Towards Realistic and Trustworthy Super-Resolution for Multispectral Remote Sensing Images

Luis Gómez-Chova, Cesar Aybar, Simon Donike



ISP · Image & Signal Processing
Universitat de València

IEEE GRSS IADF Image Analysis and Data Fusion SCHOOL COMPUTER VISION FOR EARTH OBSERVATION

525 anys 1499 - 2024 VNIVERSITAT DE VALÈNCIA

http://isp.uv.es

IADF | Sep 17, 2025

Super-Resolution in Remote Sensing

What is Super-Resolution?

Enhancing image spatial resolution while introducing HF details

low-resolution satellite data (LR)

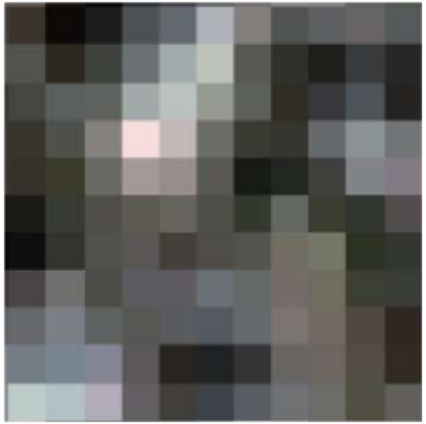
4x

high-resolution imagery (HR)

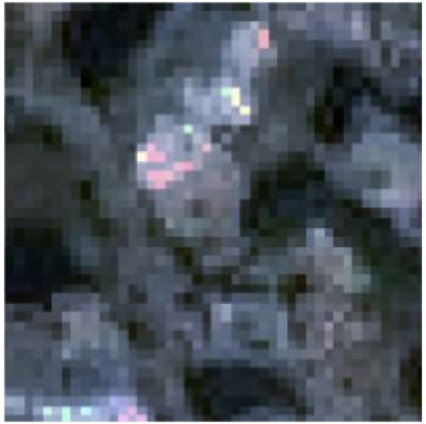
Why in Remote Sensing?

- HR imagery useful for real-world applications
- Access to HR satellite imagery is unevenly distributed:
    - Private companies provide VHR imagery but usually are too expensive
    - NASA/ESA offer global high temporal resolution but at LR resolution

SR factor x4: 16 pixels replace 1 pixel



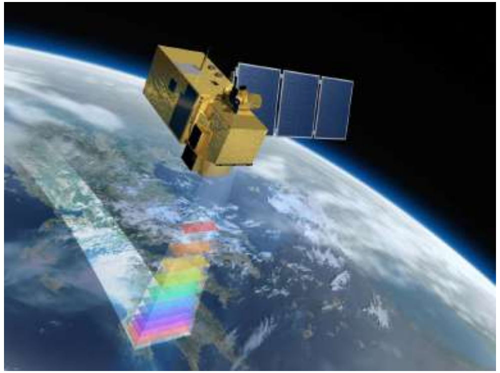
Sentinel-2 (10m)

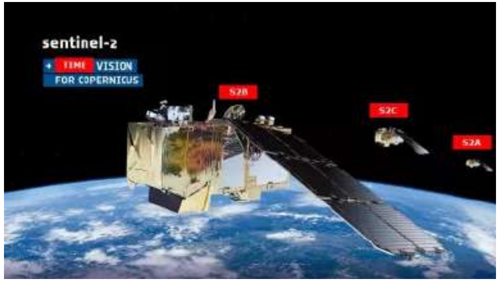


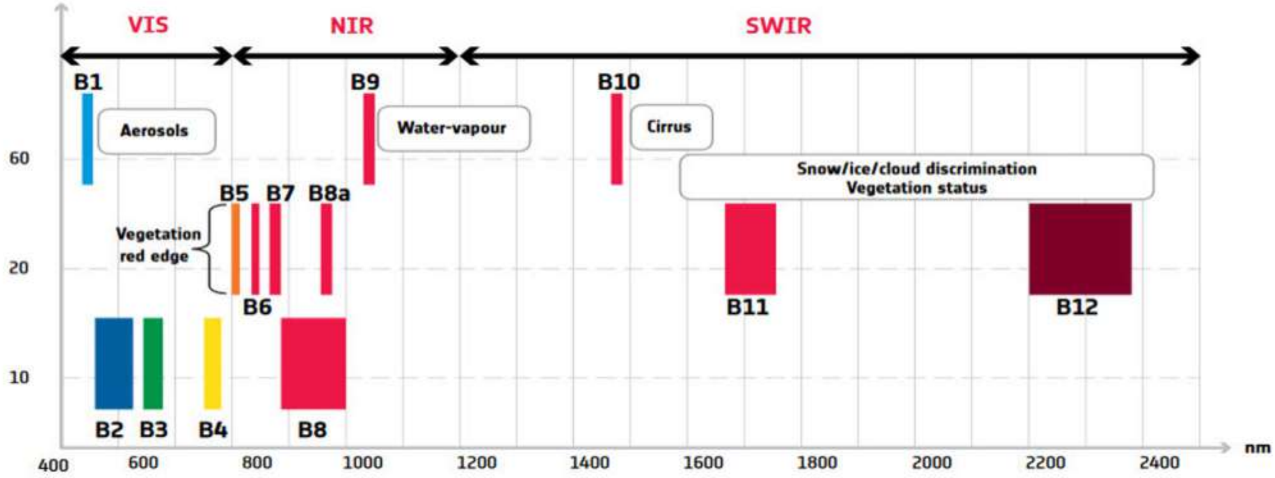
HR Reference (2.5m)

Why with Sentinel-2?

- Sentinel-2: high temporal frequency (3 satellites), global coverage, free access, broad spectral bands





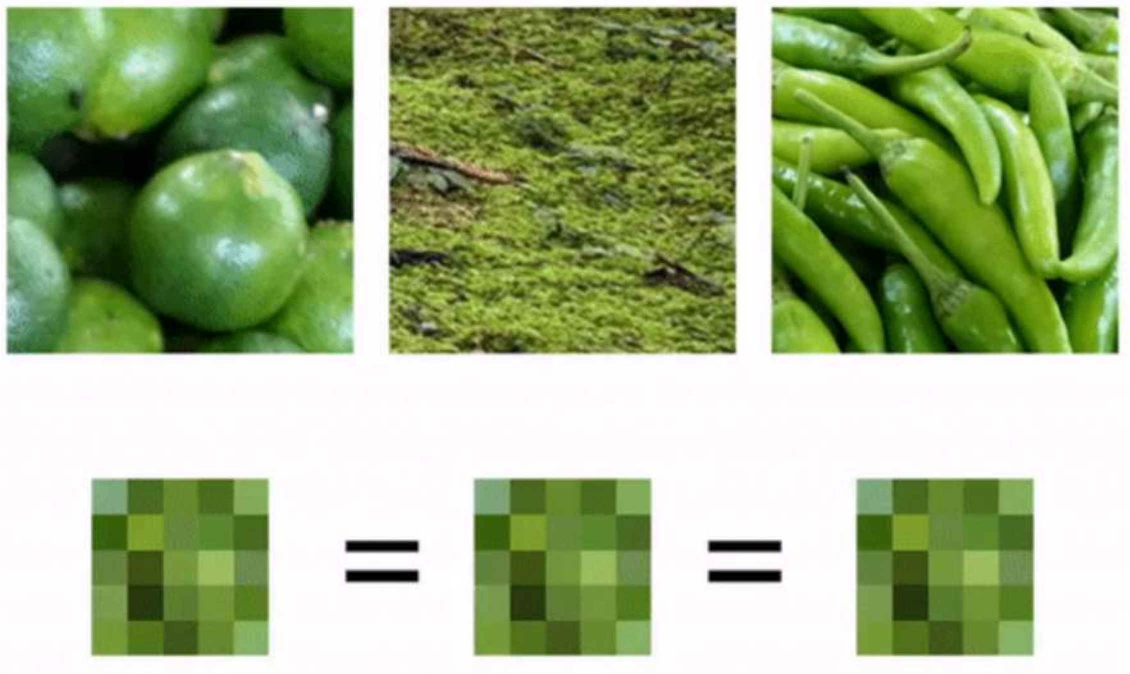


ESAOpenSR

esa

VNIVERSITAT DE VALÈNCIA

Where is the trick?



Ill-posed problem
inverse mapping from
LR to HR can yield an
infinite number of
possible solutions.

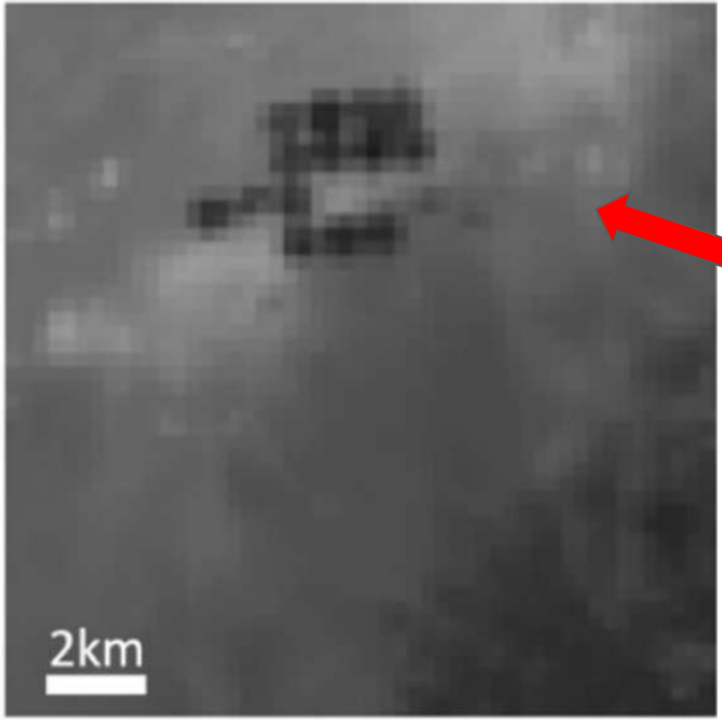
One Low- Resolution Image can come from many High-Resolution Images

NTIRE 2022 Challenge on Efficient Super-Resolution: Methods and Results

https://arxiv.org/abs/2205.05675

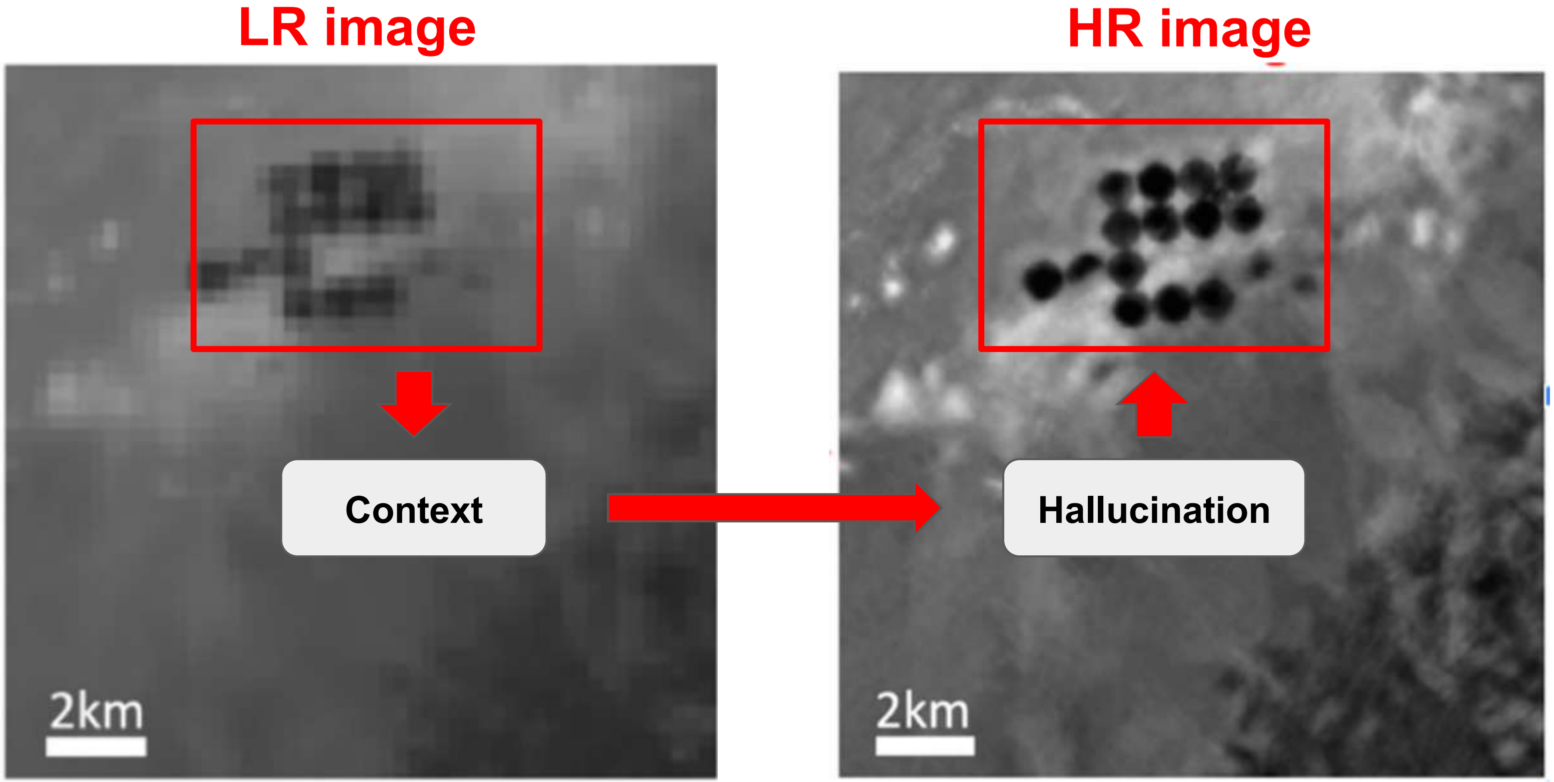
What is it?

LR image



What is it?

What is it?



Classical Methods

Interpolation-based

- Compute pixel values for the HR grid based on known values from the LR image: Nearest Neighbor, Bilinear, Bicubic
- Benchmark SR against simple interpolation baselines

Physically based

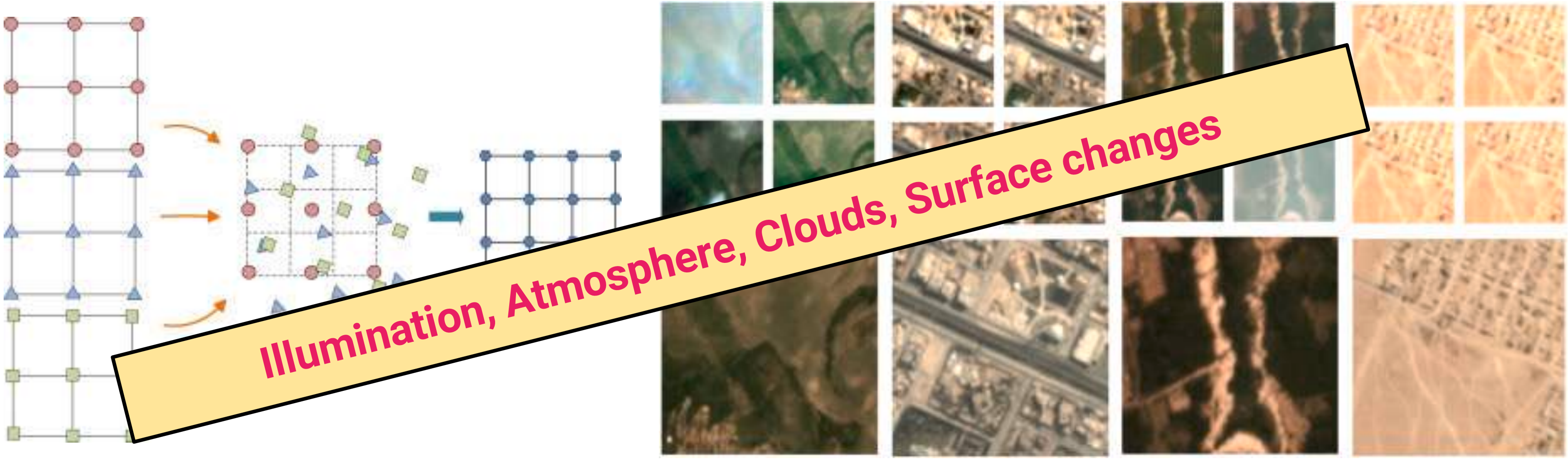
- Incorporates knowledge of the remote sensing image formation (atmospheric effects, point spread [illegible]n)

Nearest Neighbor Bilinear Bicubic

Multi-Image or Single-Image (MISR vs SISR)

Multi-Frame Super-Resolution

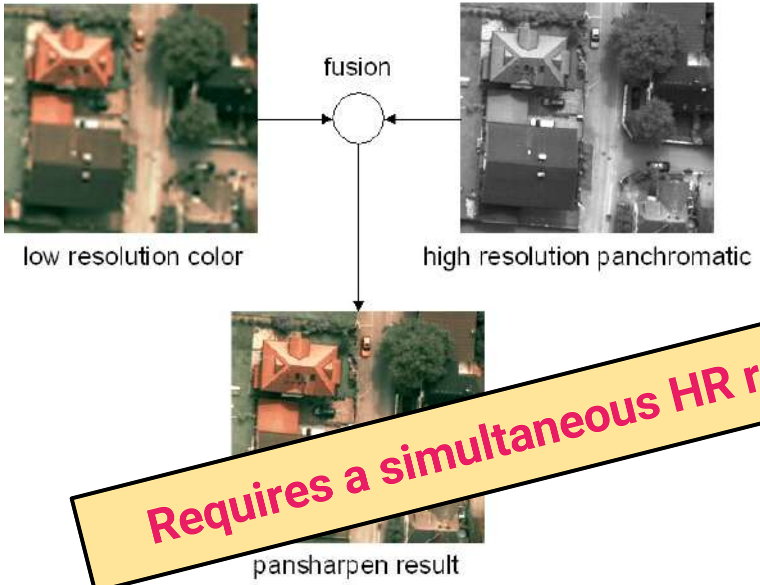
- Combine multiple LR images of the same scene to reconstruct a single HR image.
- LR images are overlapping but differently sampled (slightly different angles or times)

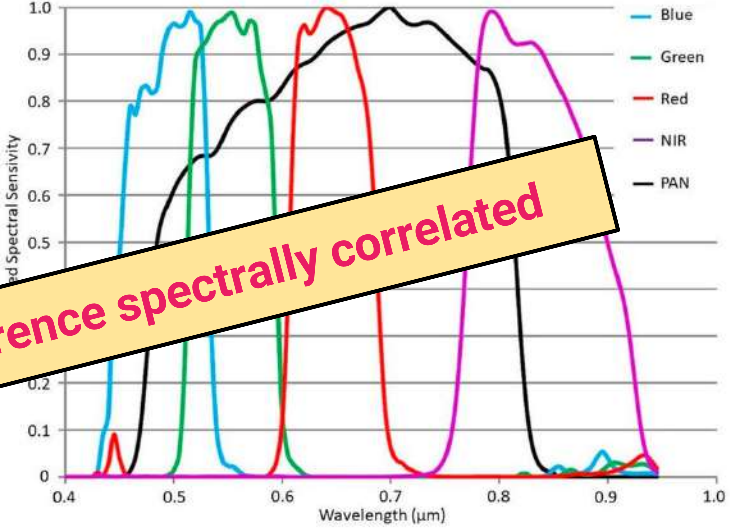


Multi-Image or Single-Image (MISR vs SISR)

Data Fusion (Pansharpening)

- Merges HR panchromatic image with a LR multi-spectral image to produce a HR multi-spectral image.





Requires a simultaneous HR reference spectrally correlated

Deep Learning Methods

Single Image Super Resolution (SISR)

- Reconstructs a HR image from a single LR input.

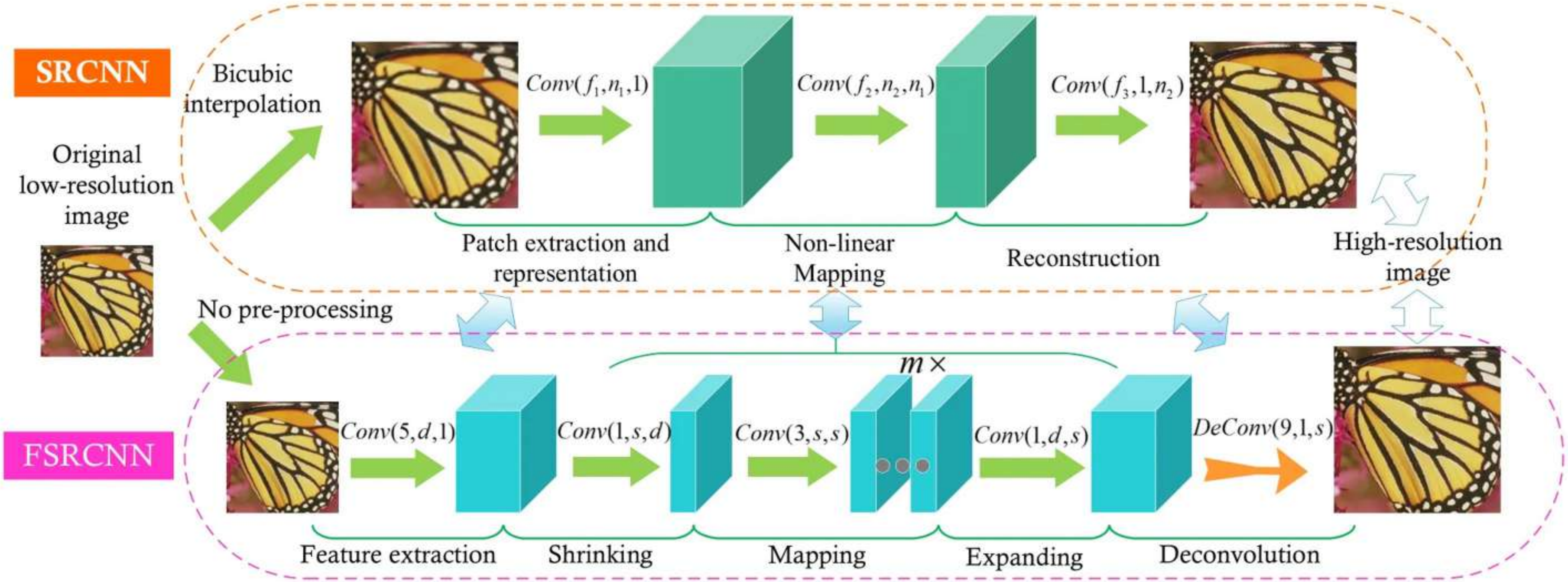
Supervised Deep Learning (DL)

- Requires large datasets of LR-HR image pairs to learn complex mappings

Common architectures

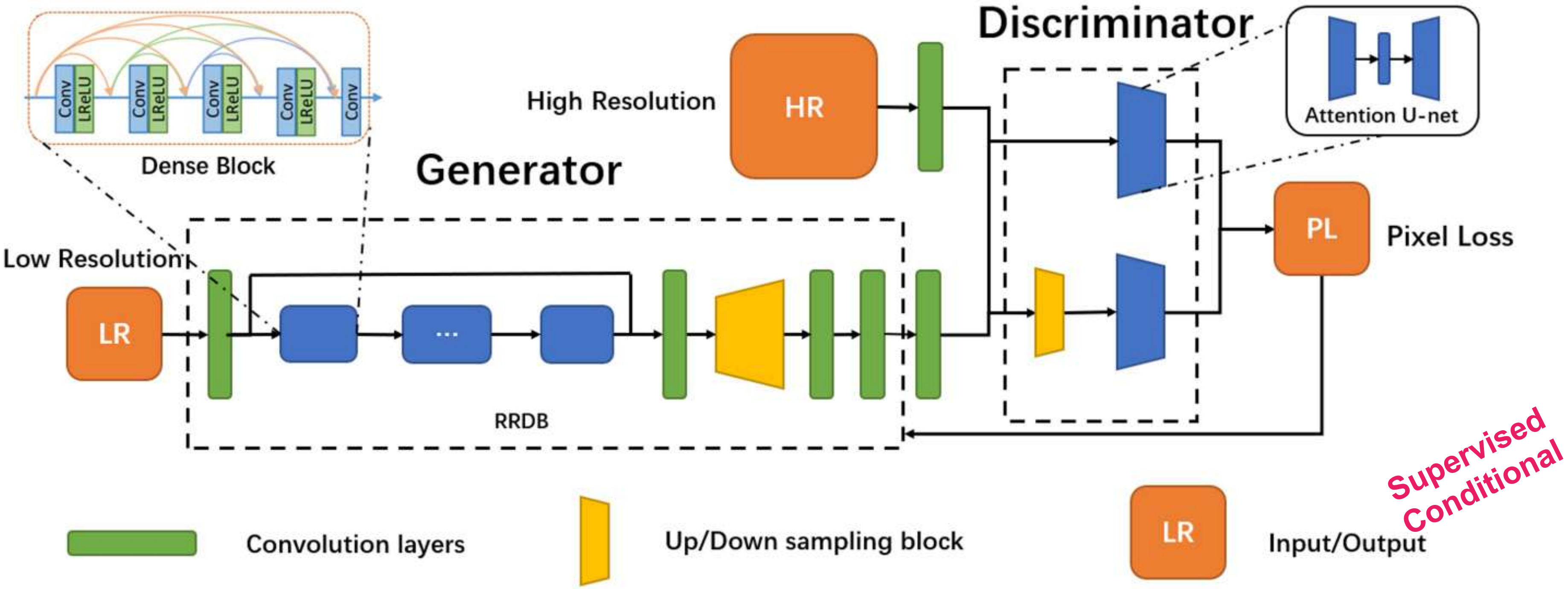
- *Convolutional Neural Networks*: end-to-end mapping LR $\longrightarrow$ HR
- *Generative Adversarial Networks*: generator-discriminator framework to produce perceptually pleasing HR
- *Diffusion Models*: generative models that progressively add noise to an image and then learn to reverse the process.

Super-Resolution Convolutional Neural Network (SRCNN)



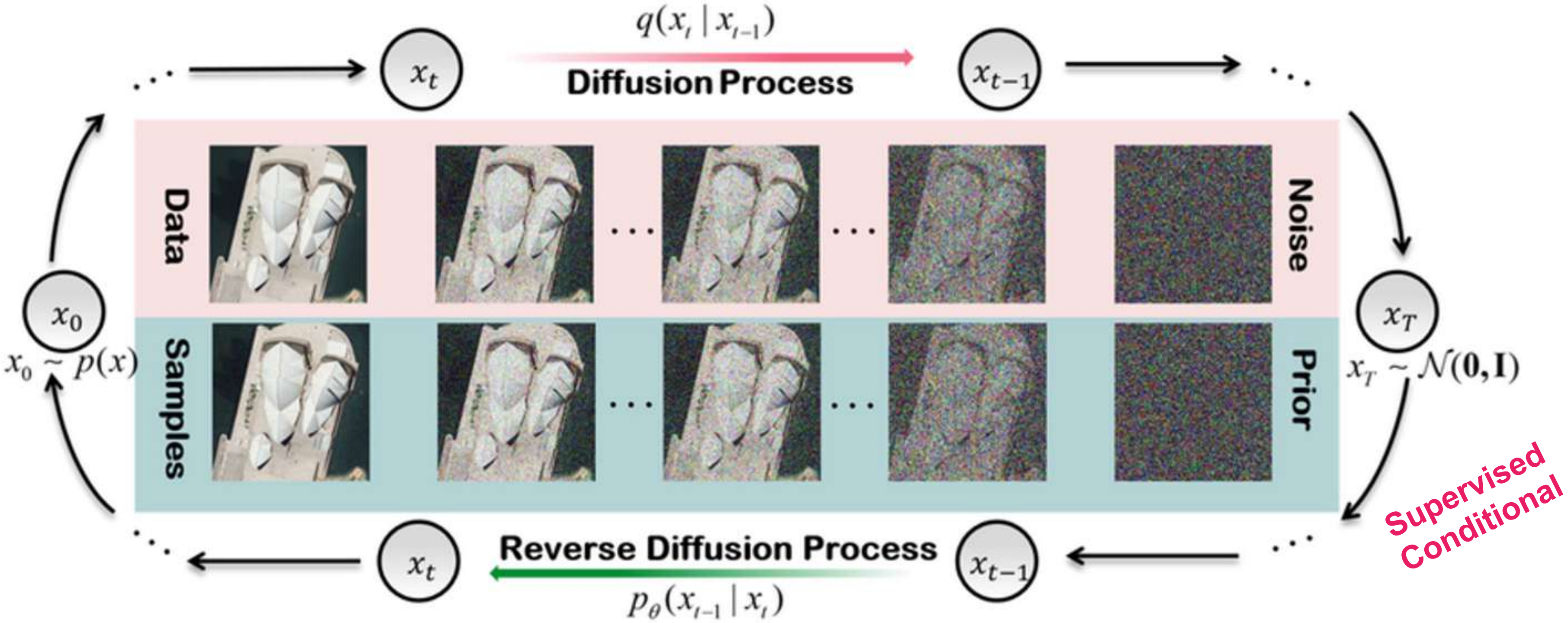
SRCNN learns end-to-end mapping from an upsampled LR image (bicubic) to the HR target.
FSRCNN introduces a deconvolution layer as the last layer.

Enhanced SR Generative Adversarial Network (ESRGAN)



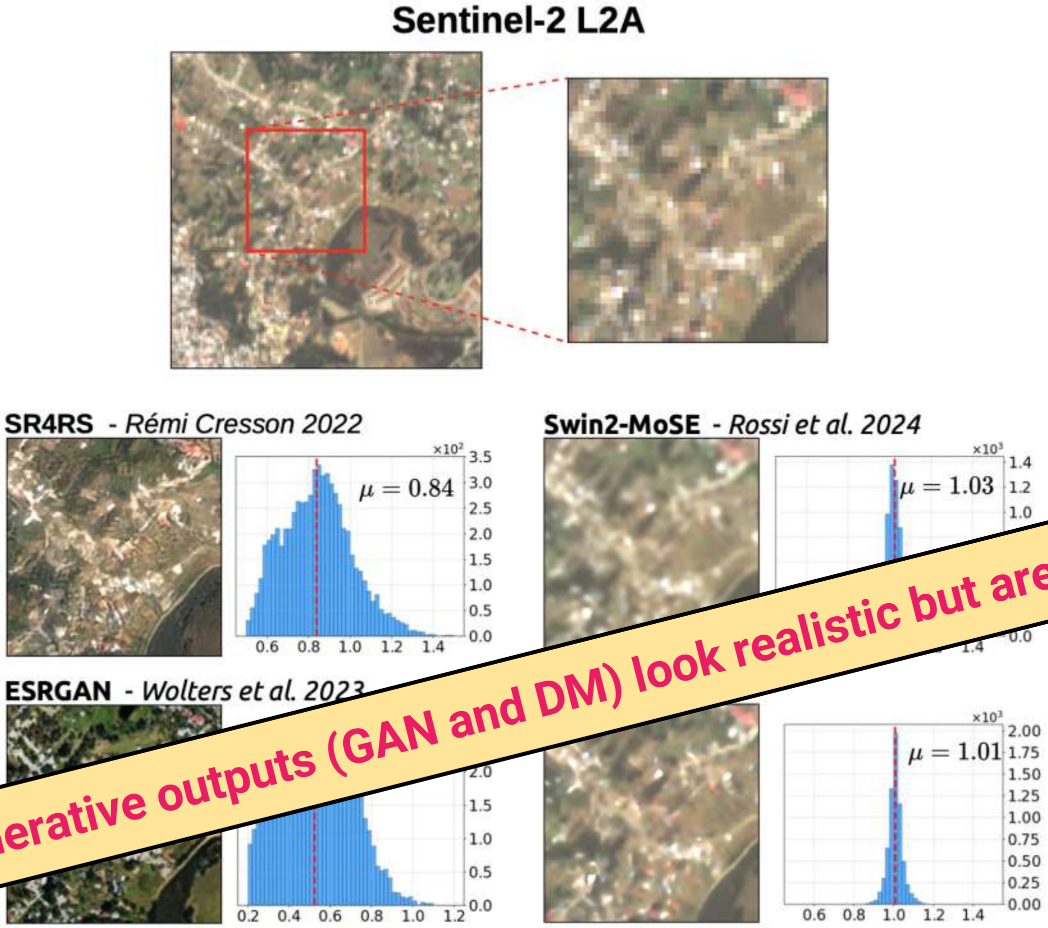
ESRGAN improves SRGAN components: architecture, adversarial loss and perceptual loss.

Diffusion Models for Super-Resolution



Diffusion probabilistic model for conditional image generation: SR through a stochastic denoising process. Inference starts with pure Gaussian noise and iteratively refines the noisy output using a U-Net trained on denoising at various noise levels.

Sentinel-2 L2A image and SOTA SR models



Generative outputs (GAN and DM) look realistic but are wrong

Outline

- **Datasets** for SR training and benchmarking.
- **SR algorithms** based on Deep Learning (DL) aiming to reconstruct HR images from LR inputs.
- **Evaluation** through pixel-wise and dedicated SR-specific metrics on benchmarking datasets.
- **Utility** of SR imagery in downstream applications (use cases).


Spoiler
Sentinel-2 image

Spoiler

LDM-SR (Benevento)

Spoiler

Google Maps RGB

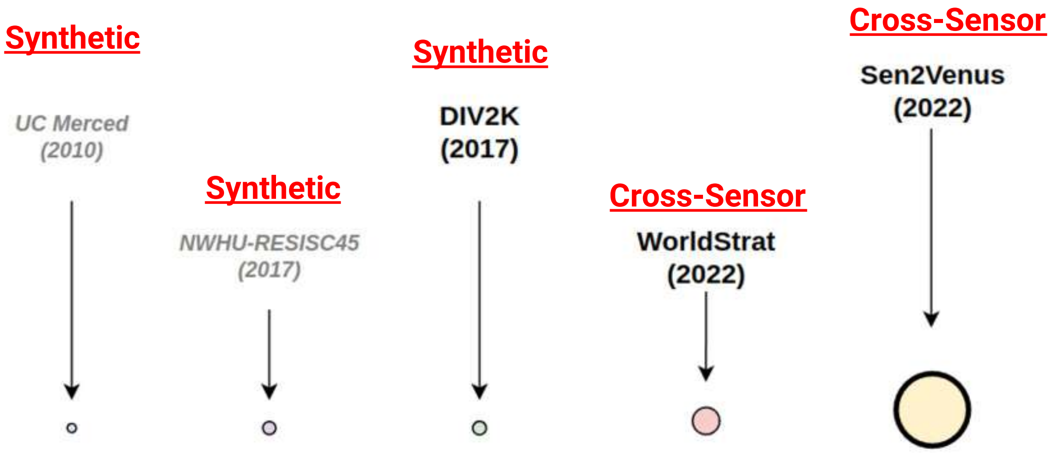
Datasets for SR in Remote Sensing

Available Datasets

	Synthetic *UC Merced (2010)*	Synthetic *NWHU-RESISC45 (2017)*	Synthetic **DIV2K (2017)**
# HR pixels	0.1×10^9	2.1×10^9	2.8×10^9
Amount	2100	31500	1000
Size	(256, 256)	(256, 256)	(1972, 1437)
Description	farmland, bushes, highways, overpasses, etc	airports, basketball, residential, ports, etc	people, scenery, animal, decoration, etc

"Synthetic" dataset: LR image generated from HR image.
"Cross-sensor" dataset: LR image from another sensor.

Available Datasets



	UC Merced (2010)	NWHU-RESISC45 (2017)	DIV2K (2017)	WorldStrat (2022)	Sen2Venus (2022)
# HR pixels	0.1×10^9	2.1×10^9	2.8×10^9	4.4×10^9	8.7×10^9
Amount	2100	31500	1000	3515	132 955
Size	(256, 256)	(256, 256)	(1972, 1437)	(a, b)	(256, 256)
Description	farmland, bushes, highways, overpasses, etc	airports, basketball, residential, ports, etc	people, scenery, animal, decoration, etc	**worldwide**	**worldwide**

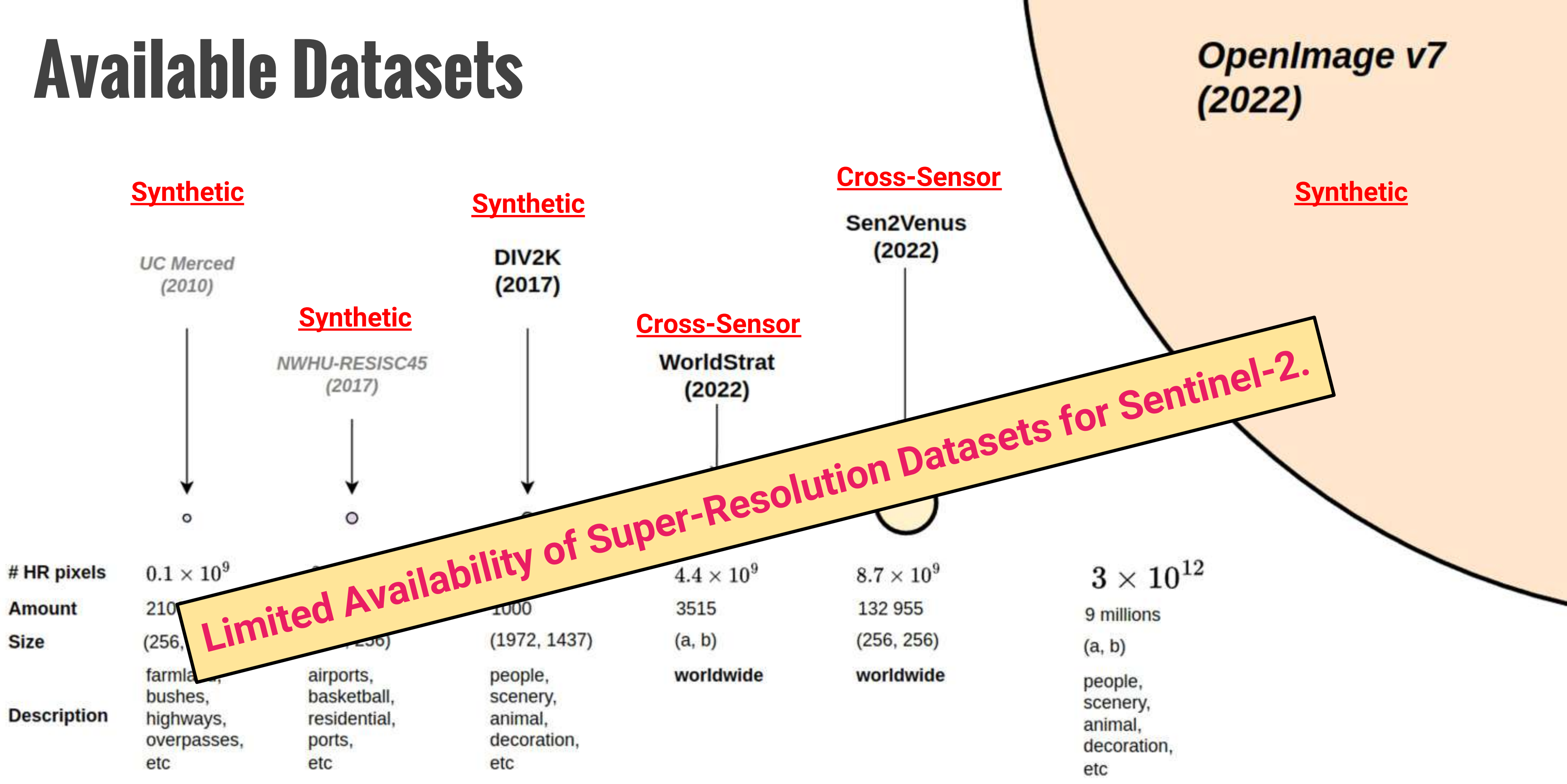
"Synthetic" dataset: LR image generated from HR image.
"Cross-sensor" dataset: LR image from another sensor.

Available Datasets

	Synthetic	Synthetic	Synthetic	Cross-Sensor	Cross-Sensor	Synthetic
	UC Merced (2010)	NWHU-RESISC45 (2017)	DIV2K (2017)	WorldStrat (2022)	Sen2Venus (2022)	OpenImage v7 (2022)
# HR pixels	0.1×10^9	2.1×10^9	2.8×10^9	4.4×10^9	8.7×10^9	3×10^{12}
Amount	2100	31500	1000	3515	132 955	9 millions
Size	(256, 256)	(256, 256)	(1972, 1437)	(a, b)	(256, 256)	(a, b)
Description	farmland, bushes, highways, overpasses, etc	airports, basketball, residential, ports, etc	people, scenery, animal, decoration, etc	**worldwide**	**worldwide**	people, scenery, animal, decoration, etc

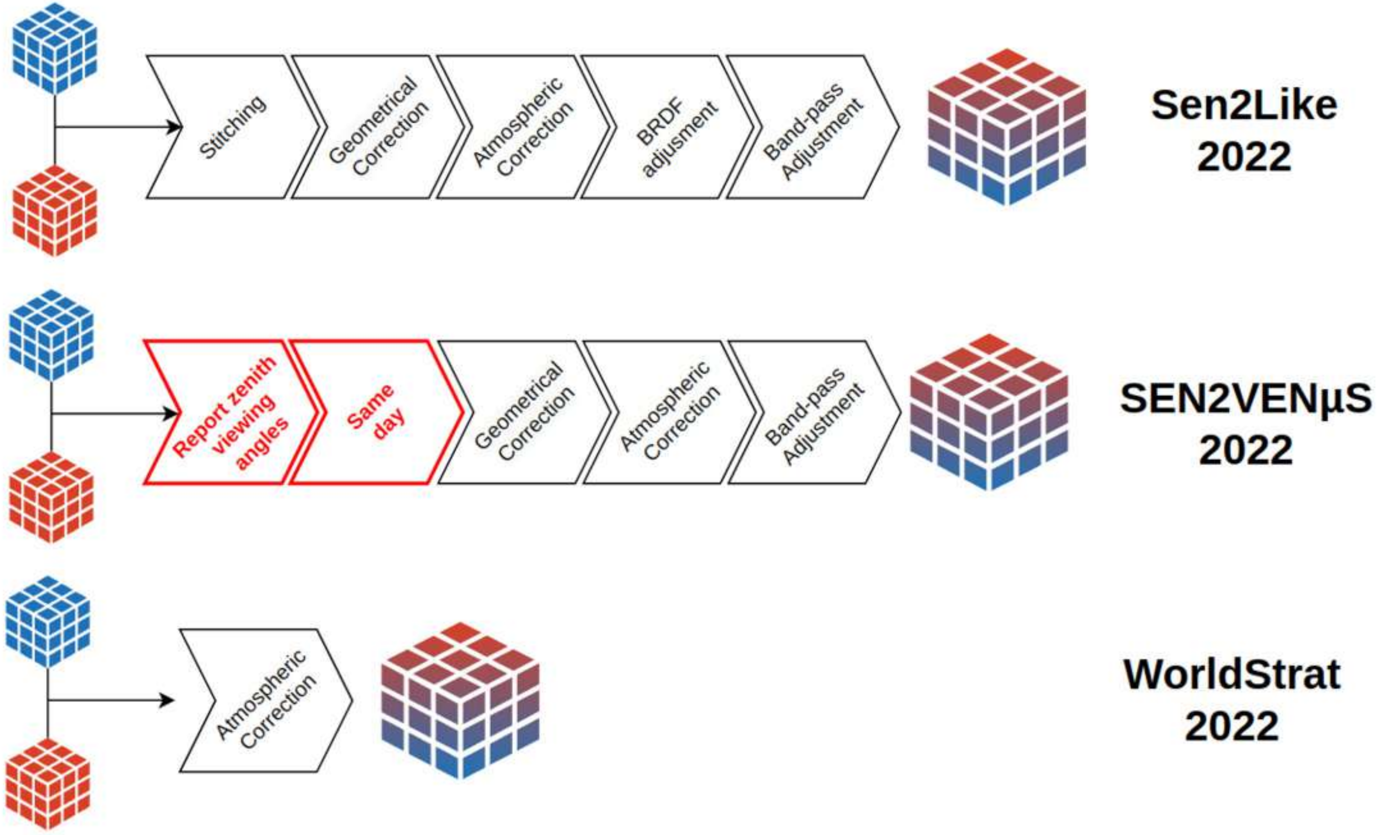
"Synthetic" dataset: LR image generated from HR image.
"Cross-sensor" dataset: LR image from another sensor.

Available Datasets

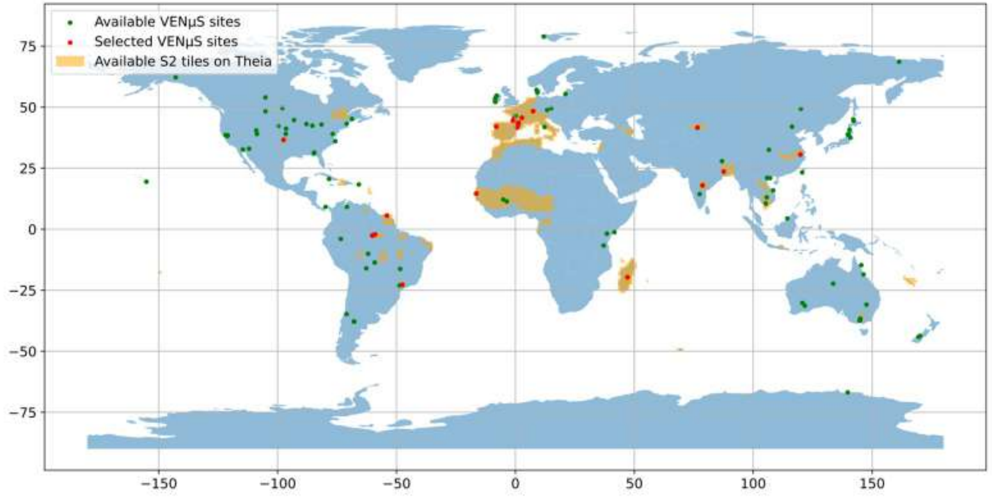


"Synthetic" dataset: LR image generated from HR image.
"Cross-sensor" dataset: LR image from another sensor.

Harmonization in cross-sensor datasets

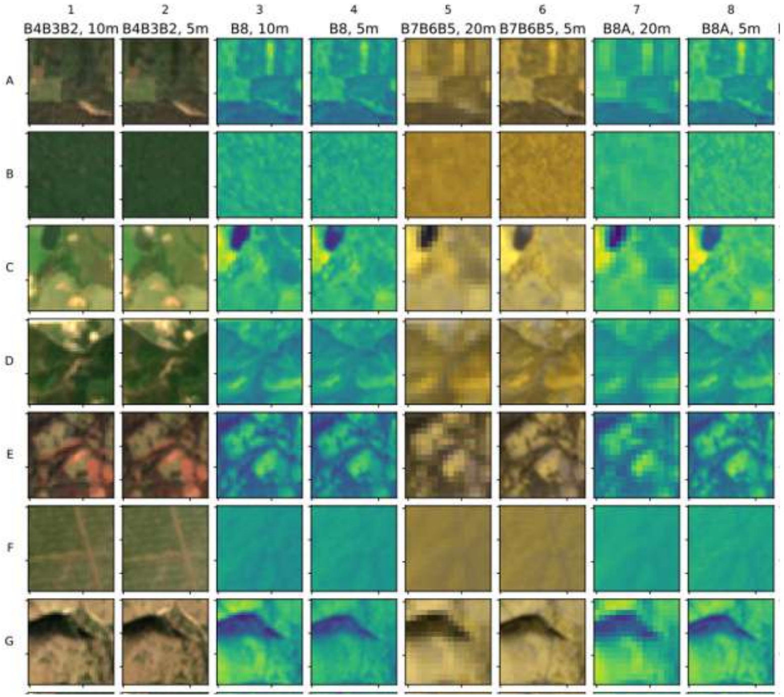


SEN2VENµS



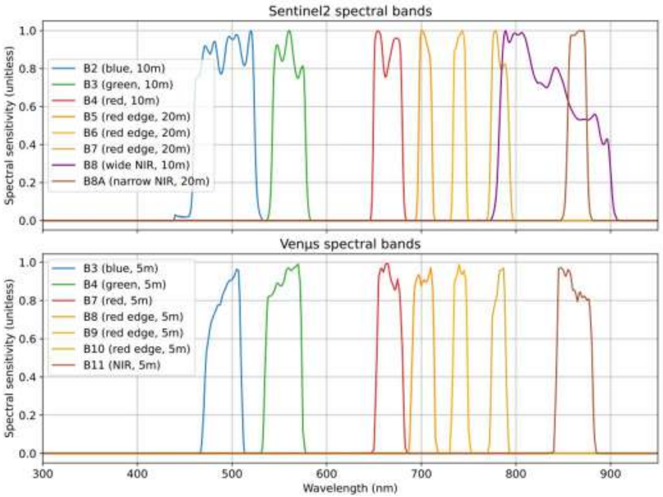
Sentinel-L2A (10/20m) -> VENµS (5m)

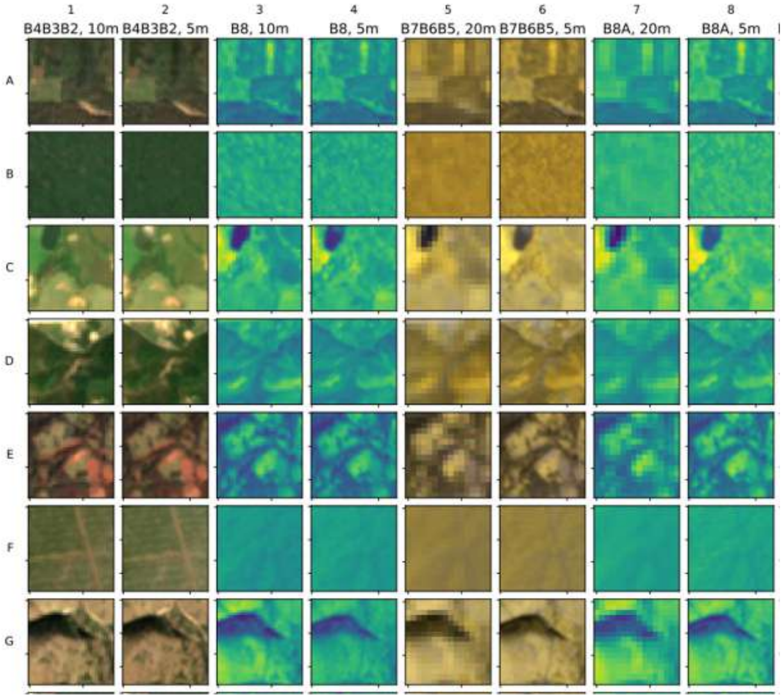
- Spatial collocation (SIFT)
- Spectral collocation (LLS)
- Friendly format



- Small and geographically biased.
- No SWIR bands

SEN2VENµS



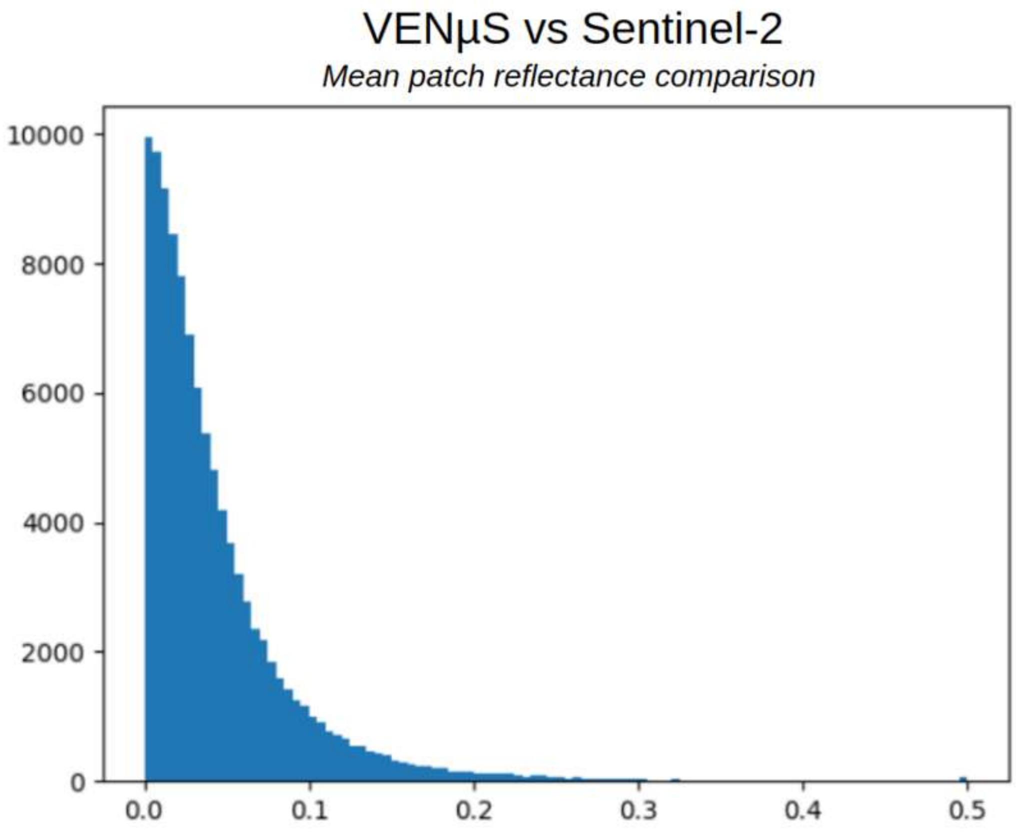


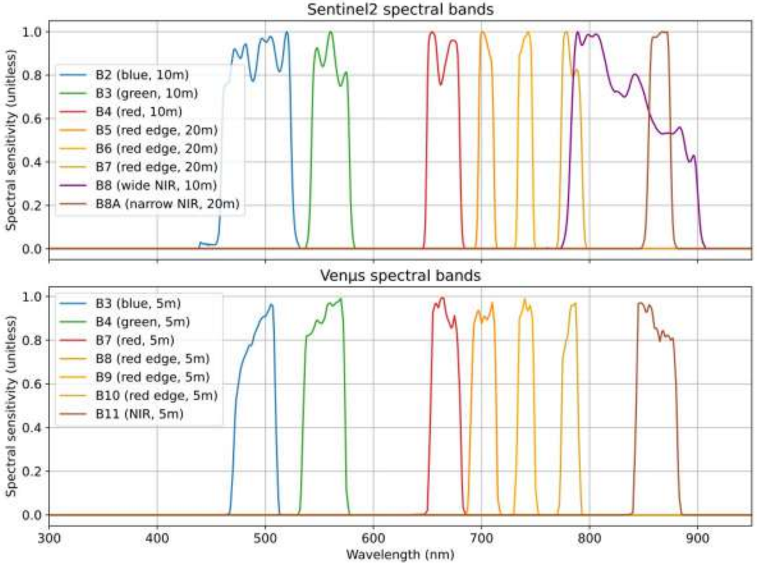
Sentinel-L2A (10/20m) -> VENµS (5m)

- Spatial collocation (SIFT)
- Spectral collocation (LLS)
- Friendly format.

- Small and geographically biased.
- No SWIR bands

SEN2VENµS





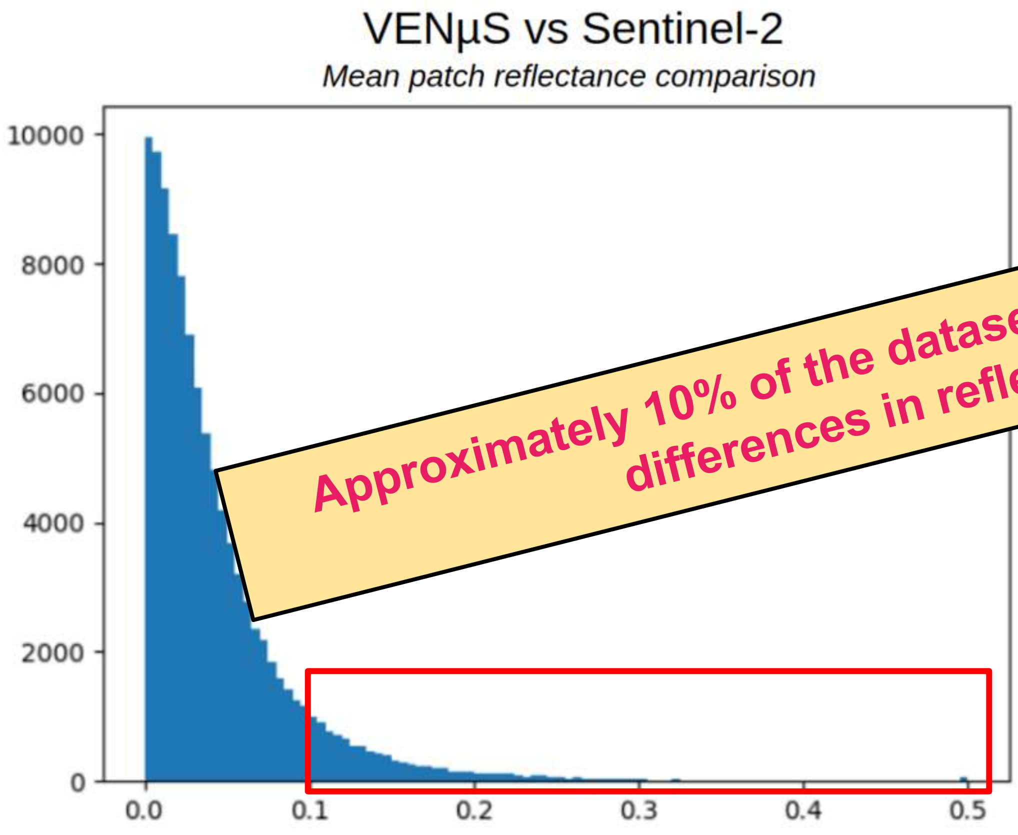
$$W^{\star} = \underset{W}{argmin} ||VW - S||_2^2$$

$$W \in R^{5\times 4}$$

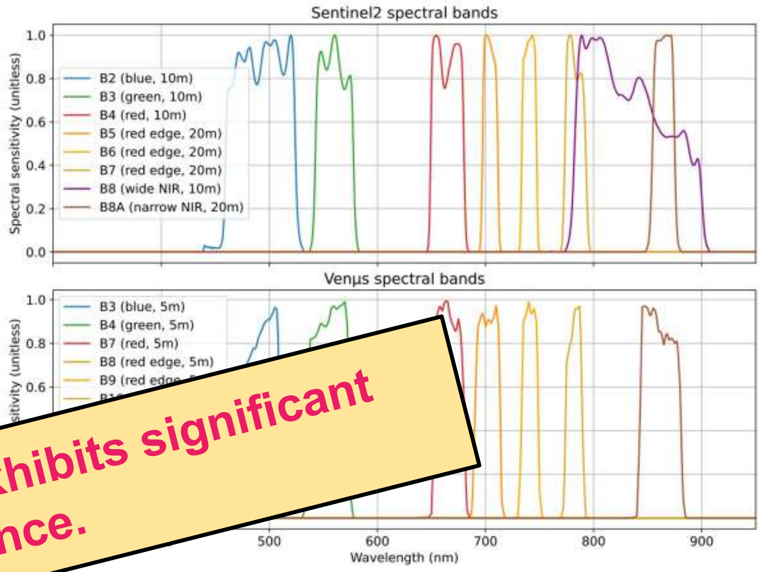
$$V = \begin{bmatrix} 1 & \rho_1^{ven\mu s,b2} & \rho_1^{ven\mu s,b4} & \rho_1^{ven\mu s,b7} & \rho_1^{ven\mu s,b11} \\ 1 & \rho_2^{ven\mu s,b2} & \rho_2^{ven\mu s,b4} & \rho_2^{ven\mu s,b7} & \rho_2^{ven\mu s,b11} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \rho_n^{ven\mu s,b2} & \rho_n^{ven\mu s,b4} & \rho_n^{ven\mu s,b7} & \rho_n^{ven\mu s,b11} \end{bmatrix}$$

$$S = \begin{bmatrix} \rho_1^{sentinel2,b2} & \rho_1^{sentinel2,b3} & \rho_1^{sentinel2,b4} & \rho_1^{sentinel2,b8} \\ \rho_2^{sentinel2,b2} & \rho_2^{sentinel2,b3} & \rho_2^{sentinel2,b4} & \rho_2^{sentinel2,b8} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_n^{sentinel2,b2} & \rho_n^{sentinel2,b3} & \rho_n^{sentinel2,b4} & \rho_n^{sentinel2,b8} \end{bmatrix}$$

SEN2VENµS



Approximately 10% of the dataset exhibits significant differences in reflectance.



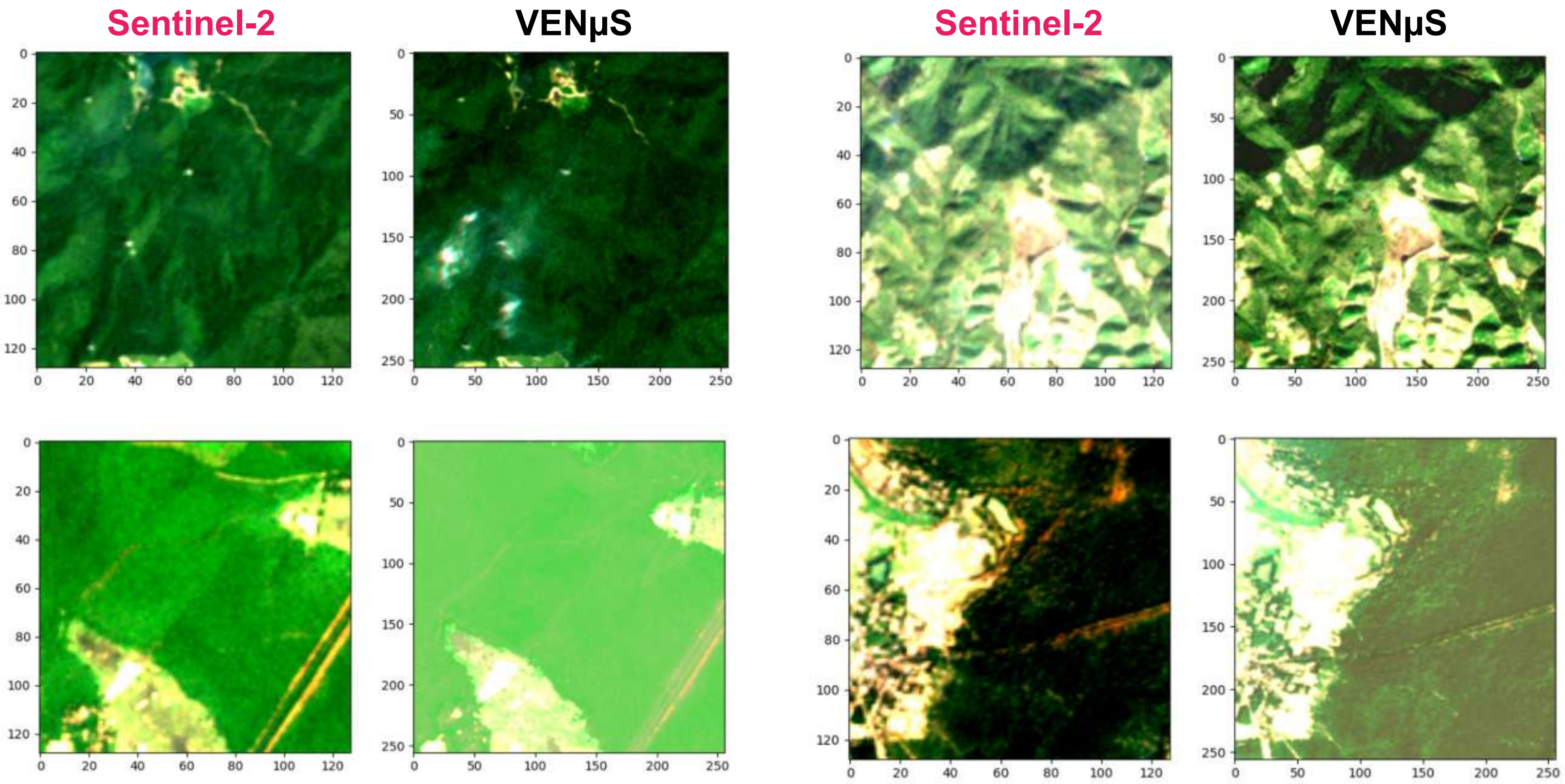
$$W^{\star} = \underset{W}{argmin} ||VW - S||_2^2$$

$$W \in R^{5\times 4}$$

$$V = \begin{bmatrix} 1 & \rho_1^{venus,b2} & \rho_1^{venus,b4} & \rho_1^{venus,b7} & \rho_1^{venus,b11} \\ 1 & \rho_2^{venus,b2} & \rho_2^{venus,b4} & \rho_2^{venus,b7} & \rho_2^{venus,b11} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \rho_n^{venus,b2} & \rho_n^{venus,b4} & \rho_n^{venus,b7} & \rho_n^{venus,b11} \end{bmatrix}$$

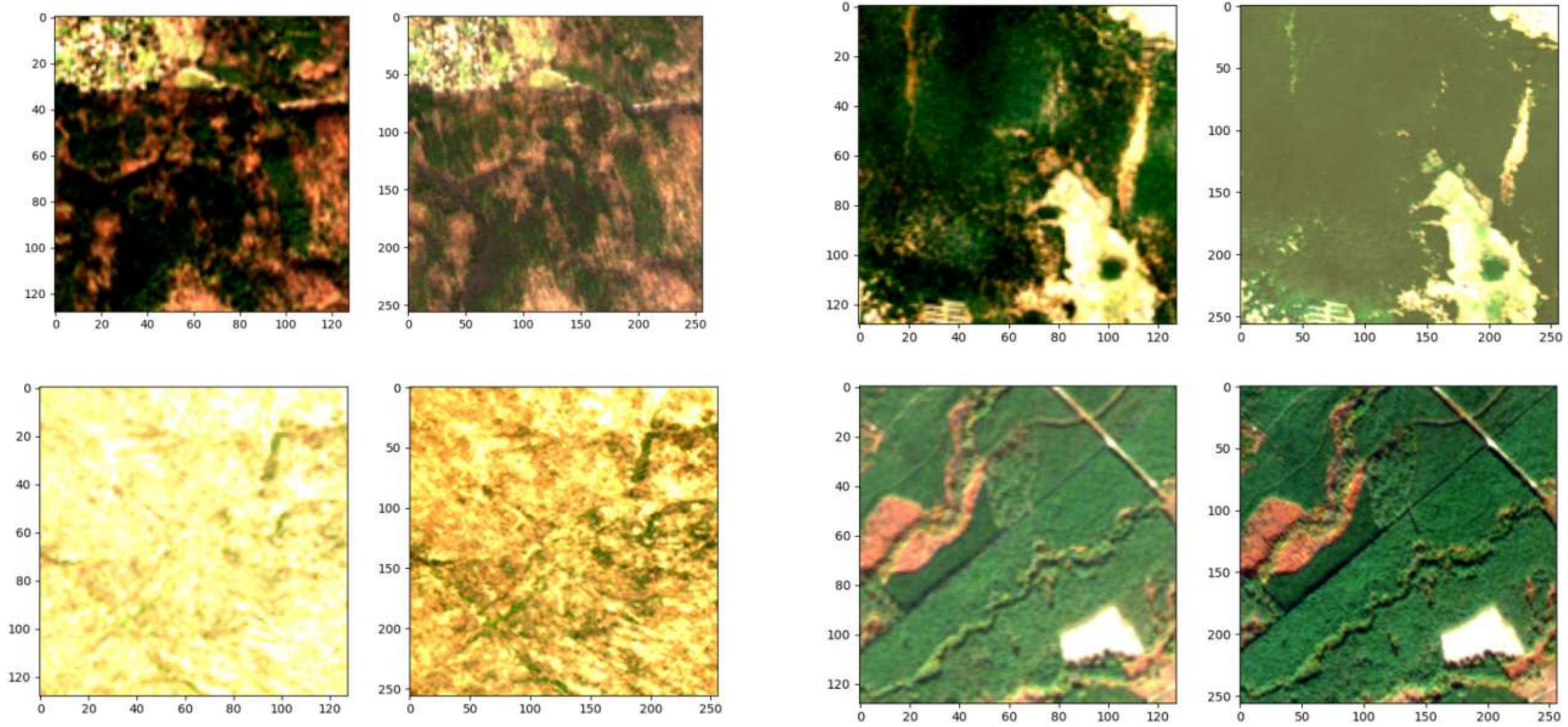
$$S = \begin{bmatrix} \rho_1^{sentinel2,b2} & \rho_1^{sentinel2,b3} & \rho_1^{sentinel2,b4} & \rho_1^{sentinel2,b8} \\ \rho_2^{sentinel2,b2} & \rho_2^{sentinel2,b3} & \rho_2^{sentinel2,b4} & \rho_2^{sentinel2,b8} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_n^{sentinel2,b2} & \rho_n^{sentinel2,b3} & \rho_n^{sentinel2,b4} & \rho_n^{sentinel2,b8} \end{bmatrix}$$

SEN2VENμS

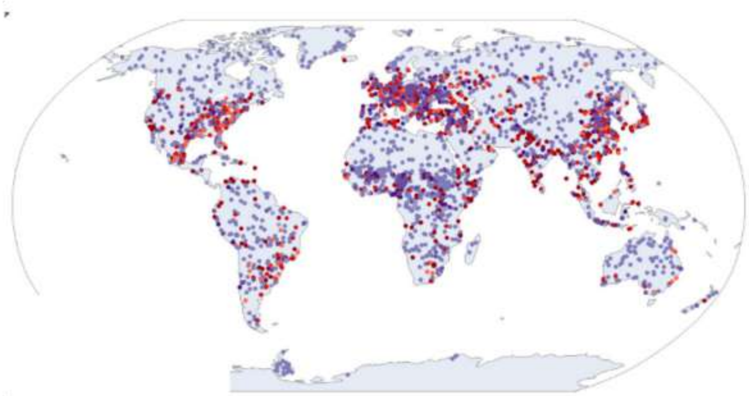


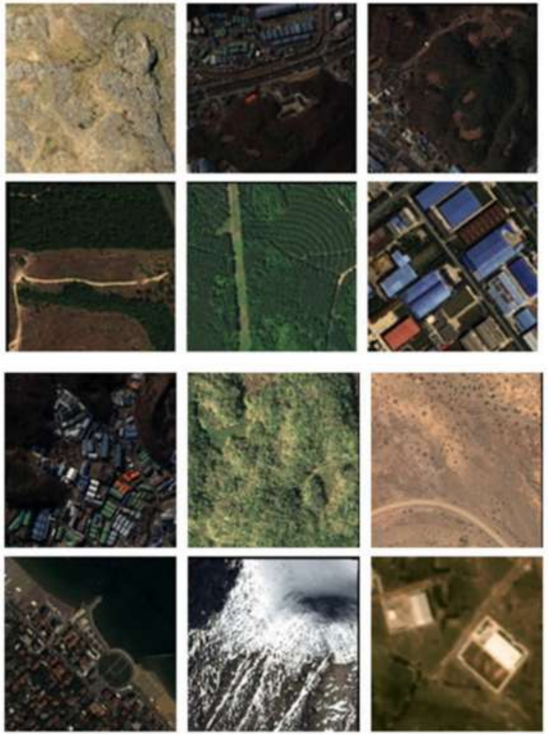
SEN2VENμS

Sentinel-2 | VENμS | Sentinel-2 | VENμS



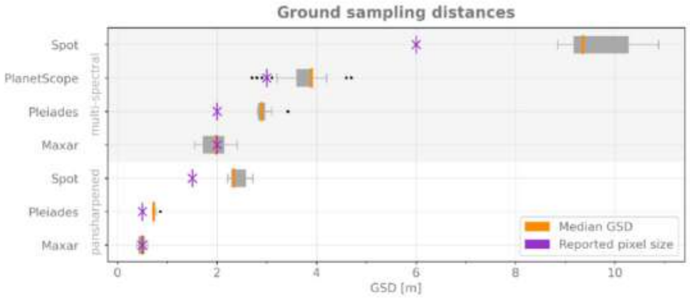
WorldStrat





Sentinel-L2A (10/20m) -> **SPOT6/7 (1.5m)***

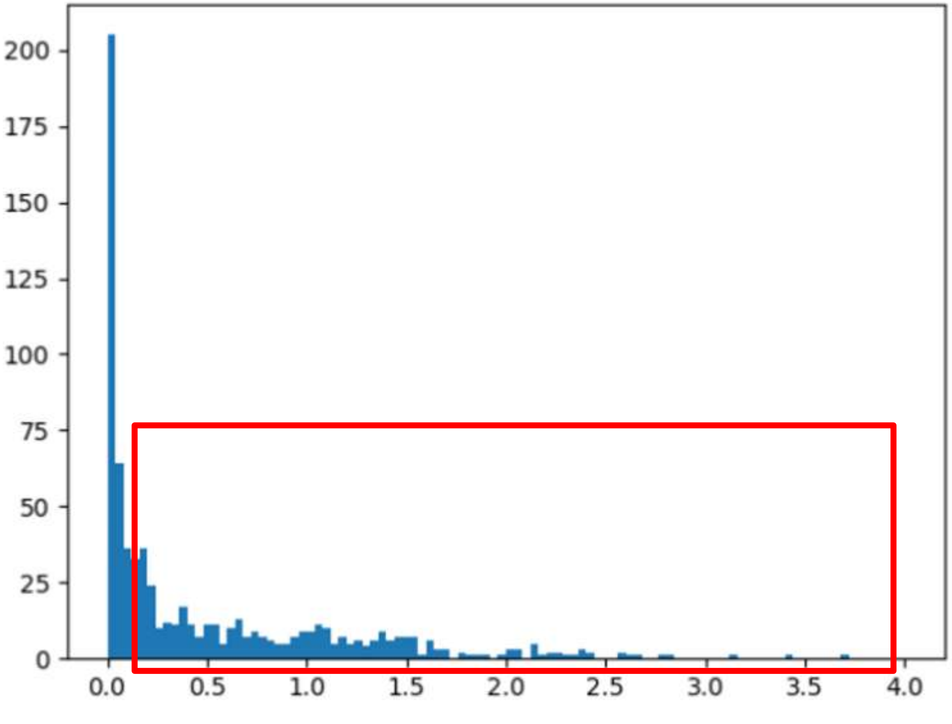
- Global coverage
- multi-temporal
- Panchromatic bands 1.5 m, but RGB NIR 6.0 m
- No spatial corregistration
- No full spectral overlap
- No red edge and SWIR bands

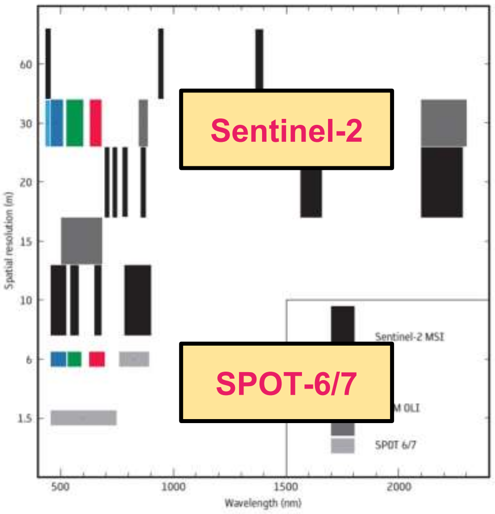


WorldStrat

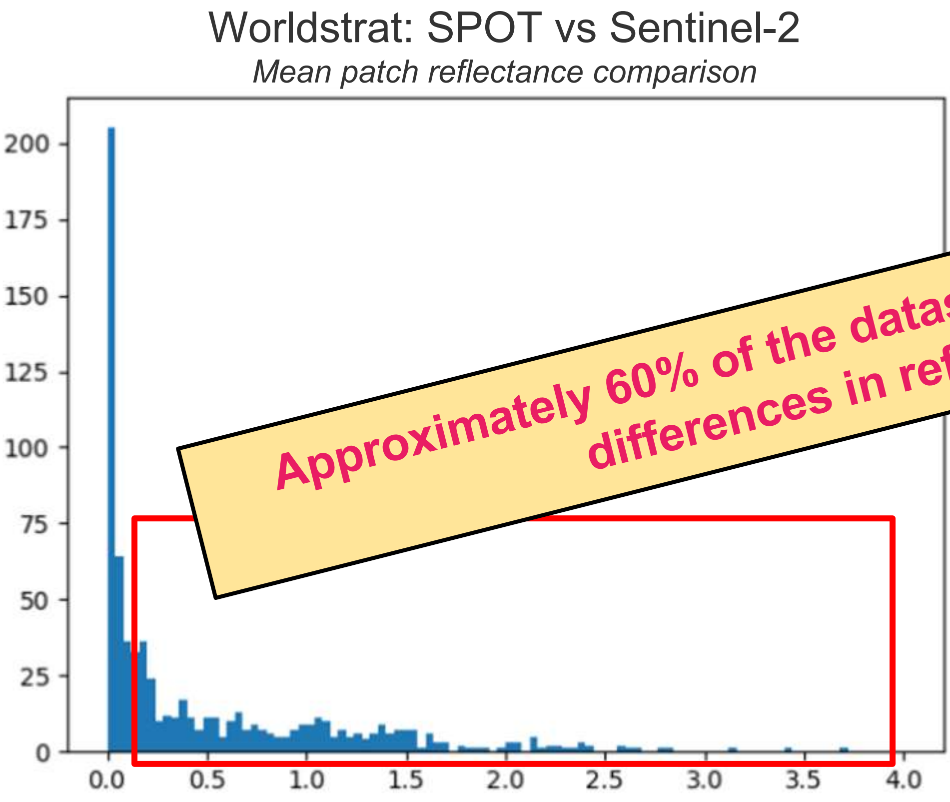
Worldstrat: SPOT vs Sentinel-2

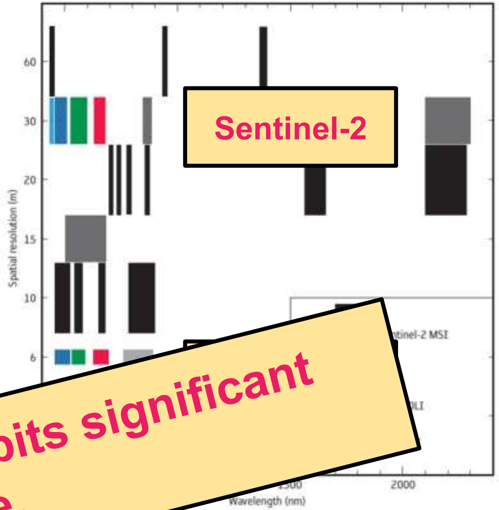
Mean patch reflectance comparison





WorldStrat

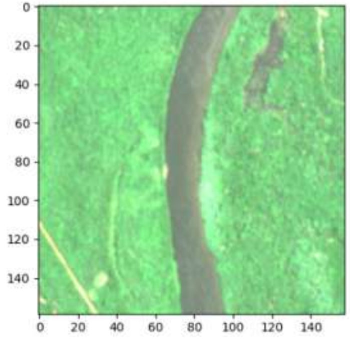
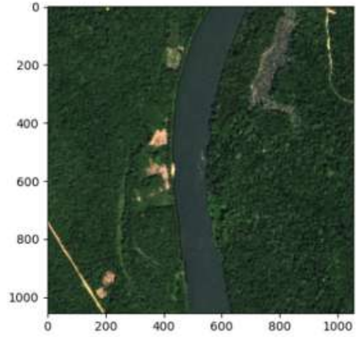
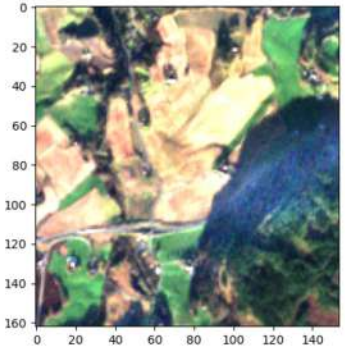
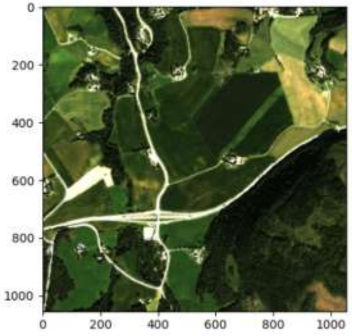
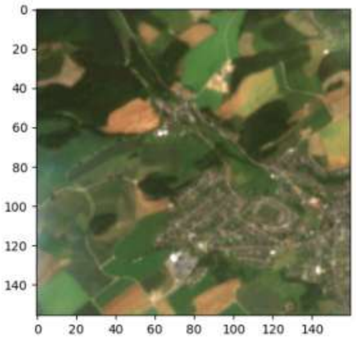
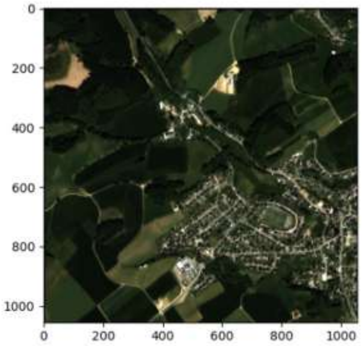
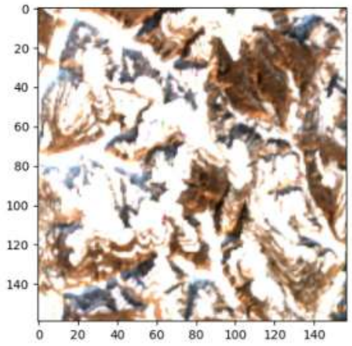
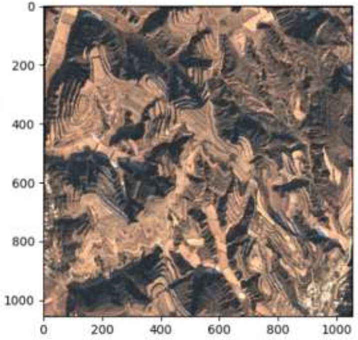




Approximately 60% of the dataset exhibits significant differences in reflectance.

WorldStrat

Sentinel-2 | **SPOT** | **Sentinel-2** | **SPOT**

Generation of SR Datasets

	“Synthetic”	“Cross-sensor”
Reference Image	• LR image: From HR image.	• LR image: From another sensor.
Challenges	• Learn the distribution shift, i.e., training dataset could be different to the real-world cases.	• Spectral bands mismatch. • Variations in atmospheric conditions during acquisition. • Differences in zenith viewing angle. • Spatial alignment is requirement.

Generation of SR Datasets

	"Synthetic"	"Cross-sensor"
Reference Image	• LR image: From HR image.	• LR image: From another sensor.
Challenges	• Learn th... ...ng dataset ...uld be differen... real...	• Spectral bands... • Vari... ...ne ...during acqui... • Differences... ...ngle. • S... ...requirement.

They might not work well in real world

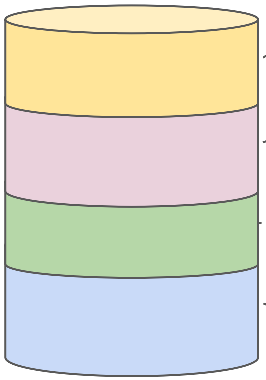
Easy to Scale

Harmonization

Hard to Scale

OpenSR Dataset

S2NAIPv2

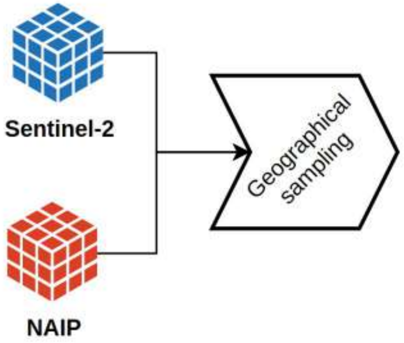


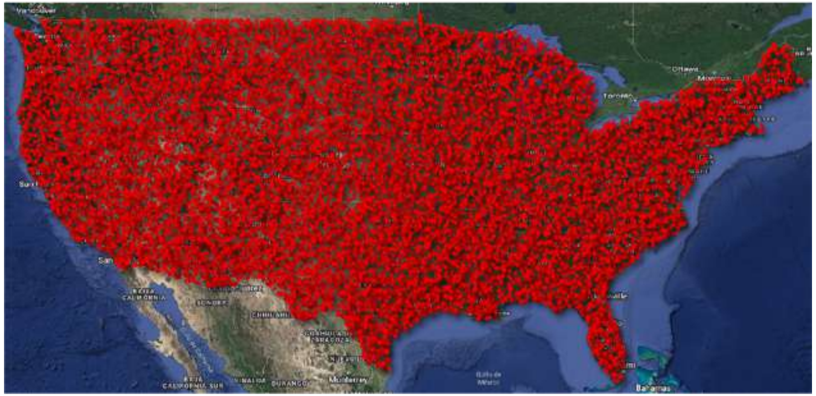
~200k samples

- *Synthetic I LR subset*
  A **60k** dataset with a corresponding LR pair created synthetically from HR **using a ML harmonization methodology.**
- *Synthetic II LR subset*
  A **60k** dataset with a corresponding LR pair created synthetically from HR **using a closer S2 imagery (HM with 30 days time frame)**.
- *Real S2 LR subset*
  A **8k** dataset with a corresponding original S2 image with the HR created synthetically by **injecting high-frequency from NAIP to S2.**
- *Real S2 LR subset **(MISR)***
  A **40k** dataset with time-series consisting of **16 images S2 image**, with the nearest HR image captured always within 0–10 days.

OpenSR - Cross-sensor Dataset

OpenSR Cross-sensor Dataset



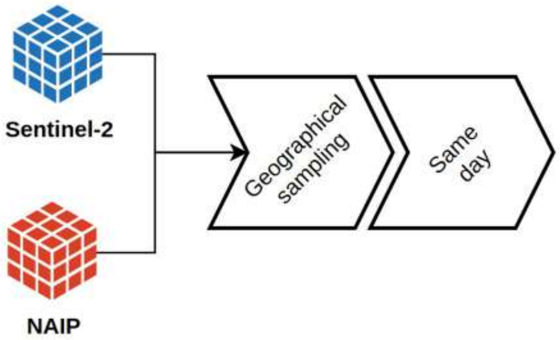


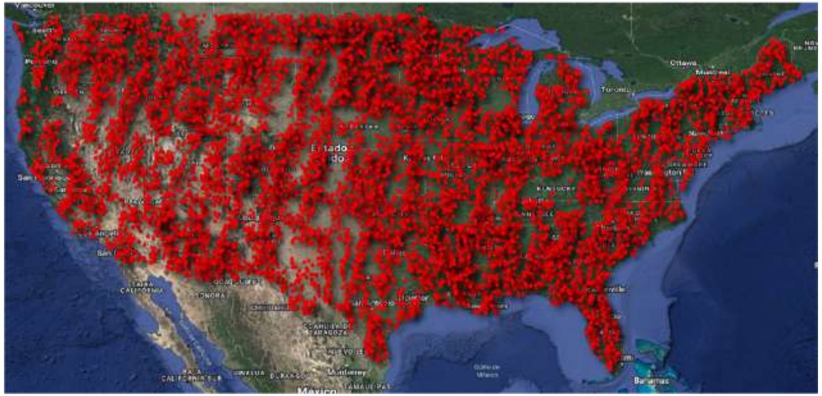
18k ROIs

NAIP: National Agriculture Imagery Program

NAIP
- >20 years (since 2002)
- RGBNIR orthophotos
- <2.5 meters
- Continental USA

OpenSR Cross-sensor Dataset



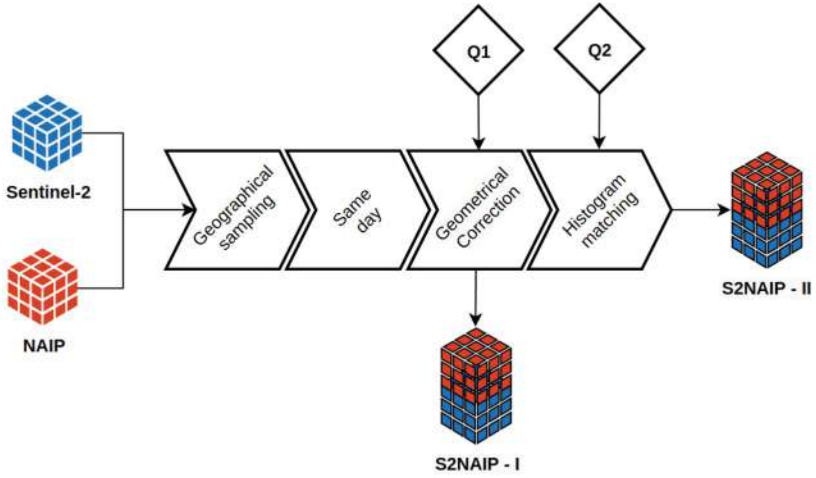


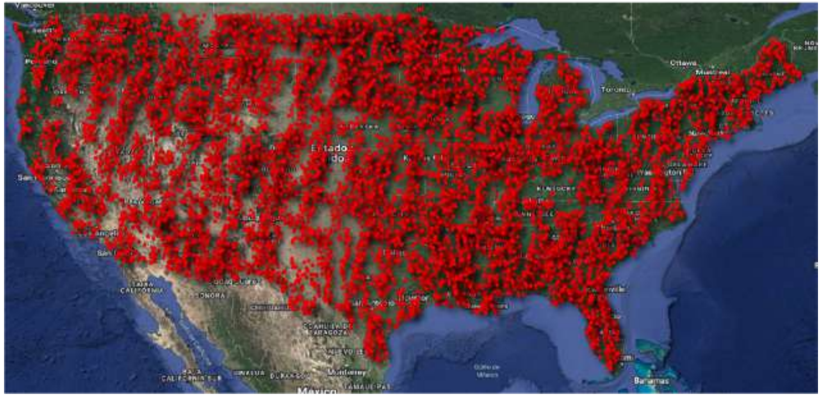
7k ROIs

NAIP: National Agriculture Imagery Program

NAIP
- >20 years (since 2002)
- RGBNIR orthophotos
- <2.5 meters
- Continental USA

OpenSR Cross-sensor Dataset





7k ROIs

SuperGlue + SuperPoint

PE Sarlin et al, 2020

Q1

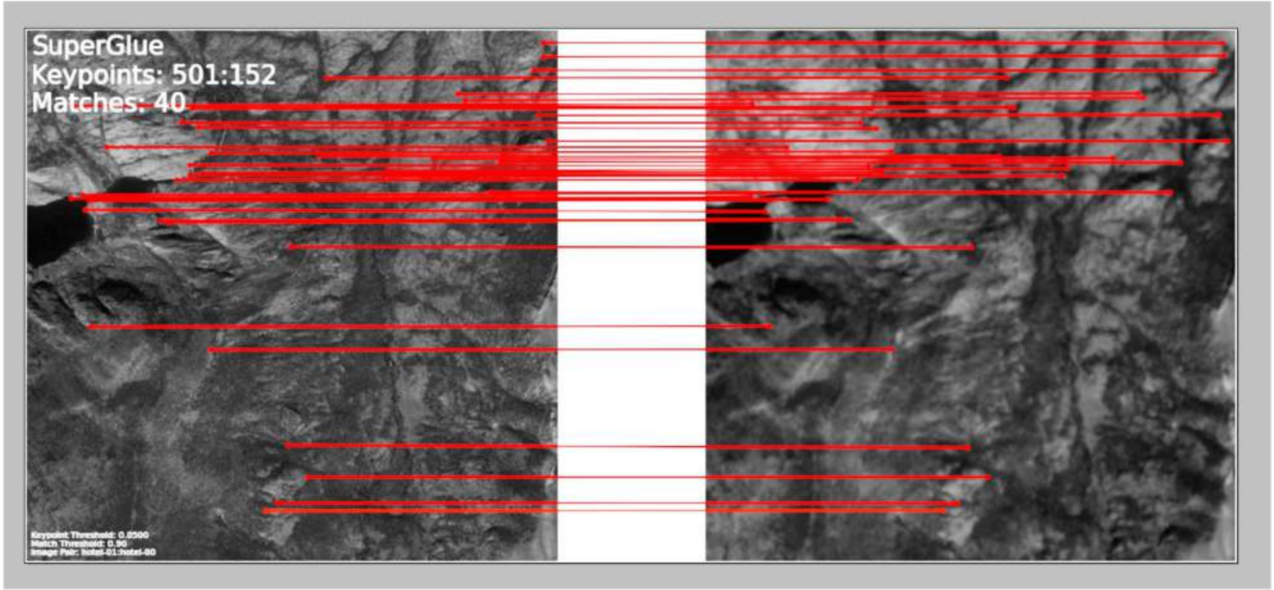
Histogram Matching

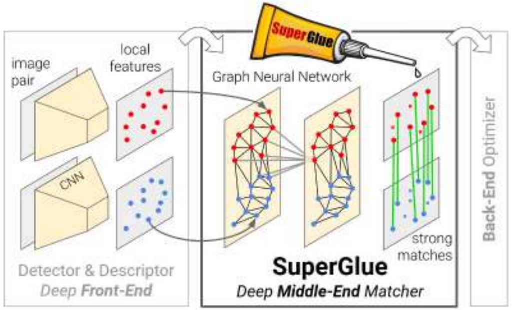
scikit-image

Q2

OpenSR Cross-sensor Dataset

5k ROIs





PE Sarlin et al, 2020

- No fine-tuned*
- Discard matches that are further that 10 m apart.

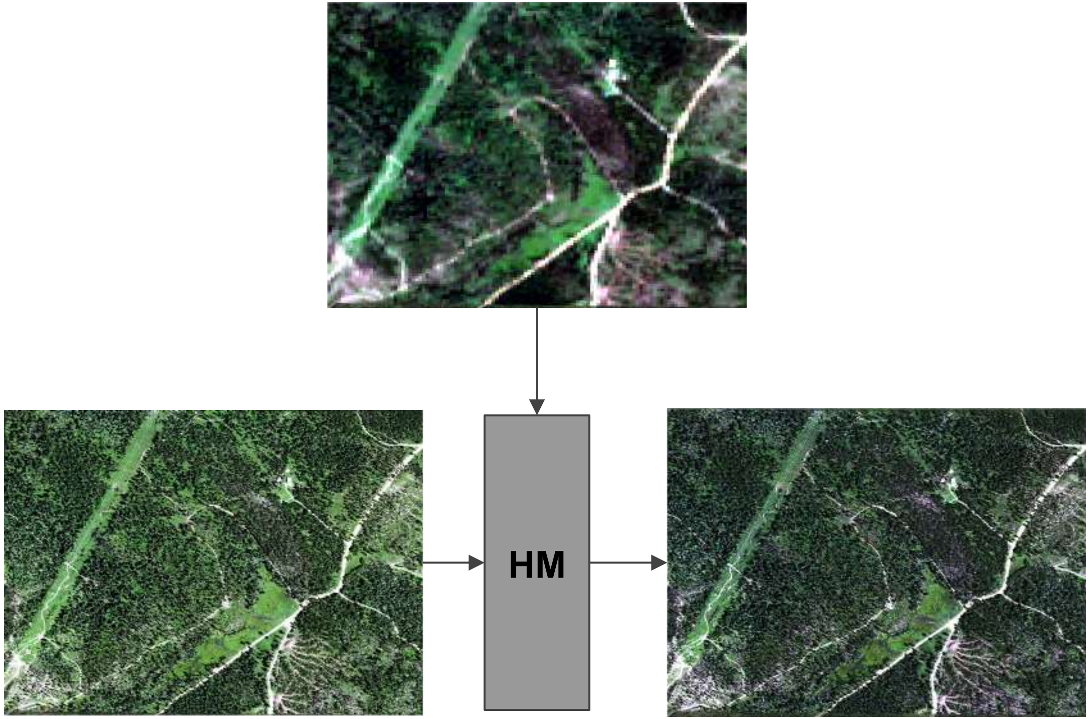
Quality test Q1

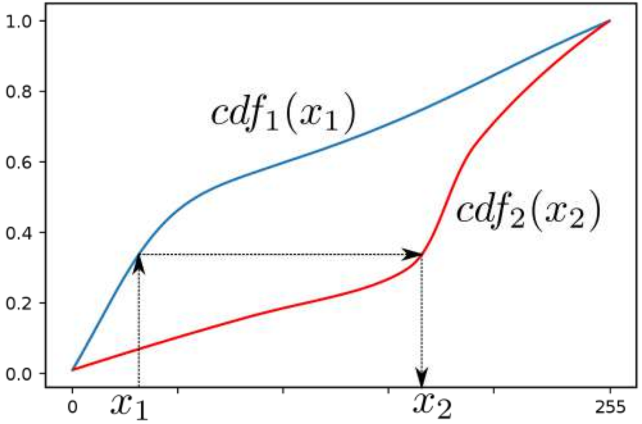
SuperGlue + SuperPoint

If more than 1 LR (10 m) pixel of difference

OpenSR Cross-sensor Dataset

3k ROIs

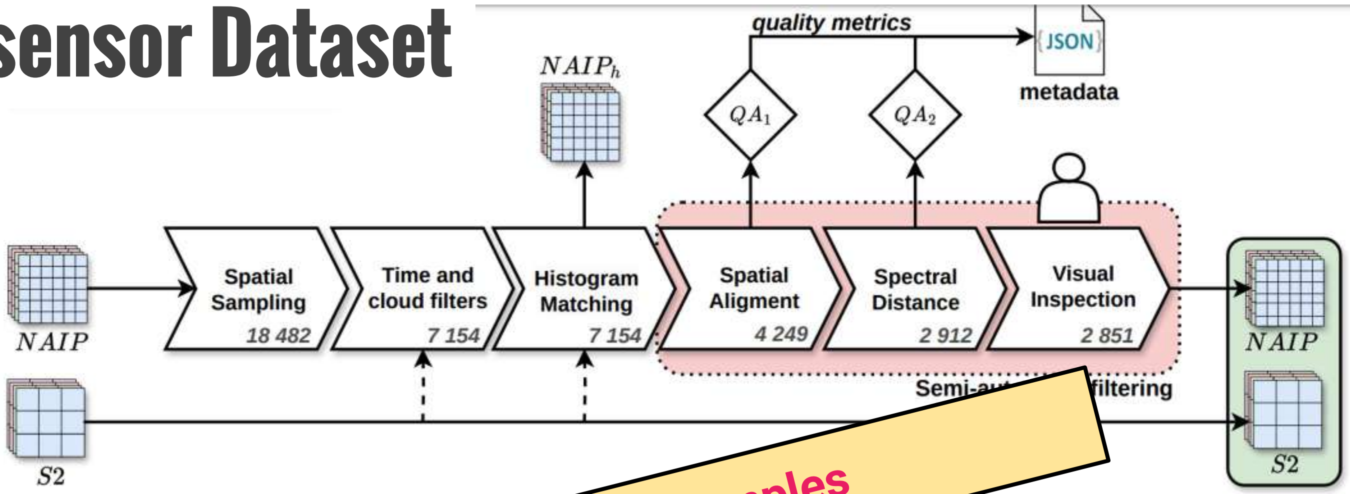




Quality test Q2

If spectral angle distance > 2 deg

OpenSR Cross-sensor Dataset



High quality dataset but very few samples

We find

2851

NAIP - S2

image pairs

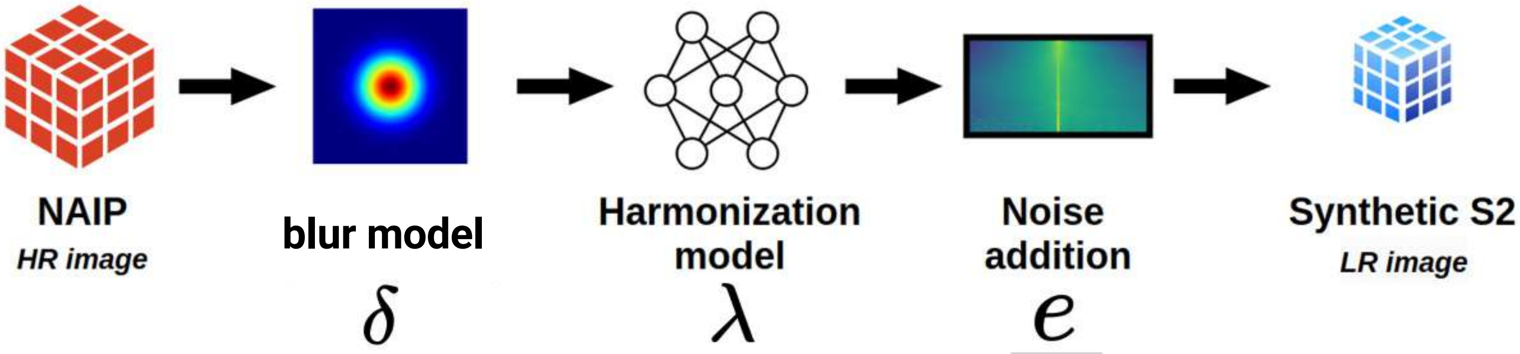
S2 NAIP S2 NAIP S2 NAIP

OpenSR - Synthetic LR Dataset

OpenSR Synthetic Dataset Degradation Model

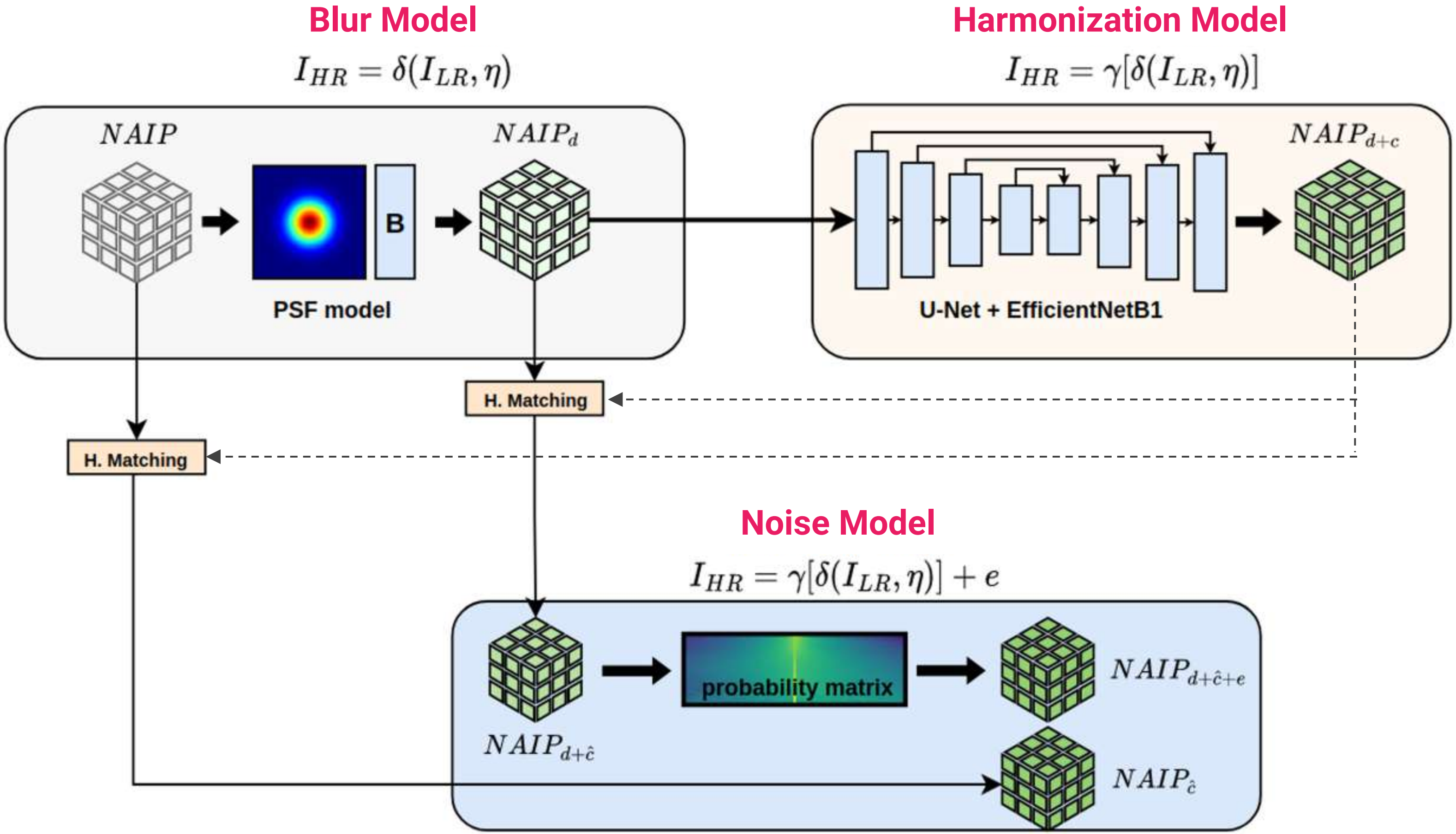
SEN2NAIP: Sentinel-2 Super-Resolution Dataset Using a Realistic Degradation Model

- **Learn the degradation** between VHR NAIP (2.5 m) and Sentinel-2 (10m)

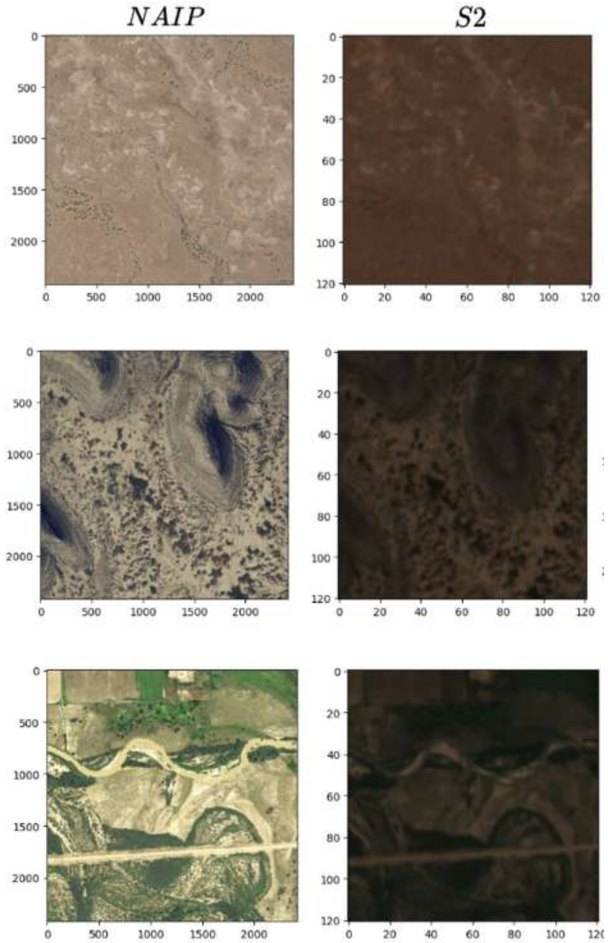


$$I_{LR} = \lambda[\delta(I_{HR}, n)] + e$$

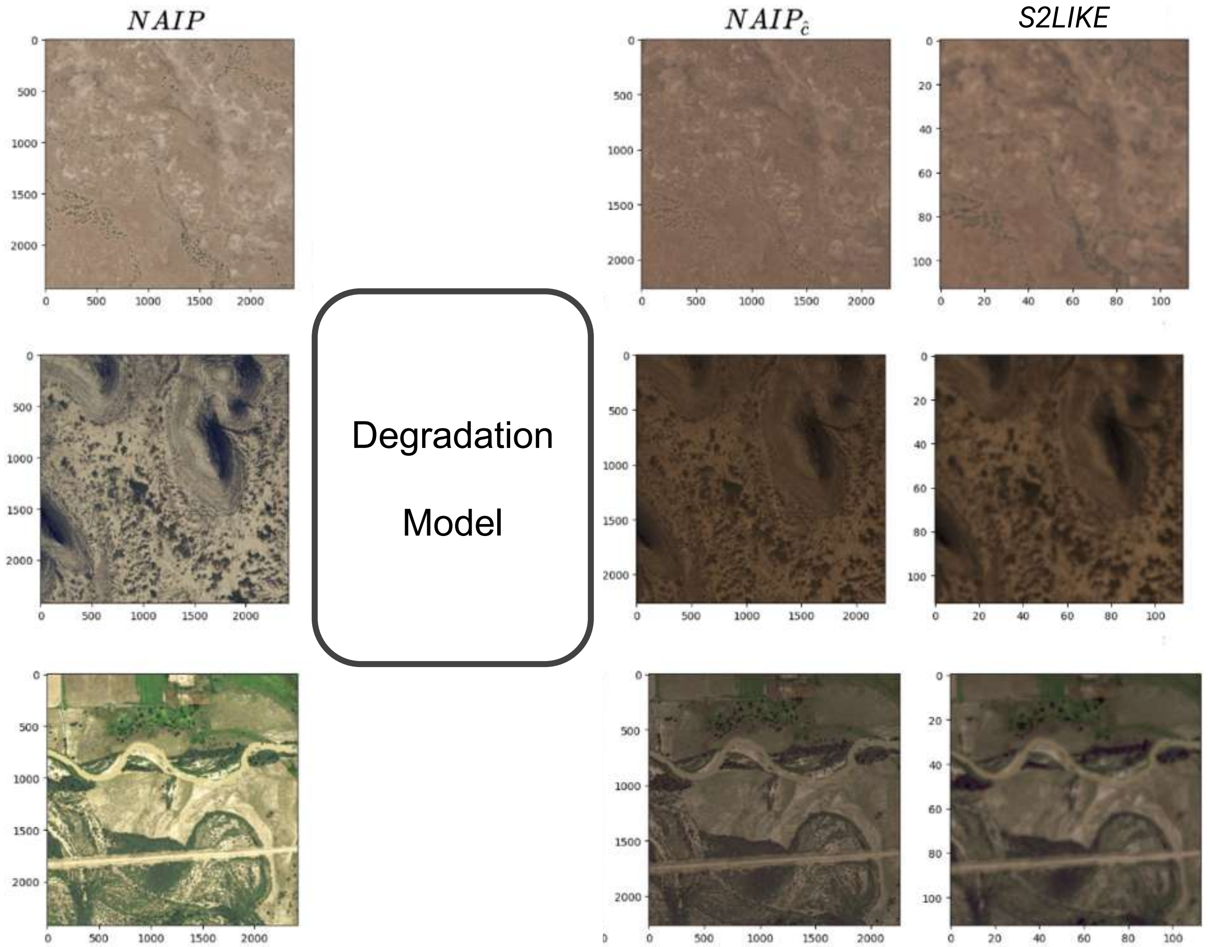
OpenSR Synthetic Dataset Degradation Model



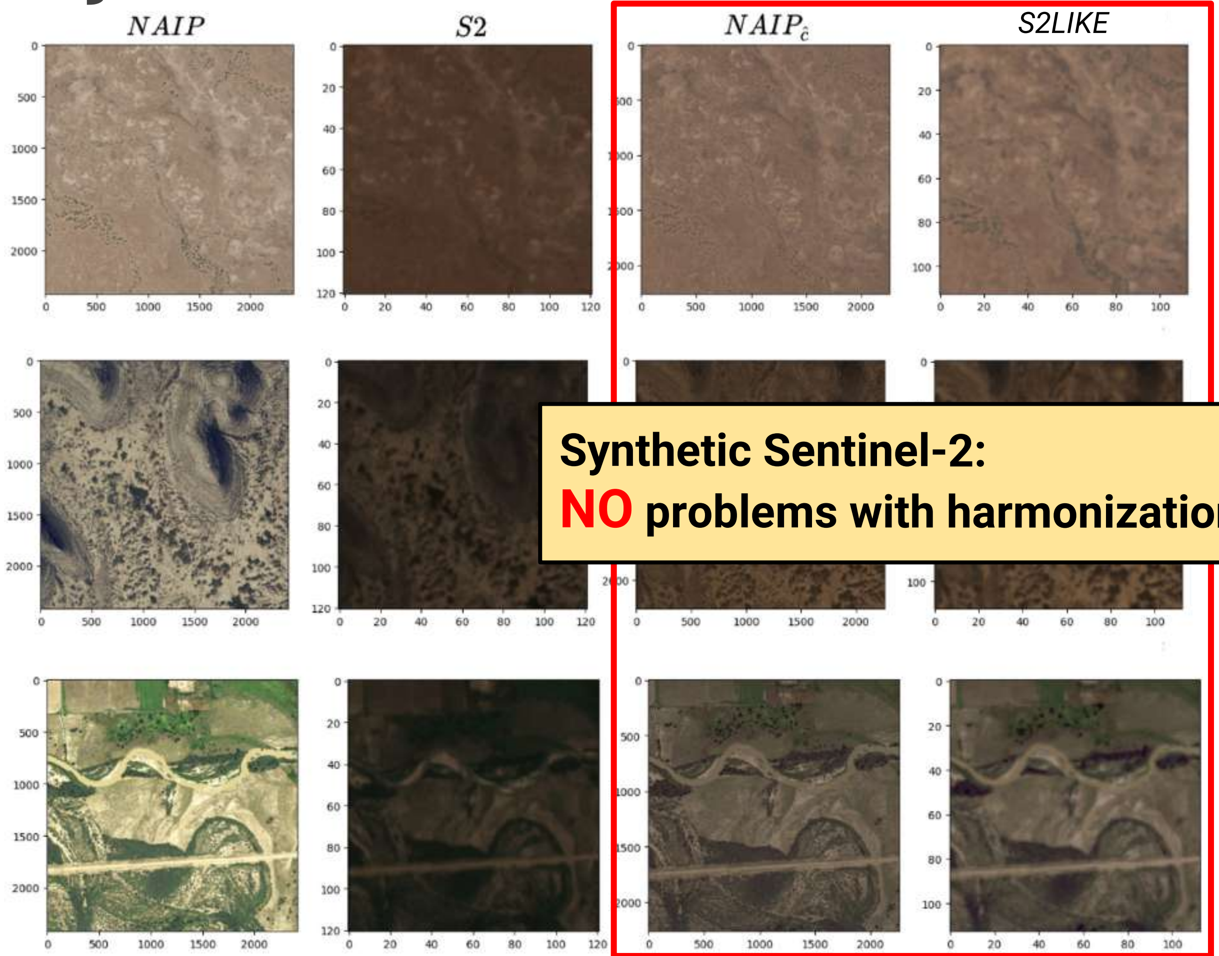
OpenSR Synthetic Dataset



OpenSR Synthetic Dataset

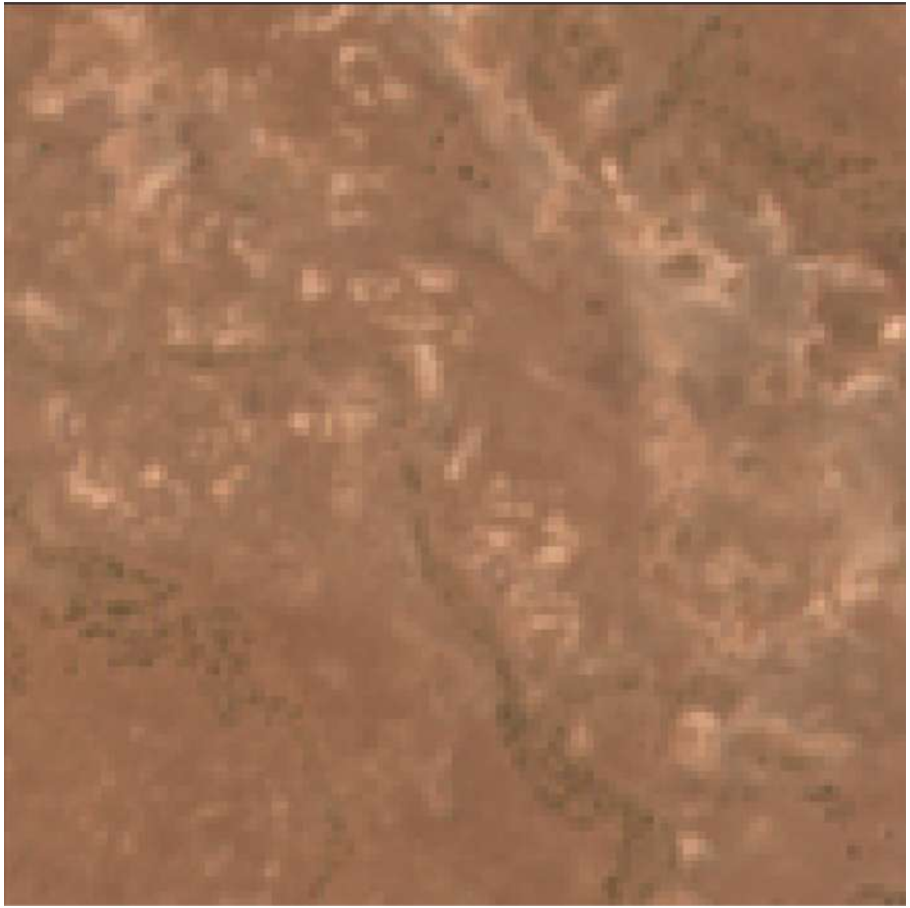


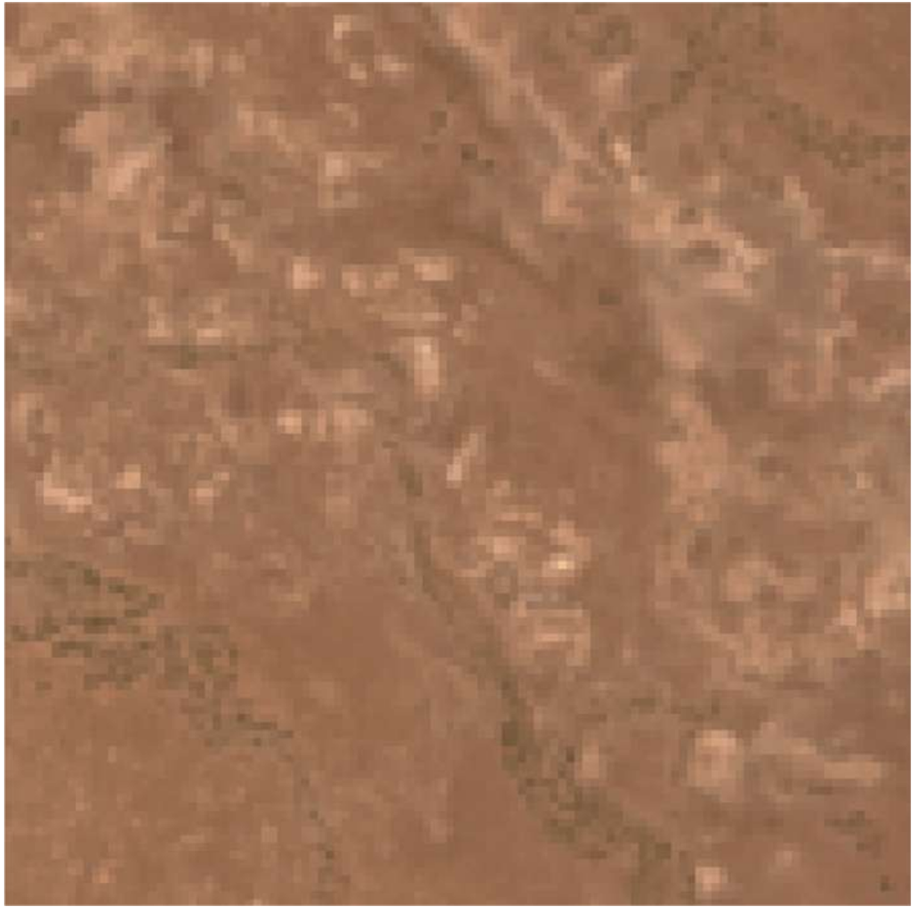
OpenSR Synthetic Dataset



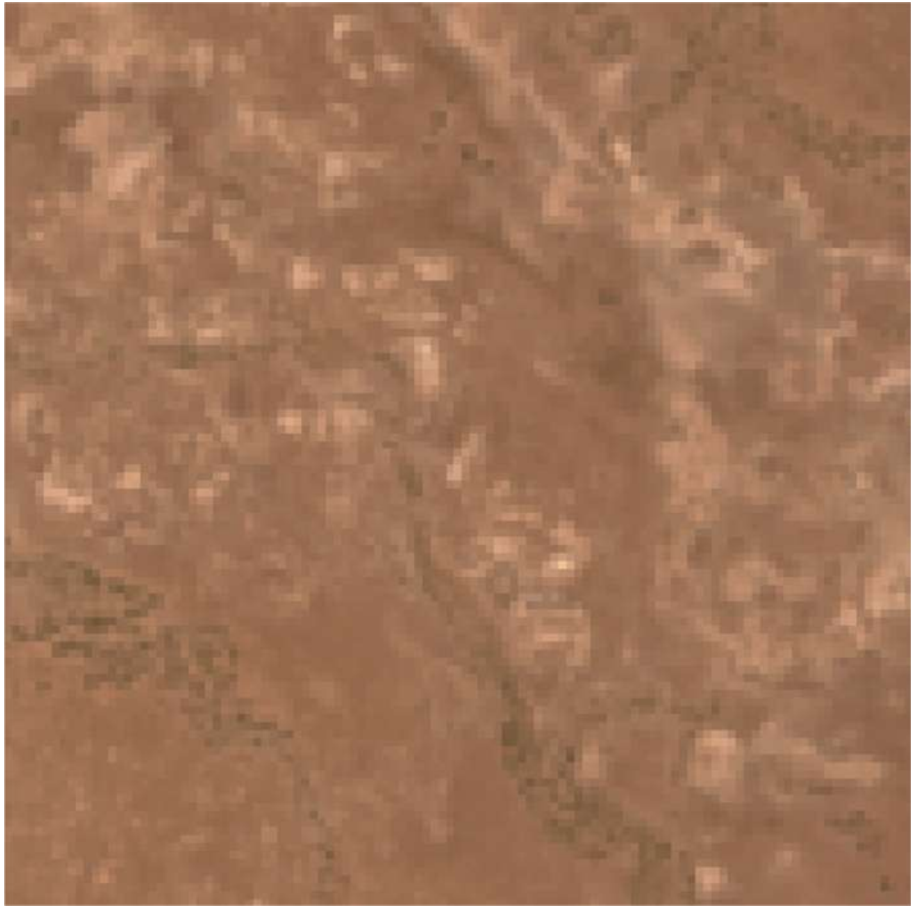
Synthetic Sentinel-2:
NO problems with harmonization

Which of both images is Sentinel-2?

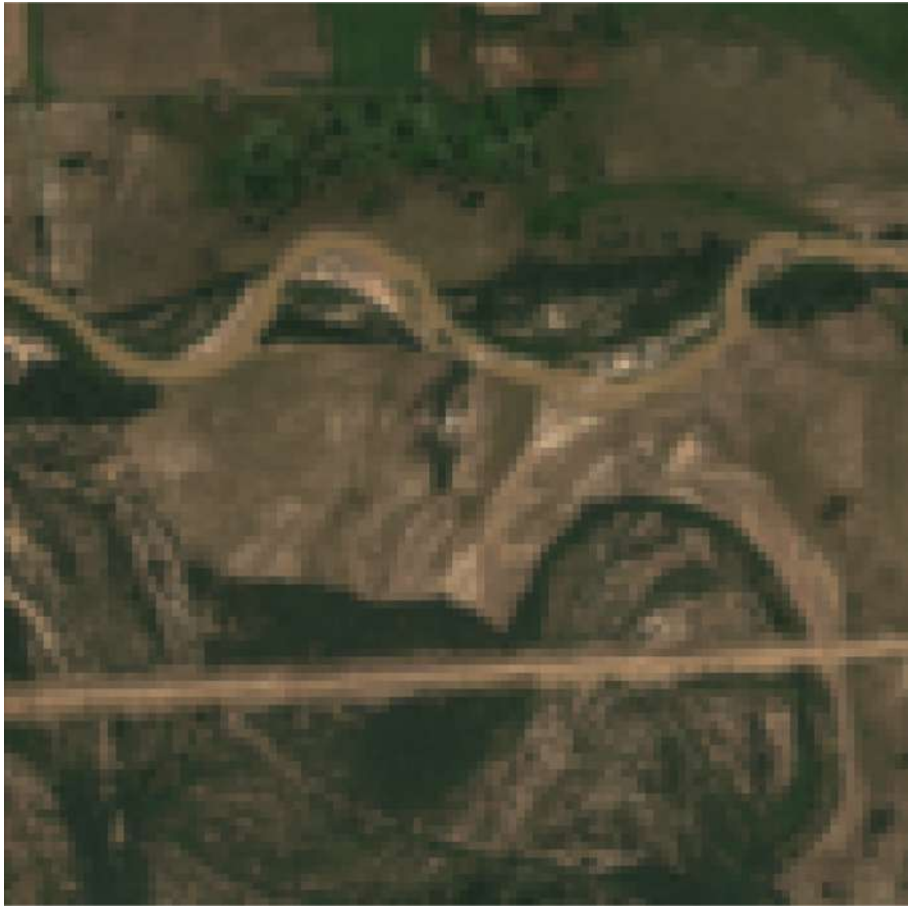




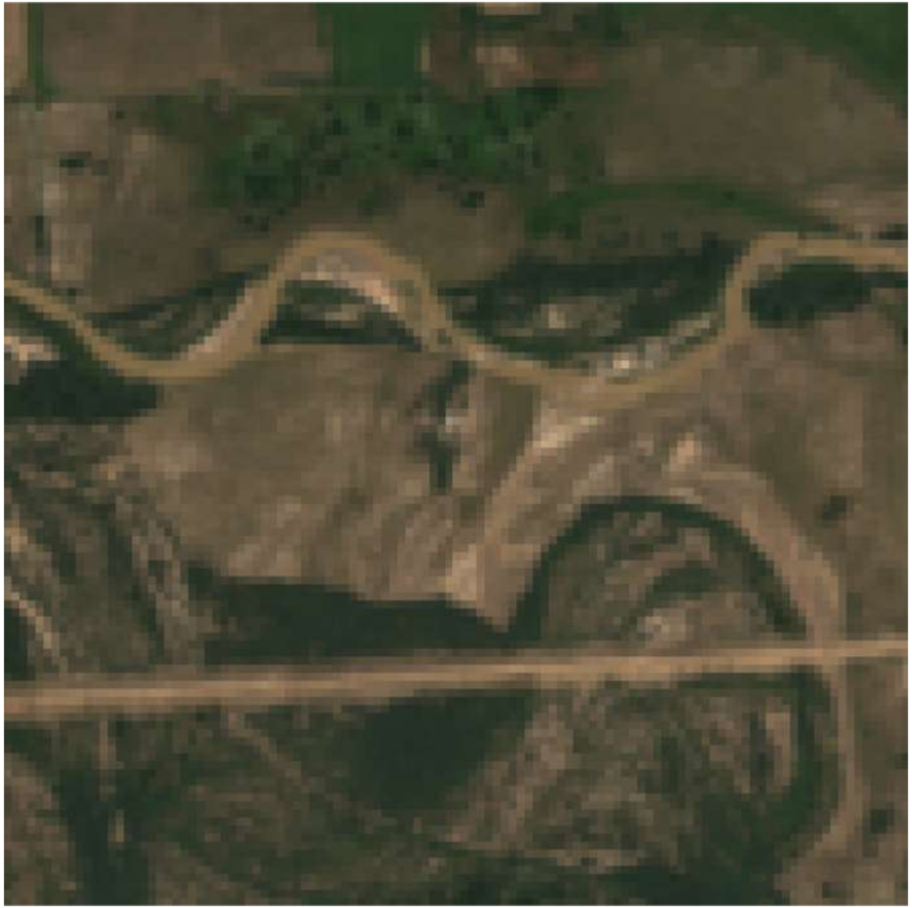
Which of both images is Sentinel-2?

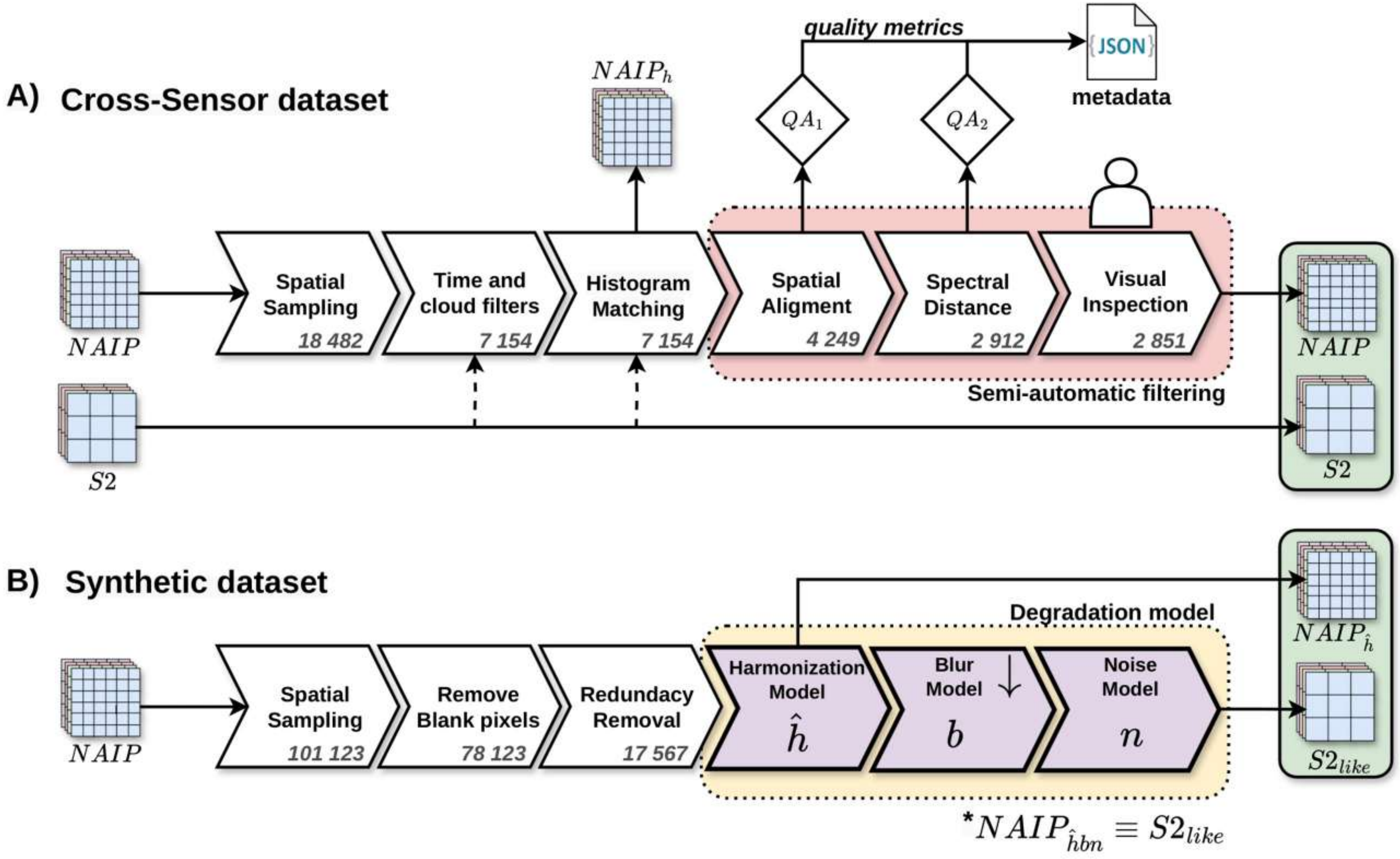


Which of both images is Sentinel-2?



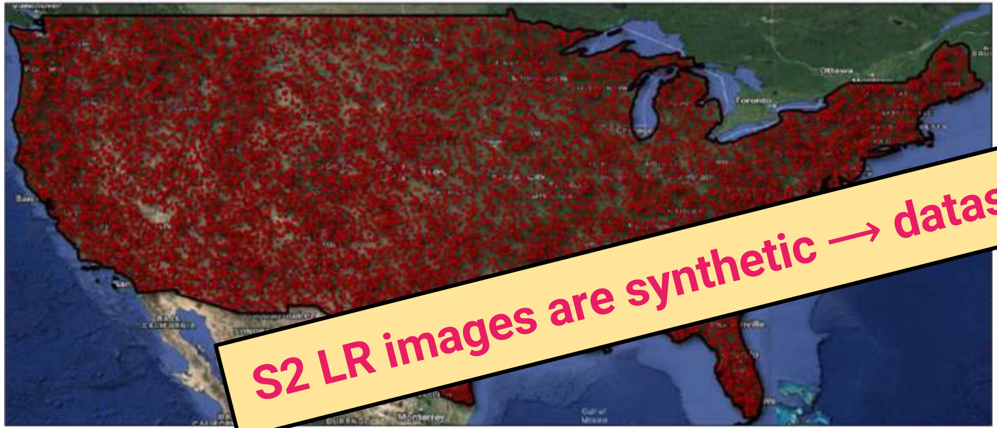
Which of both images is Sentinel-2?




A) Cross-Sensor dataset
quality metrics
JSON
metadata
$NAIP_h$
QA_1
QA_2
$NAIP$
Spatial Sampling 18 482
Time and cloud filters 7 154
Histogram Matching 7 154
Spatial Aligment 4 249
Spectral Distance 2 912
Visual Inspection 2 851
Semi-automatic filtering
$NAIP$
$S2$
$S2$
B) Synthetic dataset
$NAIP$
Spatial Sampling 101 123
Remove Blank pixels 78 123
Redundancy Removal 17 567
Degradation model
Harmonization Model $\hat{h}$
Blur Model $\downarrow$ b
Noise Model n
$NAIP_{\hat{h}}$
$S2_{like}$
$^*NAIP_{\hat{h}bn} \equiv S2_{like}$

(A) **Cross-Sensor Dataset (2 851 ROIs)**

(B) **Synthetic Dataset (17 657 ROIs)**



The locations of cross-sensor (A) and synthetic (B) regions of interest (ROIs) within the SEN2NAIP dataset.

S2 LR images are synthetic ⟶ dataset bias & distribution shift?

OpenSR - Real S2 LR Dataset

OpenSR Dataset Real S2

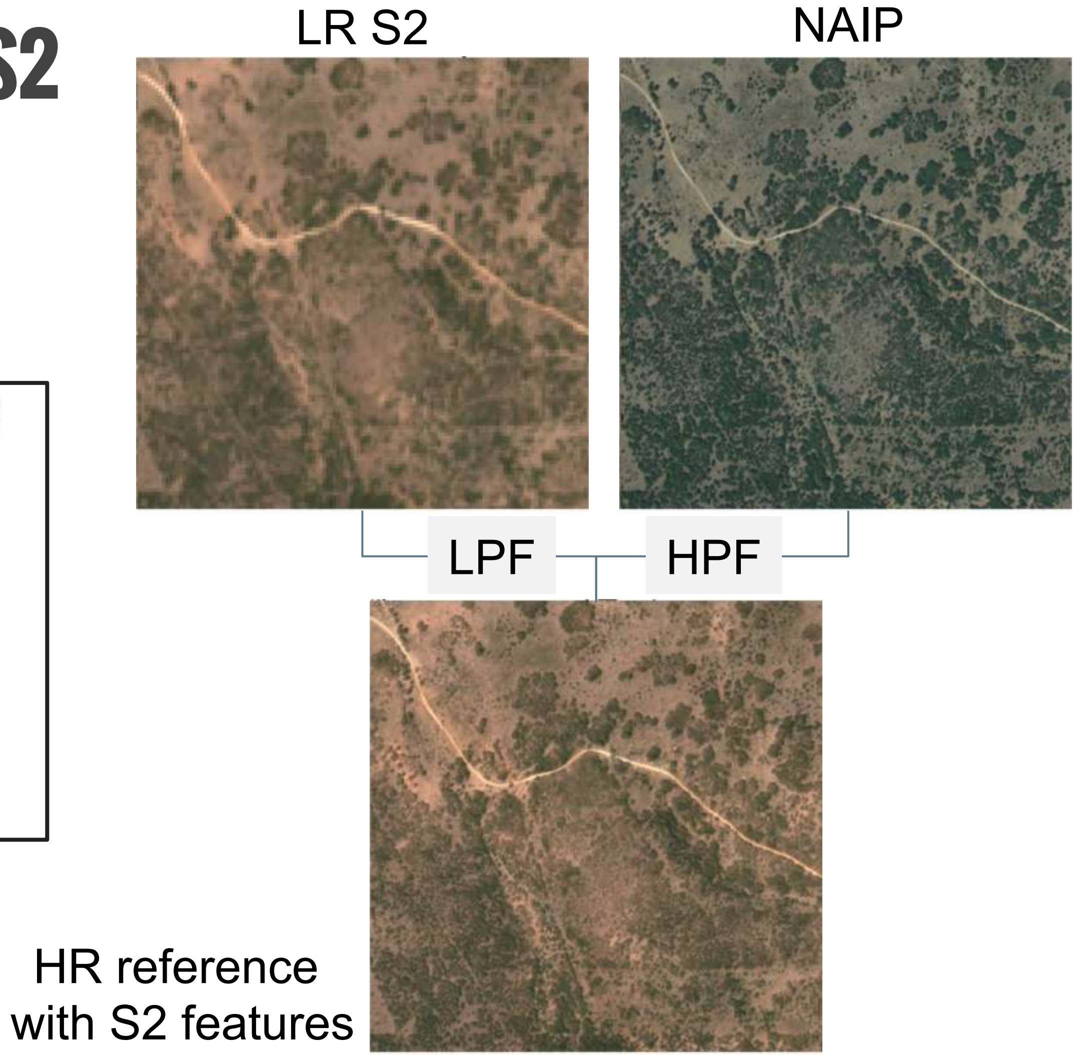
Real LR component

1. $\mathrm{LR\_up} = \mathrm{BicubicInterpolation}(LR, \mathrm{size} = SR\_size)$
2. $F_{SR} = \mathcal{F}(SR)$

   $F_{LR} = \mathcal{F}(\mathrm{LR\_up})$
3. $F_{SR}^{shifted} = \mathrm{FFTShift}(F_{SR})$

   $F_{LR}^{shifted} = \mathrm{FFTShift}(F_{LR})$
4. $F_{low} = F_{LR}^{shifted} \times \mathrm{LowPassMask}$

   $F_{high} = F_{SR}^{shifted} \times (1 - \mathrm{LowPassMask})$
5. $F_{filtered} = F_{low} + F_{high}$
6. $\mathrm{HybridImage} = \mathrm{Real}(\mathcal{F}^{-1}(\mathrm{IFFTShift}(F_{filtered})))$



OpenSR Dataset Real S2

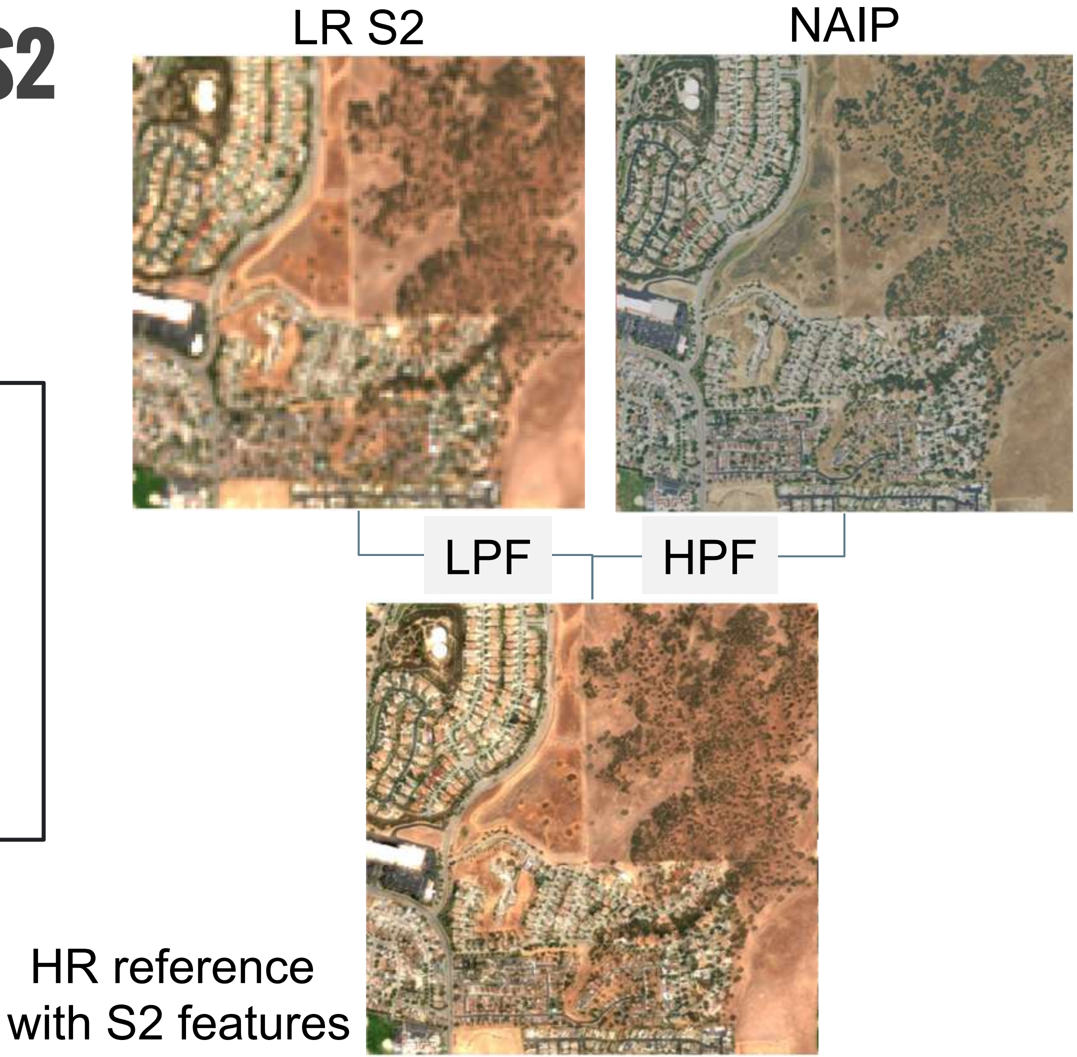
Real LR component

1. $\text{LR\_up} = \text{BicubicInterpolation}(LR, \text{size} = SR\_size)$
2. $F_{SR} = \mathcal{F}(SR)$

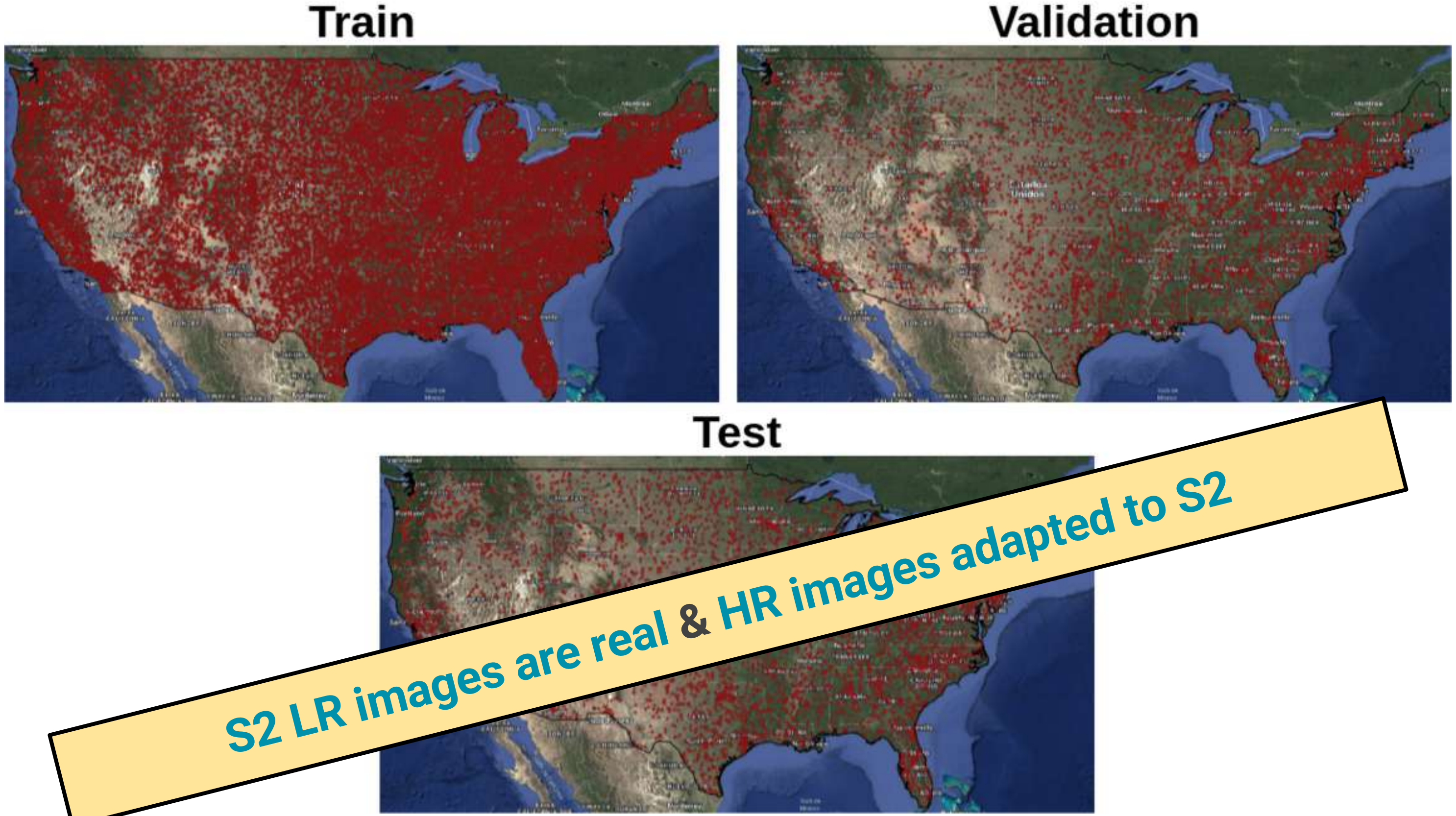
   $F_{LR} = \mathcal{F}(\text{LR\_up})$
3. $F_{SR}^{shifted} = \text{FFTShift}(F_{SR})$

   $F_{LR}^{shifted} = \text{FFTShift}(F_{LR})$
4. $F_{low} = F_{LR}^{shifted} \times \text{LowPassMask}$

   $F_{high} = F_{SR}^{shifted} \times (1 - \text{LowPassMask})$
5. $F_{filtered} = F_{low} + F_{high}$
6. $\text{HybridImage} = \text{Real}(\mathcal{F}^{-1}(\text{IFFTShift}(F_{filtered})))$

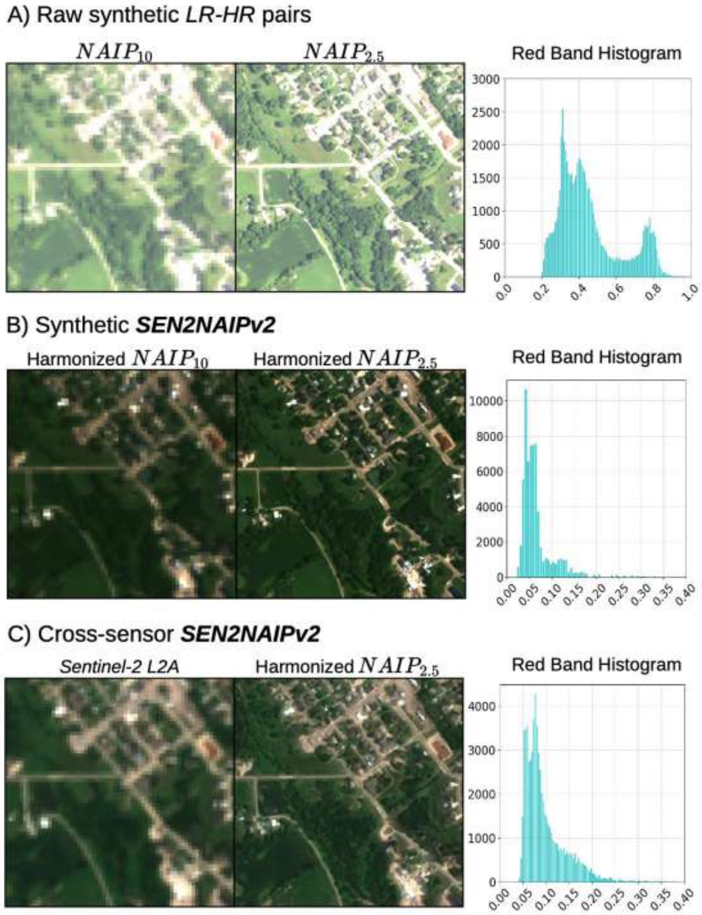


OpenSR Dataset Real S2



S2 LR images are real & HR images adapted to S2

Generation of Datasets



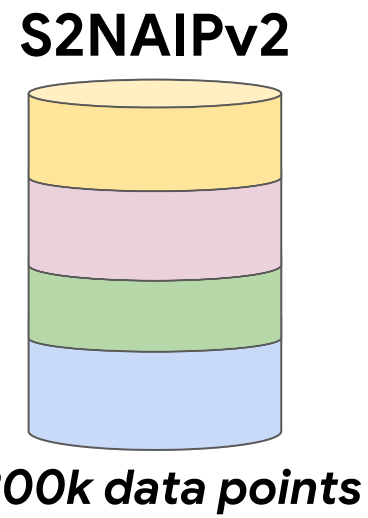
Standard synthetic approach

S2 LR are synthetic & HR harmonized to S2

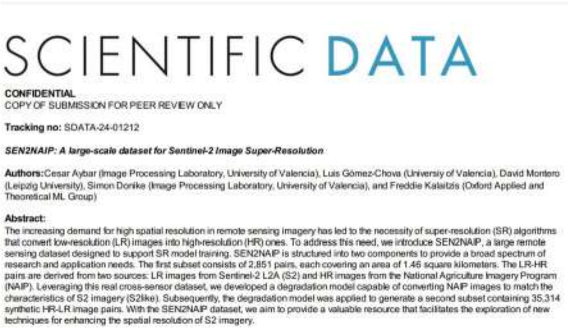
S2 LR images are real & HR adapted to S2

OpenSR Dataset

https://huggingface.co/ESA-philab



- ★ ***FAIR-compliant*** *dataset.*
- ★ *Cloud-optimized,* ***partial-read*** *capabilities.*
- ★ *Includes rich* ***metadata****.*
- ★ *The* ***largest dataset*** *up to date.*
- ★ ***STAC compliant***
- ★ ***Temporal*** support.



SCIENTIFIC DATA

CONFIDENTIAL
COPY OF SUBMISSION FOR PEER REVIEW ONLY

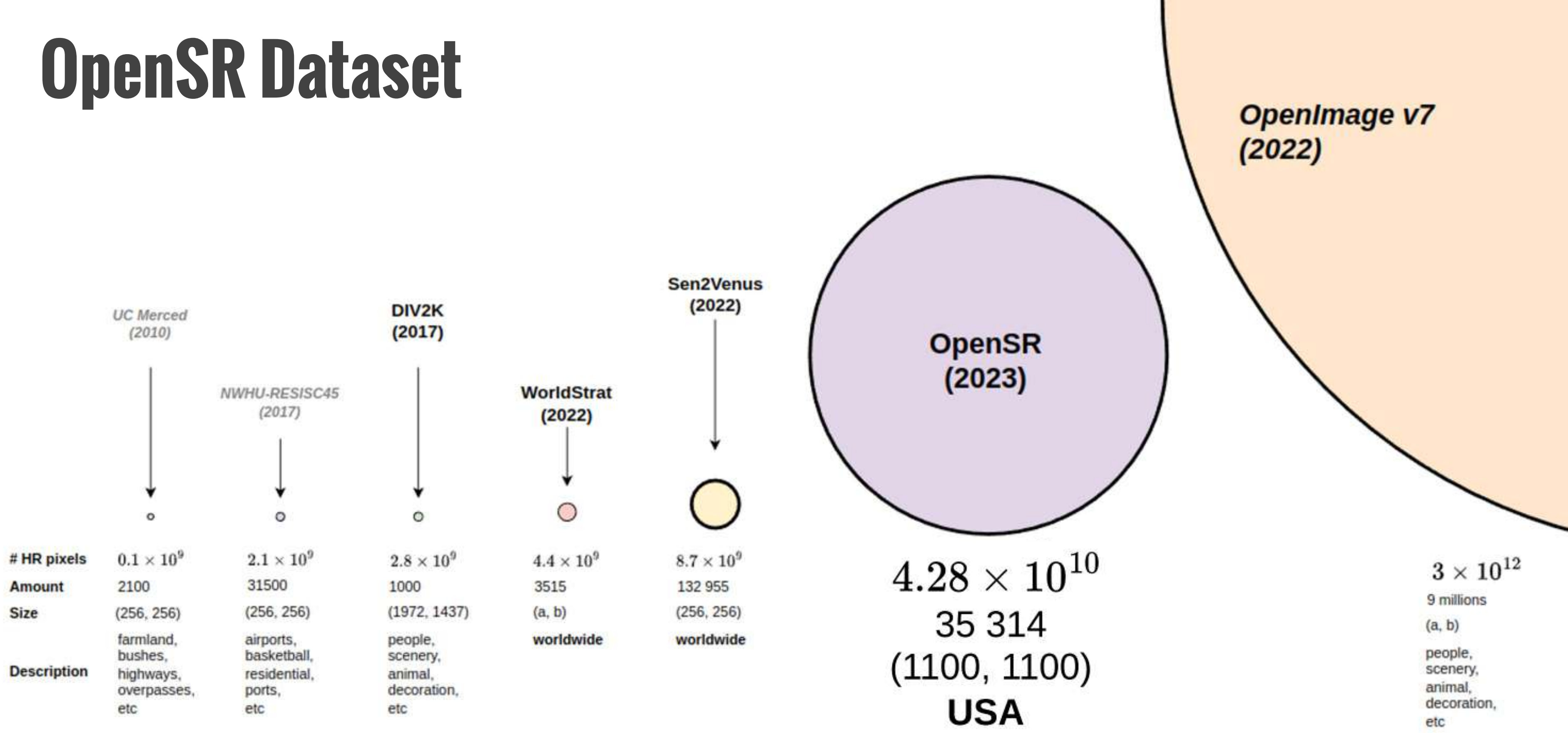
Tracking no: SDATA-24-01212

SEN2NAIP: A large-scale dataset for Sentinel-2 Image Super-Resolution

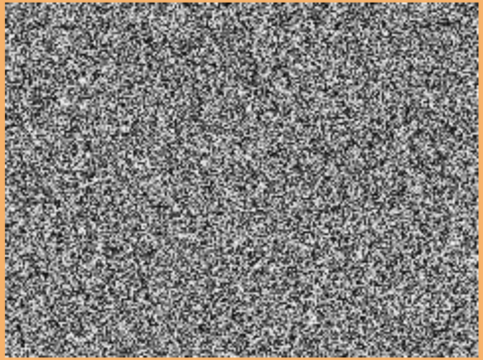
Authors: Cesar Aybar (Image Processing Laboratory, University of Valencia), Luis Gómez-Chova (Universiy of Valencia), David Montero (Leipzig University), Simon Donike (Image Processing Laboratory, University of Valencia), and Freddie Kalaitzis (Oxford Applied and Theoretical ML Group)

Abstract:
The increasing demand for high spatial resolution in remote sensing imagery has led to the necessity of super-resolution (SR) algorithms that convert low-resolution (LR) images into high-resolution (HR) ones. To address this need, we introduce SEN2NAIP, a large remote sensing dataset designed to support SR model training. SEN2NAIP is structured into two components to provide a broad spectrum of research and application needs. The first subset consists of 2,851 pairs, each covering an area of 1.46 square kilometers. The LR-HR pairs are derived from two sources: LR images from Sentinel-2 L2A (S2) and HR images from the National Agriculture Imagery Program (NAIP). Leveraging this real cross-sensor dataset, we developed a degradation model capable of converting NAIP images to match the characteristics of S2 imagery (S2like). Subsequently, the degradation model was applied to generate a second subset containing 35,314 synthetic HR-LR image pairs. With the SEN2NAIP dataset, we aim to provide a valuable resource that facilitates the exploration of new techniques for enhancing the spatial resolution of S2 imagery.

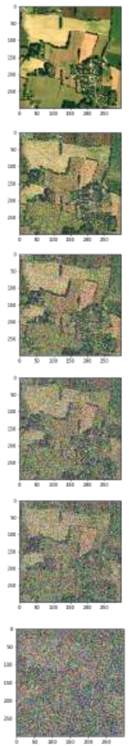
OpenSR Dataset



Diffusion Models

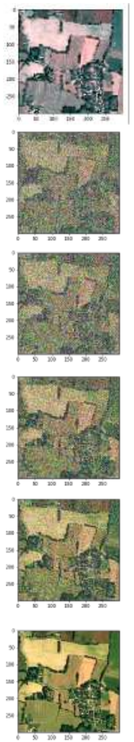


Diffusion Process



Forward Diffusion

- Gradually adds Gaussian noise
  - variance is hyperparameter
- Markov chain of *t* iterations



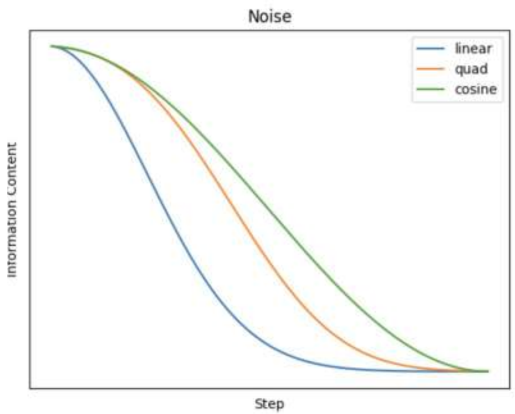
Reverse Diffusion

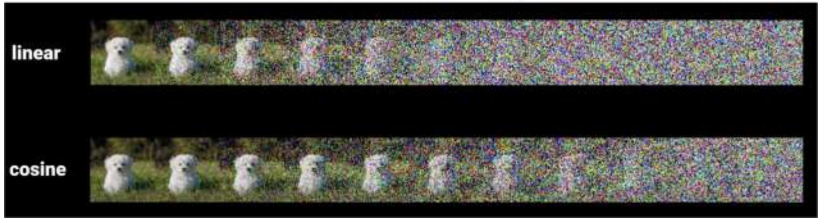
- Gradually removes noise
- Conditioned on LR input
- Reverse markov chain

https://arxiv.org/abs/2104.07636v2

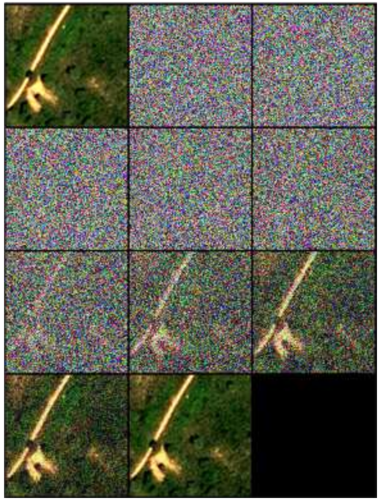
Optimization

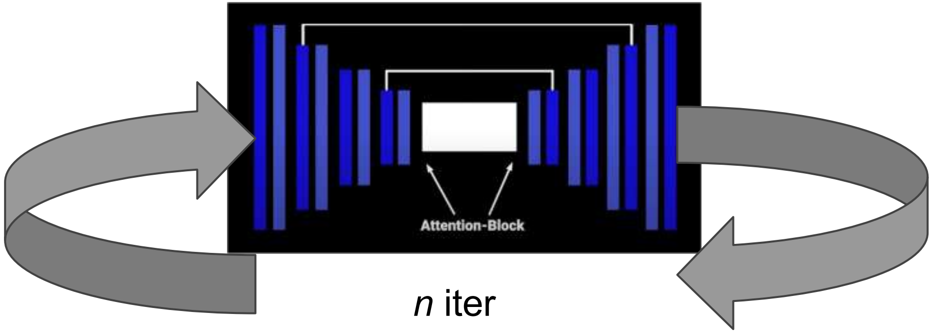
- Hyperparameter tuning
  - UNet architecture
  - Diffusion steps & noise schedule optimization





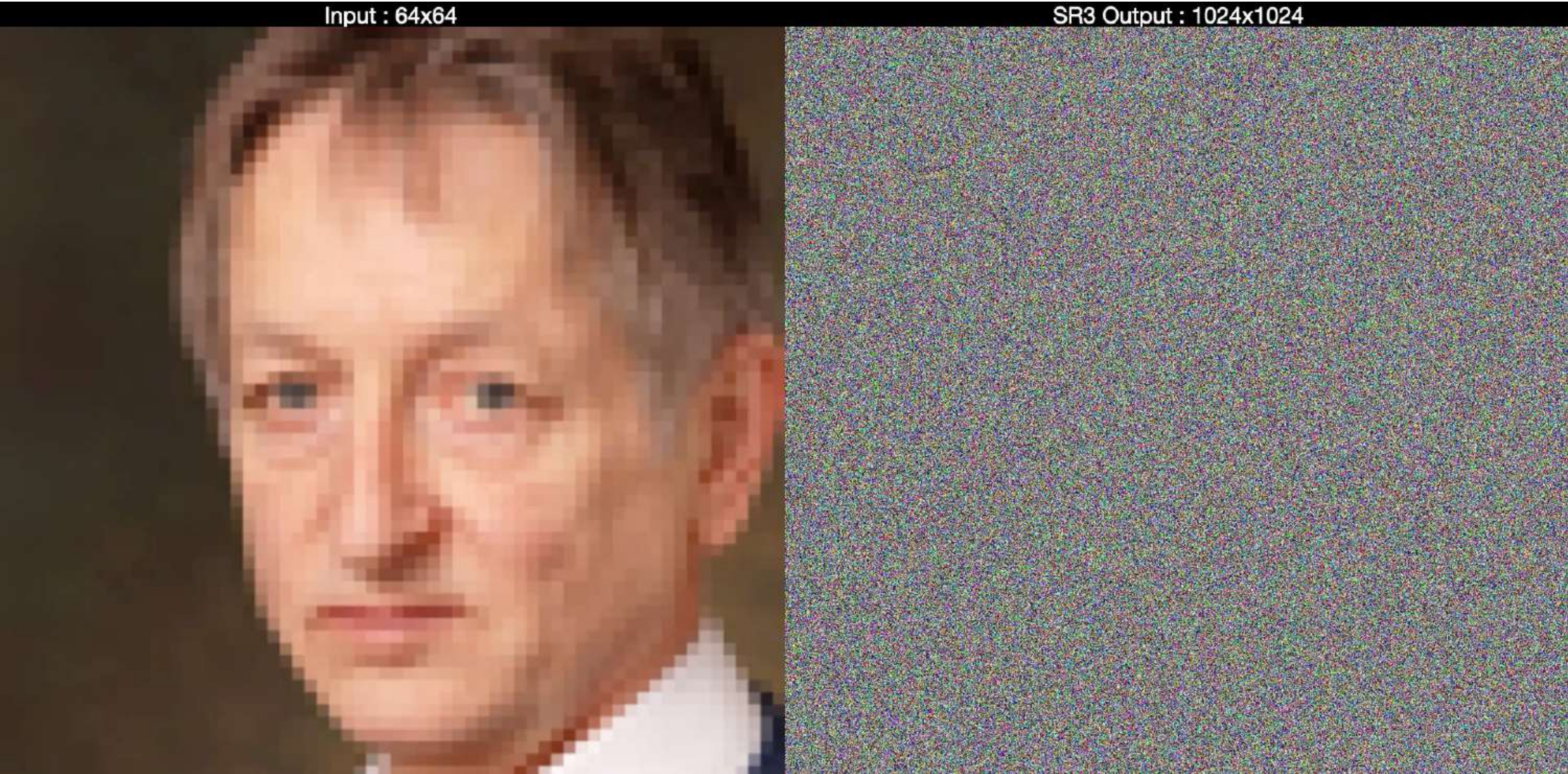
https://arxiv.org/pdf/2102.09672.pdf



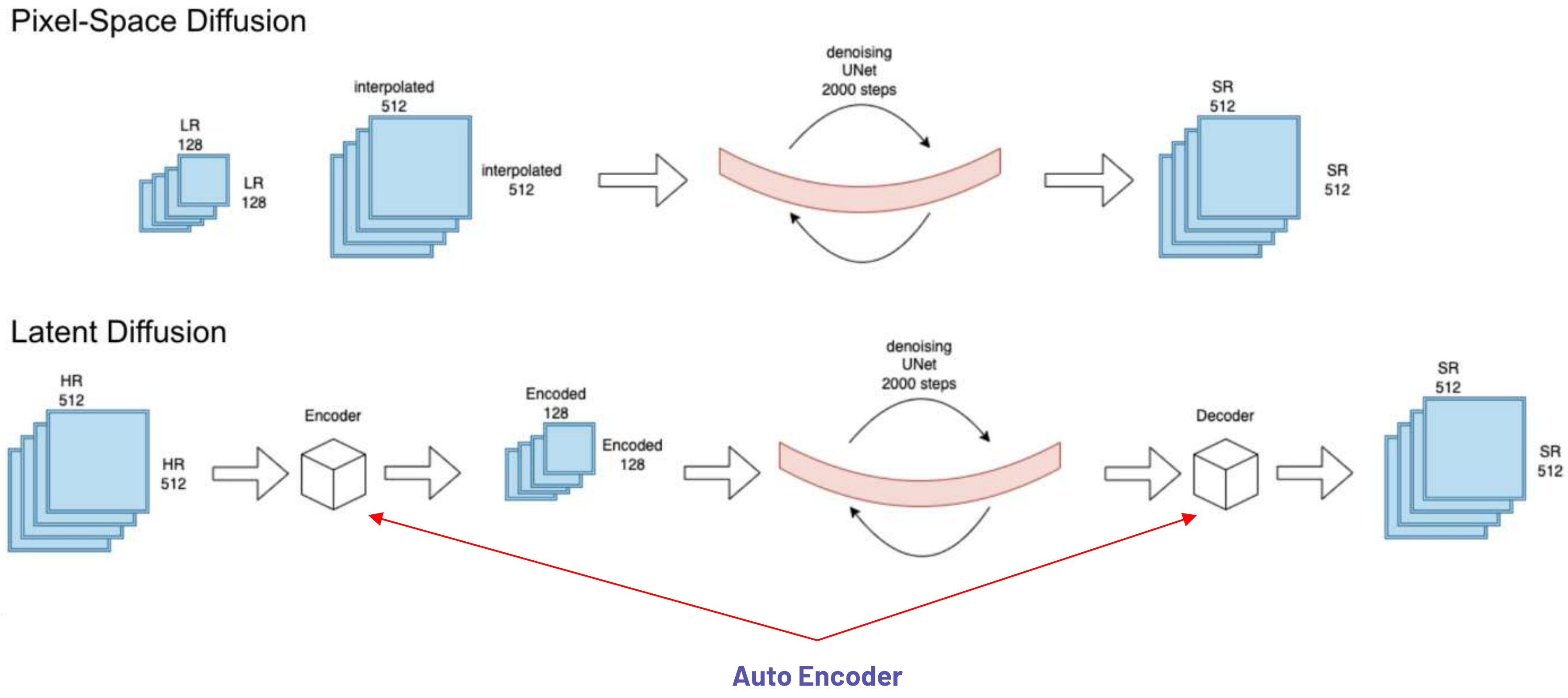


n iter

Why diffusion models for SR?



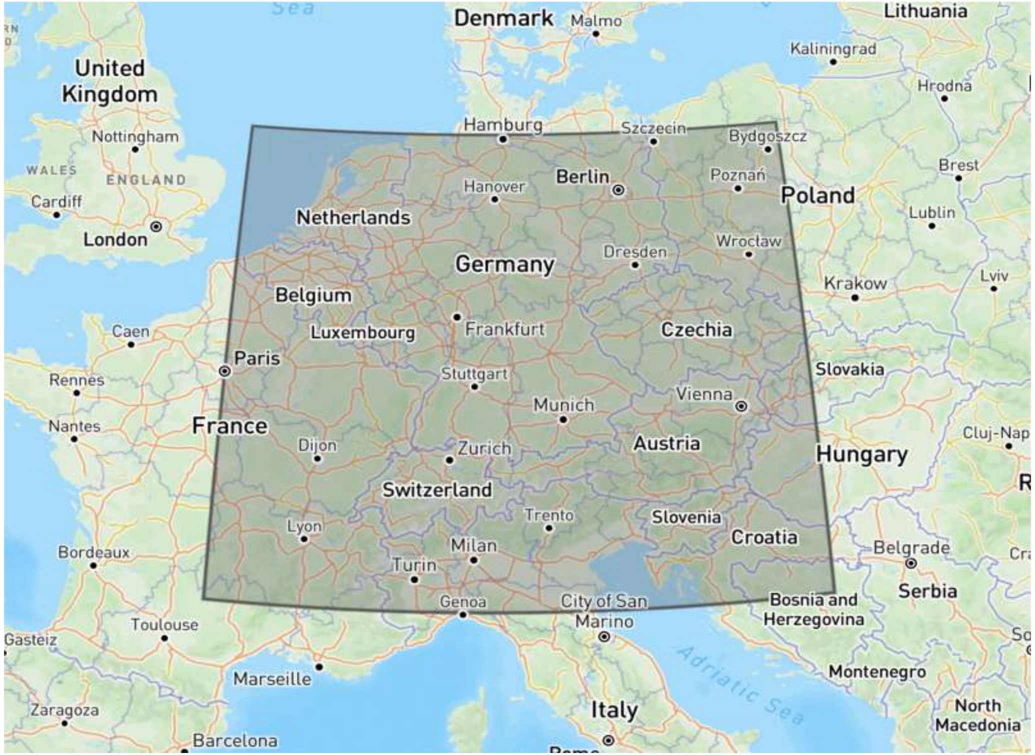
Why Latent Diffusion SR in RS?



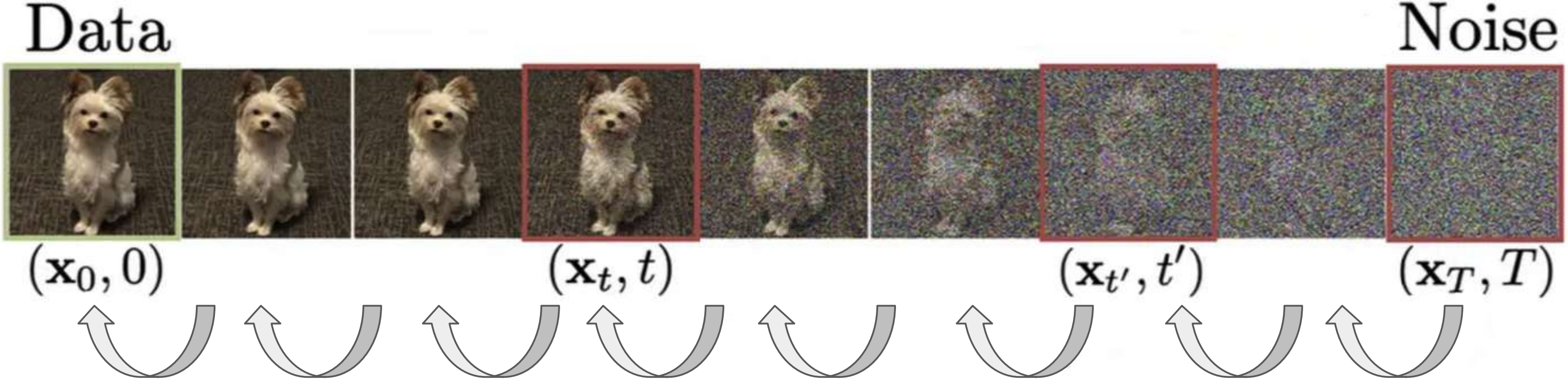
Why Latent Diffusion SR in RS?

Inference Speed for 1M sqkm

Latent-Space Diffusion	Pixel-Space Diffusion
≈ 16 hrs	≈ 11 weeks



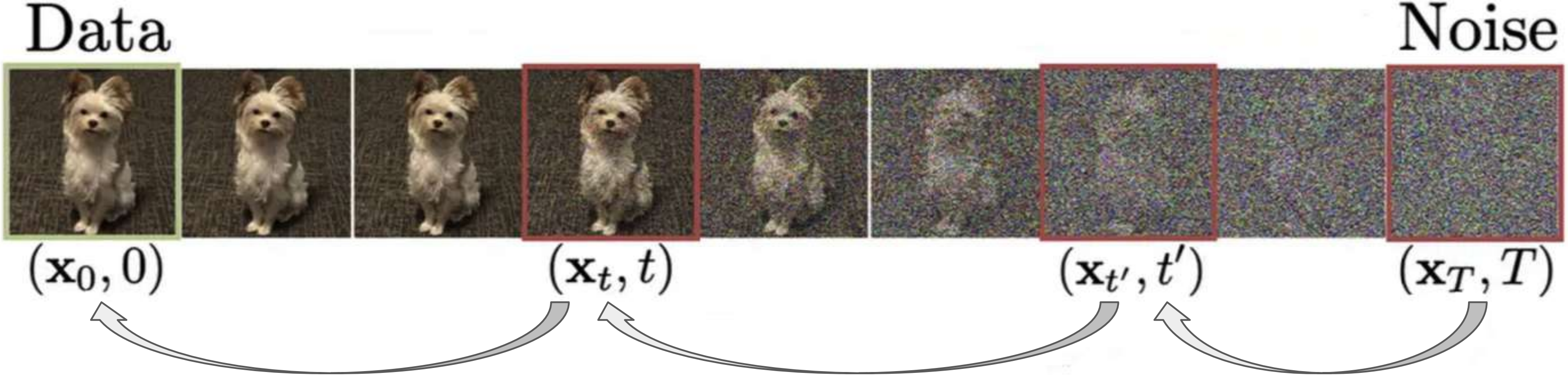
Denoising Diffusion Probabilistic Models - Sampling



Markov Chain:

- Each step individually depends on each previous step
- Noise randomly sampled at each step ⟶ traverse whole chain for 1 inf step

Denoising Diffusion Implicit Models - Sampling

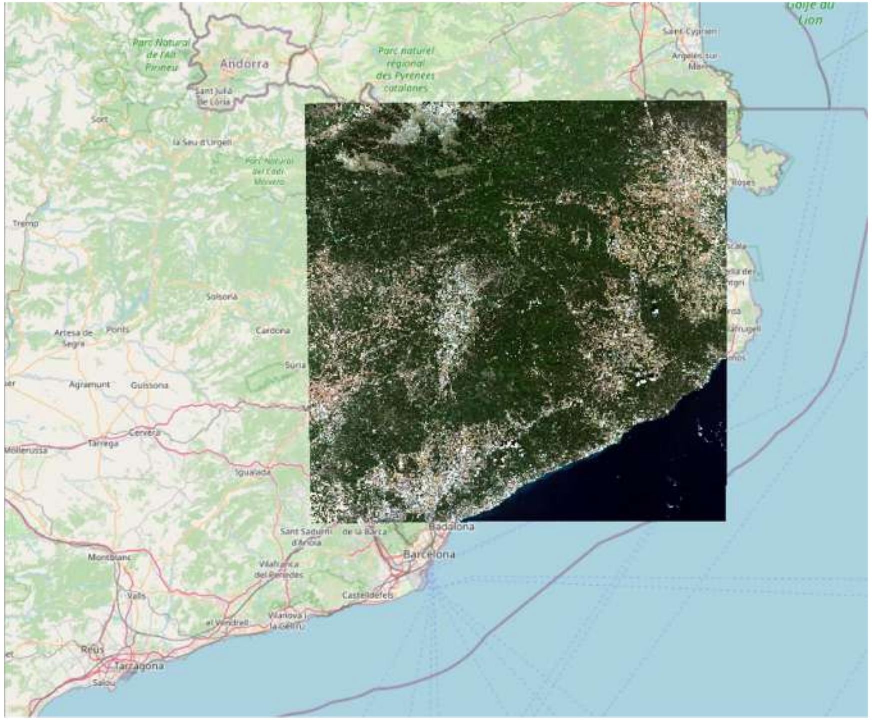


Non-Markovian:

- Random noise only at $x_t \longrightarrow$ step depends on all previous steps
- More deterministic $\longrightarrow$ less sampling diversity

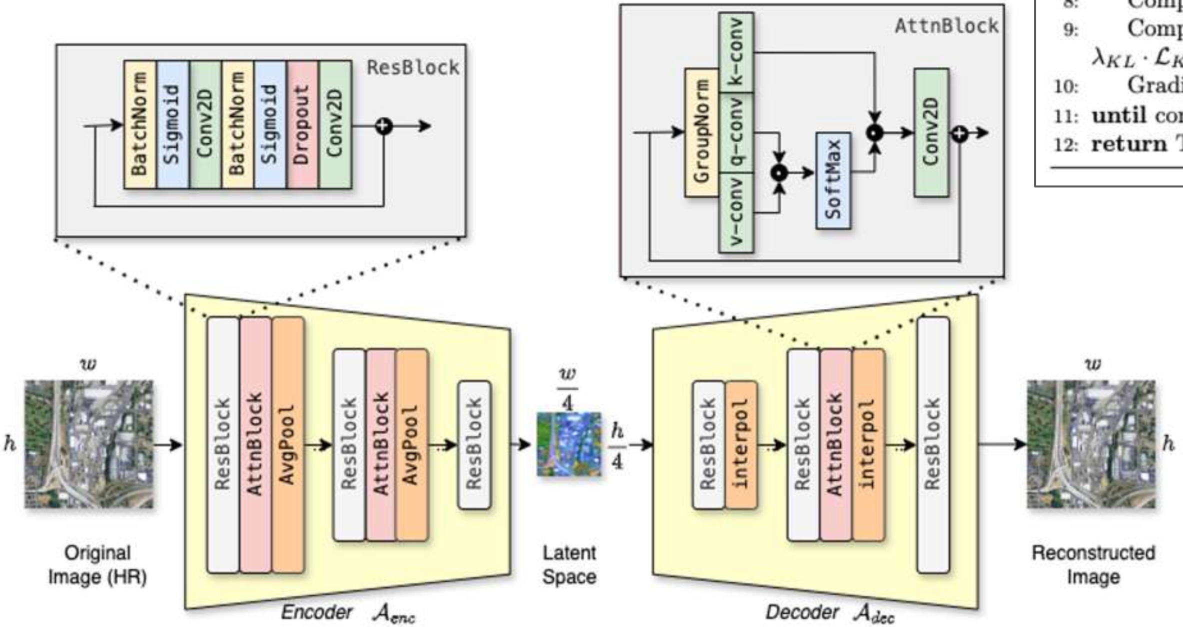
DDIM Inference Speed

Method	S2 Tile Inference Time (hh:mm)
Pixel-Space DDPM Sampling	1656:50
Pixel-Space DDIM Sampling	96:35
Latent-Space DDPM Sampling	06:20
Latent-Space DDIM Sampling - LDSR-S2	0:22



Latent Difussion SR

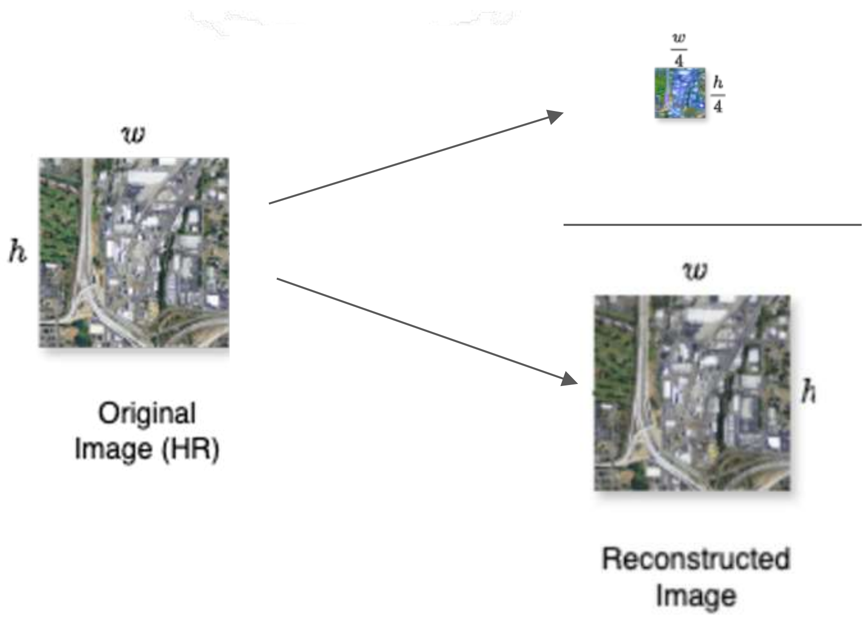
AutoEncoder Training



Algorithm 1 Training of Autoencoder

Input: HR image x_{HR}
Load pretrained LPIPS weights $\mathcal{W}_{LPIPS}$
Initialize encoder and decoder parameters $\mathcal{A}_{enc,\theta}$ $\mathcal{A}_{dec,\theta}$
repeat
  Encode $z = \mathcal{A}_{enc}(x_{HR})$
  Decode $\hat{x}_{HR} = \mathcal{A}_{dec}(z)$
  Shuffle and randomly select 3 bands of $\hat{x}_{HR}$ to form $\hat{x}_{HR}^{\text{shuffled}}$
  Compute $\mathcal{L}_{LPIPS}$ using $\mathcal{W}_{LPIPS}$ on $\hat{x}_{HR}^{\text{shuffled}}$
  Compute total loss $\mathcal{L}_{total} = \lambda_{MAE} \cdot \mathcal{L}_{MAE}(\hat{x}_{HR}) + \lambda_{GAN} \cdot \mathcal{L}_{GAN}(\hat{x}_{HR}) + \lambda_{KL} \cdot \mathcal{L}_{KL}(z) + \lambda_{LPIPS} \cdot \mathcal{L}_{LPIPS}$
  Gradient descent step to minimize $\mathcal{L}_{total}$
until converged
return Trained model parameters $\mathcal{A}_{enc,\theta}$ $\mathcal{A}_{dec,\theta}$

AutoEncoder Training



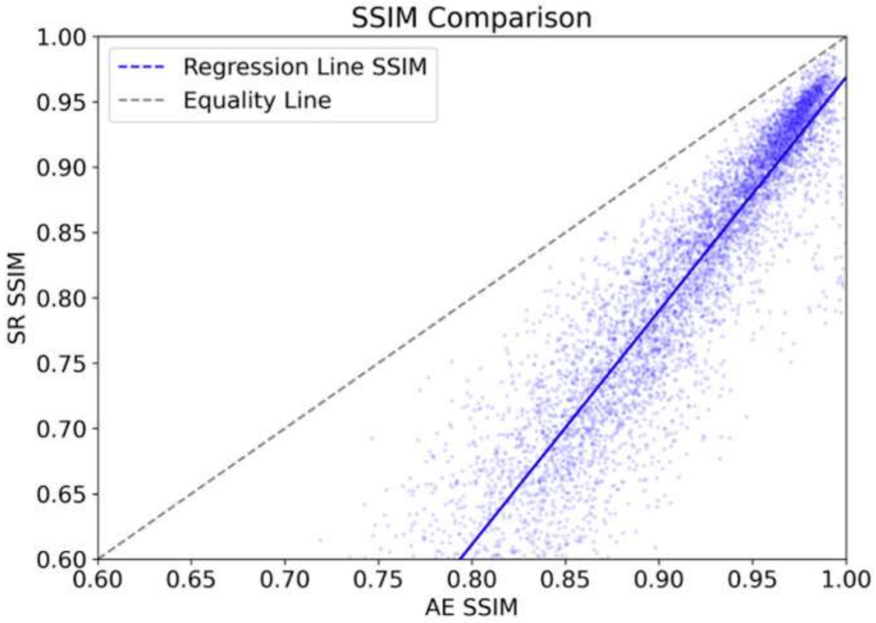
Latent space

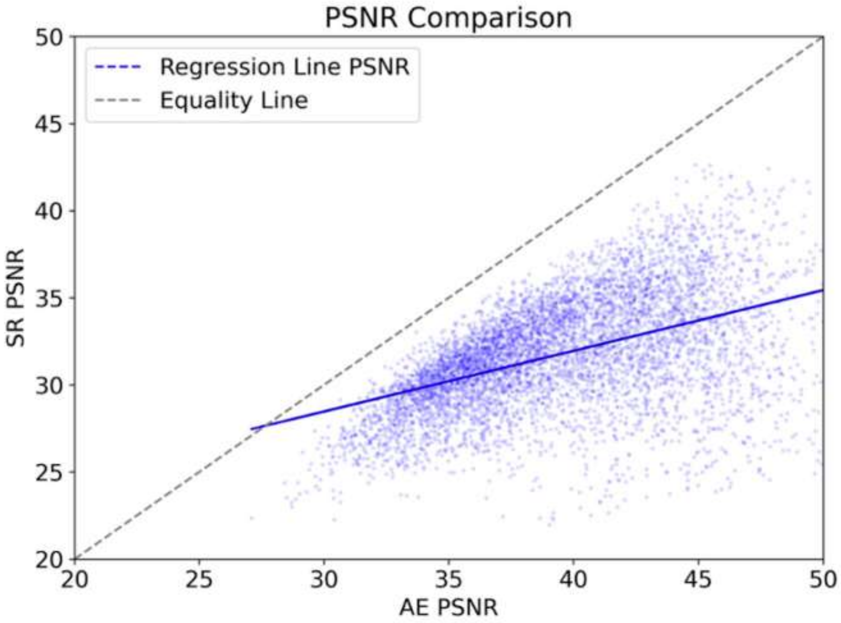
- Histogram Divergence Loss

Image space

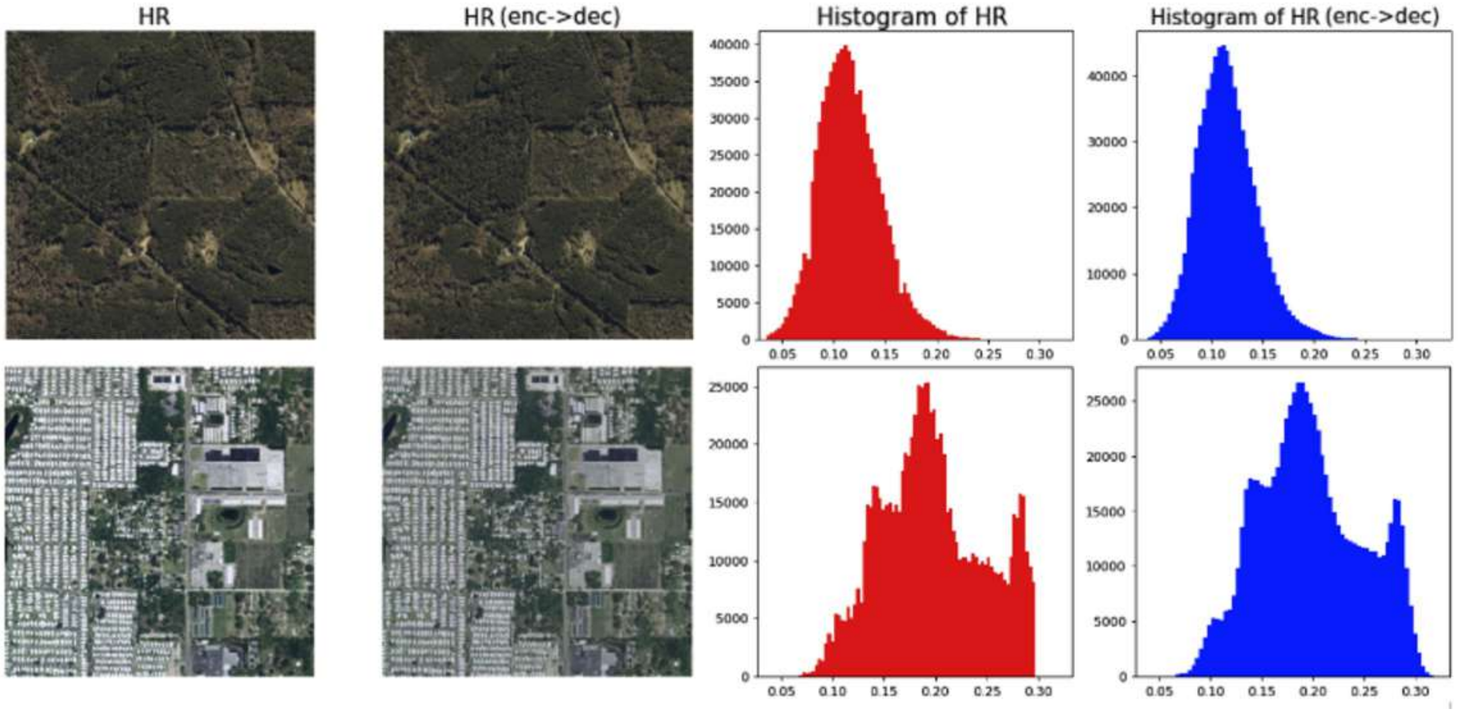
- MAE $\longrightarrow$ fidelity
- GAN $\longrightarrow$ adversarial
- LPIPS $\longrightarrow$ perceptual

AutoEncoder Training



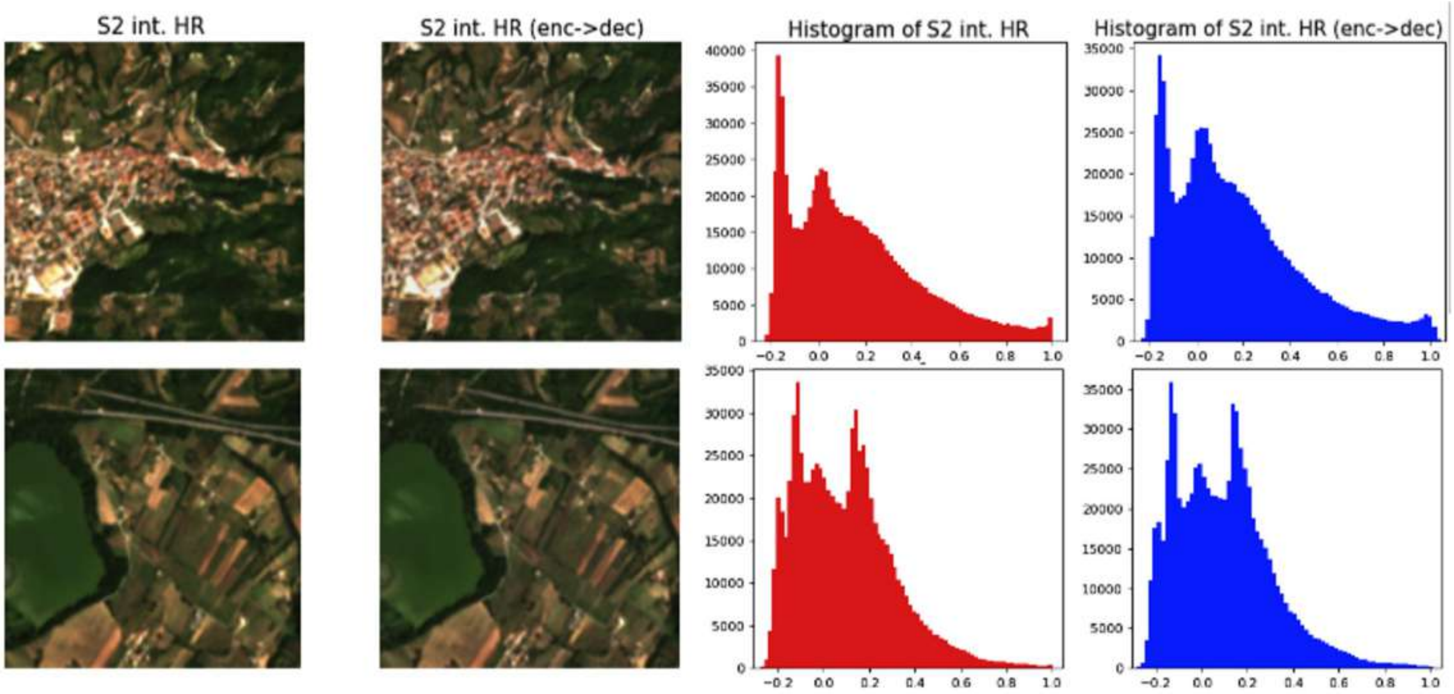


AutoEncoder Results



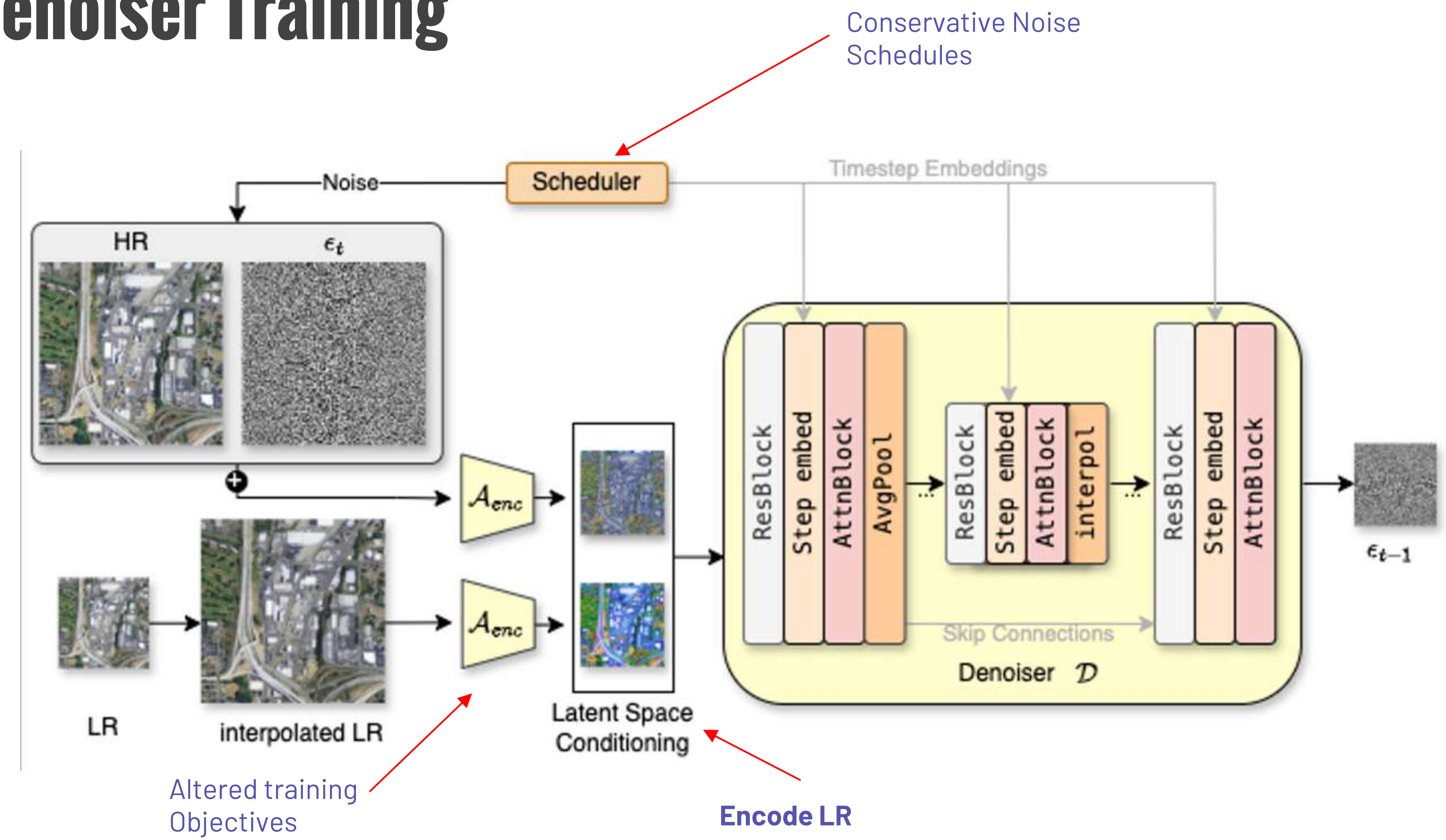
Data Type	PSNR ↑	SSIM ↑
SEN2NAIP	37.08	0.918
S2 - LR int.	45.88	0.994

AutoEncoder Results - S2



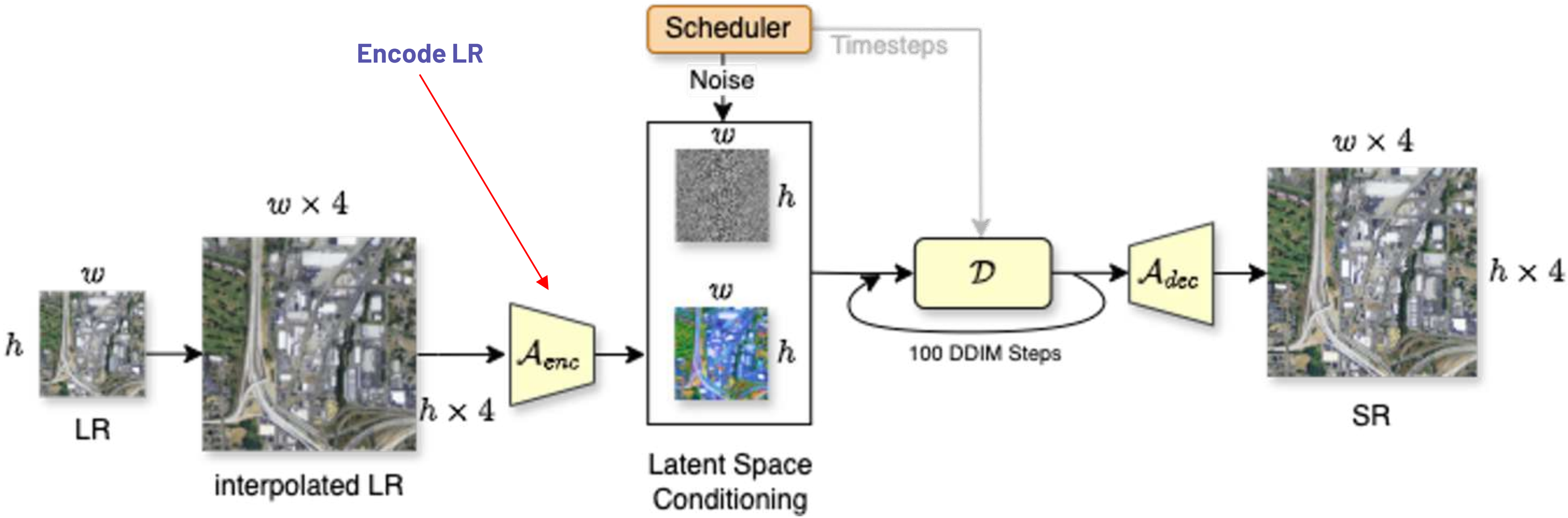
Data Type	PSNR ↑	SSIM ↑
SEN2NAIP	37.08	0.918
S2 - LR int.	45.88	0.994

Denoiser Training



Donike *et al.*, **Trustworthy Super-Resolution of Multispectral Sentinel-2 Imagery With Latent Diffusion**, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

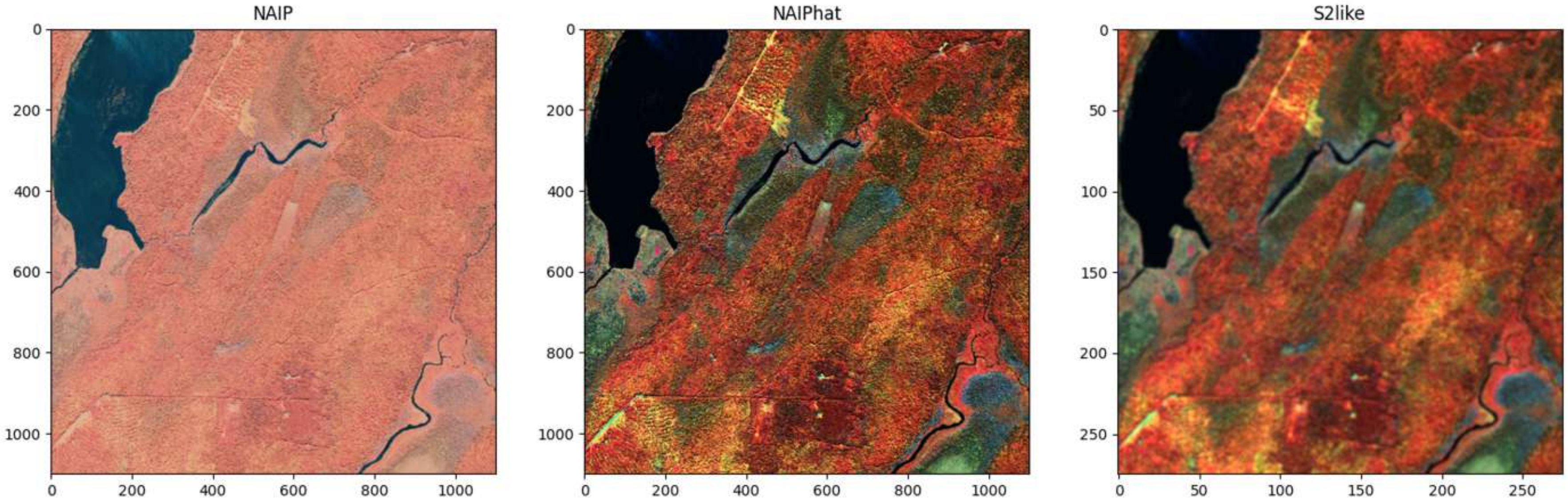
SR Inference - LDSR-S2



Donike *et al.*, **Trustworthy Super-Resolution of Multispectral Sentinel-2 Imagery With Latent Diffusion**, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

Training Data: SEN2NAIP v1

- 60k samples
- Multiple degradation kernels



Aybar *et al.*, **SEN2NAIP: A large-scale dataset for Sentinel-2 Image Super-Resolution**, *Scientific Data*, 2024

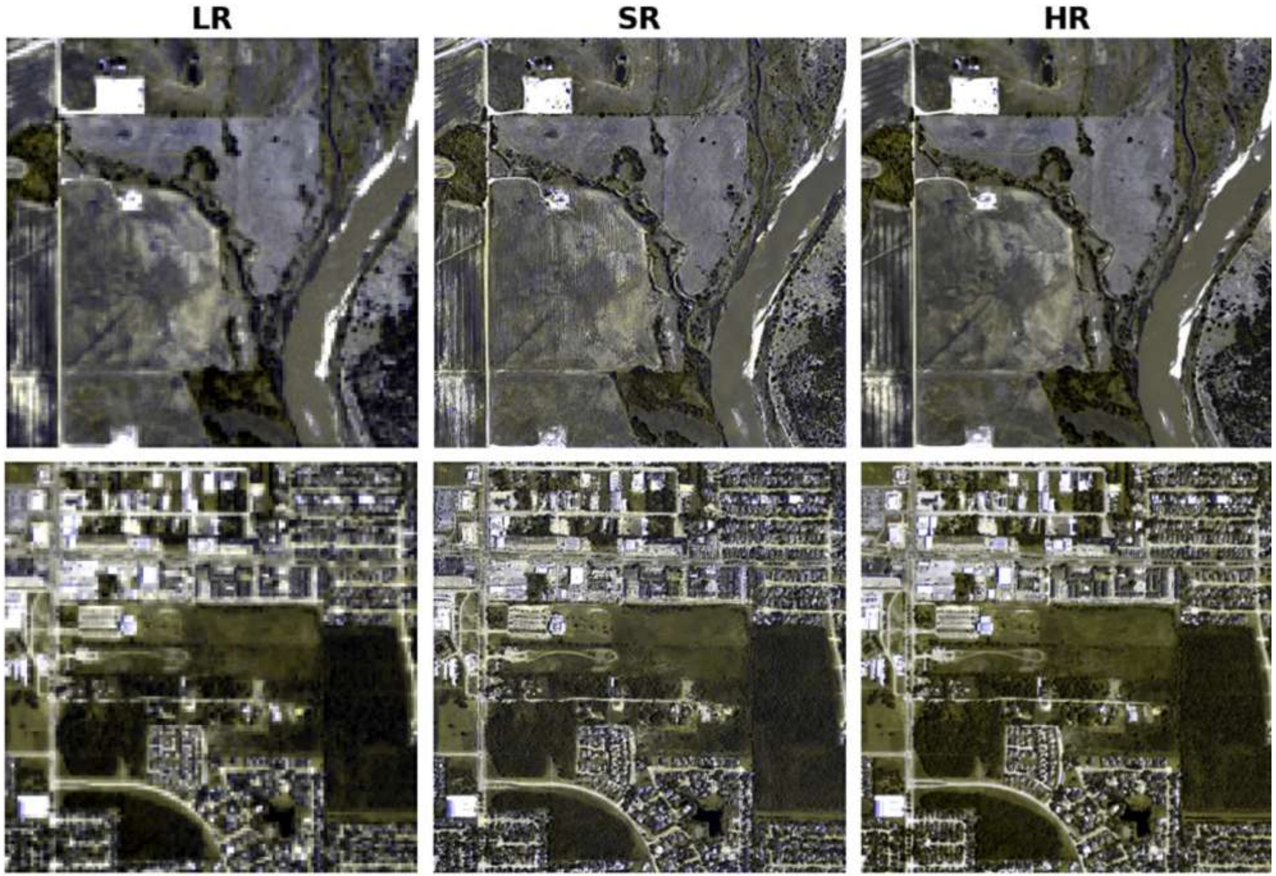
Training Data: SEN2NAIP v2

- 8k samples
- Frequency-transfer
- “real” S2

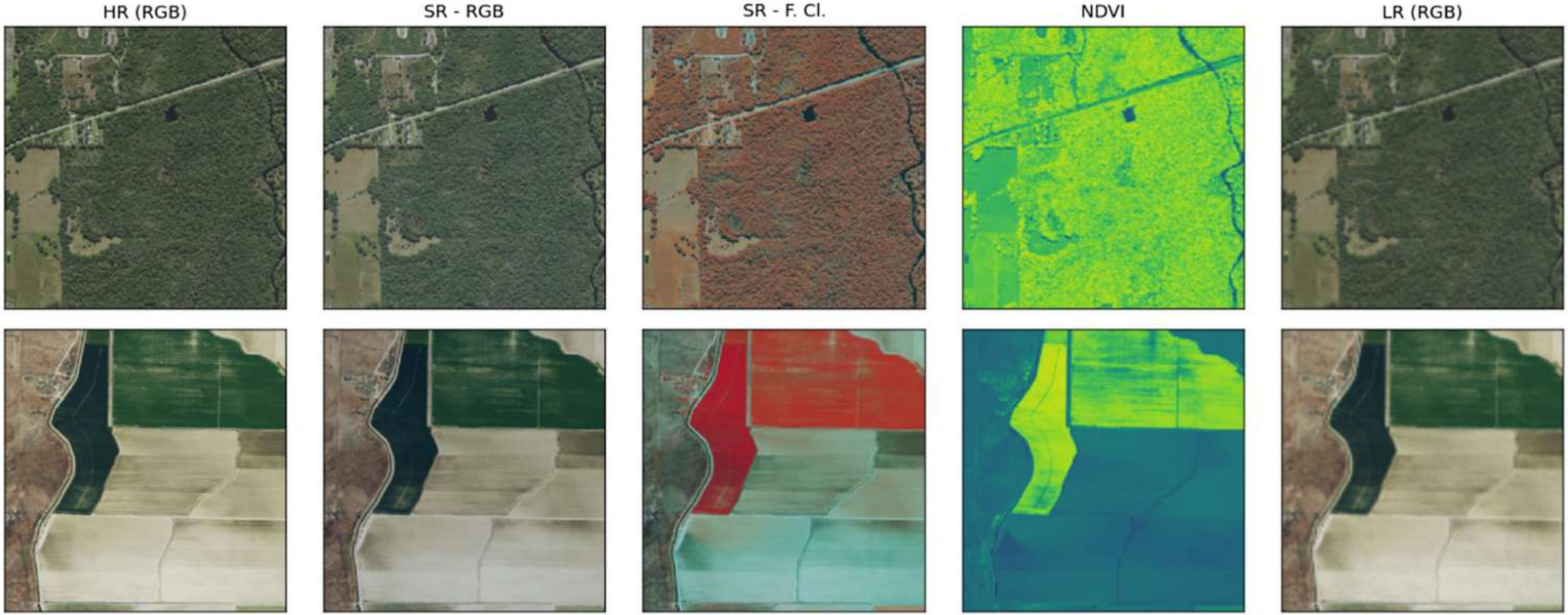
Aybar *et al.*, **SEN2NAIP: A large-scale dataset for Sentinel-2 Image Super-Resolution**, *Scientific Data*, 2024

Latent Difussion SR Results

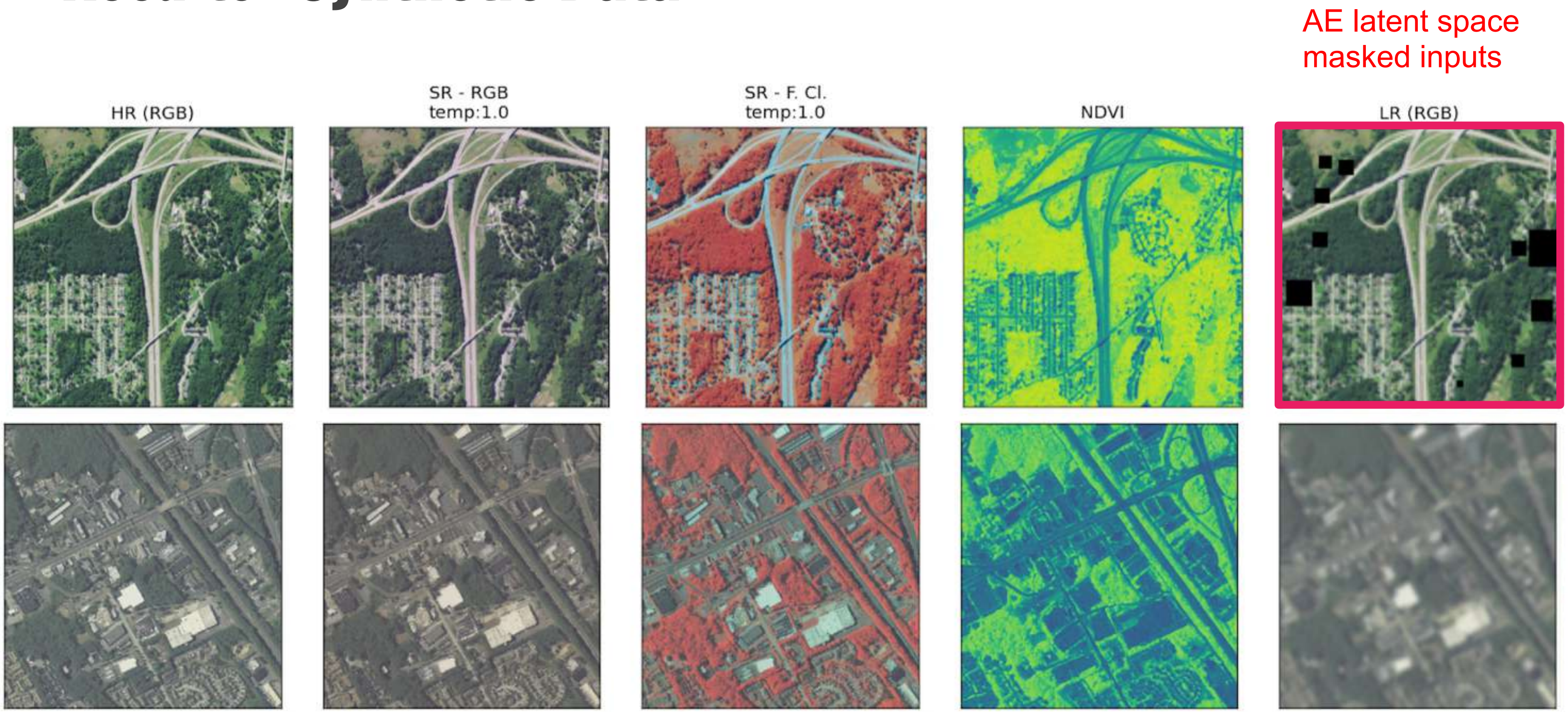
Results - Synthetic Data



Results - Synthetic Data

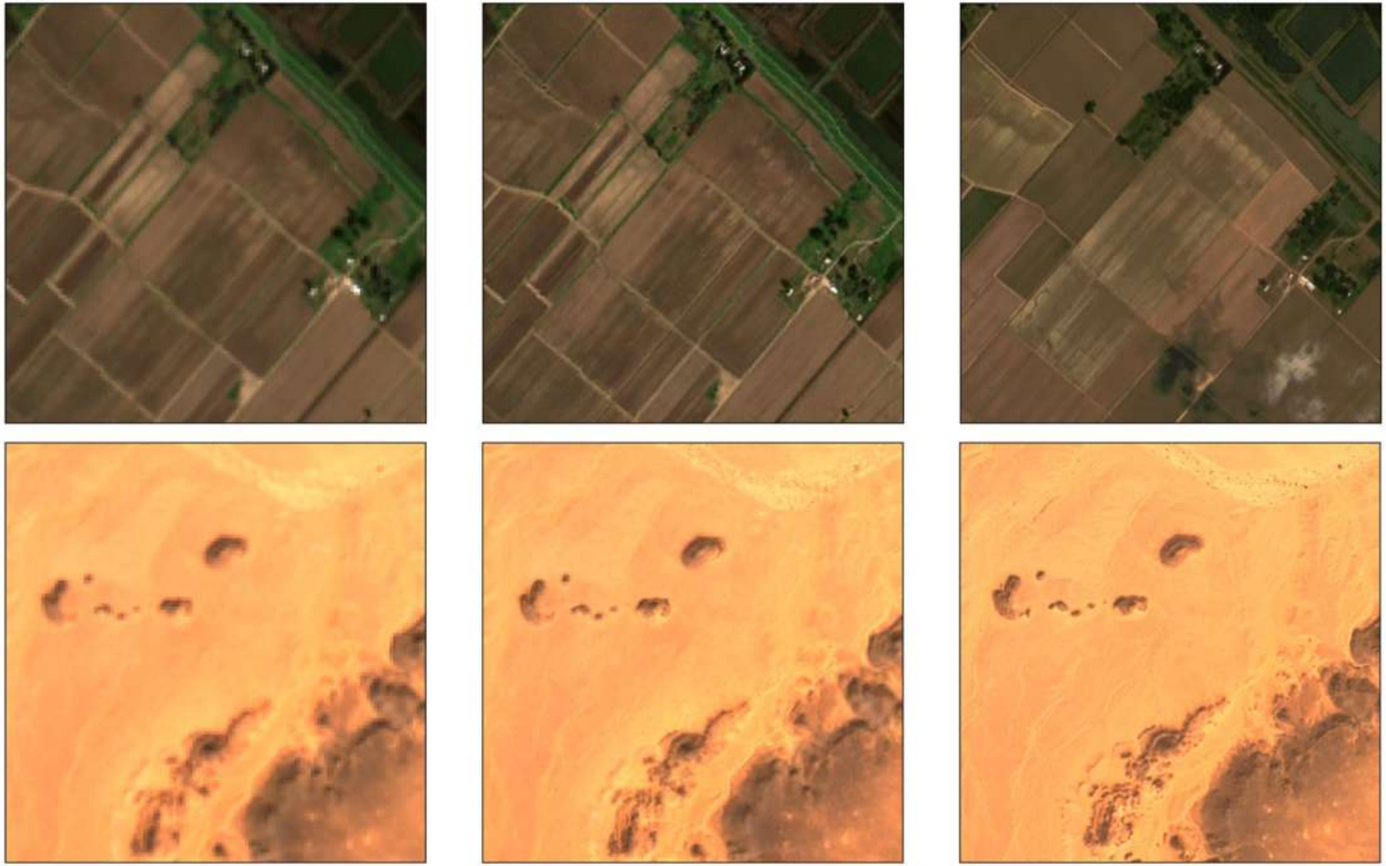


Results - Synthetic Data

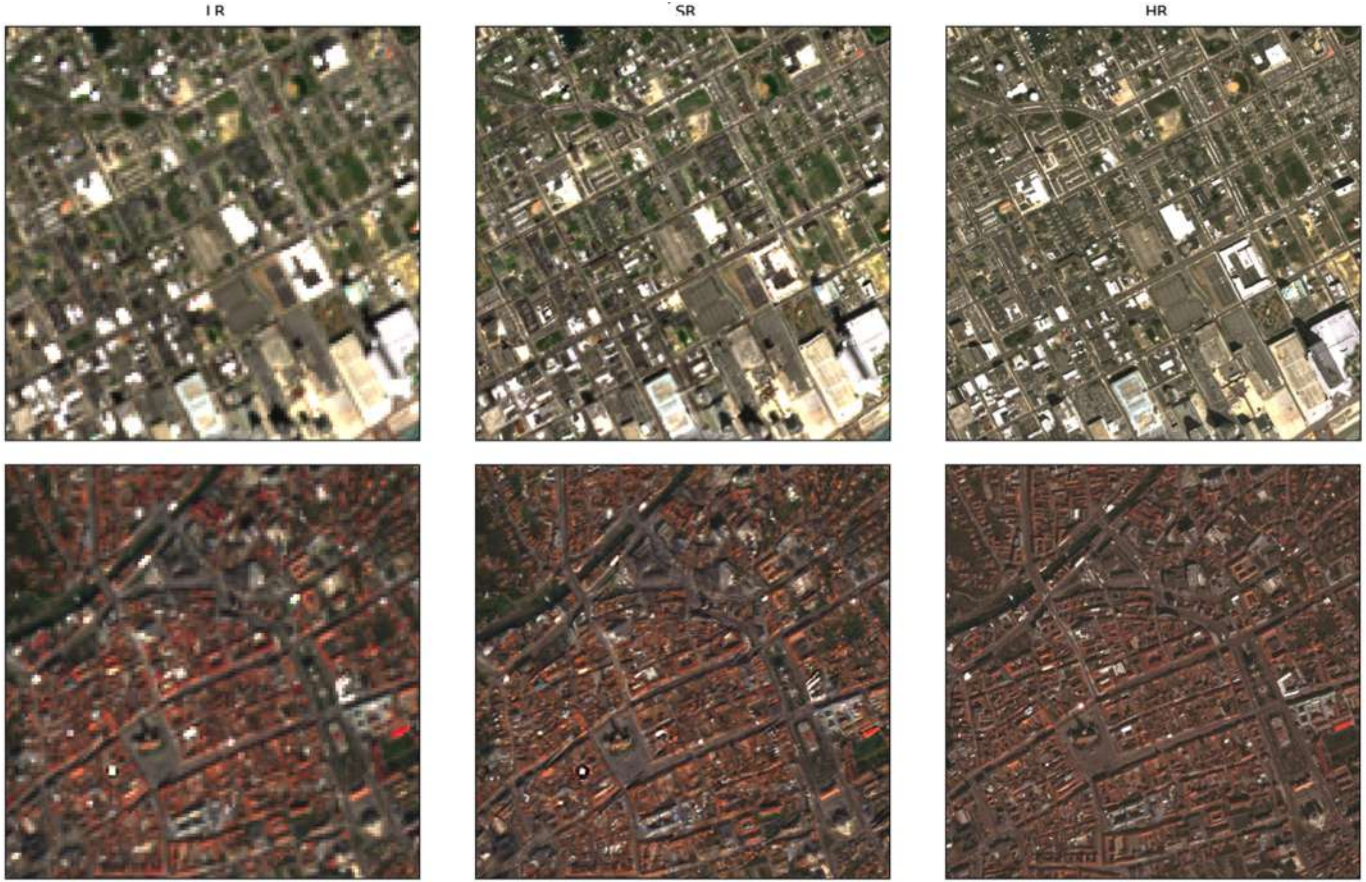


Results - S2

LR SR HR

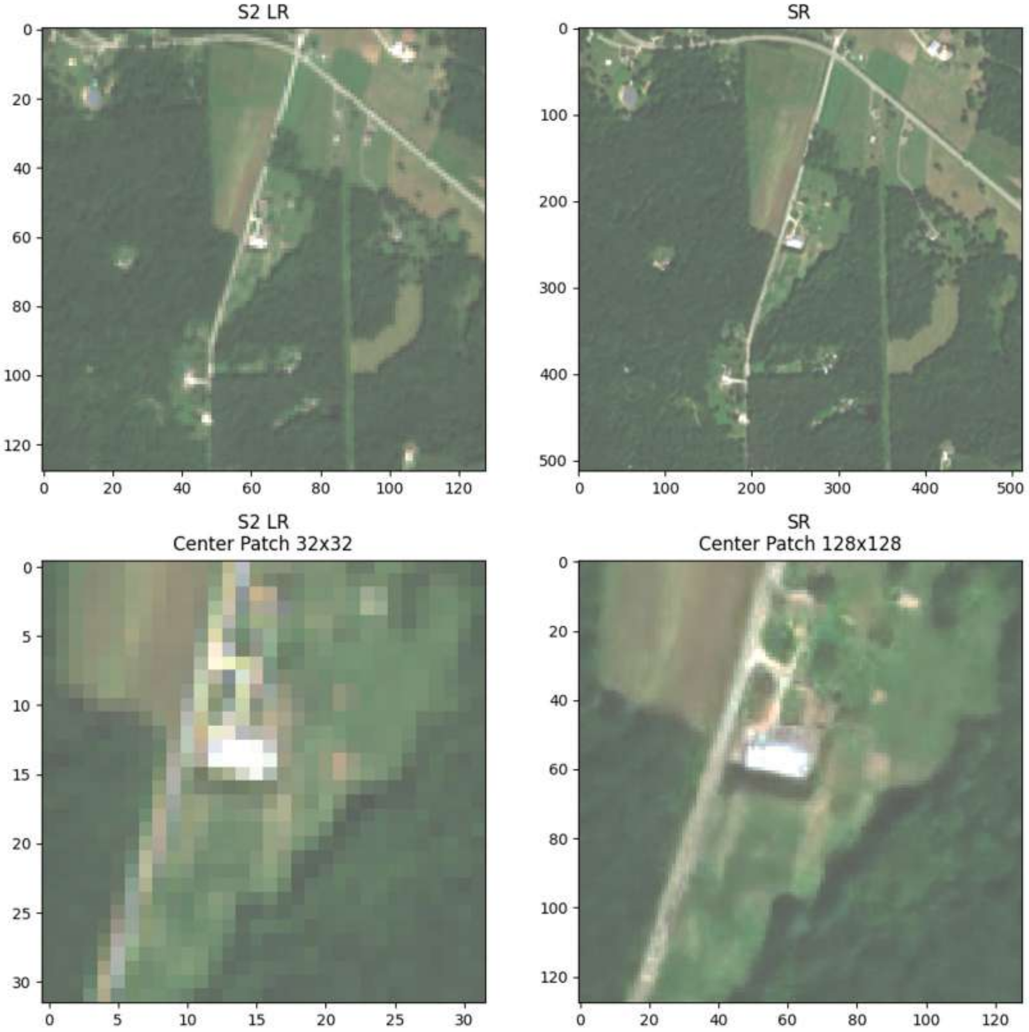


Results - S2



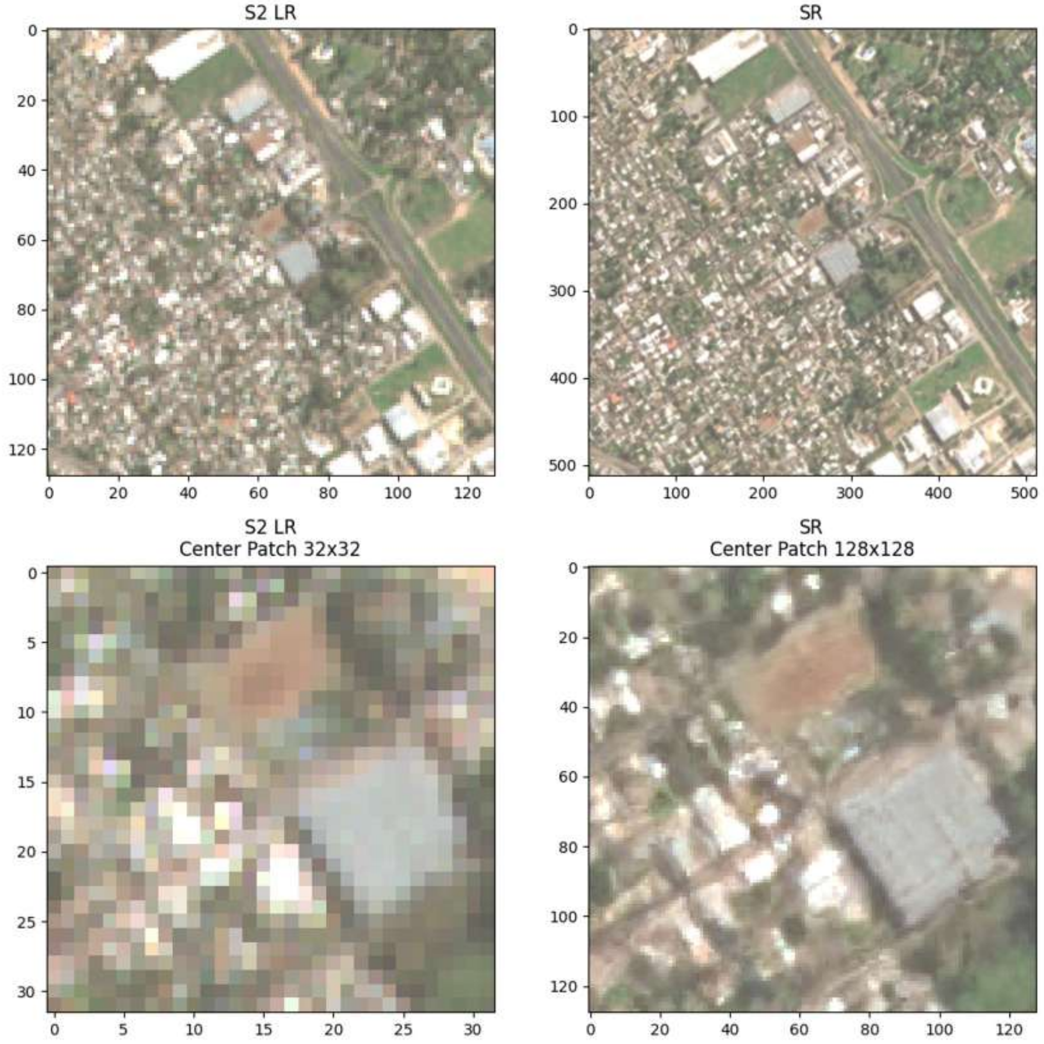
Results - S2

S2-RGB Results

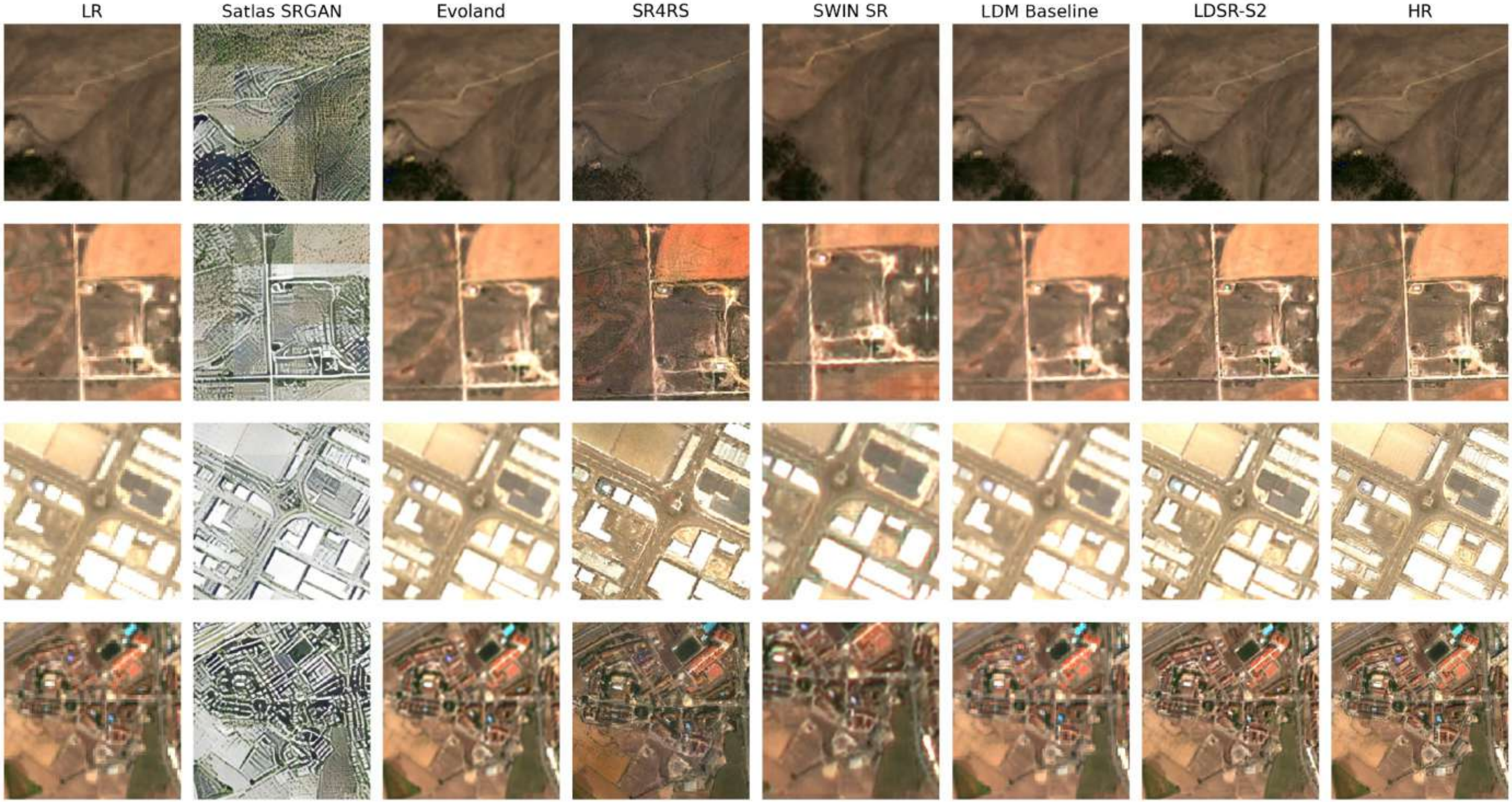


Results - S2

S2-RGB Results



Results - SOTA Comparison



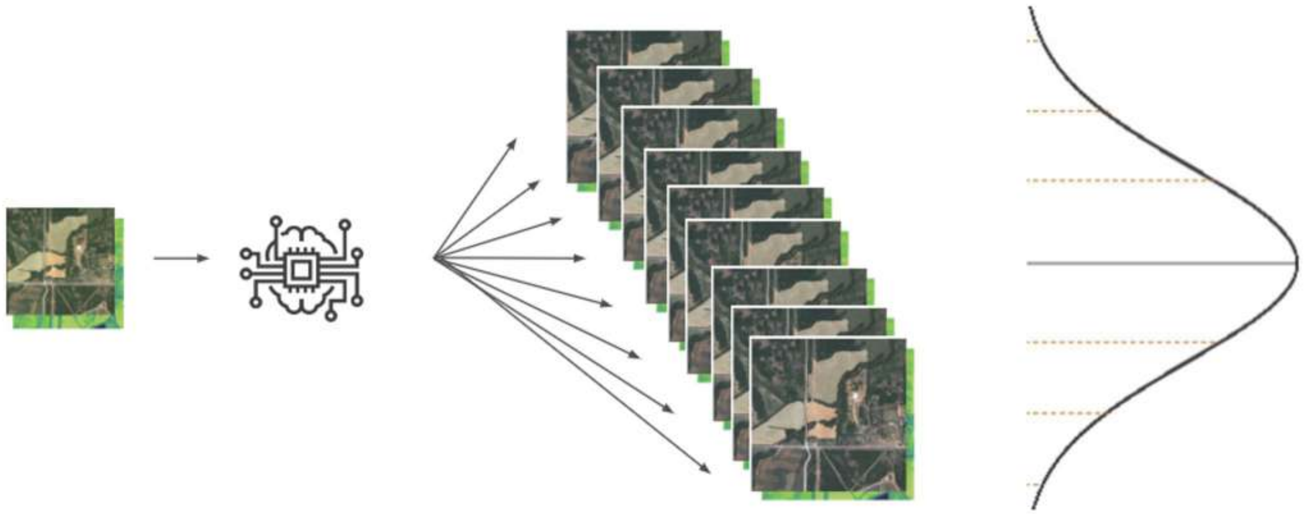
Results - SOTA Comparison

Comparison on the *opensr-test* datasets and SR models. Best results are highlighted

	LDM baseline	Pixel-Space Diffusion	LDSR-S2 model	Satlas SRGAN	SR4RS model	SWIN Transformer
Reflectance ↓	0.029	0.089	**0.003**	0.049	0.027	0.005
Spectral ↓	9.68	1.43	**1.26**	12.11	3.40	1.56
Spatial ↓	0.07	0.50	**0.01**	0.27	0.92	0.02
Synthesis ↑	**0.027**	0.013	0.007	0.022	0.013	0.005
Hallucination ↓	0.60	0.73	**0.34**	0.79	0.66	0.39
Omission ↓	0.31	0.13	0.46	**0.11**	0.20	0.59
Improvement ↑	0.09	0.14	**0.21**	0.10	0.14	0.11
PSNR ↑	30.82	36.15	**37.23**	10.78	33.55	34.20
SSIM ↑	0.63	0.83	**0.89**	0.23	0.67	0.84

LDSR Uncertainty Estimation

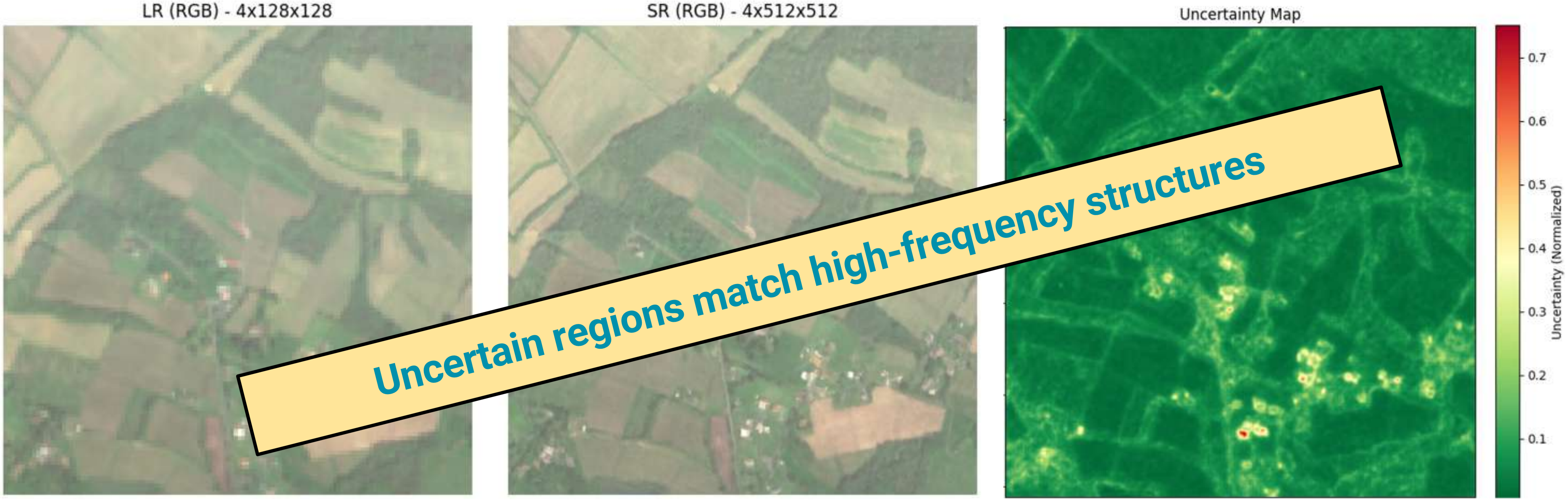
Uncertainty Estimation



$$\mathrm{CI}(x) = [\hat{y} - \epsilon, \hat{y} + \epsilon]$$
$$\epsilon = \sigma \cdot \alpha$$

Uncertainty
Pixel-wise stdev over *n* samples

Uncertainty Estimation



Donike *et al.*, **Trustworthy Super-Resolution of Multispectral Sentinel-2 Imagery With Latent Diffusion**, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

SEN2SR Framework

SEN2SR Framework

```
pip install sen2sr
```

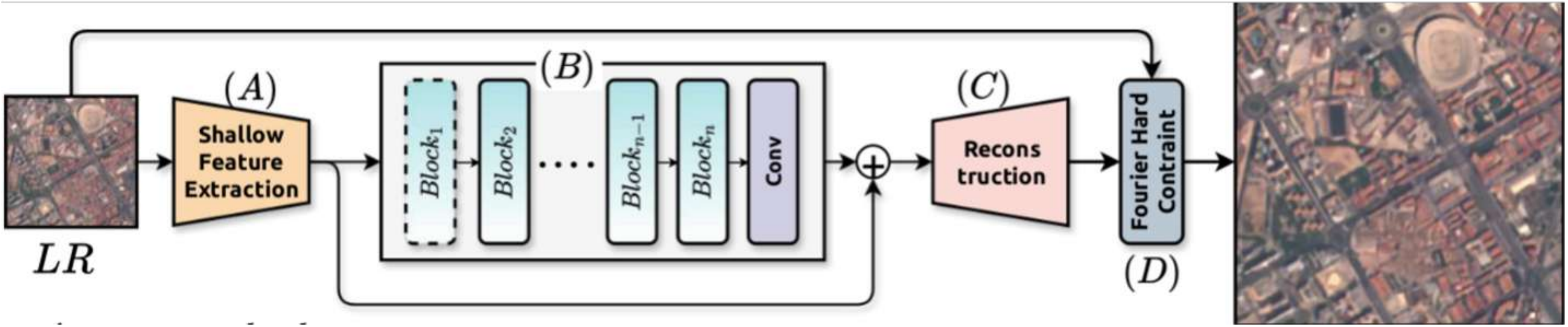
```
pip install opensr-model
```

Code, Publications, Colab Notebooks…

opensr.eu

SEN2SR Super-Resolution Framework

A radiometrically and spatially consistent super-resolution framework for Sentinel-2

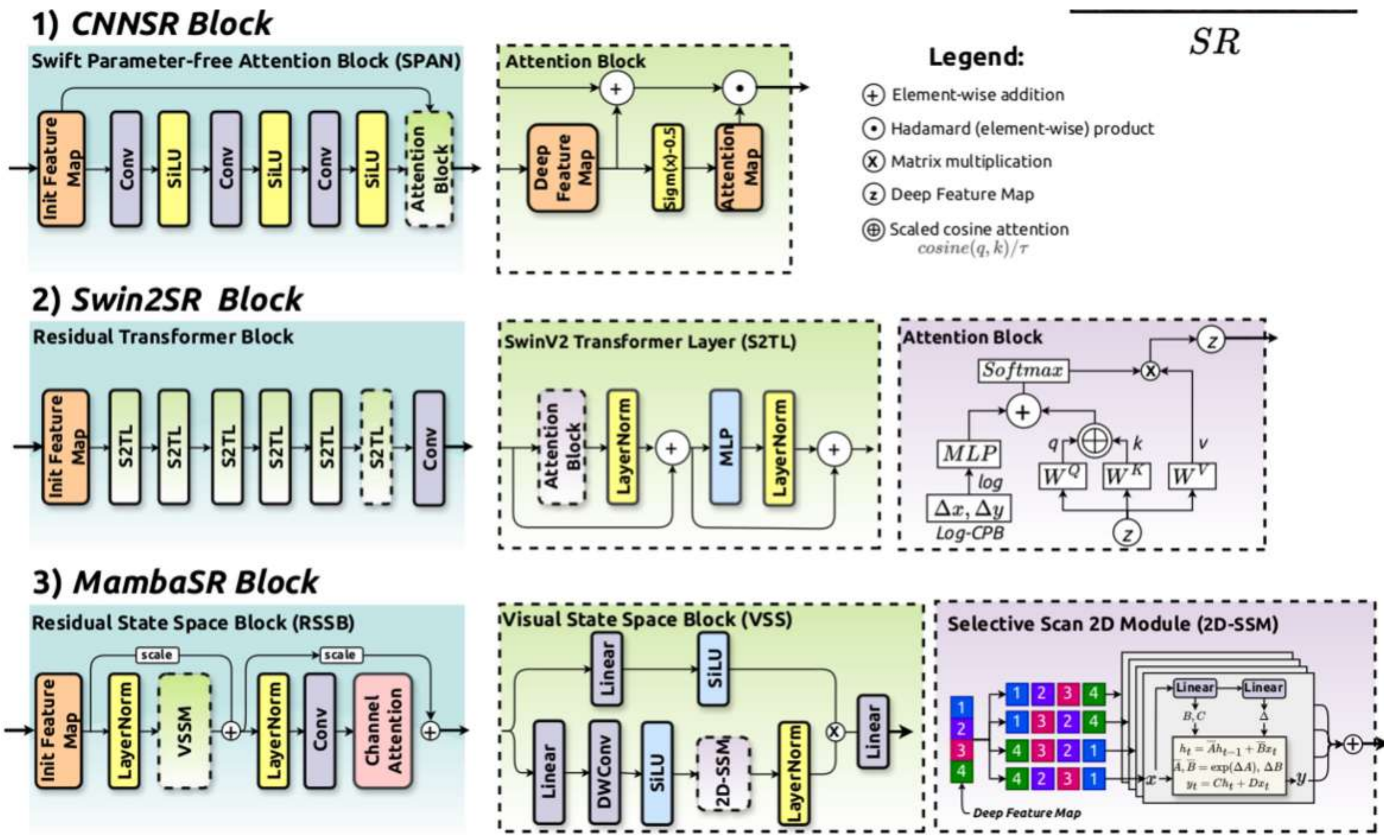


SEN2SR: Baseline Super-Resolution Architectures

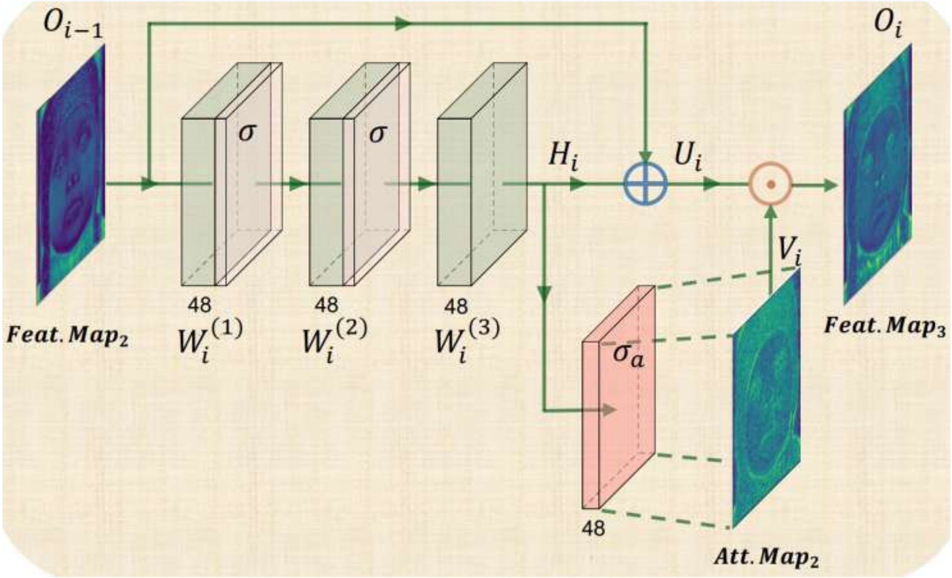
Models:
- CNNSR
- SwinSR
- MambaSR

Model size:
- Real-time (~1M)
- small (~5M)
- medium (~15M)
- large (~30M)

Losses:
- L1
- LPIPS
- CLIP
- Adversarial loss
- Combination



Trick: HF attention blocks

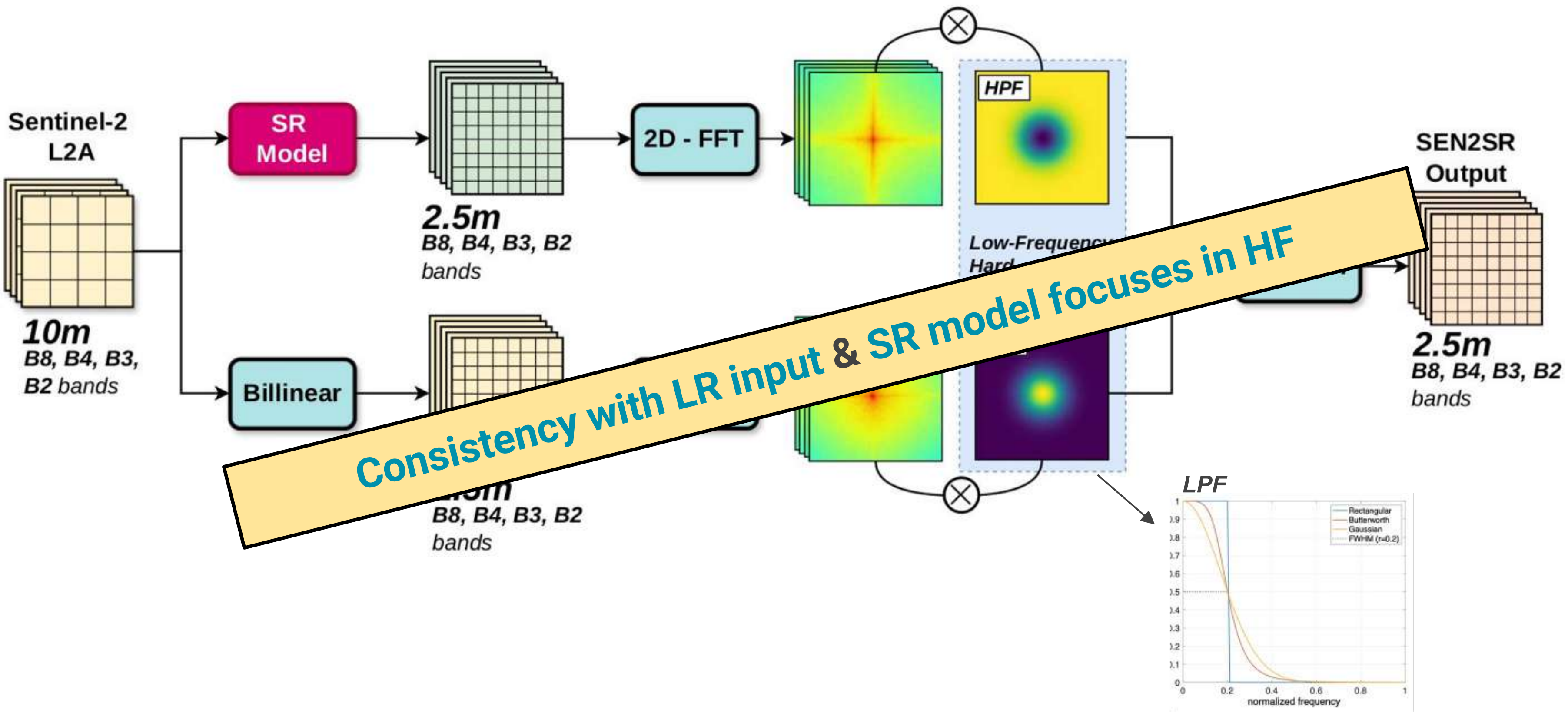


$\sigma_a(x)$	Learnable	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
Sigmoid(x)	-	28.62/0.7826	27.59/0.7366	26.08/0.7854
Sigmoid$(x) - 0.5$	-	28.63/0.7825	27.60/0.7368	26.10/0.7856
Sigmoid$(ax) - 0.5$	✓	28.62/0.7825	27.60/0.7368	26.12/0.7861
$b \times$ (Sigmoid$(ax) - 0.5$)	✓	28.62/0.7826	27.61/0.7367	26.11/0.7860

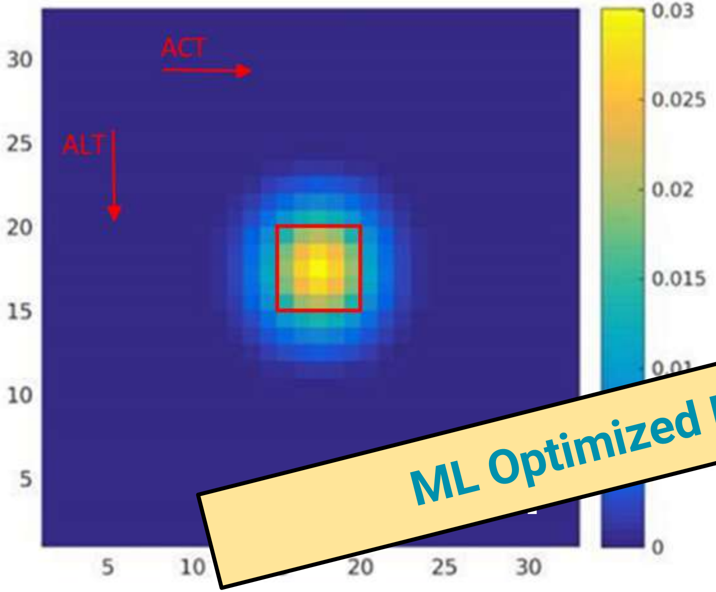
$$\begin{aligned}\frac{\partial L}{\partial W_i} &= \Pi \frac{\partial F^{(i)}_{W_i}(O_{i-1})}{\partial W_i} \\ &= \Pi \frac{\partial}{\partial W_i}(F^{(i)}_{c,W_i}(O_{i-1}) \odot \sigma_a(F^{(i)}_{c,W_i}(O_{i-1}))) \qquad (4) \\ &= \Pi \frac{\partial F^{(i)}_{c,W_i}(O_{i-1})}{\partial W_i} \odot (H_i \odot \sigma'_a(H_i) + \sigma_a(H_i)).\end{aligned}$$

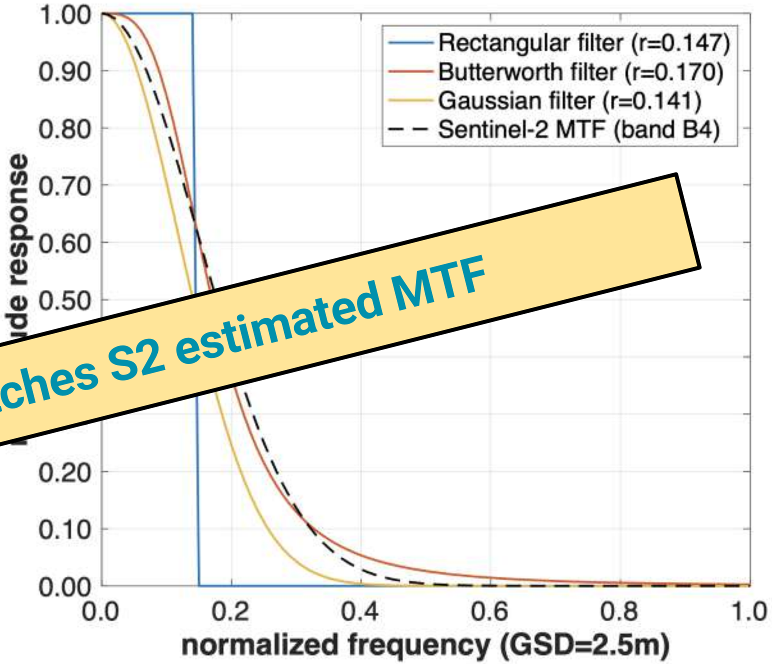
<u>Swift Parameter-free Attention</u> *Network for Efficient Super-Resolution* https://arxiv.org/pdf/2311.12770

Trick: Fourier-based Low-Frequency Hard Constraint



Trick: Fourier-based Low-Frequency Hard Constraint





ML Optimized LPF matches S2 estimated MTF

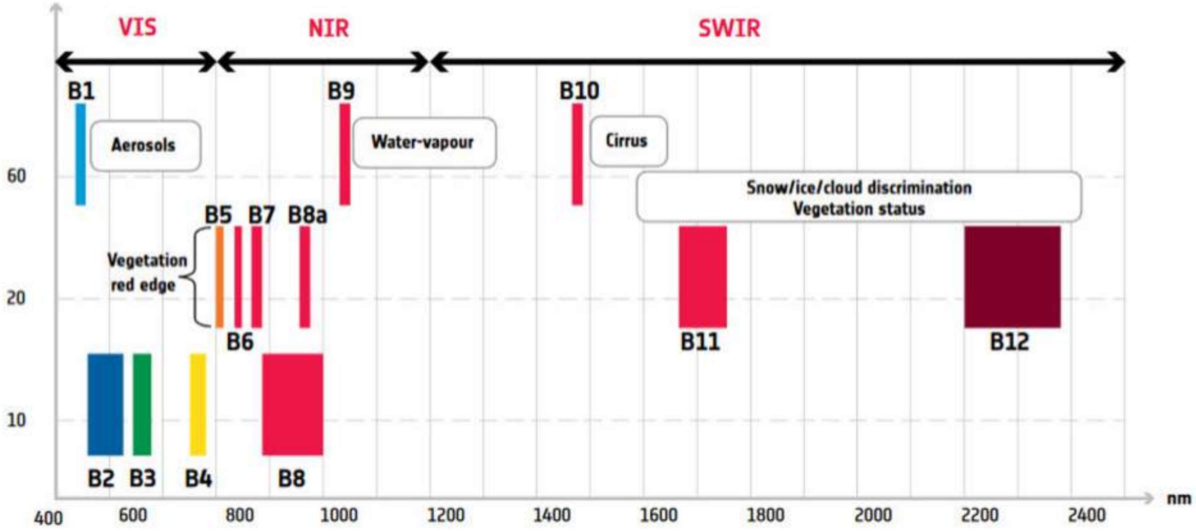
Multiband S2 SR

Sentinel-2 SR products

Objective: Super-resolve any 10m and 20m band to 2.5 meters

Problem: No Reference (HR) for S2 20m bands

- Super-resolve any 10m band to the VHR of 2.5 meters
- Super-resolve 20m to 10 meters and then to 2.5 meters



SEN2SR

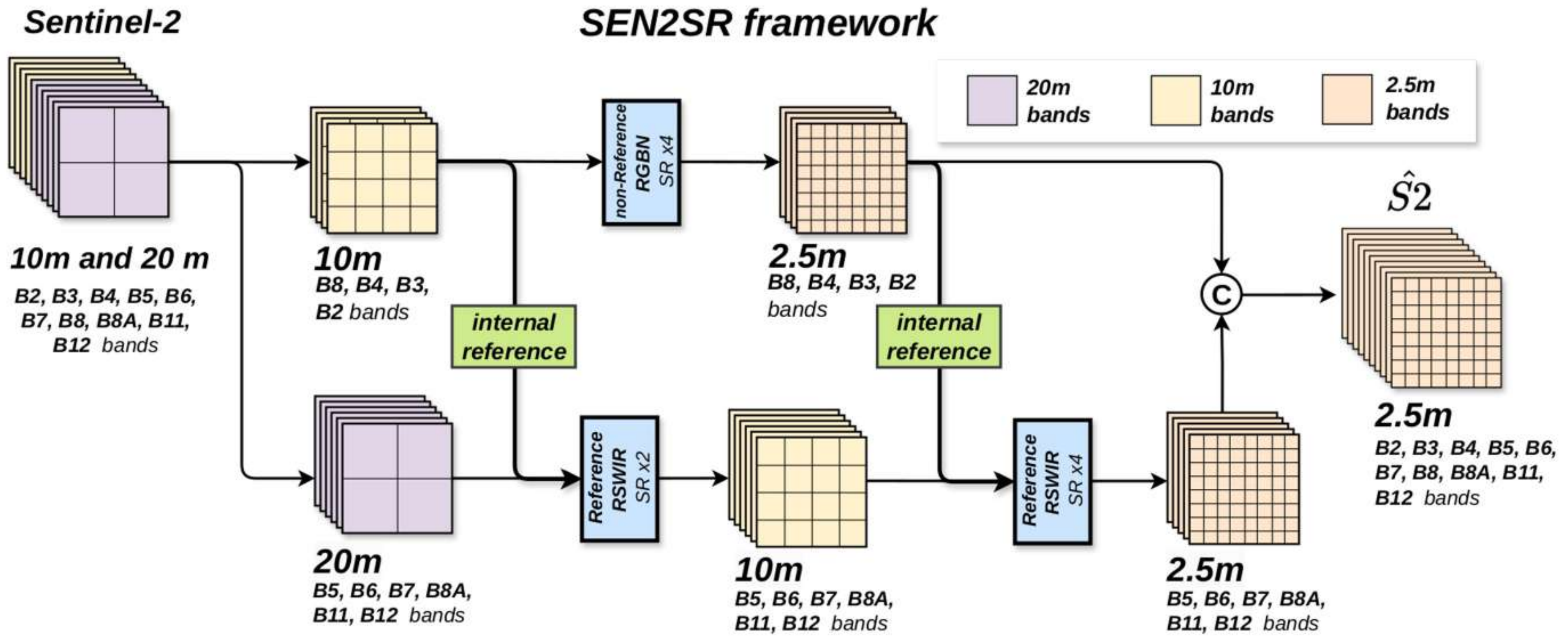
A radiometrically and spatially consistent super-resolution framework for Sentinel-2

📄 Paper: https://lnkd.in/dh8UQHYN

📦 Dataset: https://lnkd.in/dSCbNXQ8

💻 Code: https://lnkd.in/dz6aBCuZ

🧪 Colab demo: https://lnkd.in/djFekKkP



Multi-band SEN2SR Fusion

RGB
10 meters vs 2.5 meters

SR

LR

SR model
(1M model)

Multi-band SEN2SR Fusion

SWIR-Red1-Red
20 meters vs 2.5 meters

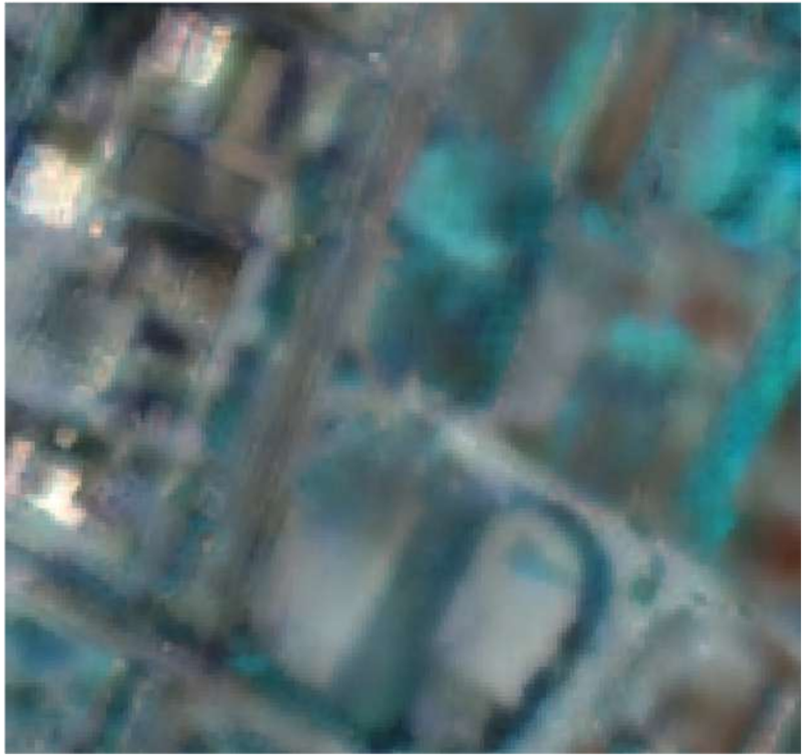
SR

LR

SR model + Fusion Framework
(1M model)

Multi-band SEN2SR Fusion

S2: Bands 9,7,5
20m



SR: Bands 9,7,5
2.5m

SEN2SR

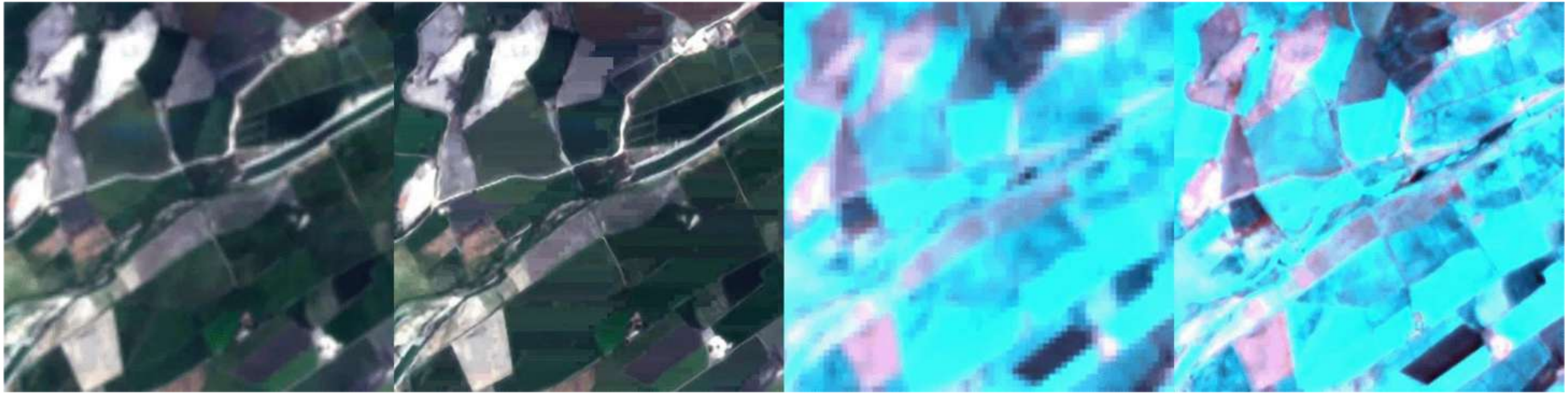
A radiometrically and spatially consistent super-resolution framework for Sentinel-2

📄 Paper: https://lnkd.in/dh8UQHYN

📦 Dataset: https://lnkd.in/dSCbNXQ8

💻 Code: https://lnkd.in/dz6aBCuZ

🧪 Colab demo: https://lnkd.in/djFekKkP



OpenSR-test Toolbox: Benchmarking Datasets & Quality Metrics

OpenSR-test: *A comprehensive benchmark for optical remote sensing image super-resolution*

OpenSR test

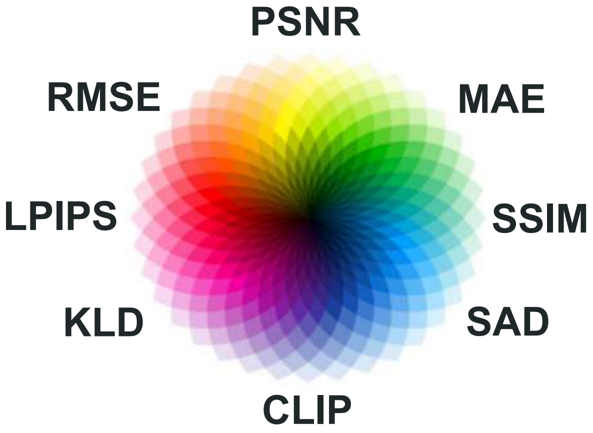
Documentation: https://esaopensr.github.io/opensr-test

GitHub: https://github.com/ESAOpenSR/opensr-test

Datasets: https://huggingface.co/datasets/isp-uv-es/opensr-test

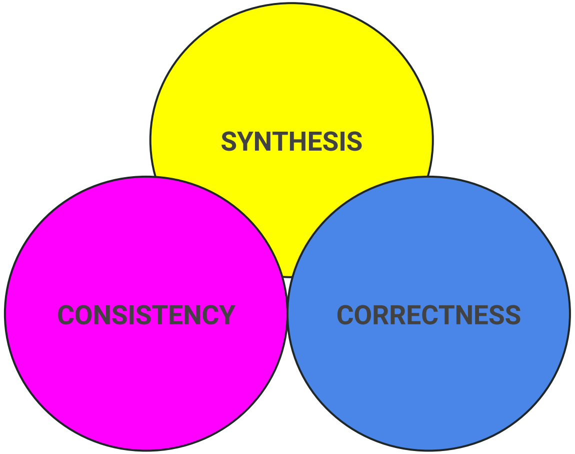
Paper: https://ieeexplore.ieee.org/abstract/document/10530998

Motivation



Traditional

one-size-fits-all



OpenSR-test

specialized for SR

OpenSR-test Benchmark Datasets

Datasets

https://huggingface.co/datasets/isp-uv-es/opensr-test

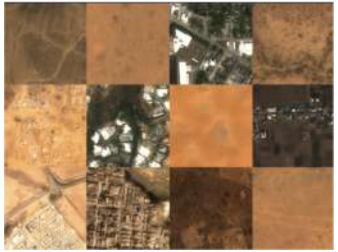
OpenSR test

A comprehensive benchmark for real-world Sentinel-2 imagery super-resolution

pypi v0.2.9 | License MIT | code style black | imports isort

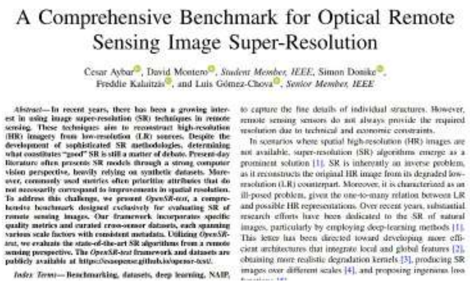
Sentinel-2 cross-sensor datasets:

- **naip:** 62 NAIP images (4x)
- **spot:** 10 SPOT images (4x)
- **venus:** 59 VENµS images (2x)
- **spain_urban:** 20 Ortho images (4x)
- **spain_crops:** 28 Ortho images (4x)



Images Available as a np.darray and GeoTIFF

```
import opensr_test
dataset = opensr_test.load("naip")
lr, hr = dataset["L2A"], dataset["HRharm"]
```



A Comprehensive Benchmark for Optical Remote Sensing Image Super-Resolution

Cesar Aybar, David Montero, *Student Member, IEEE*, Simon Donike, Freddie Kalaitzis, and Luis Gómez-Chova, *Senior Member, IEEE*

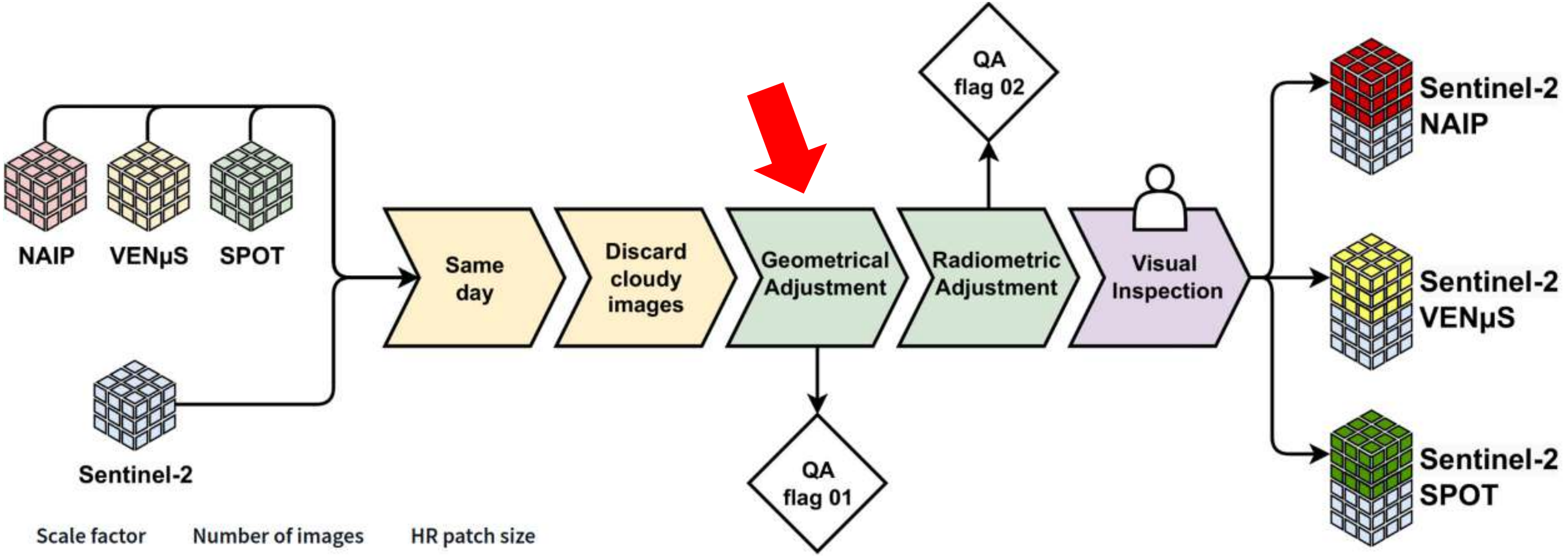
Abstract—In recent years, there has been a growing interest in using image super-resolution (SR) techniques in remote sensing. These techniques aim to reconstruct high-resolution (HR) imagery from low-resolution (LR) sources. Despite the development of sophisticated SR methodologies, determining what constitutes "good" SR is still a matter of debate. Present-day literature often presents SR models through a strong computer vision perspective, heavily relying on synthetic datasets. Moreover, commonly used metrics often prioritize attributes that do not necessarily correspond to improvements in spatial resolution. To address this challenge, we present *OpenSR-test*, a comprehensive benchmark designed exclusively for evaluating SR of remote sensing images. Our framework incorporates specific quality metrics and curated cross-sensor datasets, each spanning various scale factors with consistent metadata. Utilizing *OpenSR-test*, we evaluate the state-of-the-art SR algorithms from a remote sensing perspective. The *OpenSR-test* framework and datasets are publicly available at https://esaopensr.github.io/opensr-test/.

Index Terms—Benchmarking, datasets, deep learning, NAIP,

to capture the fine details of individual structures. However, remote sensing sensors do not always provide the required resolution due to technical and economic constraints.

In scenarios where spatial high-resolution (HR) images are not available, super-resolution (SR) algorithms emerge as a prominent solution [1]. SR is inherently an inverse problem, as it reconstructs the original HR image from its degraded low-resolution (LR) counterpart. Moreover, it is characterized as an ill-posed problem, given the one-to-many relation between LR and possible HR representations. Over recent years, substantial research efforts have been dedicated to the SR of natural images, particularly by employing deep-learning methods [1]. This letter has been directed toward developing more efficient architectures that integrate local and global features [2], obtaining more realistic degradation kernels [3], producing SR images over different scales [4], and proposing ingenious loss

Datasets - Geometrical Adjustment

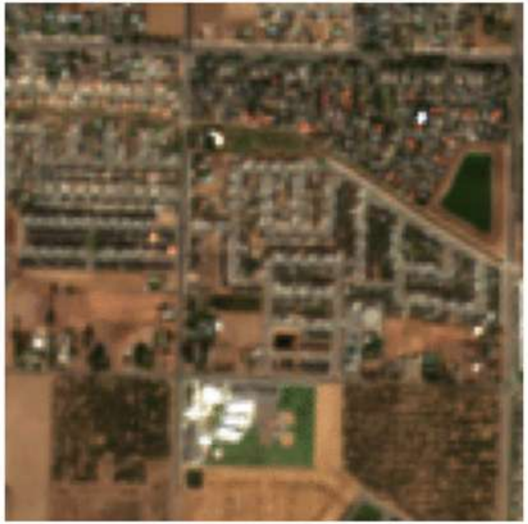


Dataset	Scale factor	Number of images	HR patch size
NAIP	x4	62	484x484
SPOT	x4	9	512x512
Venμs	x2	59	256x256
SPAIN CROPS	x4	28	512x512
SPAIN URBAN	x4	20	512x512

Datasets - Geometrical Adjustment

Original S2 Cube

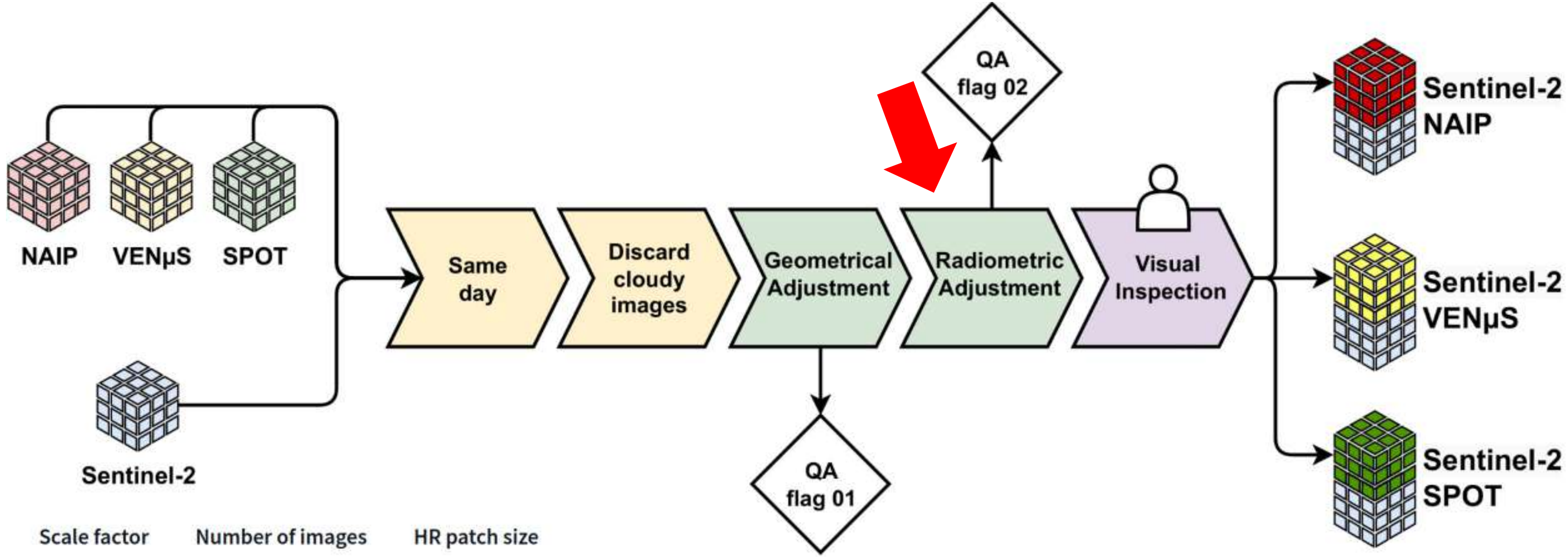
Aligned S2 Cube Date: 20151005



Three different methods to spatial align images!

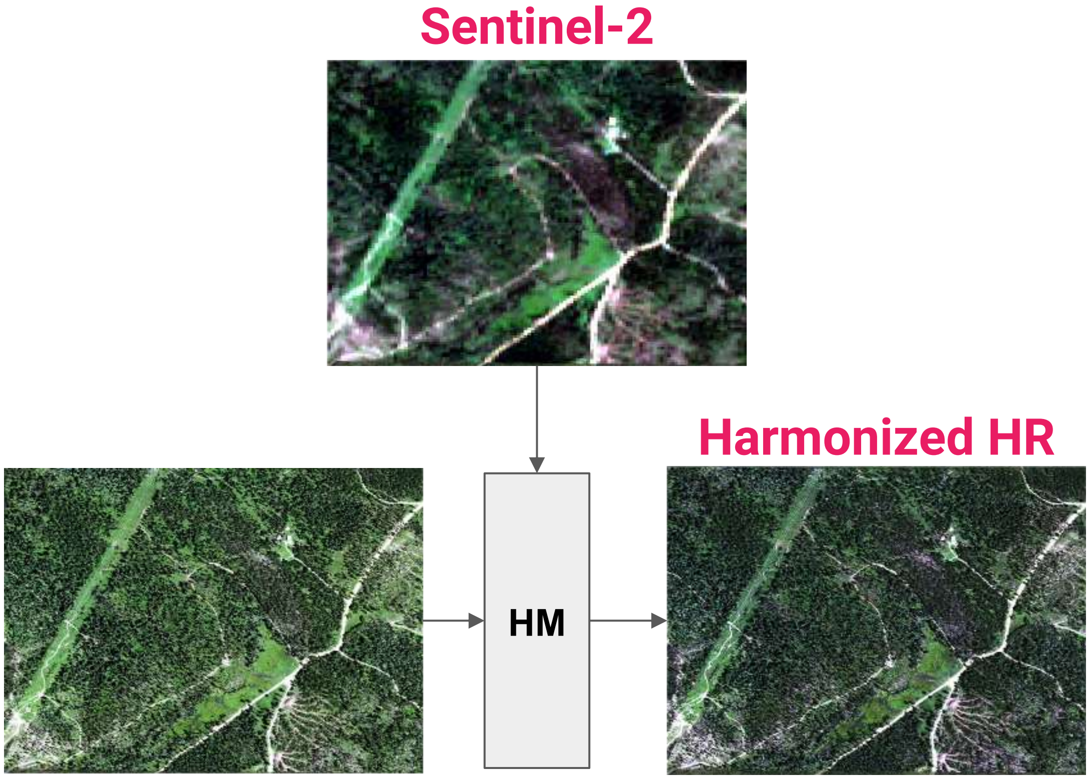
- **Enhanced Correlation Coefficient (ECC) Maximization**
- **Phase Cross Correlation (PCC)**
- **LighGlue + [DISK or SuperPoint or Xfeat]**

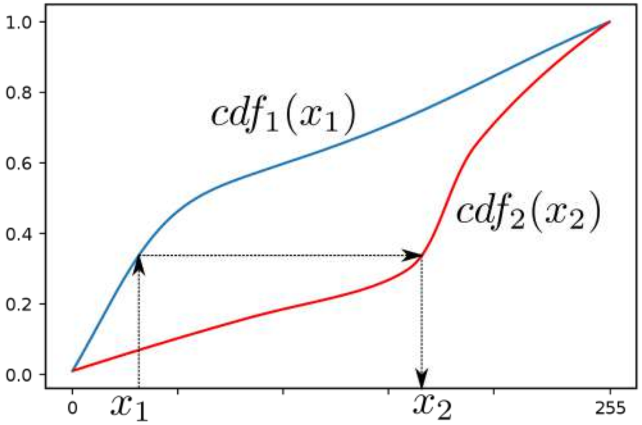
Datasets - Radiometrical Adjustment



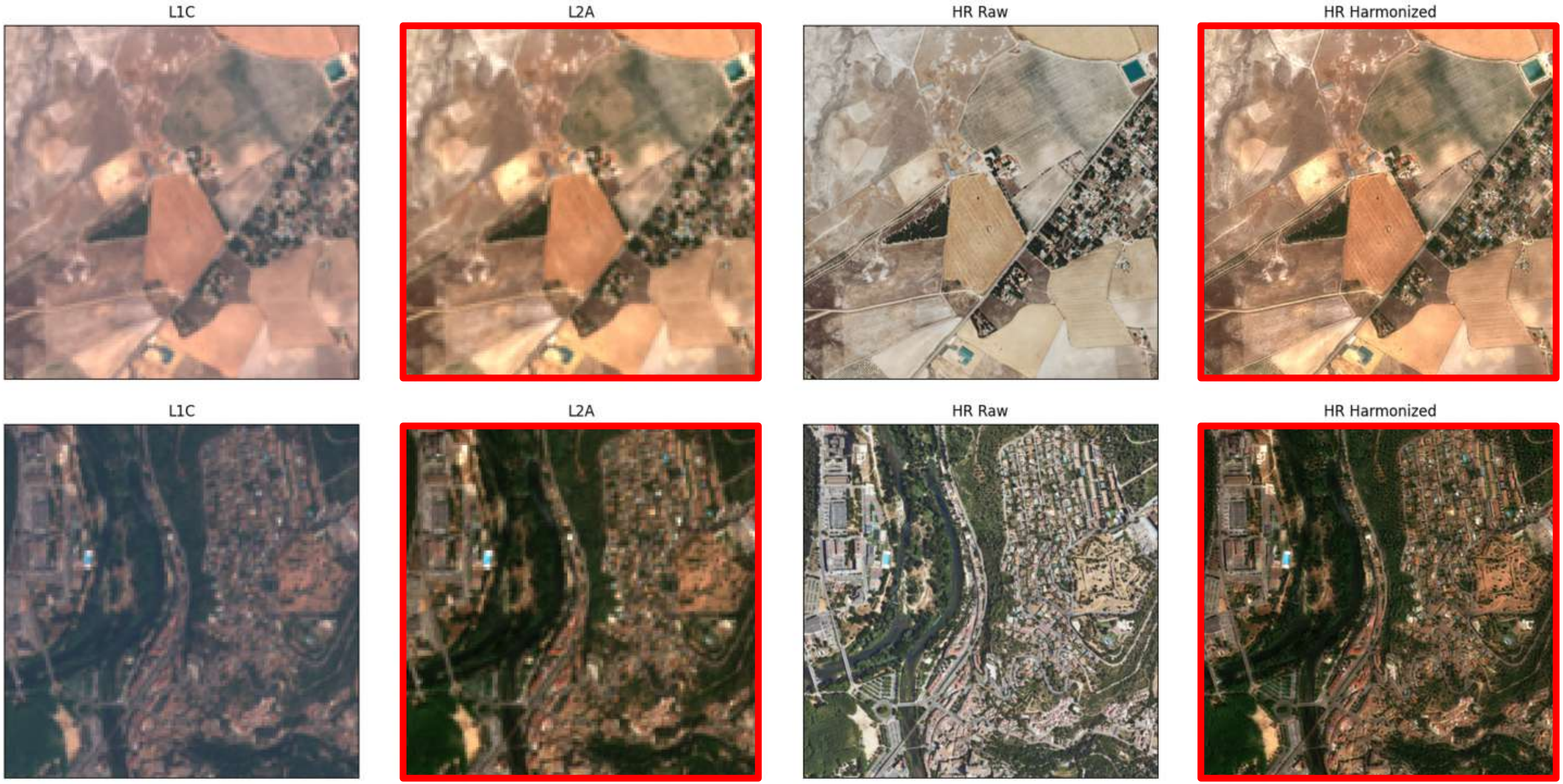
Dataset	Scale factor	Number of images	HR patch size
NAIP	x4	62	484x484
SPOT	x4	9	512x512
Venμs	x2	59	256x256
SPAIN CROPS	x4	28	512x512
SPAIN URBAN	x4	20	512x512

Datasets - Radiometrical Adjustment

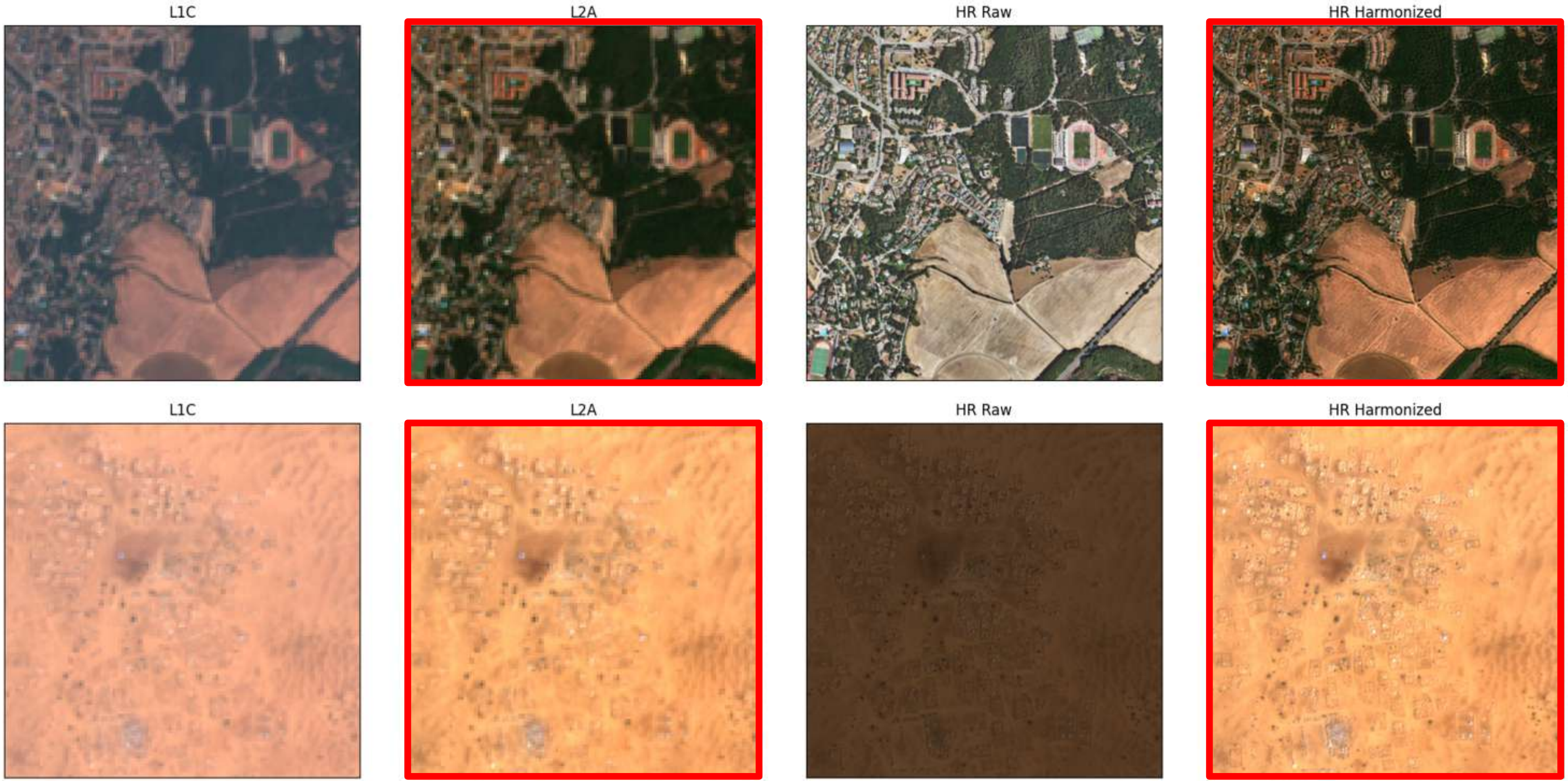




Datasets



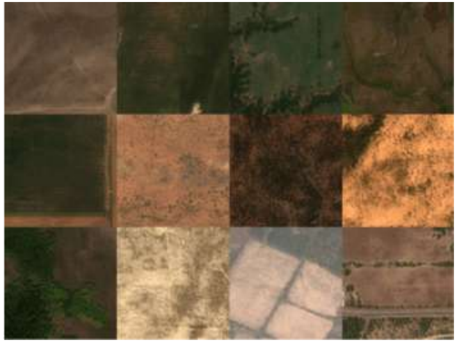
Datasets



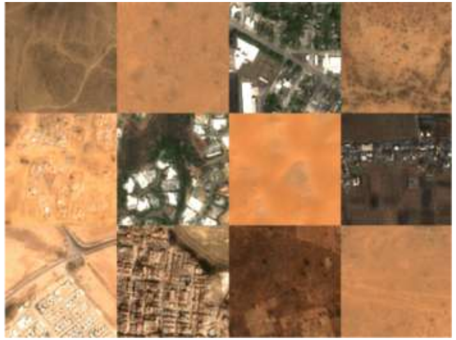
OpenSR-test datasets

https://huggingface.co/datasets/isp-uv-es/opensr-test

naip



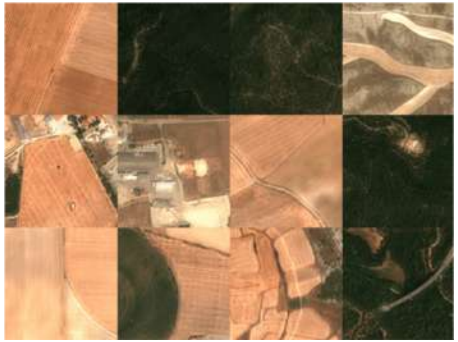
spot



venus

Dataset	Scale factor	Number of images	HR patch size
NAIP	x4	62	484x484
SPOT	x4	9	512x512
Venµs	x2	59	256x256
SPAIN CROPS	x4	28	512x512
SPAIN URBAN	x4	20	512x512

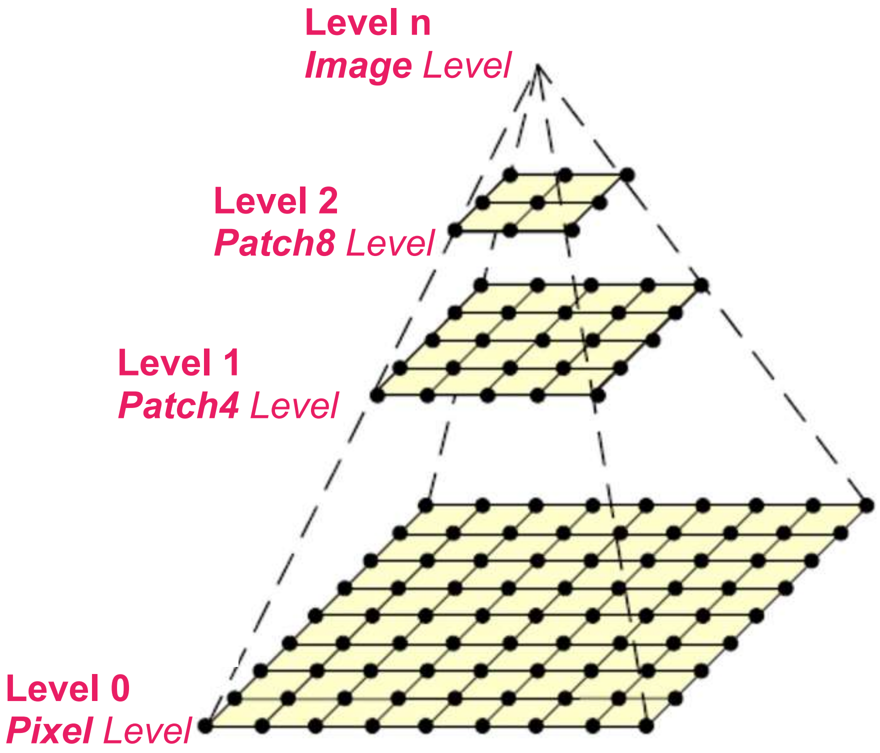
spain_urban



spain_crops

OpenSR-test Validation Metrics

Metrics



All the metrics are designed to be estimated at different levels: ***pixel, patch*** **or** ***image***

The metrics implemented are ***distance-independent***, allowing users to specify the "distance" most appropriate for their needs: L1, L2, SAD, PSNR, KLD, SSIM, LPIPS, CLIP, …

Metrics (some remarks)

Trade-off between quality metrics

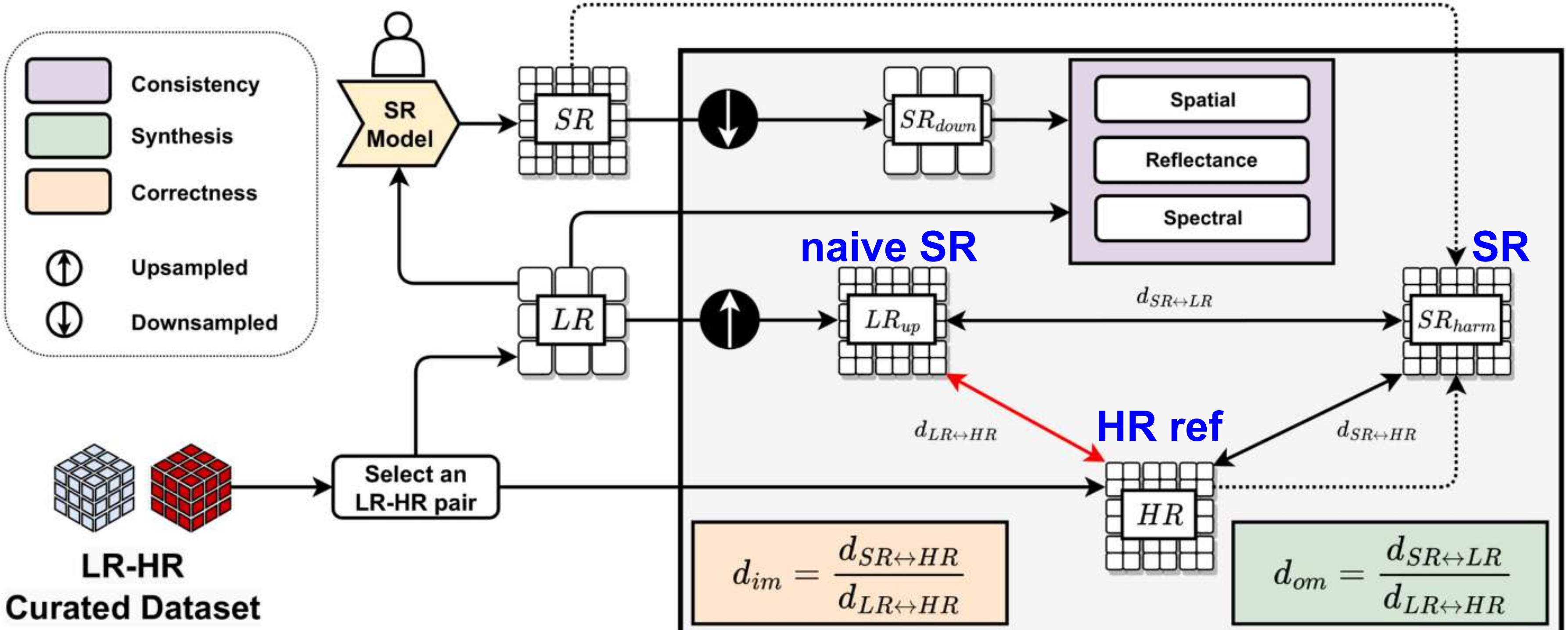
- Image fidelity: reflectance values (L1/SAD)
- Perceptual quality: visual inspection (LPIPS)
- Semantic information: consistent land covers (CLIP)

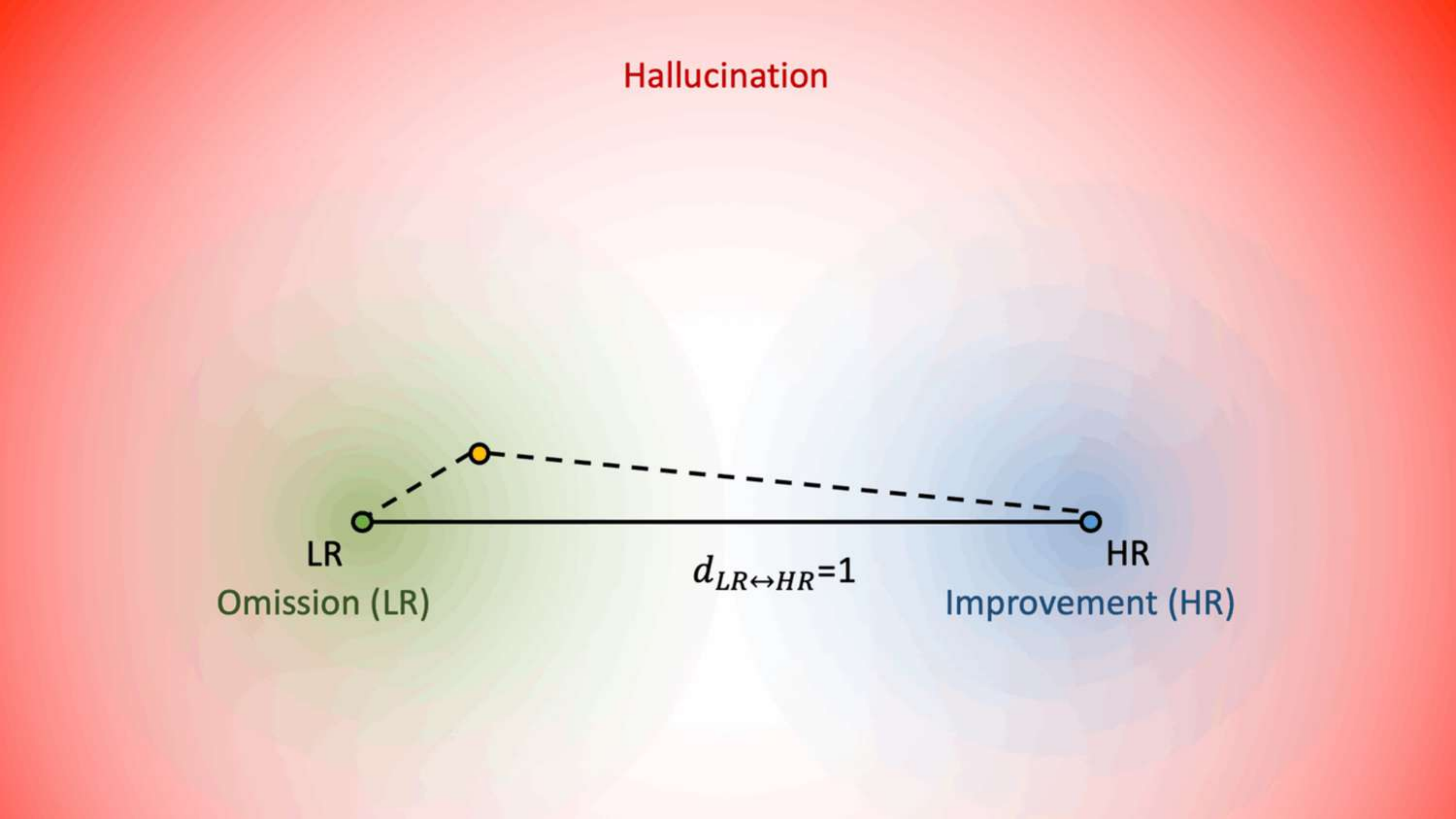
Useful for different image-to-image problems

- Pansharpening
- Image Fusion
- Image filling
- …

Metrics

IMPROVEMENT & OMISSION DISTANCE




Hallucination
LR
HR
$d_{LR \leftrightarrow HR}$=1
Omission (LR)
Improvement (HR)

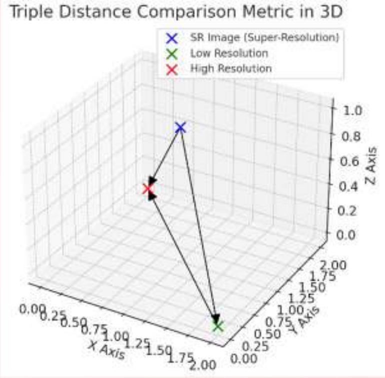
Triangle Inequality:

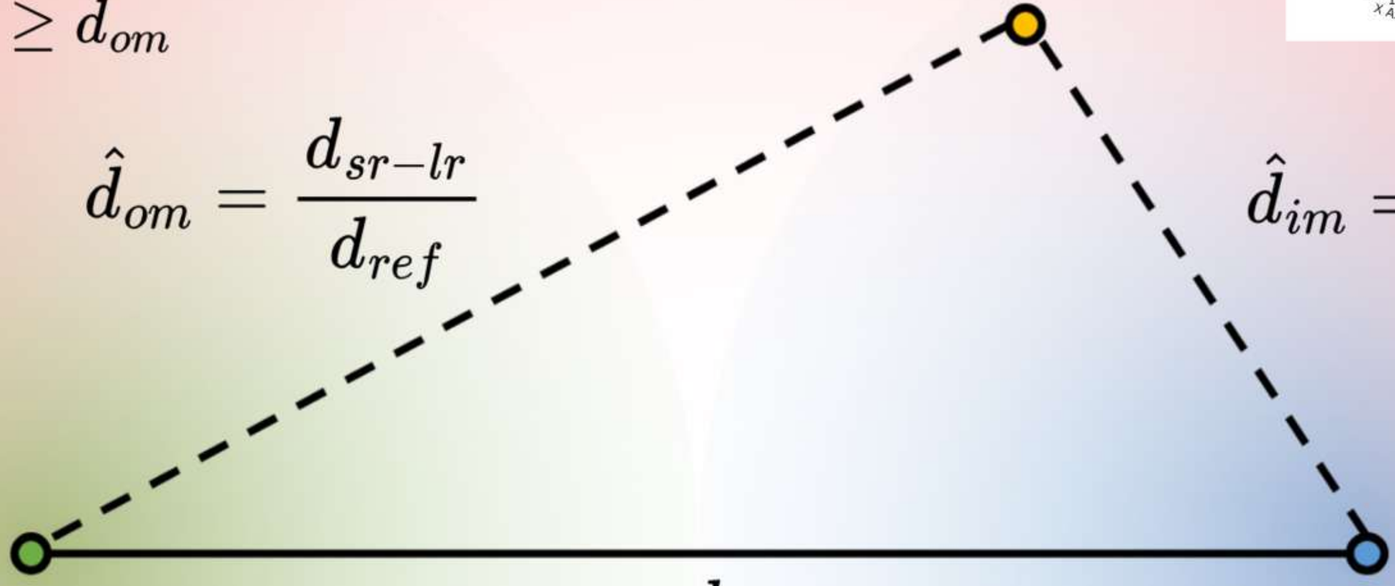
Hallucination

$$\hat{d}_{om} + \hat{d}_{im} \geq 1$$

$$\hat{d}_{om} + 1 \geq \hat{d}_{im}$$

$$\hat{d}_{im} + 1 \geq \hat{d}_{om}$$





$$\hat{d}_{om} = \frac{d_{sr-lr}}{d_{ref}}$$

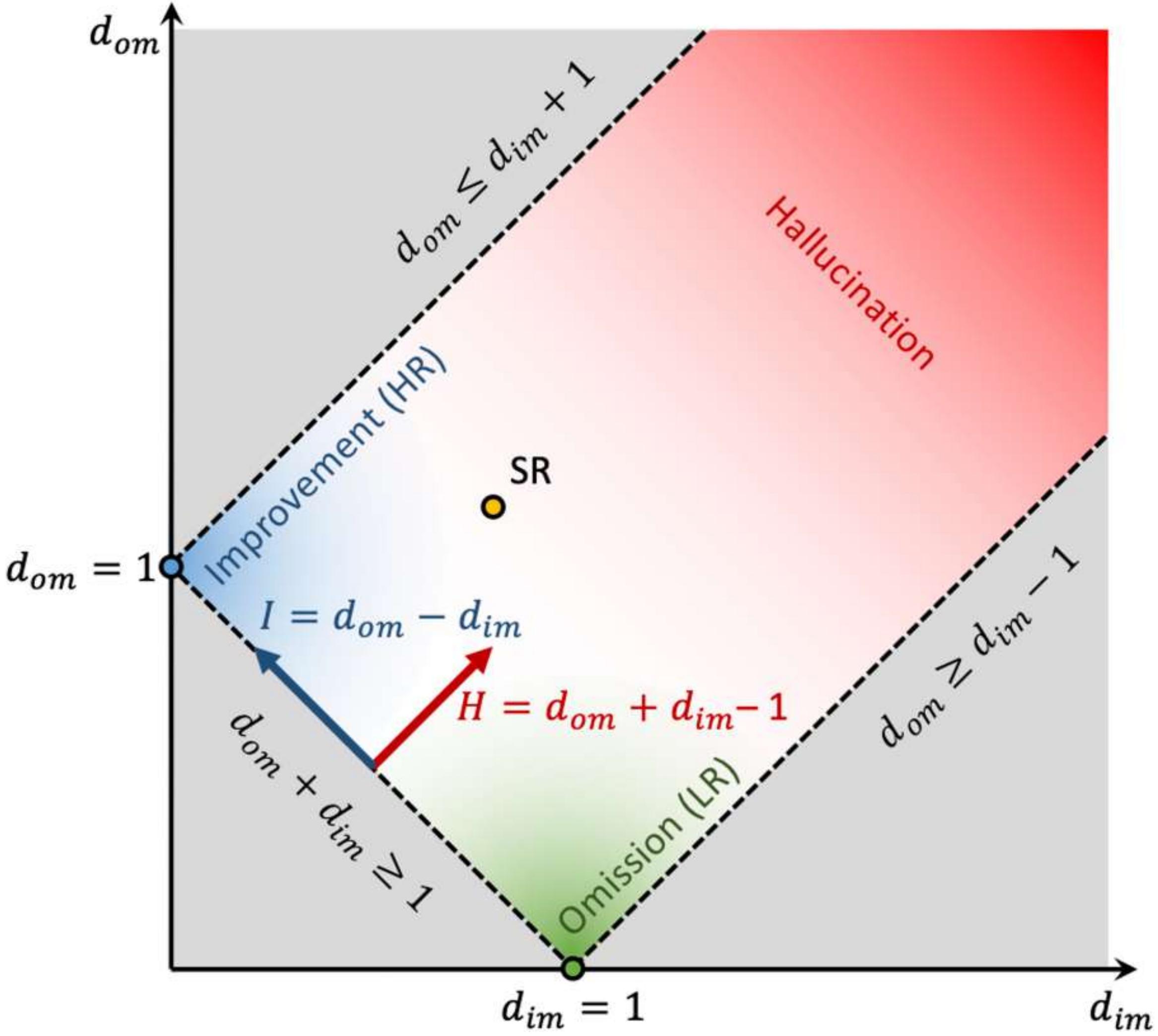
$$\hat{d}_{im} = \frac{d_{sr-hr}}{d_{ref}}$$

$$\hat{d}_{ref} = \frac{d_{lr-hr}}{d_{lr-hr}} = 1$$

Omission (LR)

Improvement (HR)

Metrics

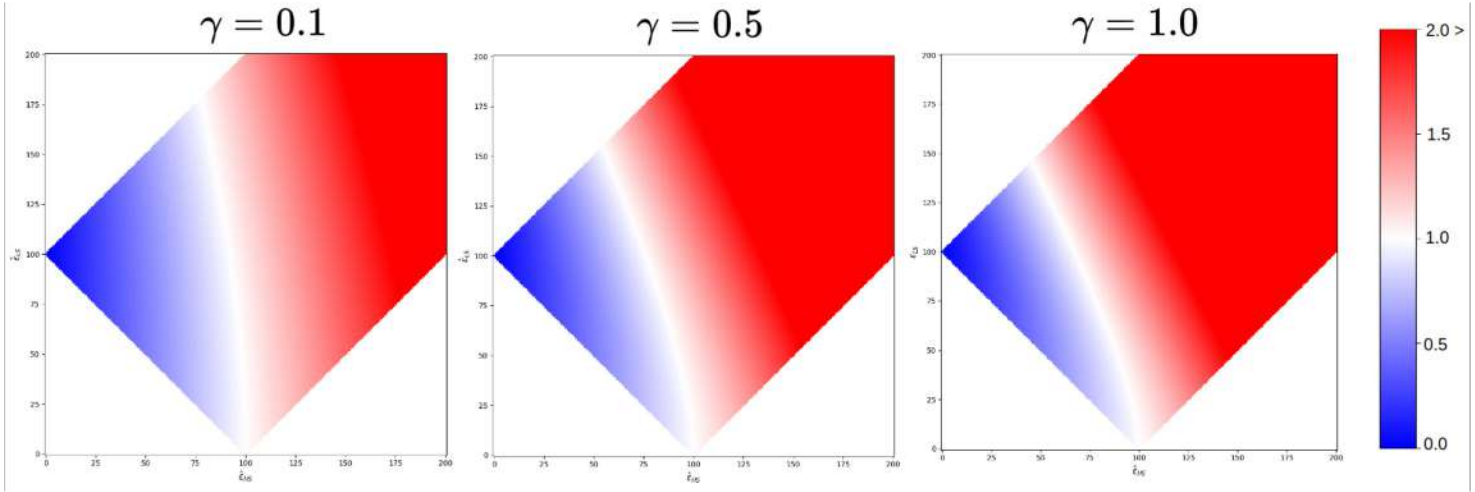


d_{om}
$d_{om} \leq d_{im} + 1$
Hallucination
Improvement (HR)
SR
$d_{om} = 1$
$I = d_{om} - d_{im}$
$H = d_{om} + d_{im} - 1$
$d_{om} \geq d_{im} - 1$
$d_{om} + d_{im} \geq 1$
Omission (LR)
$d_{im} = 1$
d_{im}

Metrics

$$H = d_{im} + d_{om} - 1$$

$$im = d_{im} + d_{om}(1 - e^{\boxed{-\gamma_{im}}H})$$

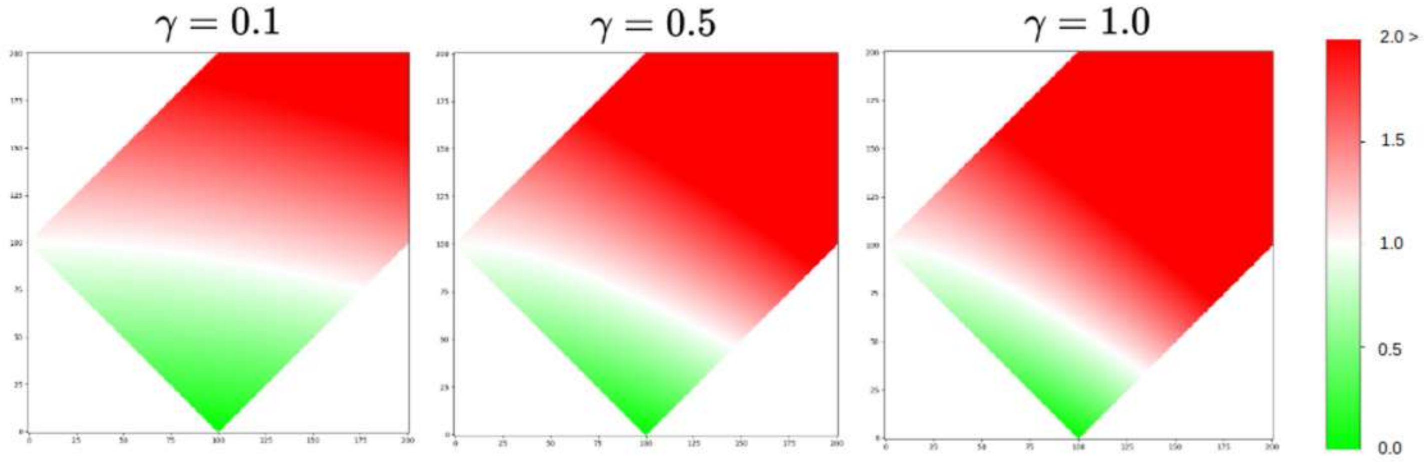


Distance to the improvement space

Metrics

$$H = d_{im} + d_{om} - 1$$

$$om = d_{om} + d_{im}(1 - e^{\boxed{-\gamma_{om}}H})$$

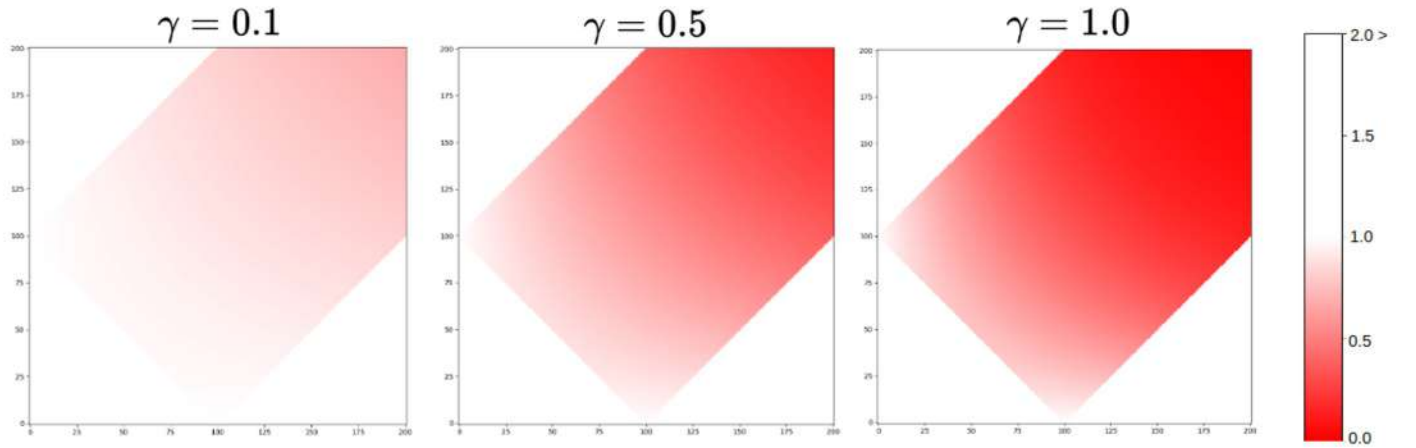


Distance to the omission space

Metrics

$$H = d_{im} + d_{om} - 1$$

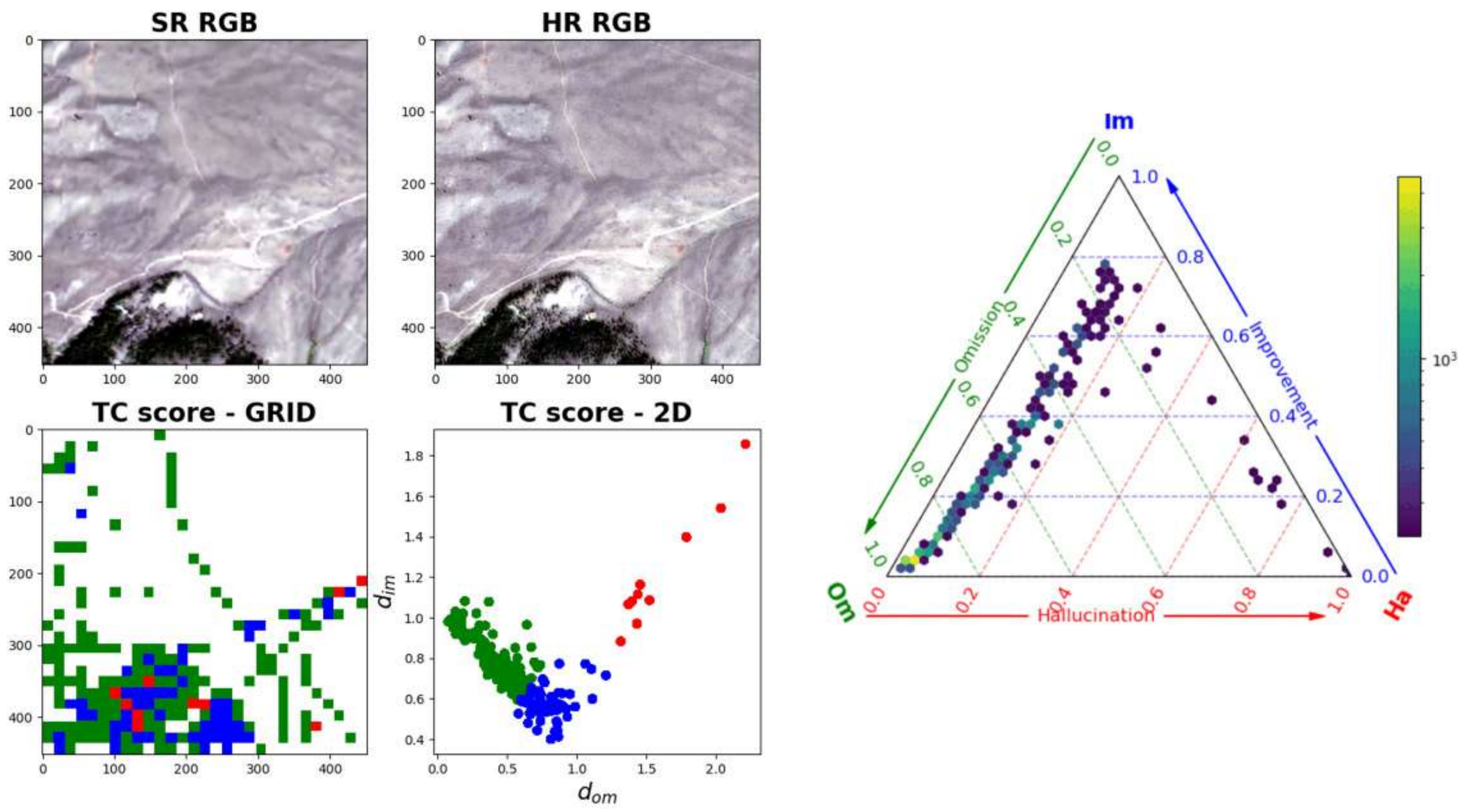
$$ha = e^{\boxed{-\gamma_{ha}} d_{im} d_{om}}$$



Distance to the hallucination space

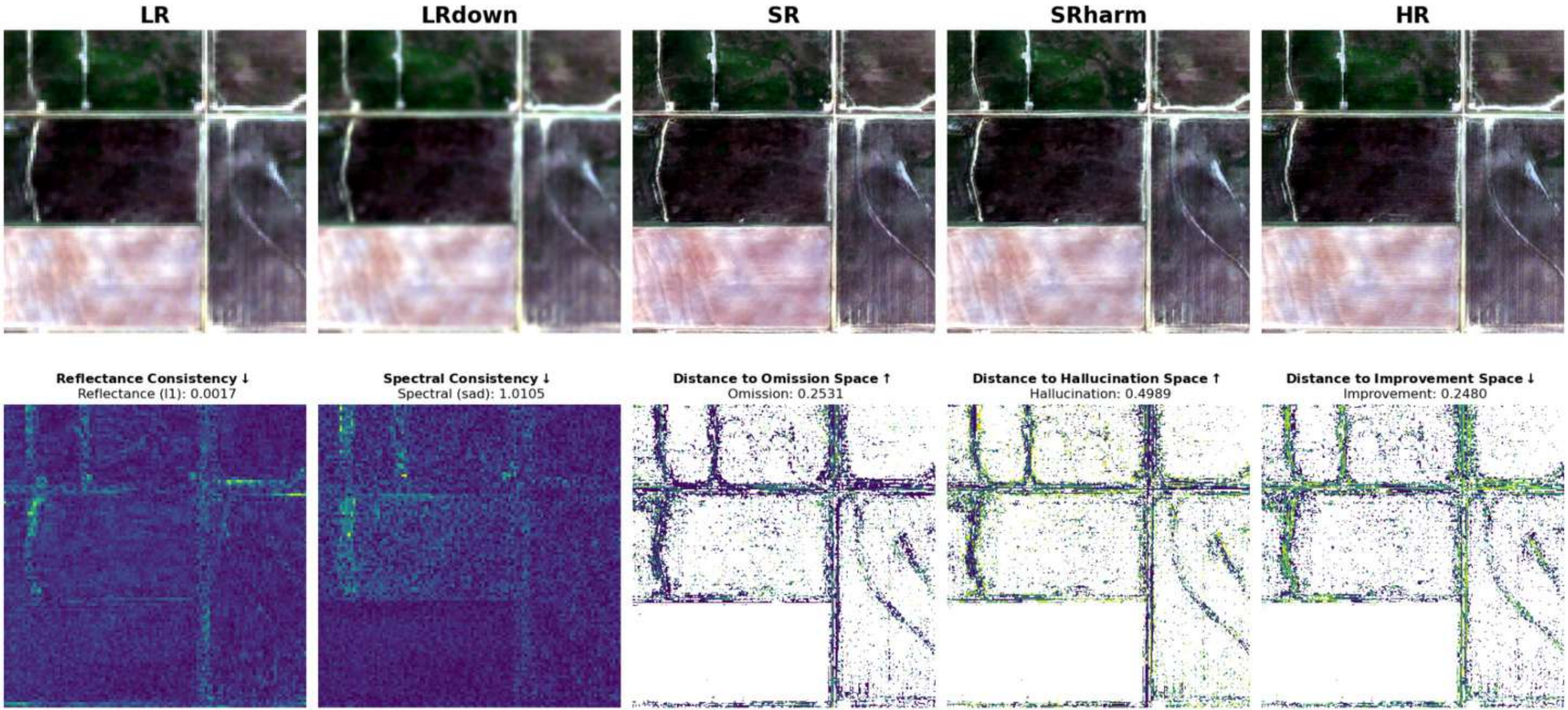
Metrics

metrics = opensr_test.Metrics(agg_method="patch", **patch_size=16**, **correctness_distance="clip"**)



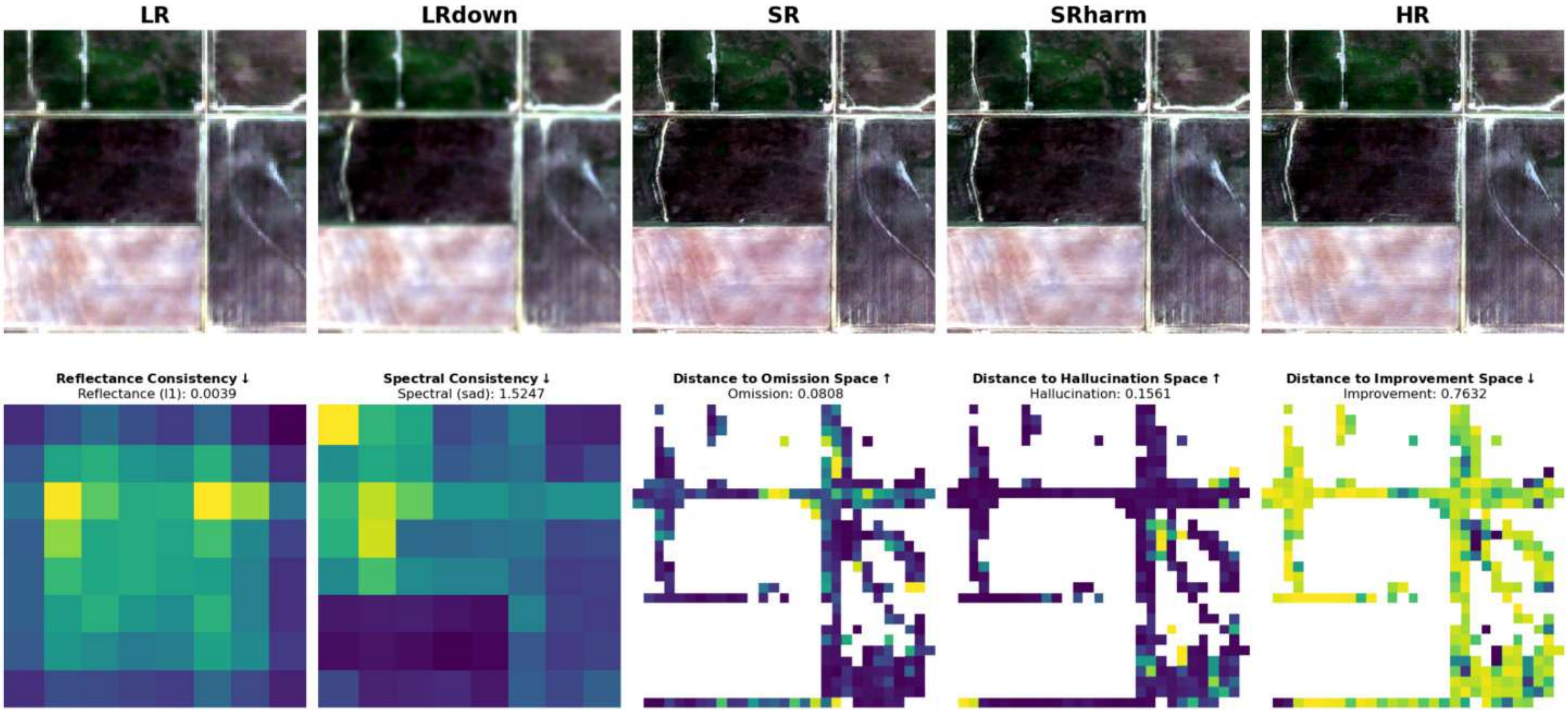
Metrics

metrics = opensr_test.Metrics(agg_method="patch", **patch_size=1**, **correctness_distance="l1"**)



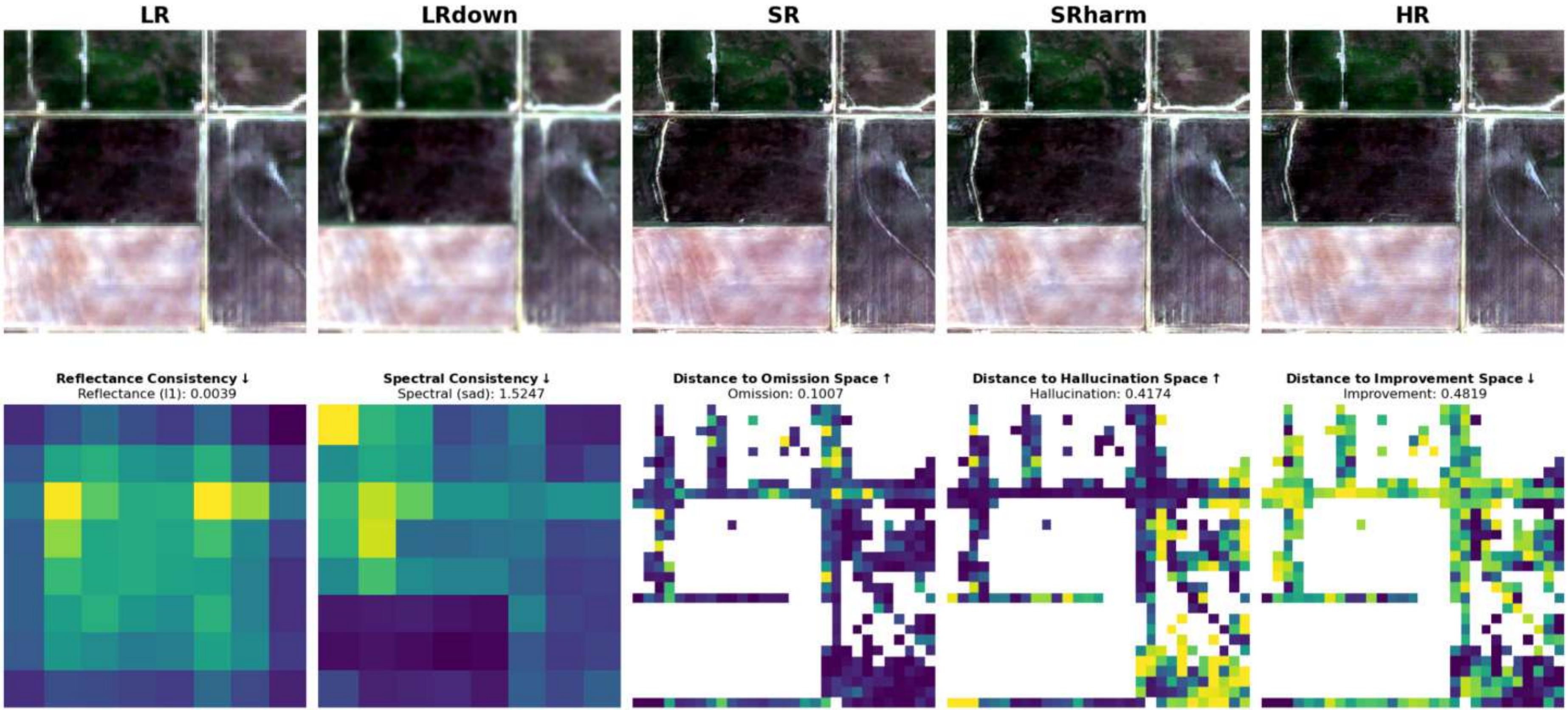
Metrics

metrics = opensr_test.Metrics(agg_method="patch", **patch_size=16**, **correctness_distance="lpips"**)



Metrics

metrics = opensr_test.Metrics(agg_method="patch", **patch_size=16**, **correctness_distance="clip"**)



Metrics

https://github.com/ESAOpenSR/opensr-test

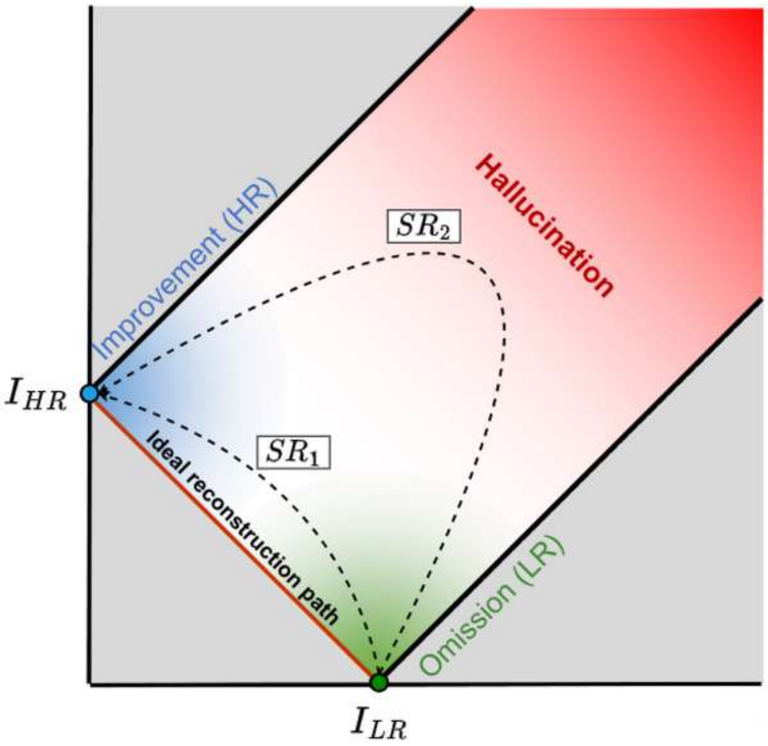
```
import torch
import opensr_test

lr = torch.rand(4, 64, 64)
hr = torch.rand(4, 256, 256)
sr = torch.rand(4, 256, 256)

metrics = opensr_test.Metrics()
metrics.compute(lr=lr, sr=sr, hr=hr)
>>> {'reflectance': 0.253, 'spectral': 26.967, 'spatial': 0.0, 'synthesis': 0.2870, 'ha_percent
```

This tool returns:

- **reflectance:** How SR affects reflectance mean values of LR image.
- **spectral:** How SR affects the spectral signature of the LR image.
- **spatial:** The spatial alignment between the SR and LR images.
- **im_distance:** The mean distance to the improvement space.
- **om_distance:** The mean distance to the omission space.
- **ha_distance:** The mean distance to the hallucination space.



Website

https://esaopensr.github.io/opensr-test/

OpenSR test

Search

OpenSR test
v.0.2.0 ☆6 ⑂0

Home API Metrics

OpenSR test

A comprehensive benchmark for real-world Sentinel-2 imagery super-resolution

pypi v0.9.0 | License MIT | code style black | imports isort

GitHub: https://github.com/ESAOpenSR/opensr-test

Documentation: https://esaopensr.github.io/opensr-test

PyPI: https://pypi.org/project/opensr-test/

Paper: https://www.techrxiv.org/users/760184/articles/735467-a-comprehensive-benchmark-for-optical-remote-sensing-image-super-resolution

Table of contents

SR Model Comparison

Model	PSNR[dB]				CLIP				LPIPS			
	Score ↑	Om% ↓	Im% ↑	Ha% ↓	Score ↓	Om% ↓	Im% ↑	Ha% ↓	Score ↓	Om% ↓	Im% ↑	Ha% ↓
MambaSR *Medium CLIP*	36.67	0.623	0.174	0.203	**0.14**	**0.477**	**0.365**	**0.158**	0.21	0.527	0.423	0.050
MambaSR *Medium LPIPS*	36.68	0.607	0.182	0.211	0.15	0.508	0.339	0.153	**0.17**	**0.398**	**0.556**	**0.046**
LDSR-S2 *Diffusion (U-Net)*	35.92	0.458	0.204	0.338	**0.14**	**0.491**	**0.348**	**0.161**	0.19	0.441	0.503	0.056
SatlasSR *ESRGAN*	21.05	0.117	0.102	0.780	0.16	0.205	0.328	0.467	0.20	0.074	0.619	0.308
SR4RS *ESRGAN*	25.32	0.183	0.120	0.697	0.15	0.237	0.418	0.346	**0.13**	**0.094**	**0.821**	**0.085**
Swin2-MoSE *Swin2*	30.56	0.509	0.133	0.358	0.18	0.734	0.128	0.138	0.37	0.929	0.045	0.026
SuperImage *HAN*	35.95	0.638	0.170	0.192	0.17	0.694	0.184	0.122	0.30	0.831	0.132	0.036
LDM *Diffusion (U-Net)*	24.59	0.272	0.073	0.656	0.20	0.148	0.163	0.689	0.34	0.156	0.266	0.578
SEN2SR-Lite *CNN Light* $\mathcal{L}_1$	**36.97**	**0.663**	**0.162**	**0.175**	0.17	0.742	0.162	0.095	0.30	0.830	0.132	0.038
SEN2SR *Mamba Medium* $\mathcal{L}_1$	**37.01**	**0.638**	**0.179**	**0.183**	0.16	0.661	0.217	0.122	0.26	0.721	0.236	0.042

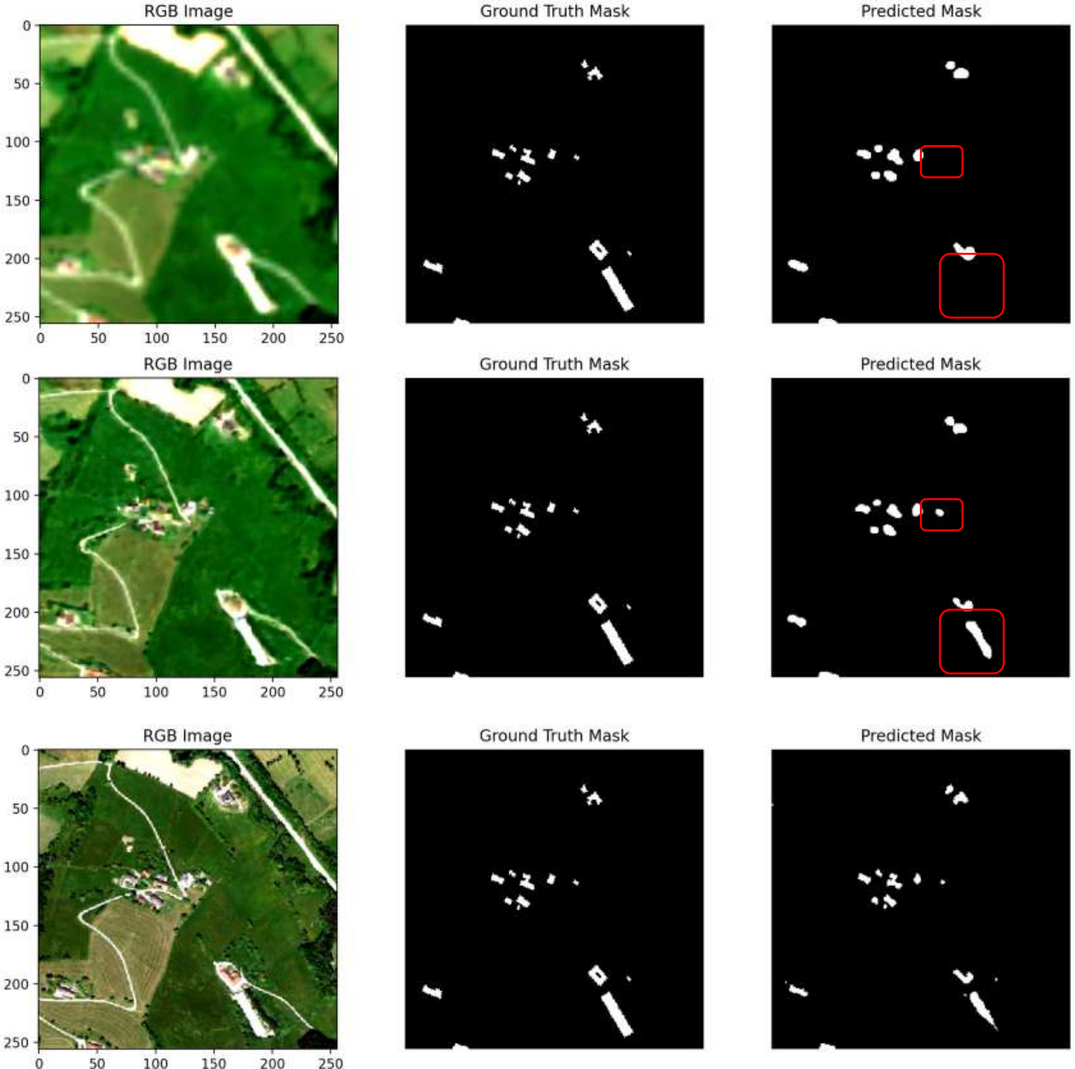
Use-Case:
Building Delineation

Use Case:
Building Detection

LR (Int.)

SR

HR

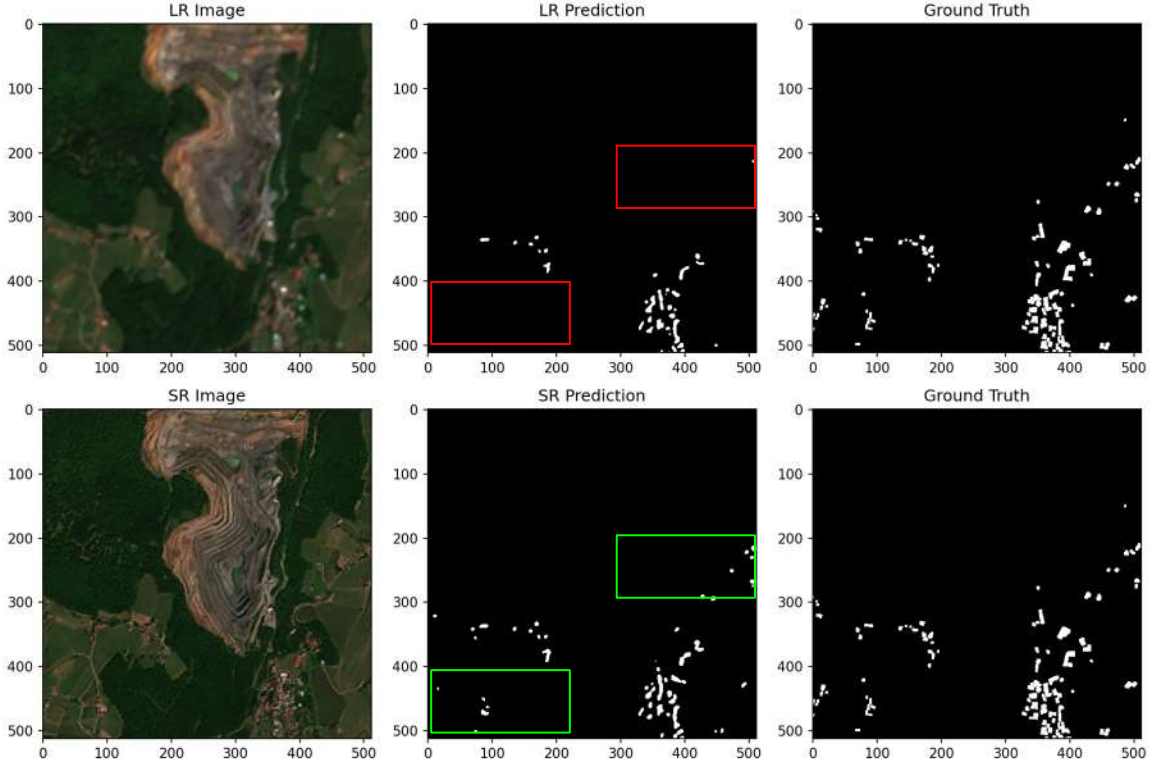


Use Case: Building Detection

```
Size Category XXS :     0 to 4 pixels.
Size Category XS :      4 to 8 pixels.
Size Category S :       8 to 12 pixels.
Size Category M :       12 to 20 pixels.
Size Category L :       20 to 50 pixels.
Size Category XL :      50 to 999 pixels.
```

Avg Object True Positive %

	LR	SR	HR
S	10.3%	14.7%	15.5%
L	42.0%	46.5%	54.6%



Use-Case:
Flood Detection

Use Case: Flood Detection

- 29.10.2024: Flash Flood
- 31.10.2024: Sen2 Acquisition

WikiMedia user "PacoPac"

Use Case: Flood Detection

October 29th, 2024

Before Flood

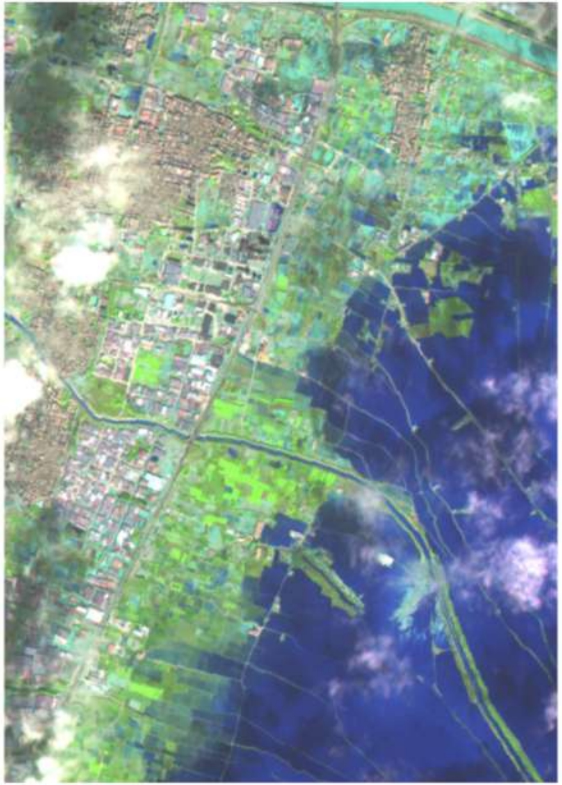
After Flood

https://isp.uv.es/ml4floods-isp/

Use Case: Flood Detection

Use Case: Flood Detection

LDSR-S2



SEN2SR

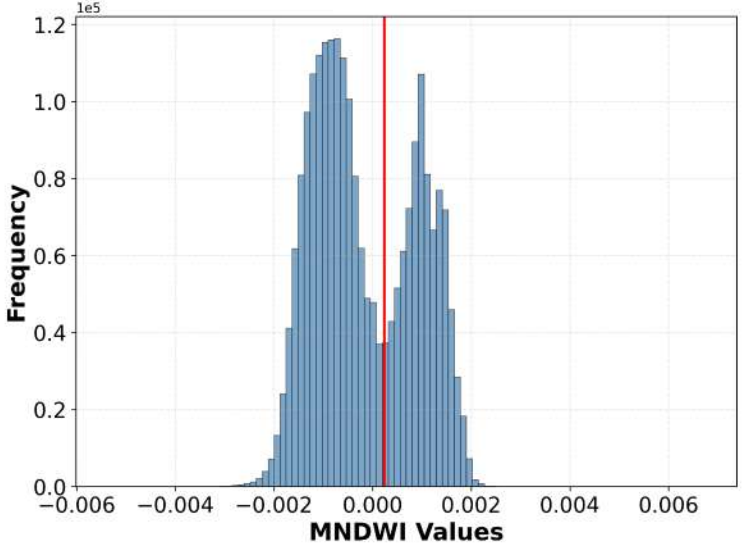
31.10.2024

Use Case: Flood Detection

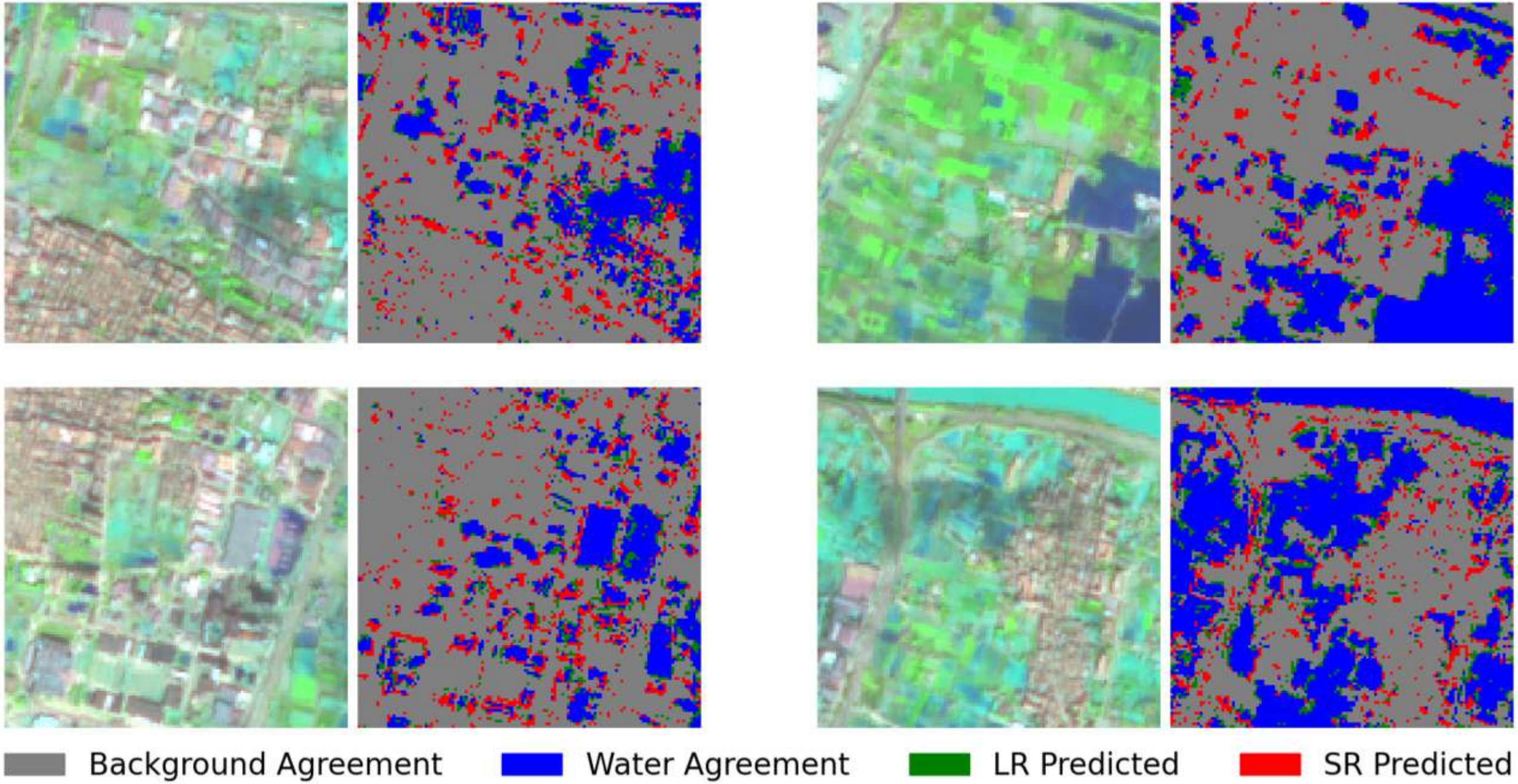
LR

SR

Modified Normalized Difference Water Index



Use Case: Flood Detection



Resources

```
pip install sen2sr
```

```
pip install opensr-model
```

Code, Publications, Colab Notebooks…

opensr.eu

Resources

https://tinyurl.com/5fc8pp2n

Towards Realistic and Trustworthy Super-Resolution for Multispectral Remote Sensing Images

Luis Gómez-Chova, Cesar Aybar, Simon Donike



ISP · Image & Signal Processing
Universitat de València

IEEE GRSS IADF SCHOOL
COMPUTER VISION FOR EARTH OBSERVATION
Image Analysis and Data Fusion

525 anys 1499-2024 Universitat de València

http://isp.uv.es