Towards Explainable AI: Application to Trustworthy Super-Resolution

ESA ITT AO/1-10948/21/I-AG

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

VNIVERSITAT DE VALÈNCIA

BROCKMANN CONSULT

UNIVERSITY OF OXFORD

SUREDOS24 – May 28, 2024

1: Datasets

Datasets Available

Synthetic

Cross-sensor

Datasets Available

Synthetic

LR: From HR.

Cross-sensor

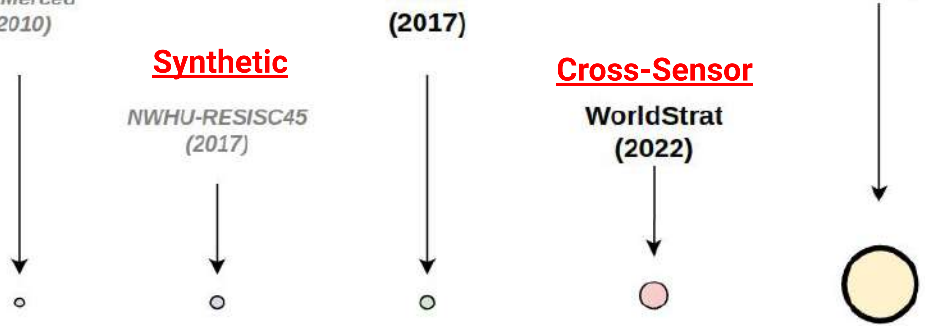
LR: From other sensor.

Datasets Available

	Synthetic	Synthetic	Synthetic
	UC Merced (2010)	*NWHU-RESISC45 (2017)*	**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

Datasets Available

	Synthetic	Synthetic	Synthetic	Cross-Sensor	Cross-Sensor
	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**



Datasets Available

	UC Merced (2010)	NWHU-RESISC45 (2017)	DIV2K (2017)	WorldStrat (2022)	Sen2Venus (2022)	OpenImage v7 (2022)
	Synthetic	Synthetic	Synthetic	Cross-Sensor	Cross-Sensor	Synthetic
# 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

Datasets Available

	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	$0.1 \times 10^{[illegible]}$	[illegible]	[illegible]	$[illegible] \times 10^9$	8.7×10^9	3×10^{12}
Amount	210[illegible]	[illegible]	[illegible]	3515	132 955	9 millions
Size	(256, [illegible]	[illegible]	(1972, 1437)	(a, b)	(256, 256)	(a, b)
Description	farmla[illegible], bushe[illegible], highways, overpasses, etc	[illegible]irports, basketball, residential, ports, etc	people, scenery, animal, decoration, etc	worldwide	worldwide	people, scenery, animal, decoration, etc

Limited Availability of Super-Resolution Datasets for Sentinel-2.

Challenges

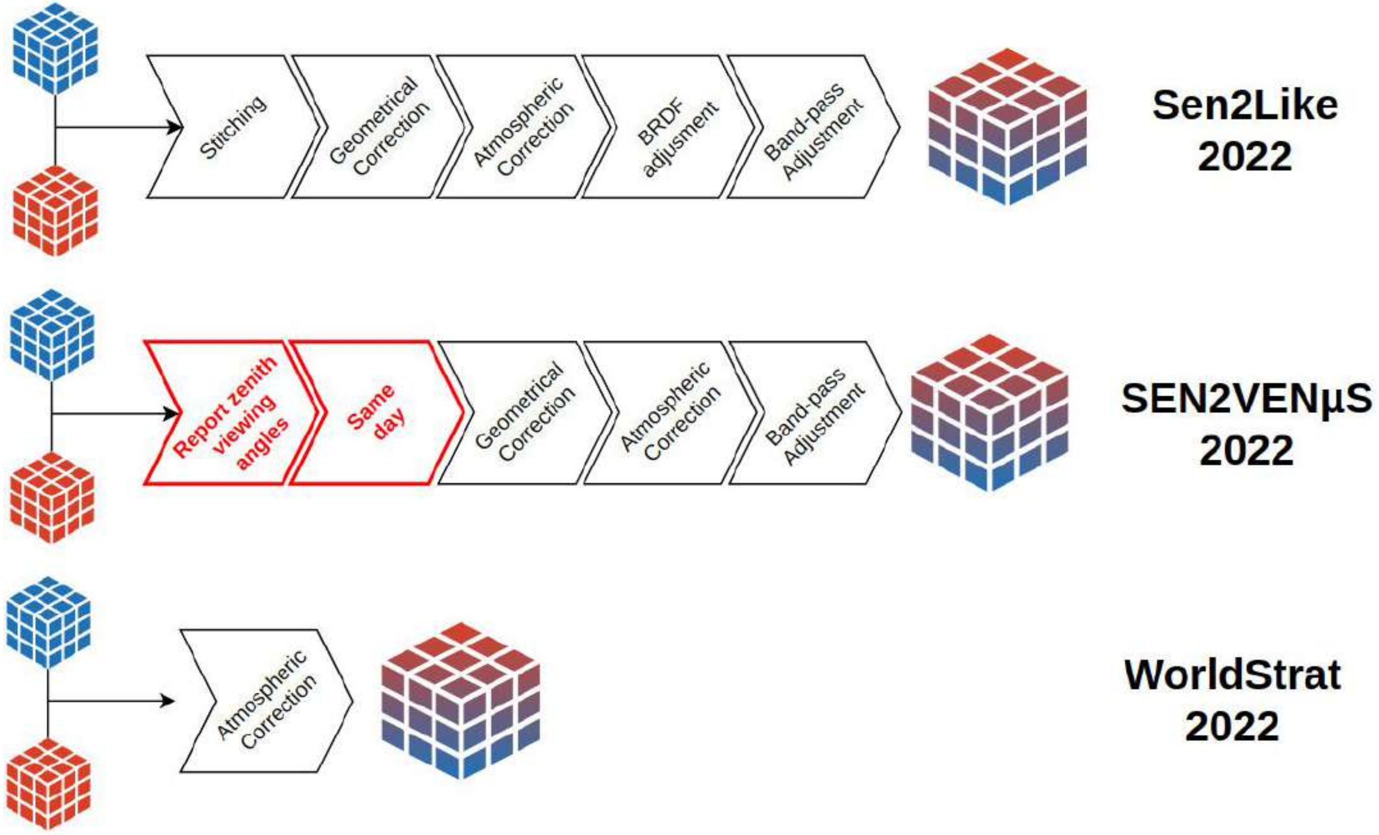
<u>Cross-Sensor</u>

- Spectral bands mismatch.
- Variations in atmospheric conditions during acquisition.
- Differences in zenith viewing angle.
- <u>Spatial alignment</u> is requirement.

<u>Synthetic</u>

- <u>**Learn the distribution shift**</u>, i.e., training dataset could be different to the real-world cases.

Harmonization



TakeAway

Challenges

Cross-Sensor

- Spectral bands mismatch.
- Variations in atmospheric conditions during acquisition.
- Differences in zenith viewing angle.
- Spatial alignment is requirement.

Synthetic

- **Learn the distribution shift**, i.e., training dataset could be different to the real-world cases.

TakeAway

Challenges

Cross-Sensor

- Spectral bands mi
- Vari nditions during
- erences in zenit
- Spatial

Harmonization

Hard to Scale

They might not work well in real world

Synthetic

Easy to Scale

- shift, i.e., ataset could be different to the real-world cases.

3: Proposed Solution

Learn the best degradation with cross-sensor;
extent it with synthetic datasets.

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

Main Goal:

Merge the best of both worlds. ❤️

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

Main Goal:

Develop an algorithm to **learn the degradation** between a very HR image (NAIP - 2.5 meters) and Sentinel-2 (10 meters).

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

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

Main Goal:

Develop an algorithm to **learn the degradation** between a very HR image (NAIP - 2.5 meters) and Sentinel-2 (10 meters).

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

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

λ : **Harmonization model**

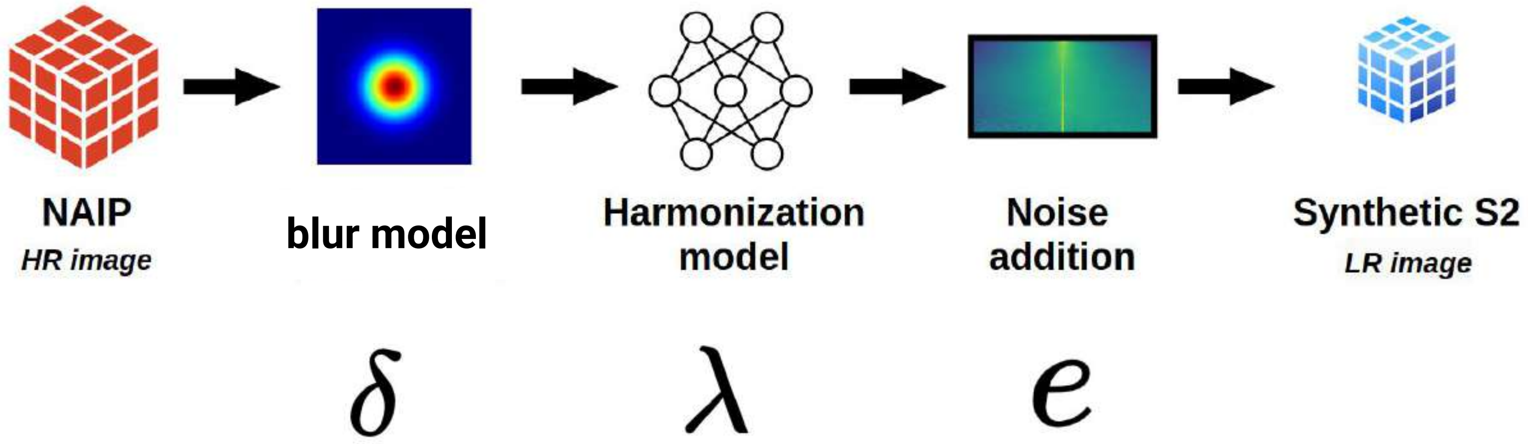
δ : **Blur model**

e : **Noise model**

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

Main Goal:

Develop an algorithm to **<u>learn the degradation</u>** between a very HR image (NAIP - 2.5 meters) and Sentinel-2 (10 meters).

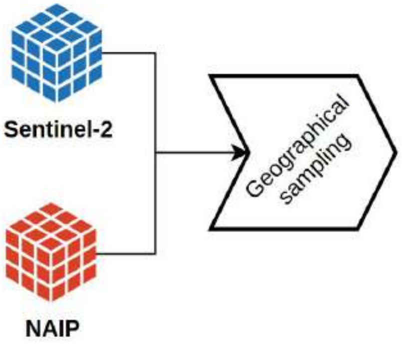


4: SEN2NAIP

Create the SEN2NAIP cross-sensor dataset

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

SEN2NAIP - CrossSensor



18k ROIs

NAIP: National Agriculture Imagery Program

Dataset Availability

2002-06-15T00:00:00Z–2022-08-31T00:00:00

Dataset Provider

USDA Farm Production and Conservation - Business Center, Geospatial Enterprise Operations

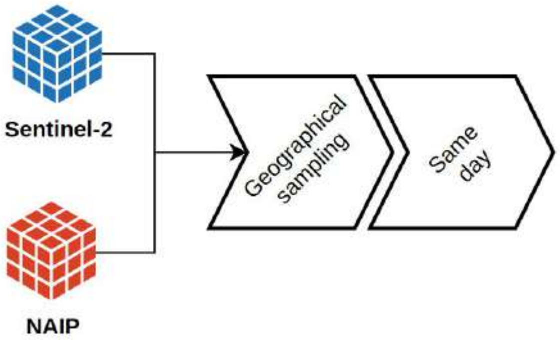
Earth Engine Snippet

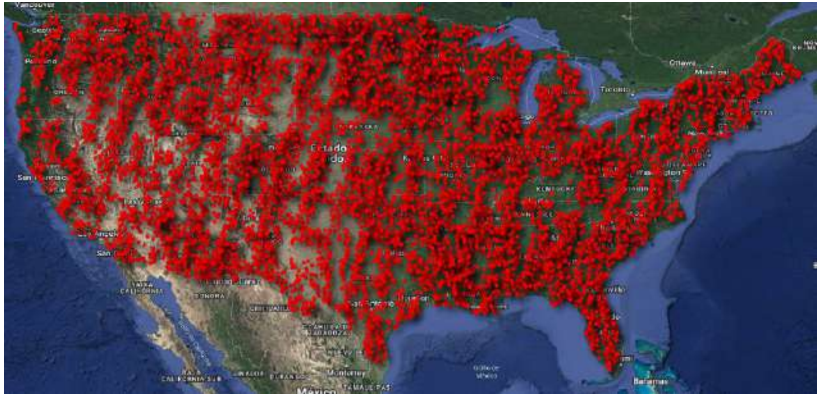
`ee.ImageCollection("USDA/NAIP/DOQQ")`

Tags

agriculture highres imagery usda aerial fpac naip

SEN2NAIP - CrossSensor





7k ROIs

NAIP: National Agriculture Imagery Program

Dataset Availability

2002-06-15T00:00:00Z–2022-08-31T00:00:00

Dataset Provider

USDA Farm Production and Conservation - Business Center, Geospatial Enterprise Operations

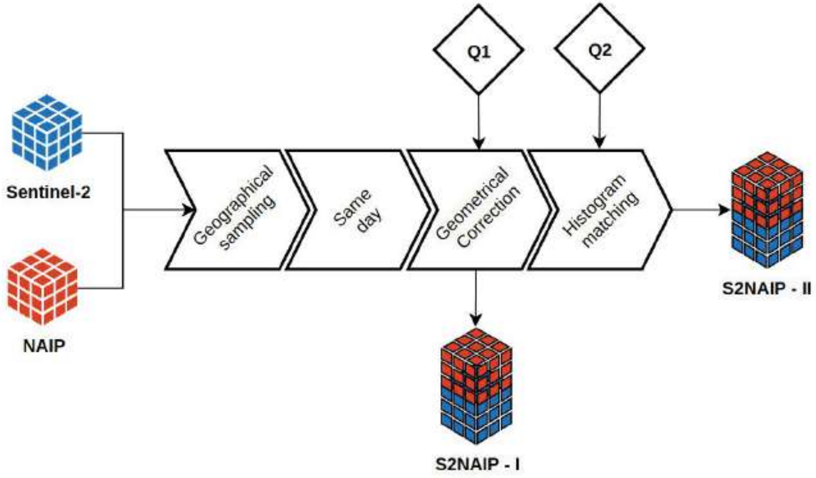
Earth Engine Snippet

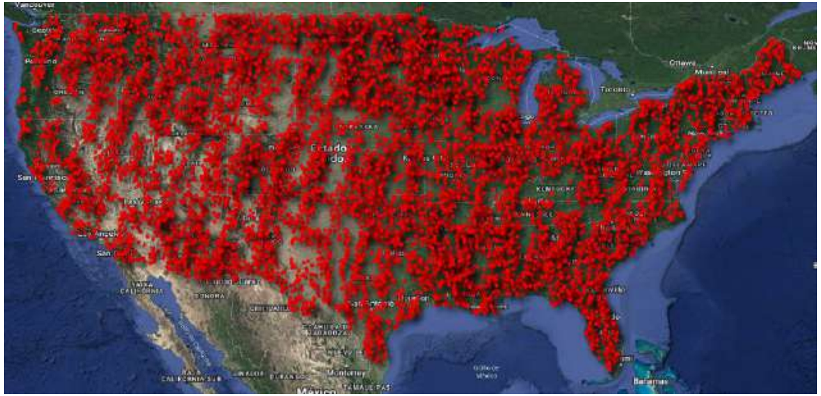
`ee.ImageCollection("USDA/NAIP/DOQQ")`

Tags

agriculture highres imagery usda aerial fpac naip

SEN2NAIP - CrossSensor





14k ROIs

SuperGlue + SuperPoint

PE Sarlin et al, 2020

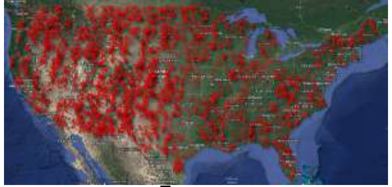
(Q1)

Histogram Matching

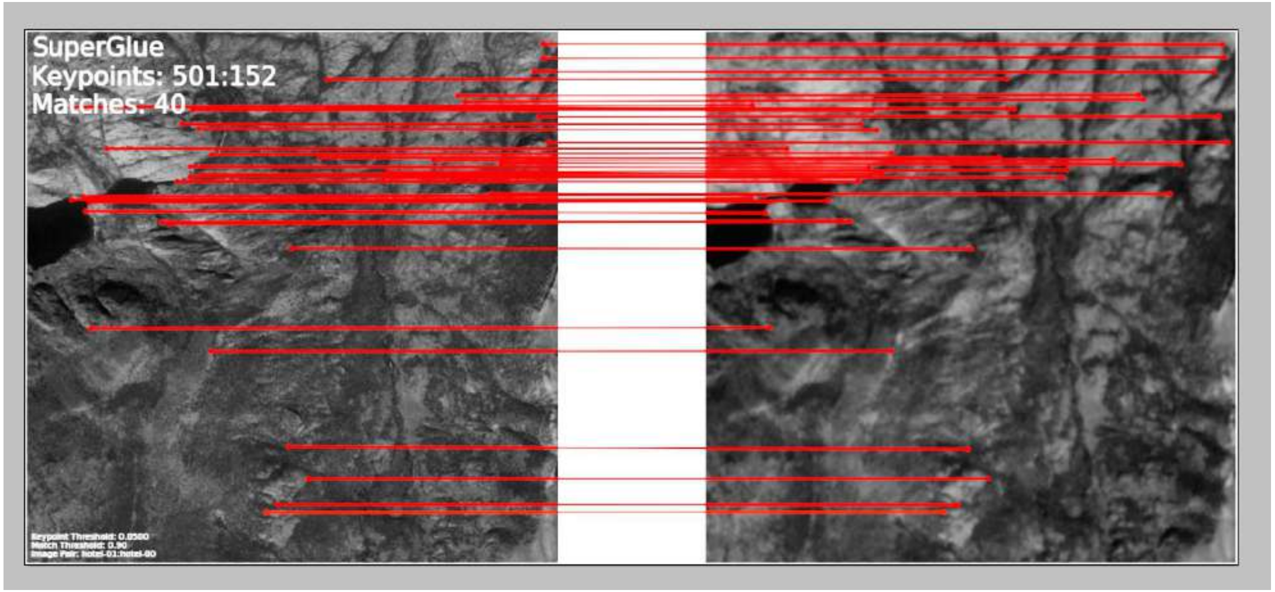
scikit-image

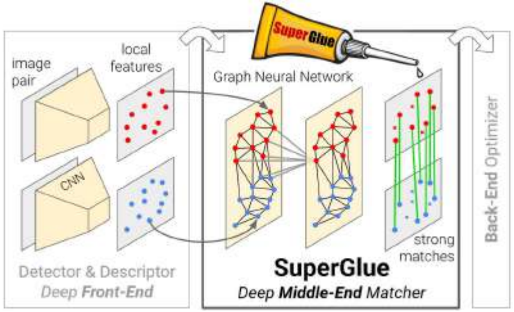
(Q2)

SEN2NAIP - CrossSensor



5k ROIs





PE Sarlin et al, 2020

- <u>No fine-tuned*</u>
- Discard matches that are further that 10 m apart.

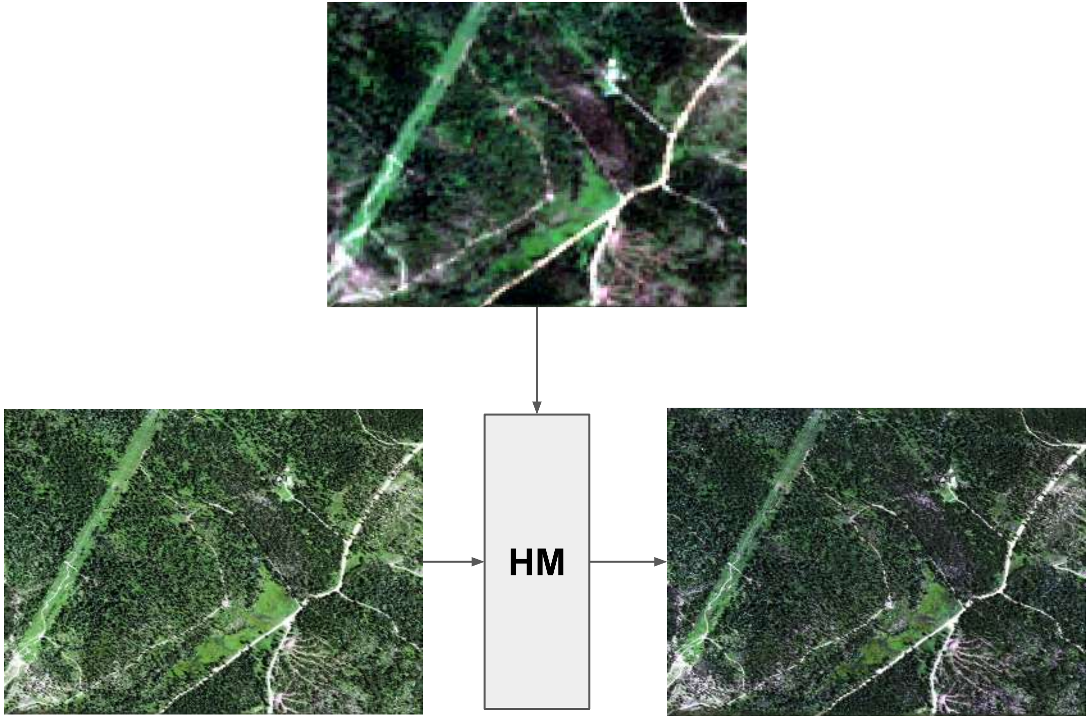
SuperGlue + SuperPoint

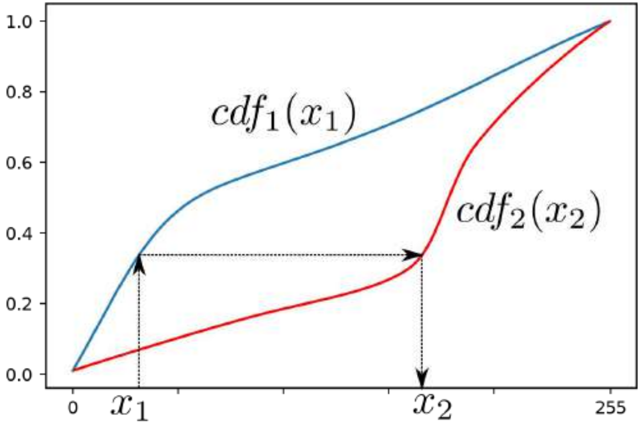
Quality test Q1

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

SEN2NAIP - CrossSensor

3k ROIs

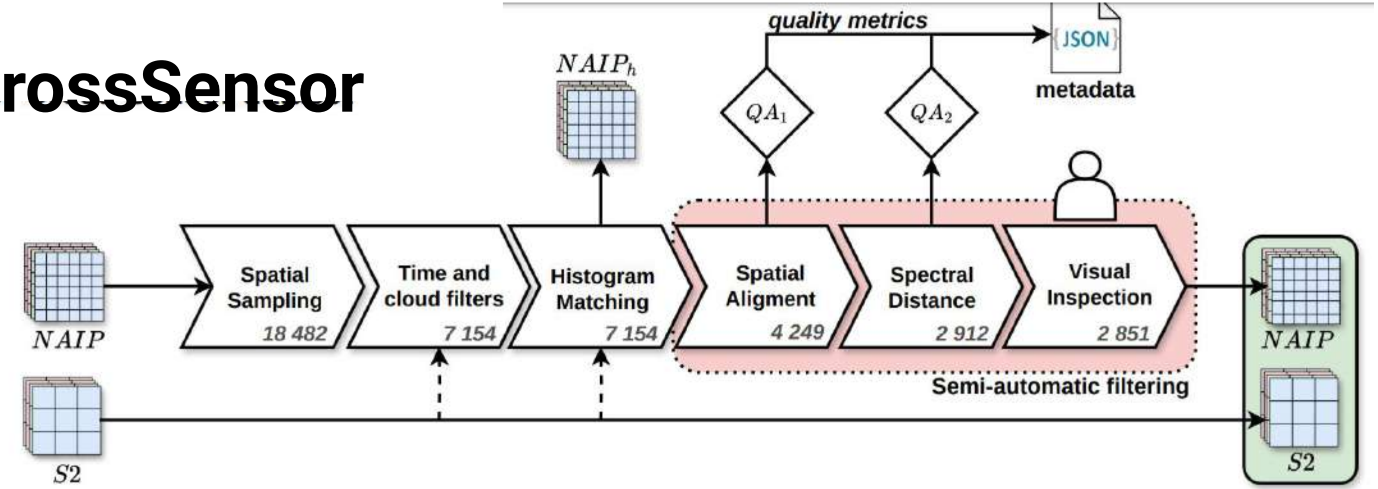




Quality test Q2

If spectral angle distance > 2 degrees

SEN2NAIP - CrossSensor

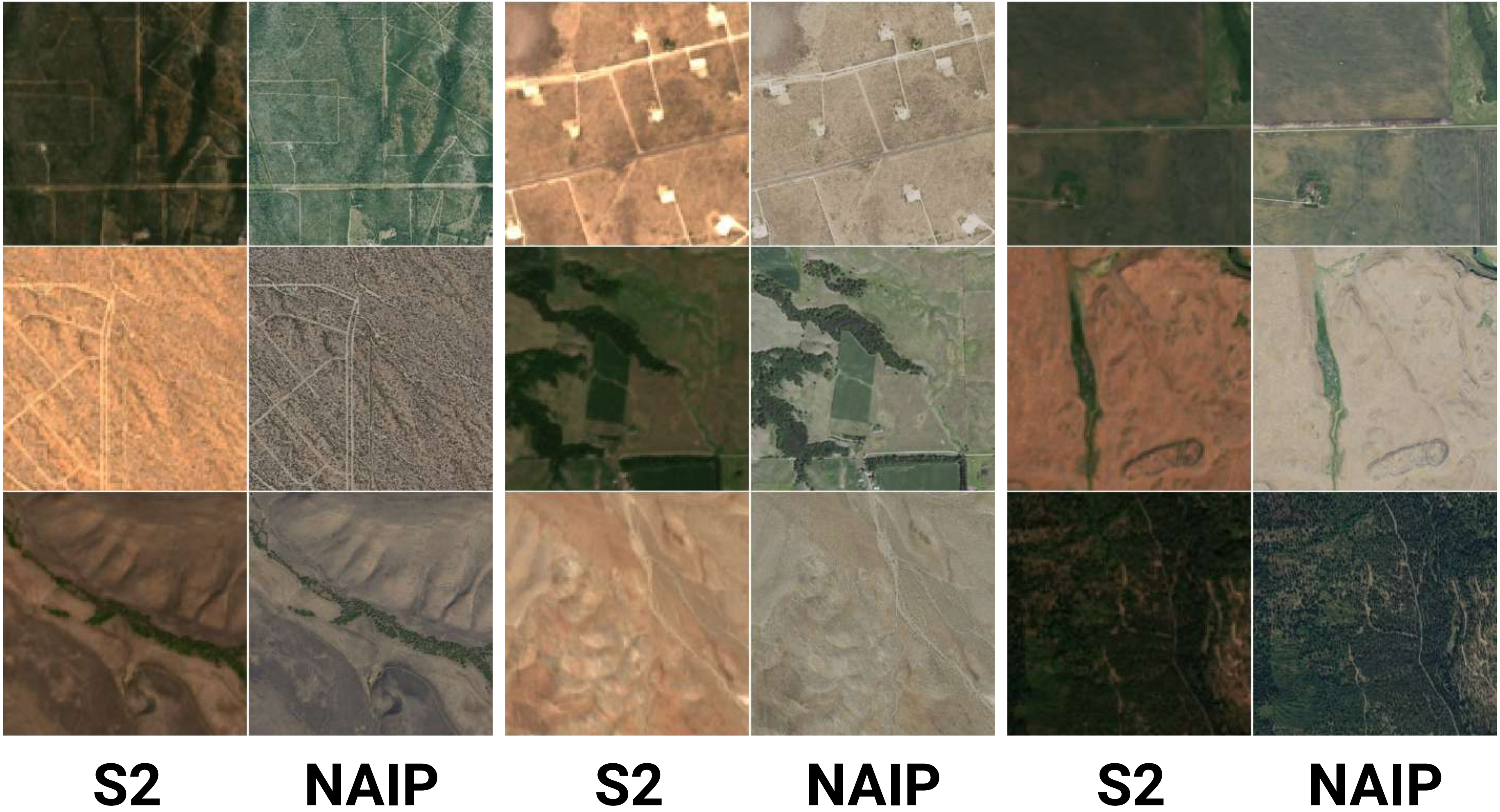


We find

2851

NAIP - S2

image pairs.

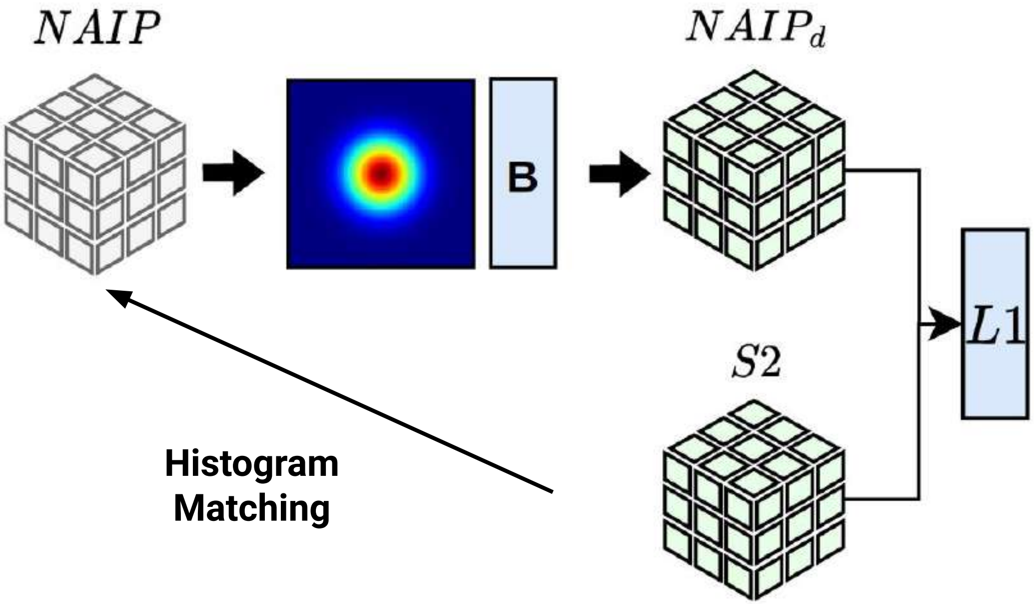

S2
NAIP
S2
NAIP
S2
NAIP

5: Degradation model

Learn how to convert NAIP to Sentinel-2 images.

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

1 Learn the blur model



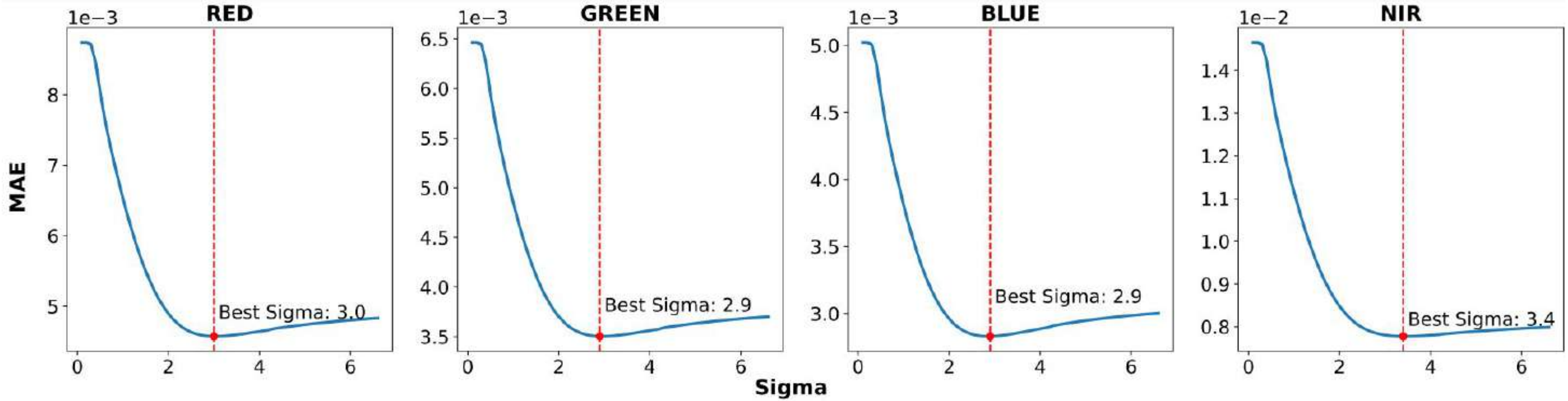
3k ROIs

We try different kernels:
Gaussian, Sigmoid and Gumbel
and different downscaling models:
bilinear, nearest, bicubic.

Best Blur model (per band):
<u>Gaussian + bilinear</u>

1 Learn the degradation

3k ROIs



Error curves (MAE) for the RGBNIR bands: each curve represents the relationship between the Gaussian kernel's sigma value (x-axis) and the associated error (y-axis). The optimal sigma value is highlighted by a dashed red line.

2 Learn the harmonization model

3k ROIs

NAIP

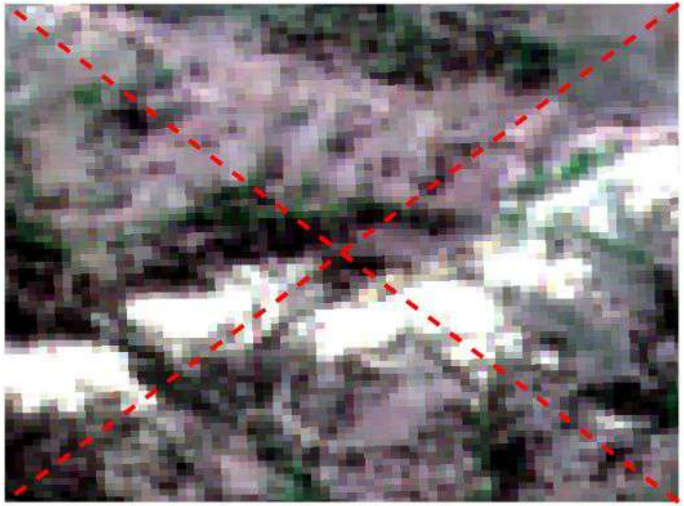
Sentinel-2

2 Learn the harmonization model

3k ROIs

NAIP

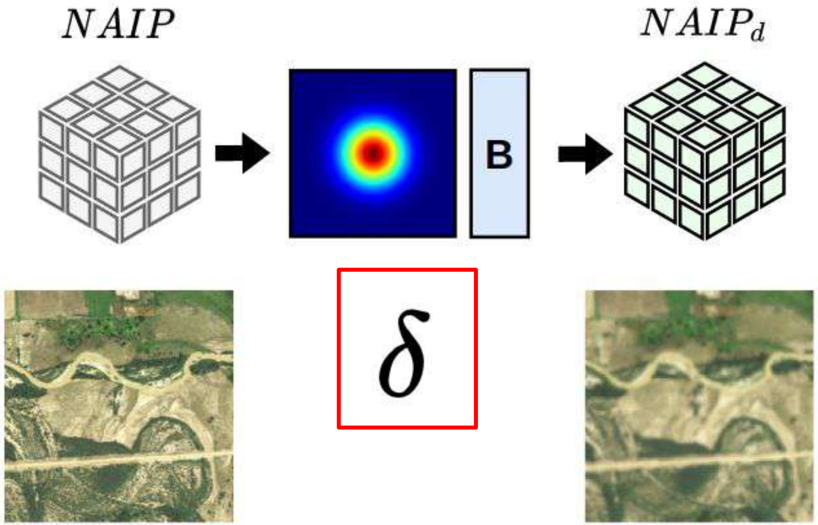
Sentinel-2



No access in inference time!

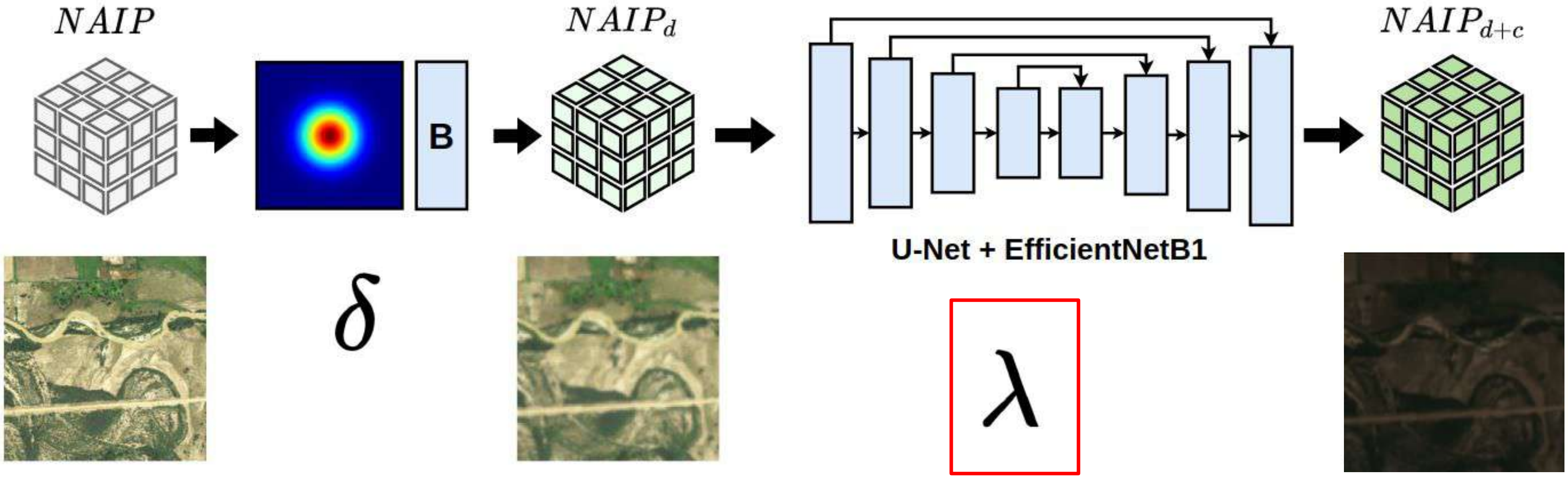
2 Learn the harmonization model

3k ROIs



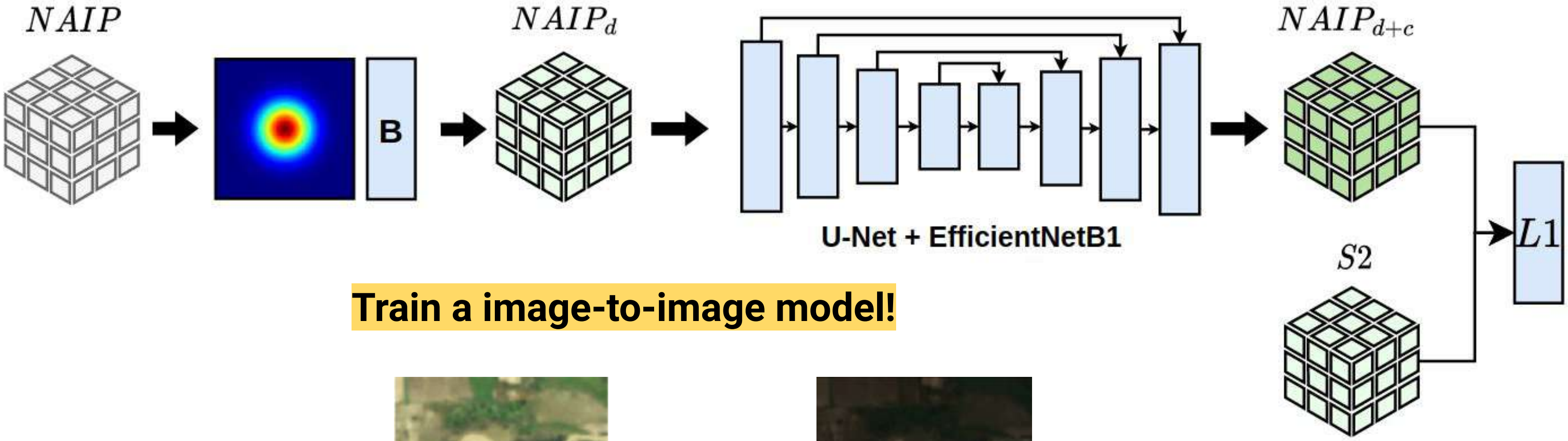
2 Learn the harmonization model

3k ROIs

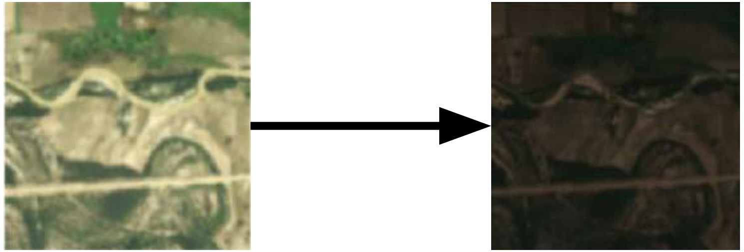


2 Learn the harmonization model

3k ROIs

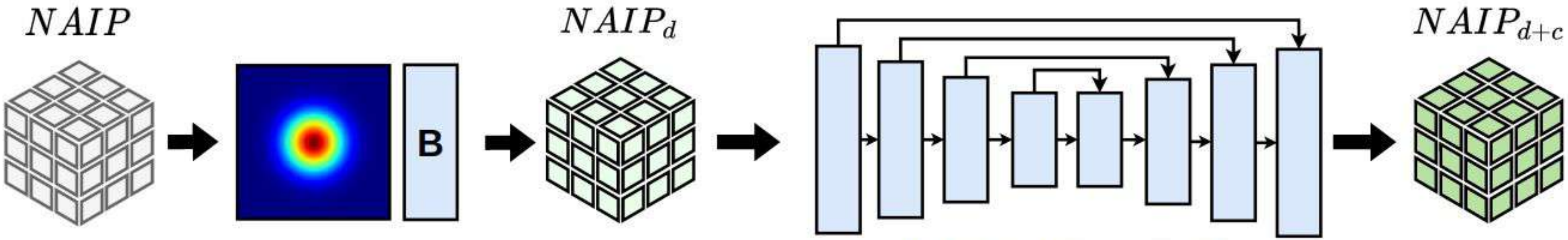


Train a image-to-image model!



2 Learn the harmonization model

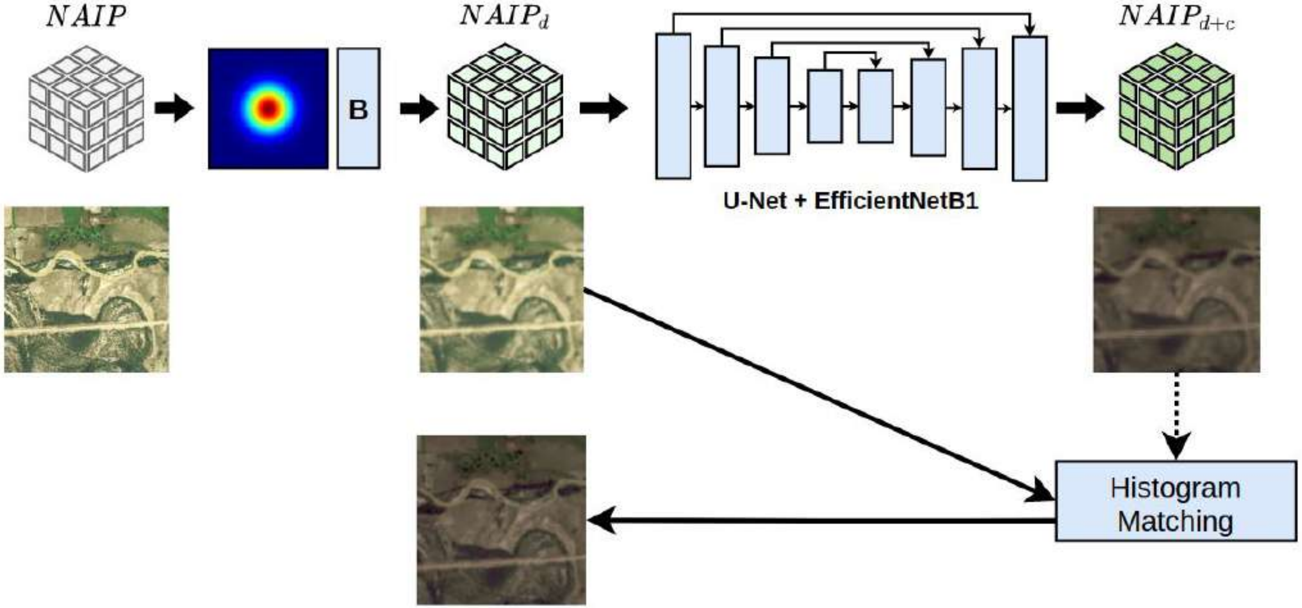
3k ROIs



The U-Net harmonization model degrades the spatial quality!

2 Learn the harmonization model

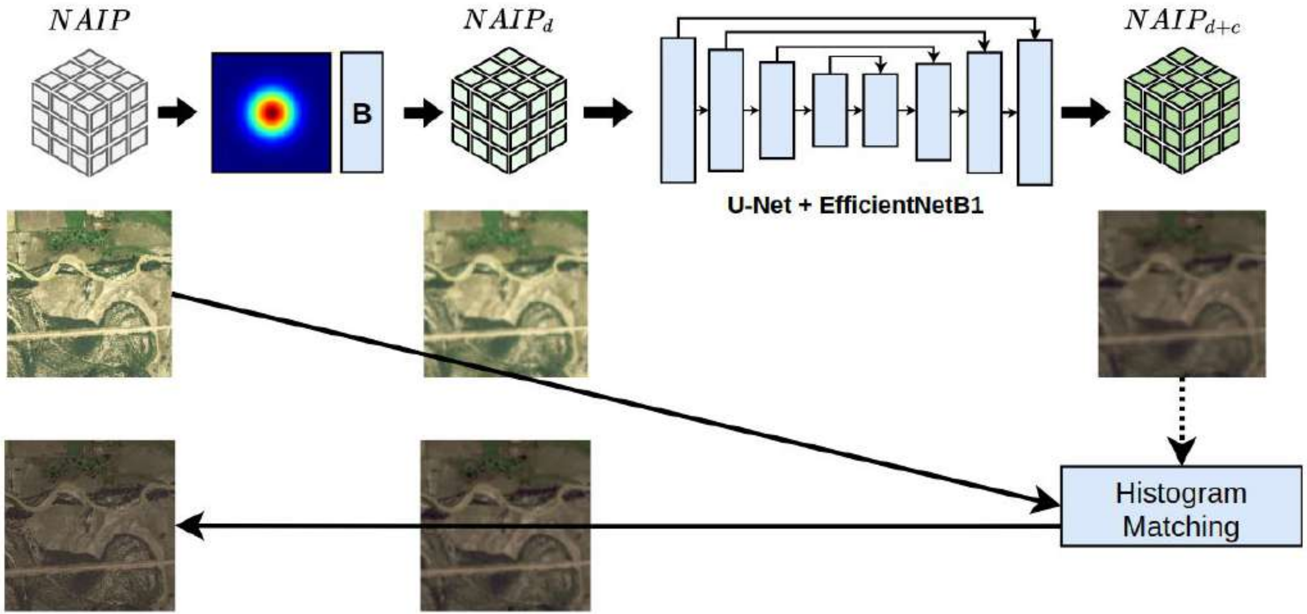
3k ROIs



The **U-Net output is used to perform a local histogram matching harmonization.**

2 Learn the harmonization model

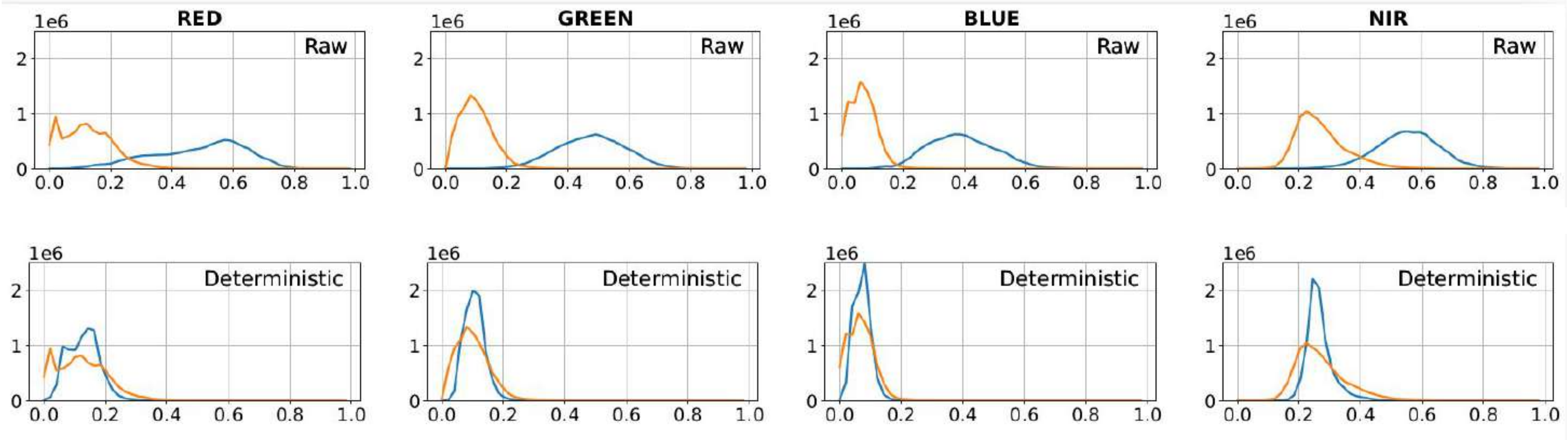
3k ROIs



The **U-Net output is used to perform a local histogram matching harmonization.**

2 Learn the harmonization model

Sentinel-2 NAIP



Spectral distribution comparisons across RGBNIR bands in the **SEN2NAIP cross-sensor test subset**. The **orange line represents the Sentinel-2 reference**, and the **blue lines indicate the proposed degradation method**.

3 Learn the noise model using CloudSEN12

3k ROIs

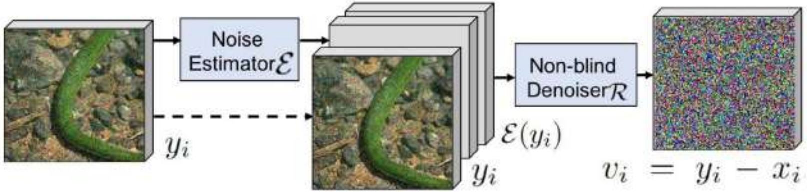
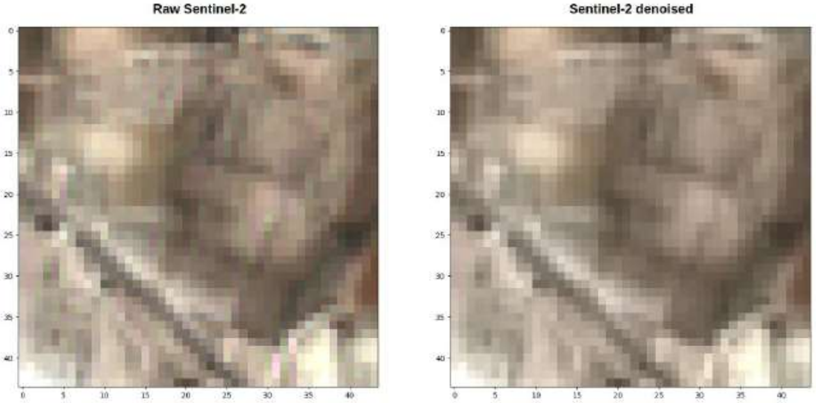
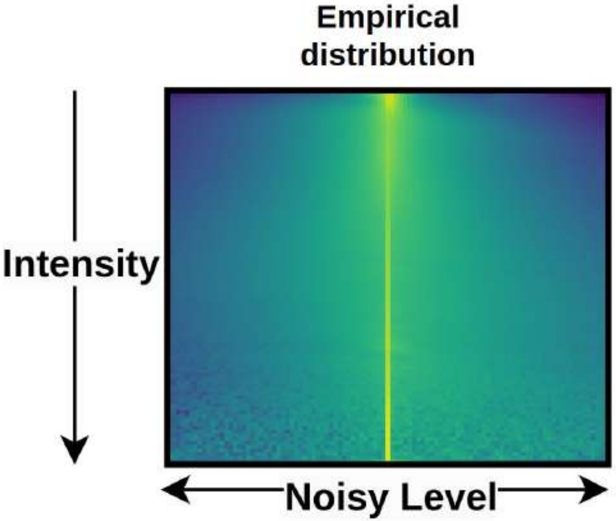


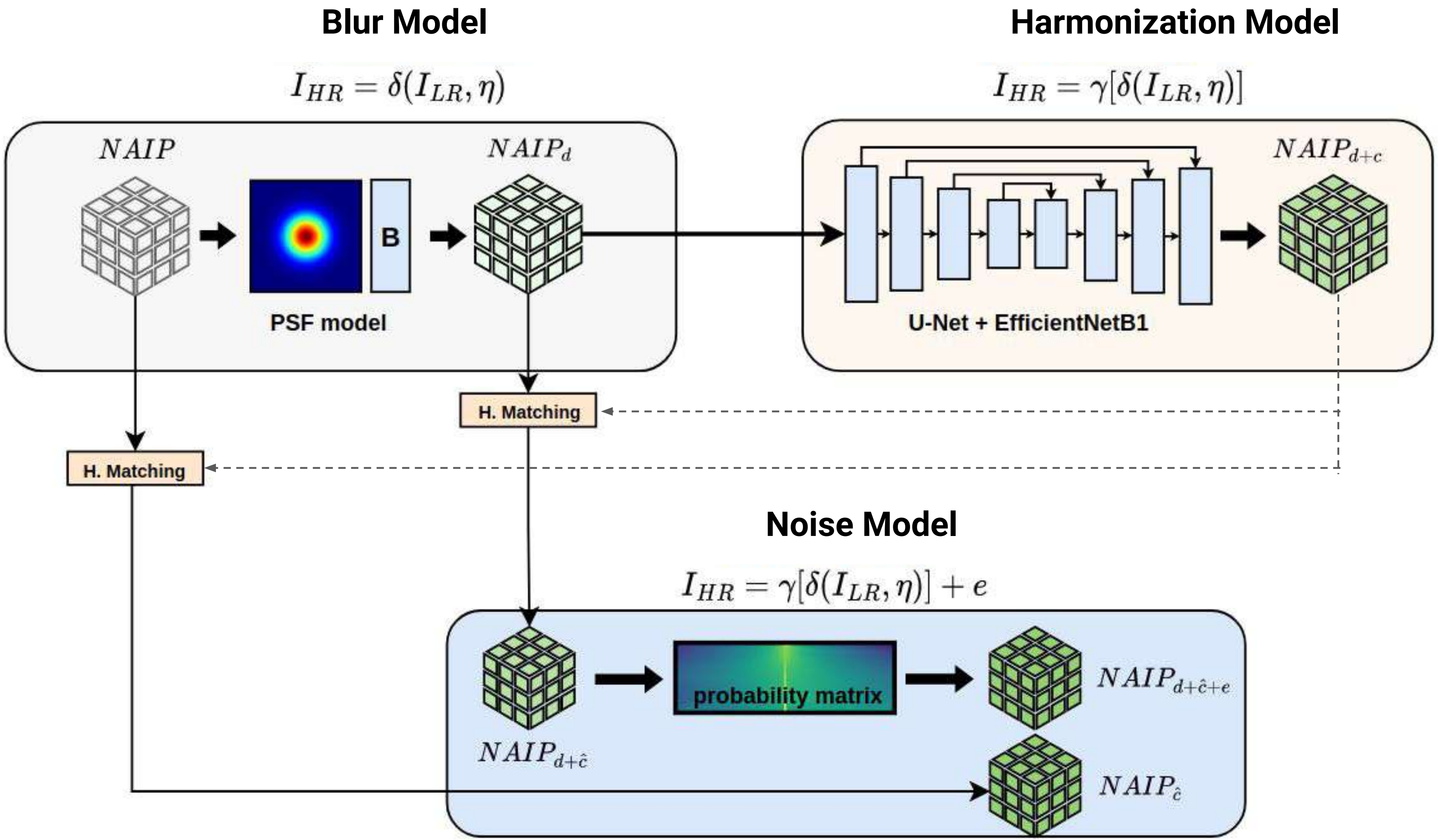
Figure 3: Structure of the proposed blind denoising model. It consists of a noise estimator $\mathcal{E}$ and a follow-up non-blind denoiser $\mathcal{R}$. The model aims to jointly learn the image residual.



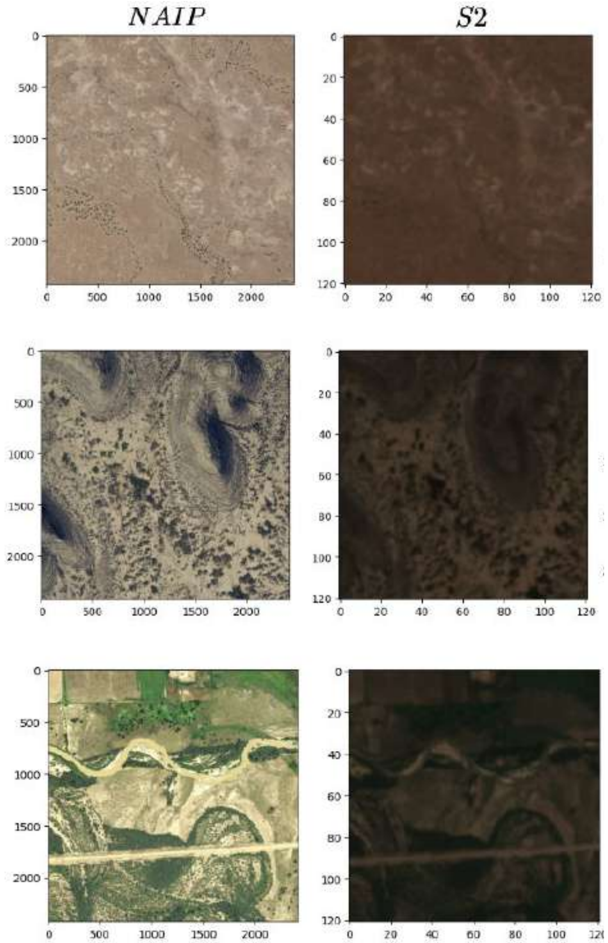


AWGN-based Denoiser, CVPR 2019
Zhou et al., 2019

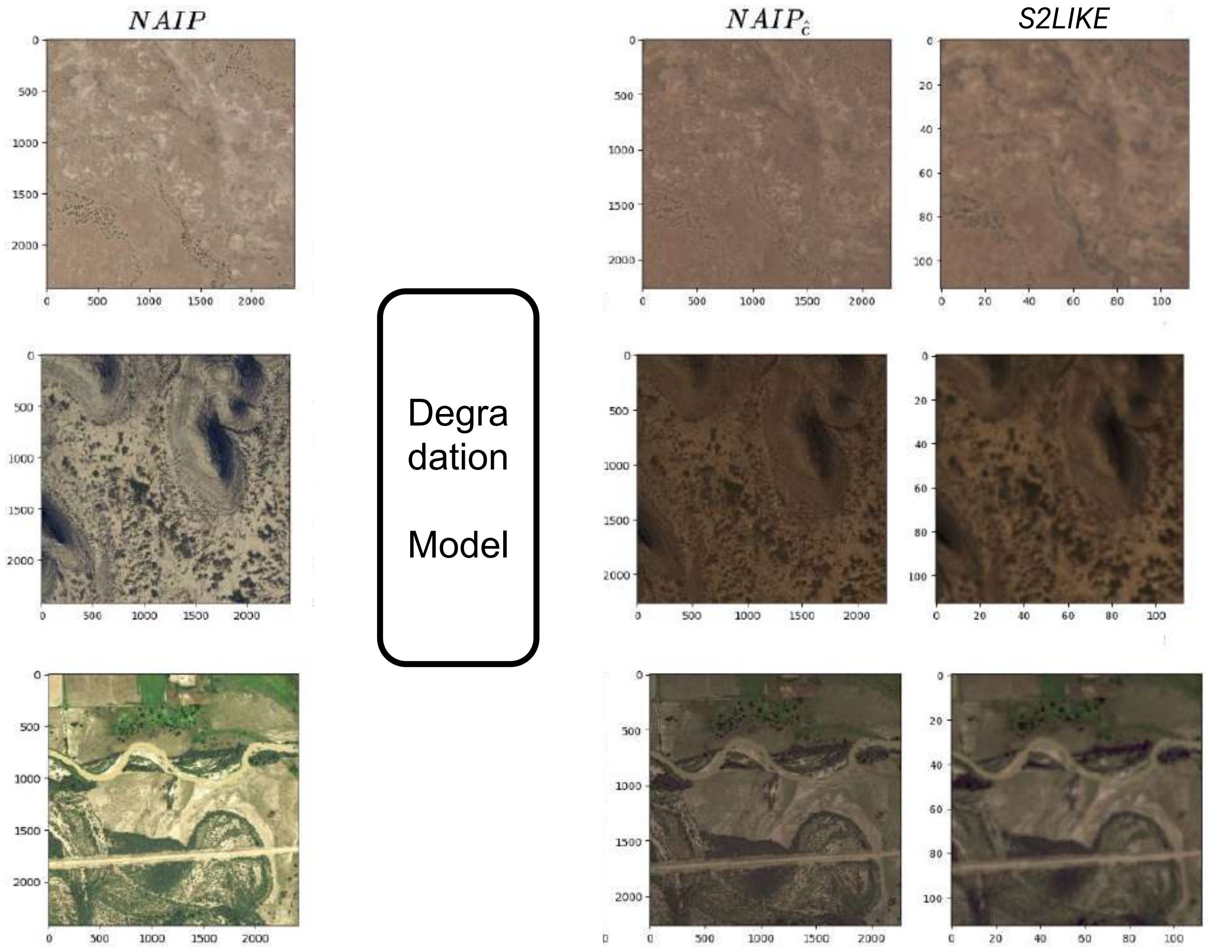
Put all together!



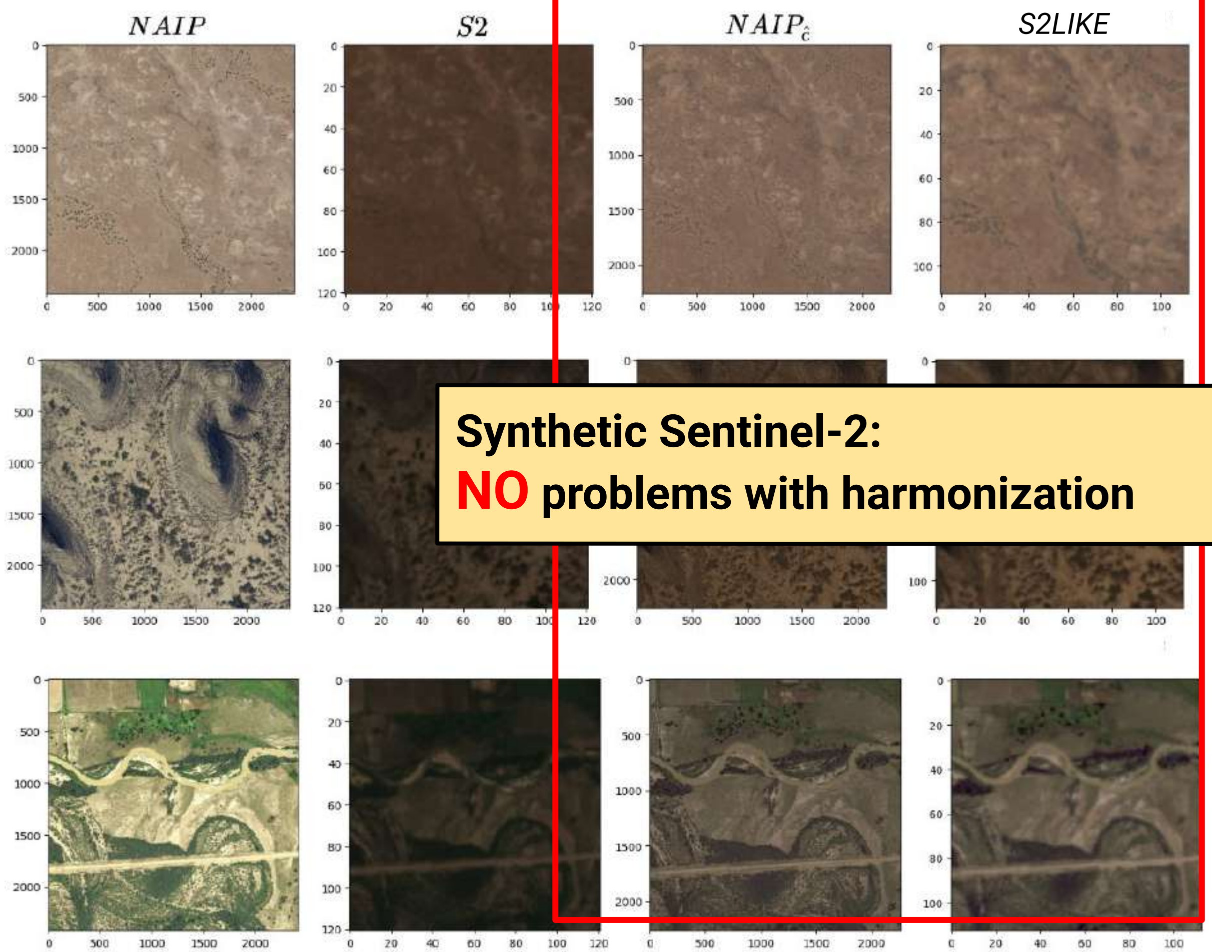
Results



Results

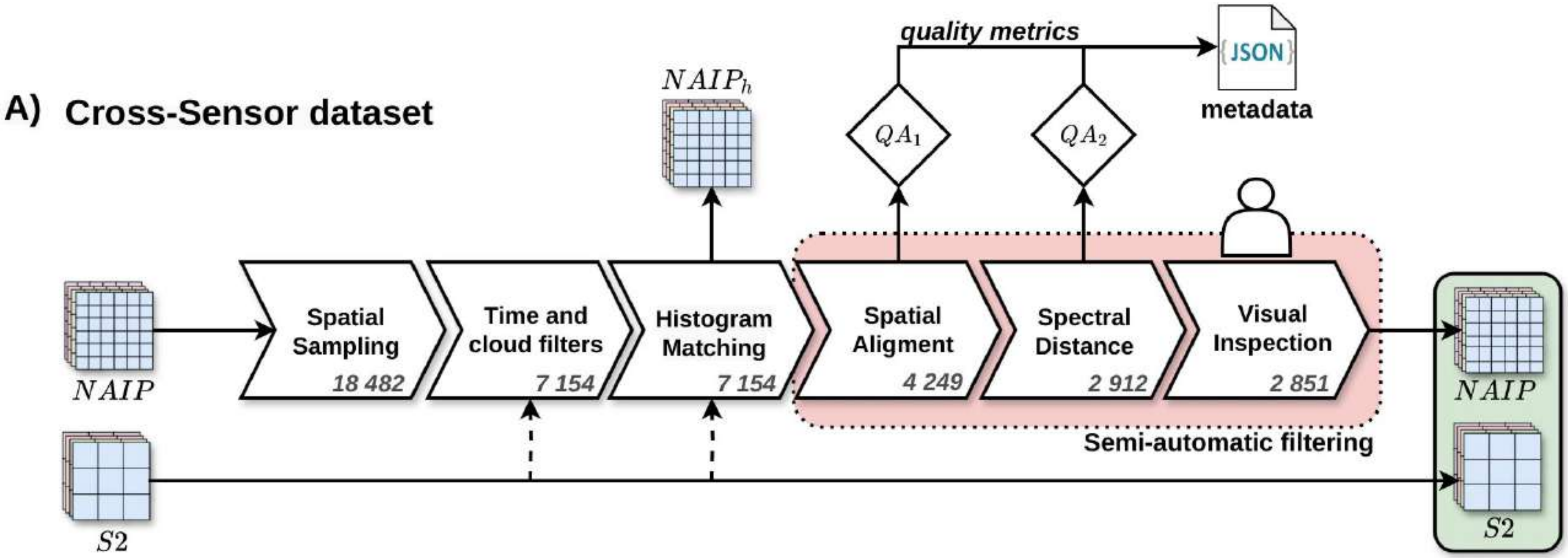


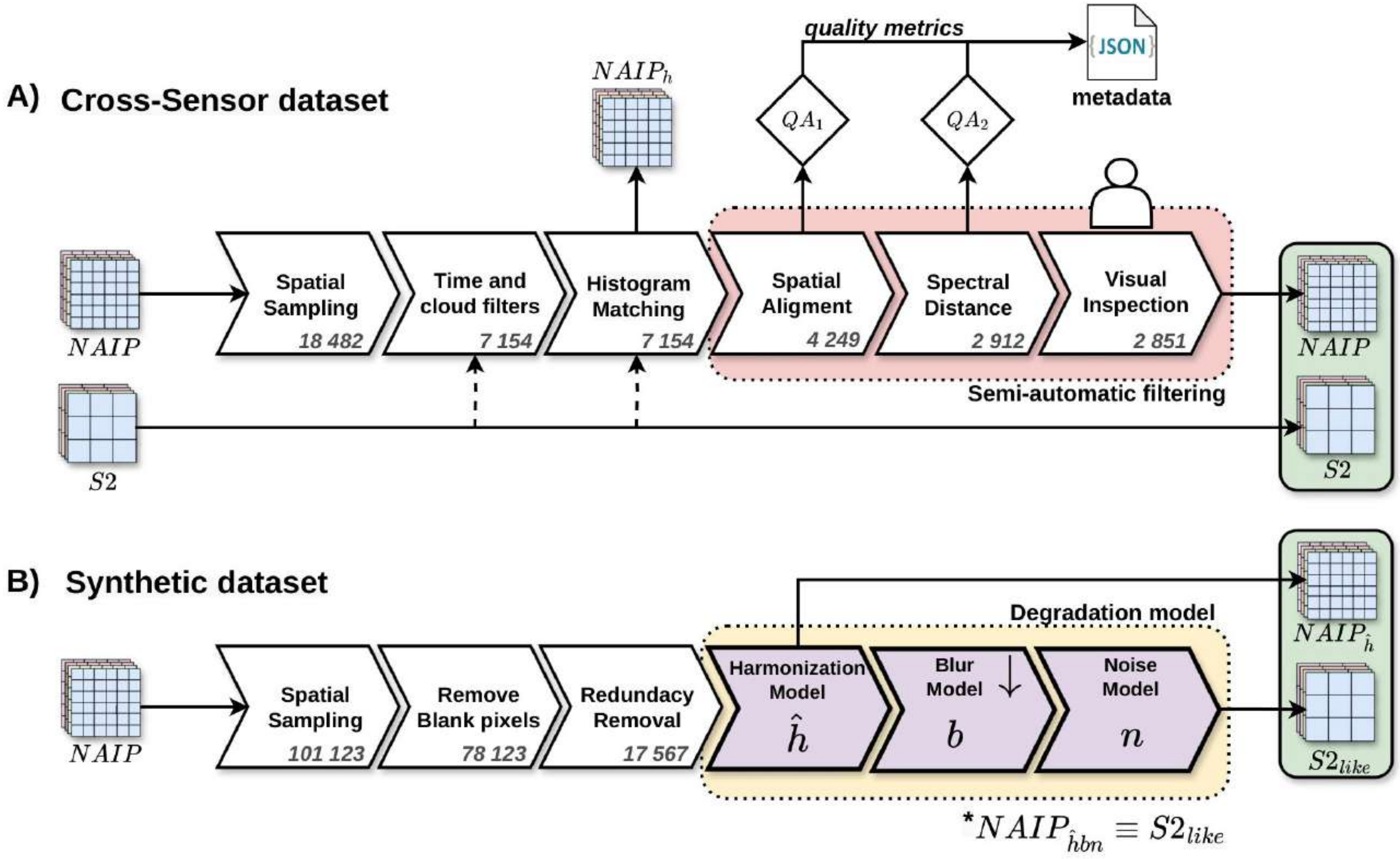
Results



Synthetic Sentinel-2:
NO problems with harmonization

4: SEN2NAIP
(Synthesis)

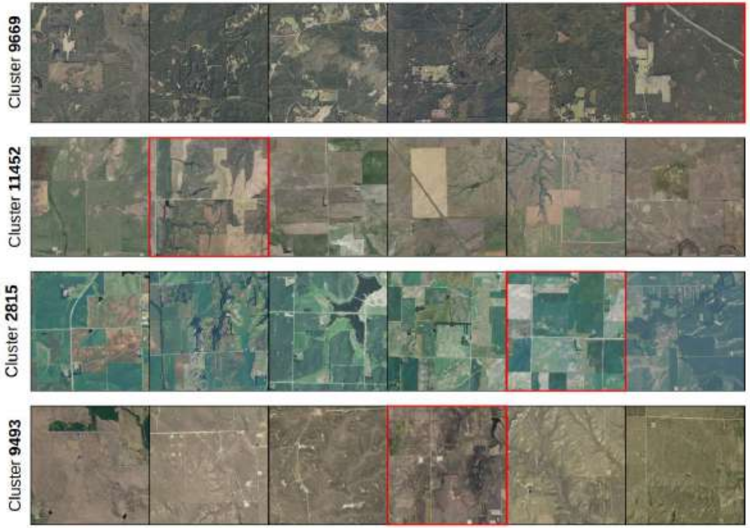

A) Cross-Sensor dataset
quality metrics
{JSON}
metadata
$NAIP_h$
QA_1
QA_2
$NAIP$
Spatial Sampling
18 482
Time and cloud filters
7 154
Histogram Matching
7 154
Spatial Aligment
4 249
Spectral Distance
2 912
Visual Inspection
2 851
Semi-automatic filtering
$S2$
$NAIP$
$S2$



A) Cross-Sensor dataset
$NAIP$
$S2$
Spatial Sampling
18 482
Time and cloud filters
7 154
Histogram Matching
7 154
$NAIP_h$
Spatial Aligment
4 249
Spectral Distance
2 912
Visual Inspection
2 851
QA_1
QA_2
quality metrics
JSON
metadata
Semi-automatic filtering
$NAIP$
$S2$
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
b
Noise Model
n
$NAIP_{\hat{h}}$
$S2_{like}$
$^*NAIP_{\hat{h}bn} \equiv S2_{like}$

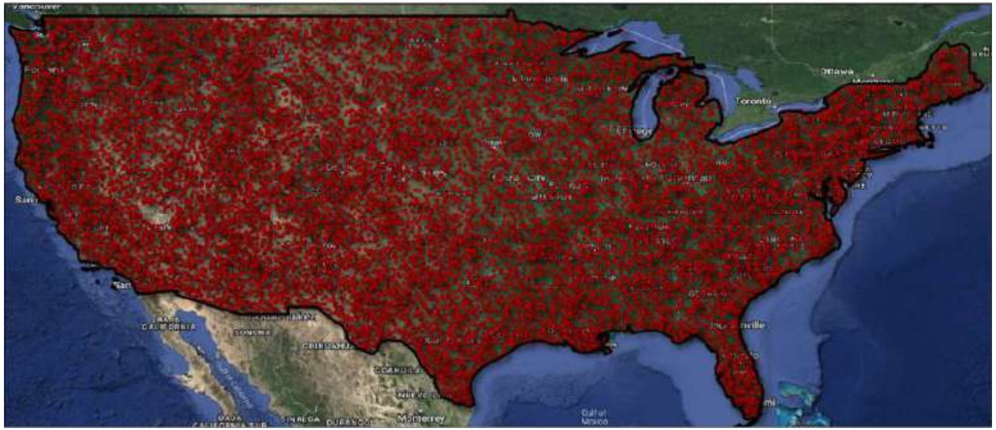
Remove Blank Pixel images

CLIP & Kmeans++



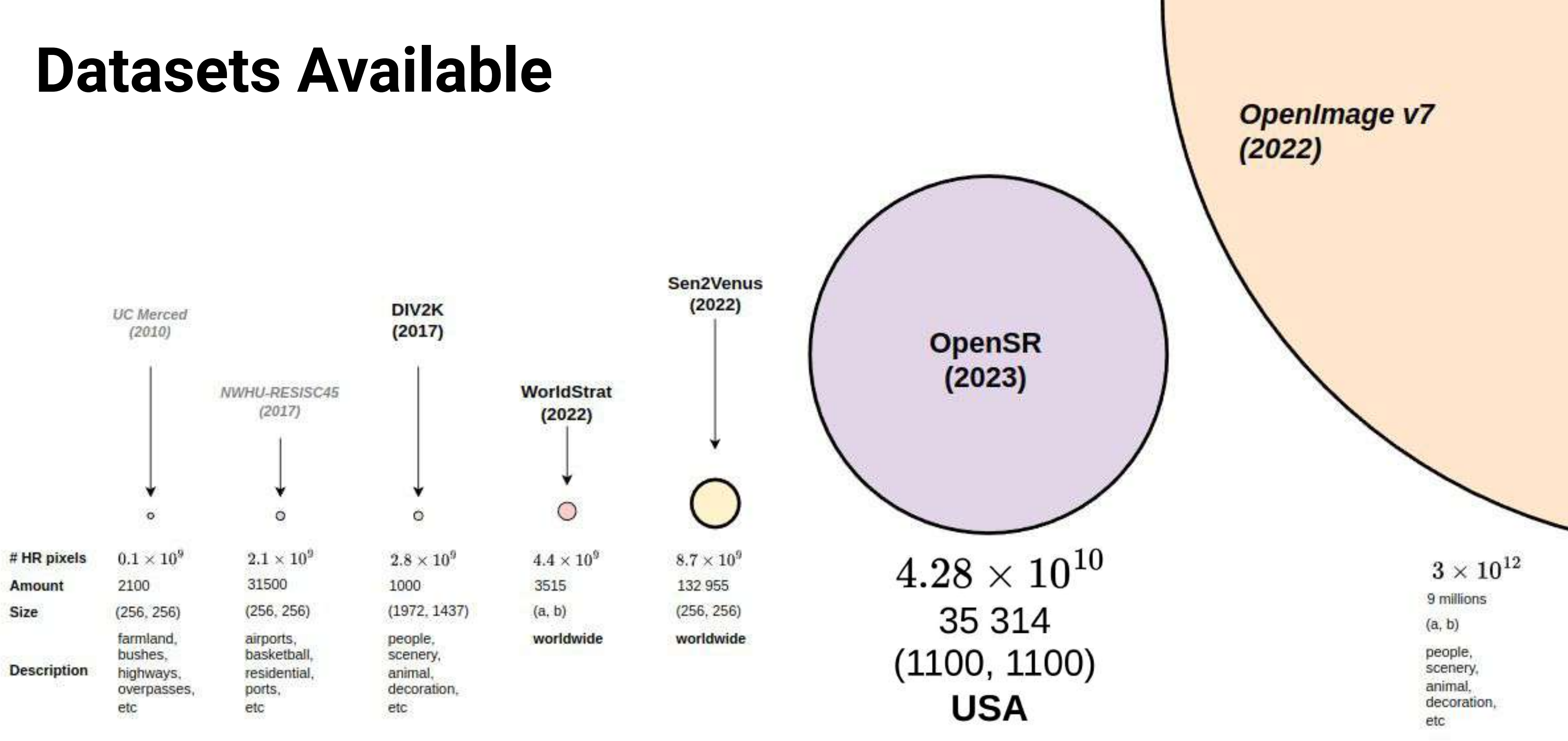
(A) **Cross-Sensor Dataset (2 851 ROIs)**

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



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

Datasets Available



5: Conclusions

TakeAway

Given its consistently high quality, the **NAIP** image is an excellent starting point to explore the training of x4 SR models.

If you want to create super-resolution real-world models using synthetic data, **learning the degradation process is as important as learning the super-resolution process**.

Only using SEN2NAIP will **make your results look like they are from the USA everywhere**.

We made a **mistake by being too strict with the filters when creating the cross-sensor dataset**. A more robust harmonization model would benefit from including more data points.

Thank you!

The “it” in AI models is the dataset.

Posted on June 10, 2023 by jbetker

I’ve been at OpenAI for almost a year now. In that time, I’ve trained a **lot** of generative models. More than anyone really has any right to train. As I’ve spent these hours observing the effects of tweaking various model configurations and hyperparameters, one thing that has struck me is the similarities in between all the training runs.

It’s becoming awfully clear to me that these models are truly approximating their datasets to an incredible degree. What that means is not only that they learn what it means to be a dog or a cat, but the interstitial frequencies between distributions that don’t matter, like what photos humans are likely to take or words humans commonly write down.

What this manifests as is – trained on the same dataset for long enough, pretty much every model with enough weights and training time converges to the same point. Sufficiently large diffusion conv-unets produce the same images as ViT generators. AR sampling produces the same images as diffusion.

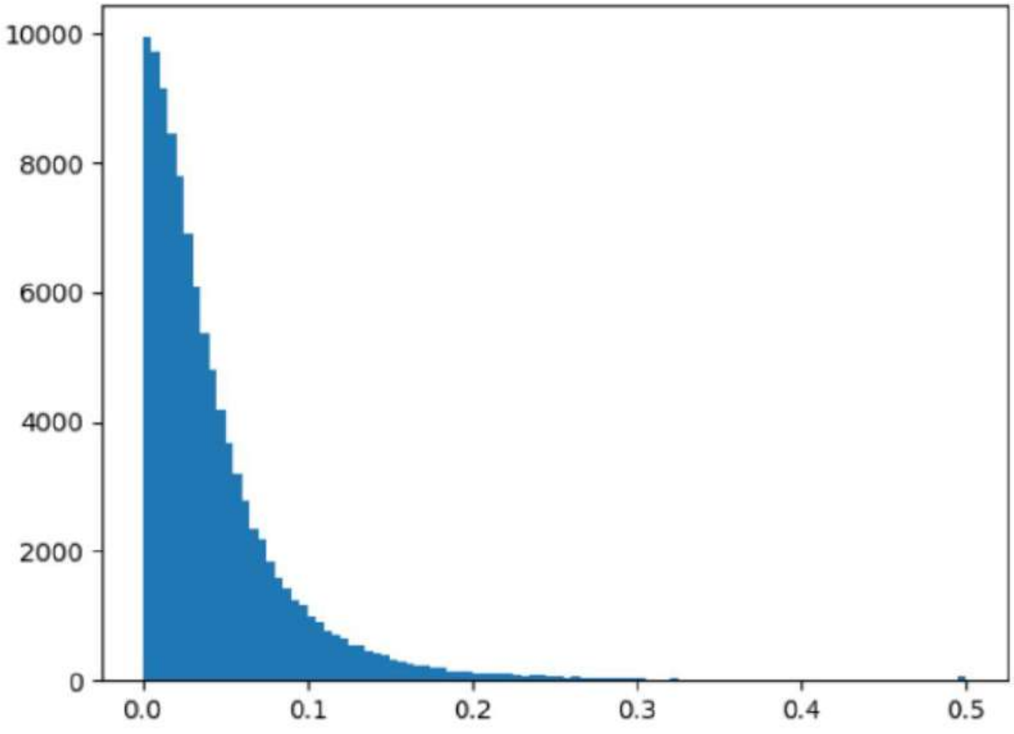
This is a surprising observation! It implies that model behavior is not determined by architecture, hyperparameters, or optimizer choices. It’s determined by your dataset, nothing else. Everything else is a means to an end in efficiently delivery compute to approximating that dataset.

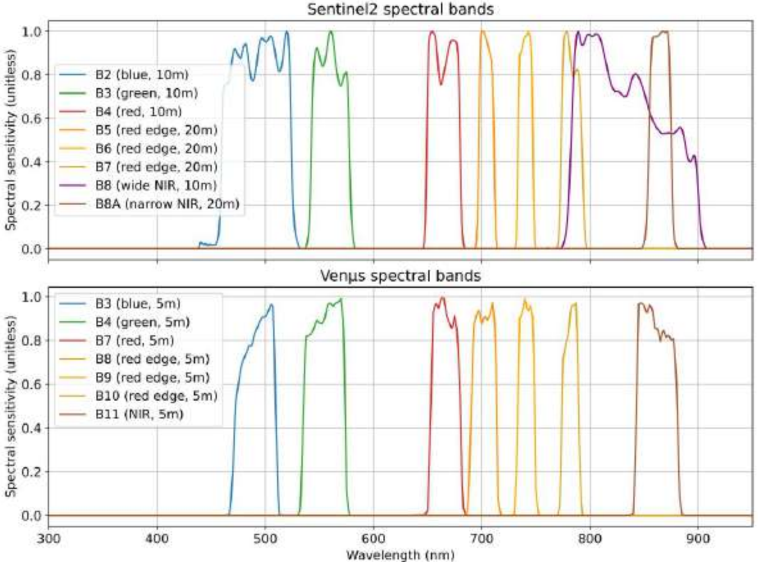
Then, when you refer to “Lambda”, “ChatGPT”, “Bard”, or “Claude” then, it’s not the model weights that you are referring to. It’s the dataset.

https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/

Harmonization

L1 norm ↓ VENμS & Sentinel-2





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

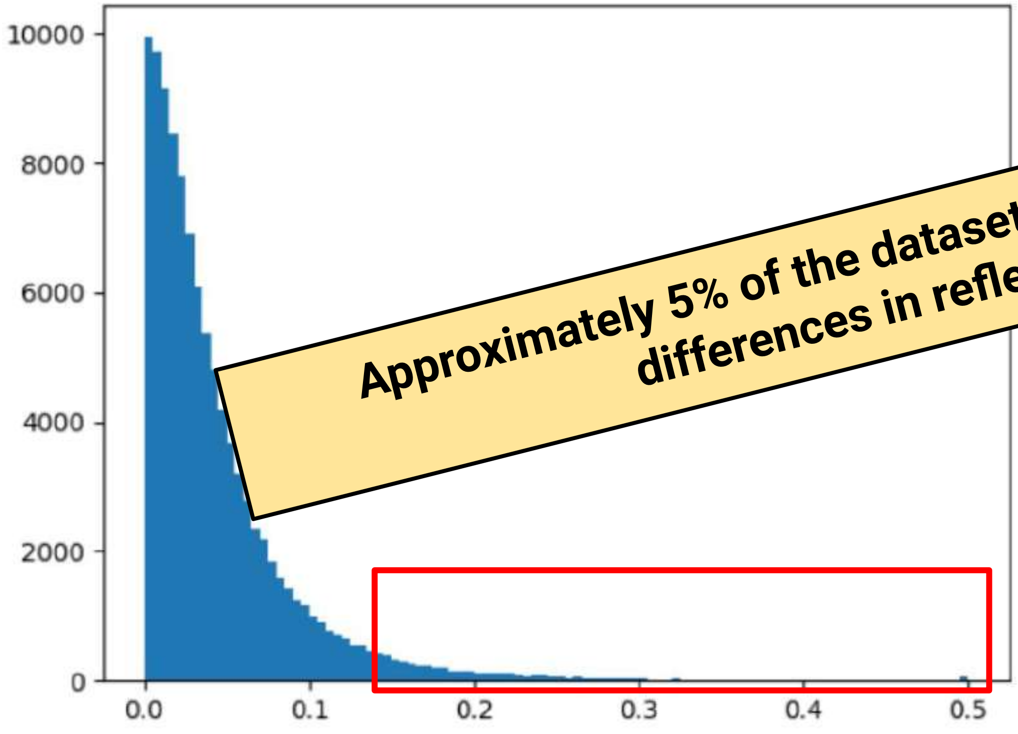
$$W \in R^{5\times 4}$$

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

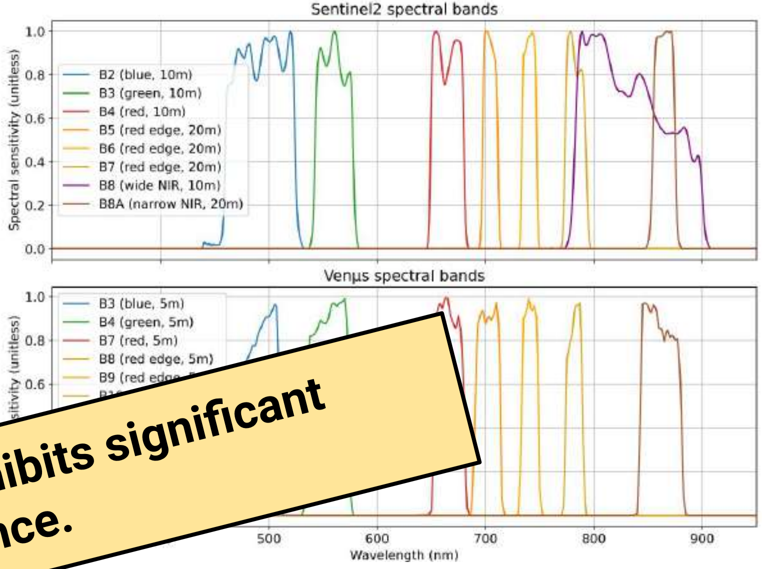
$$S = \begin{bmatrix} \rho_1^{sentinel2,b2} & \rho_1^{sentinel2,b3} & \rho_1^{sentinel2,b4} & \rho_1^{sentinel2,b8} \\ \rho_2^{sentinel2,b2} & \rho_2^{sentinel2,b3} & \rho_2^{sentinel2,b4} & \rho_2^{sentinel2,b8} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_n^{sentinel2,b2} & \rho_n^{sentinel2,b3} &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

Harmonization

L1 norm ↓VENμS & Sentinel-2



Approximately 5% of the dataset exhibits significant differences in reflectance.



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

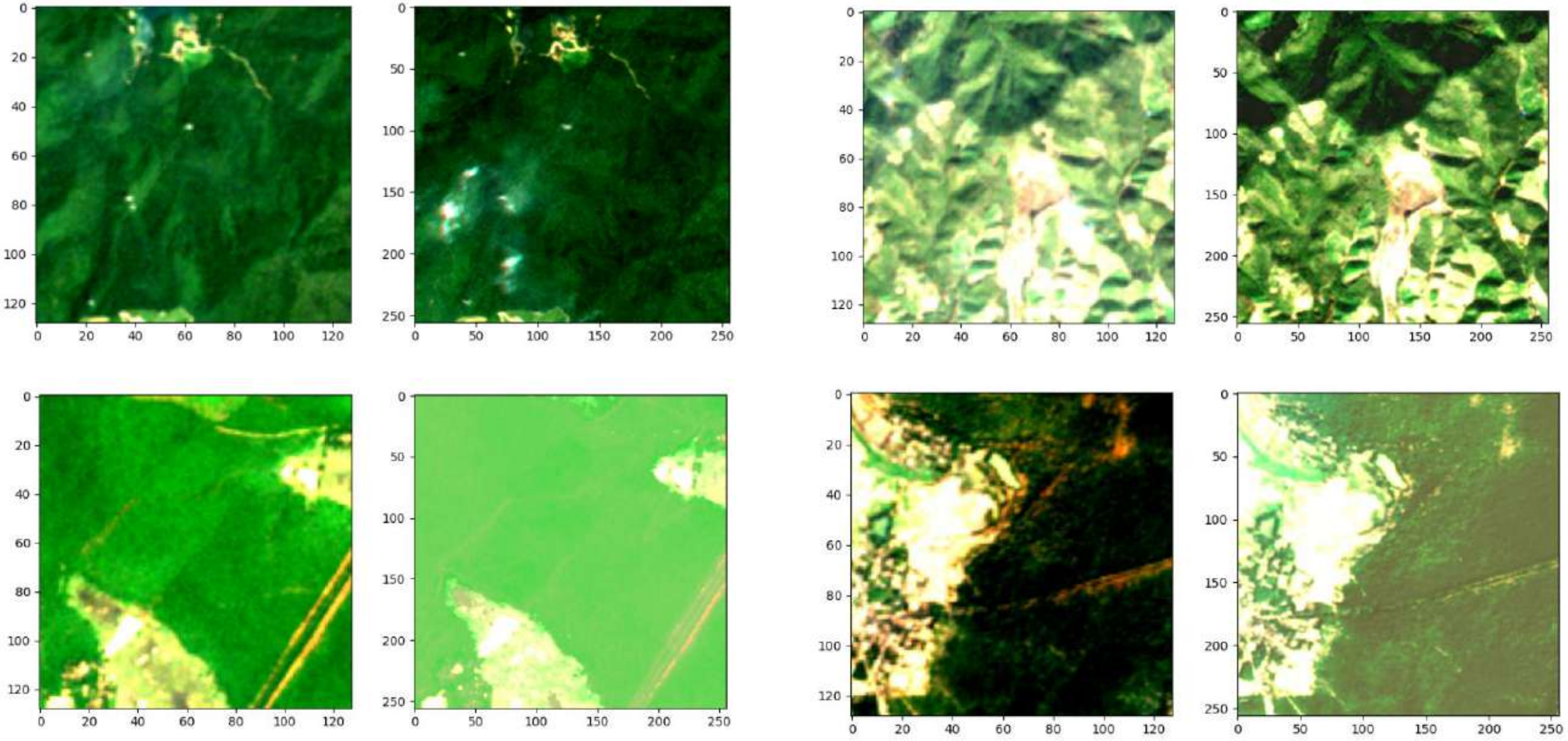
$$W \in R^{5\times 4}$$

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

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

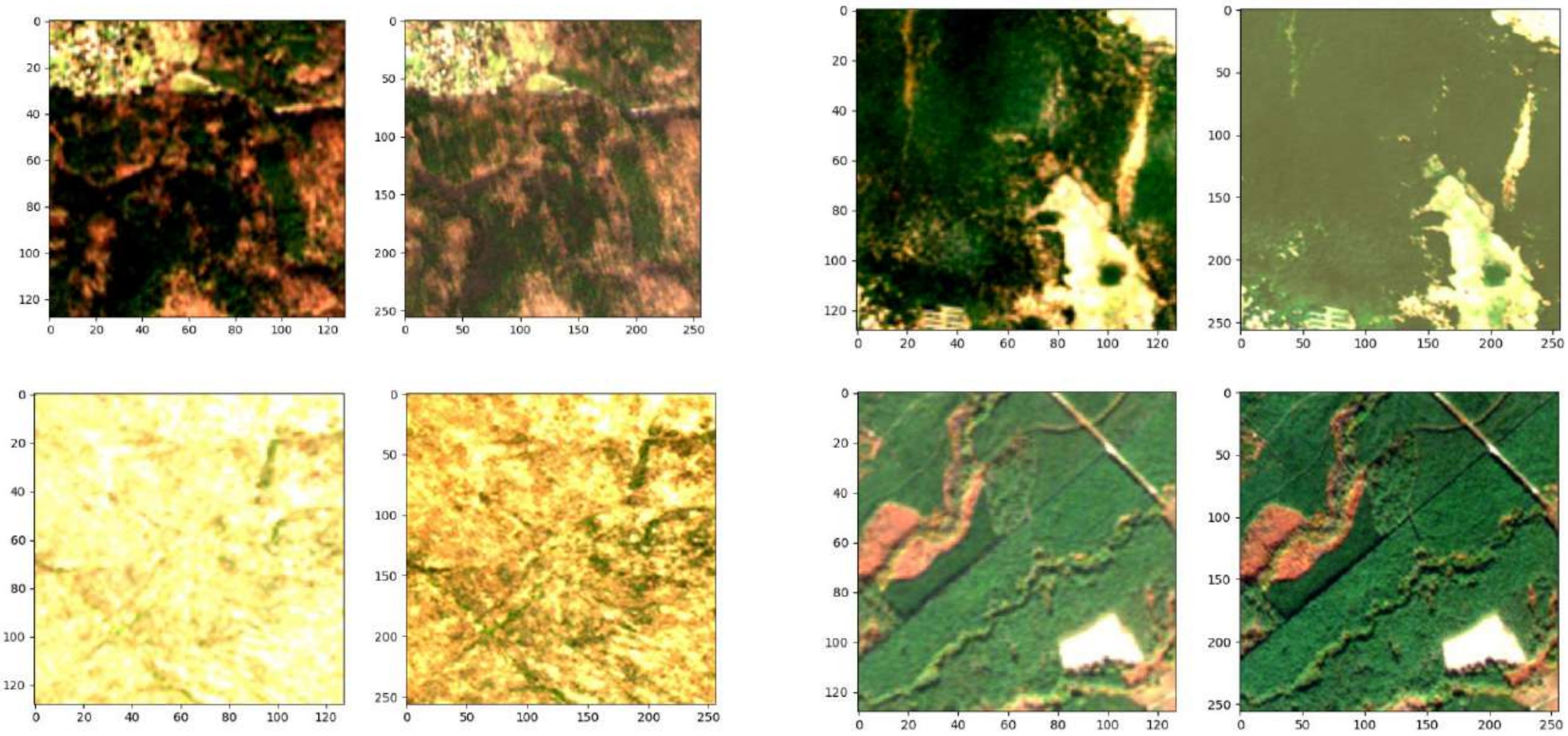
Harmonization

Sentinel-2 | **VENμS** | **Sentinel-2** | **VENμS**



Harmonization

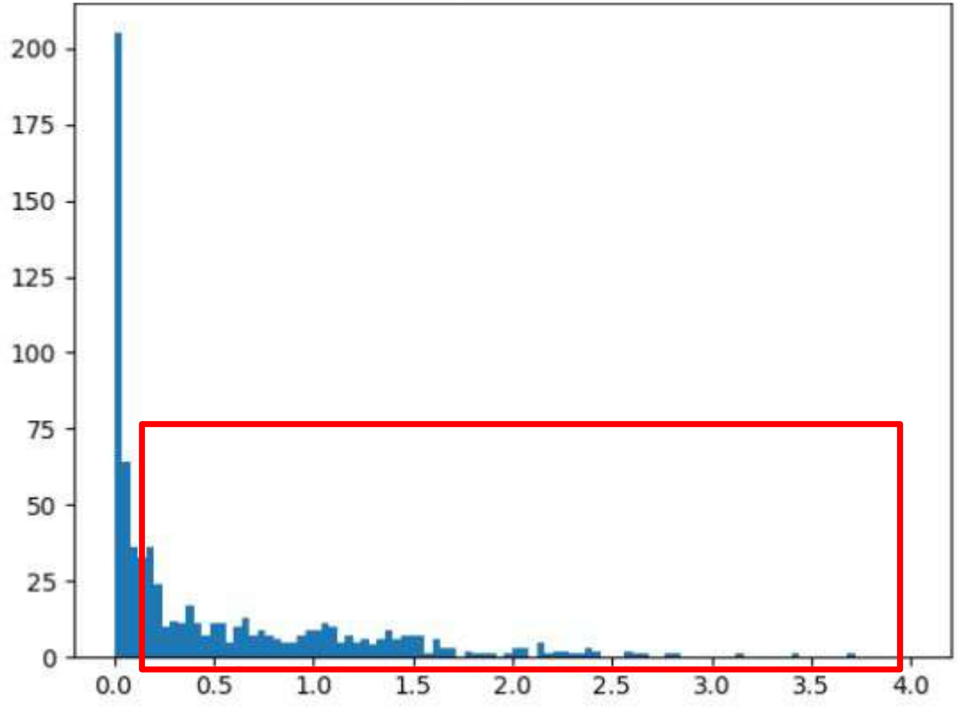
Sentinel-2 **VENμS** **Sentinel-2** **VENμS**

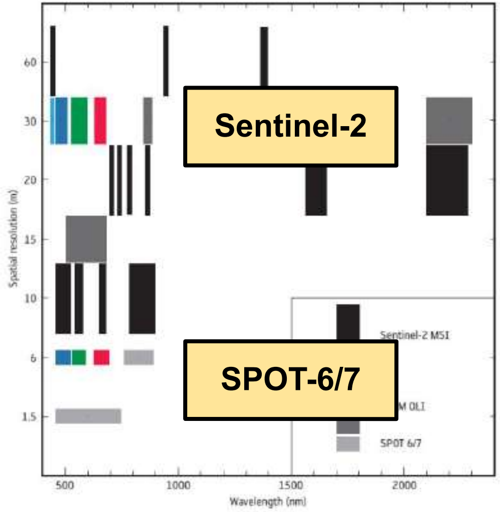


Harmonization

L1 norm

SPOT & Sentinel-2

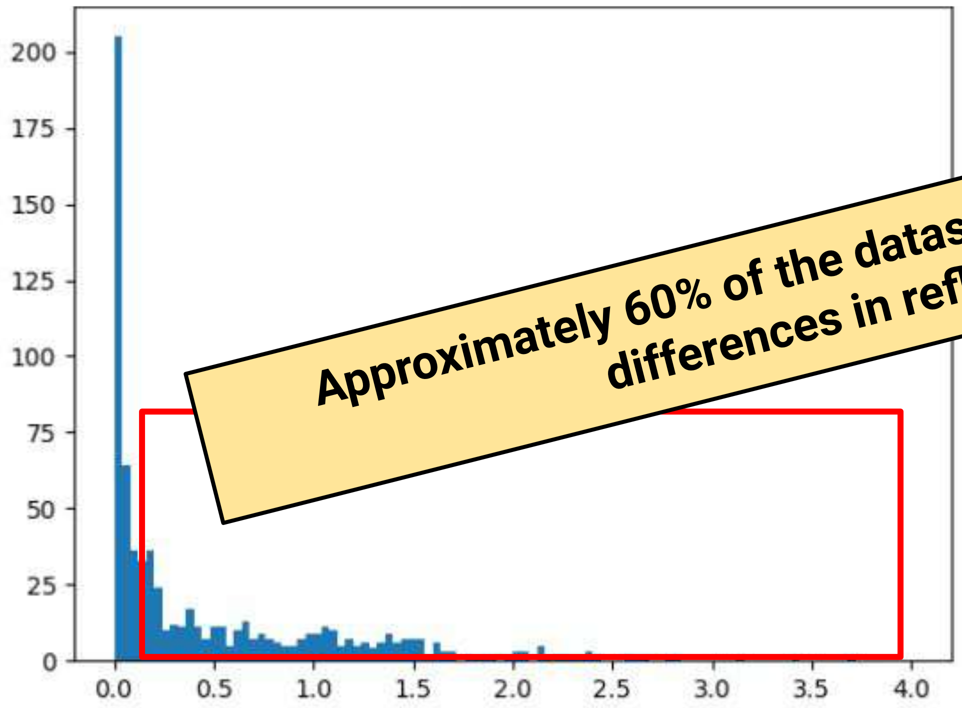




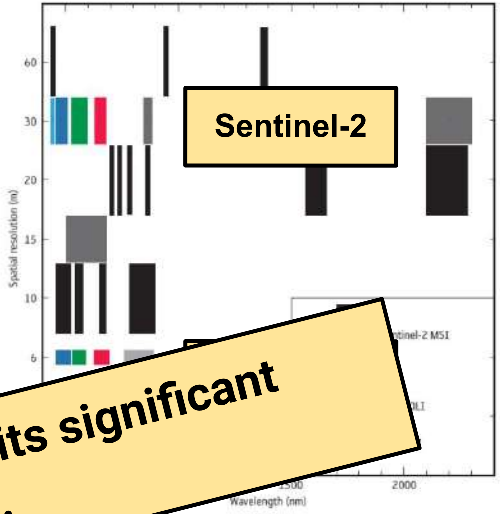
Harmonization

L1 norm

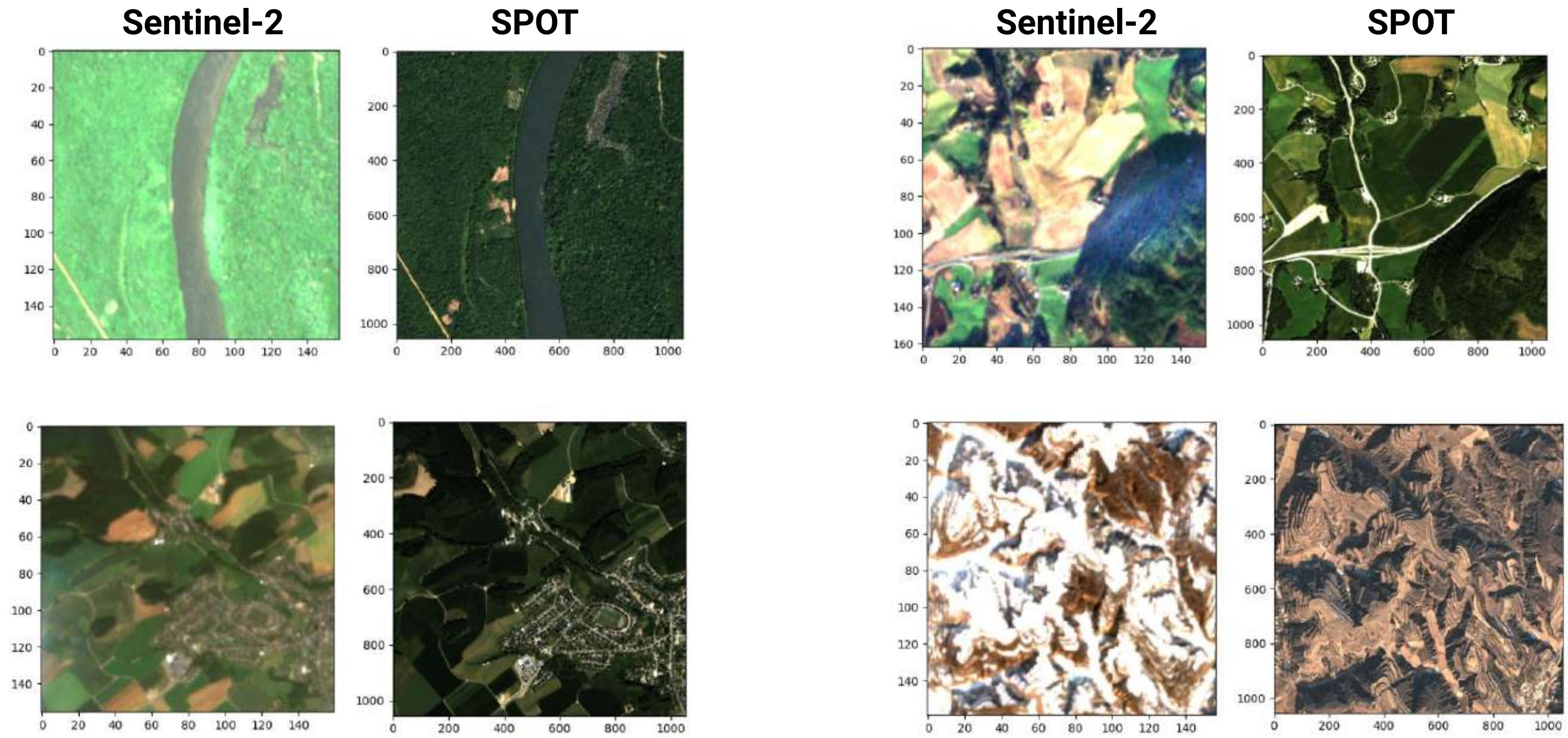
SPOT & Sentinel-2



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



Harmonization



TakeAway

We are dedicating significant effort to creating a large dataset, as we have observed that **model performance significantly improves with increased data and sufficient model parameters.** This observation aligns with findings reported in the [Wolters et al. 2024](#).

However, apart from NAIP, there are no other open, large, and **well-distributed** datasets available that match the Sentinel-2 mission period.

TakeAway

Given its consistently high quality, the **NAIP** image is an excellent starting point to explore the training of x4 SR models.

However, it will **make your results look like they are from the USA everywhere**.

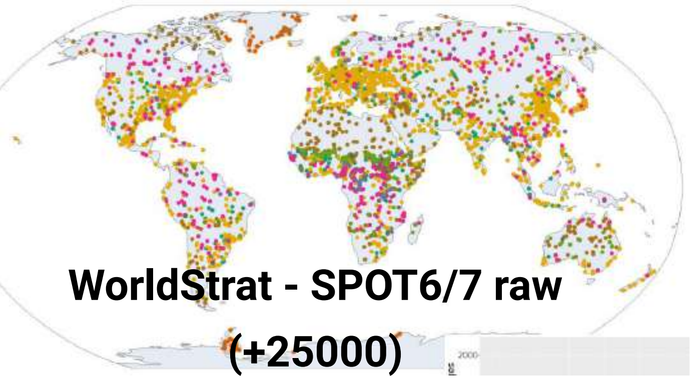
TakeAway

If you want to create super-resolution real-world models using synthetic data, **learning the degradation process is as important as learning the super-resolution process**.

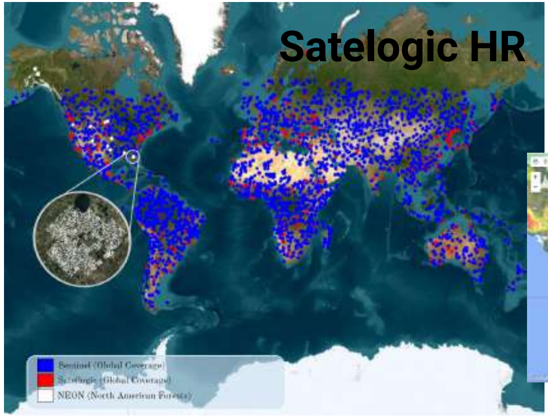
We made a **mistake by being too strict with the filters when creating the cross-sensor dataset**. A more robust harmonization model would benefit from including more data points.

More Data

WorldView 2 European Cities*
(+25000)



WorldStrat - SPOT6/7 raw
(+25000)



Open Aerial Map Subset (OAM)

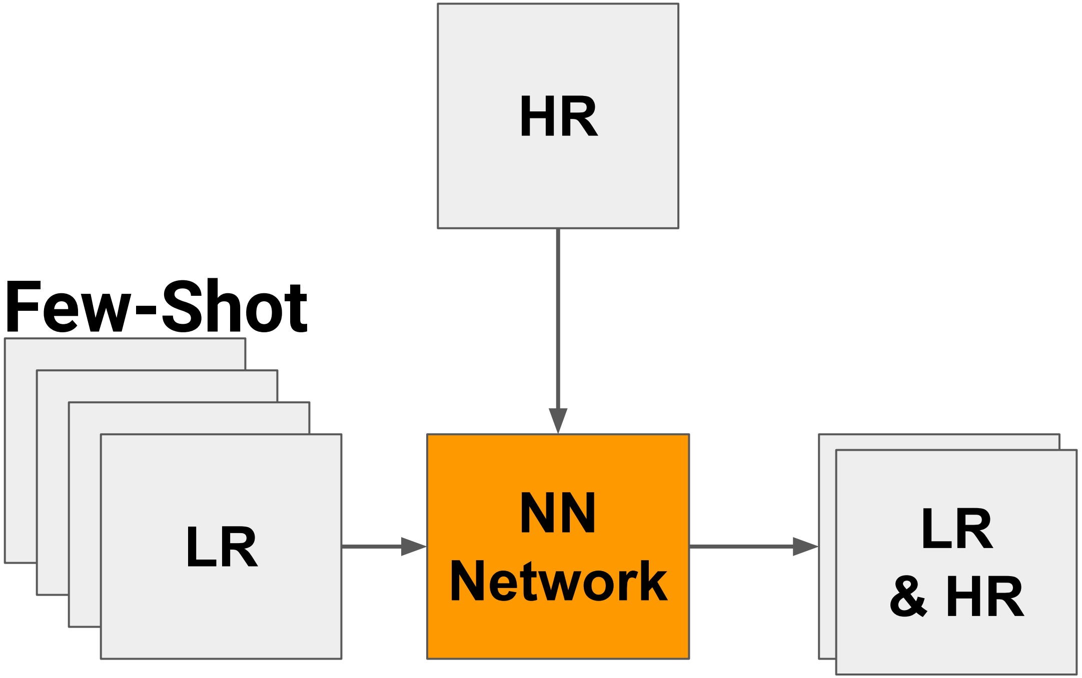
Better Spatial Alignment

Original S2 Cube

Aligned S2 Cube Date: 20151005

https://pypi.org/project/satalign/

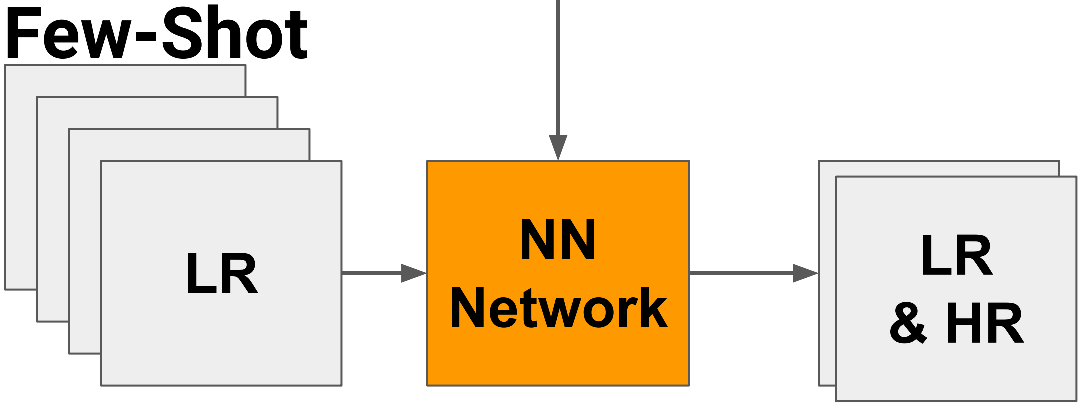
Towards BetteR Degradation Models



S2

HR

Towards BetteR Degradation Models




S2


HR