Session III: Generative models & Inverse Problems

Greg Tsagkatakis

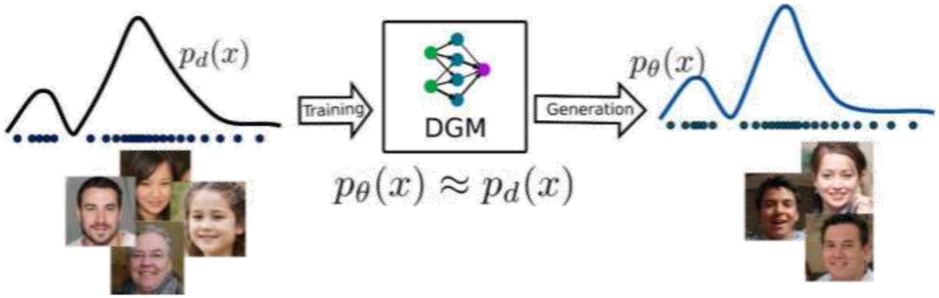
CSD - UOC

ICS - FORTH

Overview

Inverse problems

Generative models

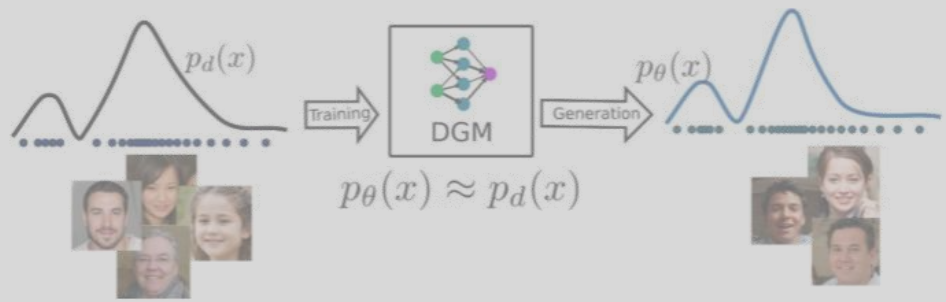


https://data-science-blog.com/blog/2022/02/19/deep-generative-modelling/

Overview

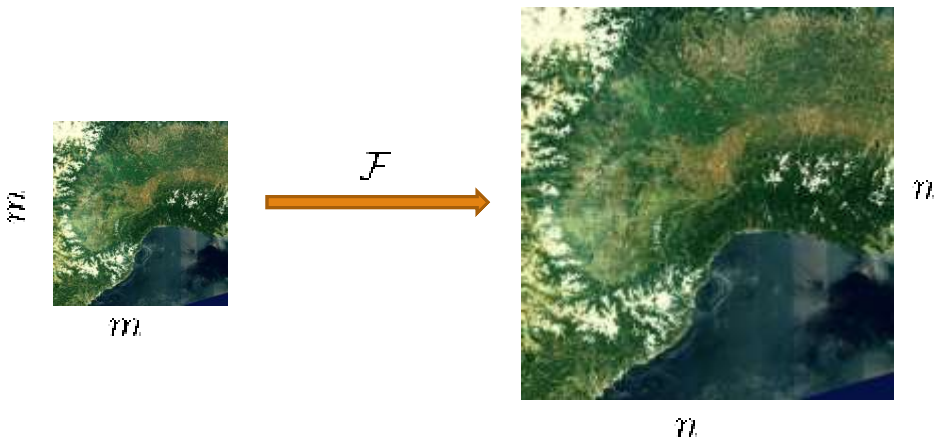
Inverse problems

Generative models

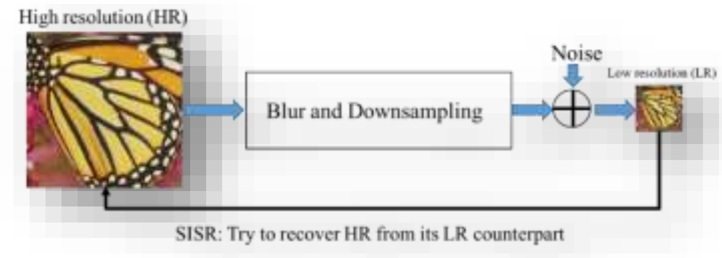


https://data-science-blog.com/blog/2022/02/19/deep-generative-modelling/

Super resolution



where $n > m$



$$MSE = \frac{\sum_{i,j=1}^{km,kn} ||Y_{true} - Y_{pred}||^2}{km * kn}$$

$$PSNR = 10 \log_{10}\left(\frac{R^2}{MSE}\right)$$

$$SSIM(x,y) = \left[l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma}\right]$$

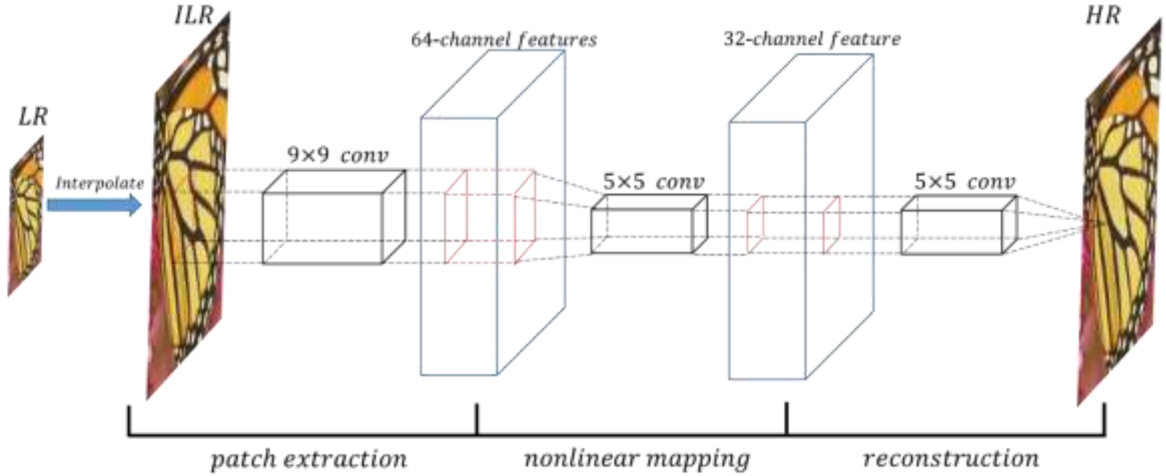
$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \qquad c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \qquad s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

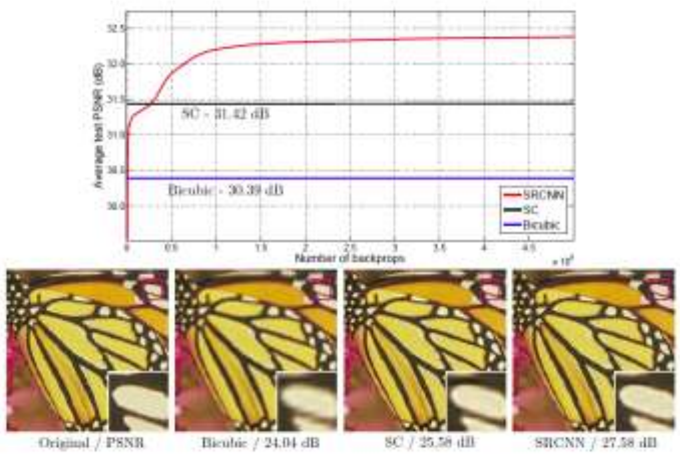
Types

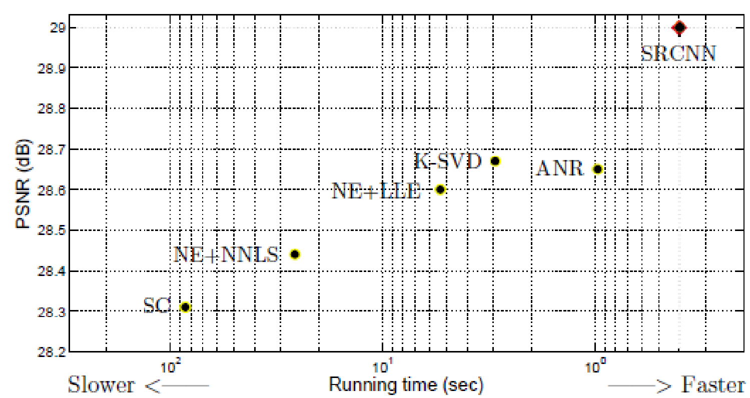
- Single image SR
- Multiframe SR
- Multispectral SR

Tsagkatakis, Grigorios, et al. "Survey of deep-learning approaches for remote sensing observation enhancement." *Sensors* 19.18 (2019).

SRCNN

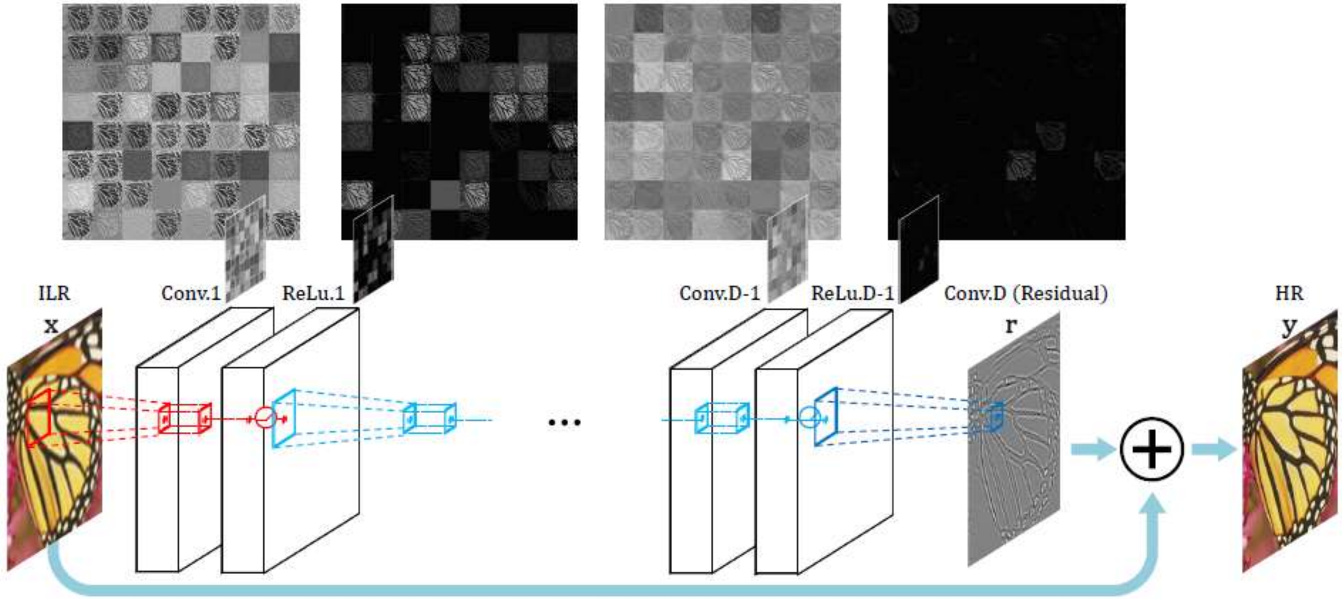


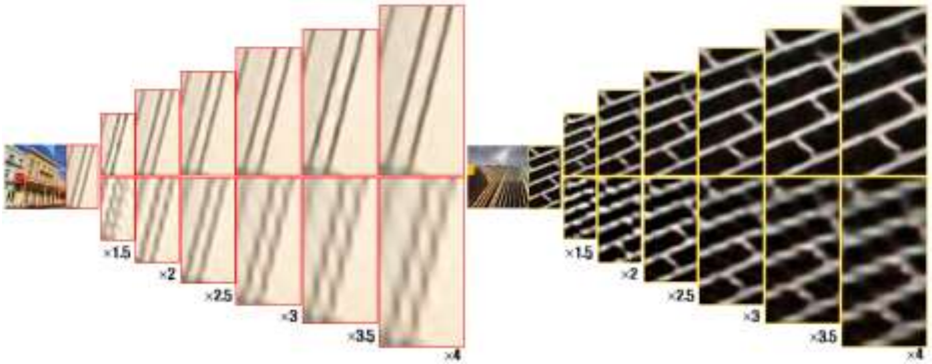




Dong, Chao, et al. "Image super-resolution using deep convolutional networks." *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015): 295-307.

VDSR





Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

EDSR

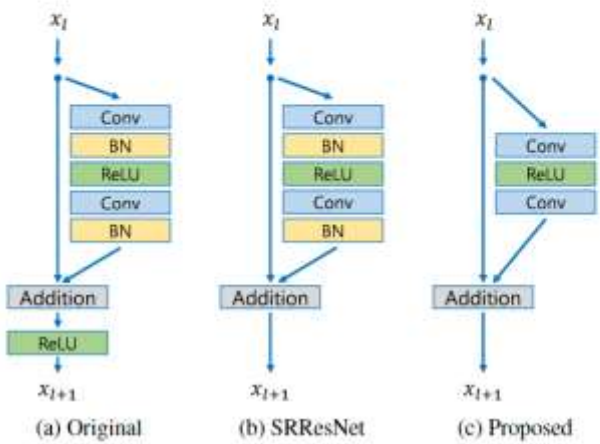
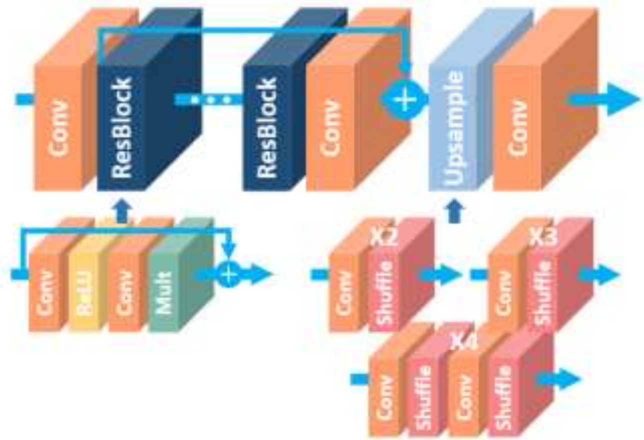
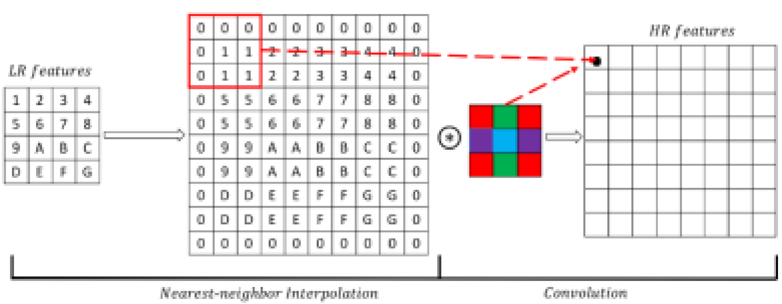
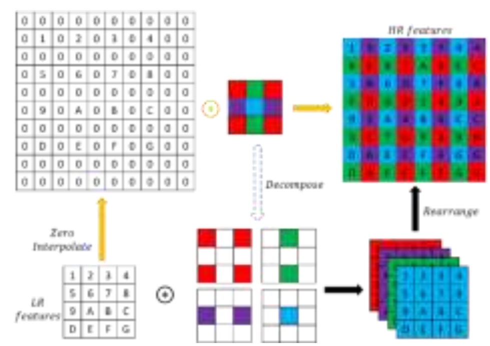


Figure 2: Comparison of residual blocks in original ResNet, SRResNet, and ours.

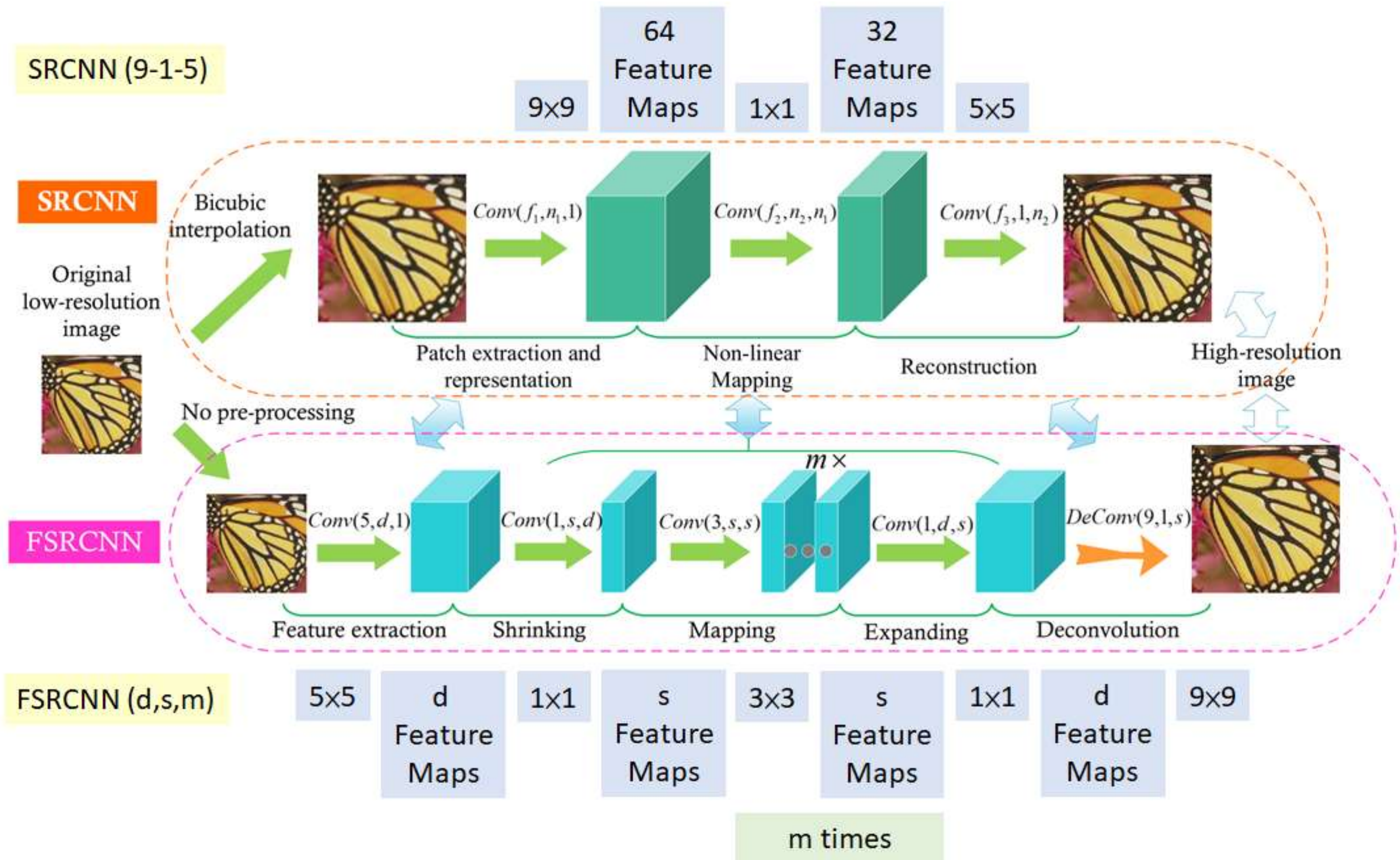






Lim, Bee, et al. "Enhanced deep residual networks for single image super-resolution." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017.

FRSCNN



Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference.*

Deep SR in RS

WV-3

Method		PSNR	SSIM	AG	LPIPS	NIQE	PI
CNN	SRCNN	32.5257	0.9612	4.8470	0.3148	6.0595	5.9935
	VDSR	32.8429	0.9661	4.9292	0.3074	5.8879	5.8466
	LGCNet	32.6579	0.9634	4.8630	0.3154	6.1080	5.9922
	PECNN	32.6076	0.9644	**5.2229**	0.3033	5.7095	5.7747
	RDN	32.8617	0.9673	5.1977	0.2938	**5.5007**	**5.6375**
	DDRN	**32.9531**	0.9677	5.1353	0.2984	**5.4723**	**5.5954**

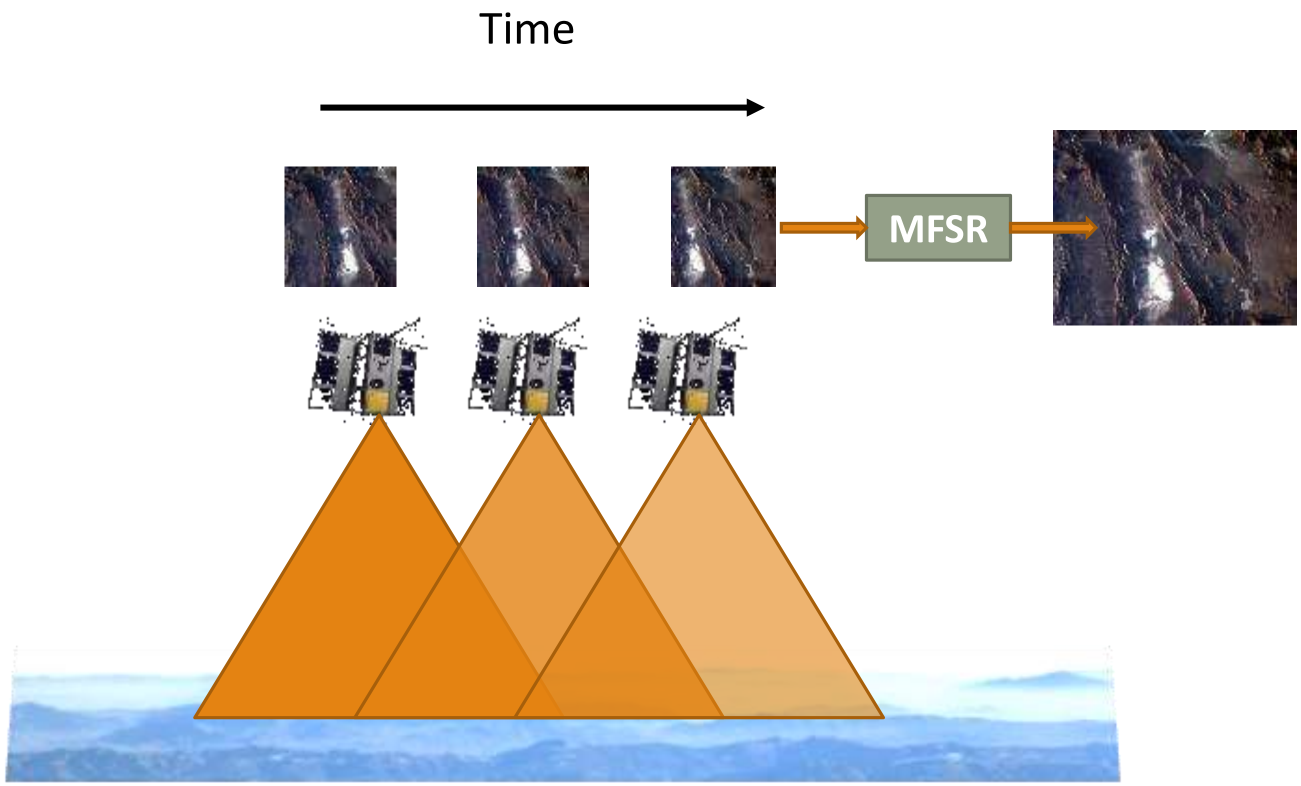
Pleiades

Method		PSNR	SSIM	AG	LPIPS	NIQE	PI
CNN	SRCNN	35.5458	0.9961	7.5978	0.1041	5.1356	4.0781
	VDSR	35.9257	0.9965	7.6702	0.0977	4.9027	3.9546
	LGCNet	35.6082	0.9963	7.5776	0.1032	4.9646	3.9894
	PECNN	35.7425	0.9965	7.6609	0.0997	4.9660	3.9906
	RDN	36.1849	**0.9968**	**7.7036**	0.0936	**4.6828**	**3.8375**
	DDRN	**36.2606**	**0.9968**	7.6390	0.0953	4.7043	3.8582

Wang, Peijuan, Bulent Bayram, and Elif Sertel. "A comprehensive review on deep learning based remote sensing image super-resolution methods." *Earth-Science Reviews* (2022): 104110.

Multi-Frame Super-Resolution

fusing short sequences of low-quality images into higher-quality ones



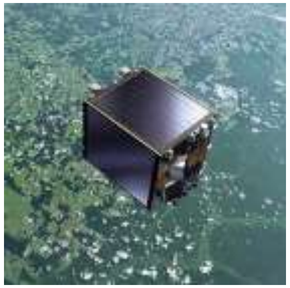
ESA Kelvin competition

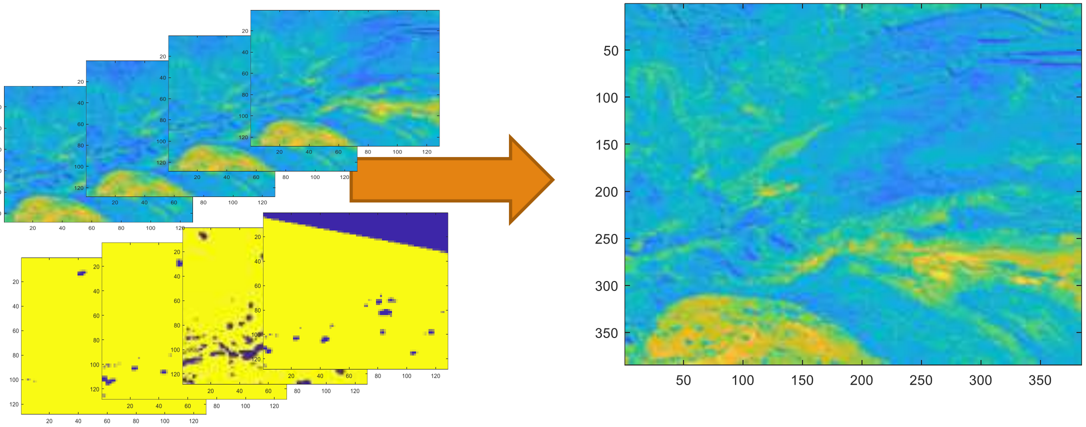
Proba-V mission

300m – daily, 100m – every 5 days

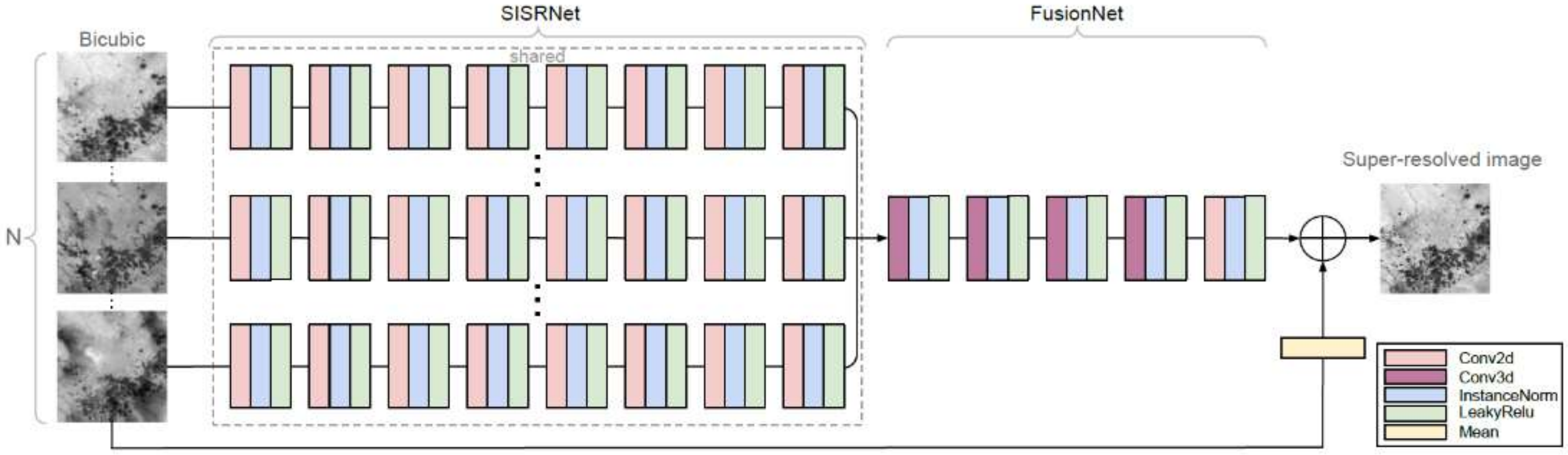
post-processing on the ground

Ended on June 2019



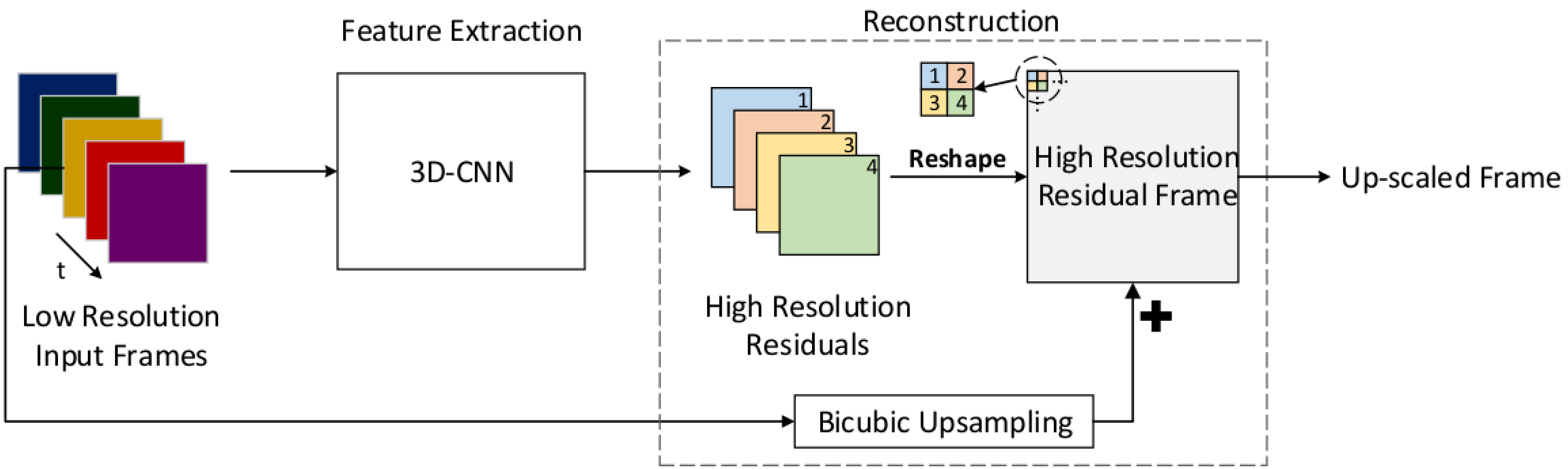


Unregistered Multi-temporal SR



Molini, A. B., Valsesia, D., Fracastoro, G., & Magli, E.. Deep learning for super-resolution of unregistered multi-temporal satellite images. In IEEE *WHISPERS 2019*

3D Wide-Activation NN

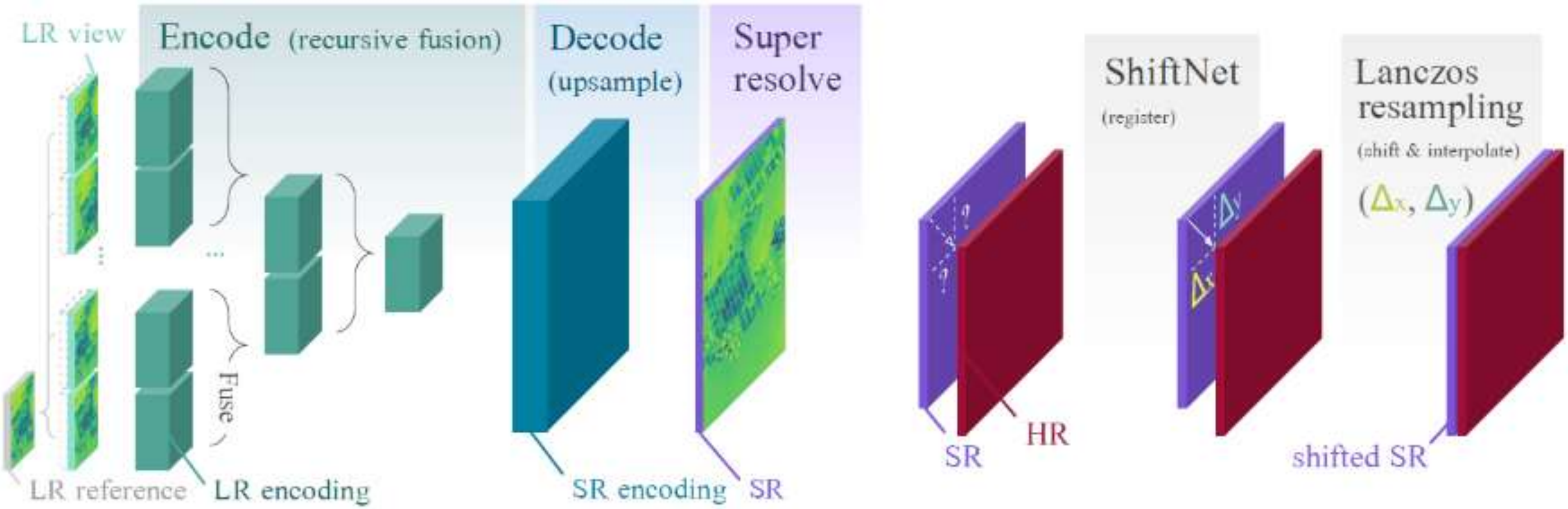


0.2 seconds to super-resolve (×3 upscaling) a scene with 32 low-resolution 128 × 128 views on an NVIDIA V100 GPU.

DL architecture that learns to fuse an arbitrary number of low-resolution frames with implicit co-registration to a reference frame.

Dorr, F. (2020). Satellite Image Multi-Frame Super Resolution Using 3D Wide-Activation Neural Networks. *Remote Sensing*, *12*(22), 3812.

HighRes-Net

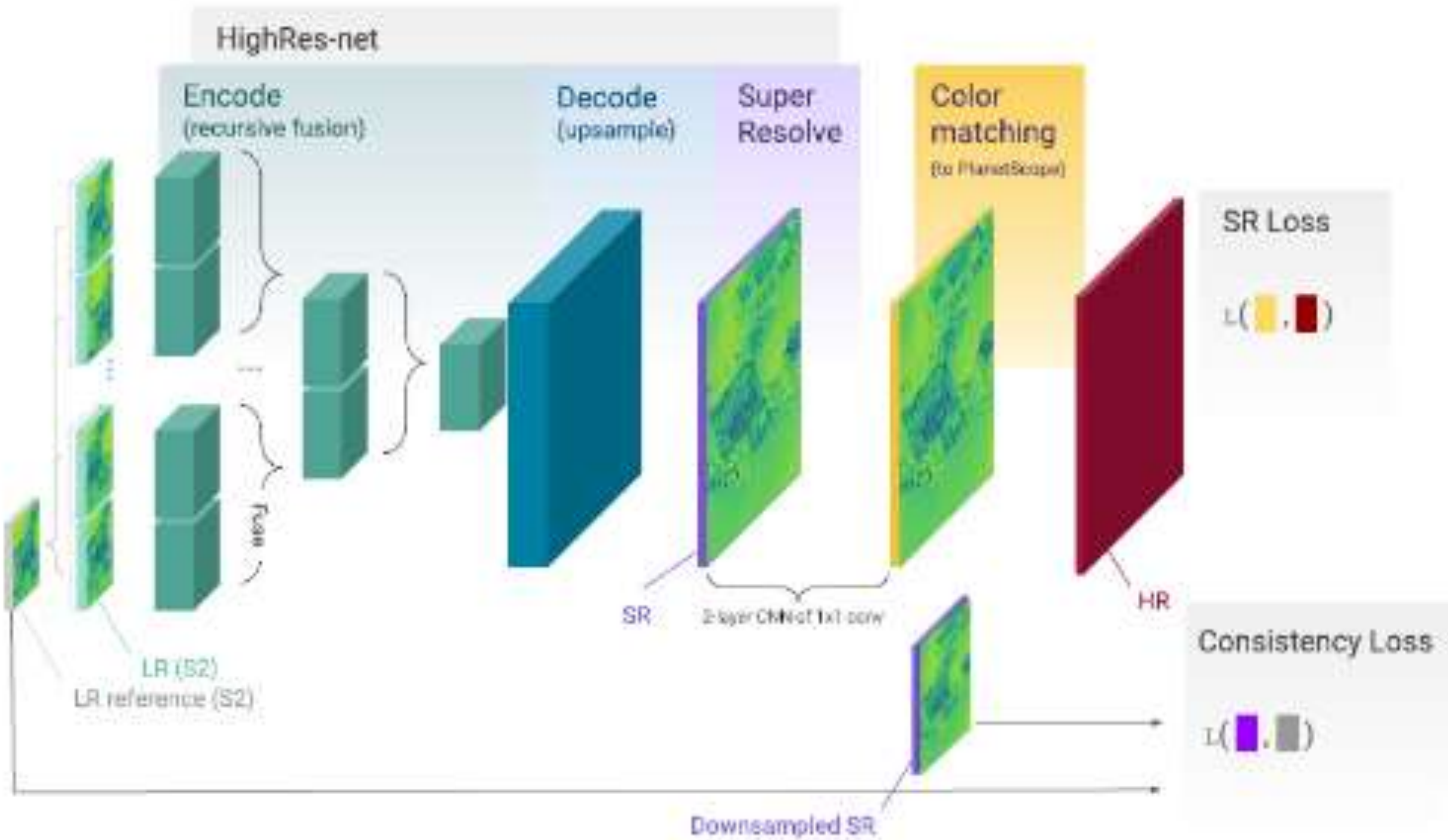


(a) HighRes-net

(b) Registered loss

Deudon, Michel, Alfredo Kalaitzis, Md Rifat Arefin, Israel Goytom, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E. Kahou, Julien Cornebise, and Yoshua Bengio. "Highres-net: Multi-frame super-resolution by recursive fusion." (2019).

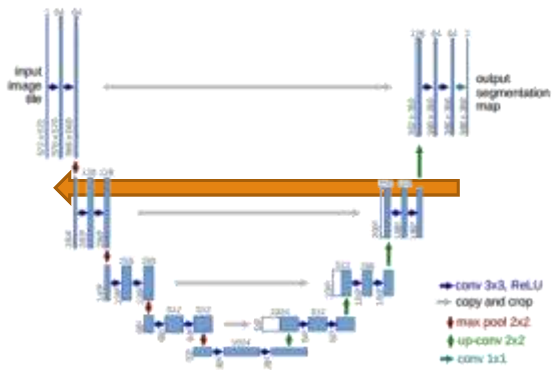
Multi-image SR on Sentinel 2



Razzak, Muhammed, et al. "Multi-Spectral Multi-Image Super-Resolution of Sentinel-2 with Radiometric Consistency Losses and Its Effect on Building Delineation." *arXiv preprint arXiv:2111.03231* (2021).

Denoising

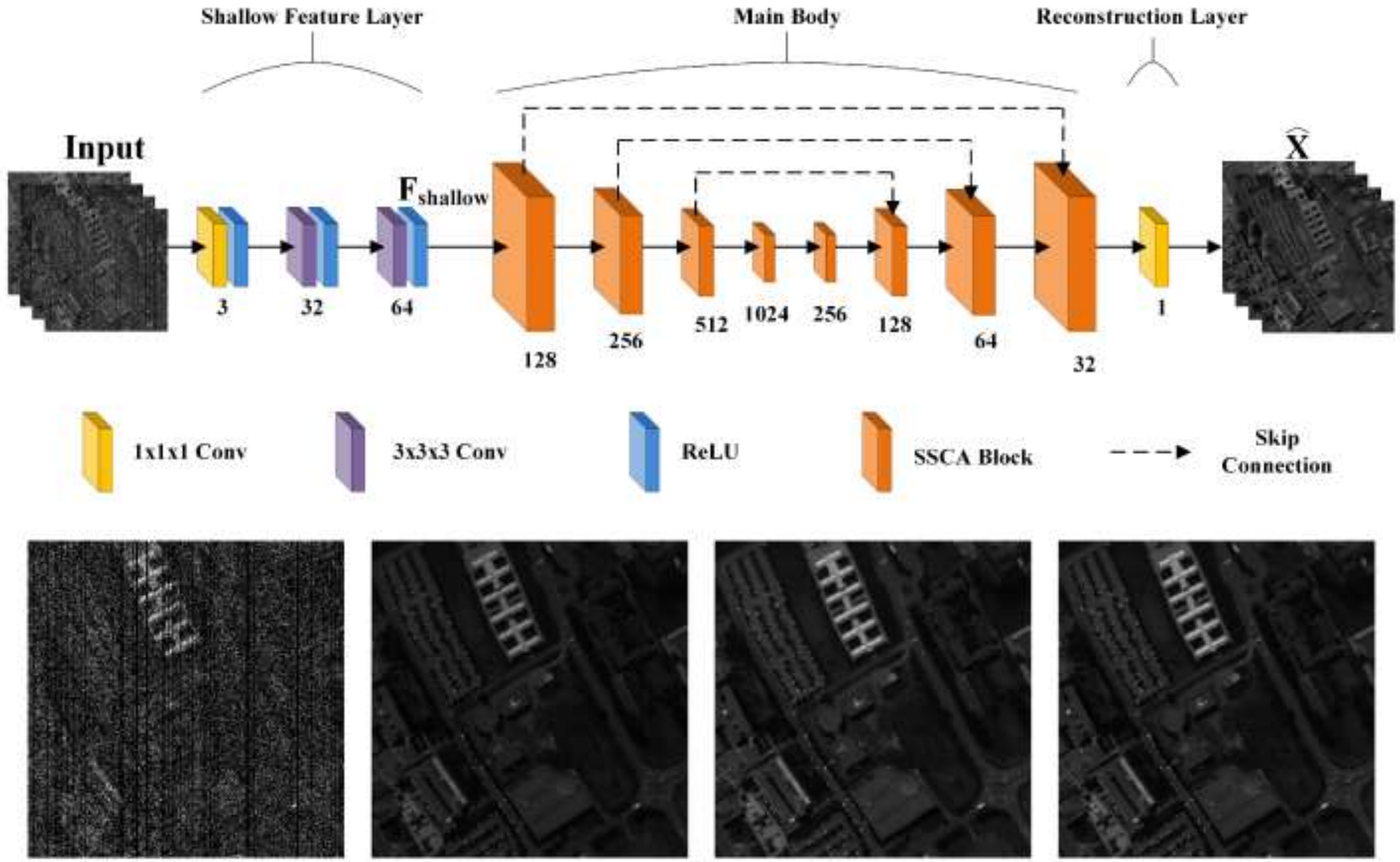
$$Y = X + n$$



$$\min_\theta \sum_{k=1}^{N} \|x_k - f(y_k;\theta)\|^2$$

SSCANet

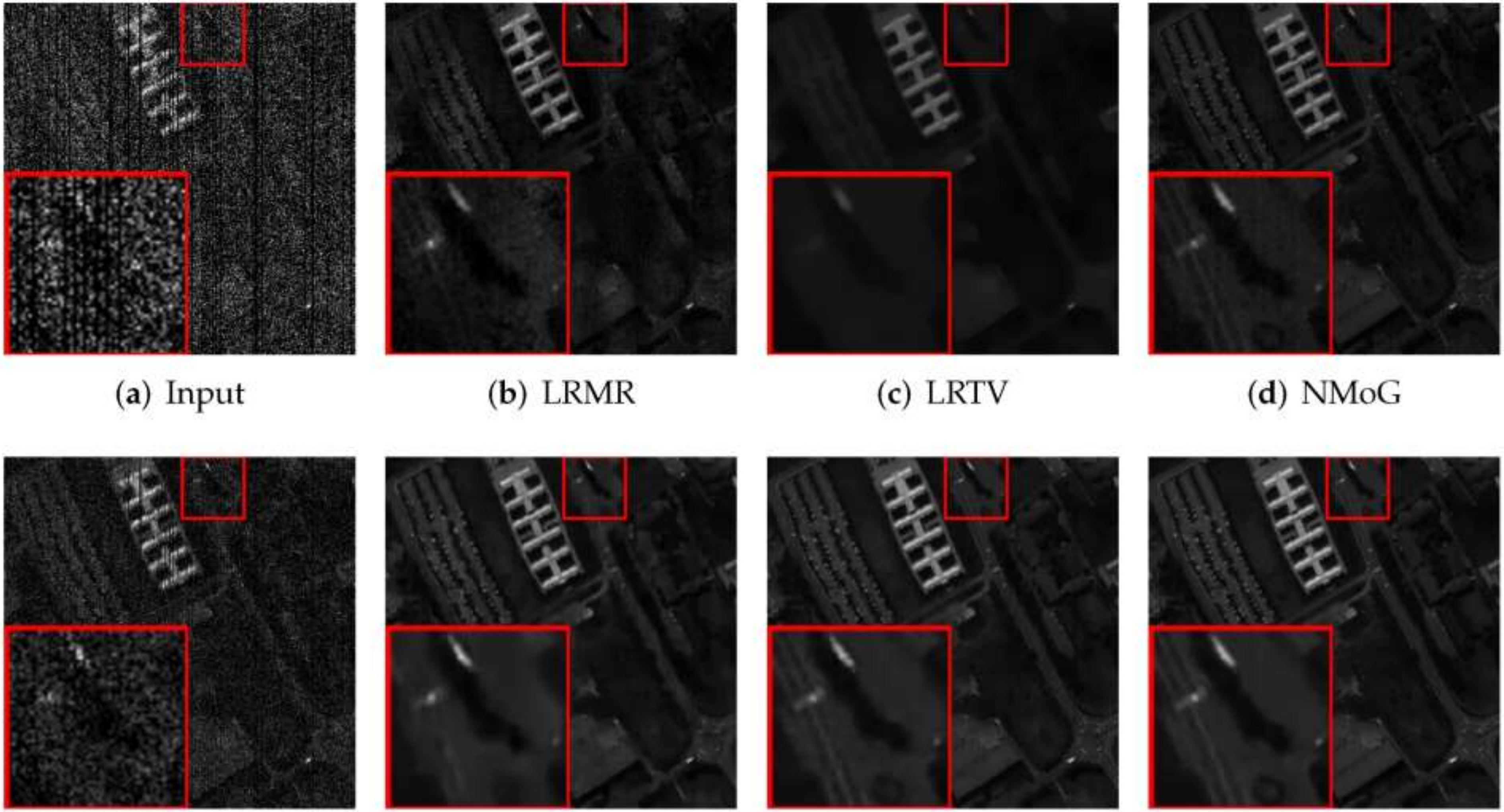
Spatial and Spectral-Channel Attention Network



(a) Input (b) NMoG (c) Ours (d) Ground-Truth (GT)

Dou, Hong-Xia, et al. "Spatial and spectral-channel attention network for denoising on hyperspectral remote sensing image." *Remote Sensing* 14.14 (2022): 3338.

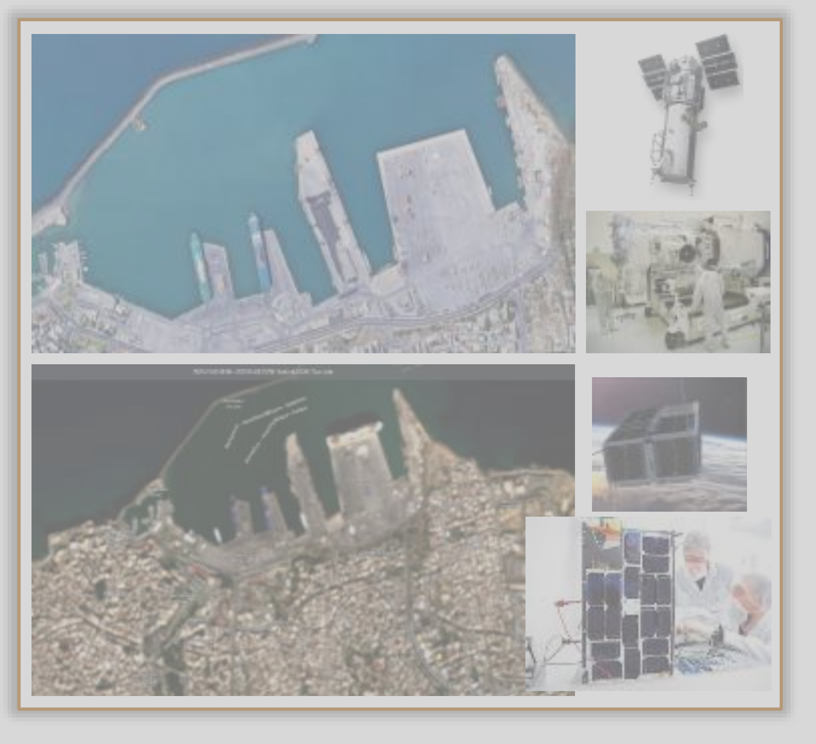
Performance



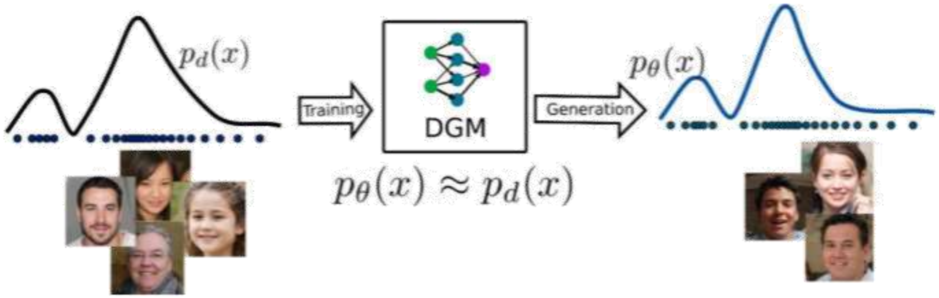
(a) Input (b) LRMR (c) LRTV (d) NMoG (e) TDL (f) HSID (g) QRNN3D (h) Ours

Overview

Inverse problems

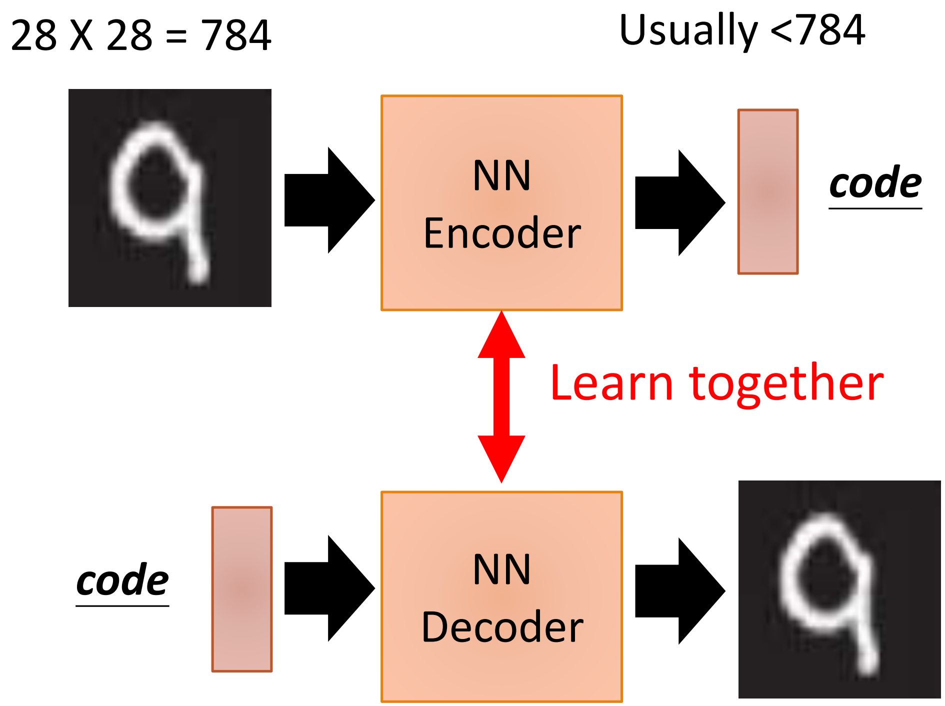


Generative models



https://data-science-blog.com/blog/2022/02/19/deep-generative-modelling/

AutoEncoders (AE)



Compact representation of the input object

Can reconstruct the original object

AutoEncoders (AE)

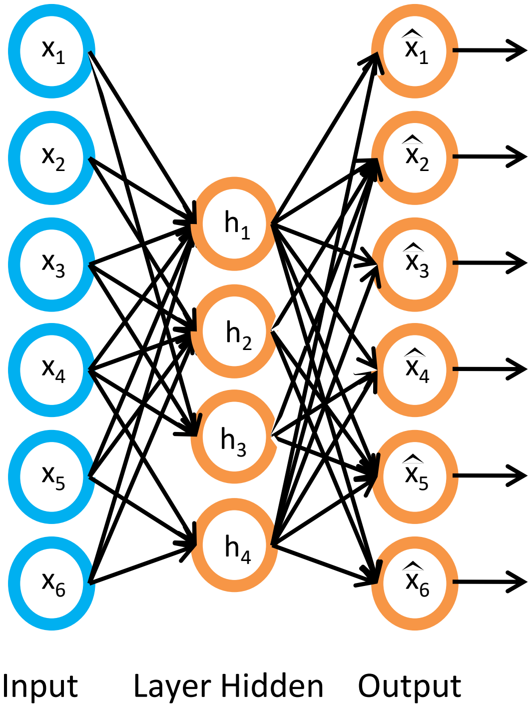
Unsupervised feature learning

Network is trained to output the input (learn identify function).

$$\boxed{J = \frac{1}{m}\sum_{i=1}^{m} \|\hat{x} - x\|_2}$$

Encoder $f(x) = \mathbf{h} = z(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$

Decoder $g(f(x)) = \hat{\mathbf{x}} = z(\mathbf{W}_2\mathbf{h} + \mathbf{b_2})$

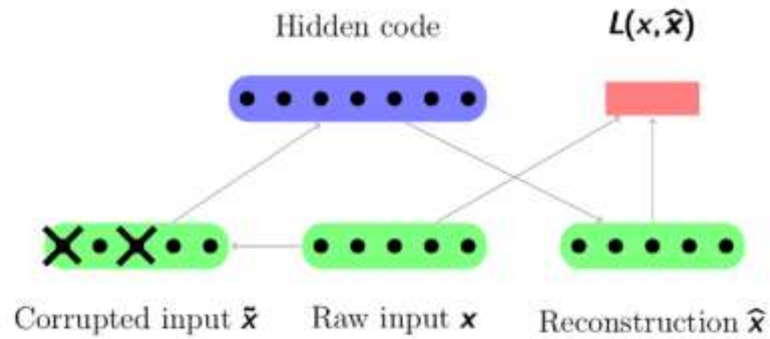


Regularized Autoencoders

Sparse neuron activation

$$J_{sparse} = \sum \|\hat{\mathbf{x}} - \mathbf{x}\|_2 + \beta \sum KL(p, \hat{p})$$

Denoising auto-encoders



Convolutional AE

$$f(x) = \mathbf{h} = z(\mathbf{W}_1 * \mathbf{x} + \mathbf{b}_1)$$

Stacked AutoEncoders (SAE)

Extended AE with multiple layers of hidden units

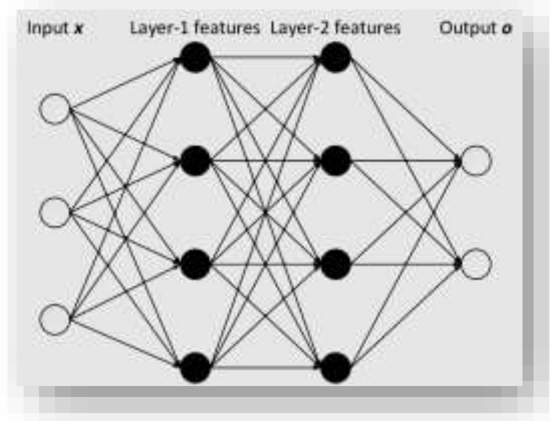
Challenges of Backpropagation

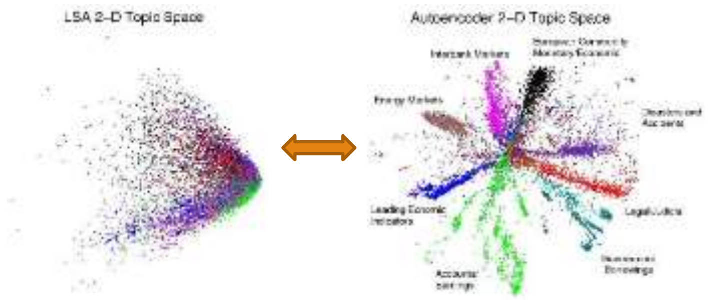
Efficient training

- Normalization of input

Unsupervised pre-training

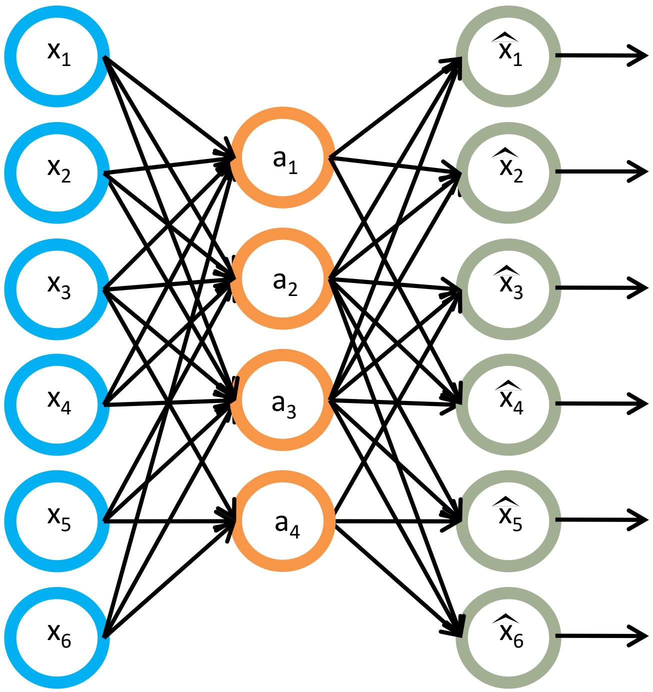
- Greedy layer-wise training
- Fine-tune w.r.t criterion





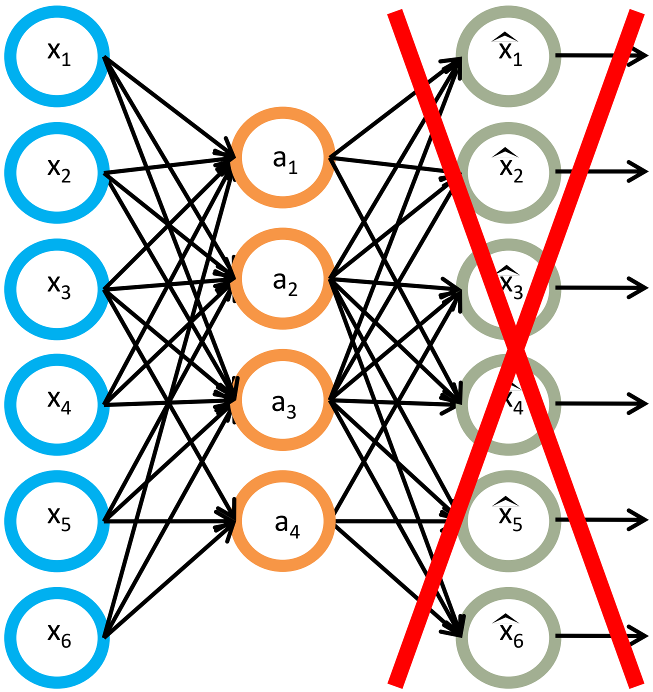
Bengio, Learning deep architectures for AI, Foundations and Trends in Machine Learning ,2009

SAE



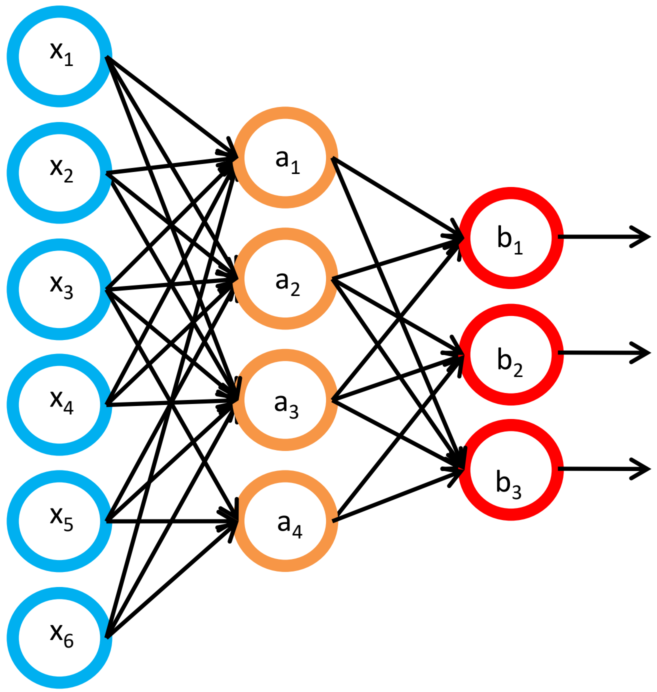
$$loss = \frac{1}{m}\sum_{i=1}^{m}\|\hat{x}_i - x_i\|_2$$

SAE

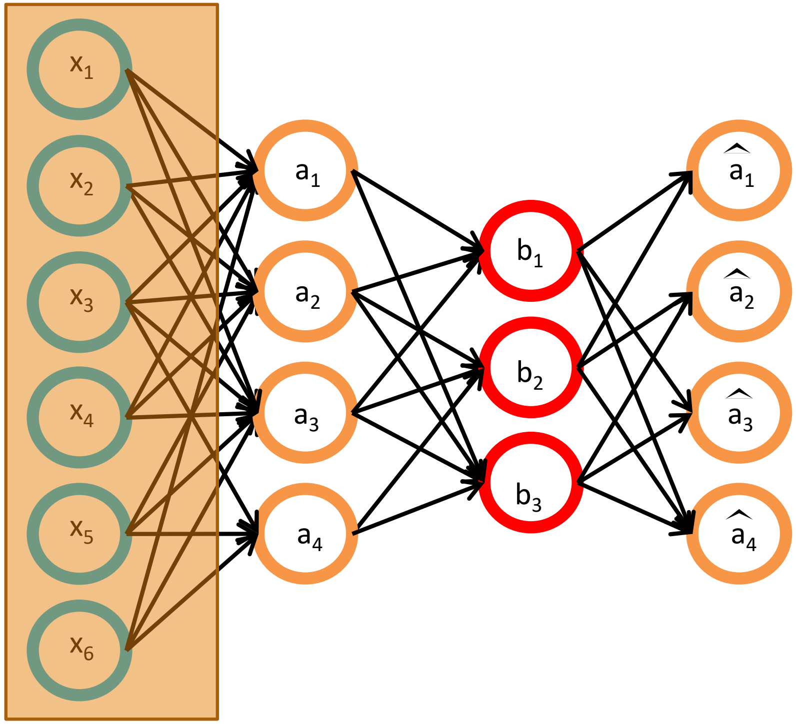


$$loss = \frac{1}{m}\sum_{i=1}^{m} \|\hat{x}_i - x_i\|_2$$

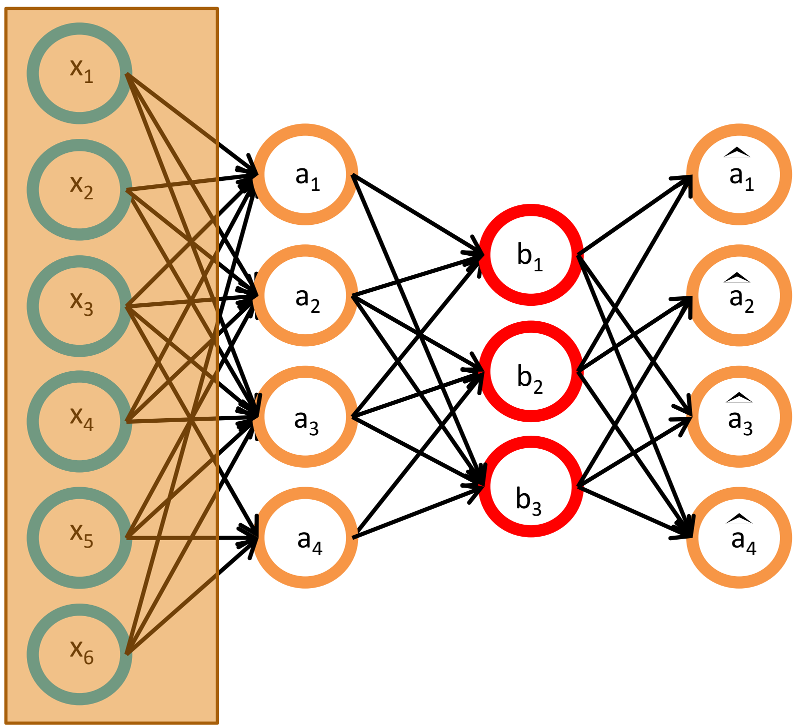
SAE



SAE

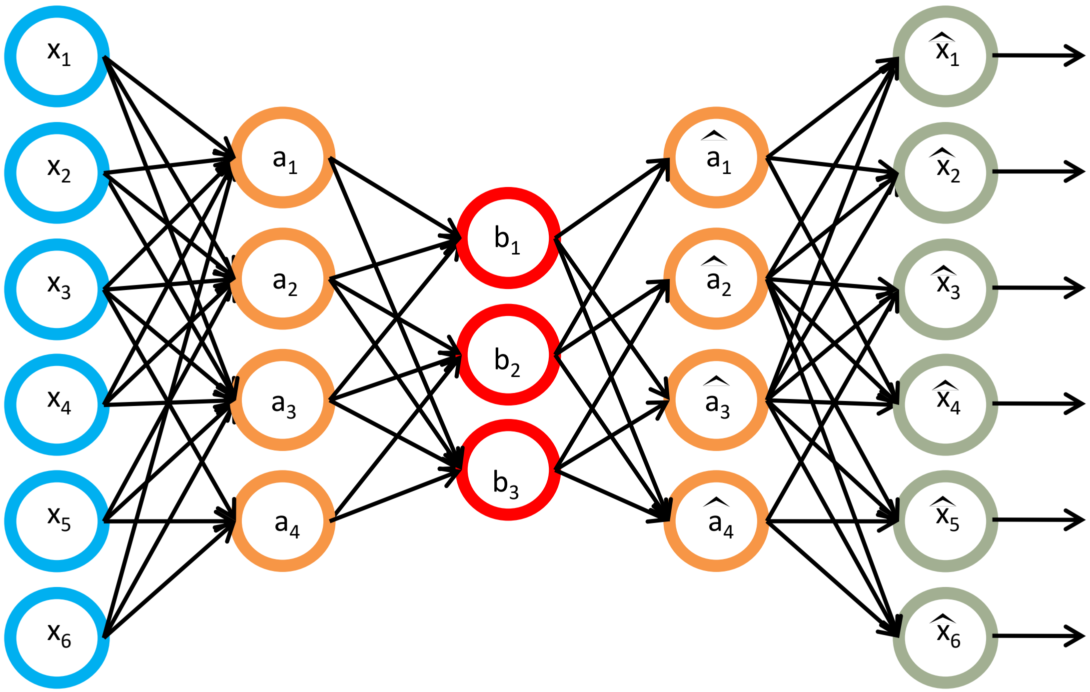


SAE



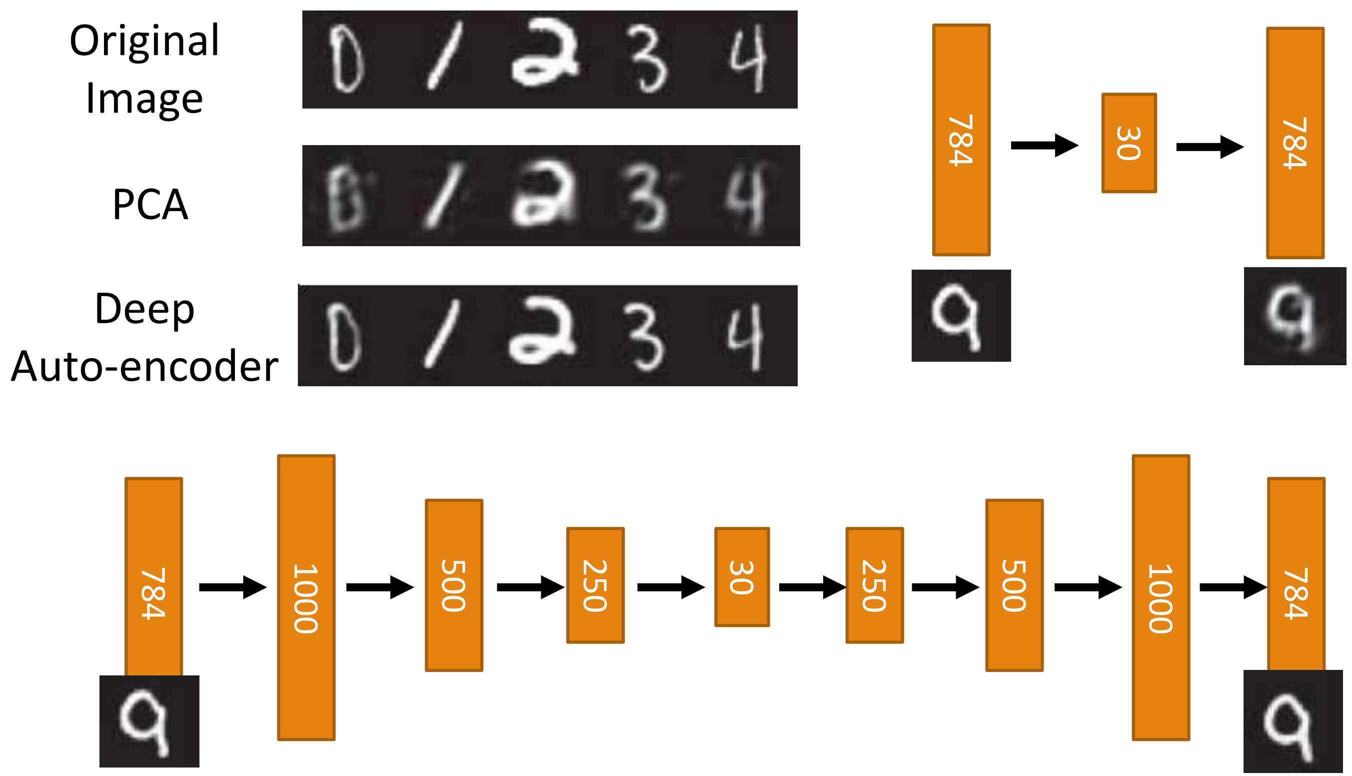
$$loss = \frac{1}{m}\sum_{i=1}^{m} \|\hat{a}_i - a_i\|_2$$

SAE

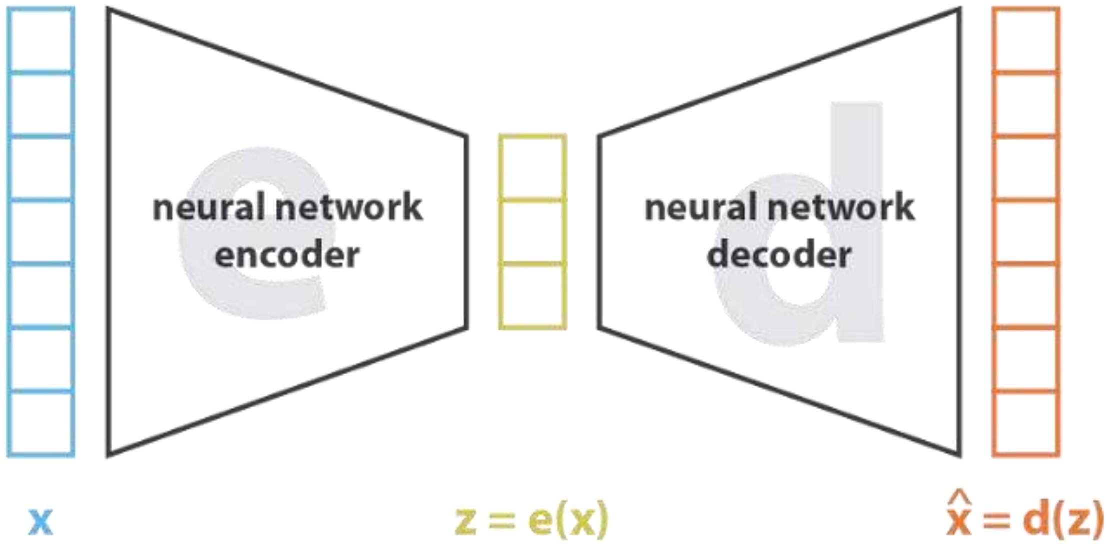


$$loss = \frac{1}{m}\sum_{i=1}^{m} \|\hat{x}_i - x_i\|_2 = \frac{1}{m}\sum_{i=1}^{m} \|\hat{a}_1(b_1(a_1(\hat{x}_i))) - x_i\|_2$$

Deep Autoencoder

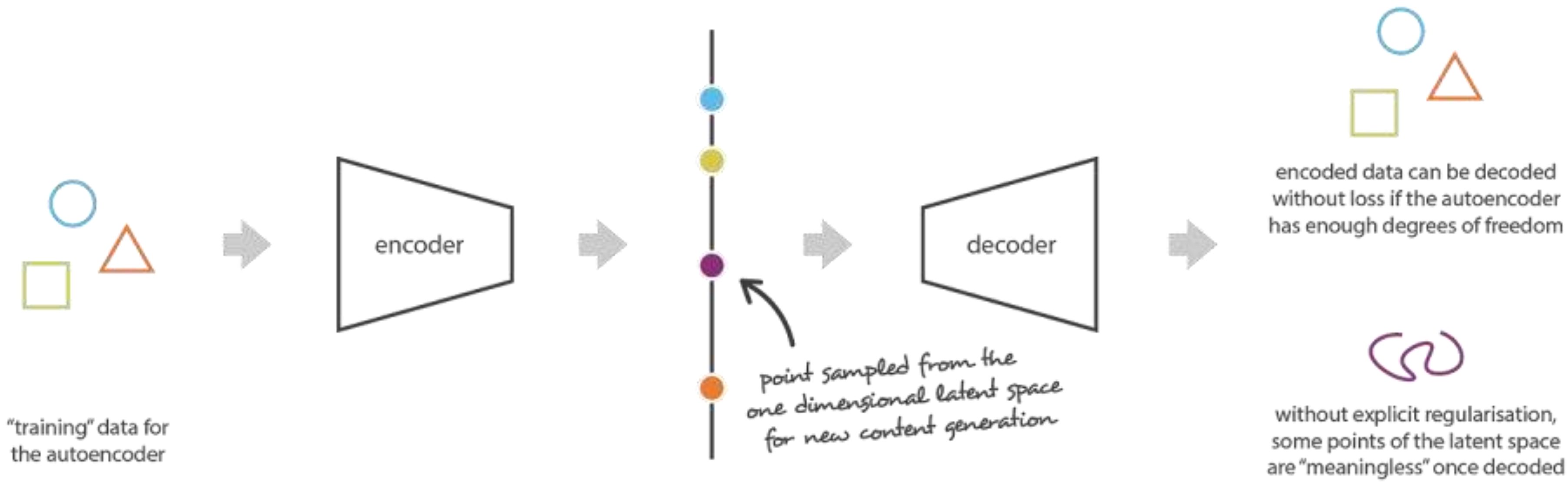


Deep Autoencoder



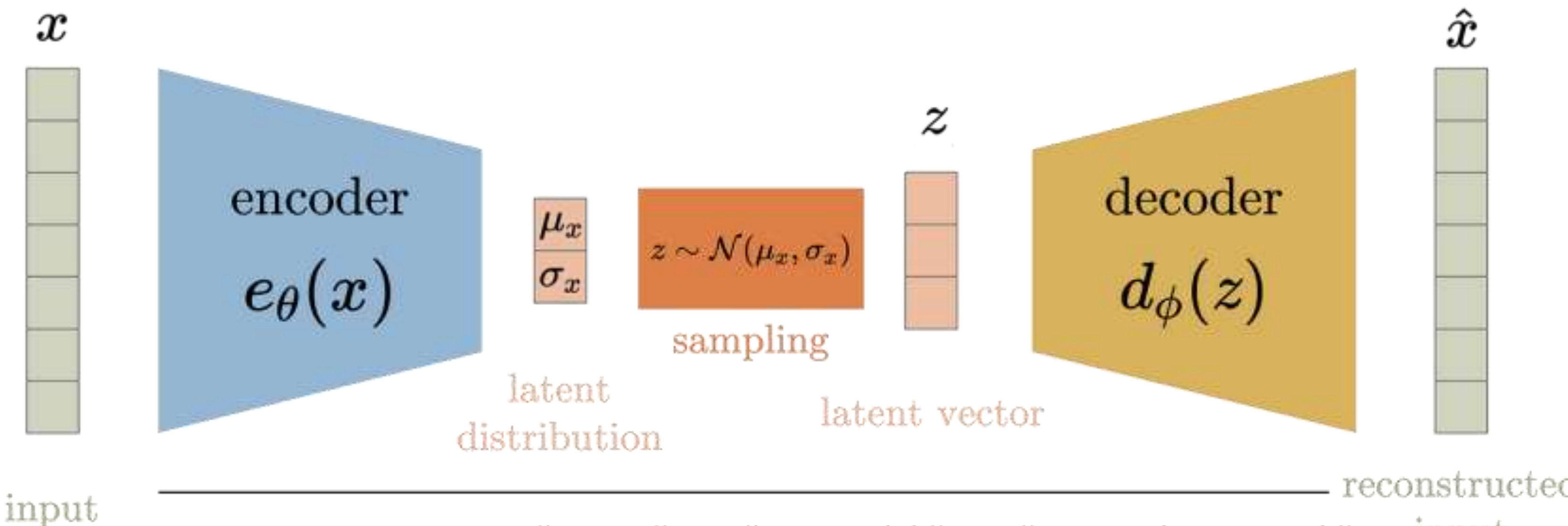
$$\text{loss} = \| x - \hat{x} \|^2 = \| x - d(z) \|^2 = \| x - d(e(x)) \|^2$$

Variational AutoEncoders



A VAE is an autoencoder whose encodings distribution is regularised during the training in order to ensure that its latent space can generate some new data.

VAE



$$\text{reconstruction loss} = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$similarity\ loss = KL\ Divergence = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$loss = reconstruction\ loss + similarity\ loss$$

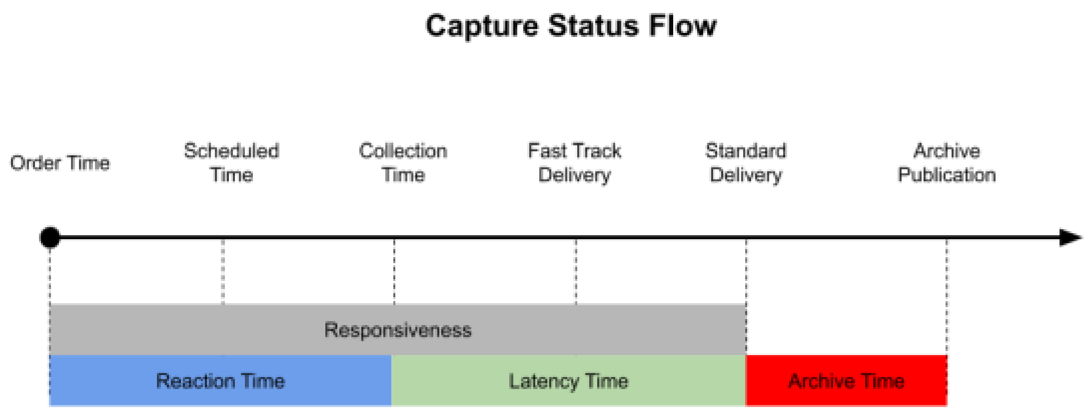
https://towardsdatascience.com/difference-between-autoencoder-ae-and-variational-autoencoder-vae-ed7be1c038f2

Onboard change detection

Applications such as disaster management enormously benefit from the rapid availability of satellite observations.

Traditionally, data analysis is performed on the ground after being transferred—downlinked—to a ground station.

Constraints on the downlink capabilities, both in terms of data volume and timing, heavily affect any downstream application's response delay.



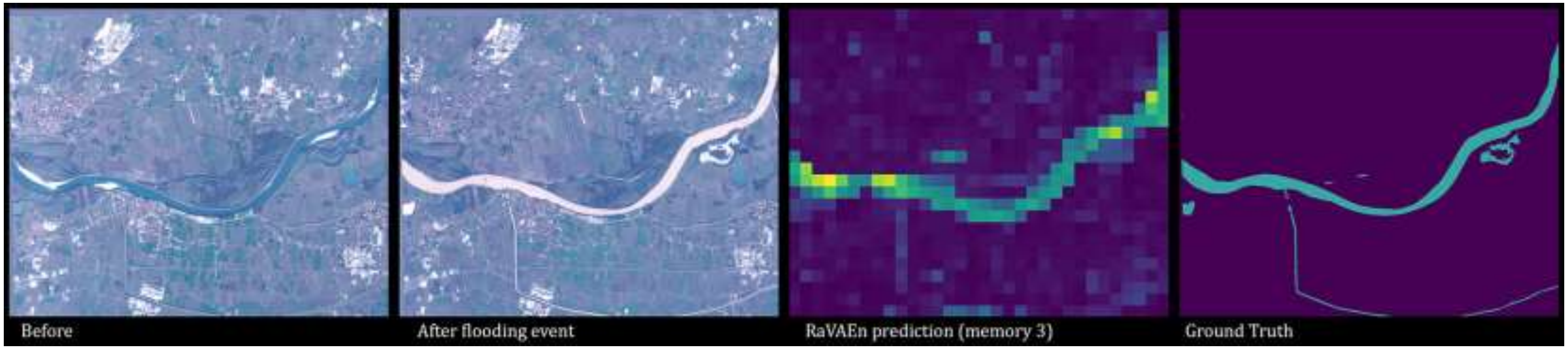
https://developers.planet.com/docs/tasking/basics/

RaVAEn

Unsupervised learning for change detection

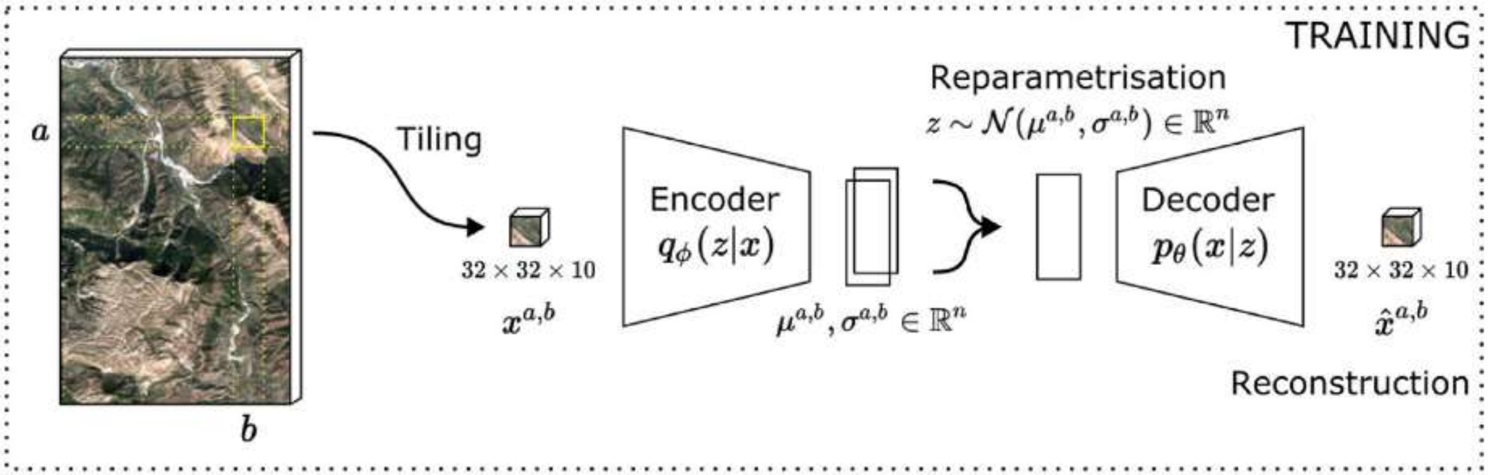
Variational AutoEncoder (VAE) model

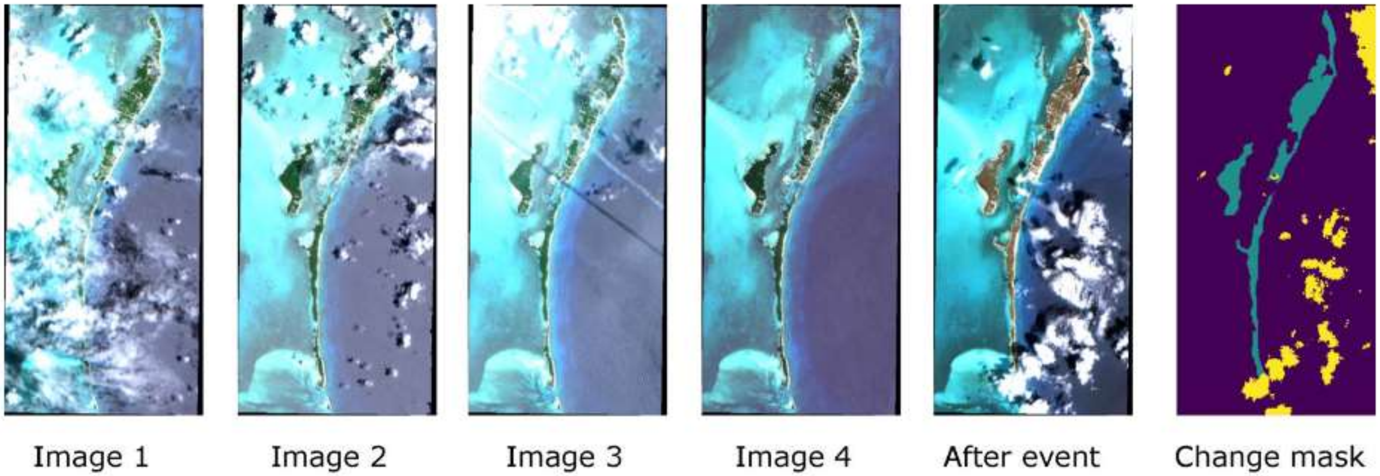
Target onboard deployment



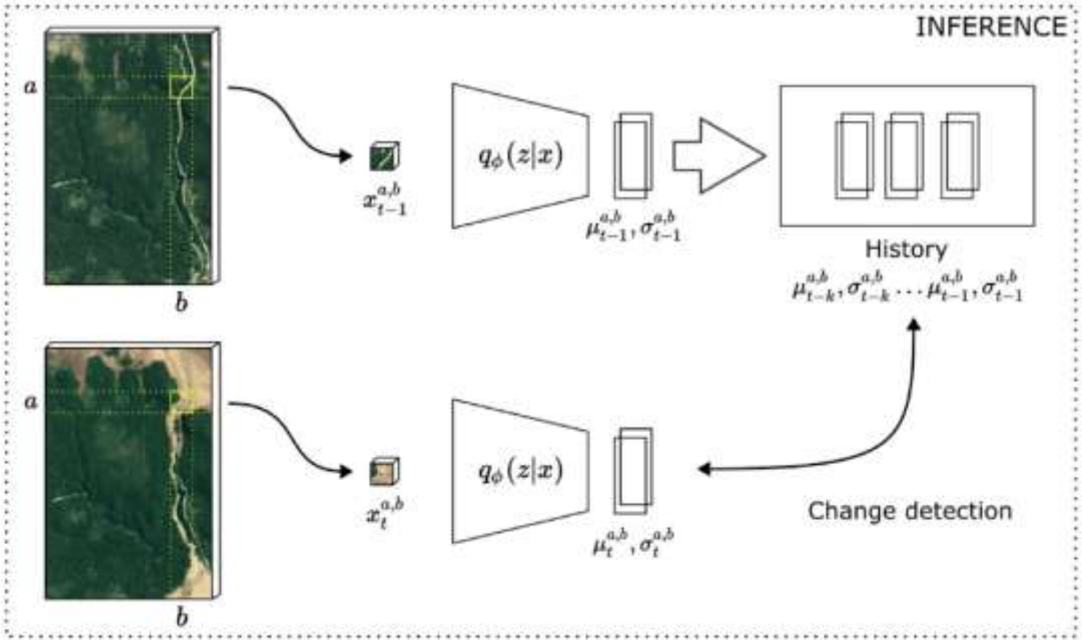
Růžička, Vít, et al. "RaVÆn: unsupervised change detection of extreme events using ML on-board satellites." *Scientific reports* 12.1 (2022): 16939.

Off-board training





On-board analysis



$$S(x_t^{a,b}) = \min_{i=1\ldots k} d(x_{t-i}^{a,b}, x_t^{a,b})$$

Dataset

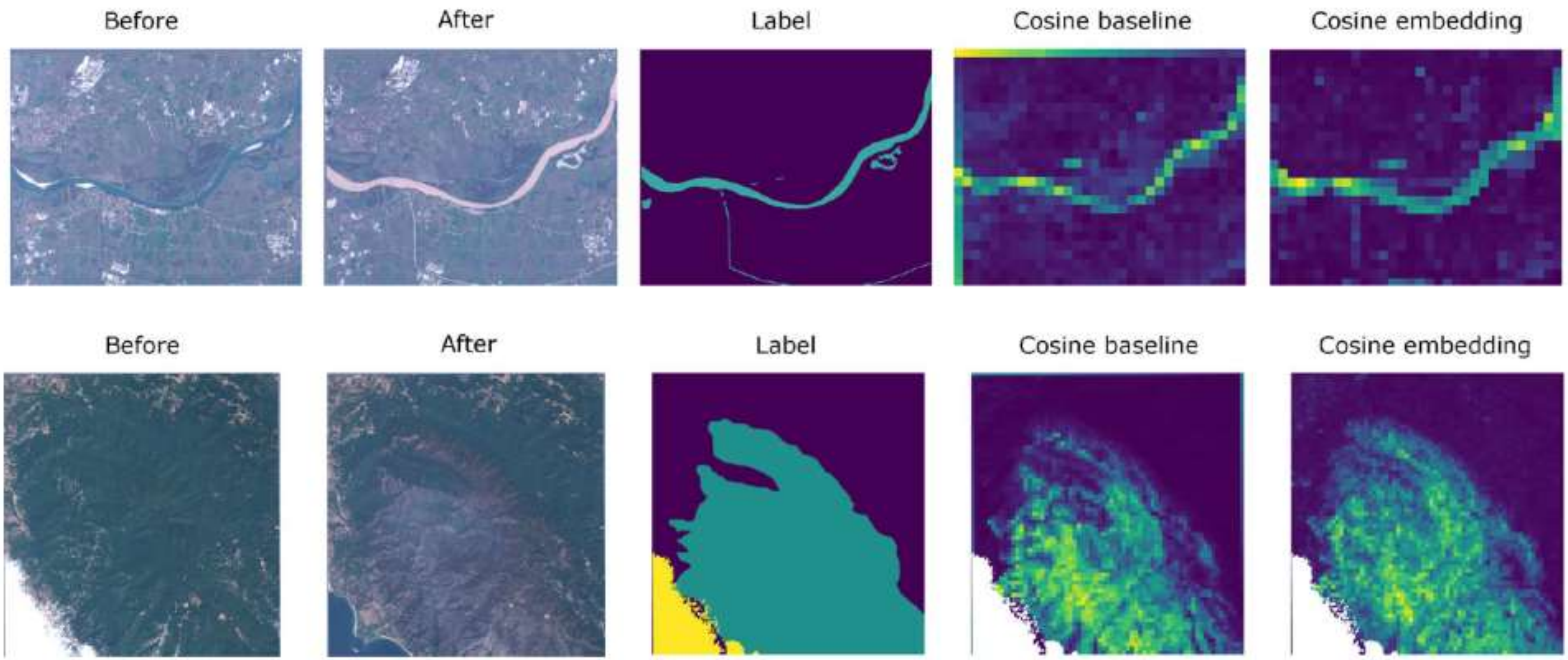
Input: Sentinel-2

Label: Copernicus EMS system

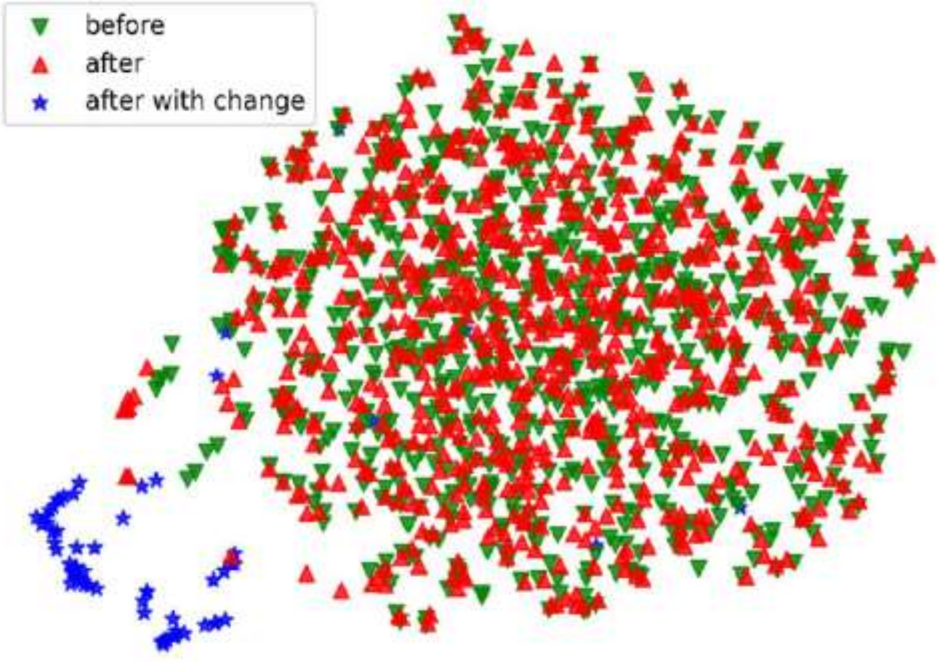
4 before + 1 after event

	Number of locations	Cumulative area (km^2)	Positive rate
Landslides	5	108	10.48
Floods	4	1301	6.74
Hurricanes	5	1622	24.31
Fires	5	3485	53.79

Performance



Visualization



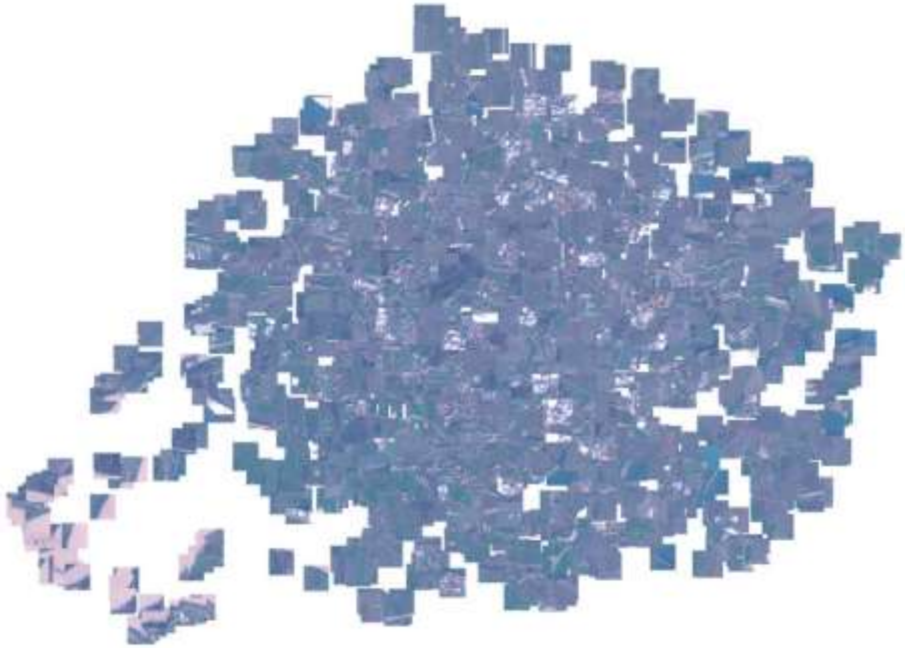


Image retrieval

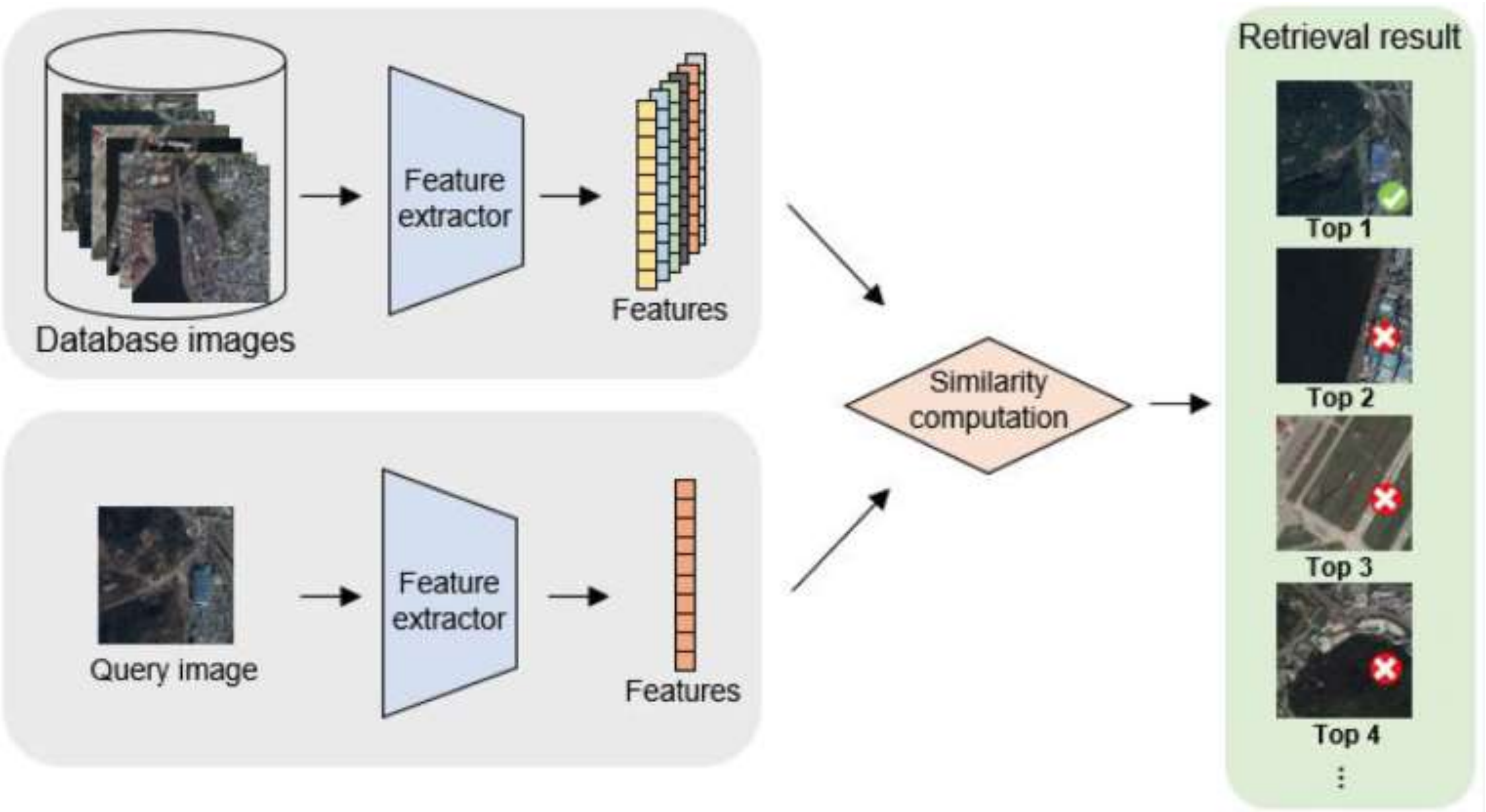


Image captioning

1. An old court is surrounded by white houses.
2. A playground is surrounded by many trees and long buildings.
3. A playground with basketball fields next to it is surrounded by many green trees and buildings.
4. Many green trees and several long buildings are around a playground.
5. This narrow, oval football field and closing basketball court, tennis court, parking lot together form this area, with plants wreathing it.

1. Four planes are stopped on the open space between the parking lot.
2. Four white planes are between two white buildings.
3. Some cars and two buildings are near four planes.
4. Four planes are parked next to two buildings on an airport.
5. Four white planes are between two white buildings.

[RSICD]
> University of Chinese Academy of Sciences
>English
> 2018
> 30 categories
> 10921 images

NWPU-Captions

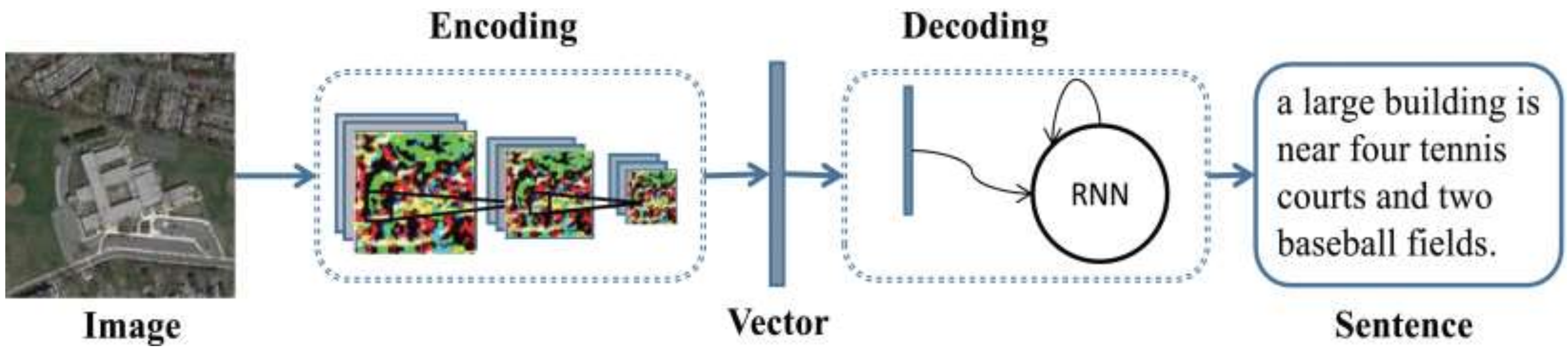
The industrial area has some blue workshops and green areas, and some roads go through the industrial area.
Houses in industrial areas vary in size.
Neatly planned plants and roads in an industrial area with good greening.
There are many roads in the industrial area.
There are some blue buildings and many black buildings in the industrial area.

The roundabout with three exits and entrances is in the residential area, and some clearings are next to the roundabout.
There are several buildings with swimming pools and some trees, and a large lawn around the roundabout with three roads.
The roundabout is next to buildings and trees.
There are many buildings of different shapes and sizes around the roundabout.
There are many buildings around the roundabout.

[NWPU-Captions]
> Huazhong University of Science and Technology
>English
> 2022
> 45 categories
> 31500 images
> 30 m - 0.2 m

https://github.com/iOPENCap/awesome-remote-image-captioning

RS -> TEXT



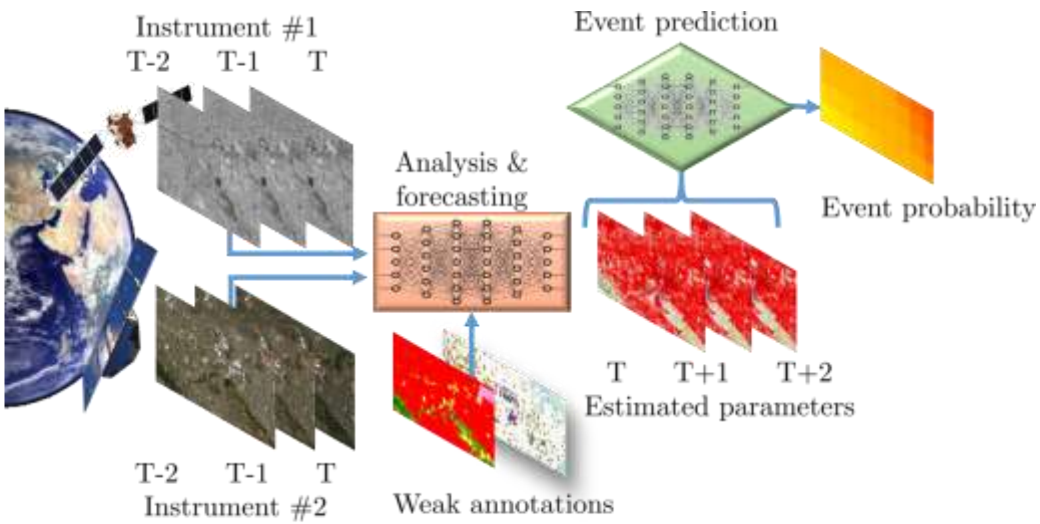
Many buildings and some green trees are in a commercial area .

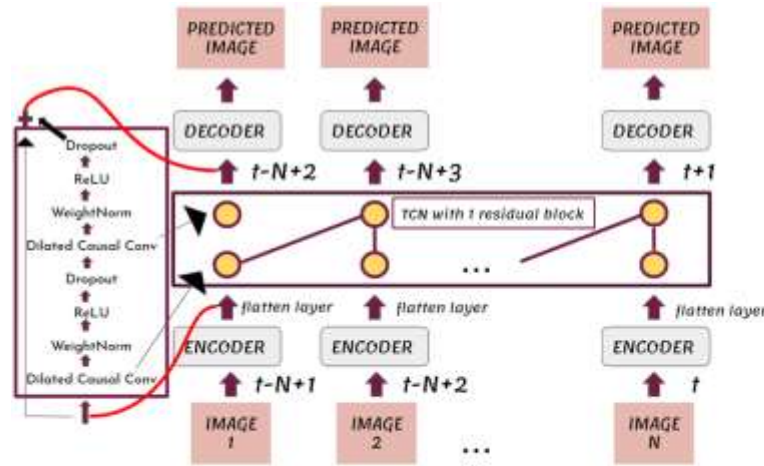
Many buildings and some green trees are in two sides of a railway station.

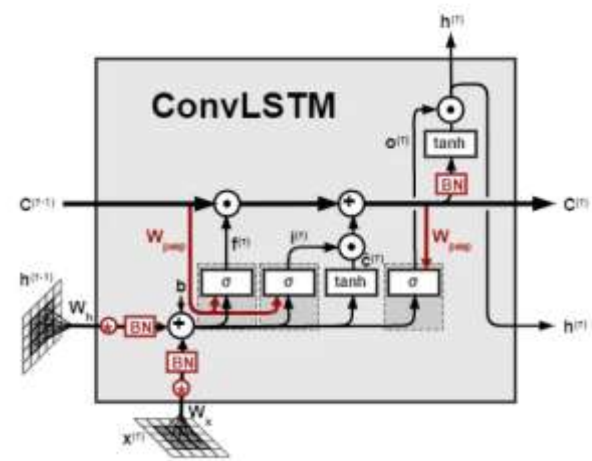
A playground with a football field in it is surrounded by some green trees and buildings.

Lu, Xiaoqiang, et al. "Exploring models and data for remote sensing image caption generation." *IEEE Transactions on Geoscience and Remote Sensing* 56.4 (2017): 2183-2195.

Forecasting

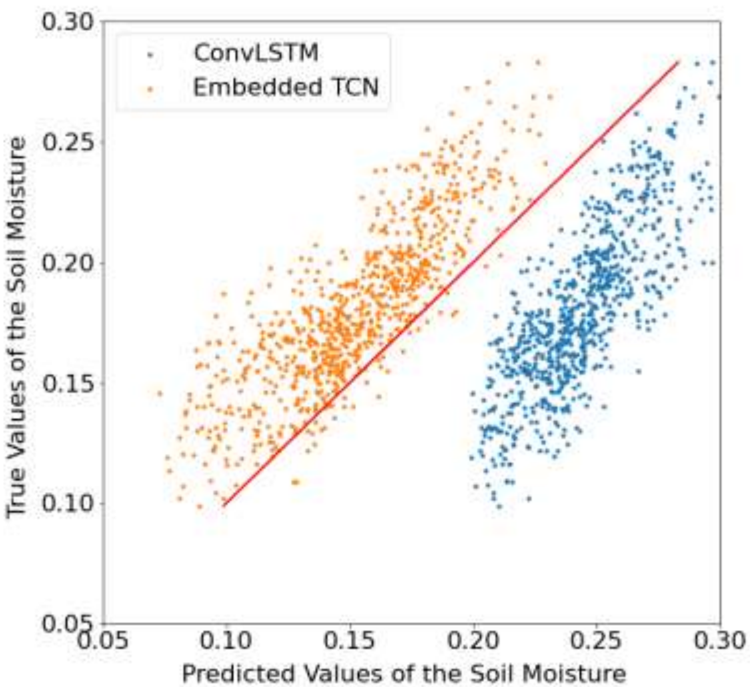




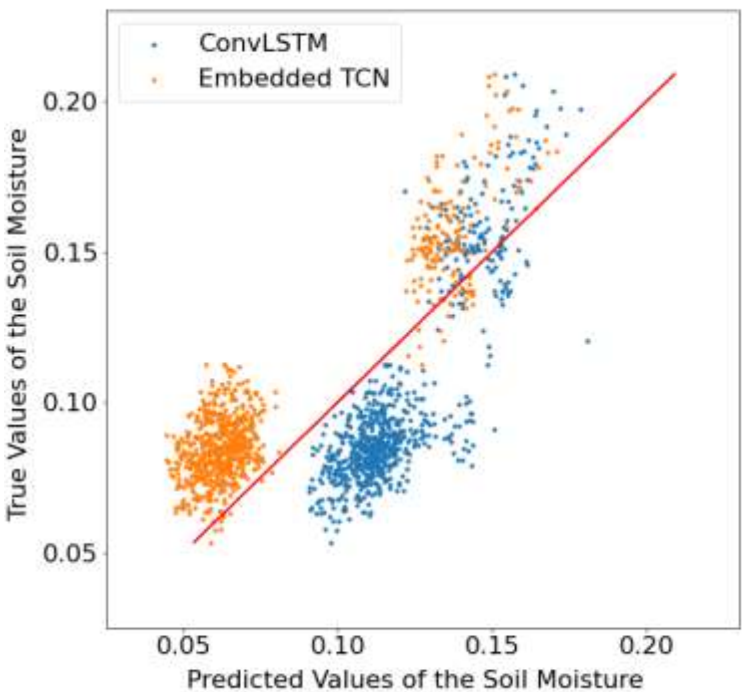


Villia, M. M., Tsagkatakis, G., Moghaddam, M., & Tsakalides, P. (2022). Embedded Temporal Convolutional Networks for Essential Climate Variables Forecasting. *Sensors*, *22*(5), 1851.

Soil moisture retrieval



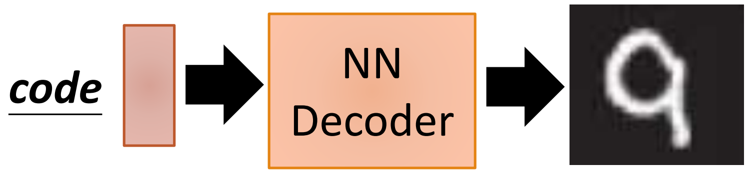
Greece

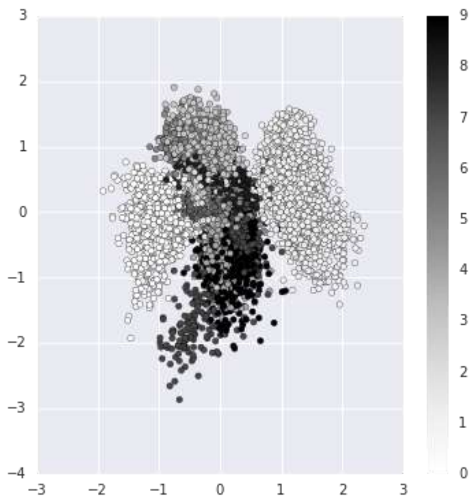


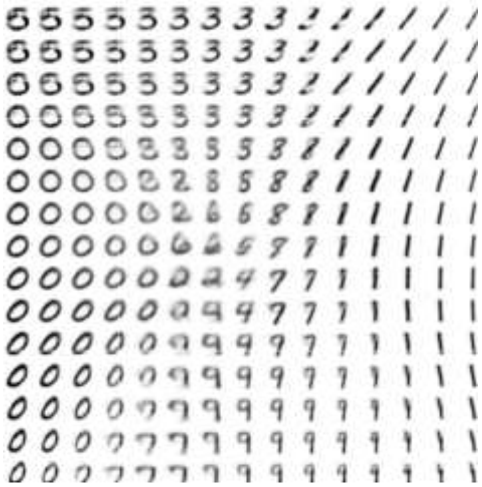
Sweden


AND NOW FOR SOMETHING COMPLETELY DIFFERENT

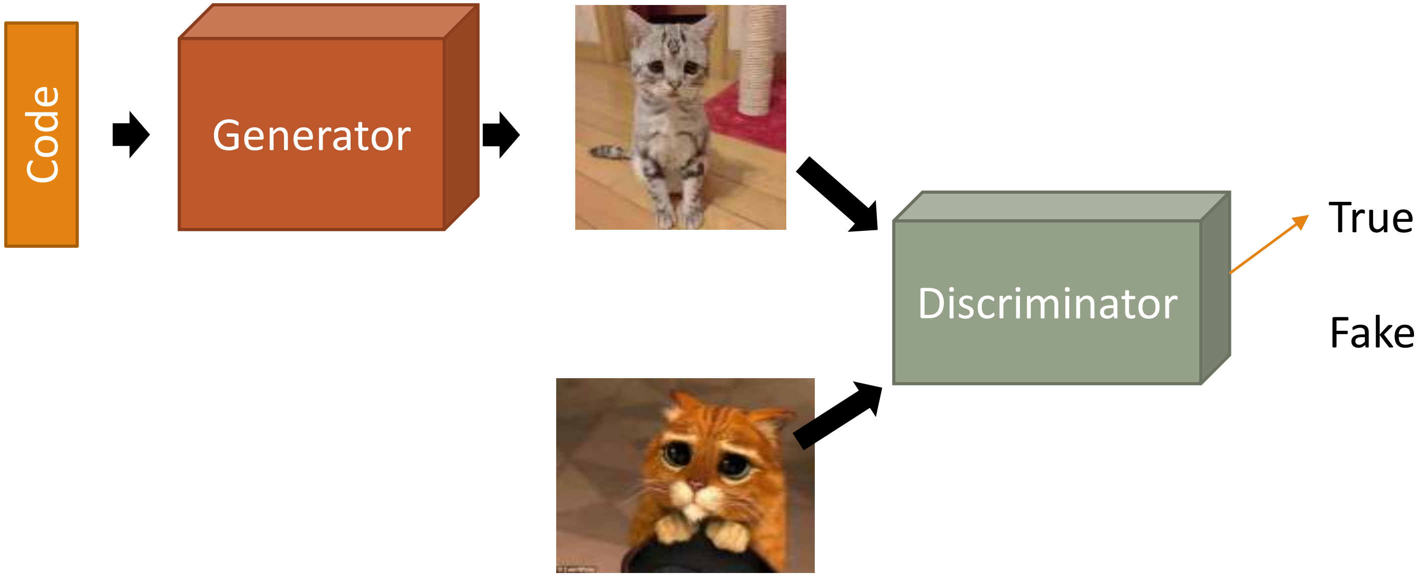
Generators?







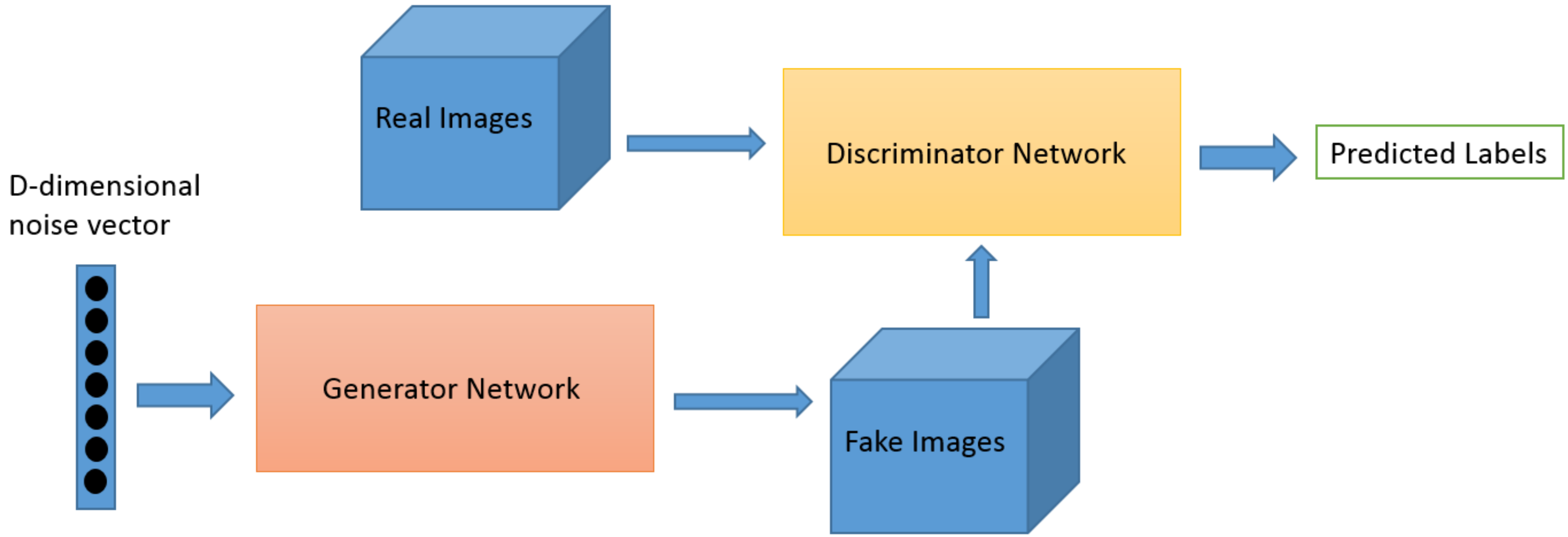
Generative Adversarial Networks



Yann LeCun, “Adversarial training is the coolest thing since sliced bread.”

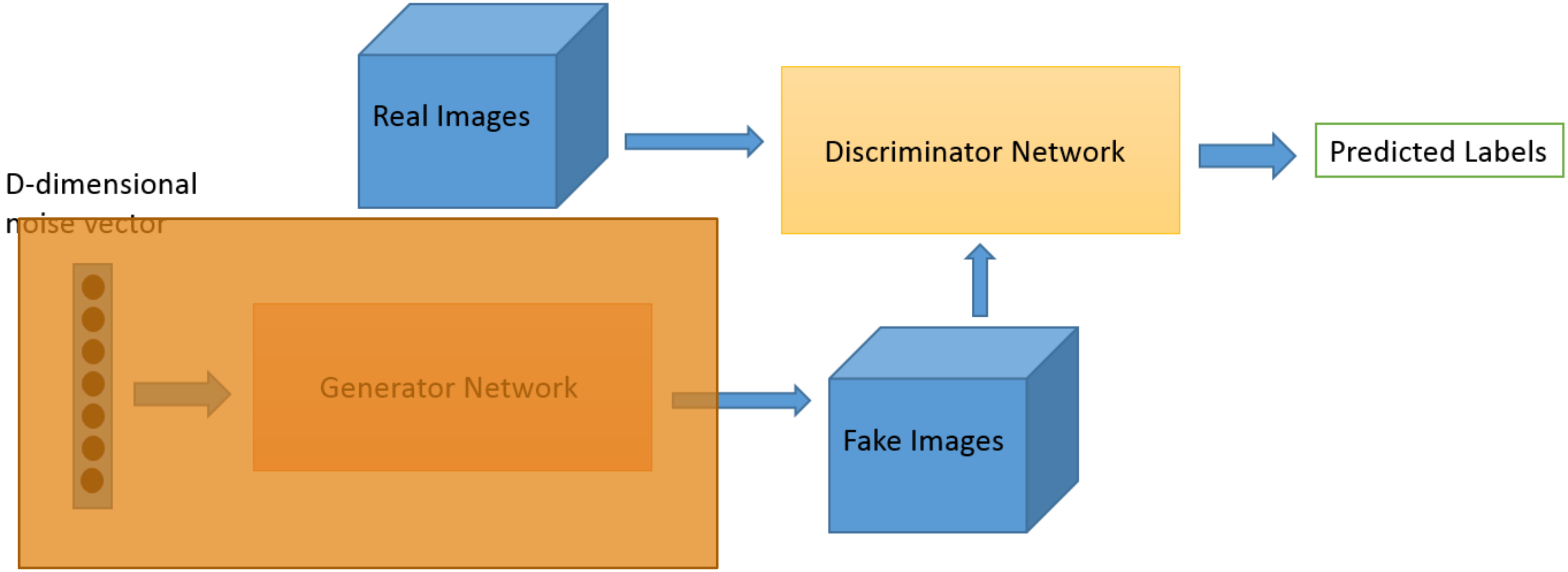
Goodfellow, Ian, et al. "Generative adversarial nets." NIPS 2014

GANs

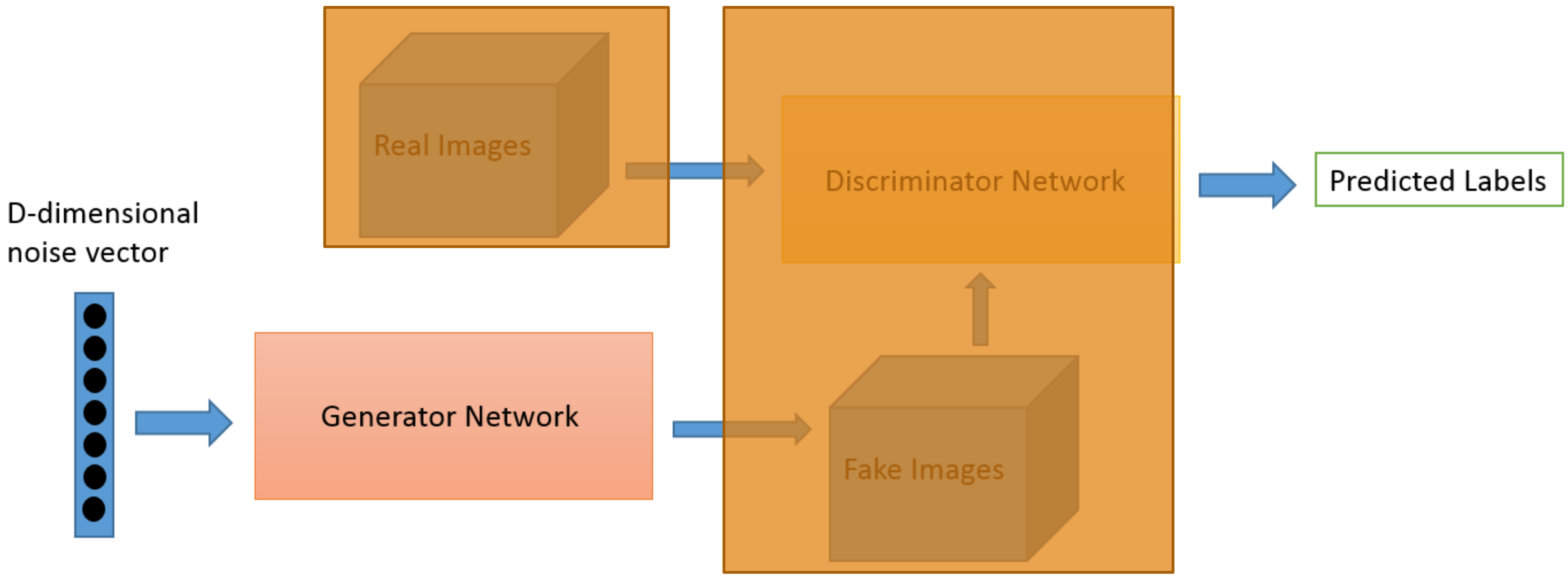


D-dimensional noise vector
Real Images
Discriminator Network
Predicted Labels
Generator Network
Fake Images

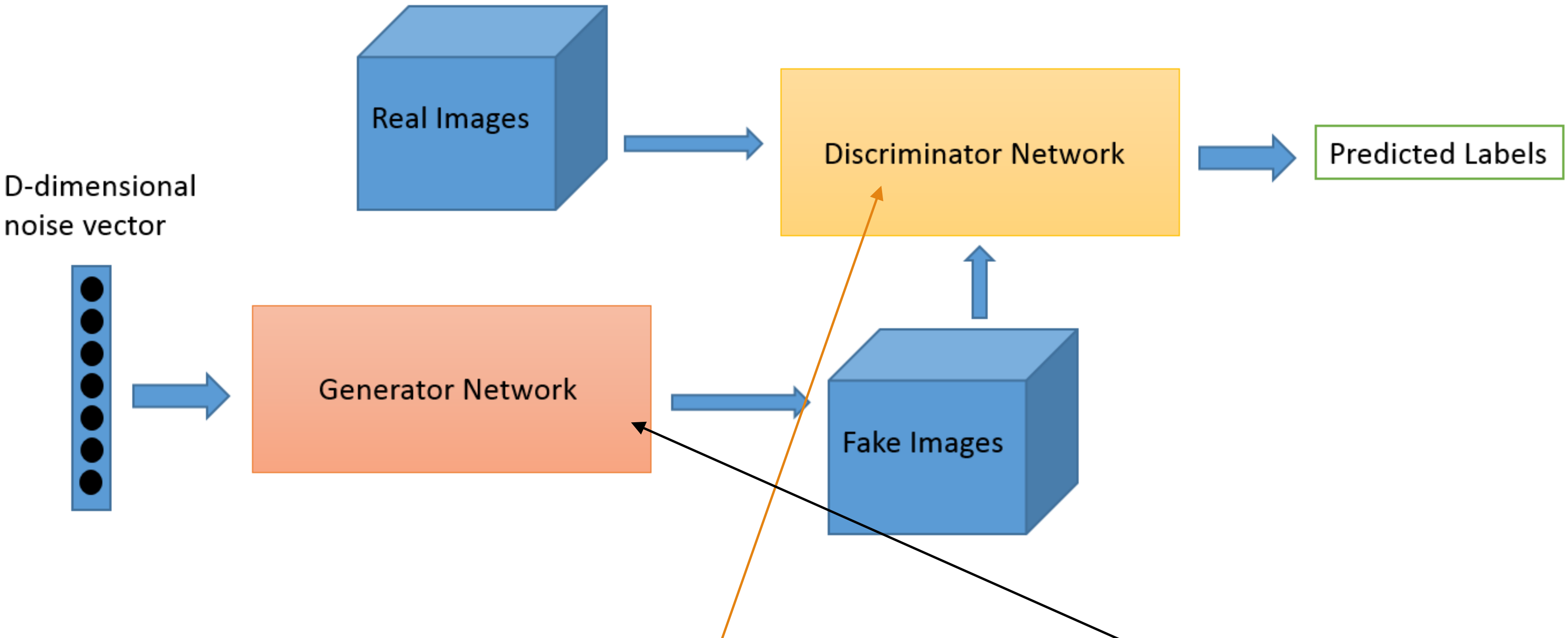
GANs


Real Images
Discriminator Network
Predicted Labels
D-dimensional
noise vector
Generator Network
Fake Images

GANs


Real Images
Discriminator Network
Predicted Labels
D-dimensional
noise vector
Generator Network
Fake Images

GANs



$$\min_{G} \max_{D} = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(z)}[1 - \log D(G(\mathbf{x}))]$$

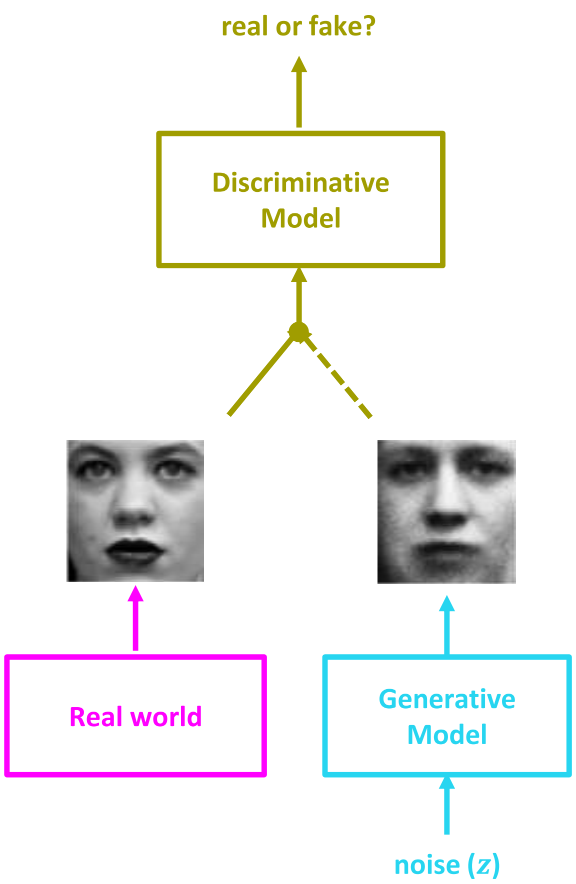
Training Procedure: Basic Idea

G tries to fool D

D tries not to be fooled

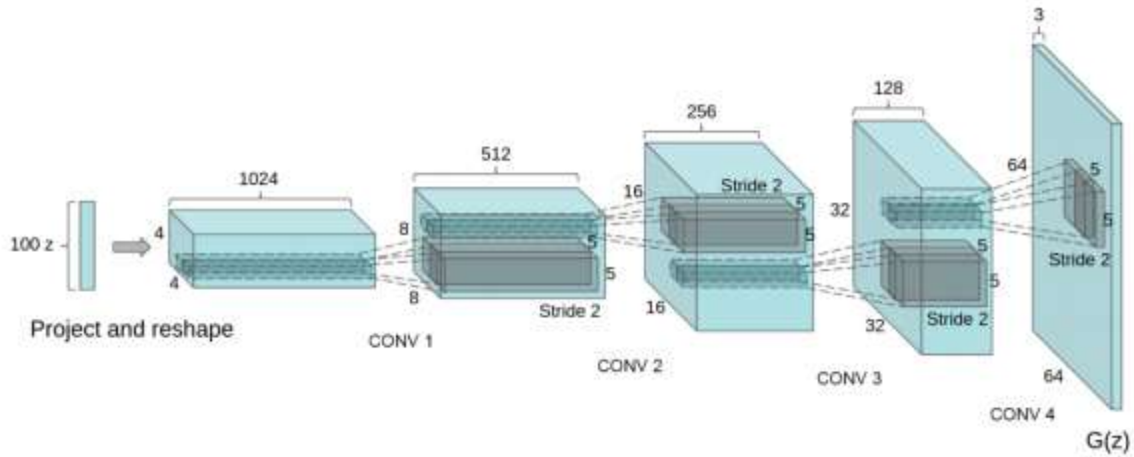
Models are trained simultaneously
- As G gets better, D has a more challenging task
- As D gets better, G has a more challenging task

Ultimately, we don't care about the D
- Its role is to force G to work harder



DCGANs

Deep Convolutional Generative Adversarial Networks



Radford et. al. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ICLR 2016

Progress

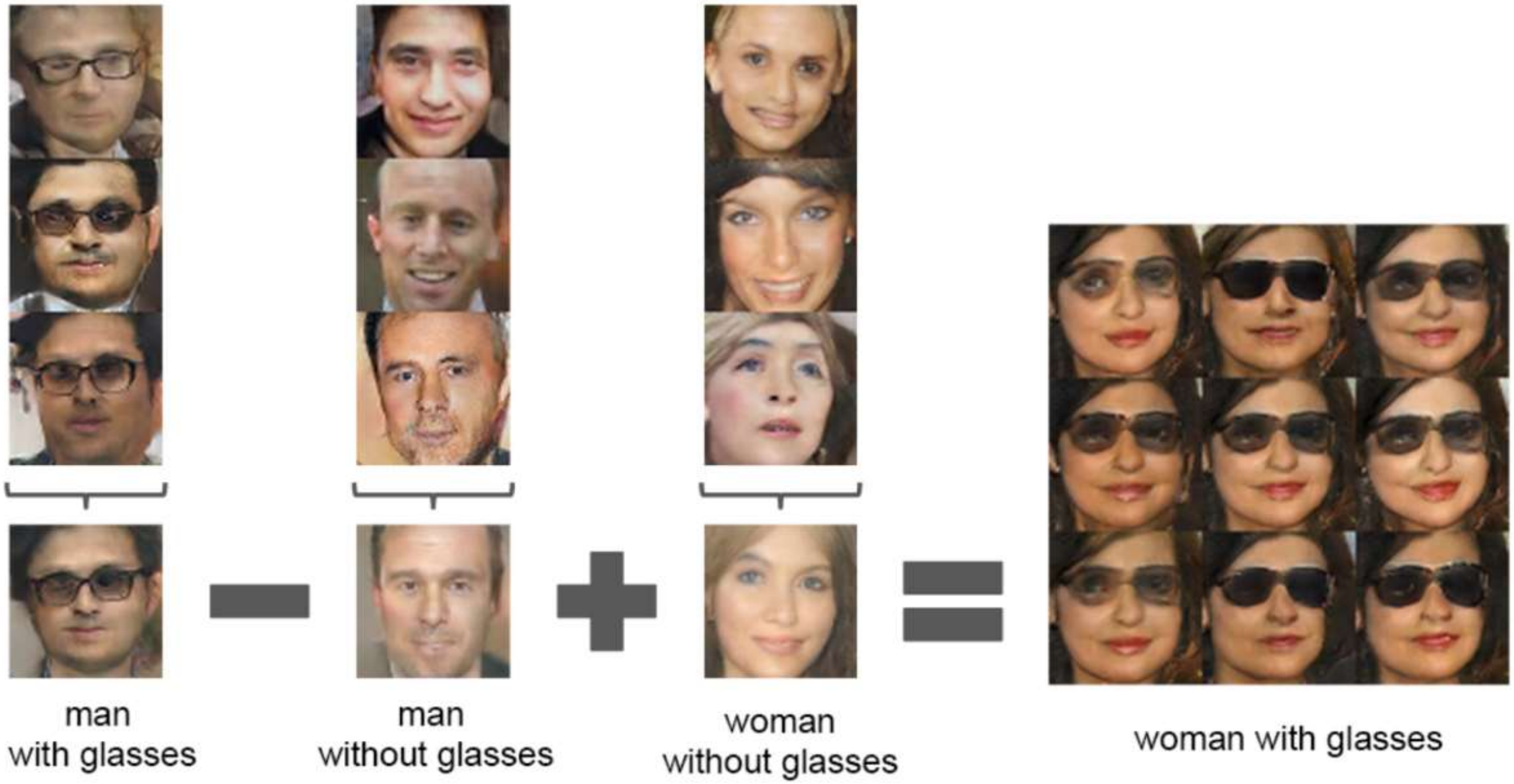
Training Data(CelebA)

Model Samples (Karras et.al., 2018)

4 years of progression on Faces

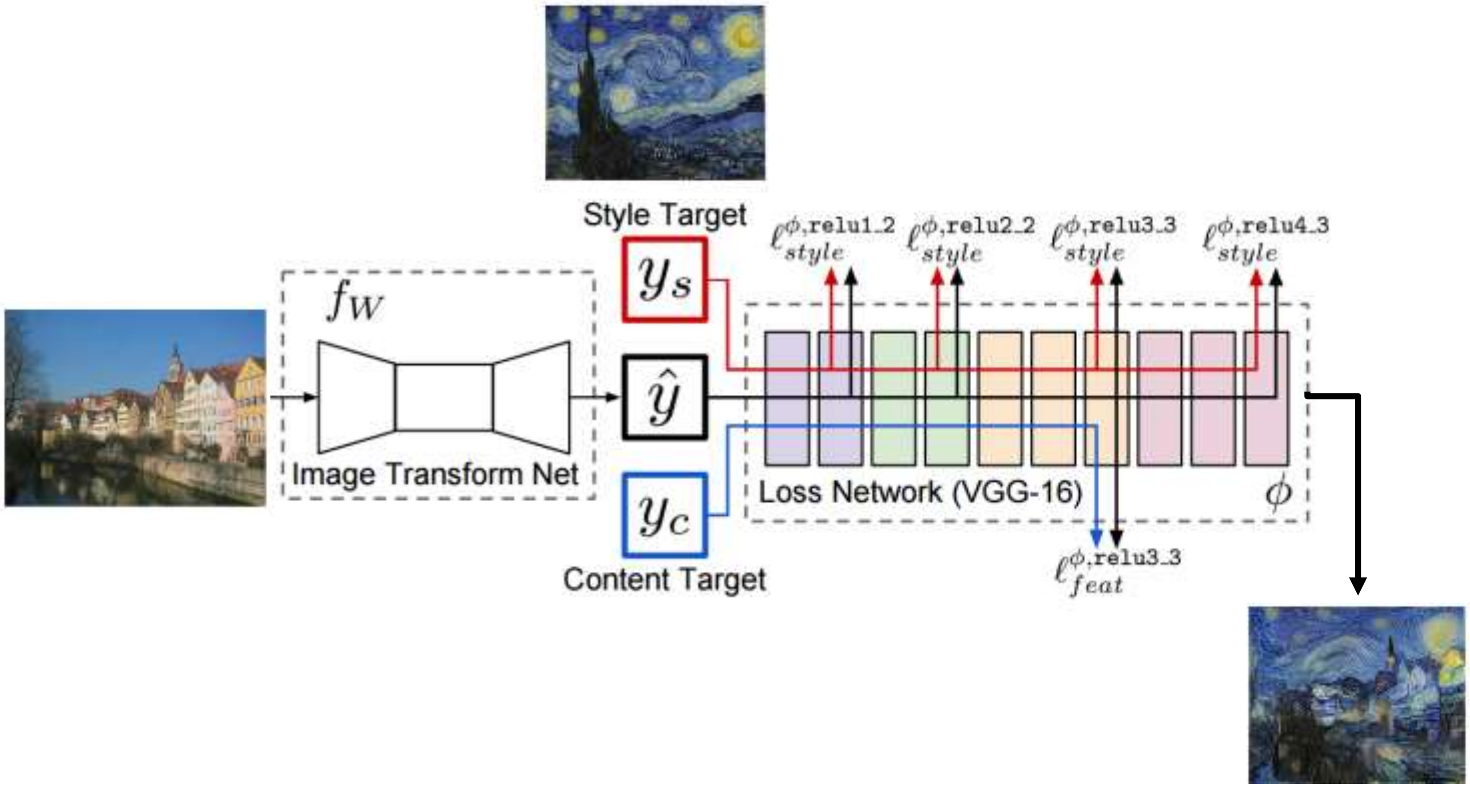
Brundage et al., 2017

Image arithmetic

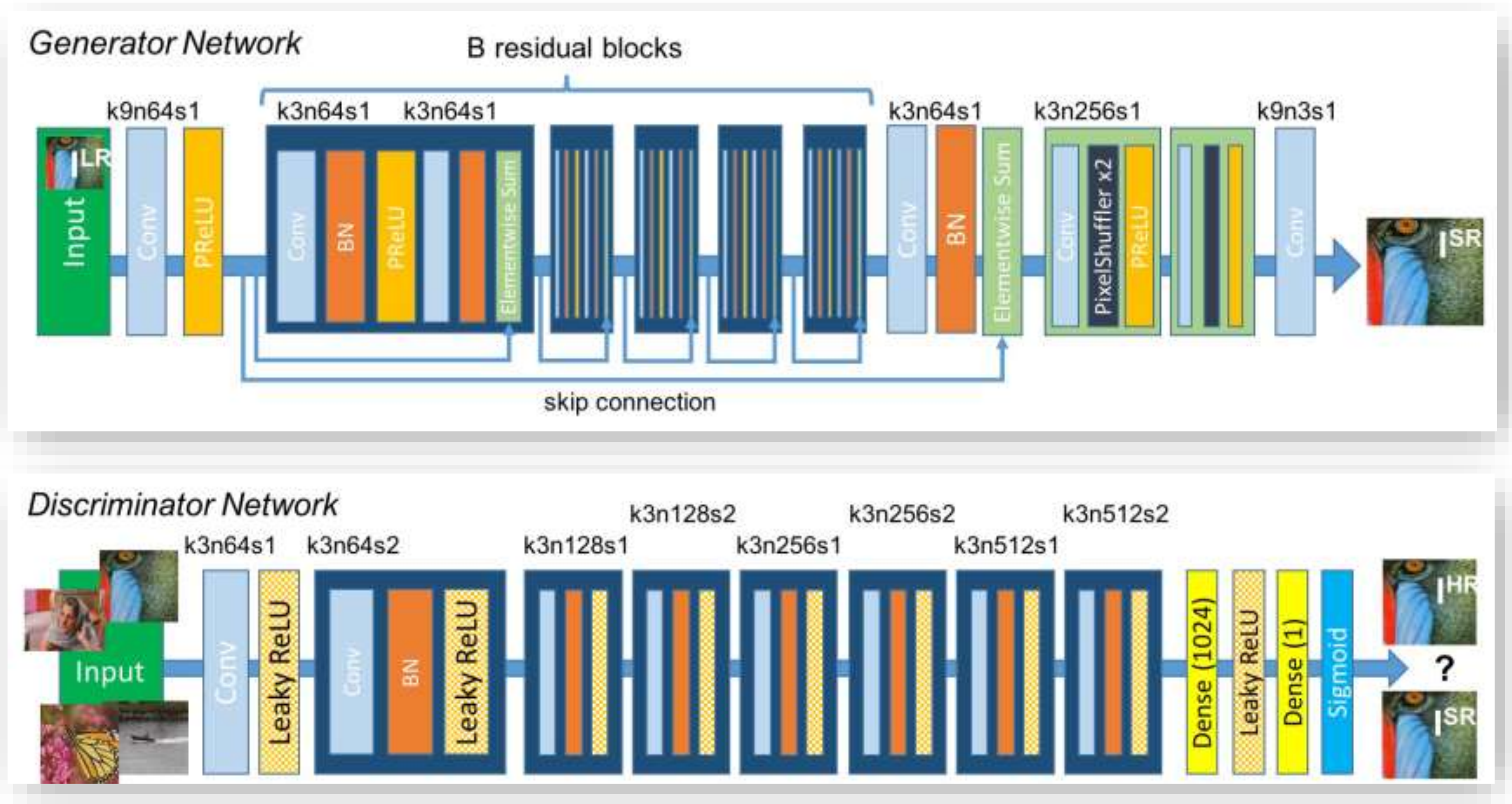


Transition

GAN based Style Transfer



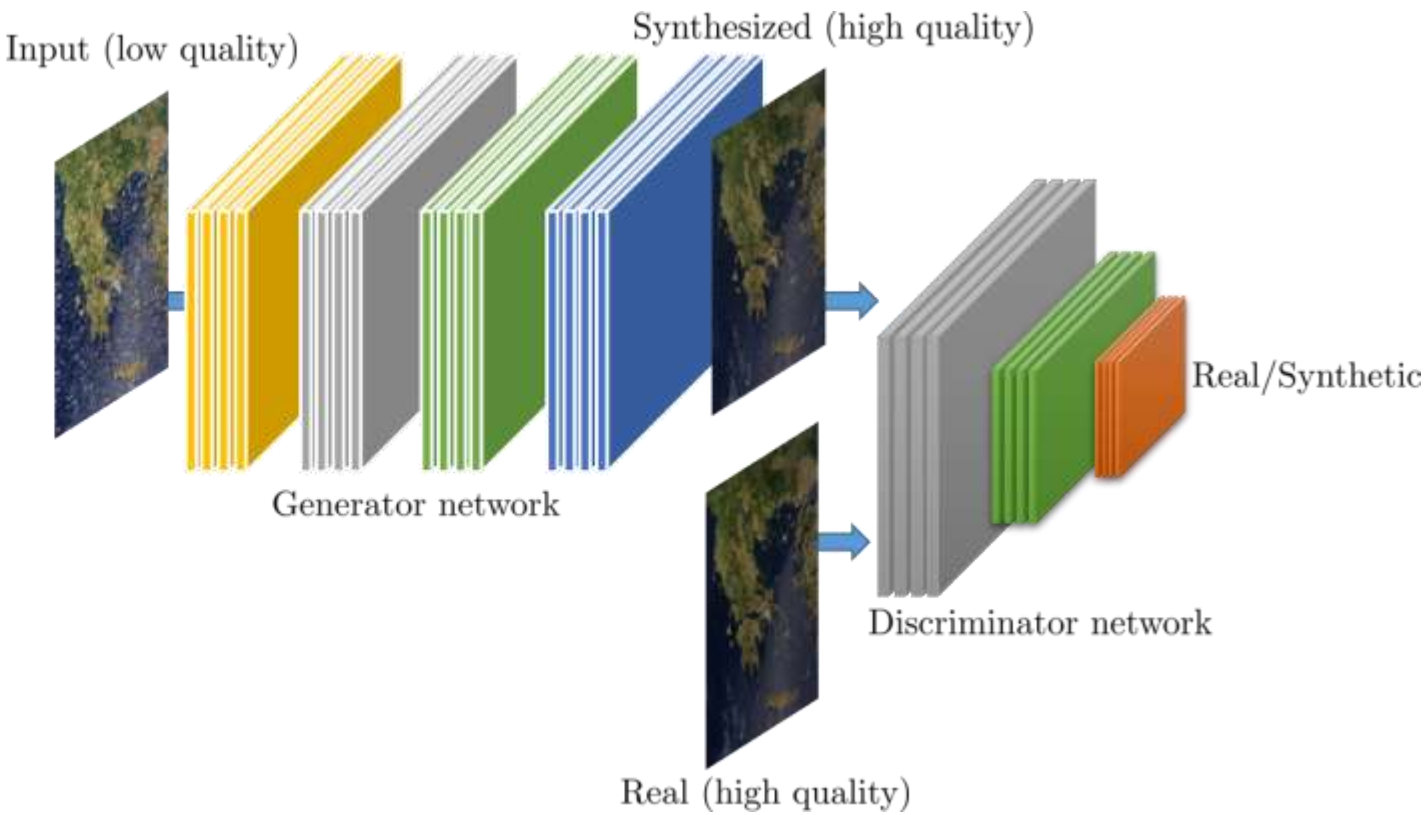
GANs for Super Resolution



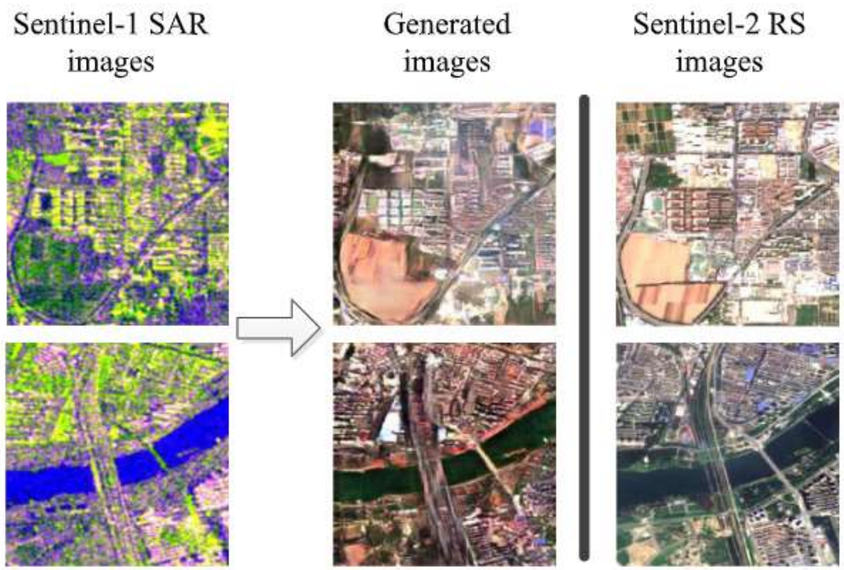
Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." *arXiv preprint arXiv:1609.04802* (2016).

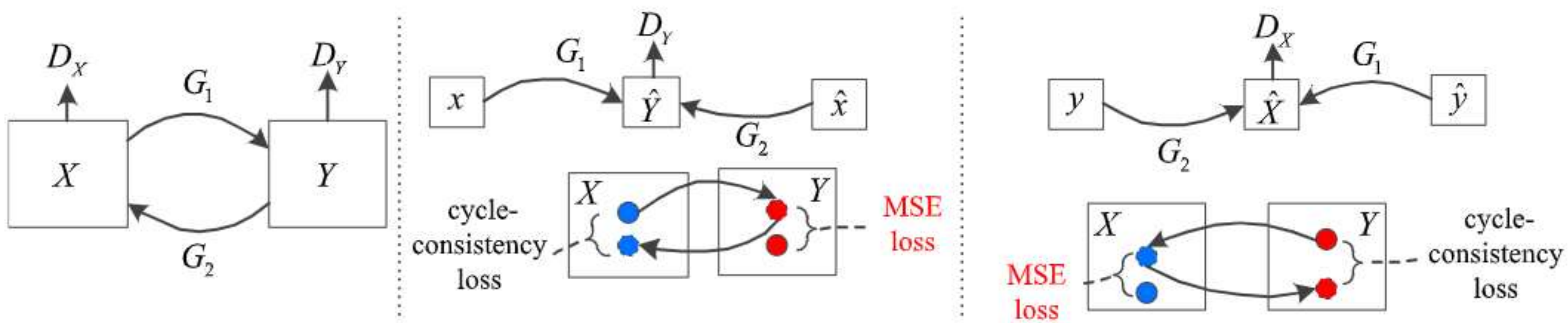
GANs for Super Resolution

SR-GAN for RS data



CycleGAN

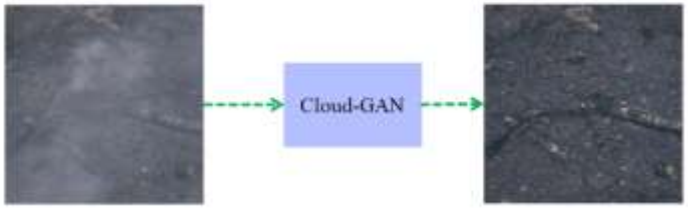


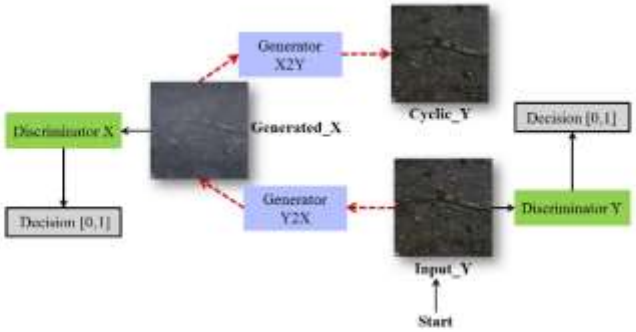


Wang, Lei, et al. "SAR-to-optical image translation using supervised cycle-consistent adversarial networks." *Ieee Access* 7 (2019): 129136-129149.

(a) (b) (c)

CLOUD-GAN

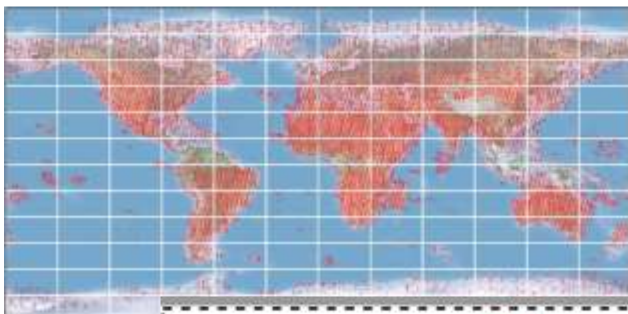


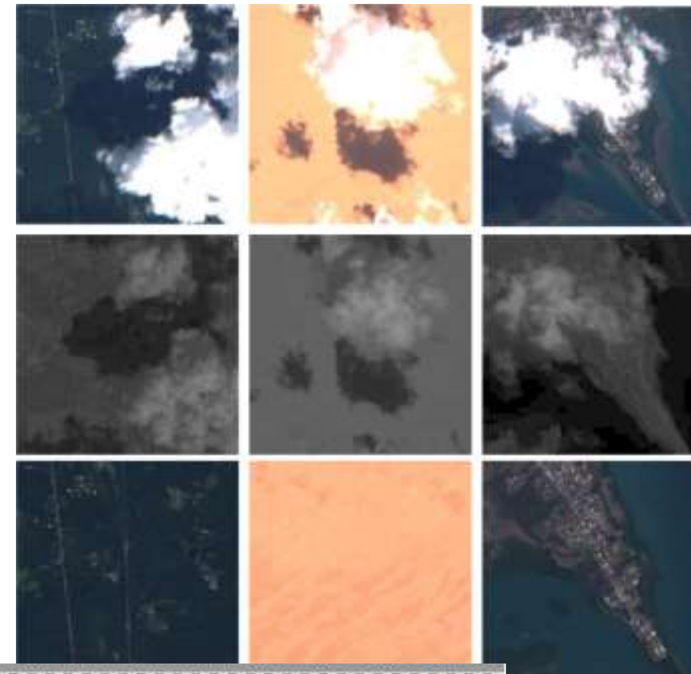


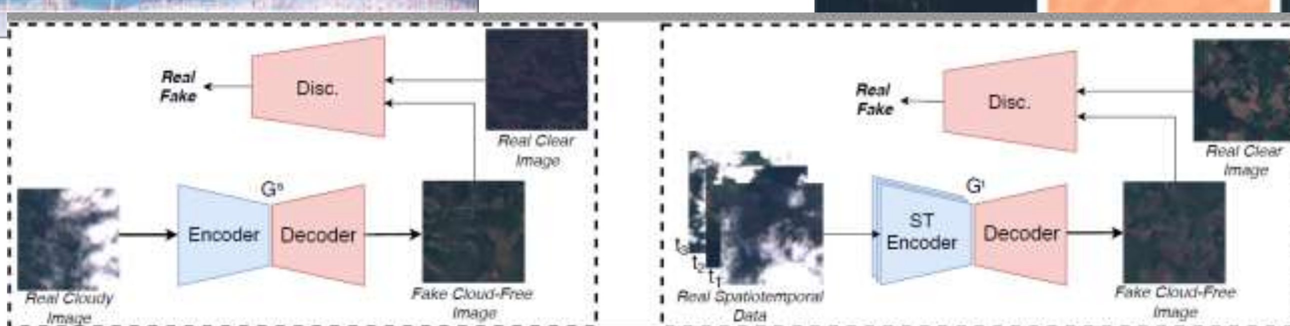
Singh, Praveer, and Nikos Komodakis. "Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks." *IGARSS 2018*

Cloud-removal

- Sentinel 2
- Two-passes
- RGB+IR
- 100K paired-images

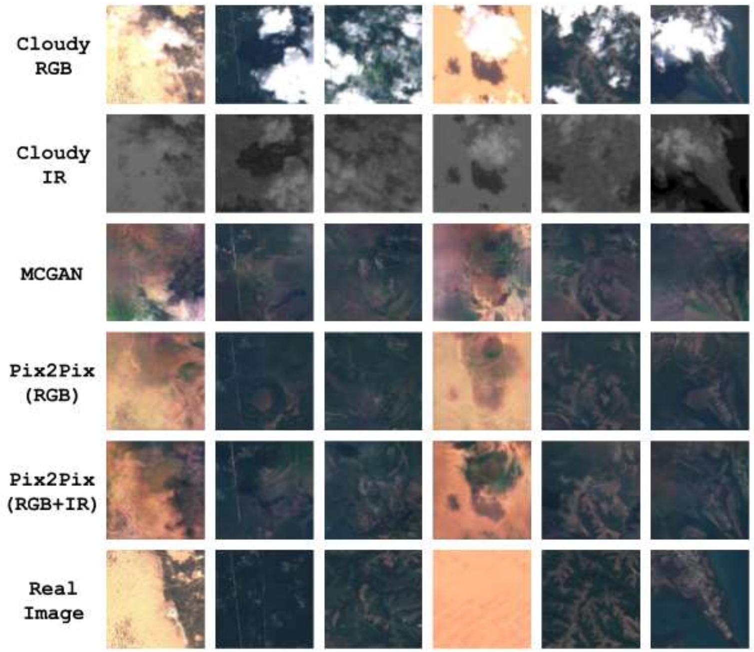


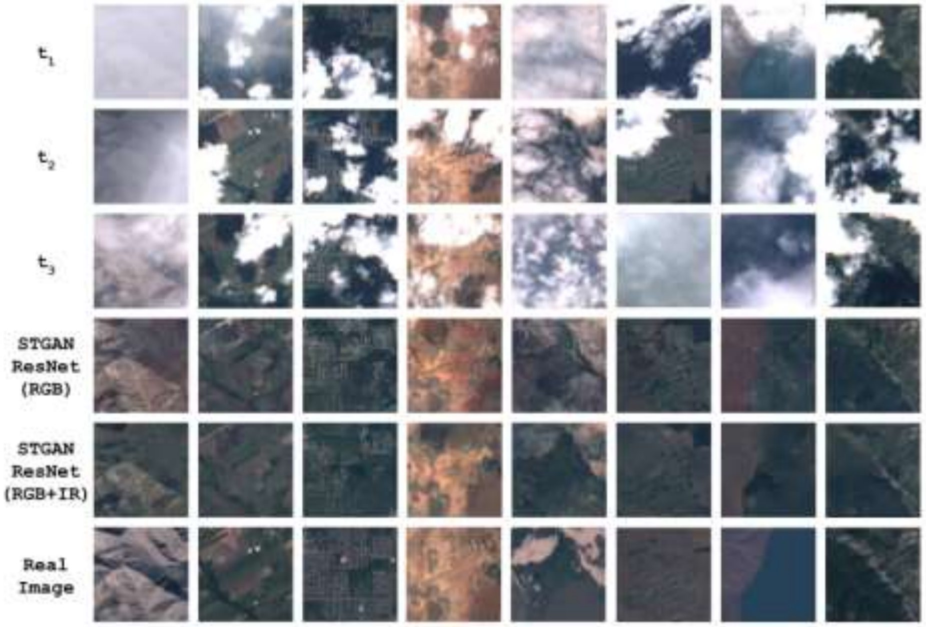




Cloud Removal in Satellite Images Using Spatiotemporal Generative Networks

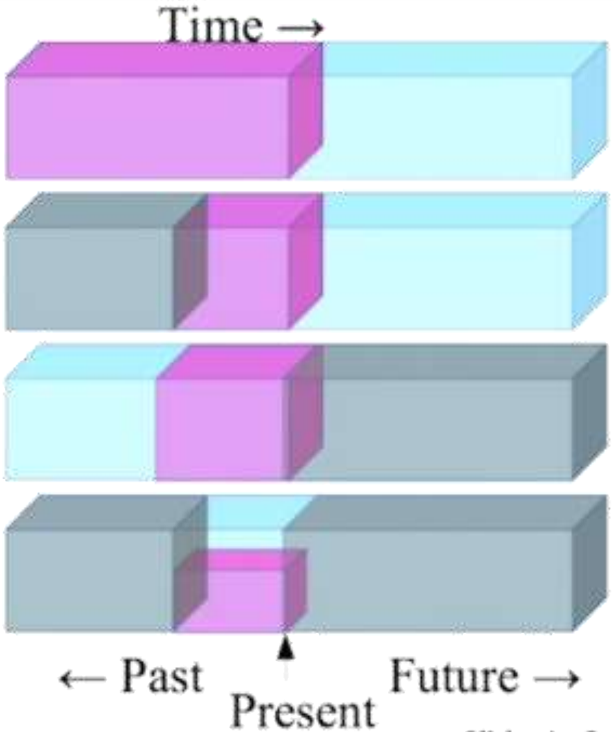
Performance





Self-supervision

- Predict any part of the input from any other part.
- Predict the future from the past.
- Predict the future from the recent past.
- Predict the past from the present.
- Predict the top from the bottom.
- Predict the occluded from the visible
- Pretend there is a part of the input you don't know and predict that.



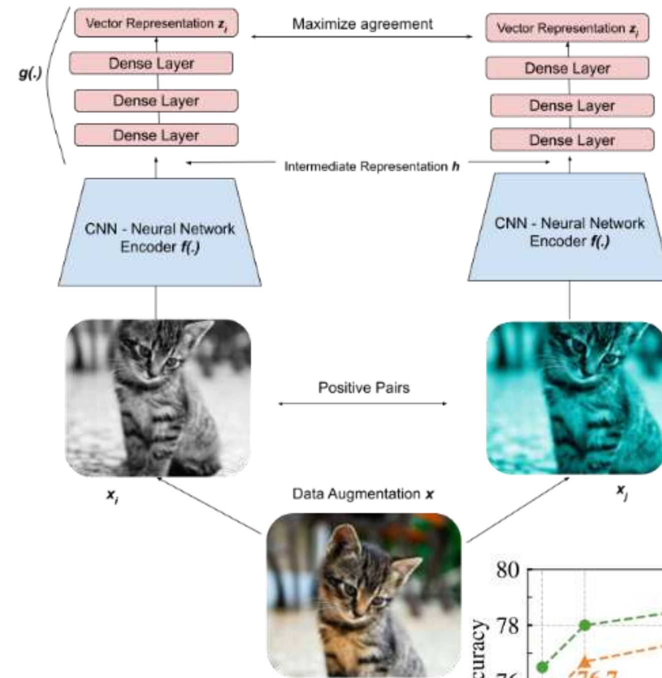
Slide: LeCun

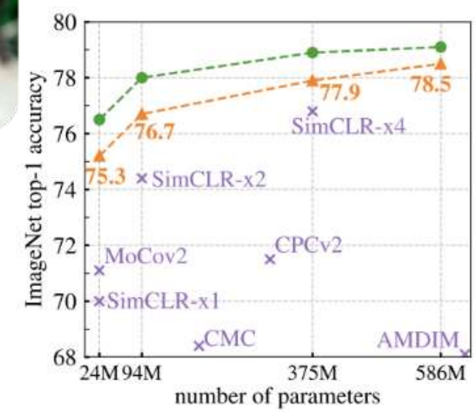
Self-supervised learning

Predicting relative location

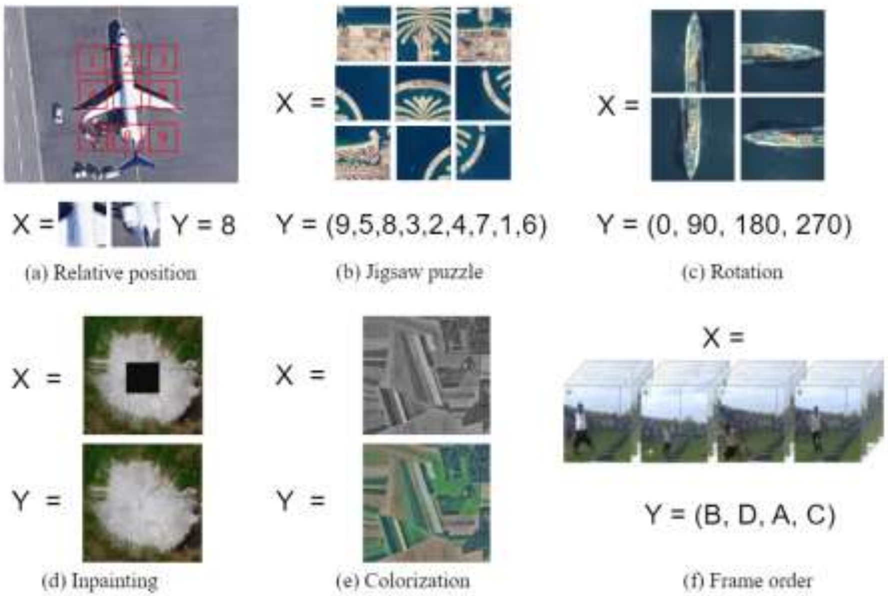
Solving a jigsaw puzzle

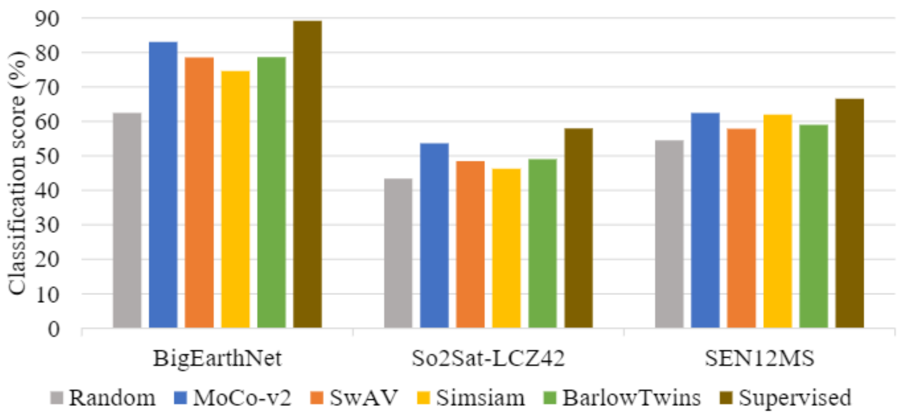
Colorizing an image





SSL

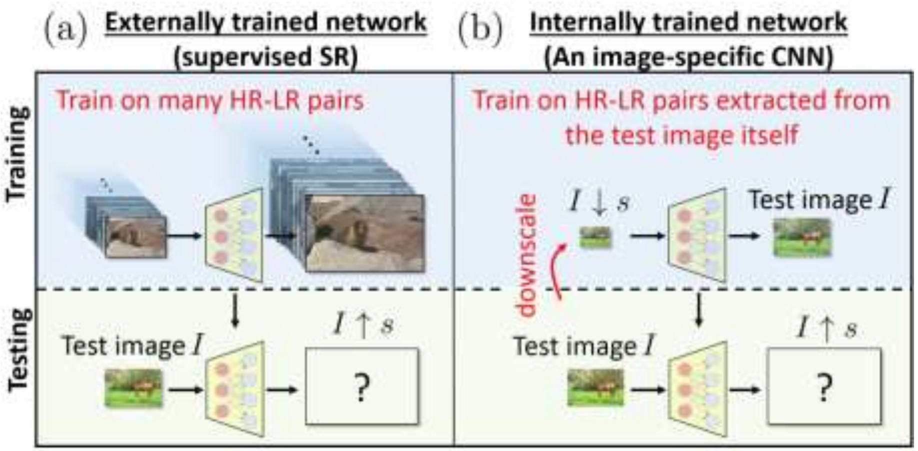




Wang, Yi, et al. "Self-supervised learning in remote sensing: A review." *arXiv preprint arXiv:2206.13188* (2022).

Zero-Shot Learning

A image-specific CNN trained at test time on internal examples extracted solely from the LR test image.

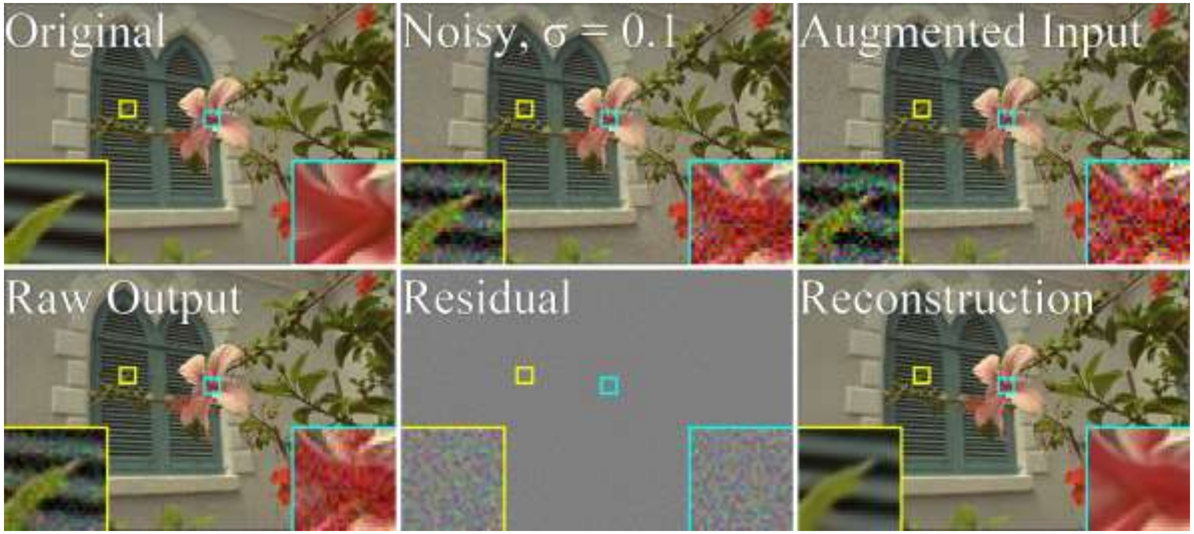


Shocher, Assaf, Nadav Cohen, and Michal Irani. "“zero-shot” super-resolution using deep internal learning." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3118-3126. 2018.

Noisier2Noise

A single noisy realization of each training example

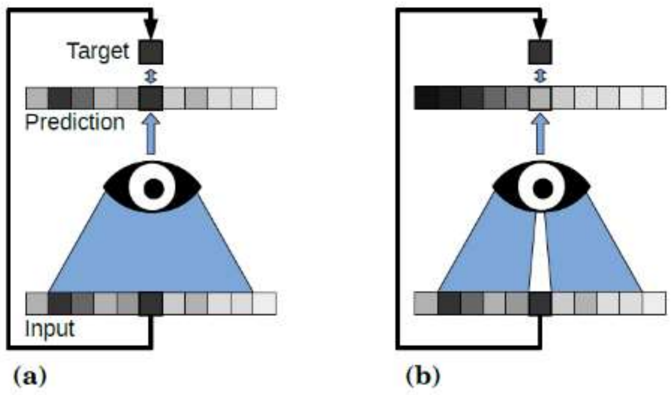
A statistical model of the noise distribution,

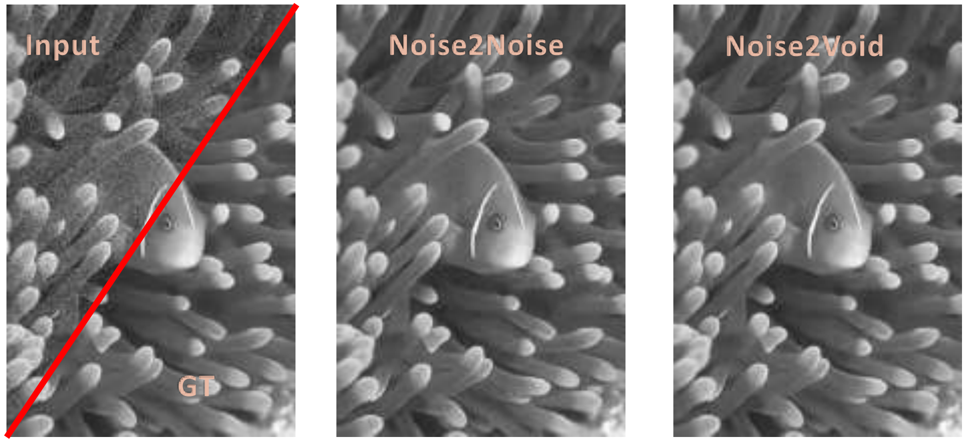


Moran, Nick, Dan Schmidt, Yu Zhong, and Patrick Coady. "Noisier2noise: Learning to denoise from unpaired noisy data." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12064-12072. 2020.

Noise2Void

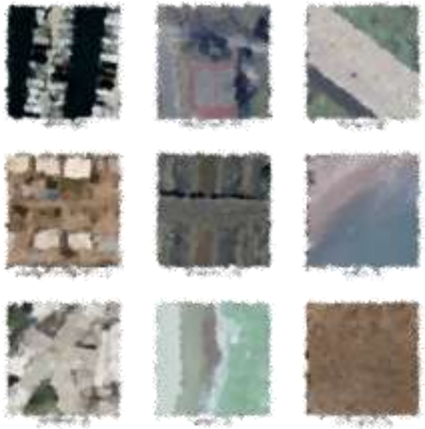
Neighboring pixels are utilized for predicting the value of a central artificially removed pixel.

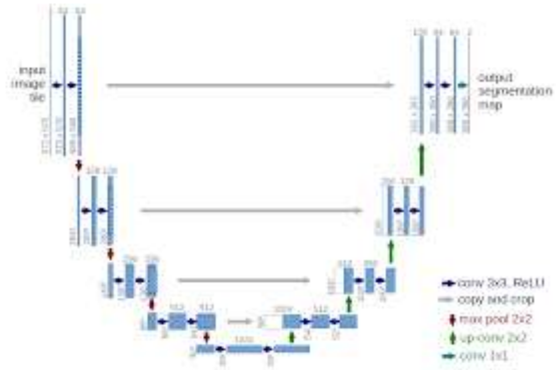


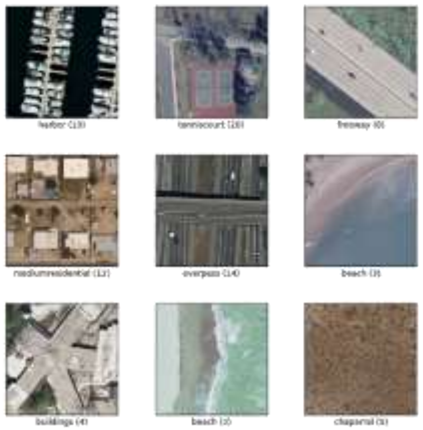


Krull, A., Buchholz, T. O., & Jug, F. (2019). Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2129-2137).

Hands-on




input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1


harbor
tenniscourt
freeway
mediumresidential
overpass
beach
buildings
beach
chaparral