


## Example #2: Known structures in complex together

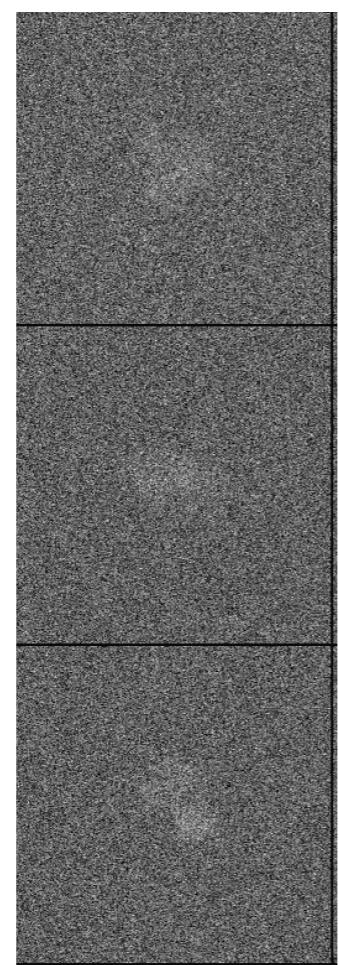
**CHD1**  
(PDB 3MWY)



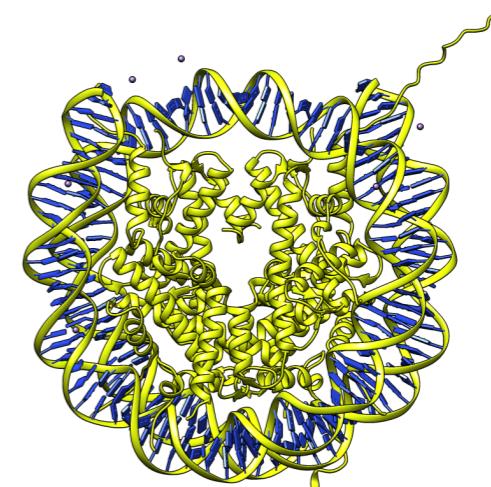
Ground truth



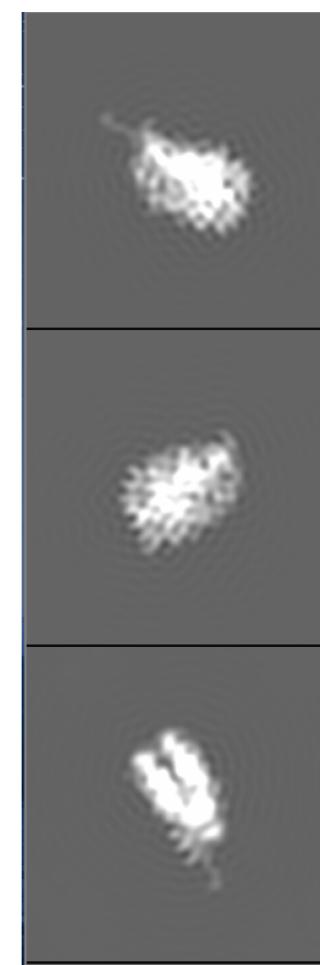
Simulated particle  
(sigma=30)



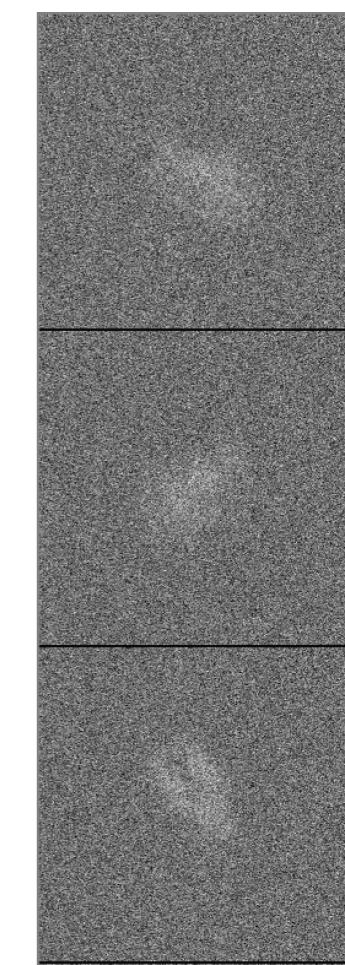
**Nucleosome**  
(PDB 1AOI)



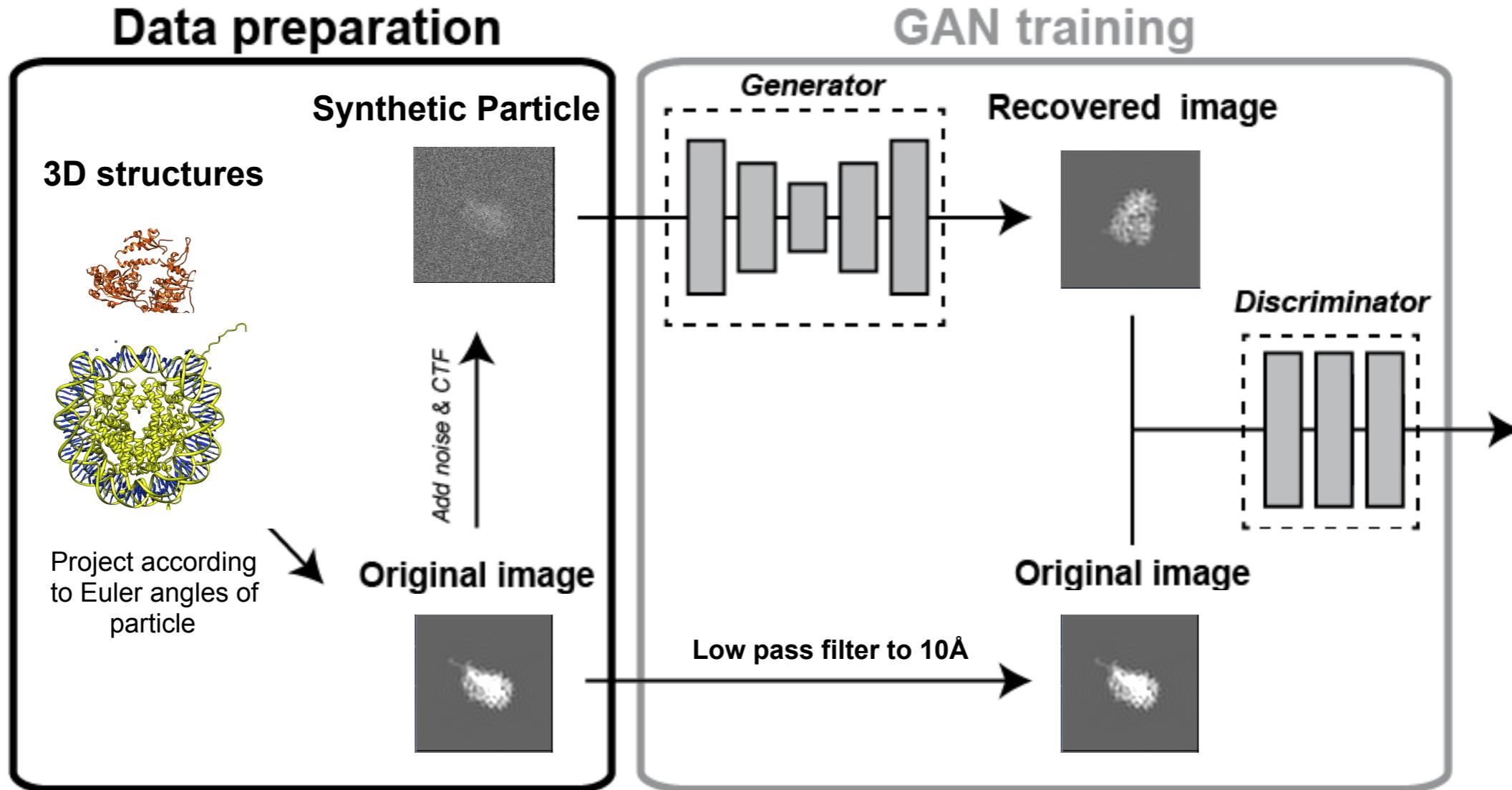
Ground truth



Simulated particle  
(sigma=30)



## Example #2: Known structures in complex together



*Training notes:*

15,000 particle / projection pairs (randomly selected)

(10,000 from nucleosome, 5,000 from CHD1)

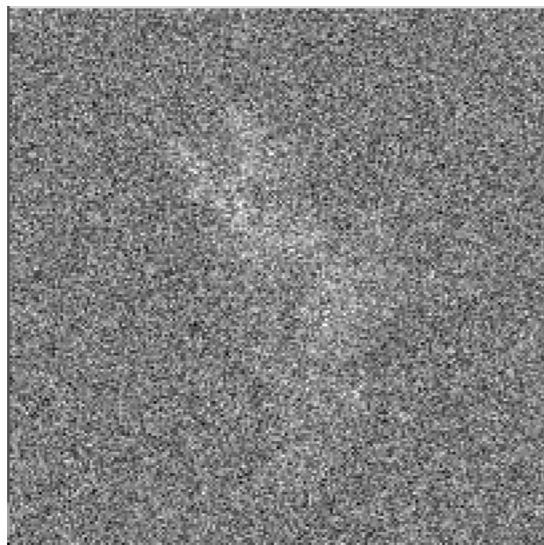
256 x 256 (1 Å/pix)

~20 hours on 1xGTX1080Ti GPU



## Example #2: Known structures in complex together

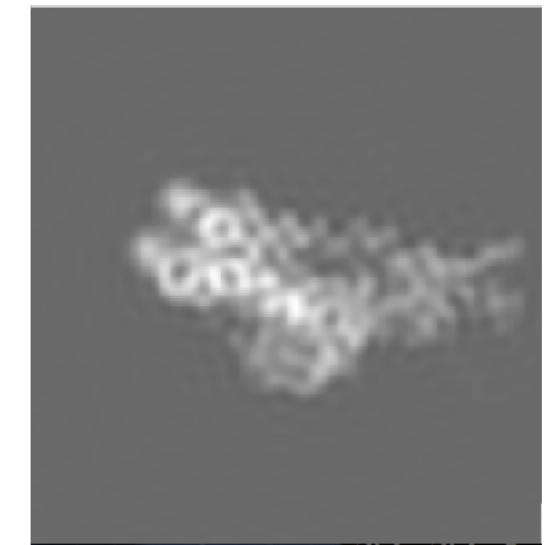
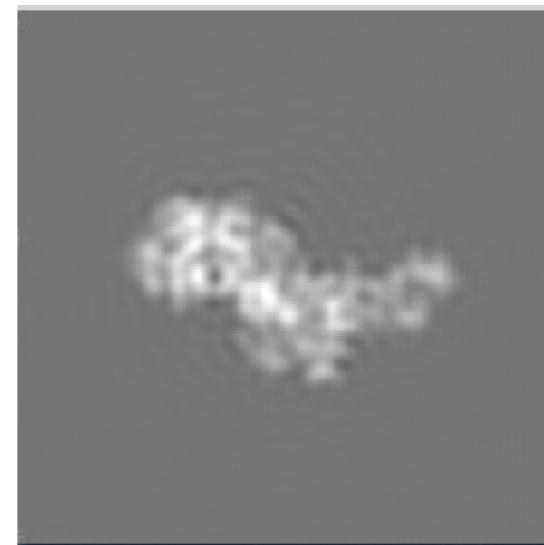
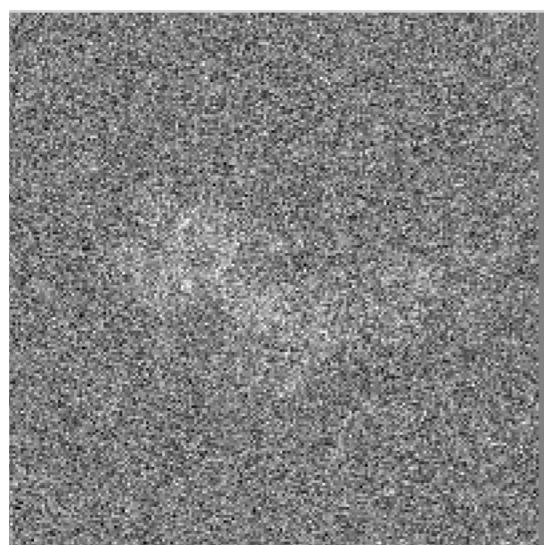
Simulated particle  
CHD1-Nuc  
(sigma=30)



GAN output

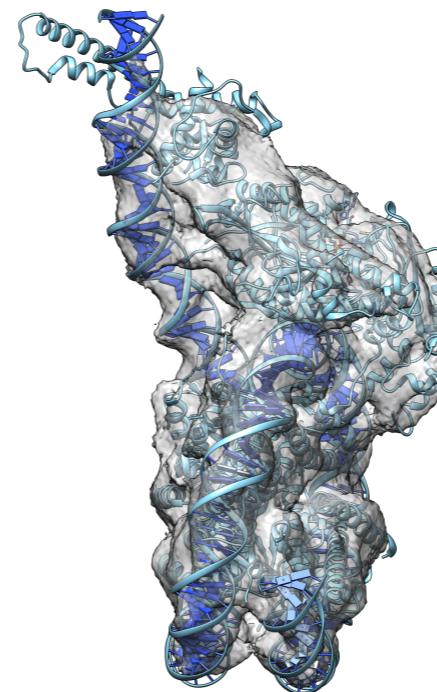
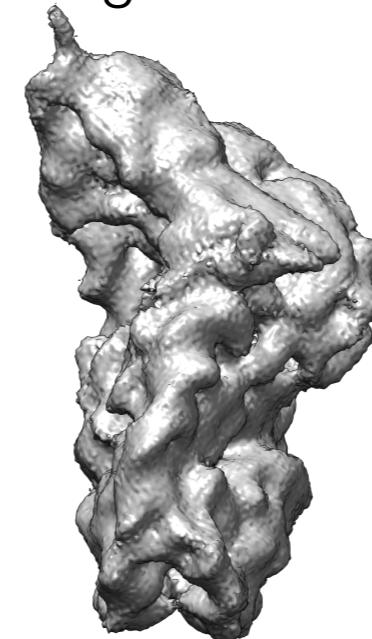
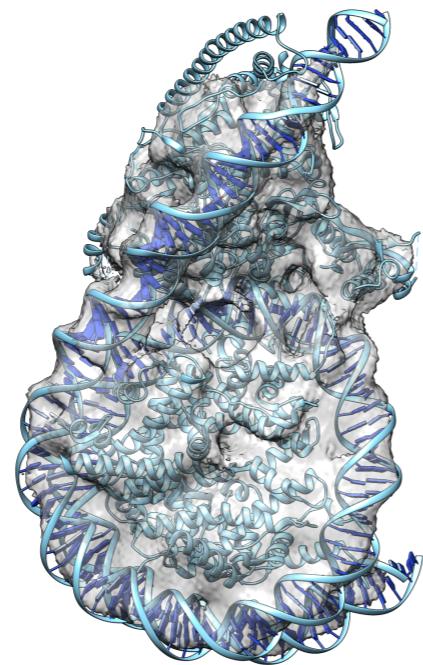
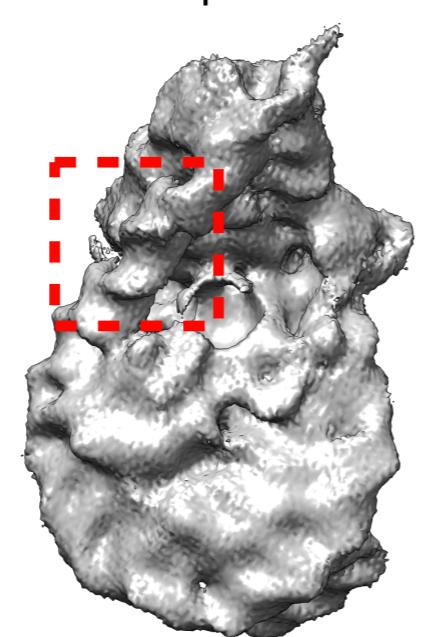
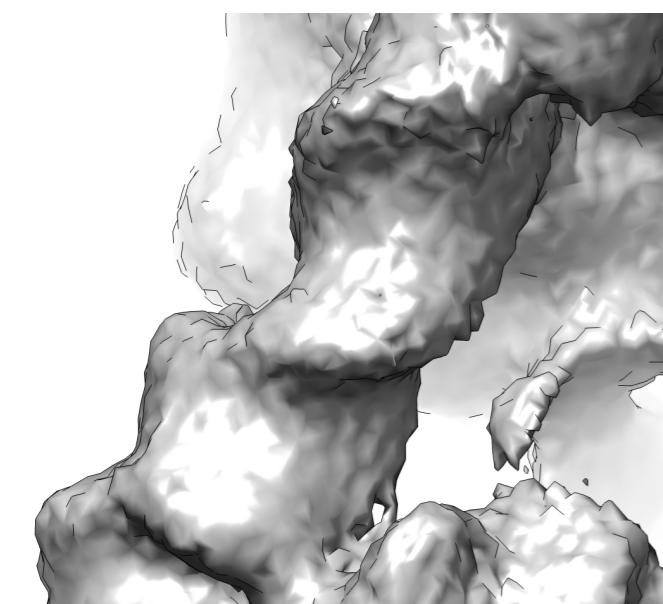


Ground truth

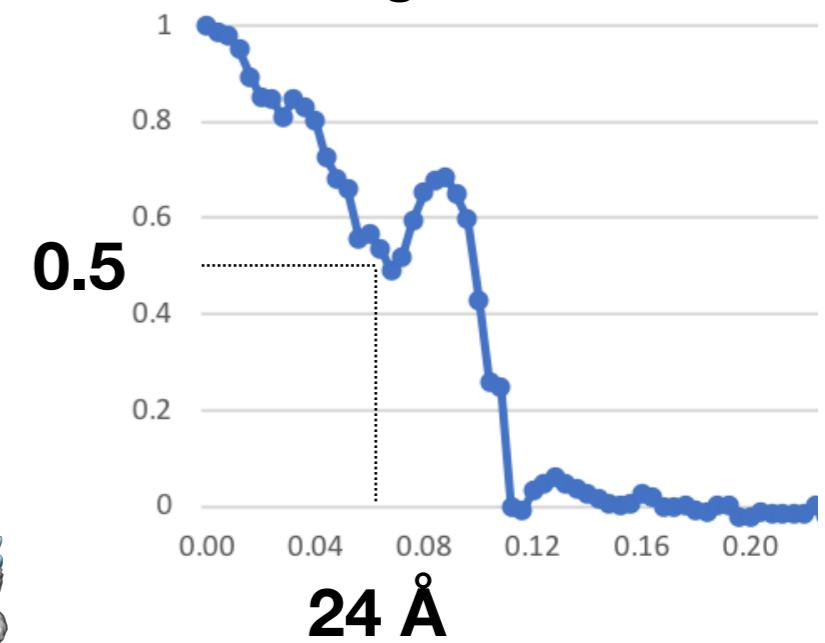


## Example #2: Known structures in complex together

Back projection of GAN particles based on original Euler angles of ground truth



**FSC between GAN 3D  
and ground truth**



# Conclusions

---

- cGANs are able to recover information from raw cryo-EM particles
- GAN outputs could aid structure determination by providing initial Euler angle assignment
- Potentially able to provide integrity information on co-complexes where individual components have known structures

## Next steps

- Continue synthetic data testing of kinesin microtubule, then apply to Liu et al. 2017 ‘FINDKIN’ program



# Acknowledgements

---



**University of Michigan**

Zhenyu Tan

Min Su

**ETH Zurich**

Ce Zhang

Hantian Zhang

Kevin Schawinski

