Why Adversarial Interaction Creates Non-Homogeneous Patterns: A Pseudo-Reaction Diffusion Model for Turing Instability

Litu Rout

Association for the Advancement of Artificial Intelligence (AAAI-21)

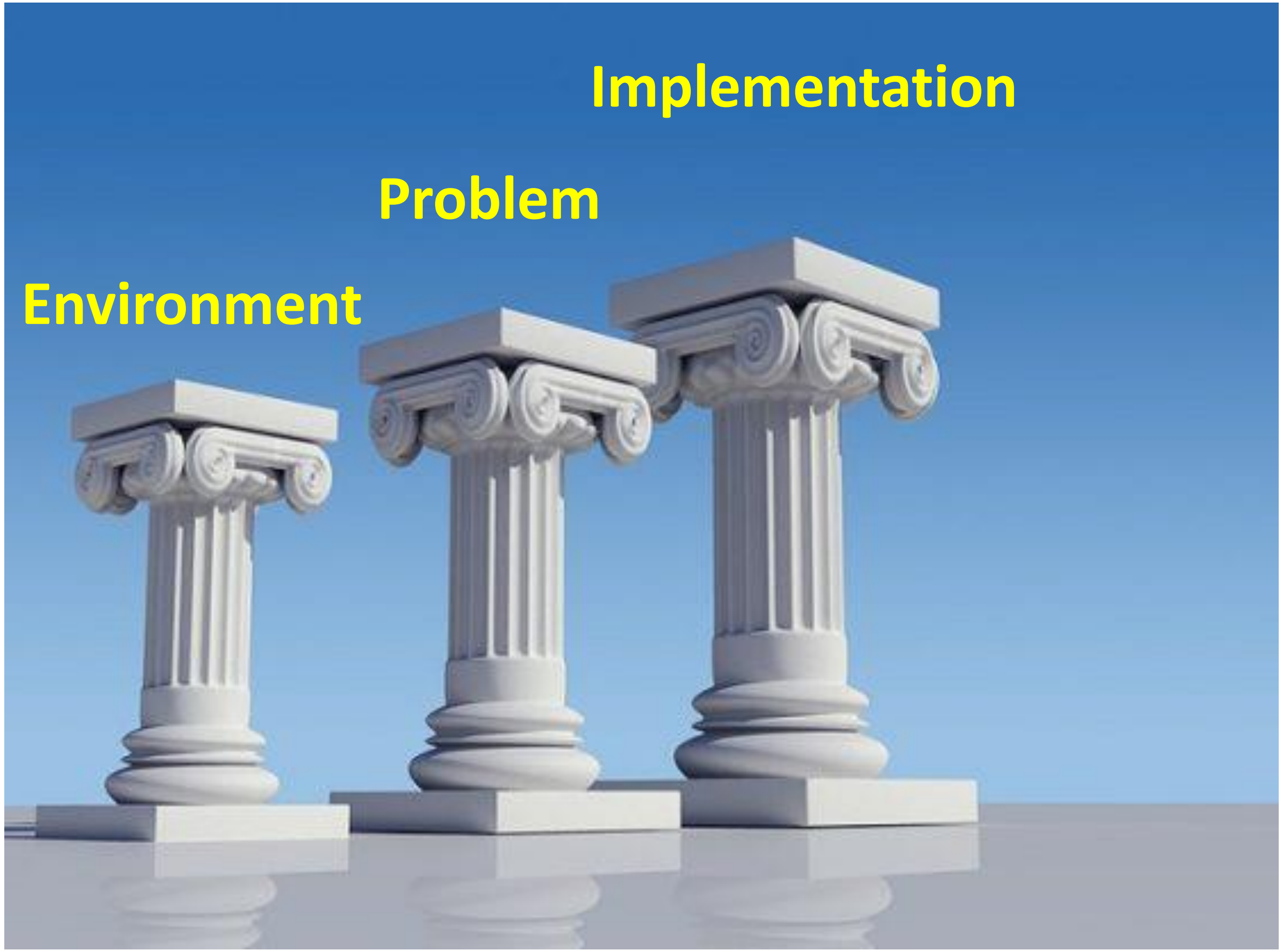
Preprint: https://liturout.github.io/data/aaai21_preprint.pdf

June 9, 2021

Why Adversarial Interaction Creates Non-Homogeneous Patterns: A Pseudo-Reaction Diffusion Model for Turing Instability

- Adversarial Interaction
  - Generative Adversarial Networks (GANs)
  - Application of conditional GANs
- Non-Homogeneous Patterns
  - Homogeneous patterns
  - Supervised learning
- Reaction-Diffusion
  - Turing's RD model (1952)
  - Gray-Scott RD model (1984)
- Turing Instability
  - Reaction dynamics
  - Diffusion dynamics

Three Pillars of Deep Learning



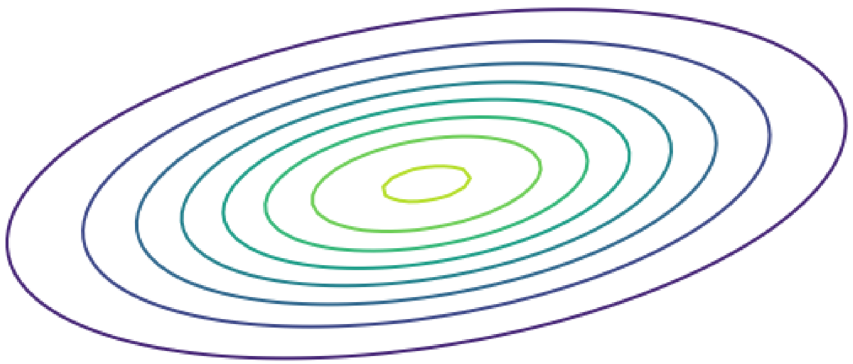
Why Adversarial Interaction Creates Non-Homogeneous Patterns: A Pseudo-Reaction Diffusion Model for Turing Instability

- Adversarial Interaction
  - Generative Adversarial Networks (GANs)
  - Application of conditional GANs
- Non-Homogeneous Patterns
  - Homogeneous patterns
  - Supervised learning
- Reaction-Diffusion
  - Turing's RD model (1952)
  - Gray-Scott RD model (1984)
- Turing Instability
  - Reaction dynamics
  - Diffusion dynamics

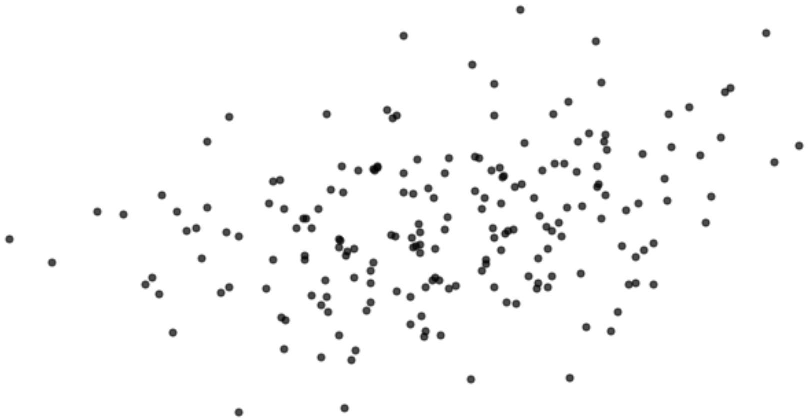
Outline

- Maximum Likelihood Estimation
- Reverse KL Divergence
- Introduction to GANs
- GANs in Remote Sensing

Generative Modeling: Density Estimation



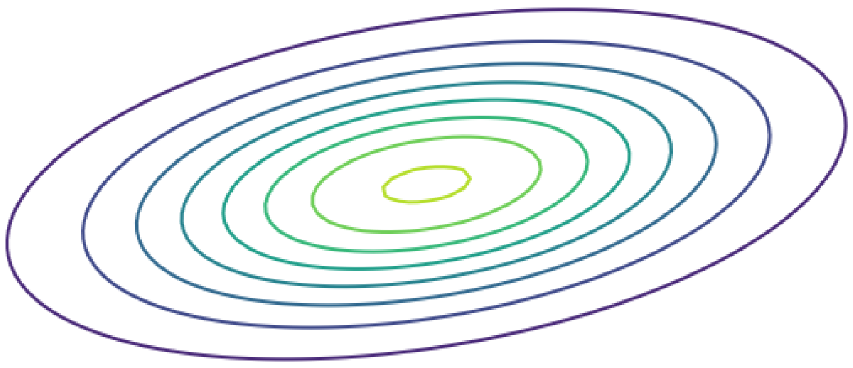
(a)



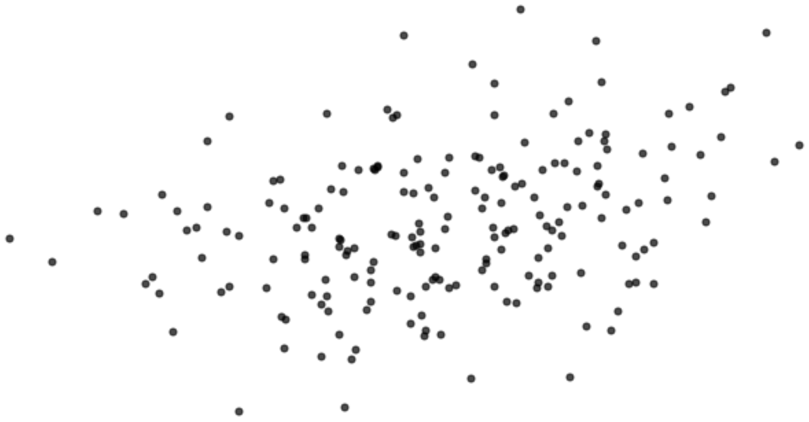
(b)

Colin Raffel, 2018

Generative Modeling: Density Estimation



(a)

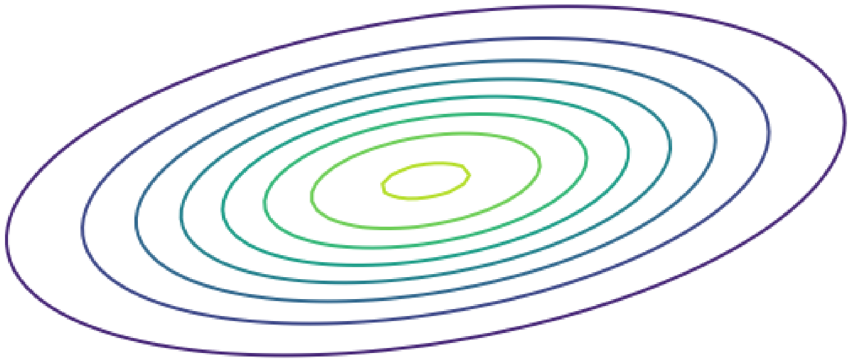


(b)

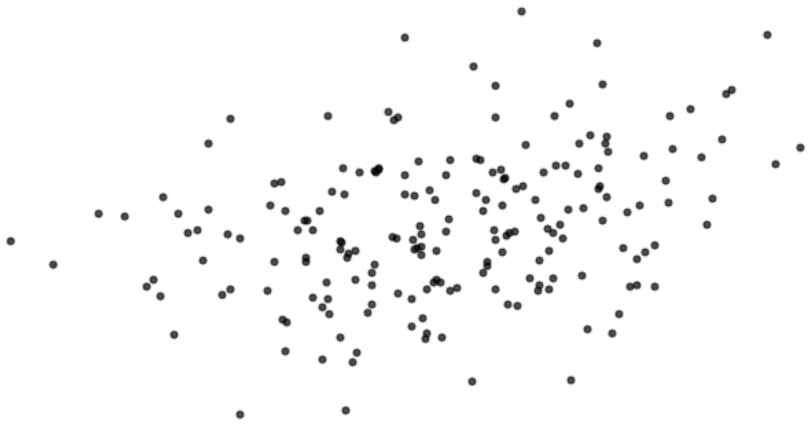
$$
\begin{aligned}
\theta^* &= \arg\min_\theta \mathrm{KL}(p(x)||q_\theta(x)) \\
&= \arg\min_\theta \mathbb{E}_{x\sim p}[\log p(x) - \log q_\theta(x)] \\
&= \arg\min_\theta \mathbb{E}_{x\sim p}[\log p(x)] - \mathbb{E}_{x\sim p}[\log q_\theta(x)]
\end{aligned}
$$

Colin Raffel, 2018

Generative Modeling: Density Estimation



(a)

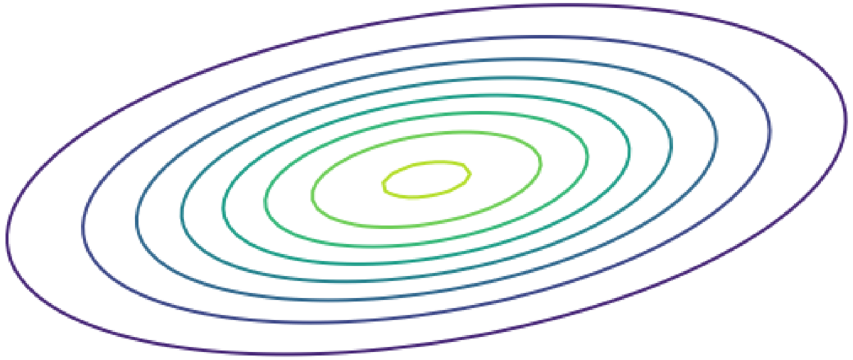


(b)

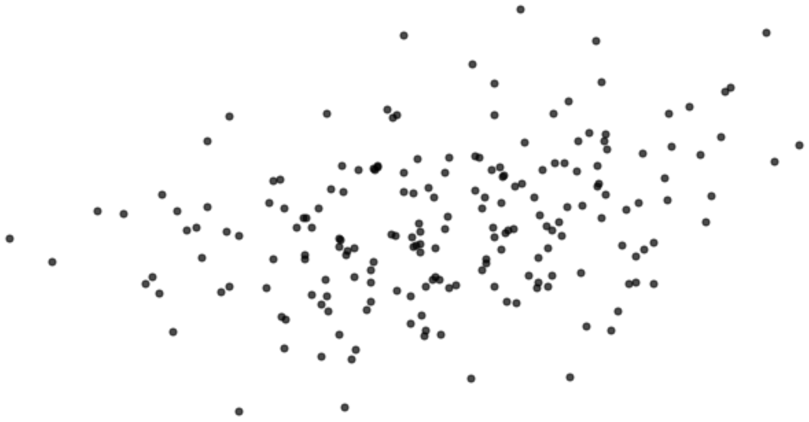
$$
\begin{aligned}
\theta^* &= \arg\min_\theta \cancel{\mathbb{E}_{x\sim p}[\log p(x)]} - \mathbb{E}_{x\sim p}[\log q_\theta(x)] \\
&= \arg\min_\theta -\mathbb{E}_{x\sim p}[\log q_\theta(x)] \\
&= \arg\max_\theta \mathbb{E}_{x\sim p}[\log q_\theta(x)]
\end{aligned}
$$

Colin Raffel, 2018

Generative Modeling: Density Estimation



(a)



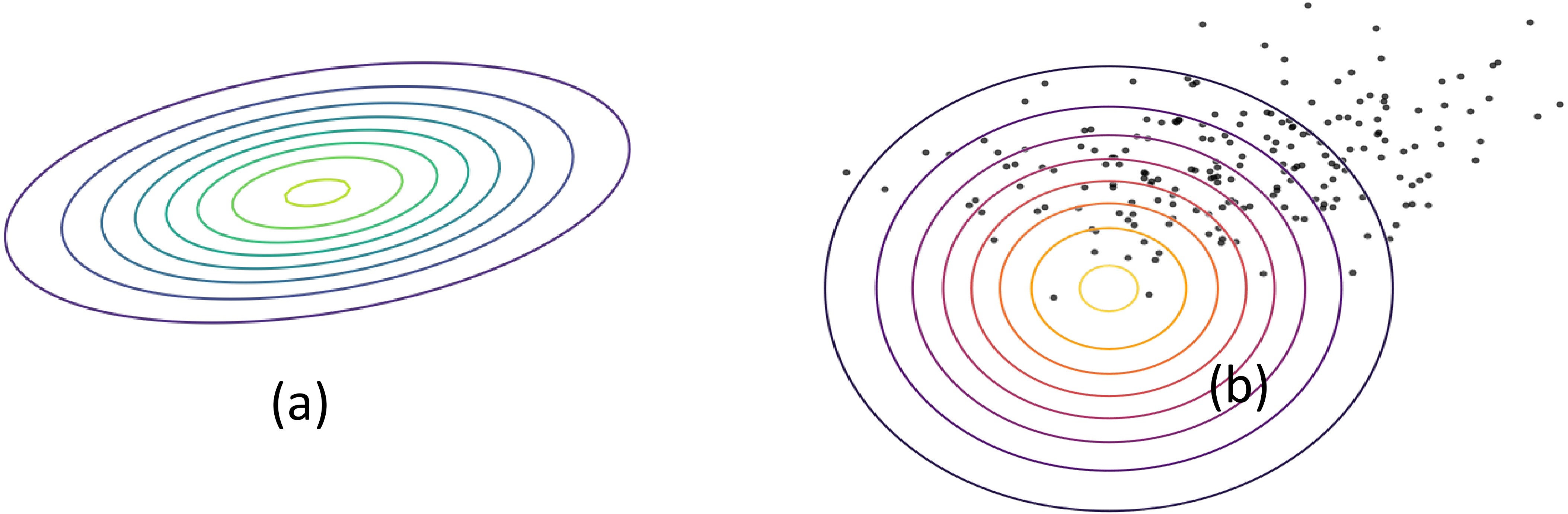
(b)

$$
\begin{aligned}
\theta^* &= \arg\min_\theta \cancel{\mathbb{E}_{x\sim p}[\log p(x)]} - \mathbb{E}_{x\sim p}[\log q_\theta(x)] \\
&= \arg\min_\theta -\mathbb{E}_{x\sim p}[\log q_\theta(x)] \\
&= \arg\max_\theta \mathbb{E}_{x\sim p}[\log q_\theta(x)]
\end{aligned}
$$

$$\boxed{\theta(t+1) = \theta(t) + \eta \frac{\partial L(\theta)}{\partial \theta(t)}}$$

Colin Raffel, 2018

Generative Modeling: Density Estimation



(a) (b)

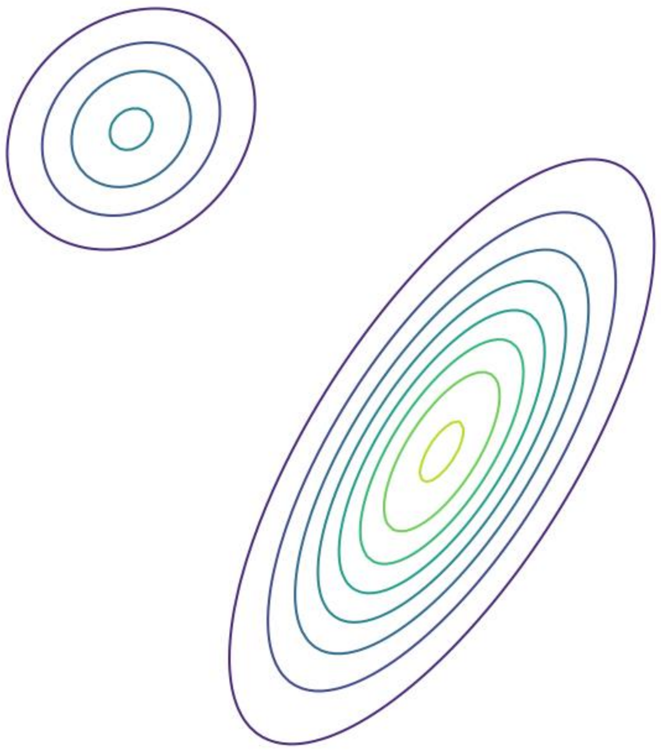
Colin Raffel, 2018

Generative Modeling: Density Estimation

(a) (b)

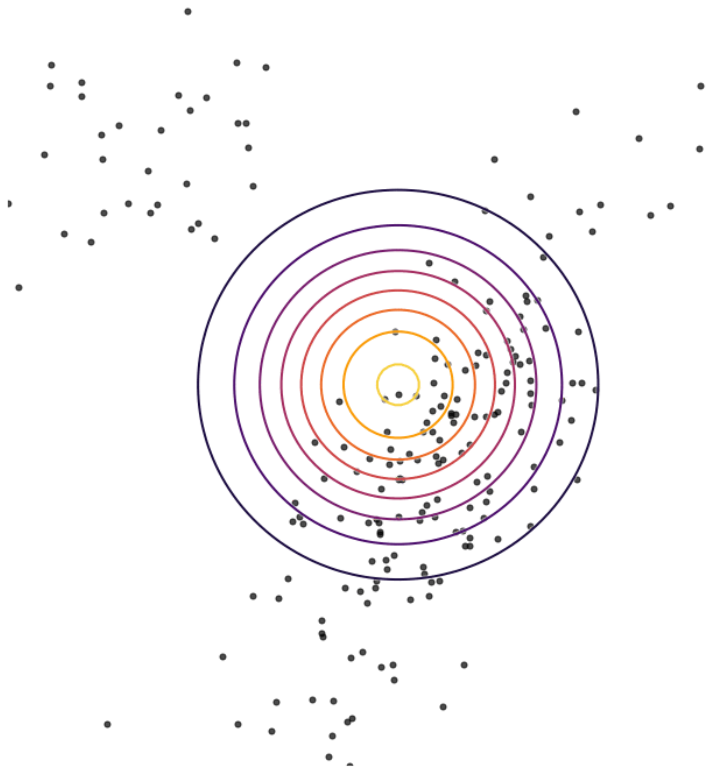
Colin Raffel, 2018

Generative Modeling: Density Estimation



(a)

$$\theta^* = \arg\max_\theta \mathbb{E}_{x \sim p}[\log q_\theta(x)]$$



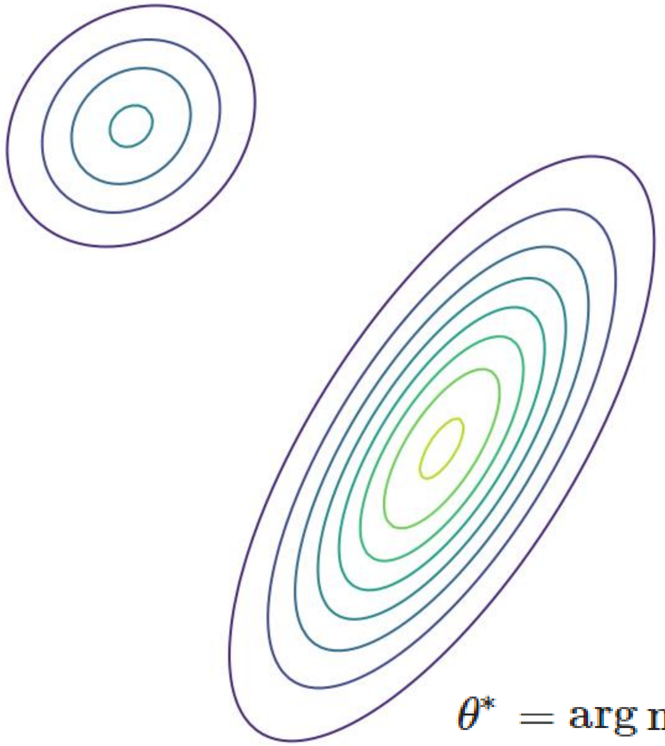
(b)

Colin Raffel, 2018

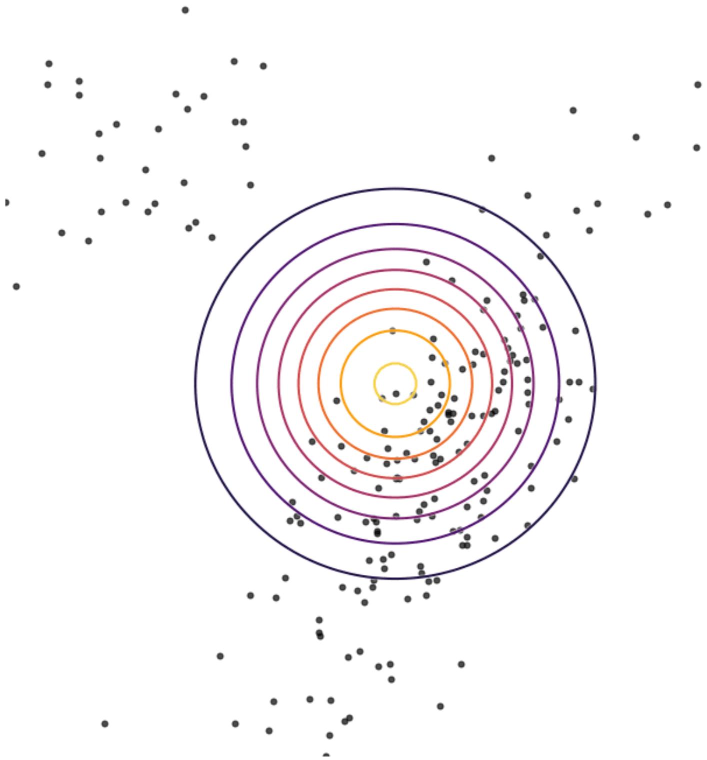
Outline

- Maximum Likelihood Estimation
- Reverse KL Divergence
- Introduction to GANs
- GANs in Remote Sensing

Generative Modeling: Density Estimation



(a)

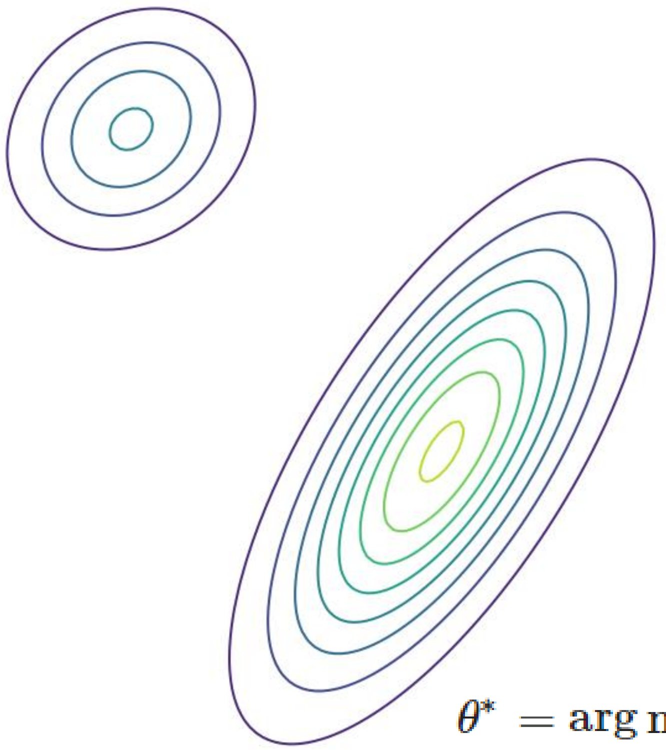


(b)

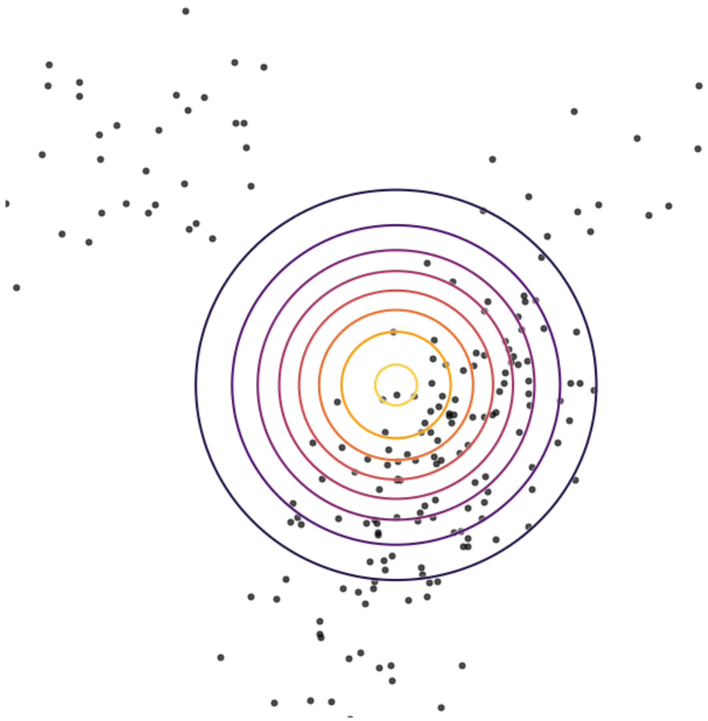
$$
\begin{aligned}
\theta^* &= \arg\min_{\theta} \mathrm{KL}(q_\theta || p) \\
&= \arg\min_{\theta} \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x) - \log p(x)] \\
&= \arg\min_{\theta} \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] - \mathbb{E}_{x\sim q_\theta}[\log p(x)] \\
&= \arg\max_{\theta} -\mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] + \mathbb{E}_{x\sim q_\theta}[\log p(x)]
\end{aligned}
$$

Colin Raffel, 2018

Generative Modeling: Density Estimation



(a)

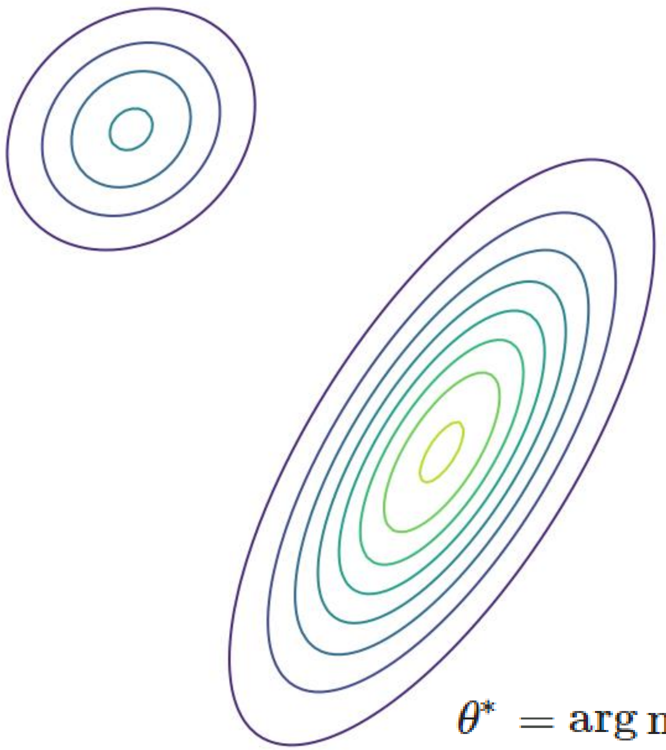


(b)

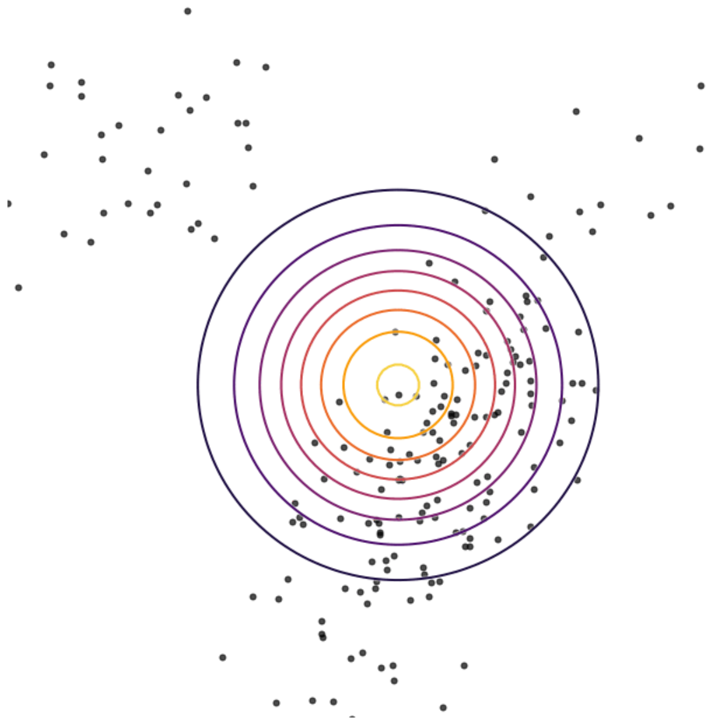
$$
\begin{aligned}
\theta^* &= \arg\min_{\theta} \mathrm{KL}(q_\theta || p) \\
&= \arg\min_{\theta} \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x) - \log p(x)] \\
&= \arg\min_{\theta} \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] - \mathbb{E}_{x\sim q_\theta}[\log p(x)] \\
&= \arg\max_{\theta} -\mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] + \mathbb{E}_{x\sim q_\theta}[\log p(x)]
\end{aligned}
$$

Colin Raffel, 2018

Generative Modeling: Density Estimation



(a)



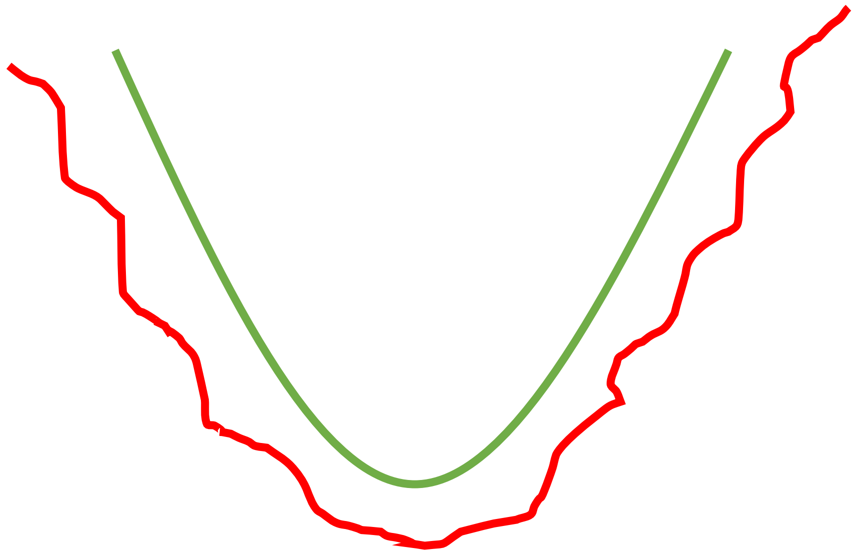
(b)

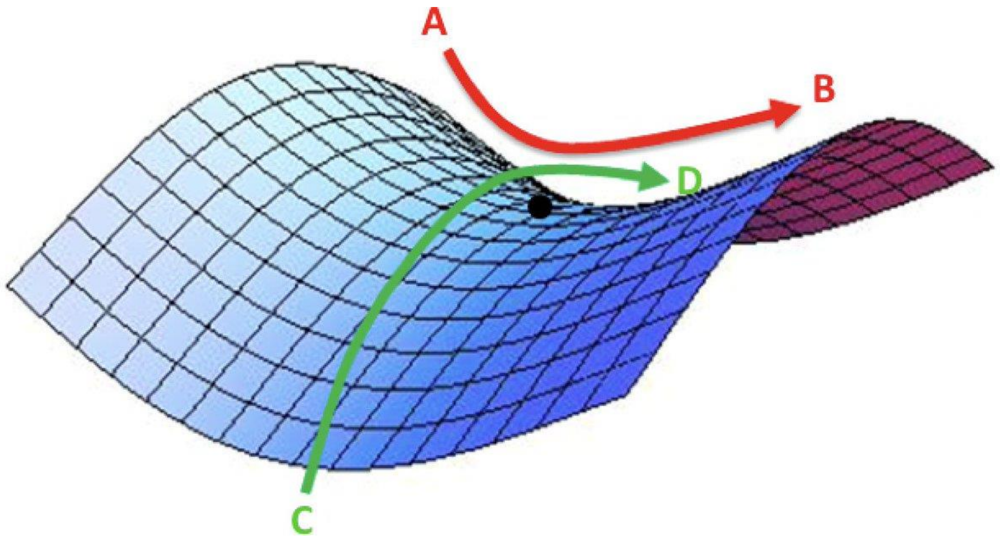
$$
\begin{aligned}
\theta^* &= \arg\min_{\theta} \mathrm{KL}(q_\theta || p) \\
&= \arg\min_{\theta} \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x) - \log p(x)] \\
&= \arg\min_{\theta} \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] - \mathbb{E}_{x\sim q_\theta}[\log p(x)] \\
&= \arg\max_{\theta} -\mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] + \boxed{\mathbb{E}_{x\sim q_\theta}[\log p(x)]}
\end{aligned}
$$

Colin Raffel, 2018

Generative Modeling: Density Estimation

$$
\begin{aligned}
\theta^* &= \arg\min_\theta \mathrm{KL}(q_\theta || p) \\
&= \arg\min_\theta \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x) - \log p(x)] \\
&= \arg\min_\theta \mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] - \mathbb{E}_{x\sim q_\theta}[\log p(x)] \\
&= \arg\max_\theta -\mathbb{E}_{x\sim q_\theta}[\log q_\theta(x)] + \boxed{\mathbb{E}_{x\sim q_\theta}[\log p(x)]}
\end{aligned}
$$



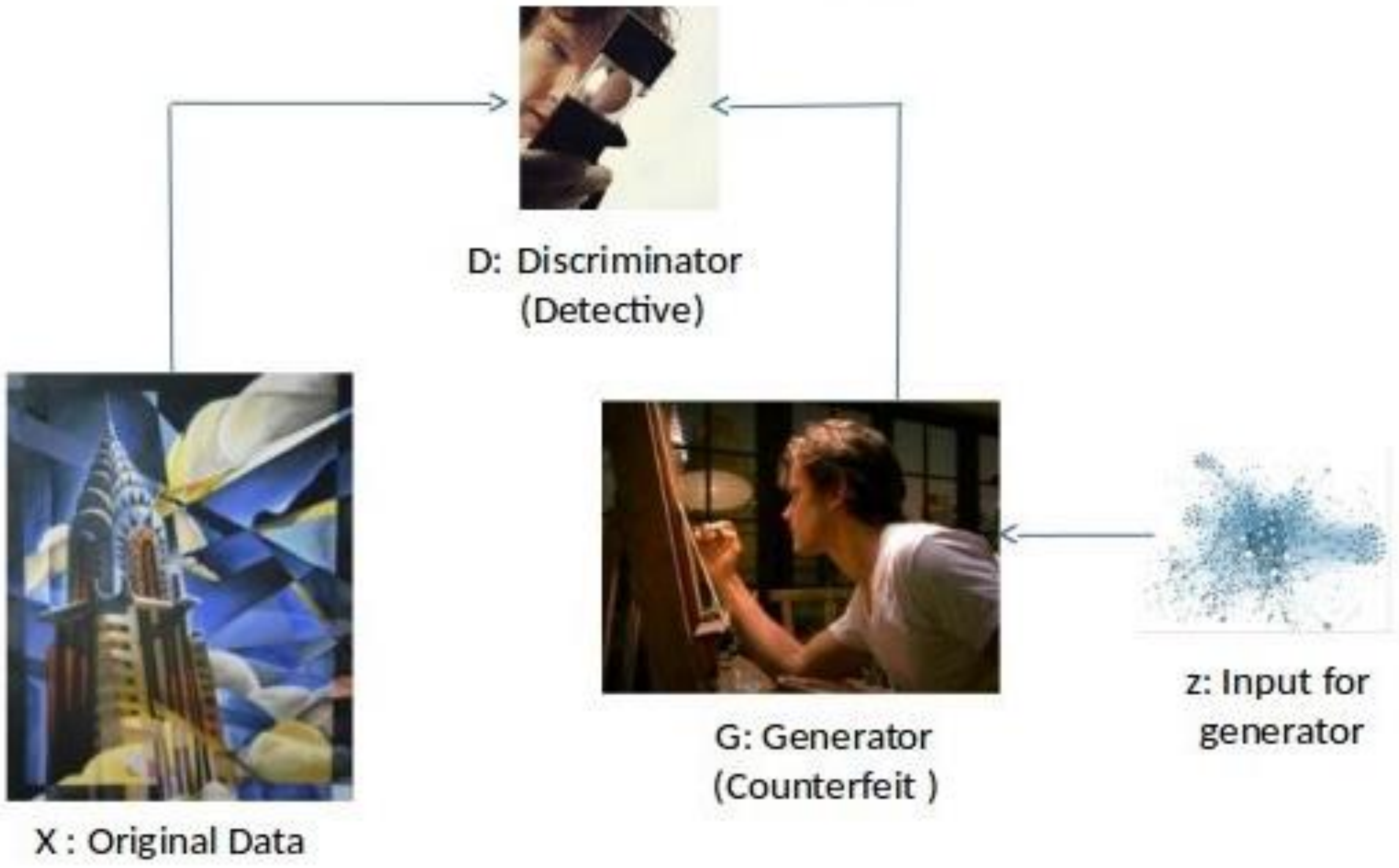


Termeh Kousha, 2019

Outline

- Maximum Likelihood Estimation
- Reverse KL Divergence
- Introduction to GANs
- GANs in Remote Sensing

Generative Modeling: Density Estimation



$$\theta^* = \arg\min_{\theta}\max_{\phi}\mathbb{E}_{x\sim p,\,\hat{x}\sim q_\theta}\mathrm{V}(f_\phi(x), f_\phi(\hat{x}))$$

$$\mathrm{V}(f_\phi(x), f_\phi(\hat{x})) = \log f_\phi(x) + \log[1 - f_\phi(\hat{x})]$$

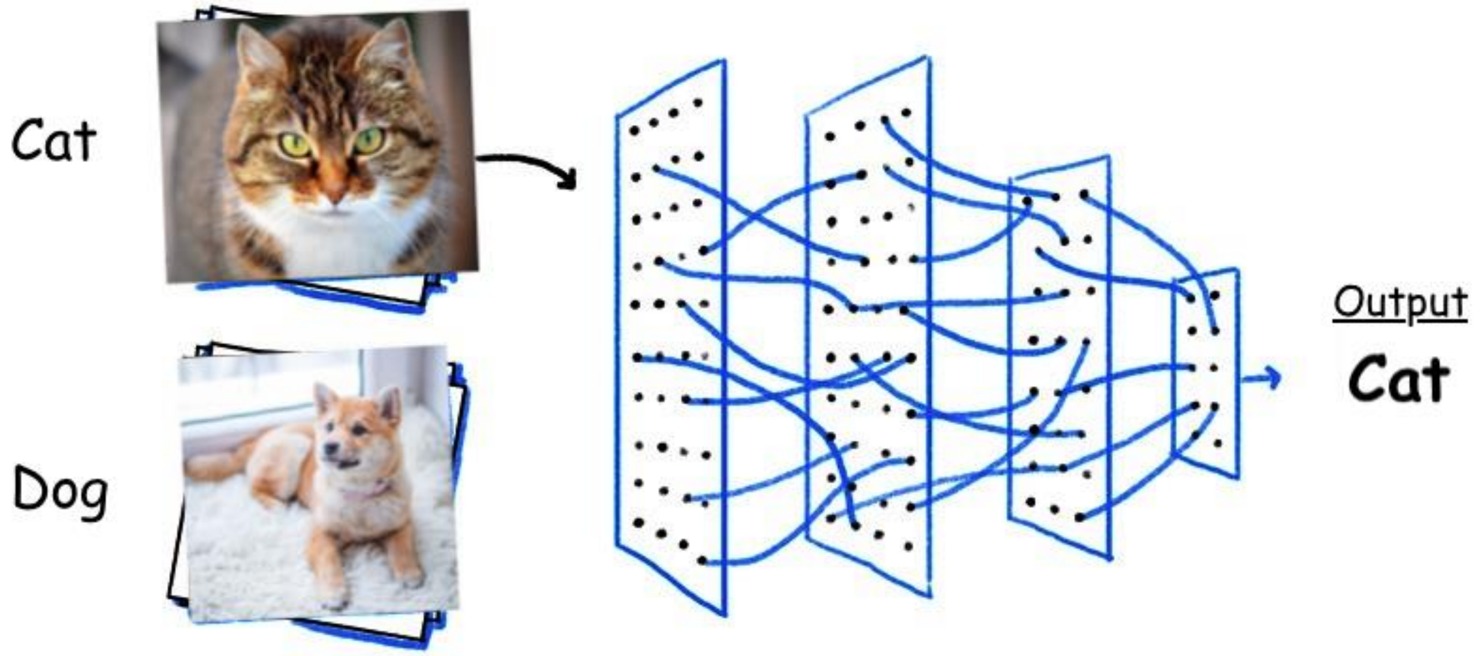
Generative Modeling: Density Estimation

(a) (b) (c)

Colin Raffel, 2018

Why Generative Adversarial Networks?

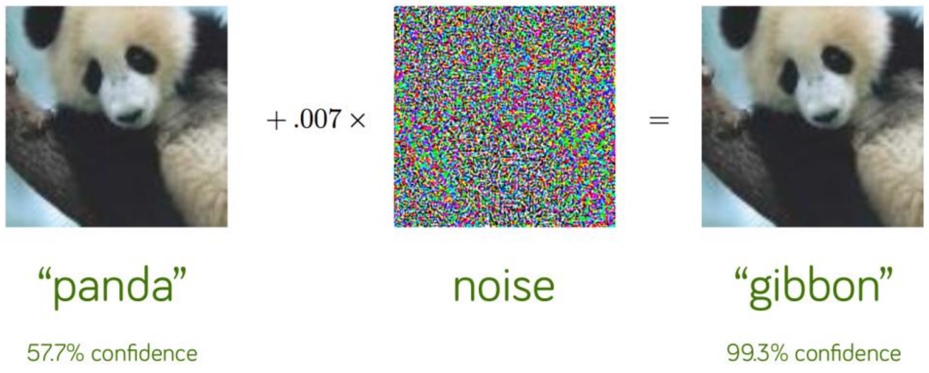
- Image Classification as Discriminator
  - Given input X, predict label Y
  - Estimate P(Y|X)

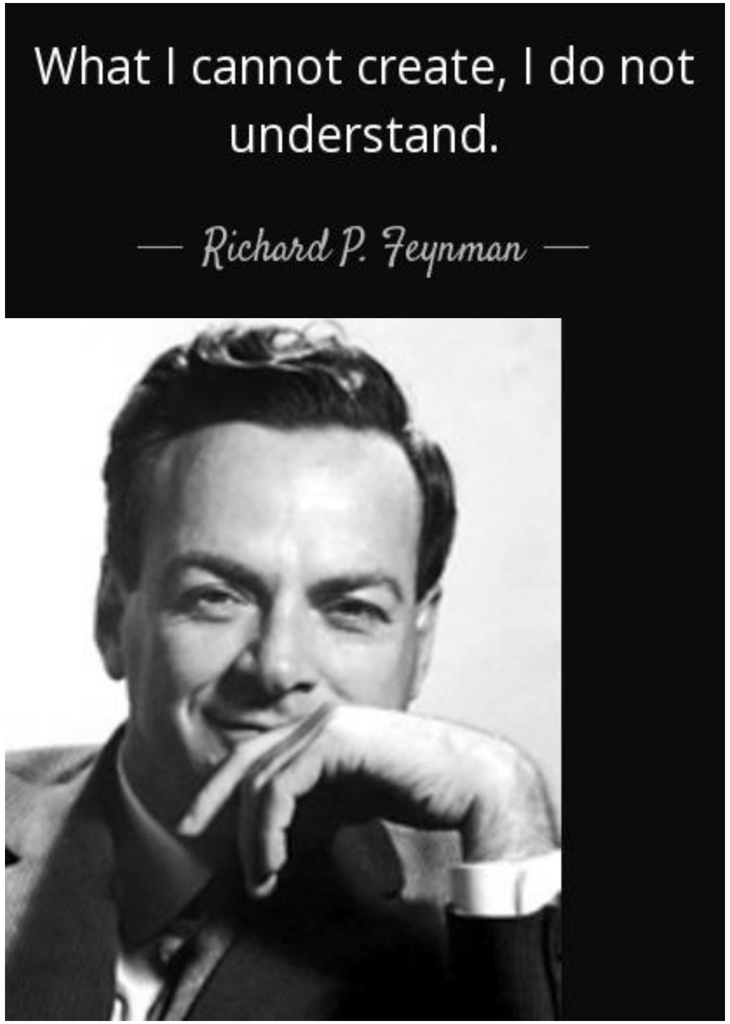


Why Generative Adversarial Networks?

- Image Classification as Discriminator
  - Given input X, predict label Y
  - Estimate P(Y|X)

- Challenges in Discriminative Models
  - Unknown P(X)
  - Can't sample from P(X)





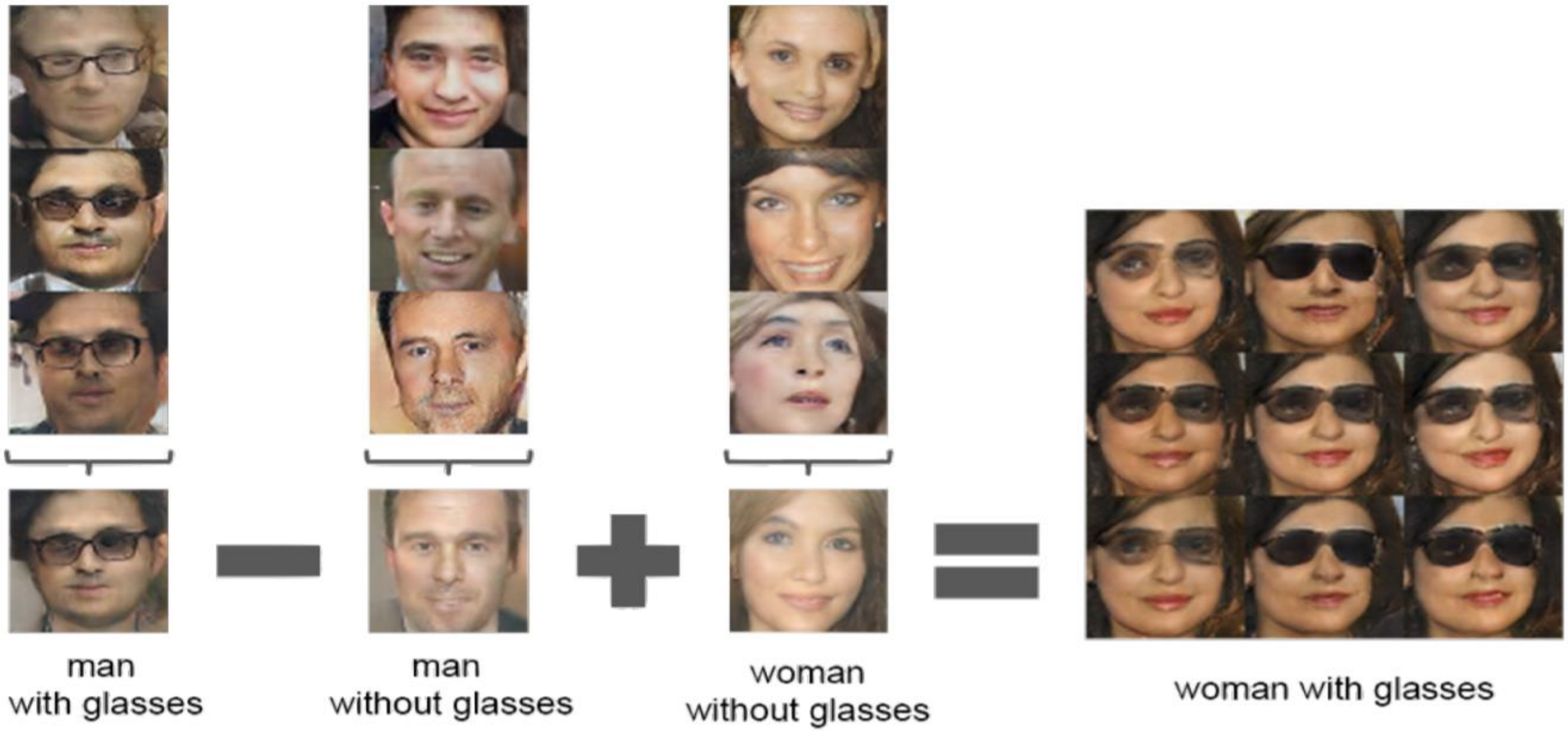
Why Generative Adversarial Networks?

- Image Classification as Discriminator
  - Given input X, predict label Y
  - Estimate P(Y|X)
- Challenges in Discriminative Models
  - Unknown P(X)
  - Can’t sample from P(X)
- Motivation for GANs
  - Model P(X)
  - Generate new samples

Faces Dataset

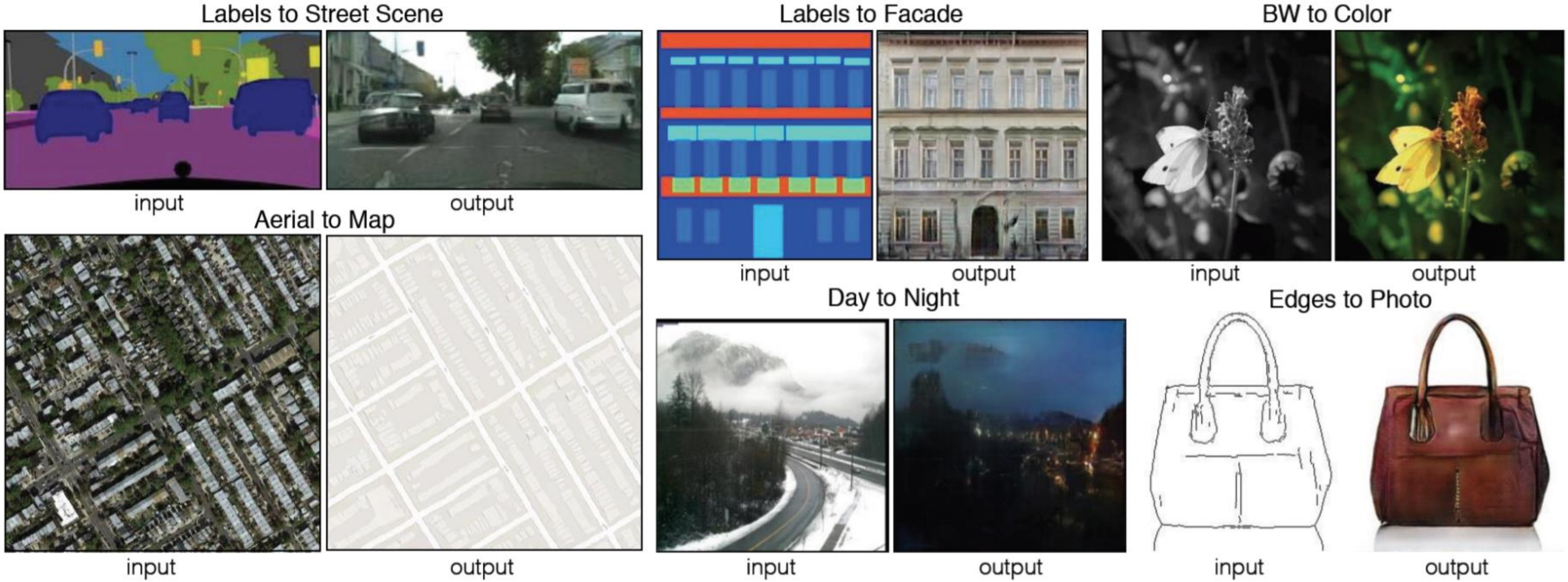
Ian Goodfellow et al., 2014

Generative Modeling: Latent Space Interpolation



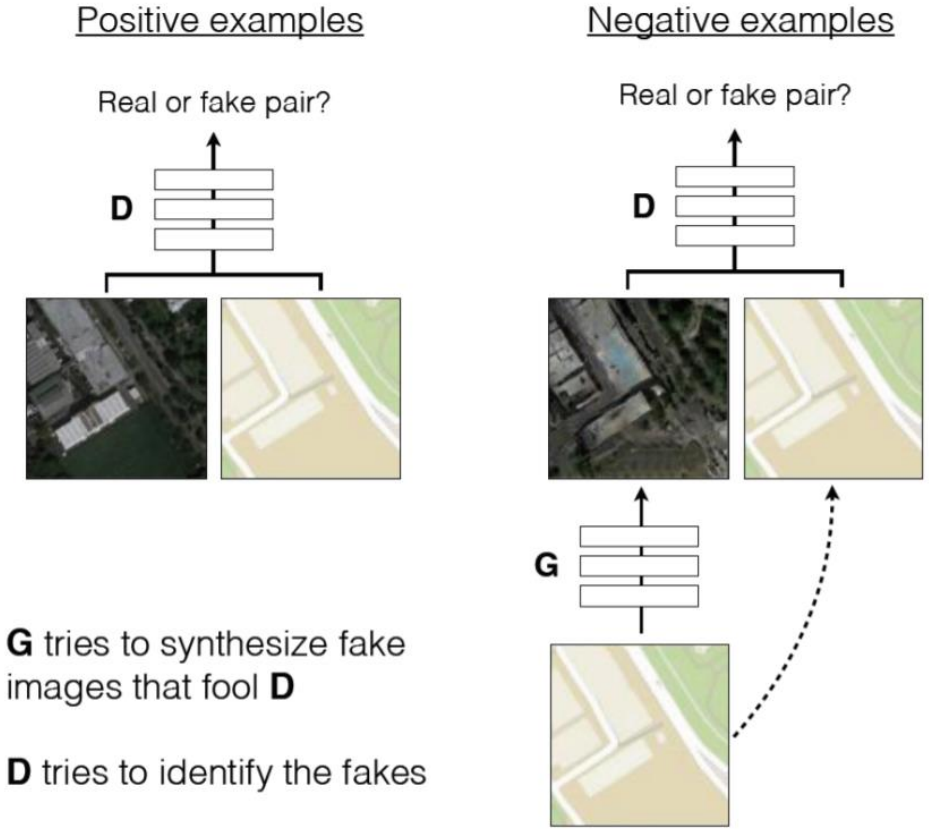
Alec Radford et al., 2015

Generative Modeling: Image-to-Image Translation



Phillip Isola et al., 2016

Generative Modeling: Image-to-Image Translation



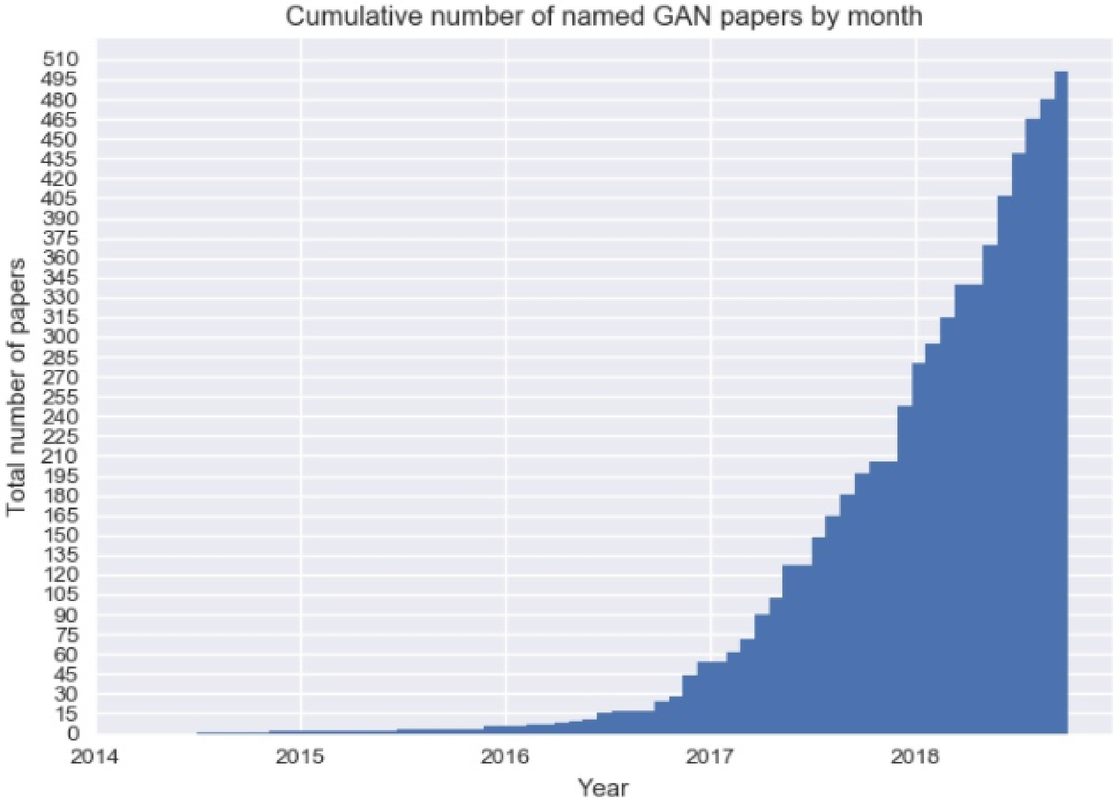
Phillip Isola et al., 2016

Generative Modeling: Sample Generation

Training Data
(CelebA)

Sample Generator
(Karras et al, 2017)

Generative Modeling: Sample Generation

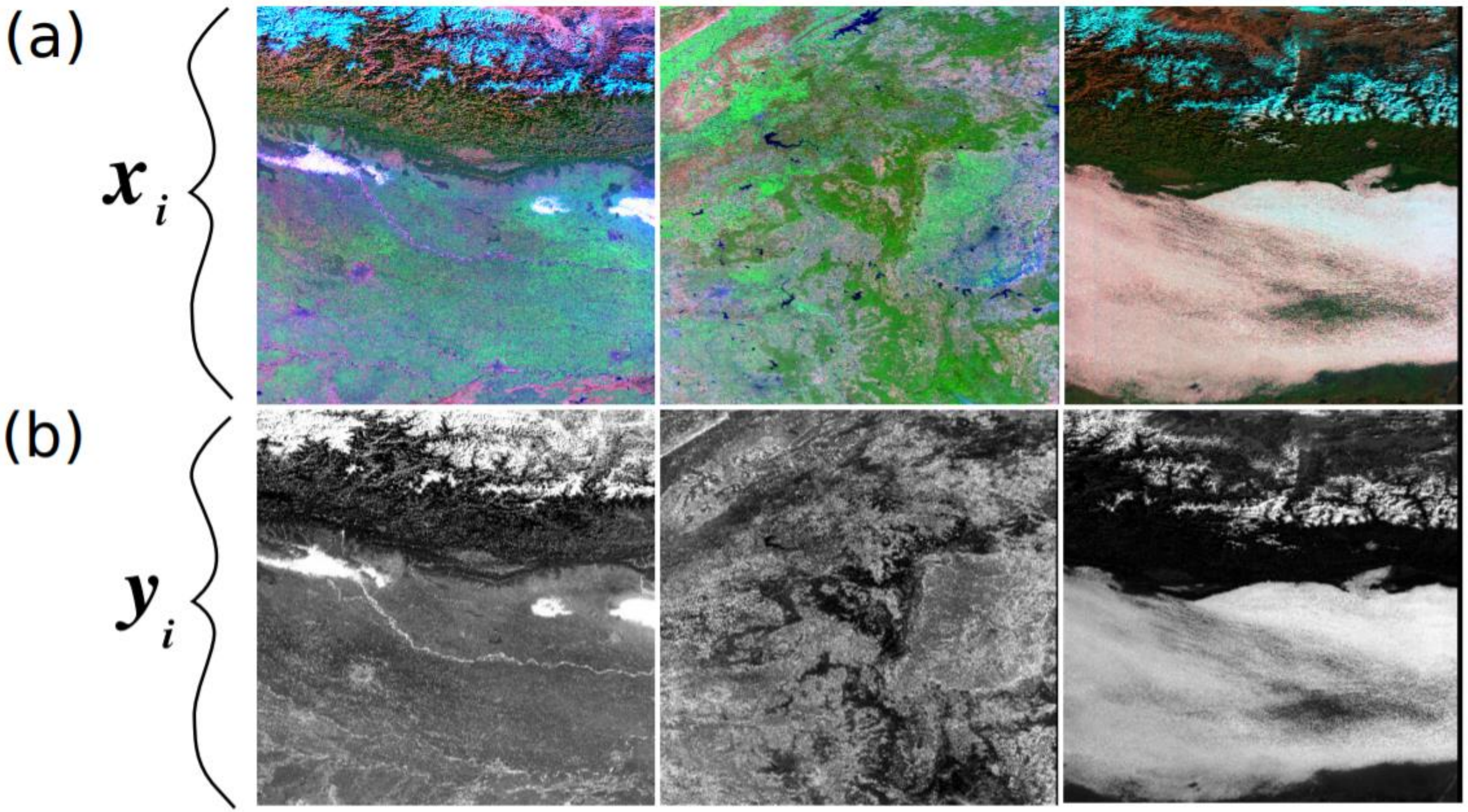


GAN Zoo: https://github.com/hindupuravinash/the-gan-zoo

Outline

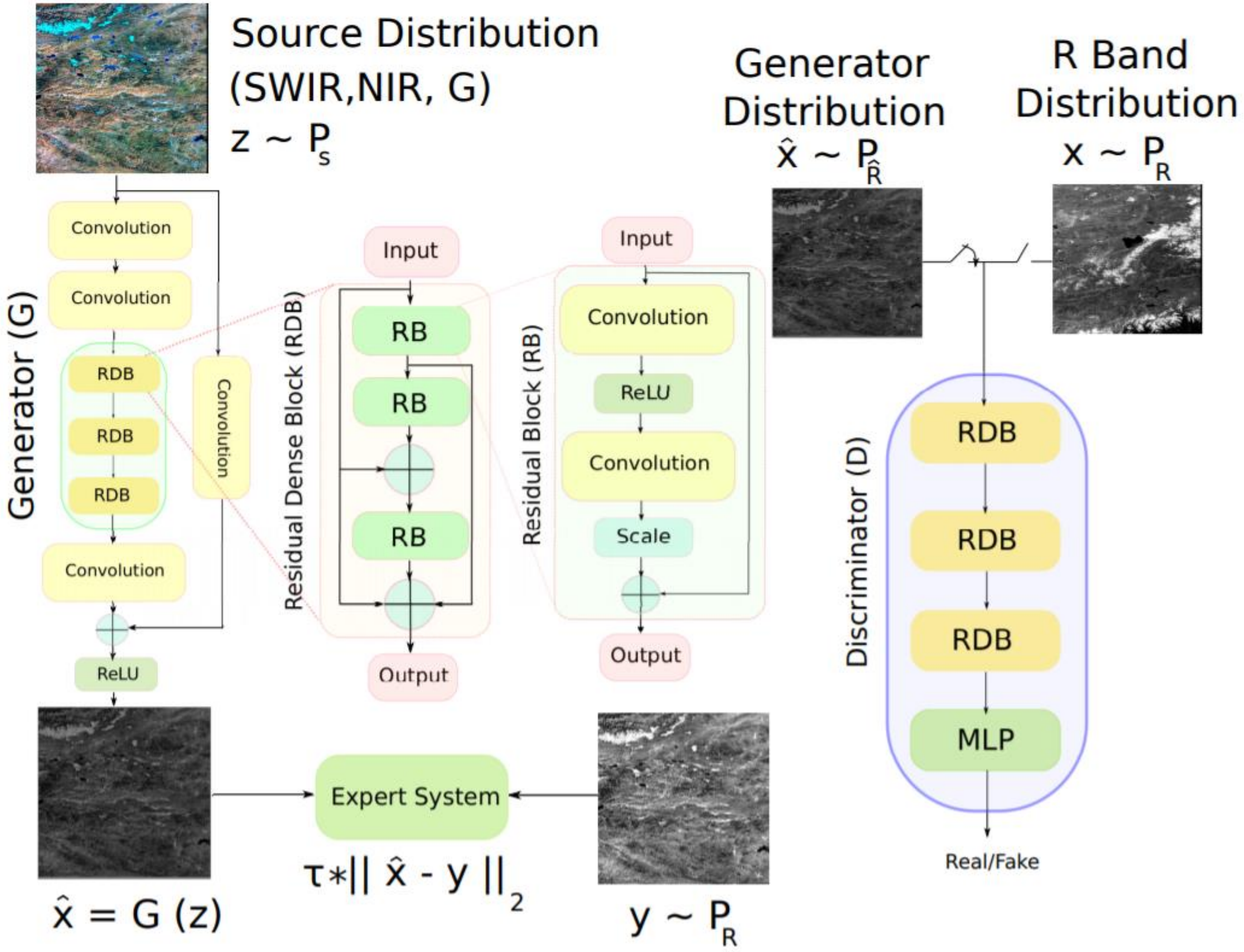
- Maximum Likelihood Estimation
- Reverse KL Divergence
- Introduction to GANs
- GANs in Remote Sensing

Generative Modeling: Missing Data Reconstruction



L.R., TGRS 2020: https://liturout.github.io/data/alert.pdf

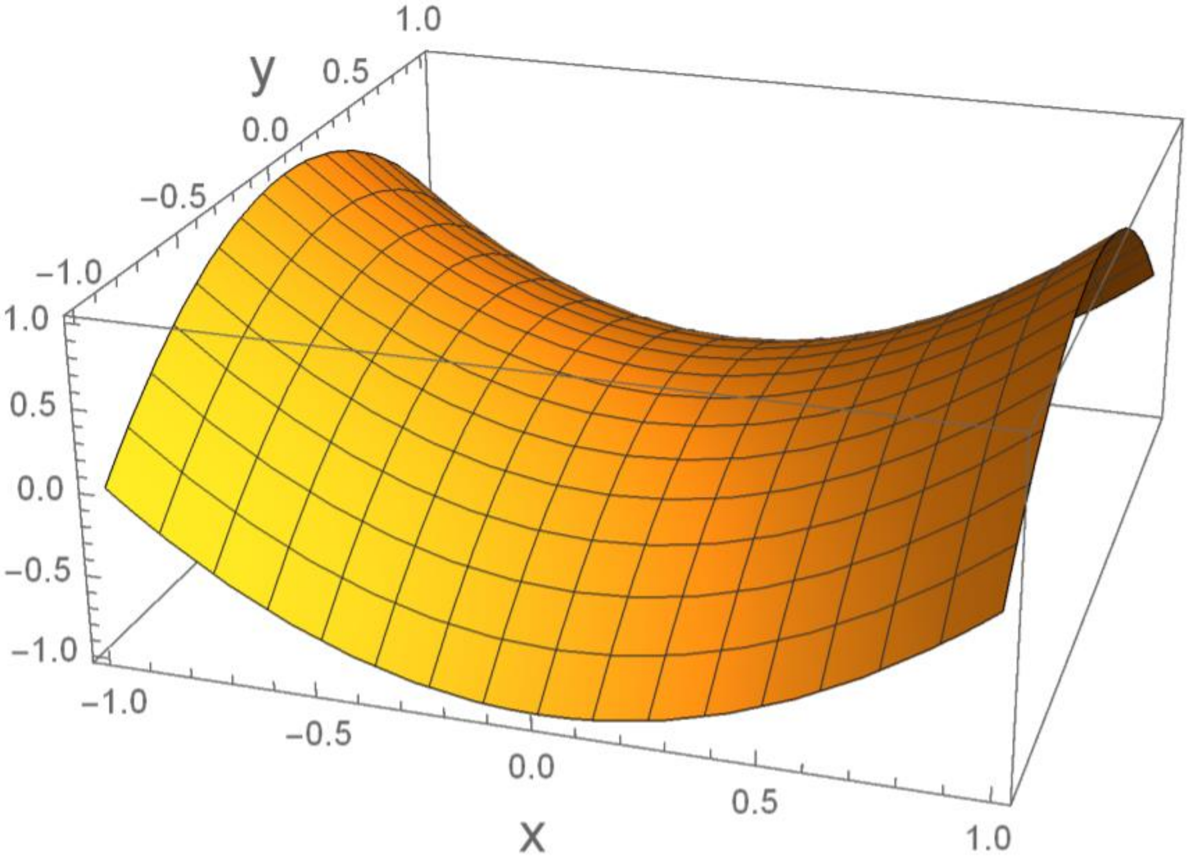
Generative Modeling: Missing Data Reconstruction



Generative Modeling: Missing Data Reconstruction

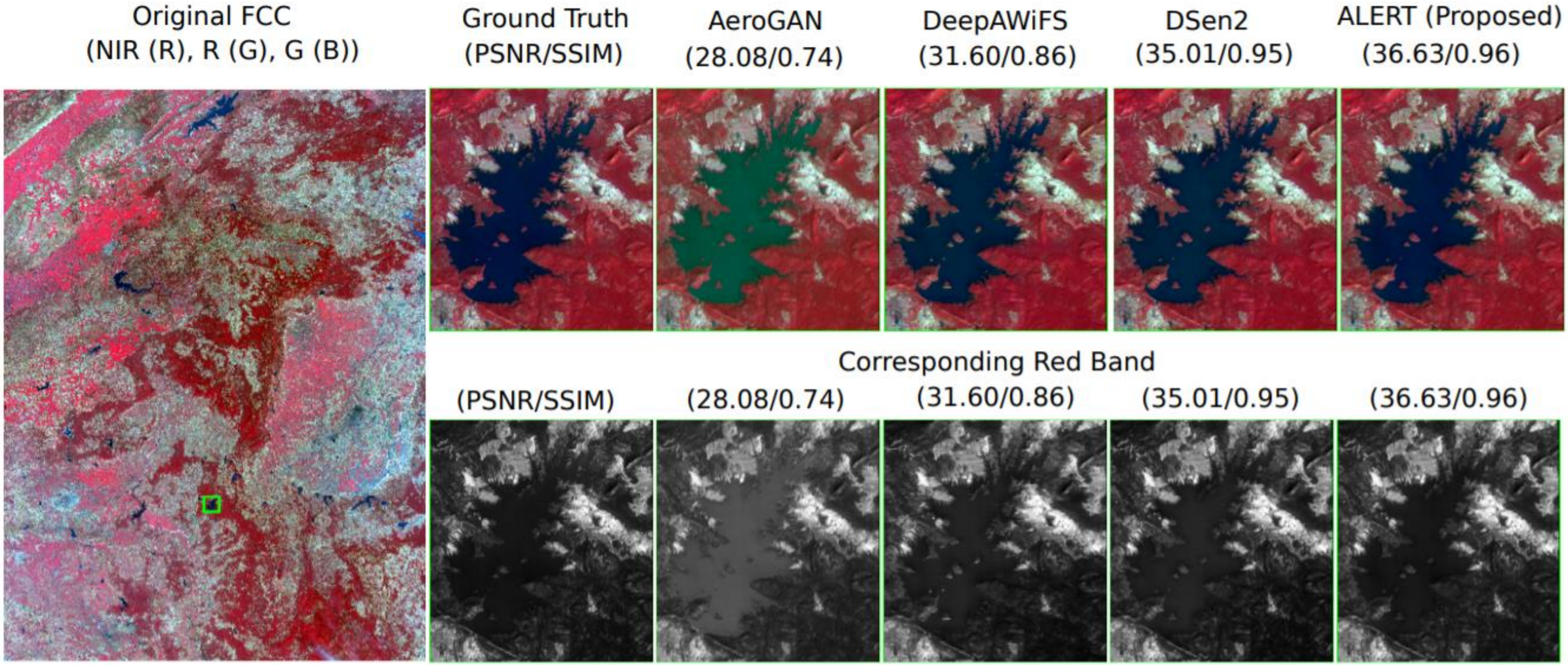
$$\min_{G} \mathbb{E}_{x\sim\mathbb{P}_R}\left[D\left(x\right)\right] - \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{R}}}\left[D\left(\hat{x}\right)\right] + \mathbb{E}_{y\sim\mathbb{P}_R}\left[\left\|\tau\left(\hat{x}-y\right)\right\|_2\right],$$

$$\min_{D} \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{R}}}\left[D\left(\hat{x}\right)\right] - \mathbb{E}_{x\sim\mathbb{P}_R}\left[D\left(x\right)\right] + \lambda\mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{R}}}\left[\left(\left\|\nabla_{\tilde{x}}D\left(\tilde{x}\right)\right\|_2 - 1\right)^2\right]$$

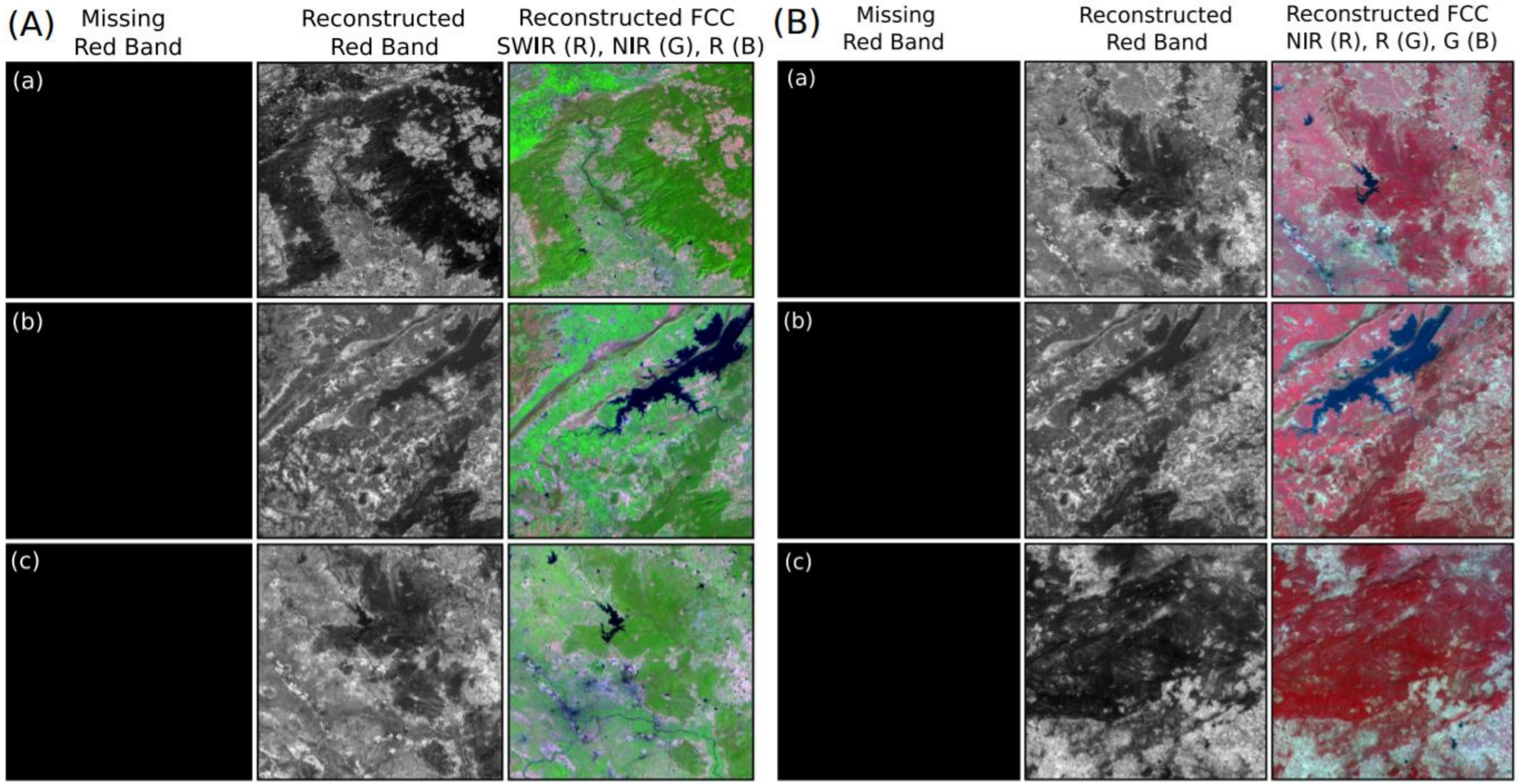


Jin et al. 2020, ICML

Generative Modeling: Missing Data Reconstruction



Generative Modeling: Missing Data Reconstruction

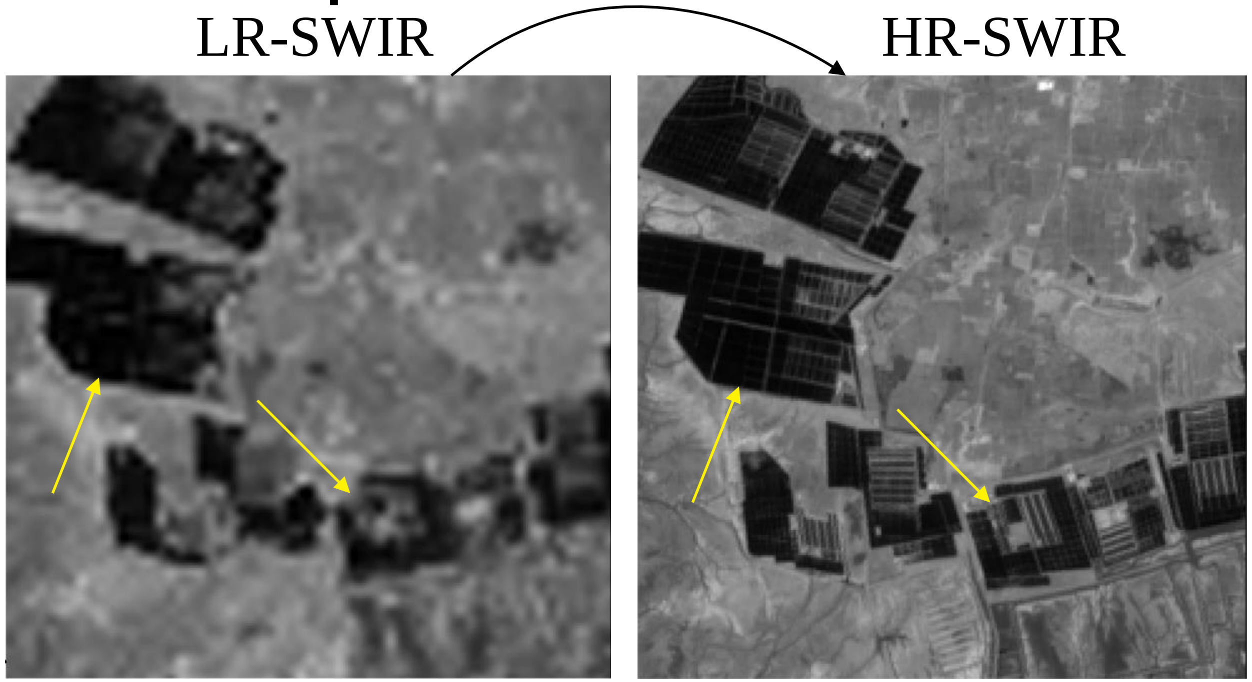


Generative Modeling: Missing Data Reconstruction

Outline

- Maximum Likelihood Estimation
- Reverse KL Divergence
- Introduction to GANs
- GANs in Remote Sensing
    - Problem Formulation

Super-resolution as conditional band synthesis



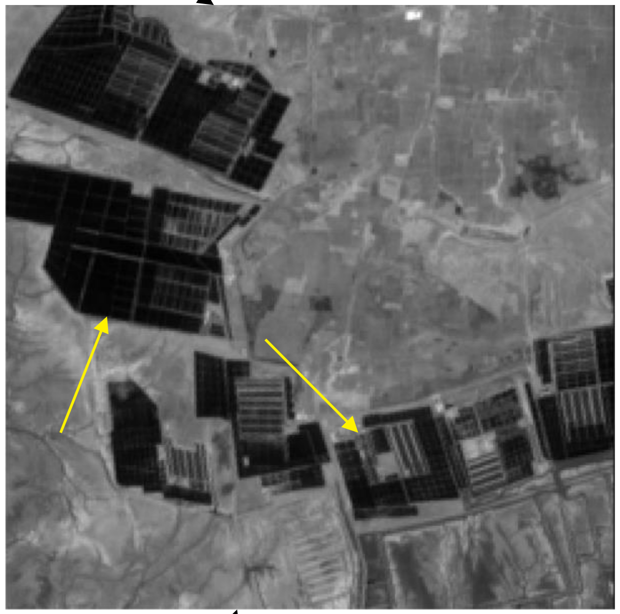
- Direct super-resolution is intractable.
- Lack necessary geometric attributes.

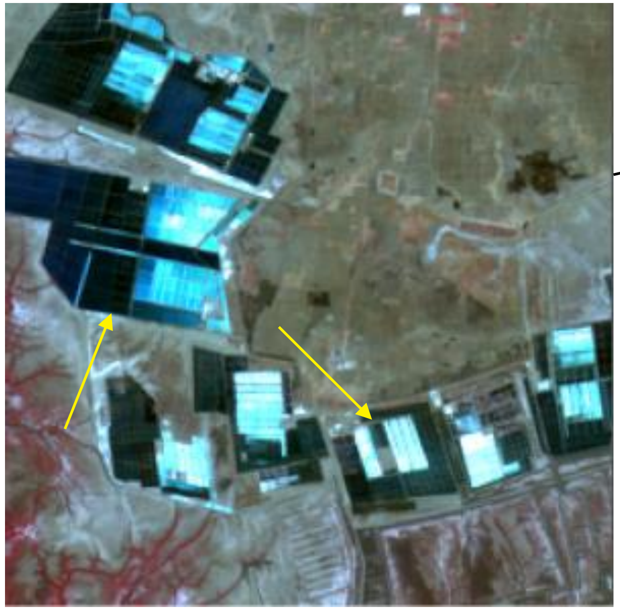
L.R., I. Misra, S. M. Moorthi and D. Dhar, CVPR-EV 2020

Super-resolution as conditional band synthesis

LR-SWIR

HR-SWIR





FCC: NIR (R), R (G), G(B)

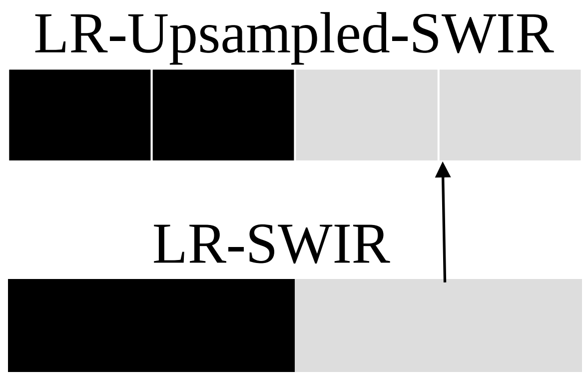
- Direct super-resolution is intractable.
- Lack necessary geometric attributes.

- Reformulate as conditional band synthesis.
- Geometry from existing high resolution bands: HR-NIR, R, G.
- Radiometry from corresponding low resolution band: LR-SWIR.

Traditional Approach

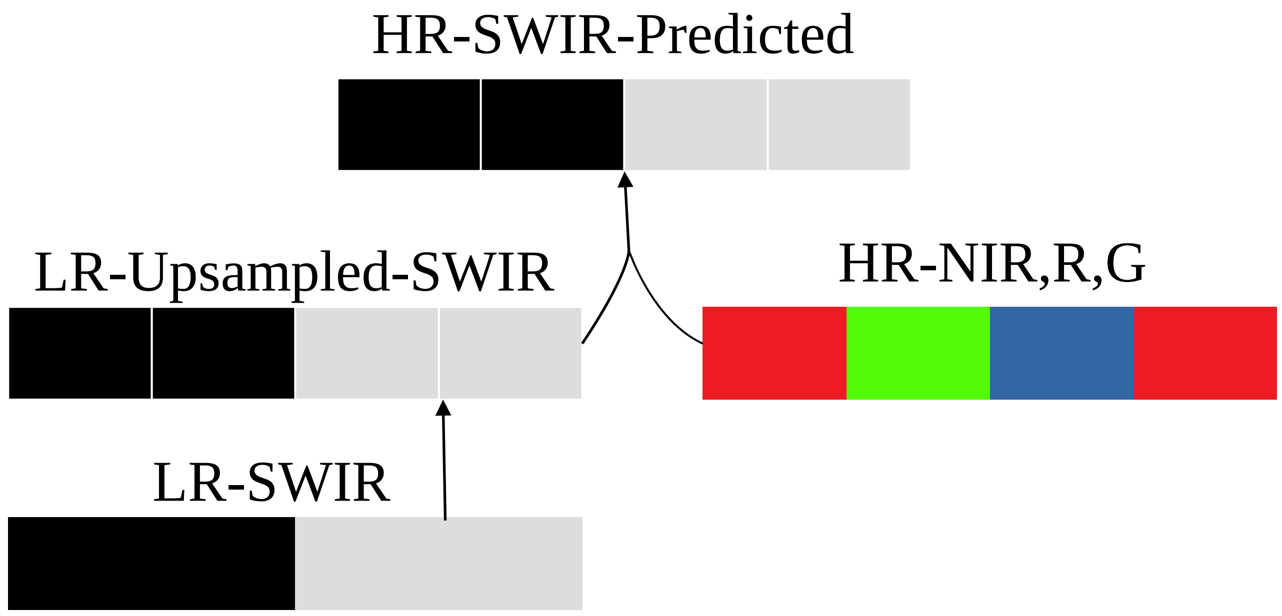
LR-SWIR

Traditional Approach

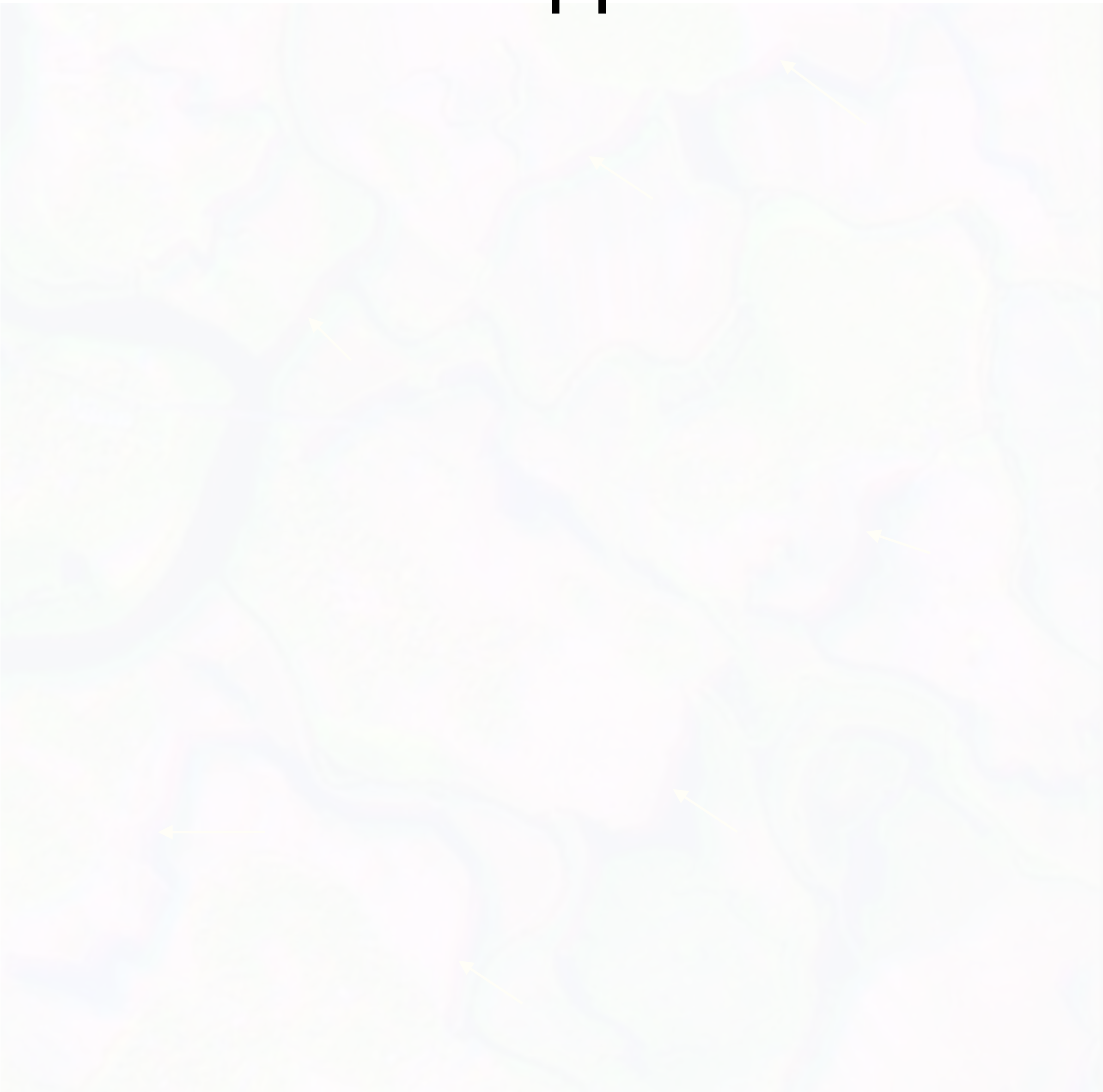


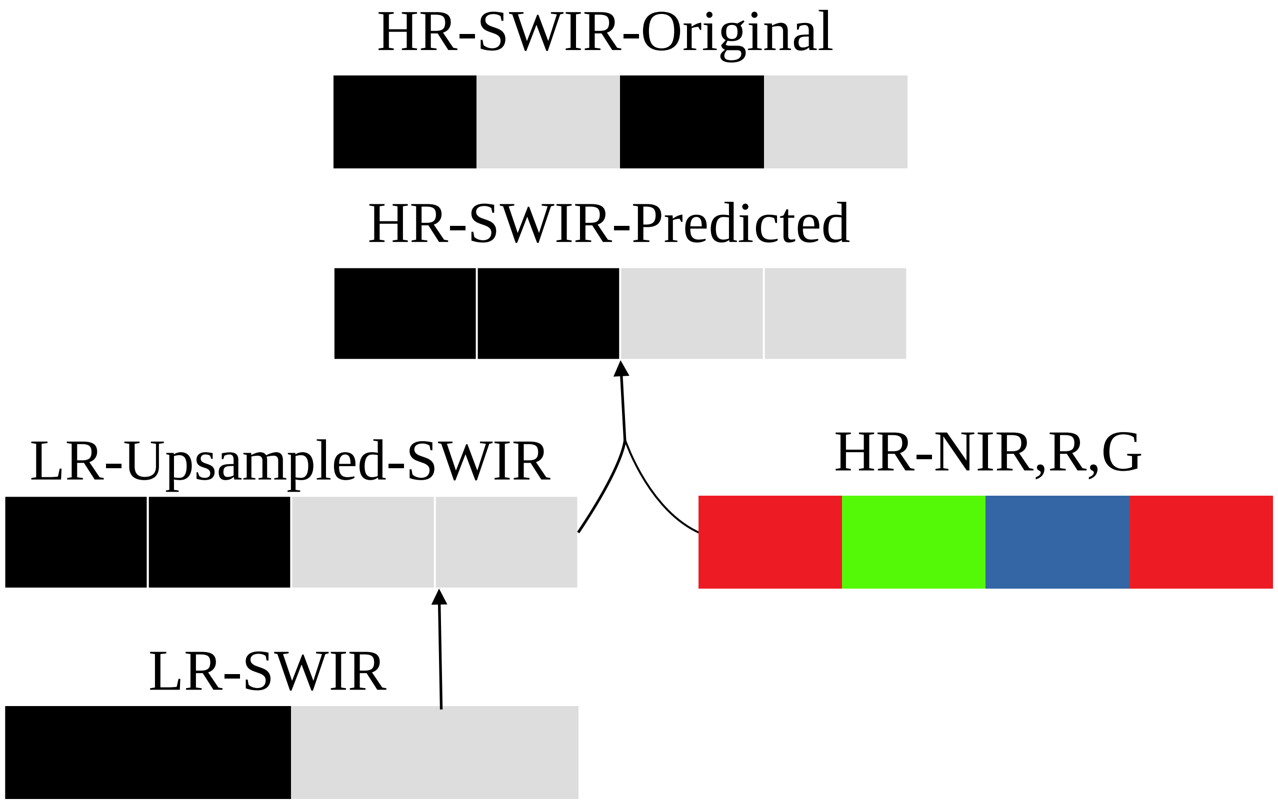
Traditional Approach

Traditional Approach

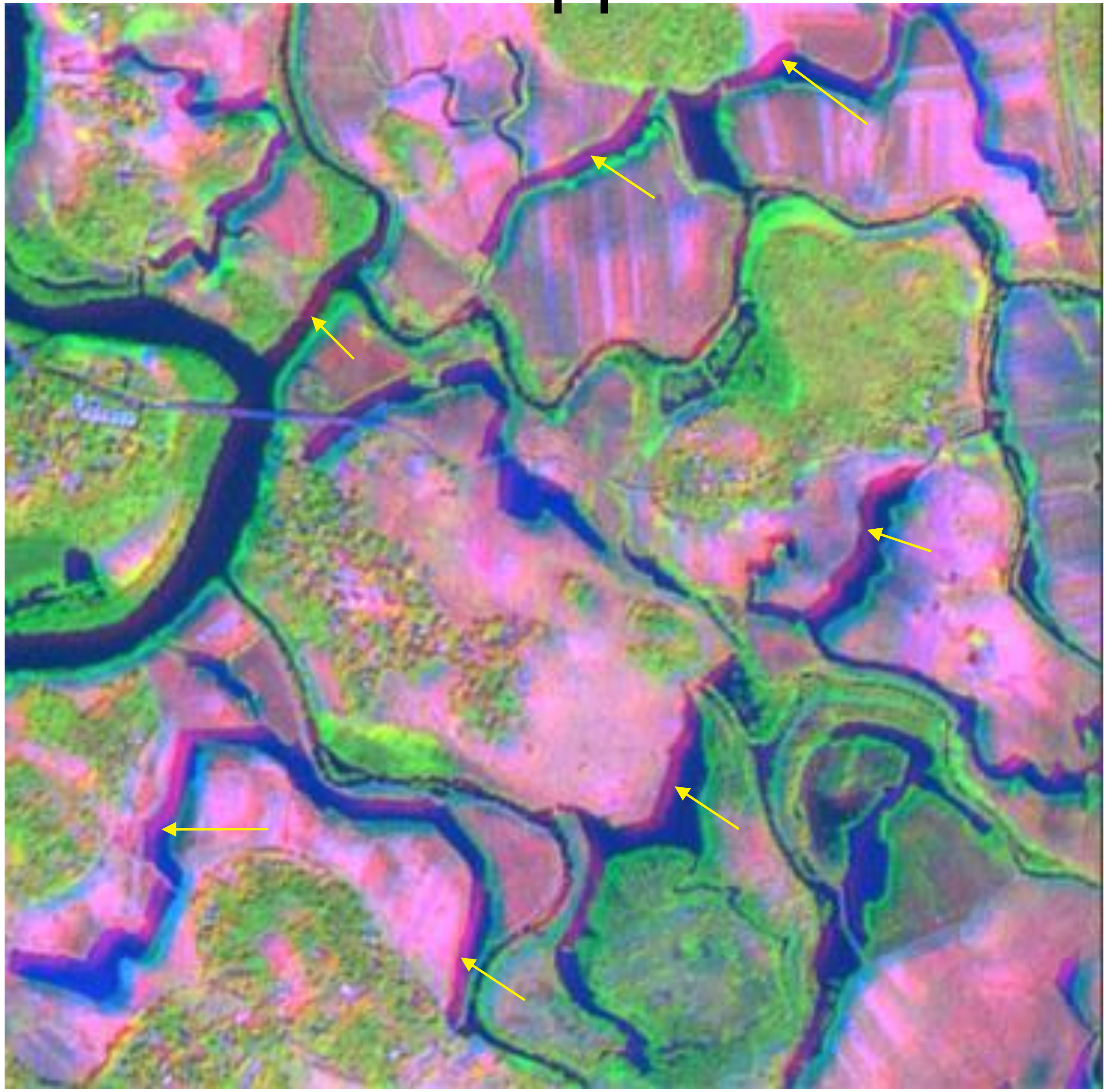


Traditional Approach

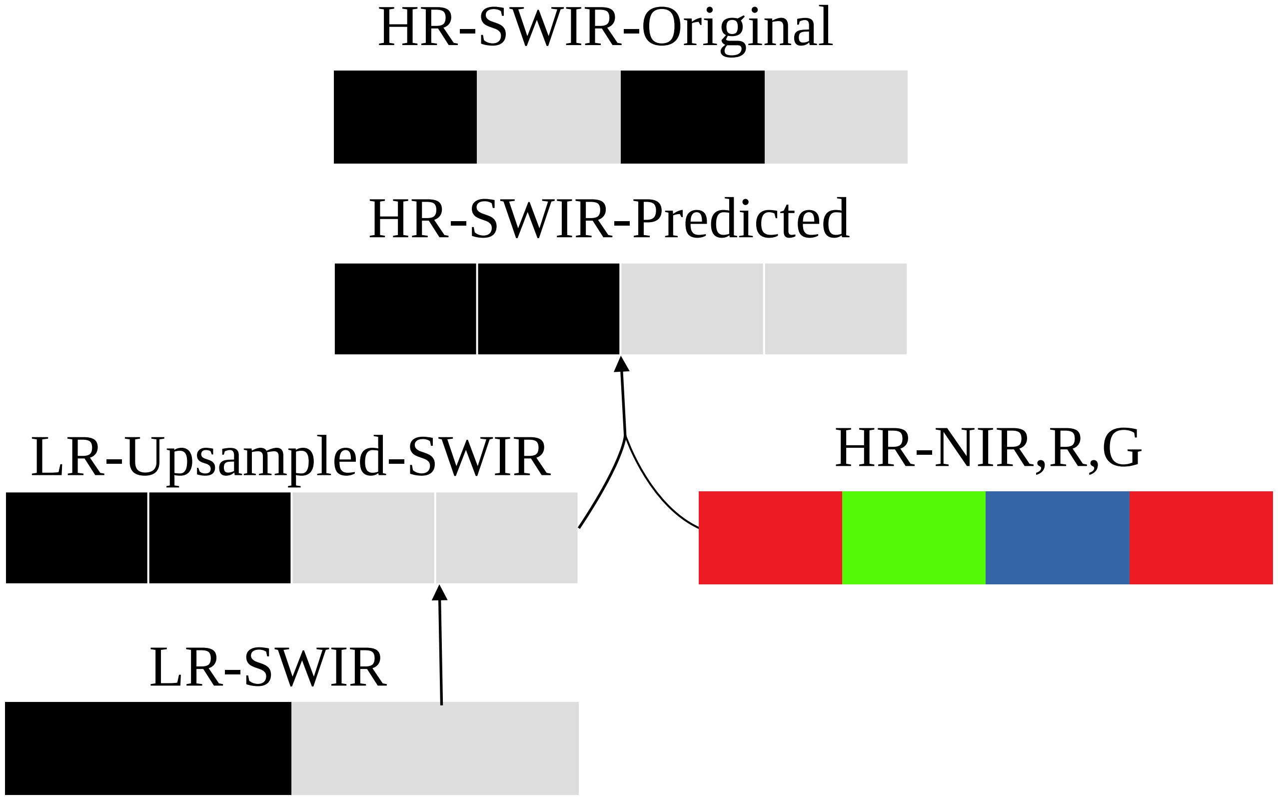




Traditional Approach



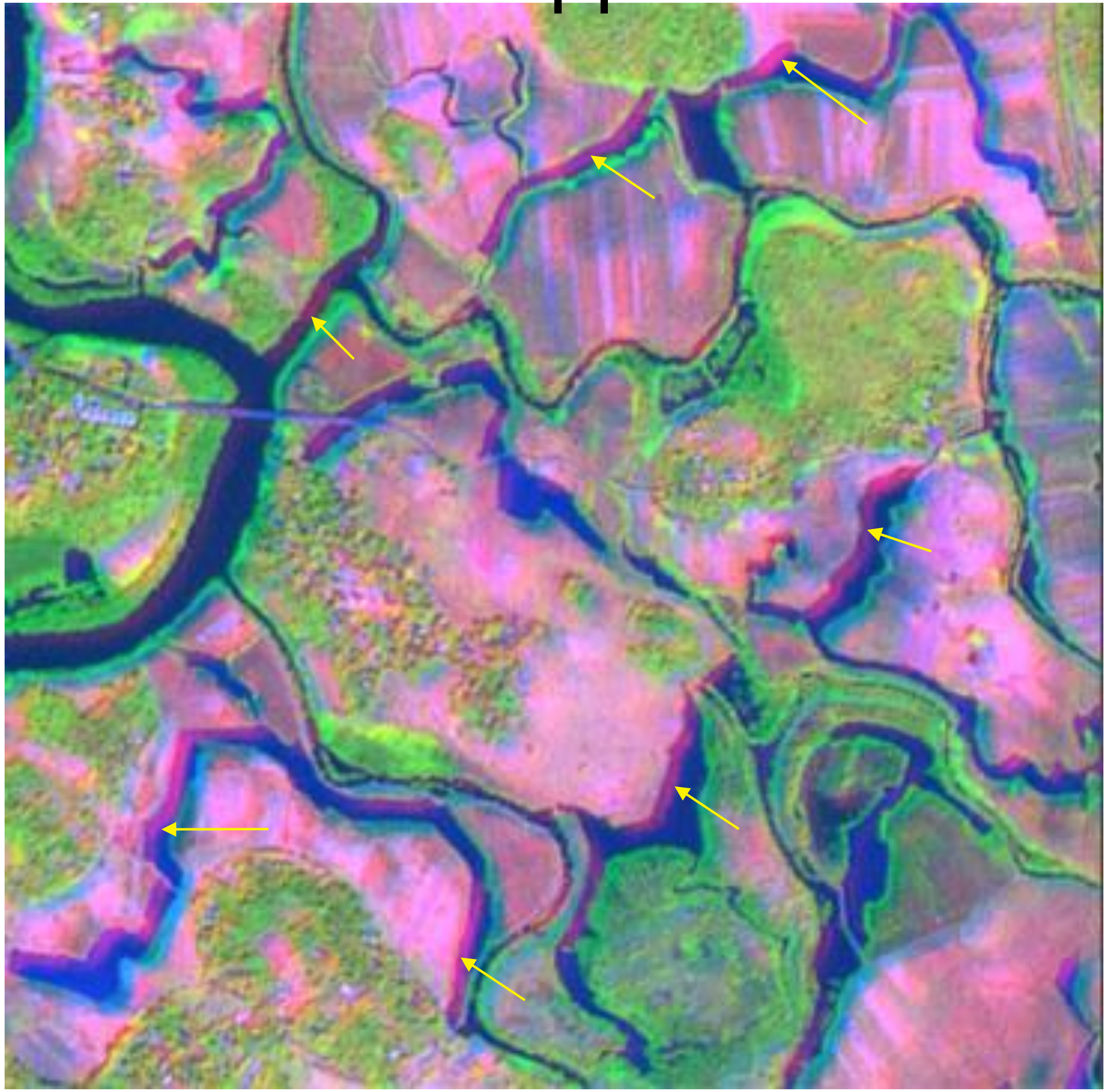
FCC: SWIR (R), NIR (G), Red (B)



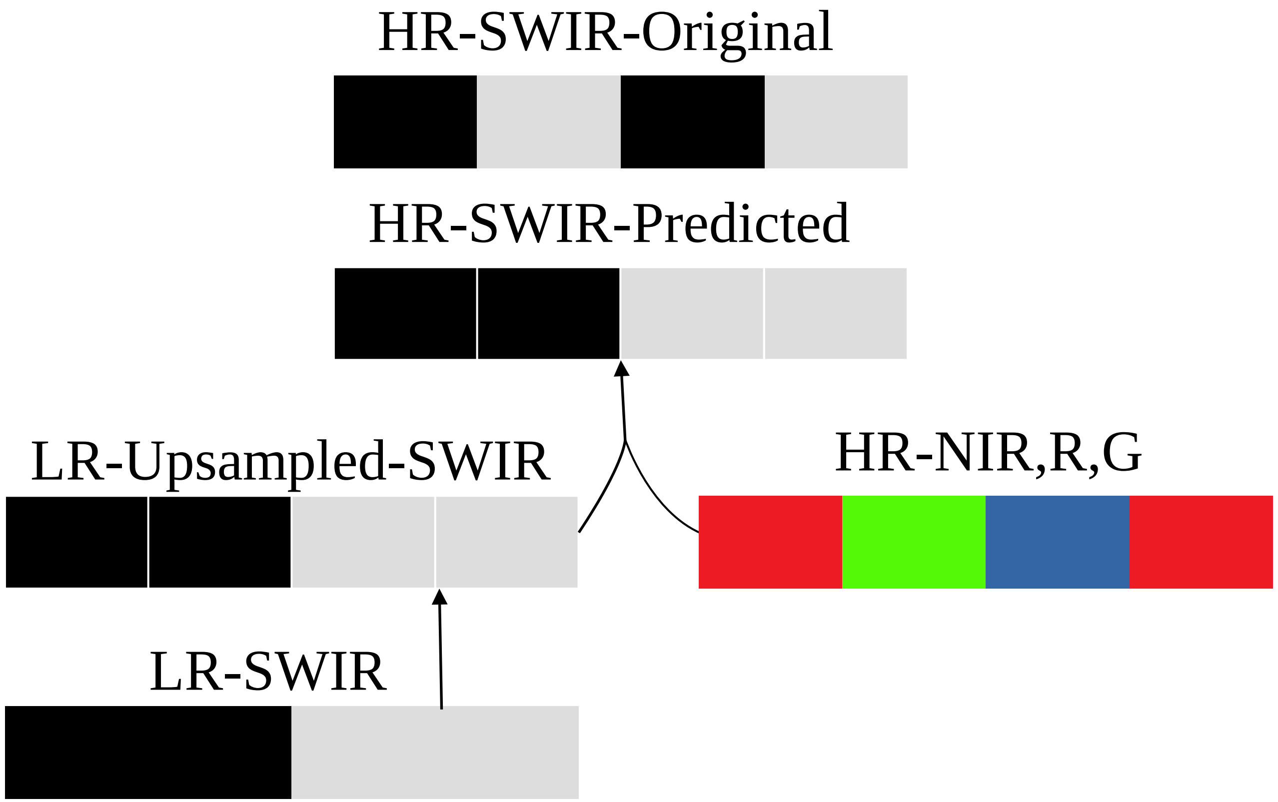
Over dependence on upsampled coarse resolution band results in unpleasant artifacts.

- Geometric distortion
- Radiometric imbalance

Traditional Approach



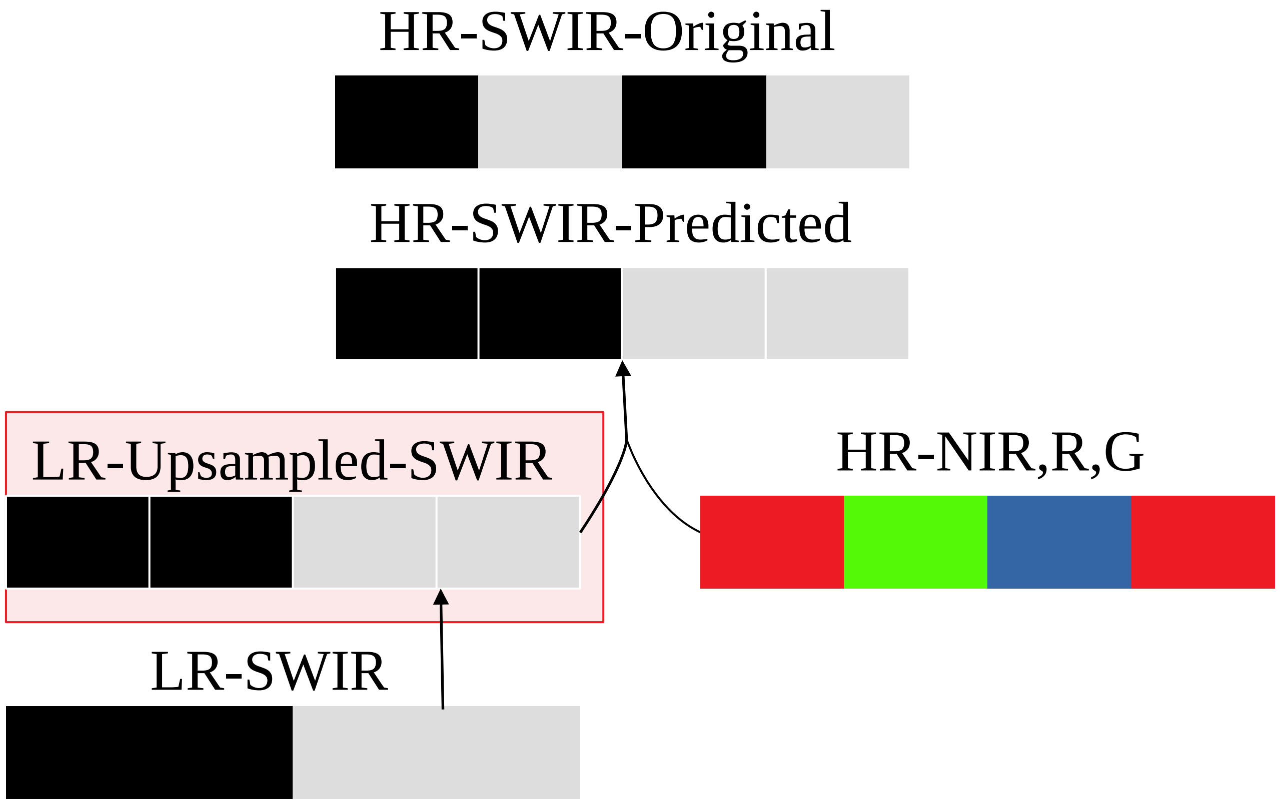
FCC: SWIR (R), NIR (G), Red (B)



Over dependency on upsampled coarse resolution band results in unpleasant artifacts.

- Geometric distortion
- Radiometric imbalance

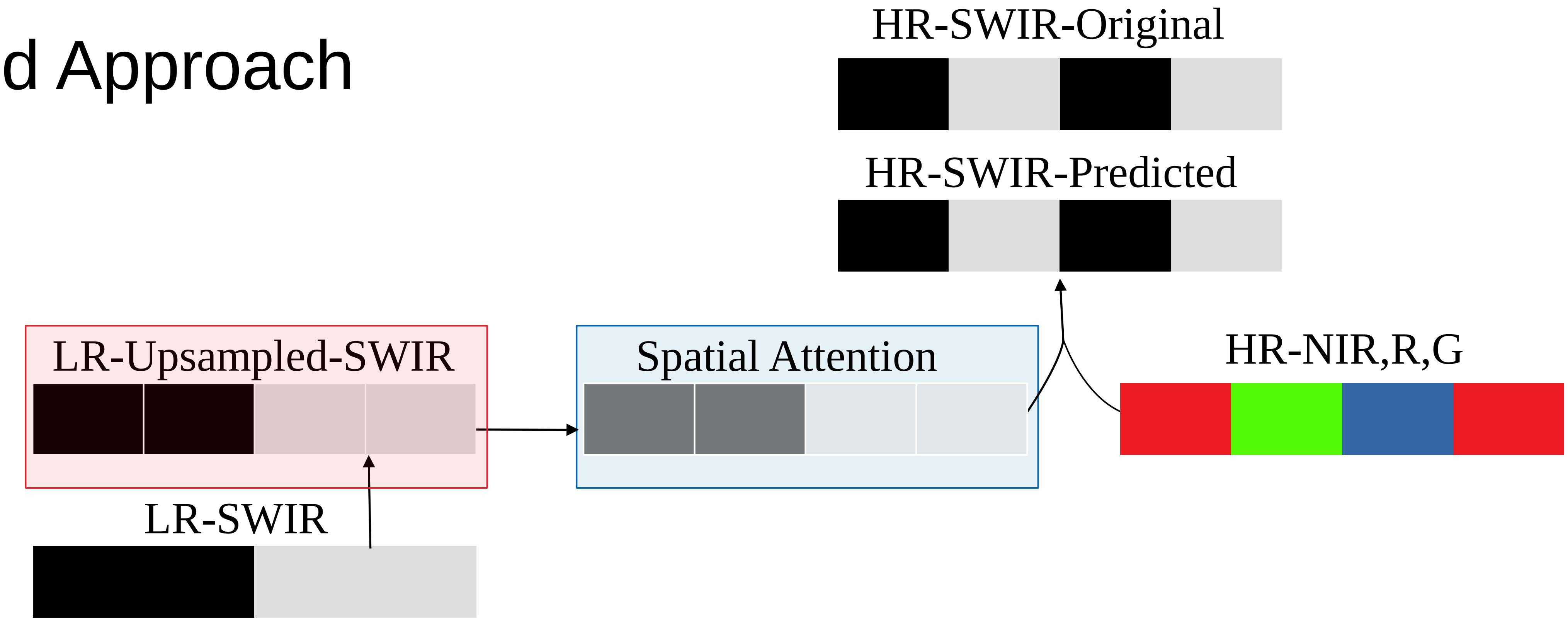
Traditional Approach



Over dependency on upsampled coarse resolution band results in unpleasant artifacts.

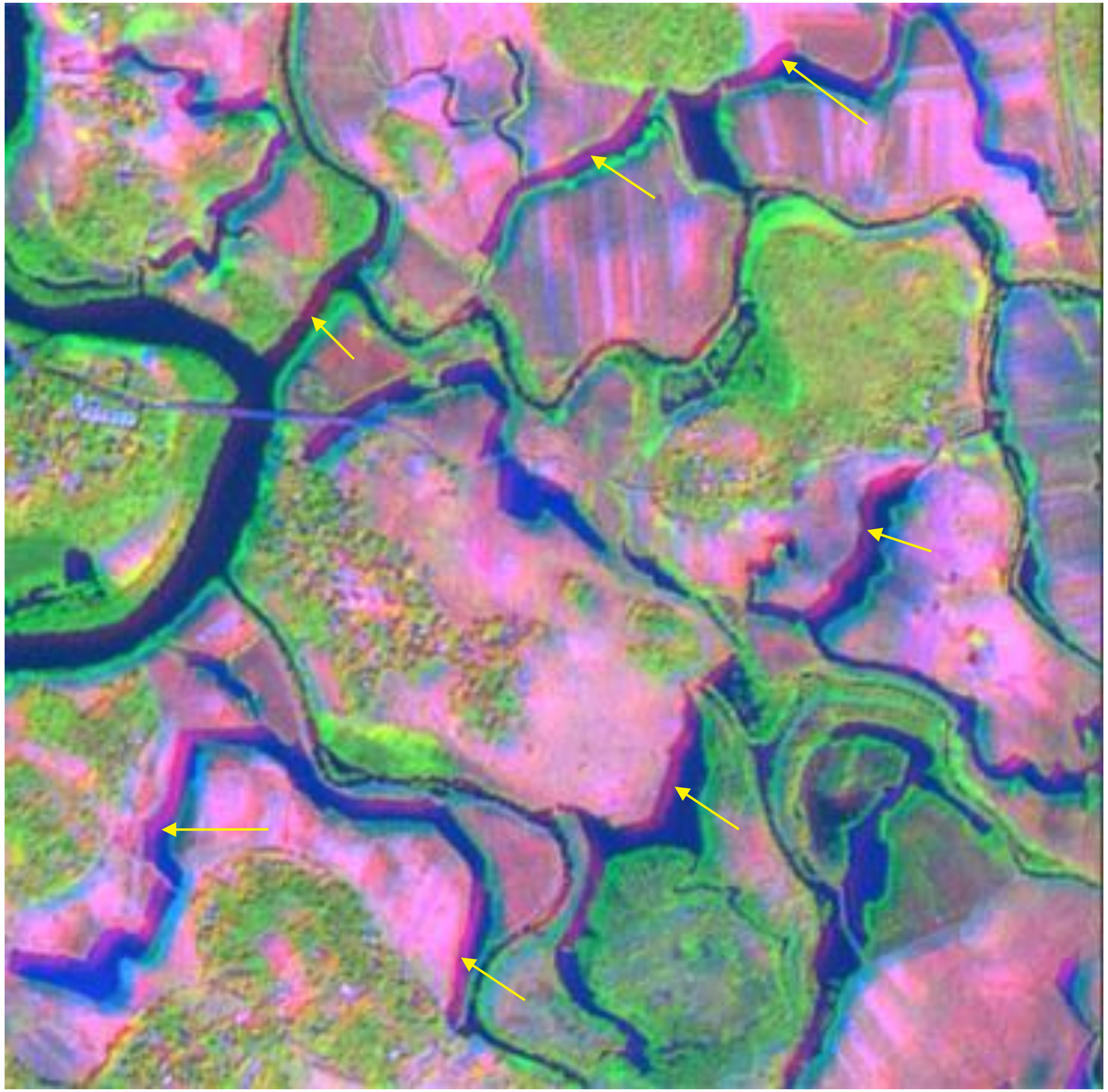
- Geometric distortion
- Radiometric imbalance

Proposed Approach

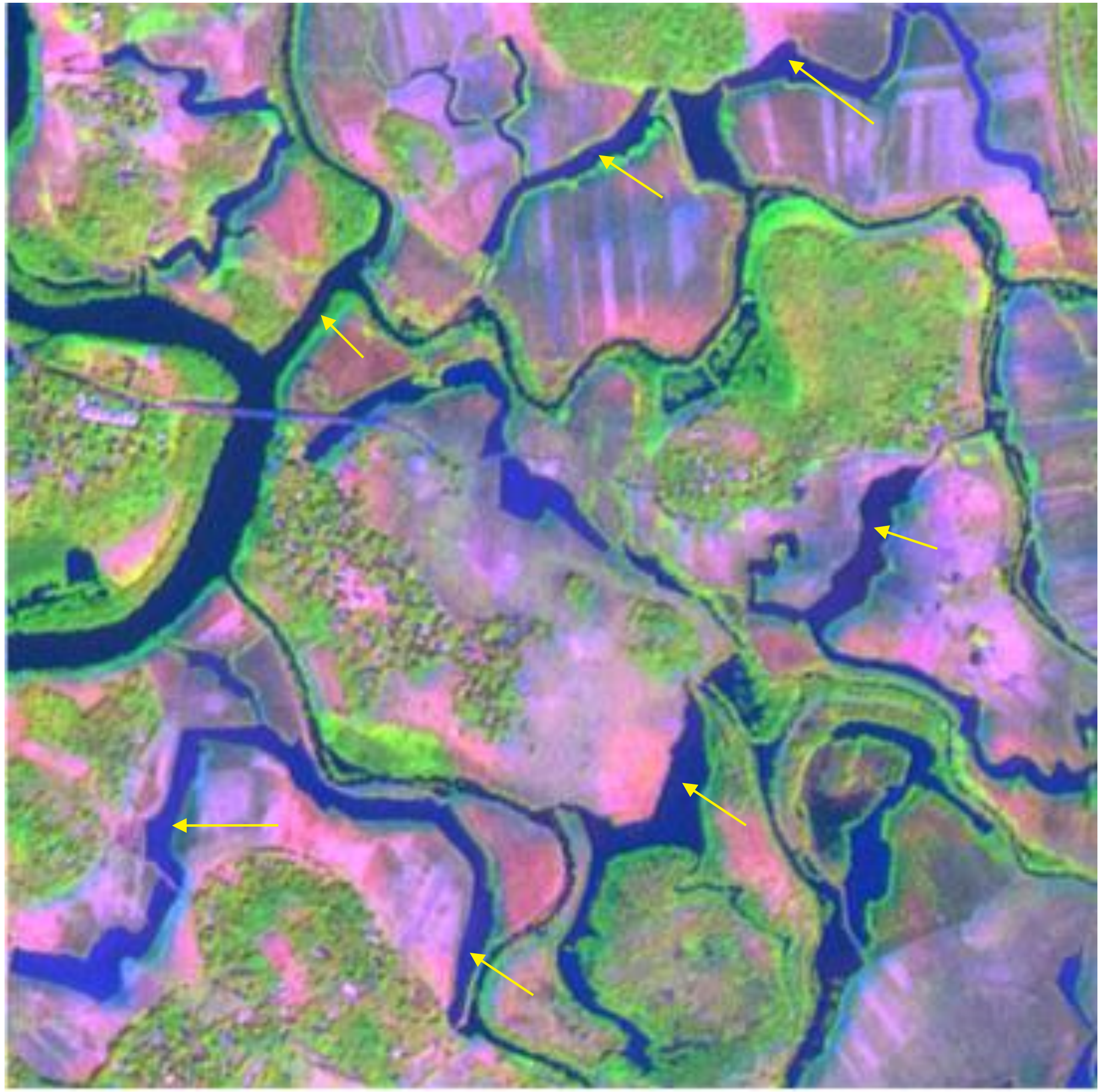


Over dependency on upsampled coarse resolution band can be suppressed by replacing it with spatial attention map.

Traditional Approach

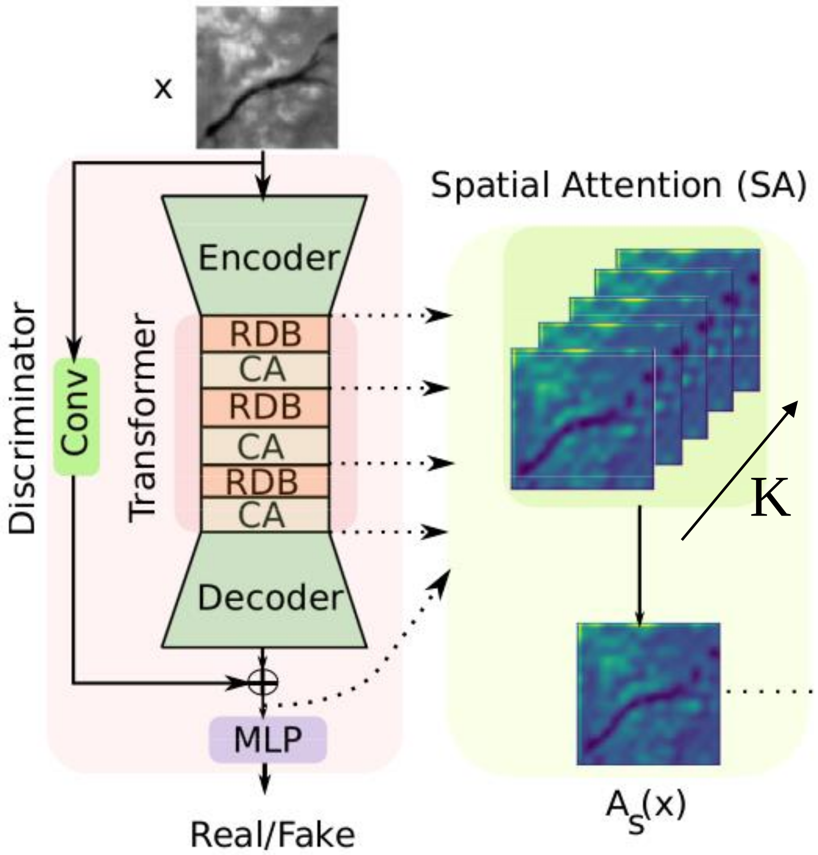


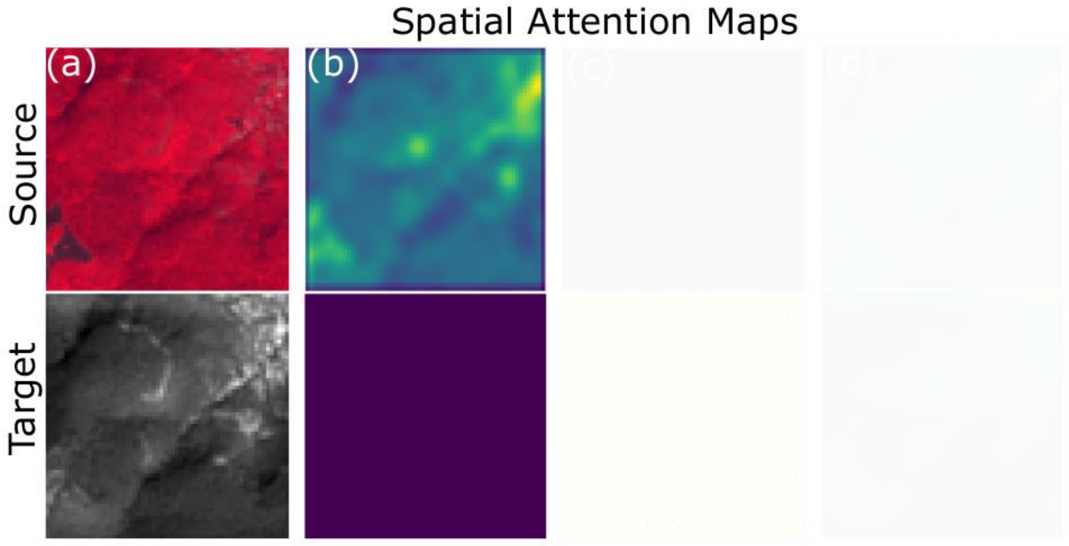
Proposed Approach



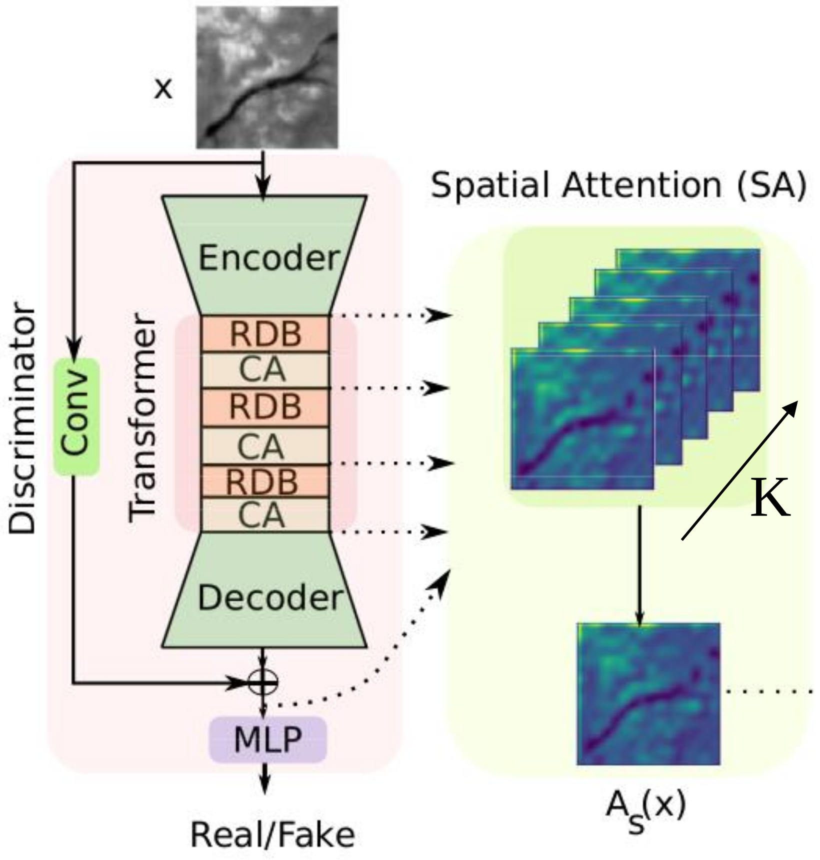
FCC: SWIR (R), NIR (G), Red (B)

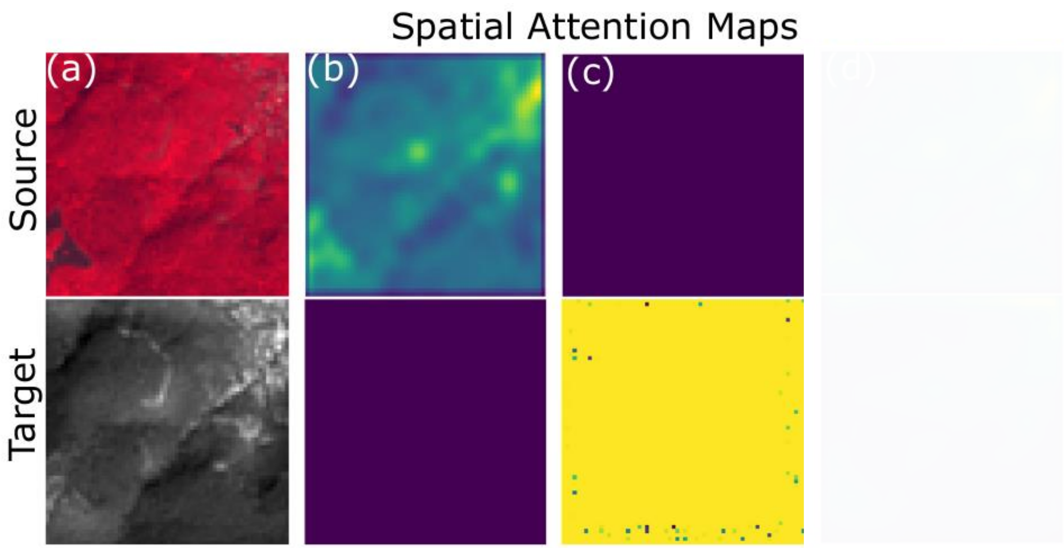
Spatial Attention from Discriminator



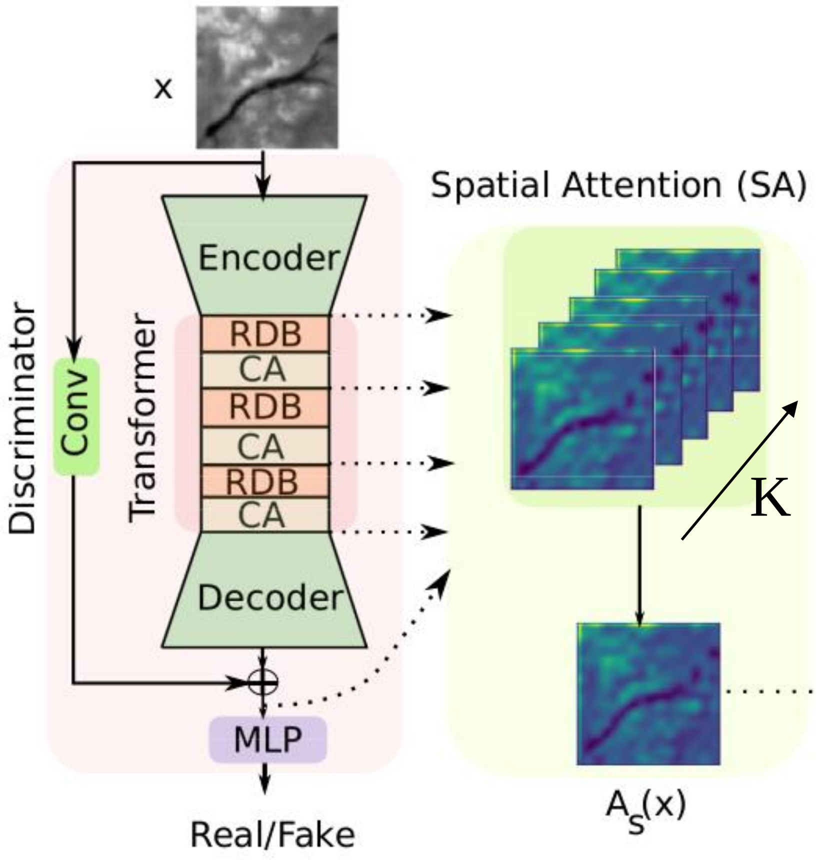


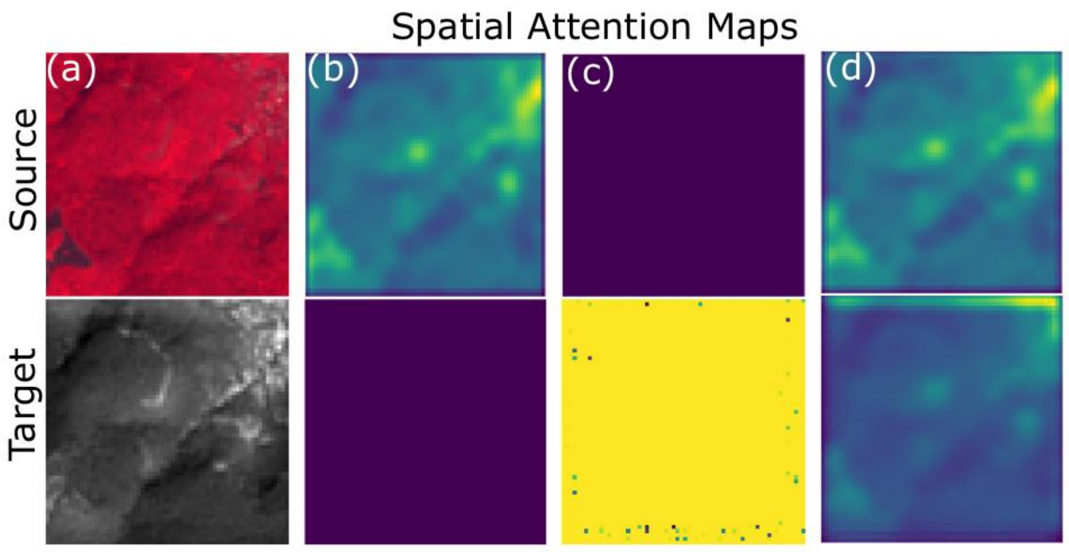
Spatial Attention from Discriminator





Spatial Attention from Discriminator

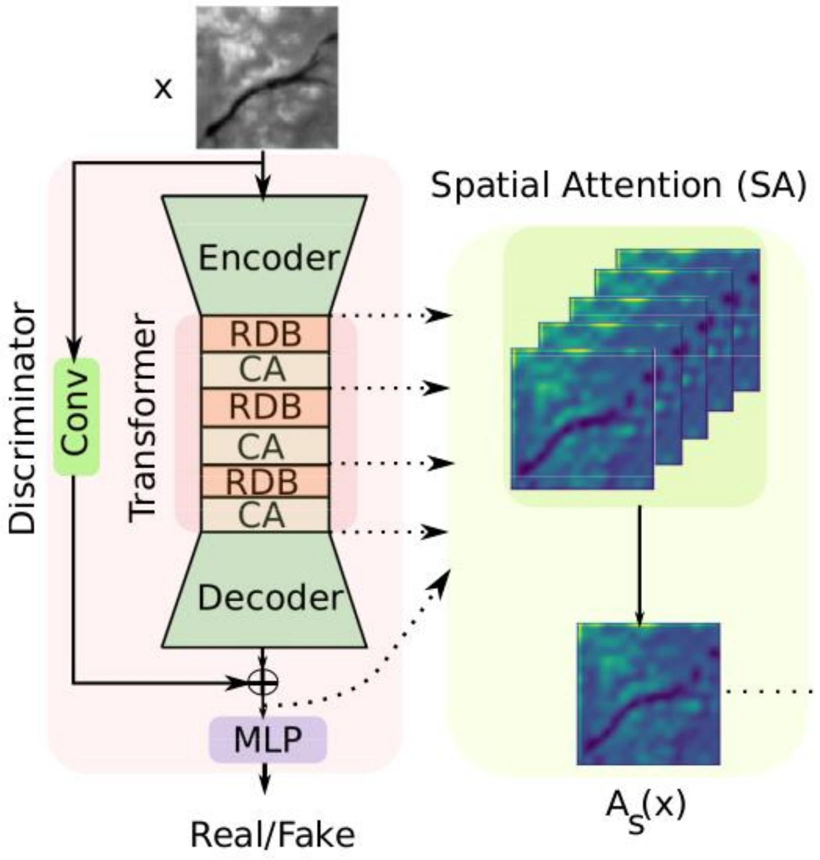




$$A_s(x) = \mathcal{N}\left(D_s(x)\right),$$

$$D_s(x) = \sum_{i=1}^{K} \mathcal{N}\left(\sum_{j=1}^{C} |A_{ij}(x)|\right)$$

Spatial Attention from Discriminator



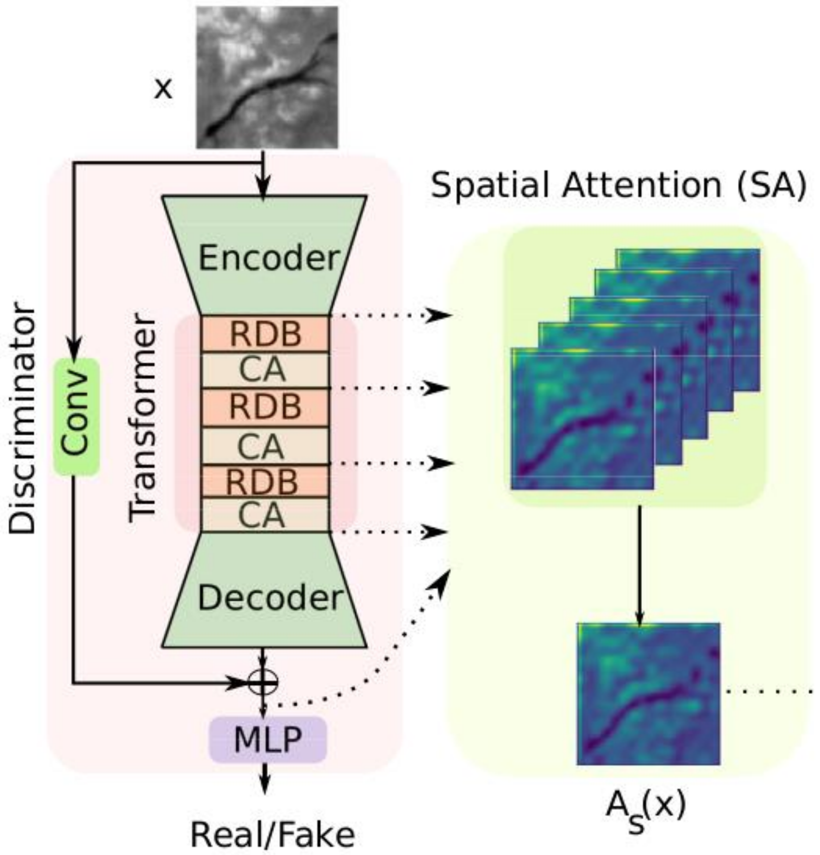
Spatial Attention Loss

$$\mathscr{L}_{sa} = \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}},\, y\sim\mathbb{P}_y}\left[\|A_s(\hat{x}) - A_s(y)\|_2^2\right]$$

Domain Adaptation Loss

$$\mathscr{L}_{da} = \mathbb{E}_{\tilde{y}\sim\mathbb{P}_{\tilde{y}},\, y\sim\mathbb{P}_y}\left[\|A_s(\tilde{y}) - A_s(y)\|_2^2\right]$$

Spatial Attention from Discriminator



Spatial Attention Loss

$$\mathcal{L}_{sa} = \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}},y\sim\mathbb{P}_y}\left[\|A_s(\hat{x}) - A_s(y)\|_2^2\right]$$

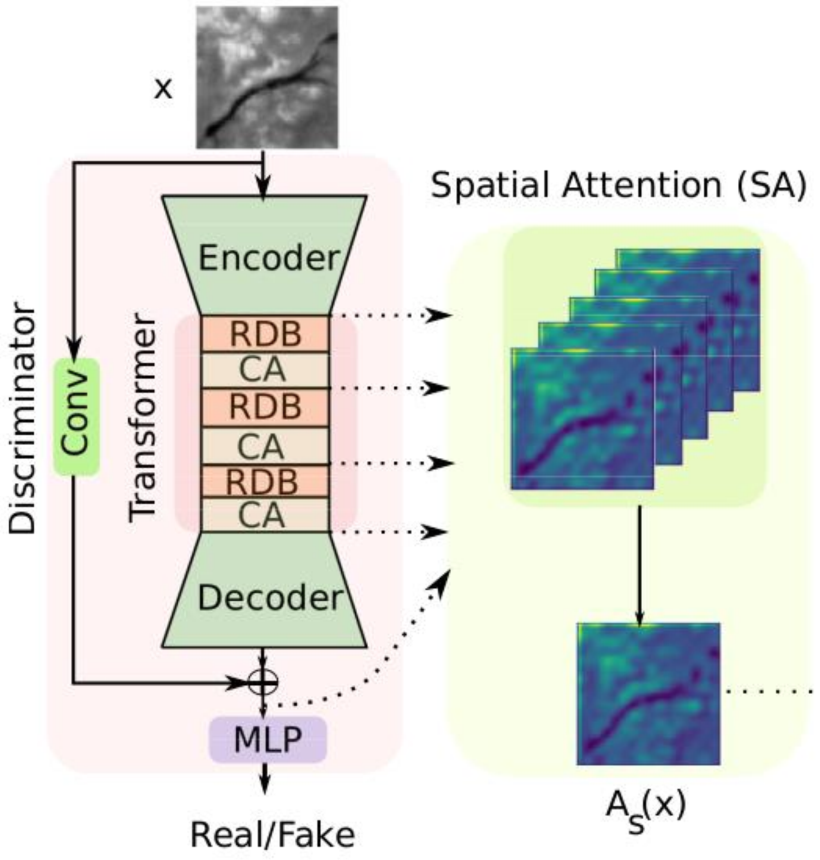
Domain Adaptation Loss

$$\mathcal{L}_{da} = \mathbb{E}_{\tilde{y}\sim\mathbb{P}_{\tilde{y}},y\sim\mathbb{P}_y}\left[\|A_s(\tilde{y}) - A_s(y)\|_2^2\right]$$

Discriminator Objective

$$\min_D \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}[D(\hat{x})] - \mathbb{E}_{x\sim\mathbb{P}_x}[D(x)]$$
$$+ \lambda_{gp}\mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}}}\left[(\|\nabla_{\tilde{x}}D(\tilde{x})\|_2 - 1)^2\right]$$
$$+ \lambda_{sa}\mathcal{L}_{sa} + \lambda_{da}\mathcal{L}_{da},$$

Spatial Attention from Discriminator



Spatial Attention Loss

$$\mathcal{L}_{sa} = \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}},y\sim\mathbb{P}_y}\left[\|A_s(\hat{x}) - A_s(y)\|_2^2\right]$$

Domain Adaptation Loss

$$\mathcal{L}_{da} = \mathbb{E}_{\tilde{y}\sim\mathbb{P}_{\tilde{y}},y\sim\mathbb{P}_y}\left[\|A_s(\tilde{y}) - A_s(y)\|_2^2\right]$$

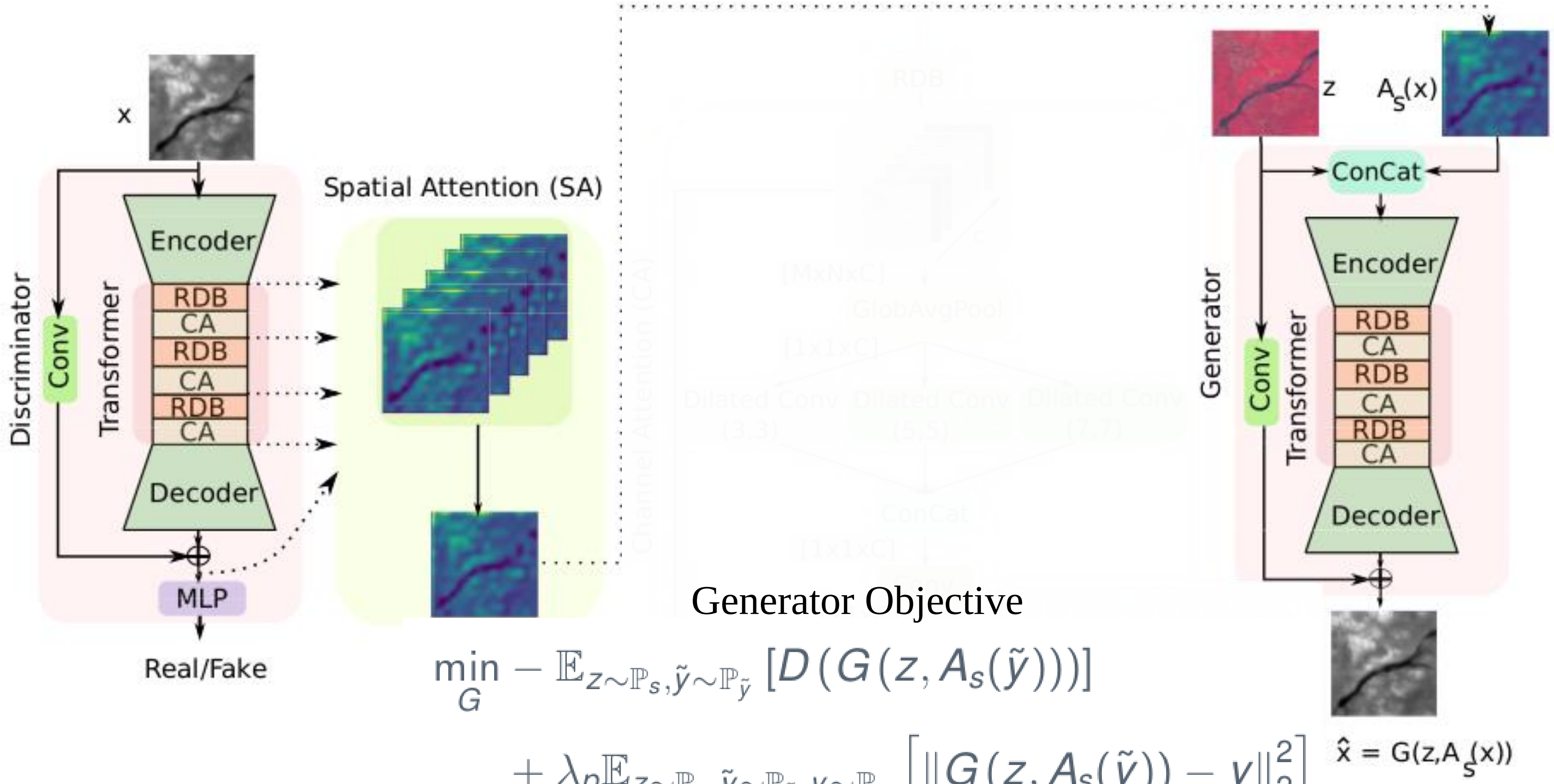
Discriminator Objective

$$\min_D \mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[D(\hat{x})\right] - \mathbb{E}_{x\sim\mathbb{P}_x}\left[D(x)\right]$$

$$+ \lambda_{gp}\mathbb{E}_{\tilde{x}\sim\mathbb{P}_{\tilde{x}}}\left[\left(\|\nabla_{\tilde{x}}D(\tilde{x})\|_2 - 1\right)^2\right]$$

$$+ \lambda_{sa}\mathcal{L}_{sa} + \lambda_{da}\mathcal{L}_{da},$$

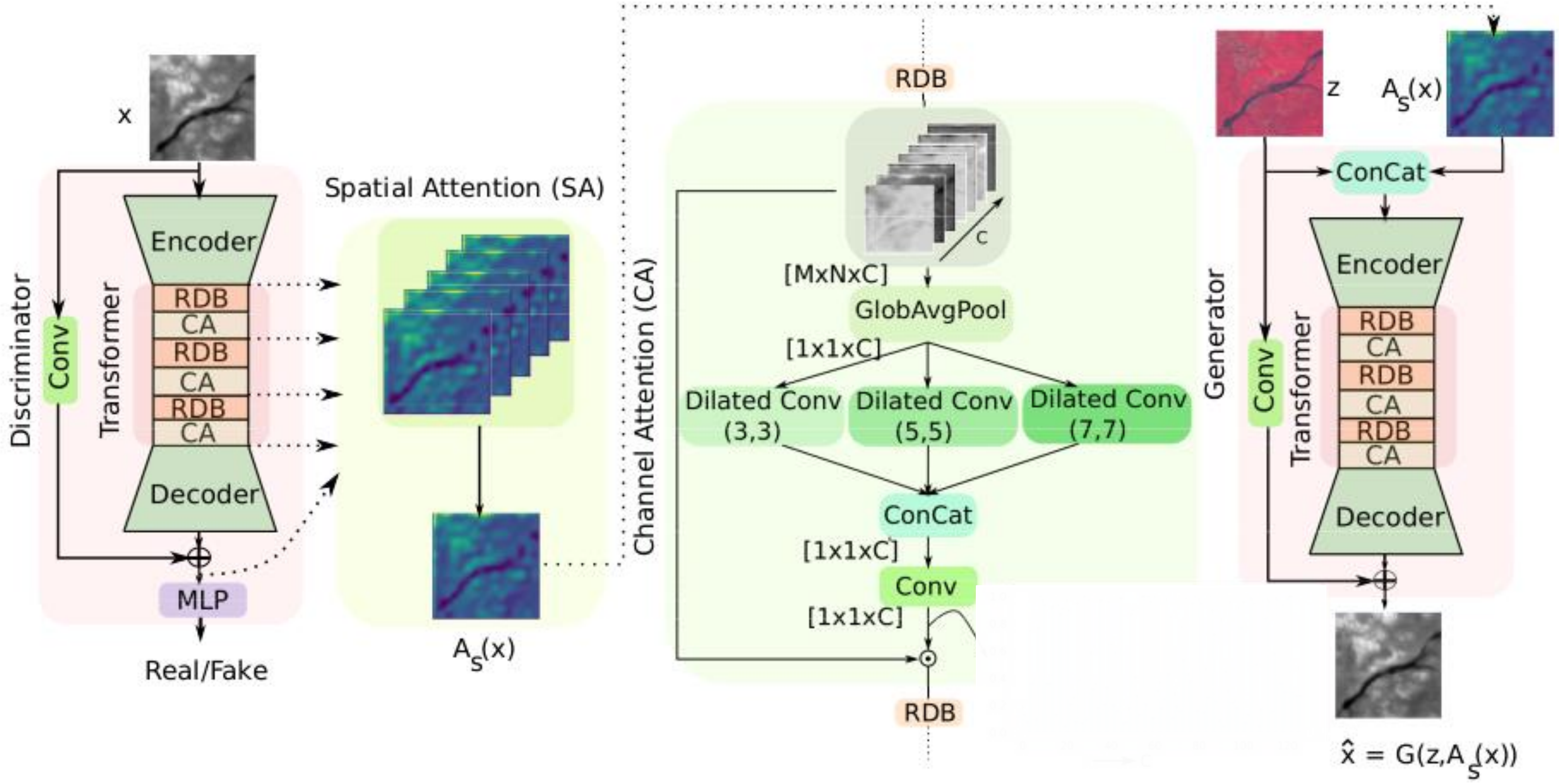
Spatial Attention from Discriminator



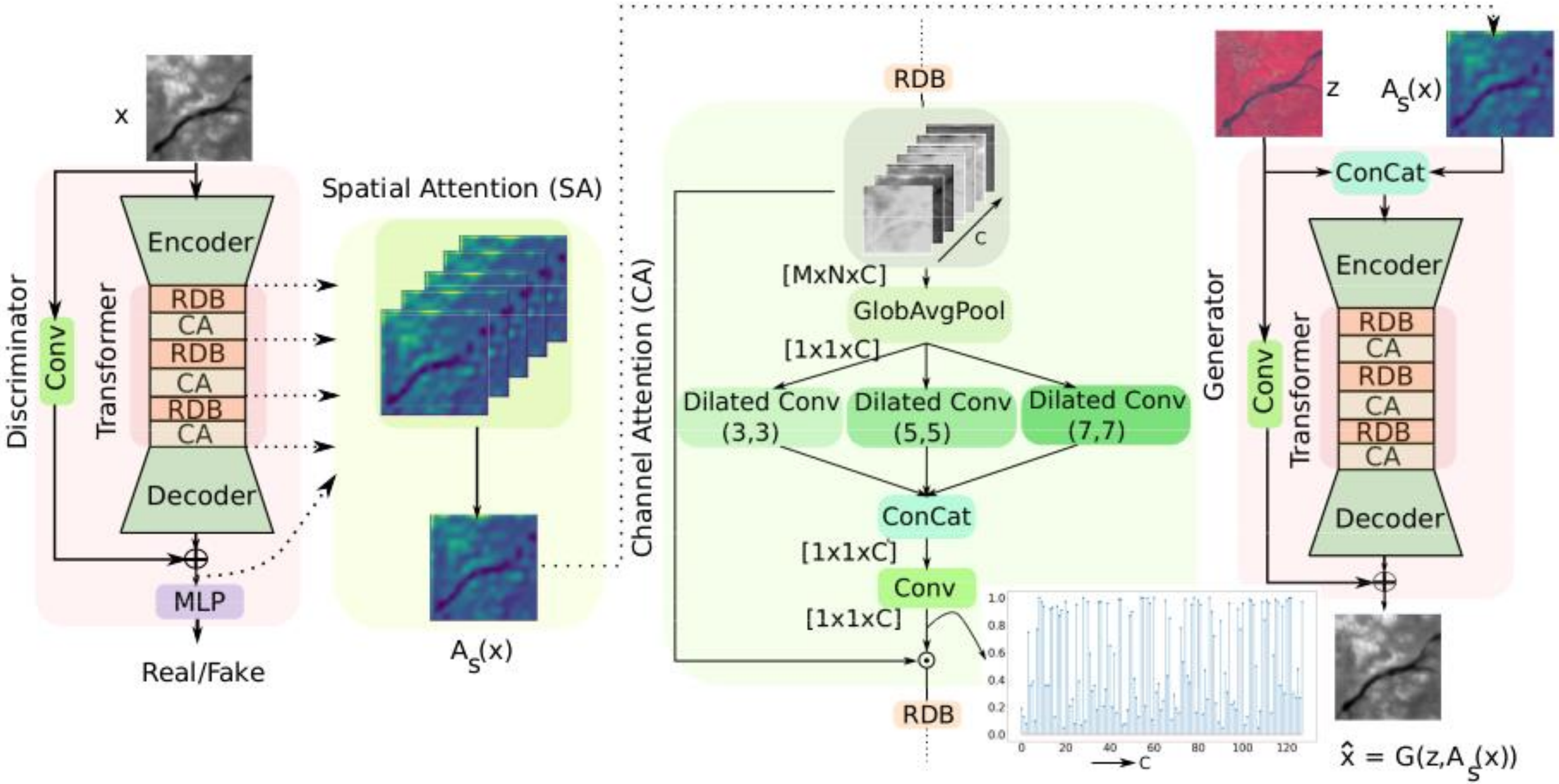
Generator Objective

$$\min_{G} -\mathbb{E}_{z\sim\mathbb{P}_s,\tilde{y}\sim\mathbb{P}_{\tilde{y}}}\left[D\left(G\left(z, A_s(\tilde{y})\right)\right)\right]$$
$$+\lambda_p\mathbb{E}_{z\sim\mathbb{P}_s,\tilde{y}\sim\mathbb{P}_{\tilde{y}},y\sim\mathbb{P}_y}\left[\left\|G\left(z, A_s(\tilde{y})\right)-y\right\|_2^2\right]$$

Spatio-Spectral Laplacian Attention

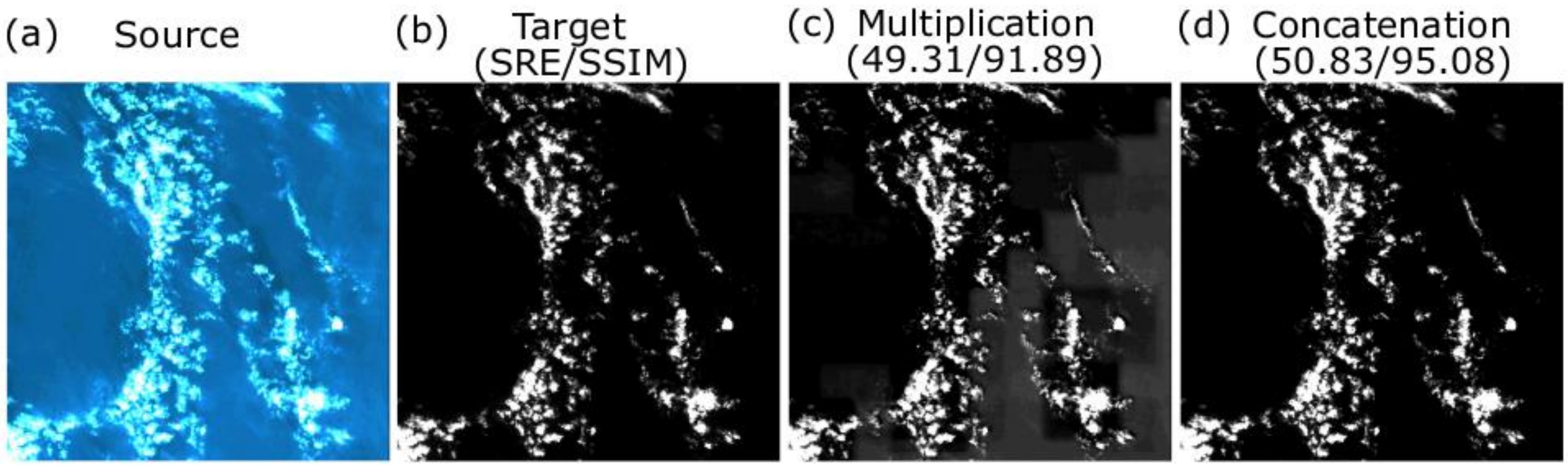


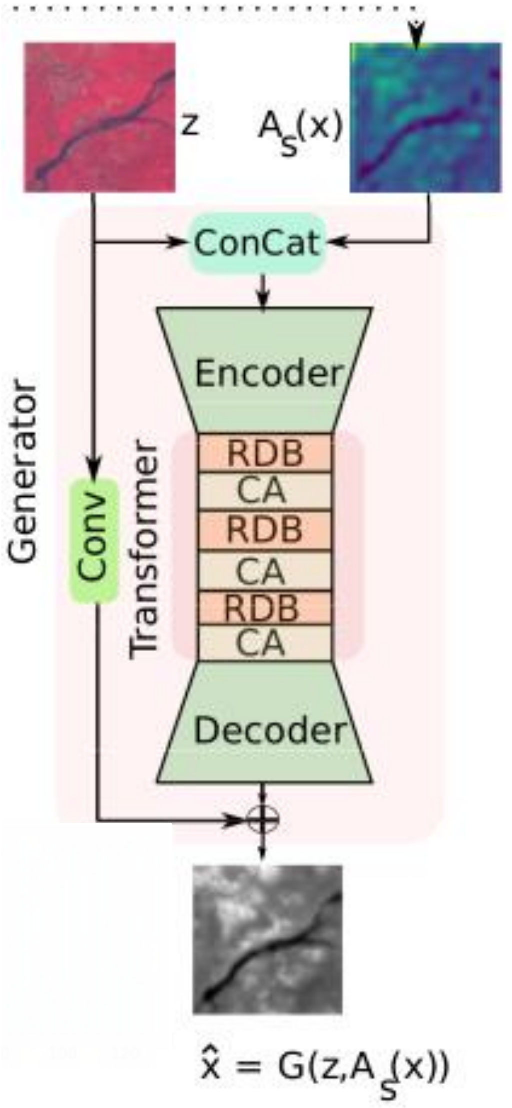
Spatio-Spectral Laplacian Attention



Spectral attention coefficients

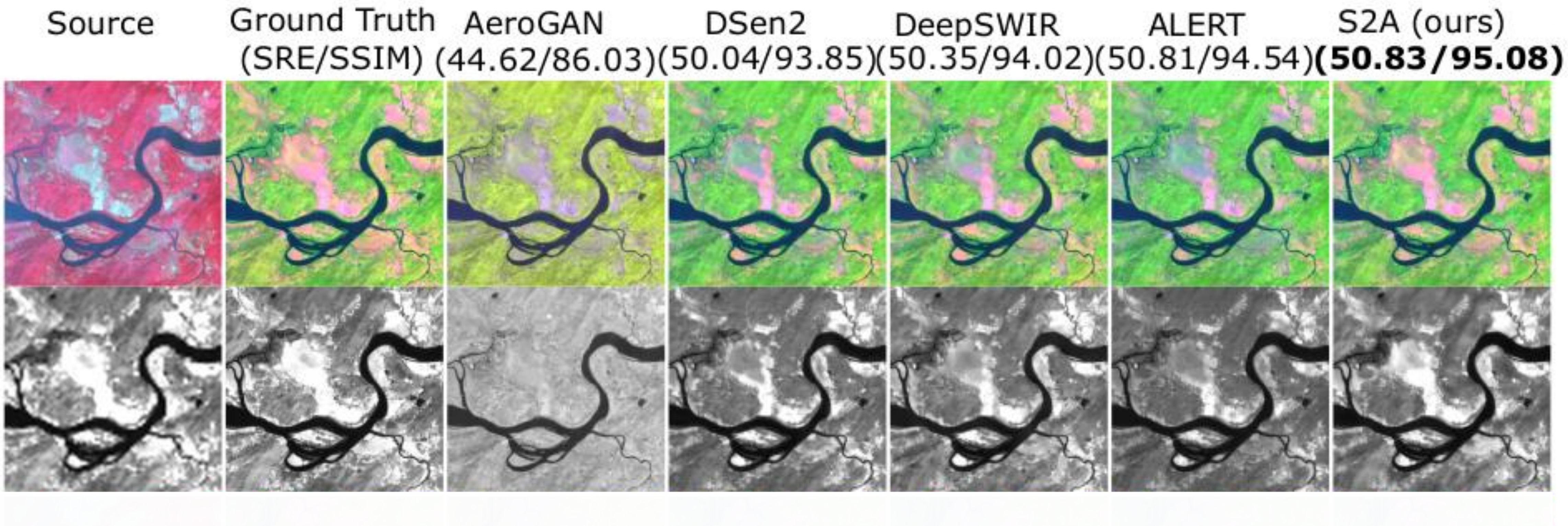
Combining Spatial Attention with Source Bands



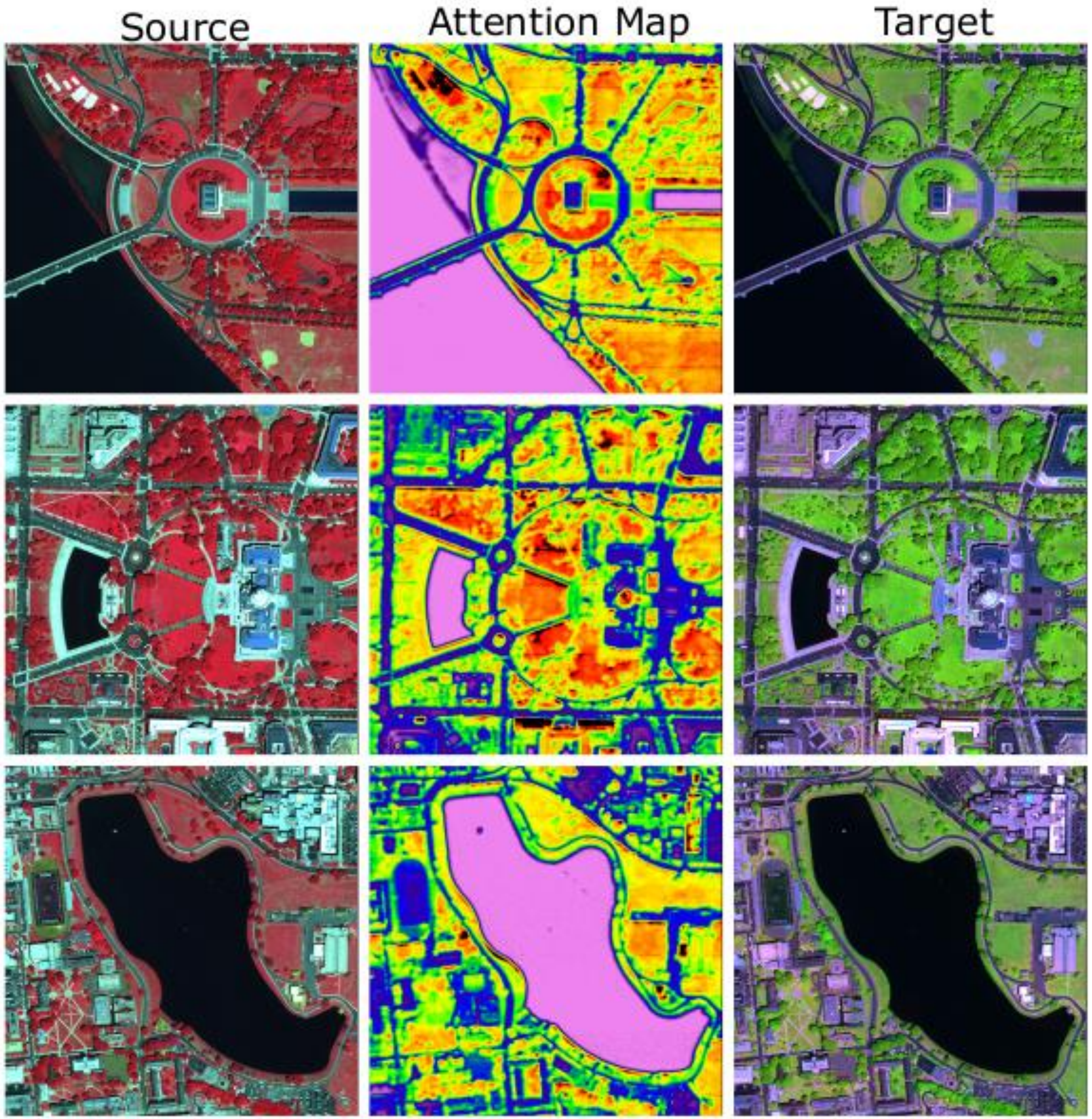


Multiplication:

- Attention module latches on to bright targets.
- Synthesized band contains blocky artifacts.

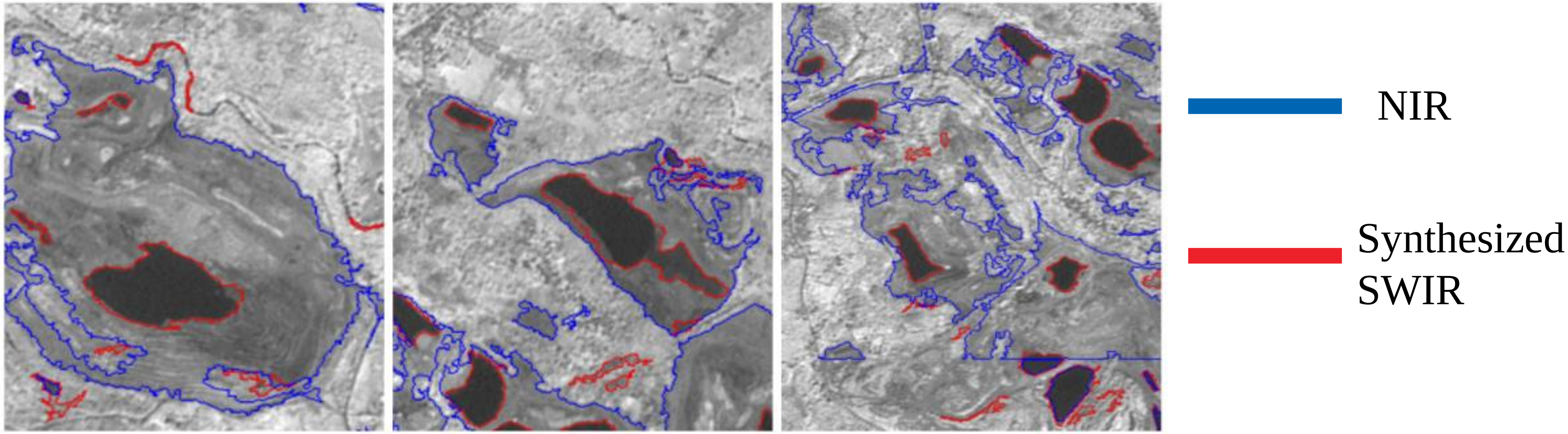


Method	RMSE	SSIM(%)	SRE(dB)	PSNR(dB)	SAM(deg)
AeroGAN [31]	21.62	86.03	44.62	36.50	12.15
DSen2 [21]	14.14	93.85	50.04	41.94	7.88
DeepSWIR [33]	13.75	94.02	50.35	42.27	7.66
ALERT [32]	12.97	94.54	50.81	42.80	7.48
S2A (ours)	11.74	95.08	50.83	42.76	6.87



- Learns to attend to relevant parts of source imagery.
- Homogeneous and heterogeneous targets are discernible.
- Similar features have similar attention coefficients

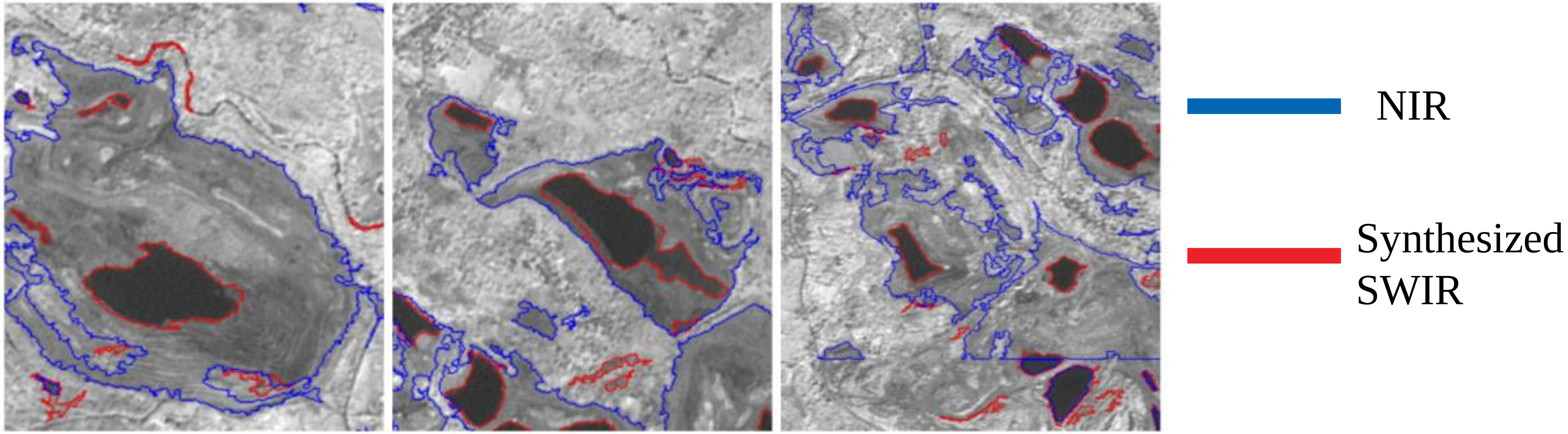
Wetland Delineation



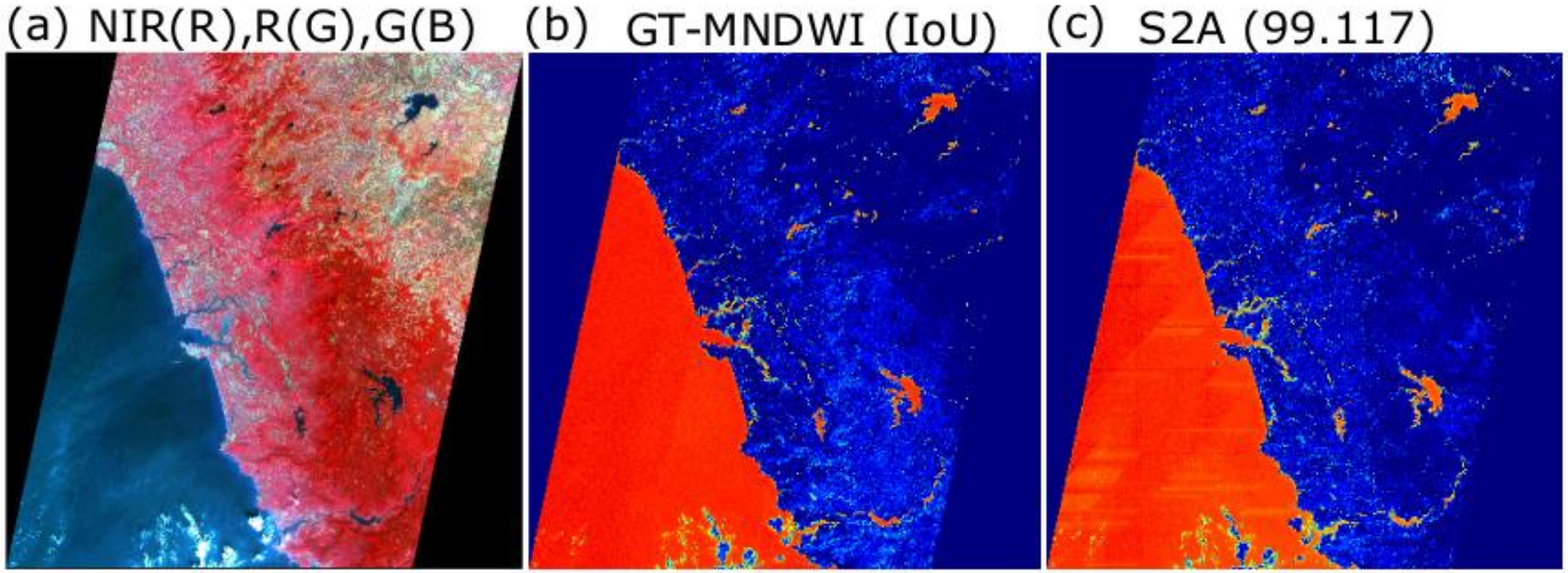
Water Segmentation

(a) NIR(R),R(G),G(B) (b) GT-MNDWI (IoU) (c) S2A (99.117)

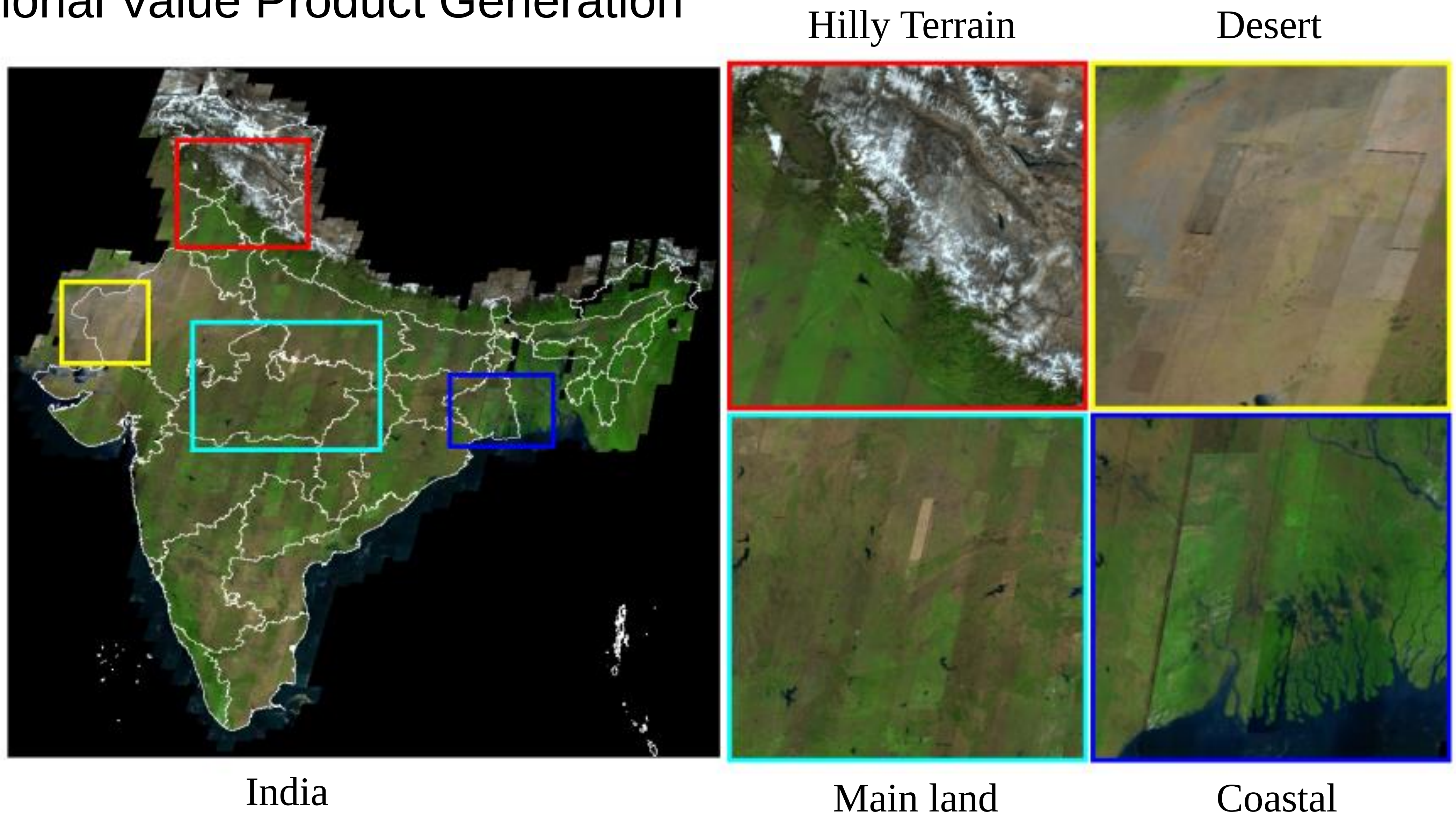
Wetland Delineation



Water Segmentation



Additional Value Product Generation



Outline

- Maximum Likelihood Estimation
- Reverse KL Divergence
- Introduction to GANs
- GANs in Remote Sensing
  - Problem Formulation

Why Adversarial Interaction Creates Non-Homogeneous Patterns: A Pseudo-Reaction Diffusion Model for Turing Instability

- Adversarial Interaction
  - Generative Adversarial Networks (GANs)
  - Application of conditional GANs
- Non-Homogeneous Patterns
  - Homogeneous patterns
  - Supervised learning
- Reaction-Diffusion
  - Turing's RD model (1952)
  - Gray-Scott RD model (1984)
- Turing Instability
  - Reaction dynamics
  - Diffusion dynamics