POLITECNICO DI MILANO

School of Industrial and Information Engineering

Master of Science in Computer Science and Engineering

Academic Year 2018 - 2019

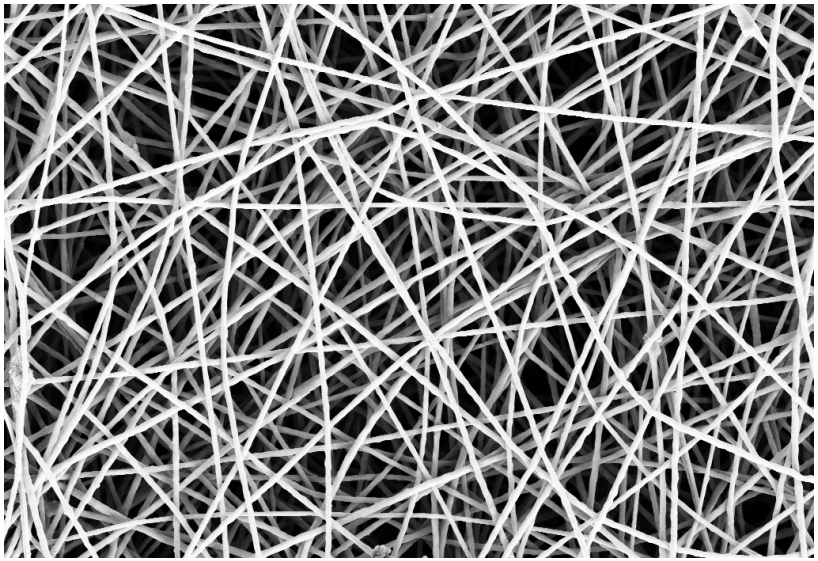
Adversarially Learned Anomaly Detection using Generative Adversarial Networks

Candidate: Şemsi Yiğit Özgümüş
Supervisor : Giacomo Boracchi

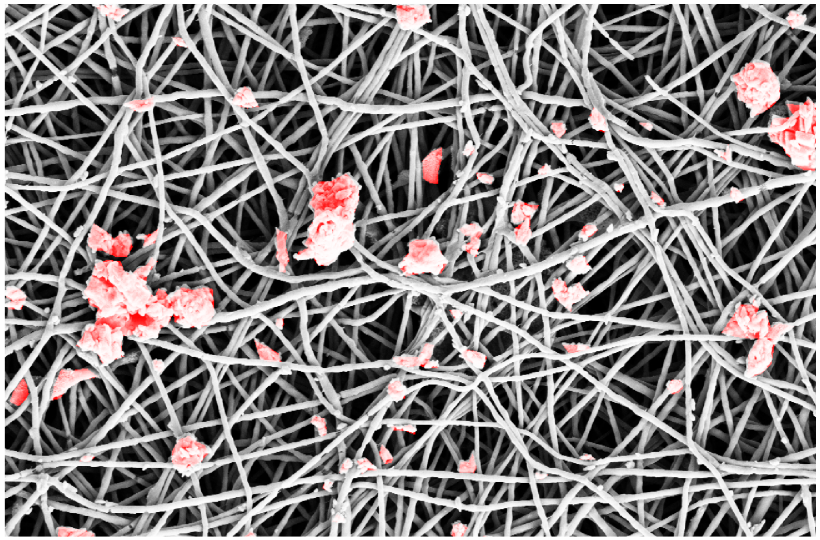
What is Anomaly Detection ?

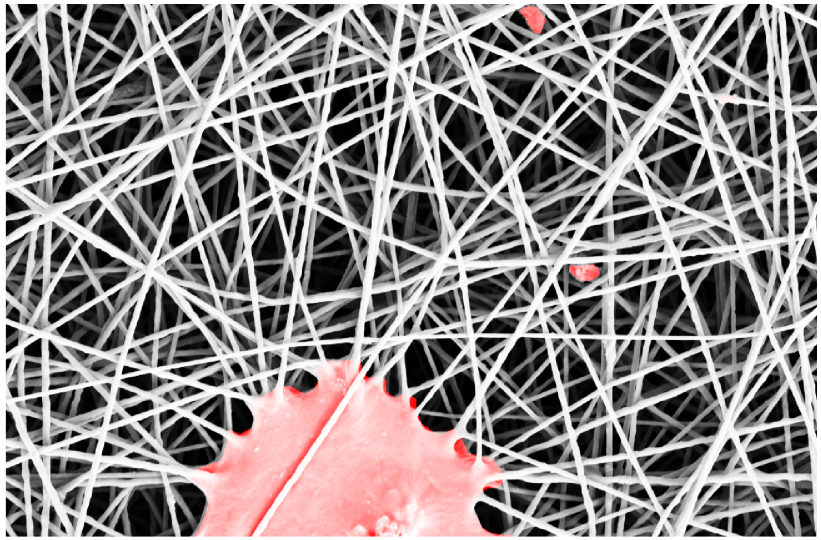
Test

Problem focus of This Thesis



- Nanofibrous material production monitoring
- Detection of defective regions occurred in the production process due to external factors
- Detection of anomalies are required to monitor quality of batches and prevent waste

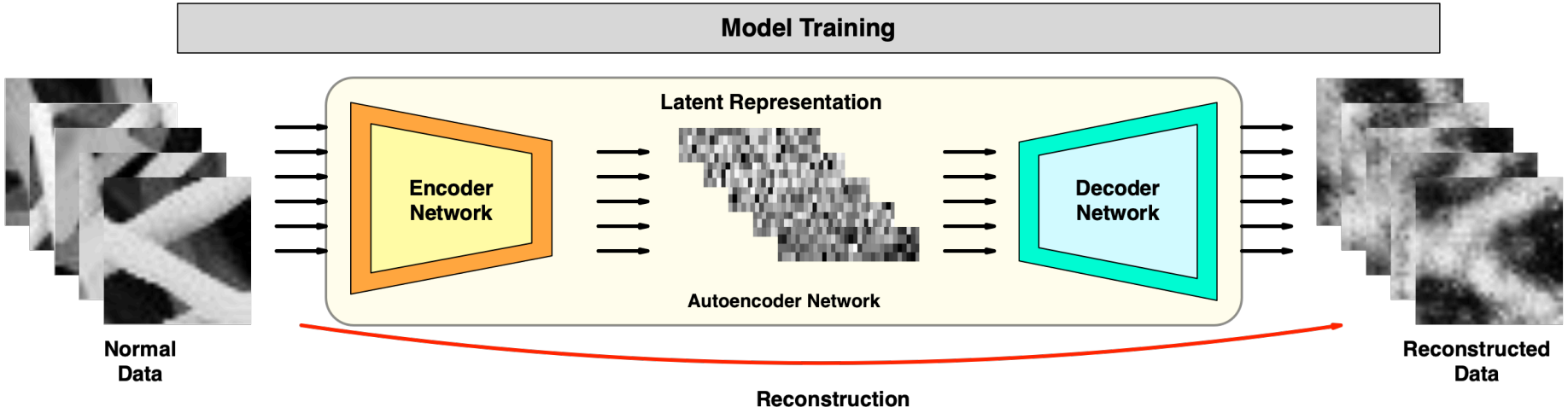




Reconstruction Based Anomaly Detection

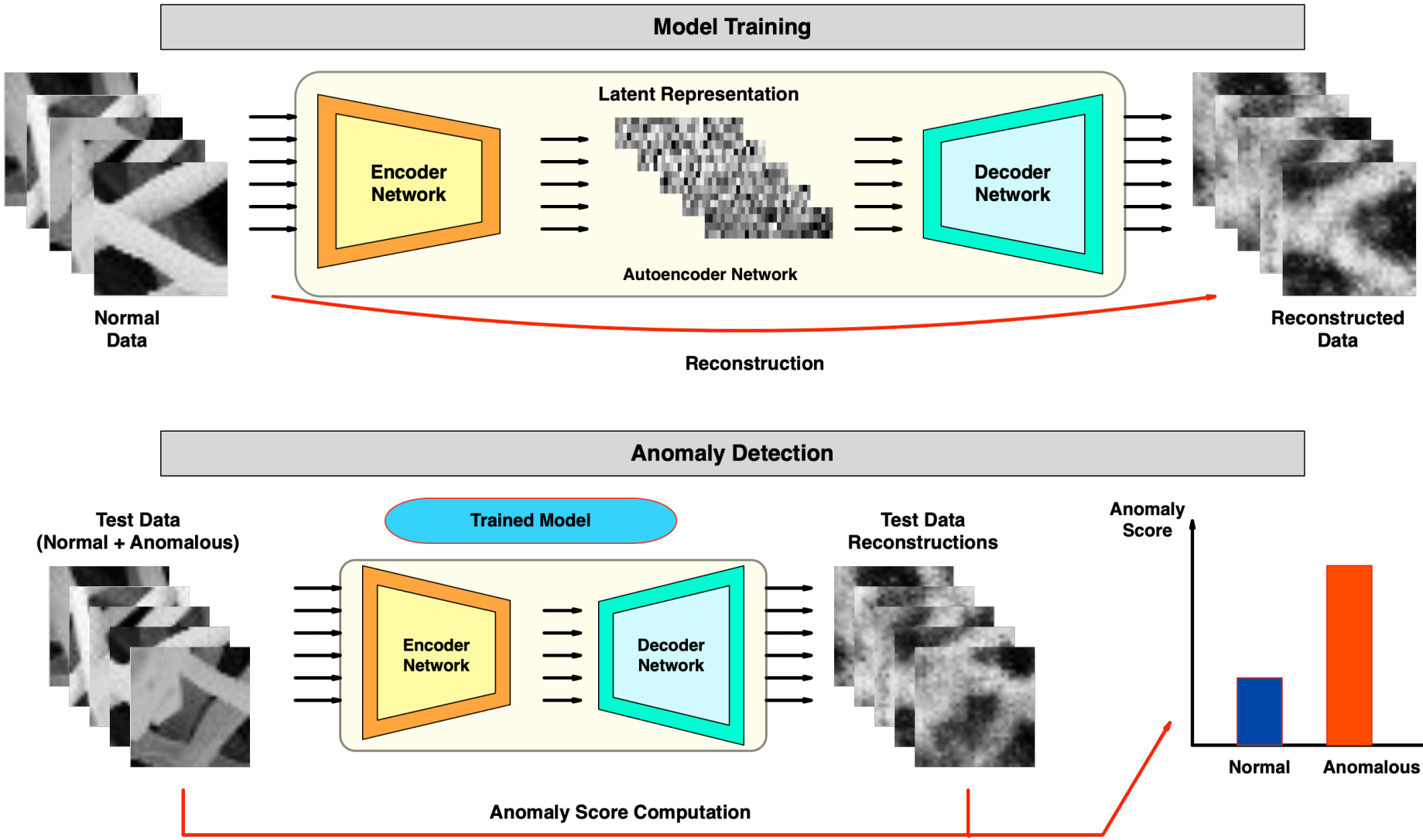
- Autoencoder networks
- Imbalance between normal and anomalous sample size
- Focusing on modelling **notion of normality**

Reconstruction Based Anomaly Detection



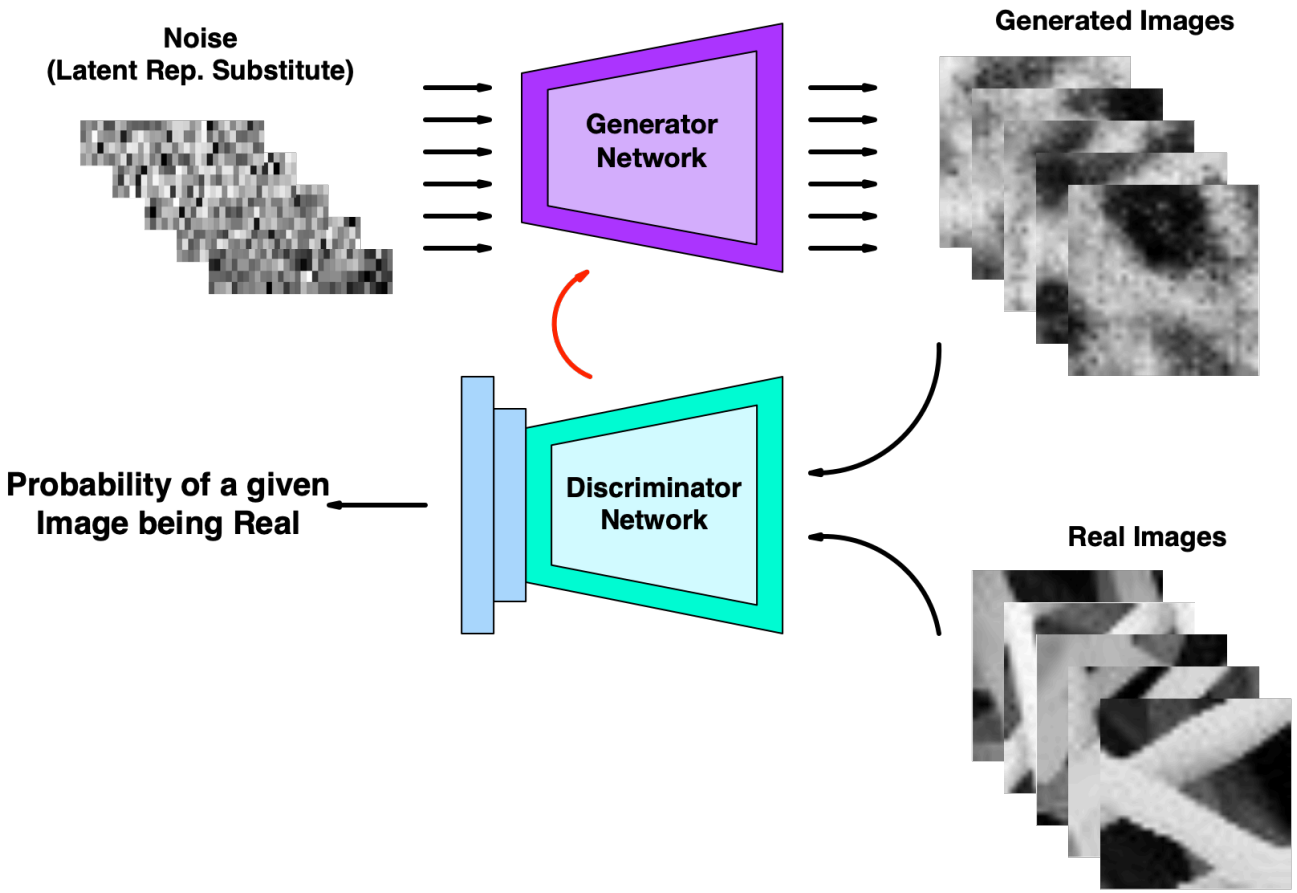
- Mapping from image space to latent dimension space
- Reconstruction from latent representation
- Computing score from the information loss

Reconstruction Based Anomaly Detection



What is Generative Adversarial Network ?

What is Generative Adversarial Network ?



- Network pair trained adversarially using a mini max game
- Generator learns the generate data similar to target data's distribution
- Mapping from latent dimension to higher dimension is learned

What is Generative Adversarial Network ?

- Training objective of GAN:

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

What is Generative Adversarial Network ?

- Training objective of GAN:

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

- Perspective of Discriminator Network:

$$V(G,D) = \mathbb{E}_{\boldsymbol{x}\sim p \text{ data}}(\boldsymbol{x})\underbrace{[\log D(\boldsymbol{x})]}_{\text{Real samples classified as being real}} + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}(\boldsymbol{z})[\log(1 - \underbrace{D(G(\boldsymbol{z}))}_{\text{Generated samples classified as being fake}})]$$

Discriminator wants to maximize this

Discriminator wants to minimize this

What is Generative Adversarial Network ?

- Training objective of GAN:

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

- Perspective of Discriminator Network:

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

Real samples classified as being real

Discriminator wants to maximize this

Generated samples classified as being fake

Discriminator wants to minimize this

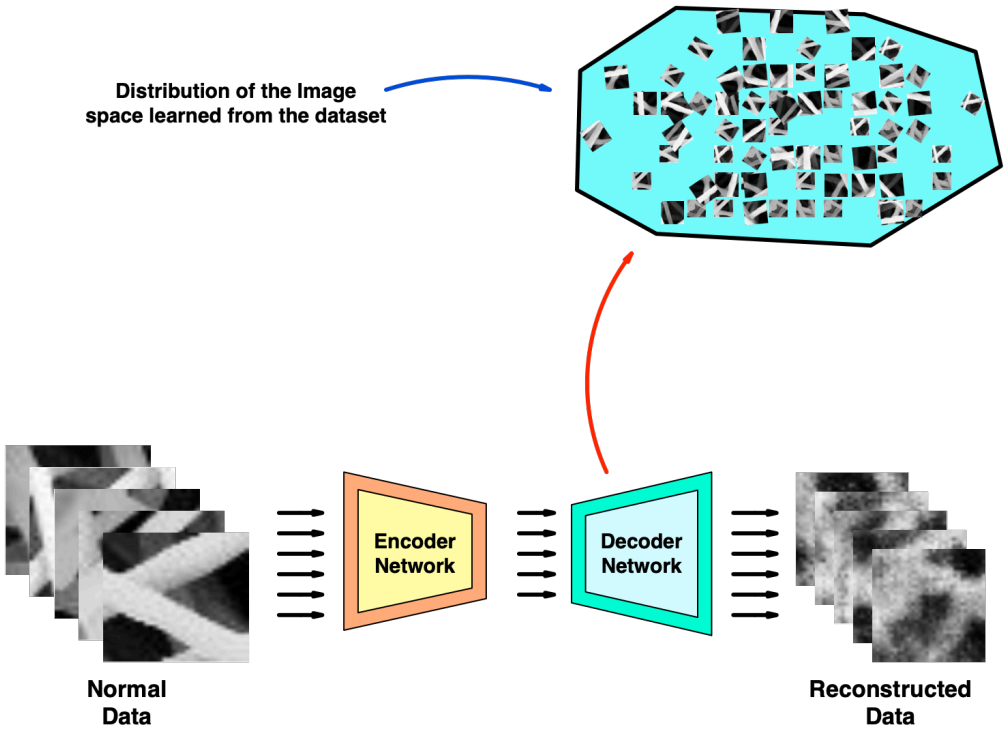
- Perspective of Generator Network:

$$V(G,D) = \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}}(\boldsymbol{z})[\log(1 - \underbrace{D(G(\boldsymbol{z}))})]$$

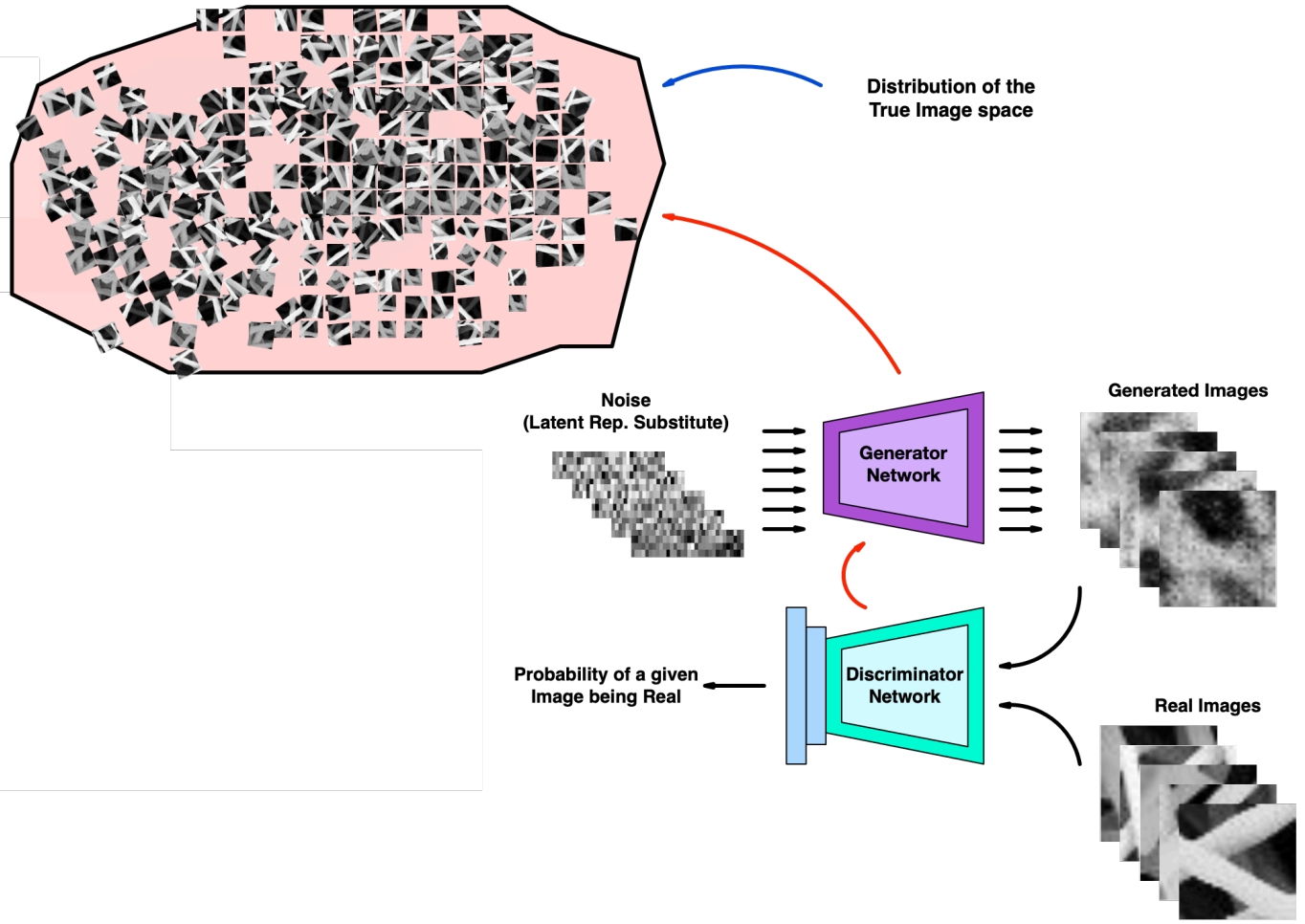
How realistic are the generated samples ?

Generator wants to maximize this

Motivation Behind Using GANs



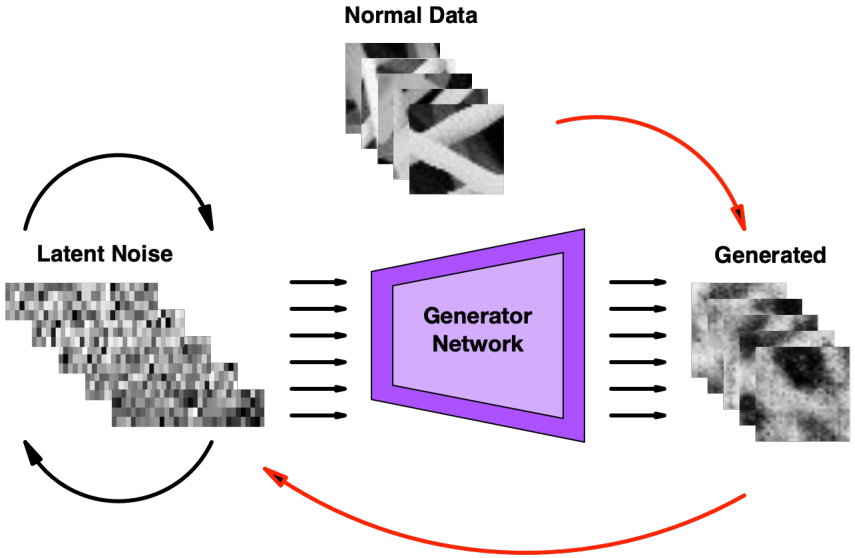
Motivation Behind Using GANs



GAN Based Anomaly Detection Models

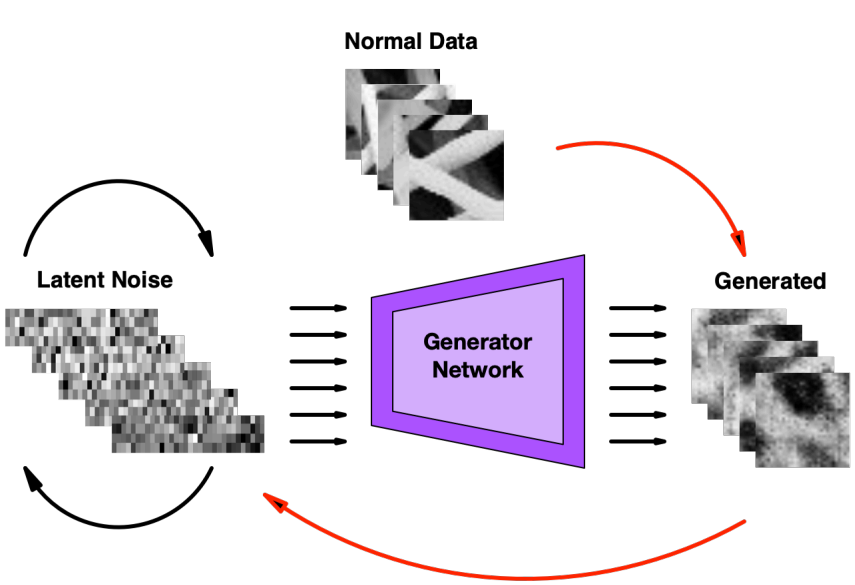
GAN Based Anomaly Detection Models

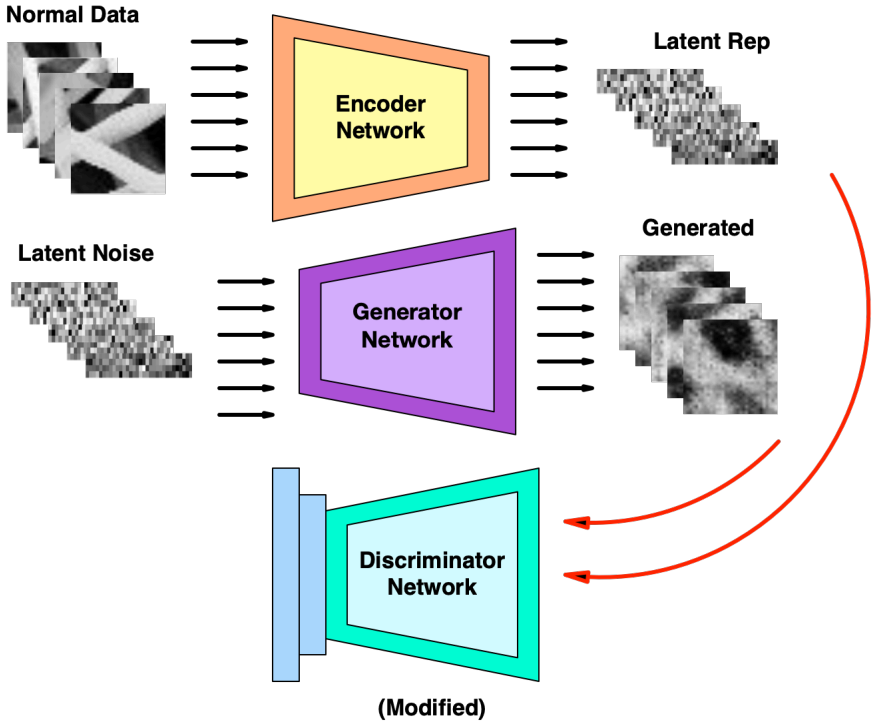
- GAN training + latent representation learning
  - Approximation of latent representation



GAN Based Anomaly Detection Models

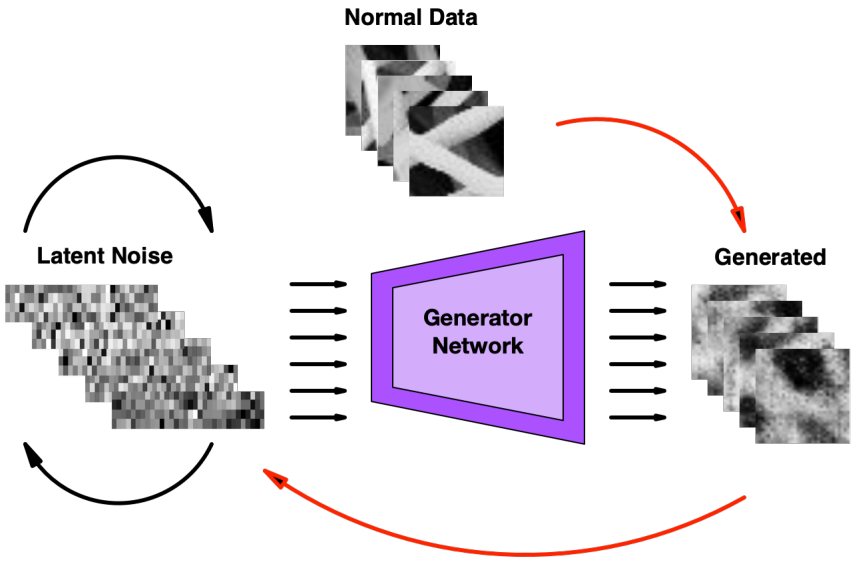
- GAN training + latent representation learning
  - Approximation of latent representation
  - Adversarial feature learning using Encoder network

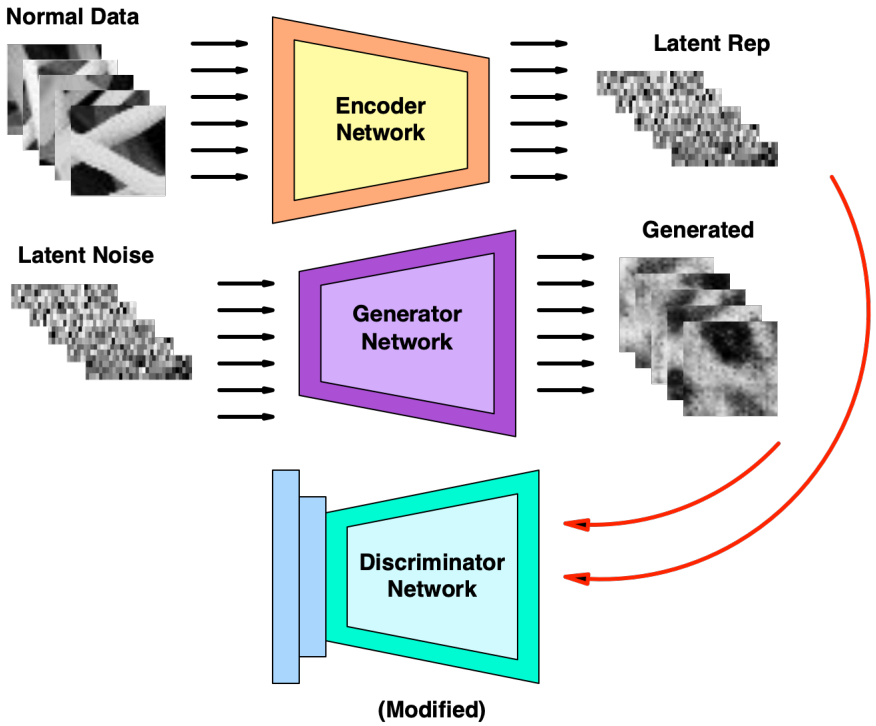




GAN Based Anomaly Detection Models

- GAN training + latent representation learning
  - Approximation of latent representation
  - Adversarial feature learning using Encoder network

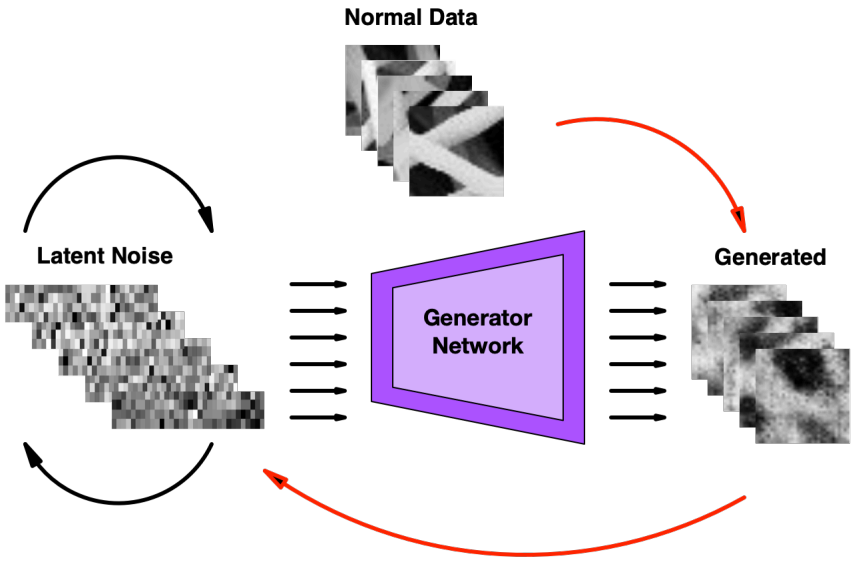




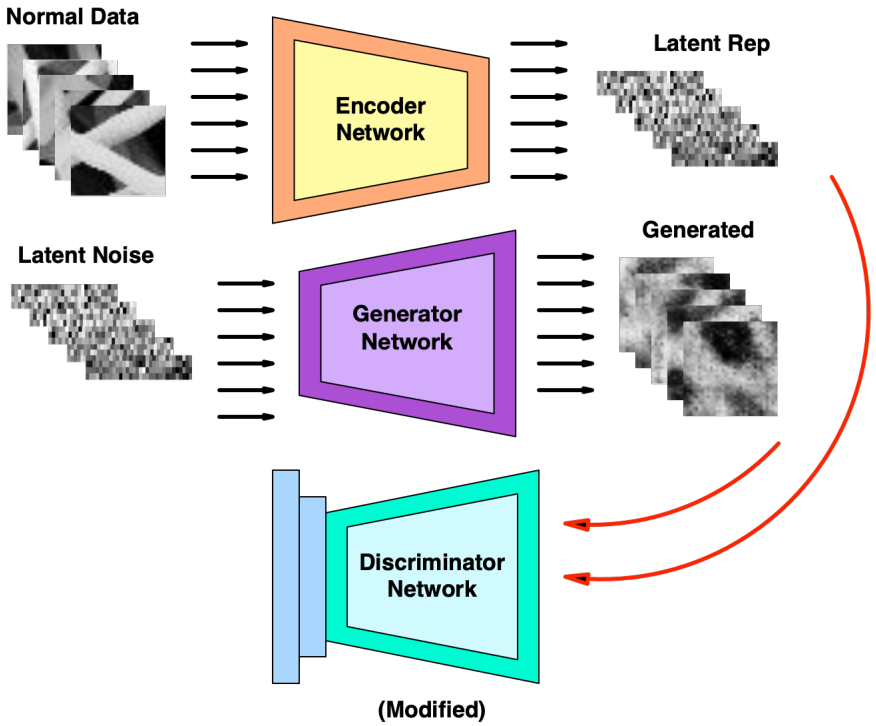
Really Slow Performance

GAN Based Anomaly Detection Models

- GAN training + latent representation learning
  - Approximation of latent representation
  - Adversarial feature learning using Encoder network



Really Slow Performance



Convergence & Stabilization Problems

Proposed Model

- Energy Based GAN for more stable training

Proposed Model

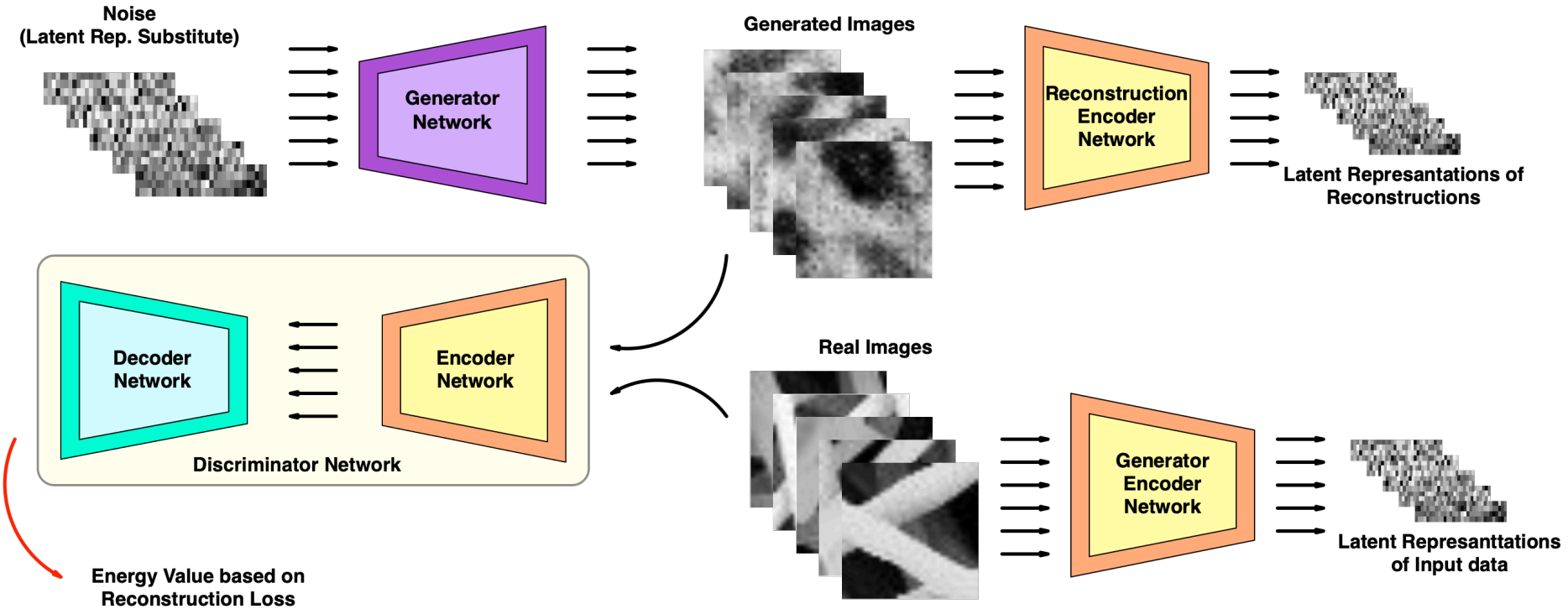
- Energy Based GAN for more stable training
- Encoder networks for latent representations of inputs and their reconstructions.

Proposed Model

- Energy Based GAN for more stable training
- Encoder networks for latent representations of inputs and their reconstructions.
- Sequential training to ensure convergence

Proposed Model

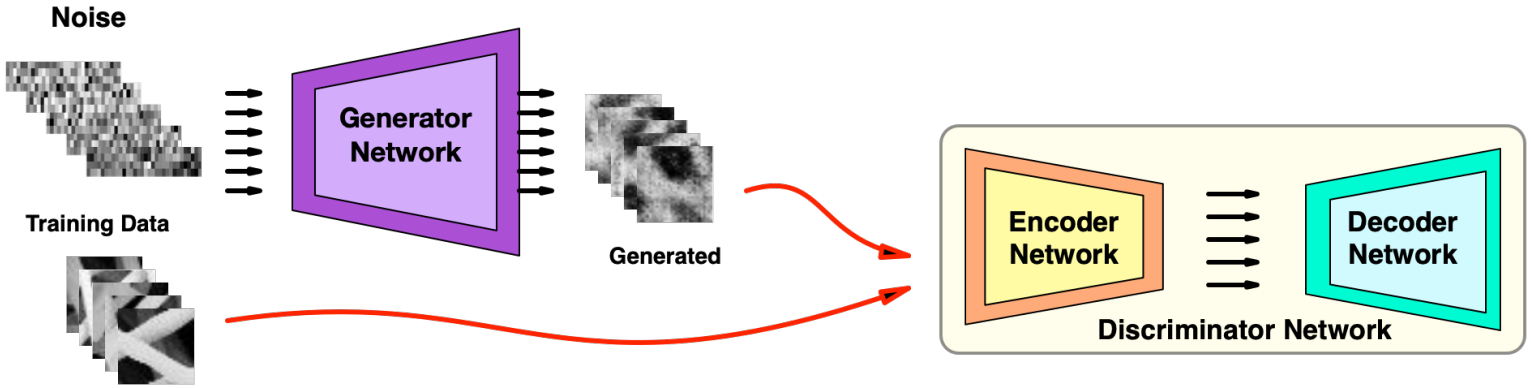
- Energy Based GAN for more stable training
- Encoder networks for latent representations of inputs and their reconstructions.
- Sequential training to ensure convergence



Energy Based Generative Adversarial Network

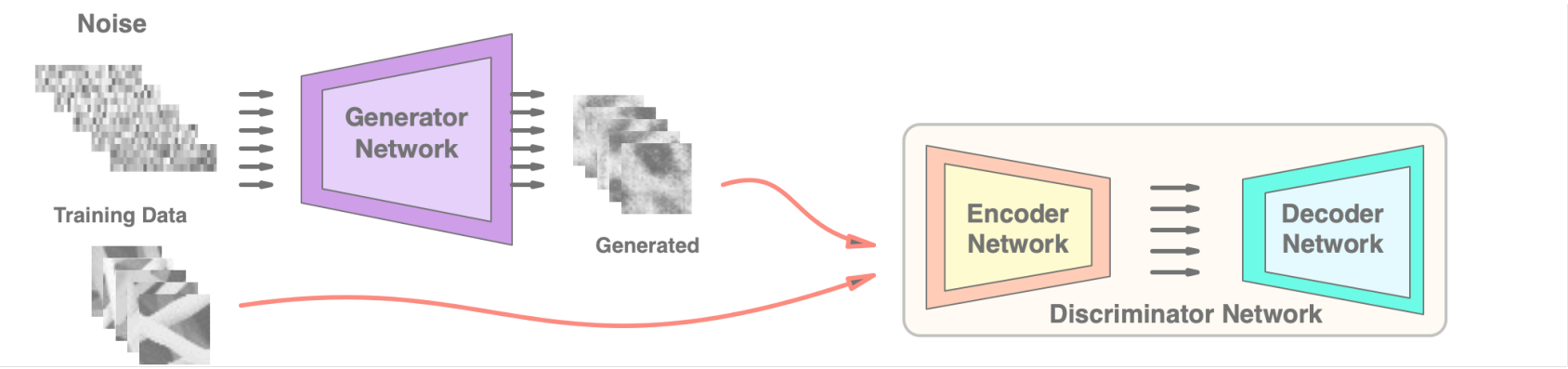
Training of the Model – Stage 1

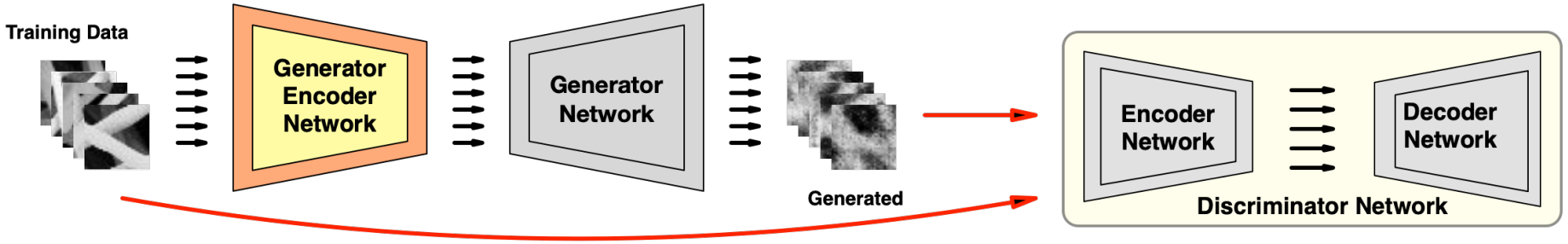
- Initial stage is energy based GAN training



Training of the Model – Stage 2

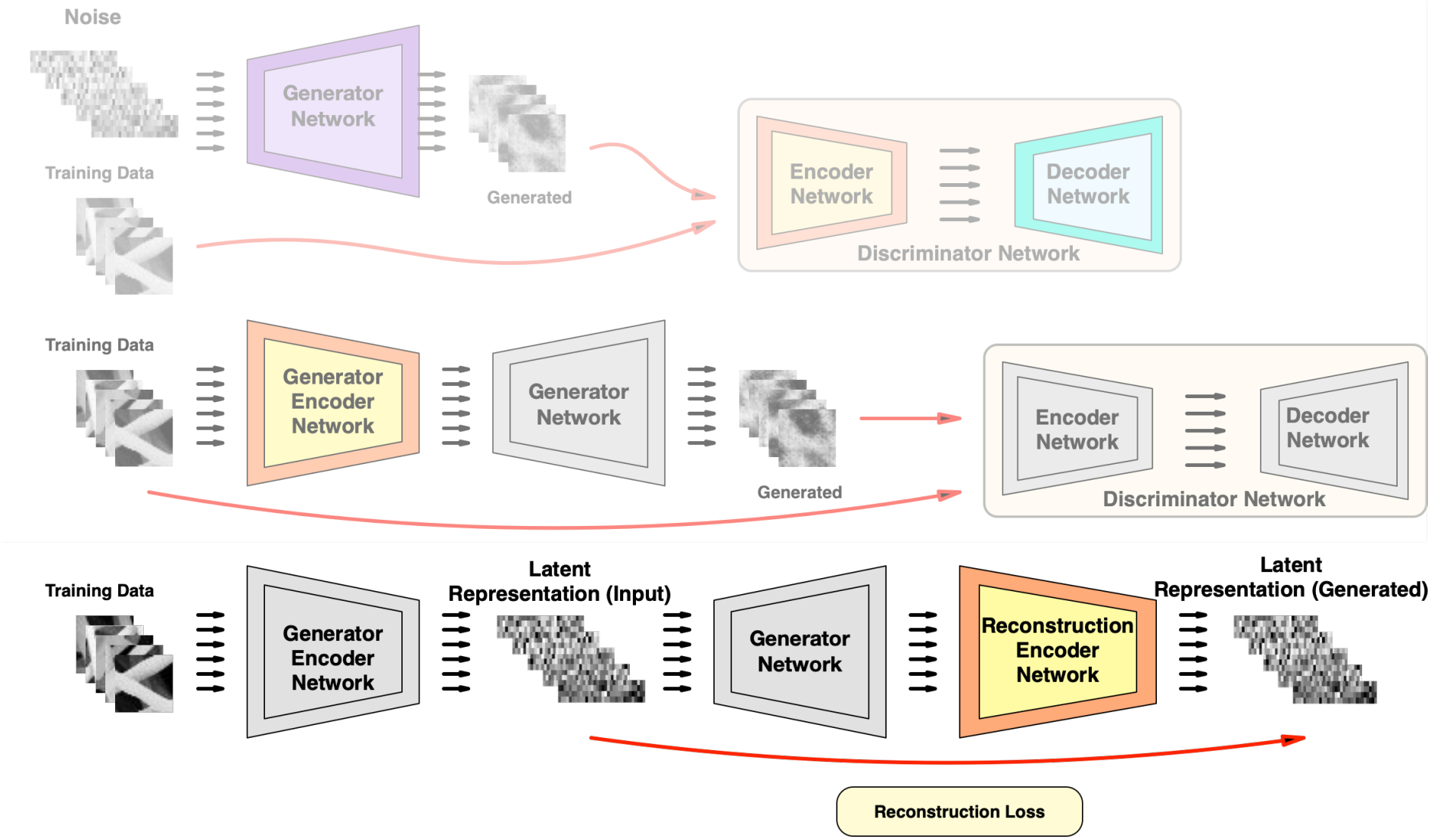
- Generator Encoder training for bi directional mapping





Training of the Model – Stage 3

- Reconstruction Encoder training for L.R. of reconstructions



Anomaly Detection of Proposed Model