RESEARCH PRESENTATION

Performance Bounds for Reliability and Hallucination Risk in Remote Sensing Super Resolution

Remote Sensing | Super Resolution | Trustworthy AI

David Iagaru, DLR – German Aerospace Center, Ecole Polytechnique
Jakob Gawlikowski, DLR – German Aerospace Center, TUM
Nina Gottschling, Oak Ridge National Laboratory
Ioannis Sgouralis, University of Tennessee Knoxville
Anders C. Hansen, University of Cambridge, UK
Ben Adcock, Simon Fraser University
Vegard Antun, University of Oslo

DLR

Inverse problem framework :

$F : \mathcal{X} \times \mathcal{E} \to \mathcal{Y}$
$(x, e) \mapsto F(x, e)$

$\mathcal{M}_1 \subset \mathcal{X}, \mathcal{M}_2^{\mathcal{E}} = F(\mathcal{M}_1, \mathcal{E})$

Given a measurement $y \in \mathcal{M}_2^{\mathcal{E}}$
Recover $x \in \mathcal{M}_1$ *st* $\mathrm{F(x, e)} = y$ for some $e \in \mathcal{E}$

Model $\phi : \mathcal{Y} \to \mathcal{X}$ (DNN, Optimization based, …)

DNN are universal approximators. But are there problem inherent performance bounds ?

DLR

Inverse problem framework :

$F : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{Y}$

$(x, e) \mapsto F(x, e)$

$\mathcal{M}_1 \subset \mathcal{X}, \mathcal{M}_2^{\mathcal{E}} = F(\mathcal{M}_1, \mathcal{E})$

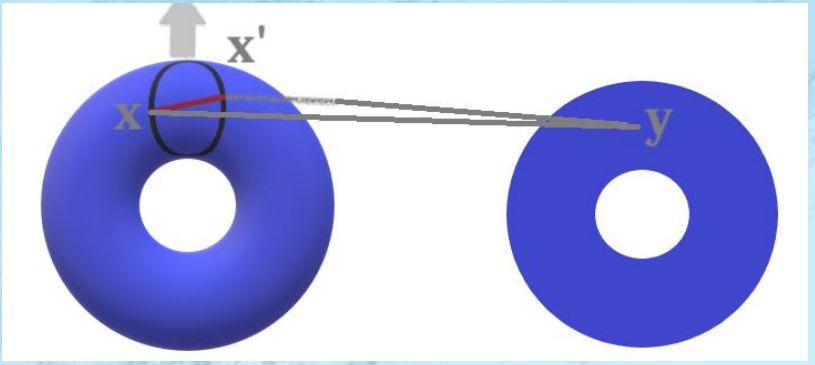
Given a measurement $y \in \mathcal{M}_2^{\mathcal{E}}$

Recover $x \in \mathcal{M}_1$ *st* $\mathrm{F(x, e)} = y$ for some $e \in \mathcal{E}$

Model $\phi : \mathcal{Y} \rightarrow \mathcal{X}$ (DNN, Optimization based, …)

DNN are universal approximators. But are there problem inherent performance bounds ?

Yes !



$F_y = \{ x \in \mathcal{M}_1 : \exists e \in \mathcal{E} \; s.t \; F(x, e) = y \}$

DLR

Dataset specific performance bounds

Consider a dataset $(x_m, y_m)_{1 \leq m \leq M} \in \left(\mathcal{M}_1 \times \mathcal{M}_2^{\mathcal{E}}\right)^M$

Re-indexed $\left(x_{k,n}, y_k\right)_{1 \leq k \leq K,\, 1 \leq n \leq N(k)}$

With $F_{y_k}^{N(k)} = \{x_m \quad \in F_{y_k}\} = \{x_{k,n} 1 \leq n \leq N(k)\}$ being the finite approximate feasible set of y_k for $1 \leq k \leq K$

And $M = \sum_{k=1}^{K} N(k)$

DLR

Dataset specific performance bounds

Consider a dataset $(x_m, y_m)_{1 \le m \le M} \in \left(\mathcal{M}_1 \times \mathcal{M}_2^{\mathcal{E}}\right)^M$

Re-indexed $\left(x_{k,n}, y_k\right)_{1 \le k \le K,\, 1 \le n \le N(k)}$

With $F_{y_k}^{N(k)} = \left\{x_m \quad \in F_{y_k}\right\} = \{x_{k,n} 1 \le n \le N(k)\}$ being the finite approximate feasible set of y_k for $1 \le k \le K$

And $M = \sum_{k=1}^{K} N(k)$

Define the average error as $\mathcal{L}(\mathcal{D}_M, \phi, p) = \left(\frac{1}{M} \sum_{m=1}^{M} d_X\left(x_m, \phi(y_m)\right)^p\right)^{\frac{1}{p}}$

Can $\mathcal{L}(\mathcal{D}_M, \phi, p)$ always be arbitrarily small ?

DLR

Dataset specific performance bounds

Consider a dataset $(x_m, y_m)_{1 \le m \le M} \in \left(\mathcal{M}_1 \times \mathcal{M}_2^{\mathcal{E}}\right)^M$

Re-indexed $\left(x_{k,n}, y_k\right)_{1 \le k \le K,\, 1 \le n \le N(k)}$

With $F_{y_k}^{N(k)} = \left\{x_m \quad \in F_{y_k}\right\} = \{x_{k,n} 1 \le n \le N(k)\}$ being the finite approximate feasible set of y_k
for $1 \le k \le K$
And $M = \sum_{k=1}^{K} N(k)$

Define the average error as $\mathcal{L}(\mathcal{D}_M, \phi, p) = \left(\frac{1}{M}\sum_{m=1}^{M} d_{\mathcal{X}}\left(x_m, \phi(y_m)\right)^p\right)^{\frac{1}{p}}$

Define the Kernel size as

$$Kersize(F, \mathcal{M}_1, \mathcal{E}, p)_K = \left(\frac{1}{M}\sum_{k=1}^{K} \frac{2}{N(k)-1} \sum_{1 \le n < n' \le N(K)} d_{\mathcal{X}}(x_{k,n}, x_{k,n'})\right)^{\frac{1}{p}}$$

DLR

Dataset specific performance bounds

Define the average error as $\mathcal{L}(\mathcal{D}_M, \phi, p) = \left(\frac{1}{M}\sum_{m=1}^{M} d_{\mathcal{X}}\left(x_m, \phi(y_m)\right)^p\right)^{\frac{1}{p}}$

Define the Kernel size as $Kersize(F, \mathcal{M}_1, \mathcal{E}, p)_K = \left(\frac{1}{M}\sum_{k=1}^{K} \frac{2}{N(k)-1}\sum_{1\leq n<n'\leq N(K)} d_{\mathcal{X}}(x_{k,n}, x_{k,n'})\right)^{\frac{1}{p}}$

Then, the average error is lower bounded :

$$\frac{1}{2}Kersize(F, \mathcal{M}_1, \mathcal{E}, p)_K \leq \mathcal{L}(\mathcal{D}_M, \phi, p)$$

- Dataset specific bound
- Forward problem inherent : applies to any decoder ϕ

Potential usages :

- Training early stopping
- Uncertainty Quantification

DLR

Dataset specific performance bounds : fast proxy lower bound for linear problems

For simplicity, $F(x, e) = Ax + e$

For each data point (x_m, y_m) add the symmetric $\left(x_m - 2\, P_{\mathcal{N}(A)} x_m, y_m\right) = (x'_m, y_m)$ to the dataset

$$\frac{1}{2}\left(\frac{1}{M}\sum_{m=1}^{M} d_{\mathcal{X}}(x_m, x'_m)^p\right)^{\frac{1}{p}} \leq \left(\frac{1}{M}\sum_{m=1}^{M} d_{\mathcal{X}}\left(x_m, \phi(y_m)\right)^p + d_{\mathcal{X}}\left(x_m{}', \phi(y_m)\right)^p\right)^{\frac{1}{p}}$$

$$\frac{1}{2} SKersize(F, \mathcal{M}_1, \mathcal{E}, p)_K \leq \mathcal{L}(\mathcal{D}_M, \phi, p)$$

- Cheap computation
- Dataset specific
- Forward problem inherent : applies to any decoder ϕ

Potential usages :

- Training early stopping
- Uncertainty Quantification

DLR

Super resolution of Satellite images

High resolution (HR) : $\mathcal{M}_1 \subset \mathcal{X} = \mathbb{R}^{4\times768\times768}$ (NAIP datset, 1m Ground Sampling Distance)
Low resolution (LR) : $\mathcal{M}_2 \subset \mathcal{Y} = \mathbb{R}^{4\times768/S\times768/S}$ (S m GSD)
Scale : $S \in \{2,3,4,6,8\}$



F

Gaussian blurring
Bilinear Downsampling
+
Gaussian noise



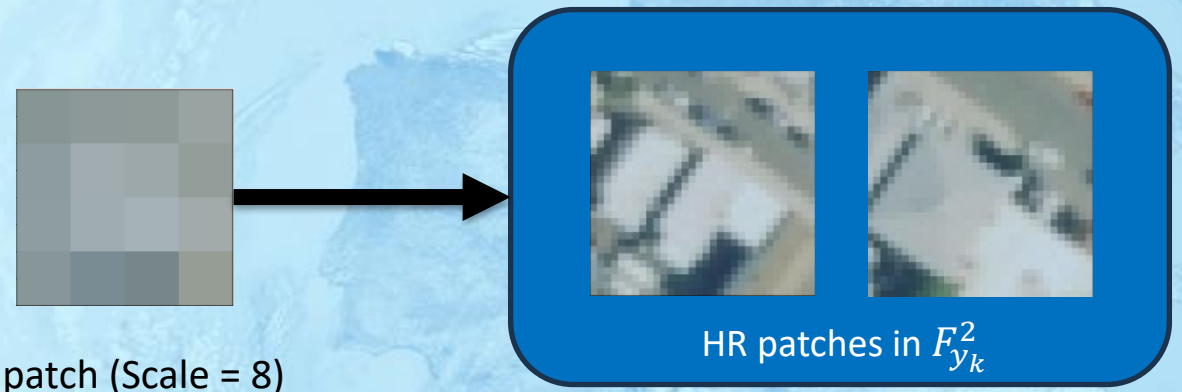
Evaluation metrics :

- L1
- Besov norm (s=1, p=1, q=2, on db4 wavelet coefficients)
- SSIM distance [3]

DLR

Super Resolution of satellite images

Determining $F_{y_k}^{N(k)}$ at detail level : split the HR images into patches of size $4 \times 32 \times 32$



LR patch (Scale = 8)

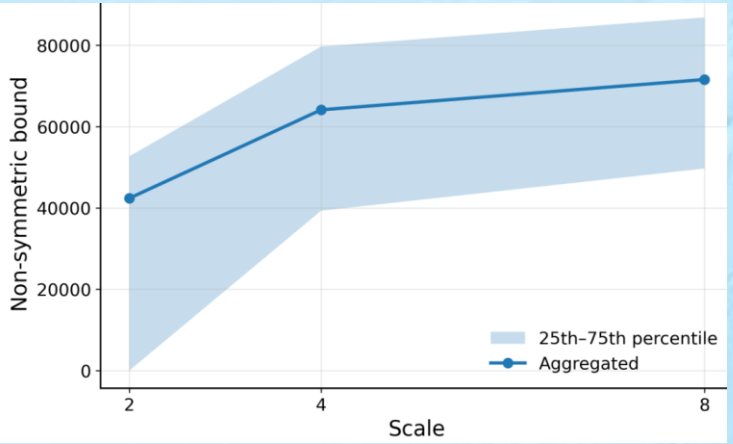
y_k

Computing bounds patchwise and aggregating over each image

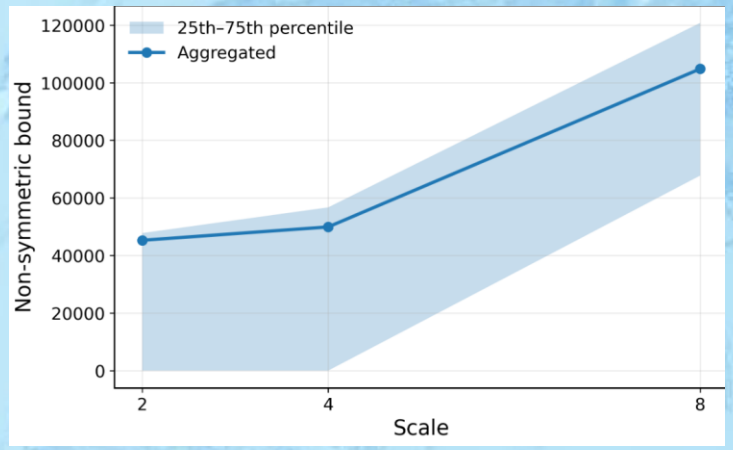
DLR

Kernel size vs Super Resolution scale

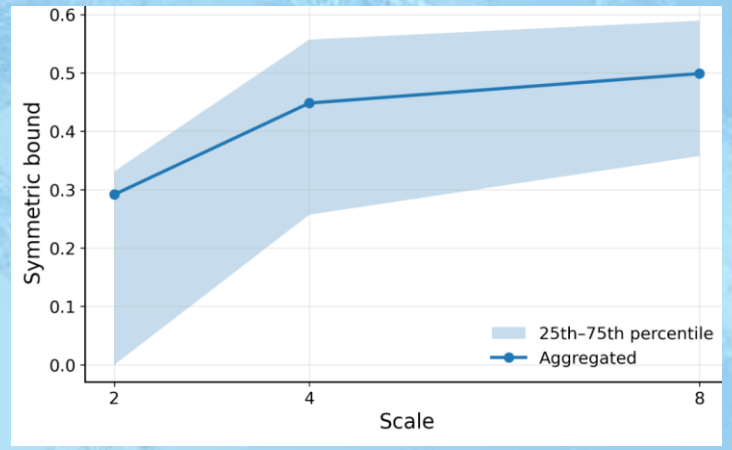
For which scales can satellite images be reliably super-resolved ?



Besov norm
(s=p=1, q=2, db4 wavelets)



L1 distance



SSIM distance

The Performance limit increases as the scale goes up

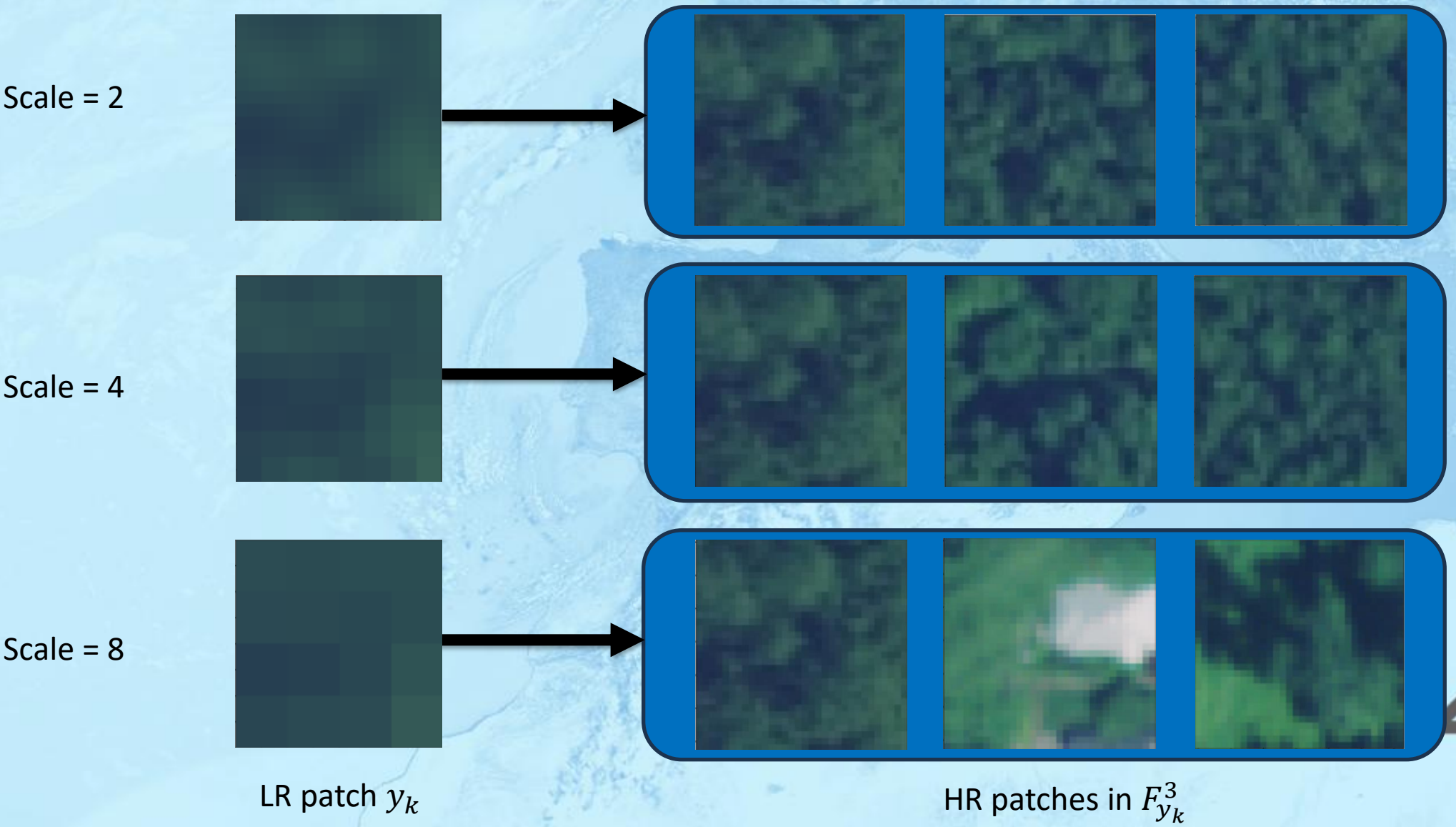
When is it still viable ?

DLR

Kernel size vs Super Resolution scale

The Performance limit increases as the scale goes up

When is it still viable ?

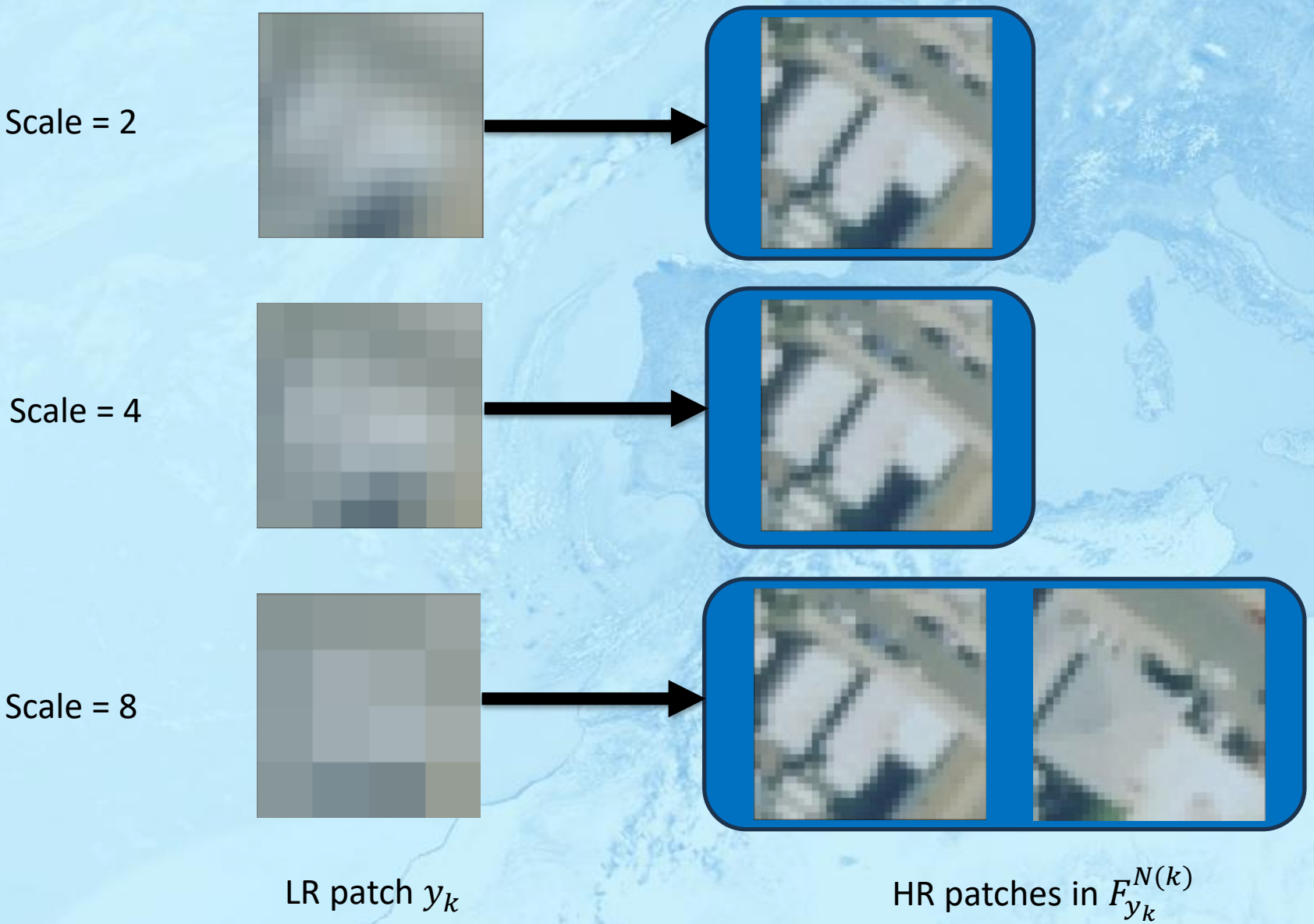


DLR

Kernel size vs Super Resolution scale

The Performance limit increases as the scale goes up

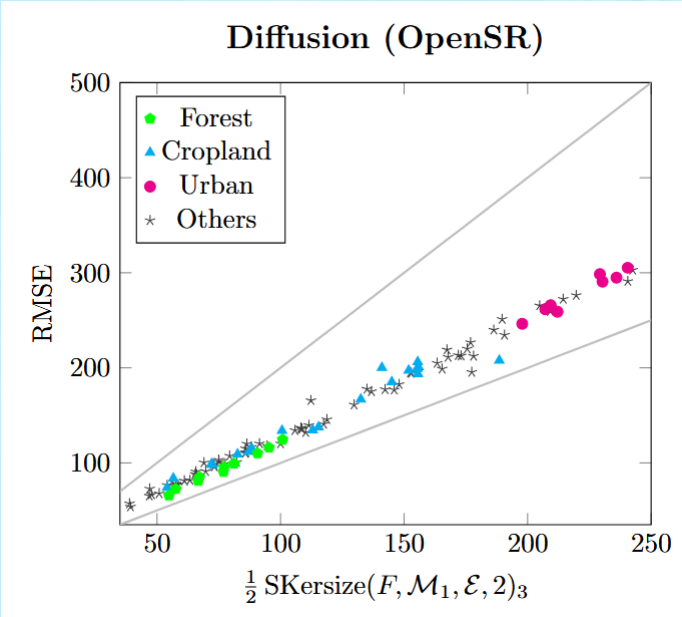
When is it still viable ?

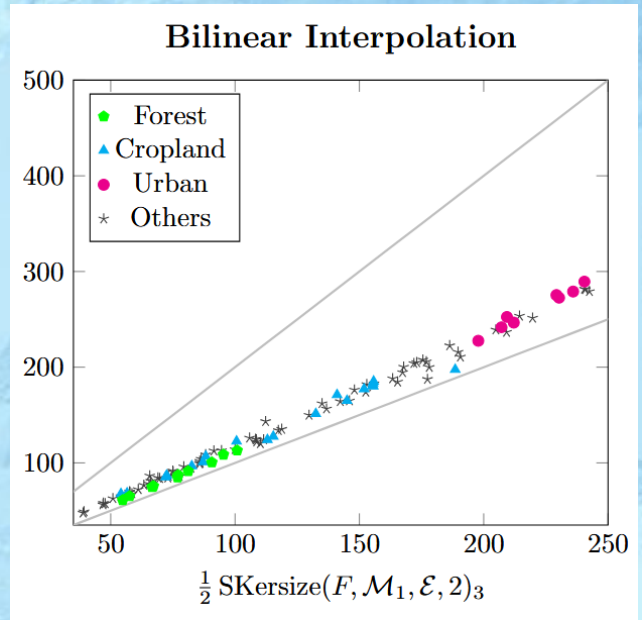


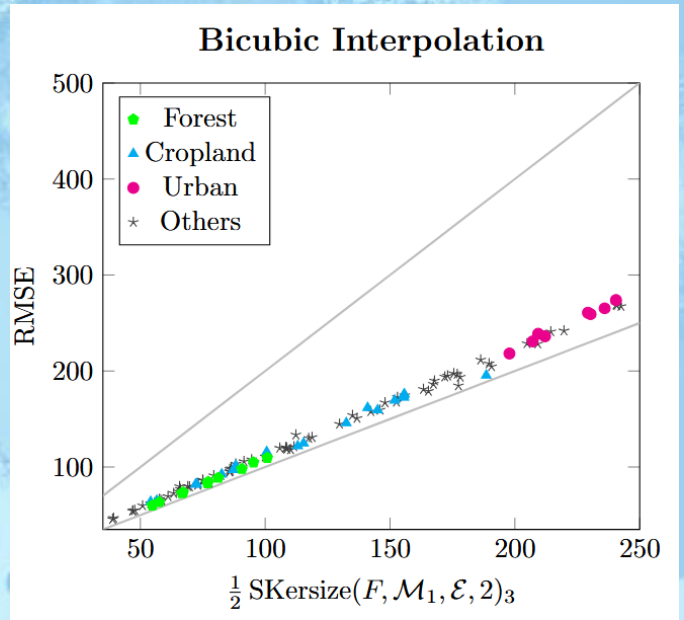
DLR

Sharpness of the performance bounds across land cover classes

Super-Resolution scale fixed to 4







- The lower bound is sharp in practice
- The problem difficulty highly depends on the considered distribution ($\mathcal{M}_1$)

DLR

Kernel size compared to other Uncertainty Quantification metrics

Metric	Patch-wise		Pixel-wise	
	Pearson r	R^2	Pearson r	R^2
Absolute Error	0.865	0.754	0.599	0.364
Hallucination [2]	0.624	0.427	0.128	0.024
Kernel projection HR	0.784	0.628	0.234	0.058

Comparison (correlation and coefficient of determination R^2) of the Std of ensemble prediction with other uncertainty maps: pixel-wise vs patch-wise

Metric	Pearson r	R^2
Absolute Error	0.956	0.915
Hallucination [2]	0.317	0.173
Kernel Projection HR	0.775	0.614

Patch-wise correlation and coefficient of determination R^2 between the Std of ensemble prediction obtained via Test Time augmentation and other uncertainty maps

By Riccardo Aliprandi, Giacomo Vescovi, Ming Yuan Lee, Timo Marek Weiss, Maximilian Wagner (TUM DI Lab)

DLR

References

[1] Gottschling et al. Average Kernelsizes – Computable Sharp Accuracy Bounds for inverse problems. 2025.

[2] Aybar et al. - A Comprehensive Benchmark for Optical Remote Sensing Image Super-Resolution. IEEE Geoscience and Remote Sensing Letters (Volume: 21). 2024.

[3]Brunet et al. On the Mathematical Properties of the Structural Similarity Index. IEEE Transactions on Image Processing (Volume: 21, Issue: 4, April 2012)