ILLINOIS

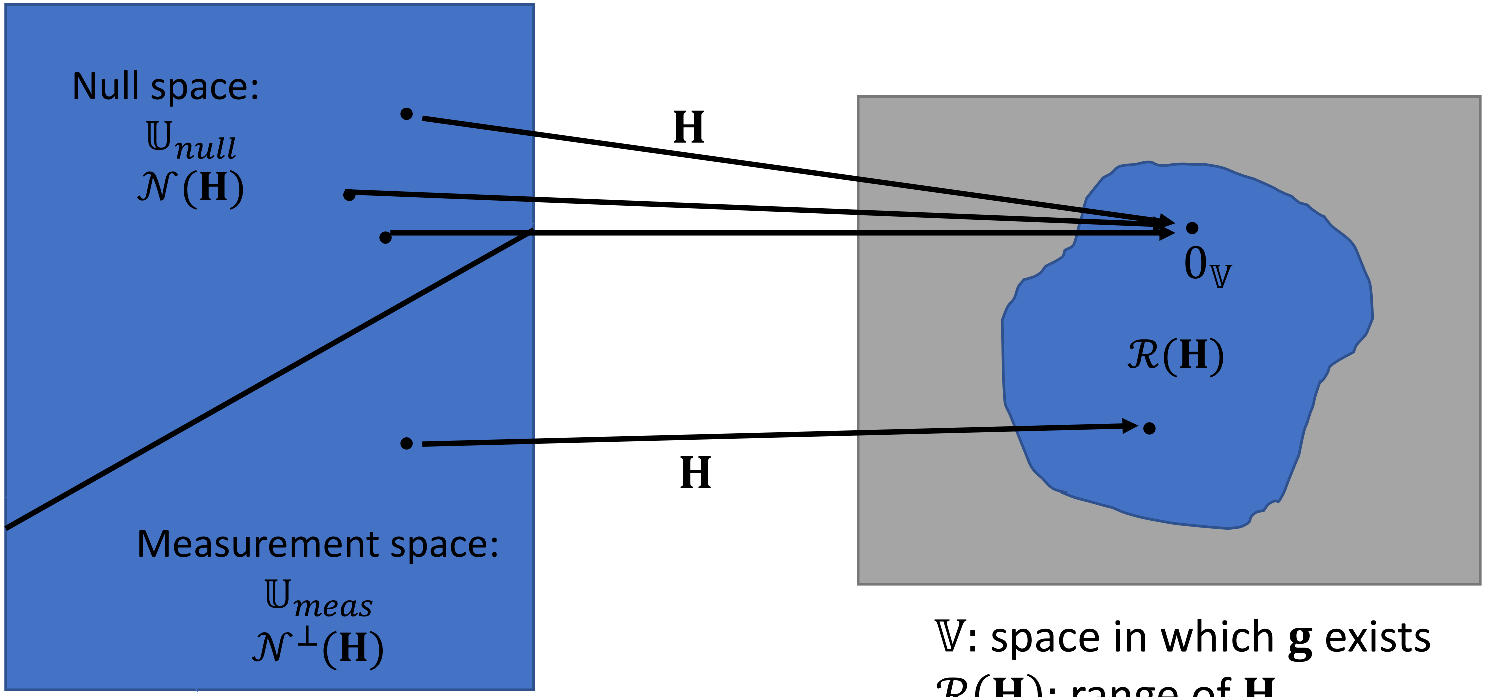
Bioengineering Department
GRAINGER COLLEGE OF ENGINEERING

BIOE 580: Foundations of Imaging Science
Lecture 13: Hallucinations in Image Reconstruction

Sayantan Bhadra

October 6, 2021

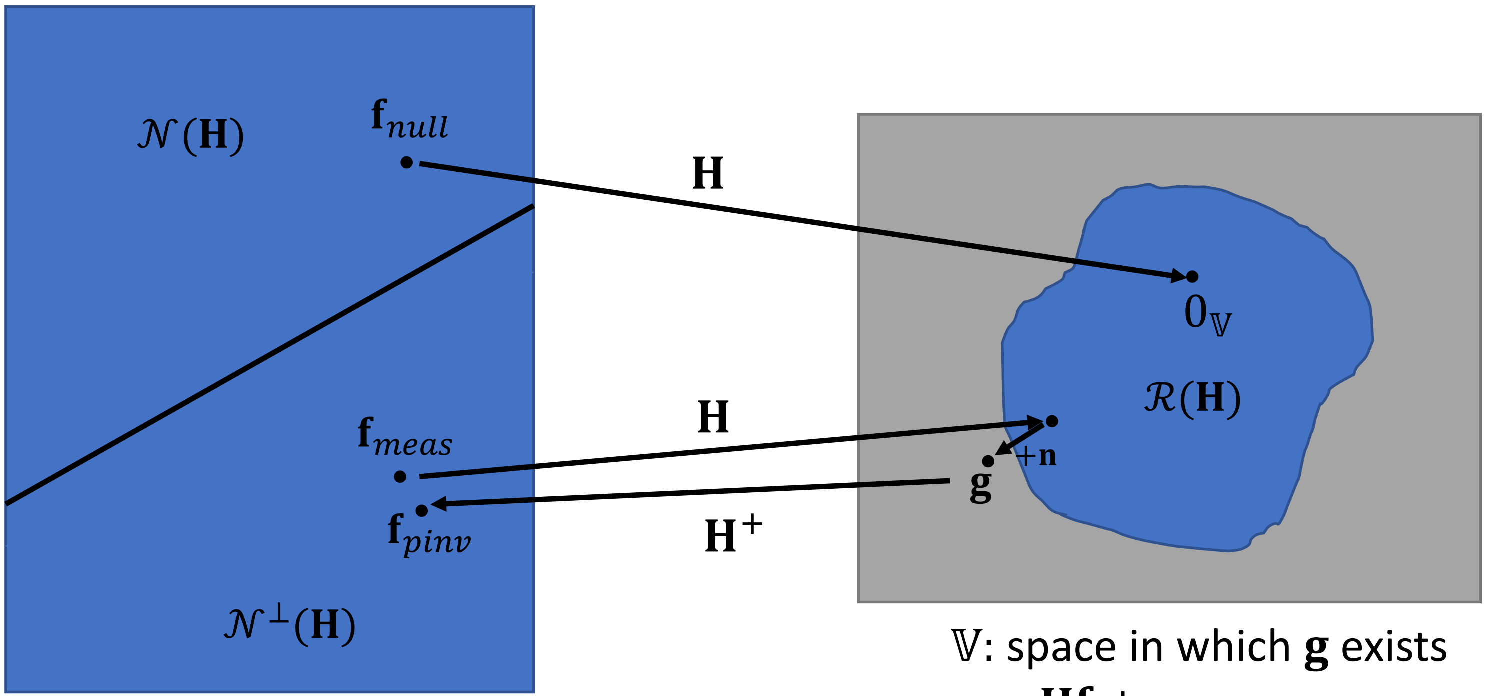
Measurement and null space



$\mathbb{U}$: domain of $\mathbf{H}$

$\mathbb{V}$: space in which $\mathbf{g}$ exists
$\mathcal{R}(\mathbf{H})$: range of $\mathbf{H}$

Measurement and null components



$\mathbb{U}$: domain of $\mathbf{H}$

$\mathbf{f} = \mathbf{f}_{meas} + \mathbf{f}_{null}$

$\mathbb{V}$: space in which $\mathbf{g}$ exists

$\mathbf{g} = \mathbf{Hf} + \mathbf{n}$

$\mathbf{f}_{meas} \approx \mathbf{H}^{+}\mathbf{g} = \mathbf{f}_{pinv}$

$\mathbf{f}_{null} = ??$ ➡ need prior knowledge

Image reconstruction from incomplete measurements

- Given a measurement $\mathbf{g} \in \mathbb{E}^M$ where

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}$$

such that $\mathbf{H} \in \mathbb{E}^{M \times N}$, $\mathbf{n} \in \mathbb{E}^M$ and $\mathbf{f} \in \mathbb{E}^N$

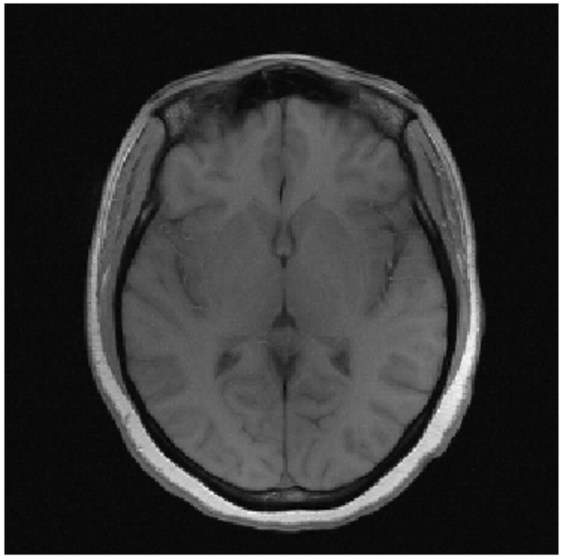
- **Goal**: estimate $\mathbf{f}$ given $M \ll N$

- Examples – limited-view X-ray CT, accelerated MRI

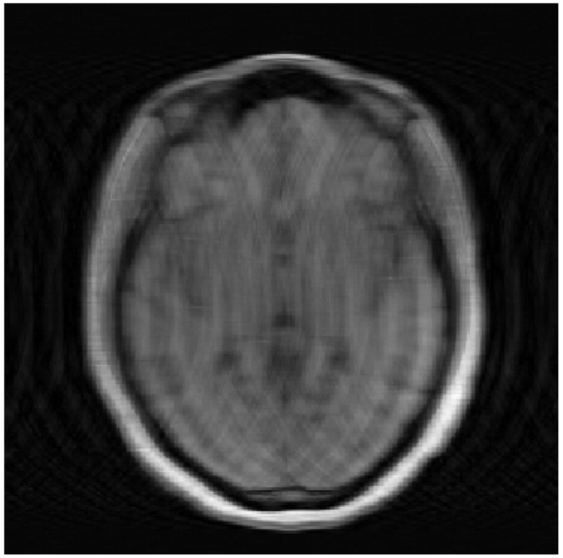
- $\mathbf{f}$ cannot be determined uniquely ➡ ill-posed inverse problem, prior knowledge needs to be imposed

Example: Accelerated MRI

True object



IFFT of zero-filled k-space



Severe aliasing artifacts due to sampling k-space at 1/3 of Nyquist rate

Definition of priors in image reconstruction

- Bayes' theorem:

likelihood

prior

$$p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{g})}$$

posterior

marginal

— $p(\mathbf{f})$: distribution over objects (***prior***)

— Common choices for $p(\mathbf{f})$: natural images are smooth or piecewise constant etc.

- "learned" priors

- A poor choice of $p(\mathbf{f})$ may lead to *false structures* in the reconstructed image

False structures due to inaccurate priors

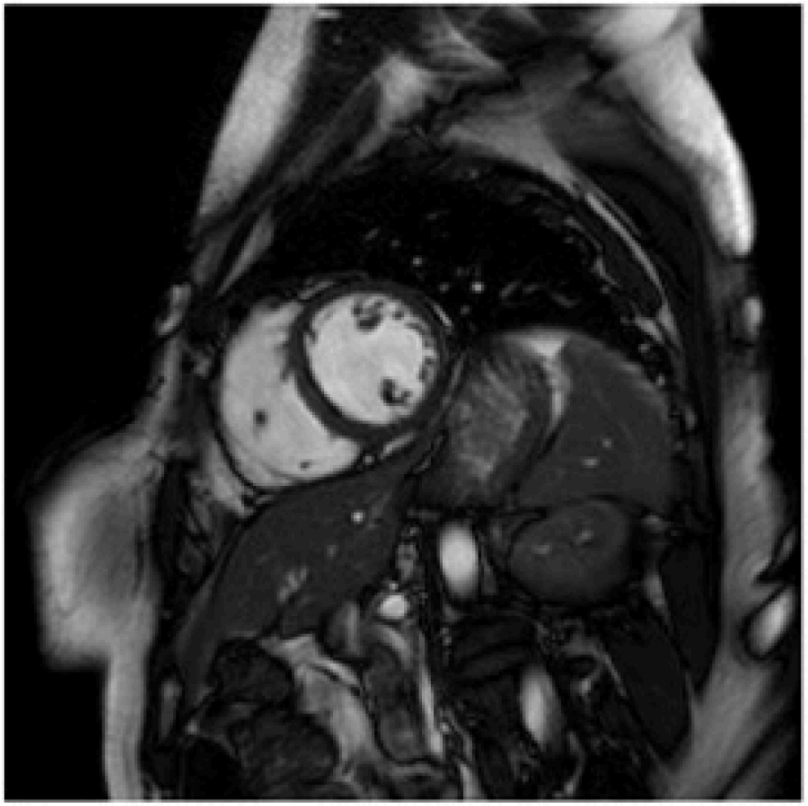
- False structures can be defined in two ways depending on its type – *false positive or false negative*

- **False positive**: Structure present in the reconstructed image $\hat{\mathbf{f}}$ but absent in the true object $\mathbf{f}$

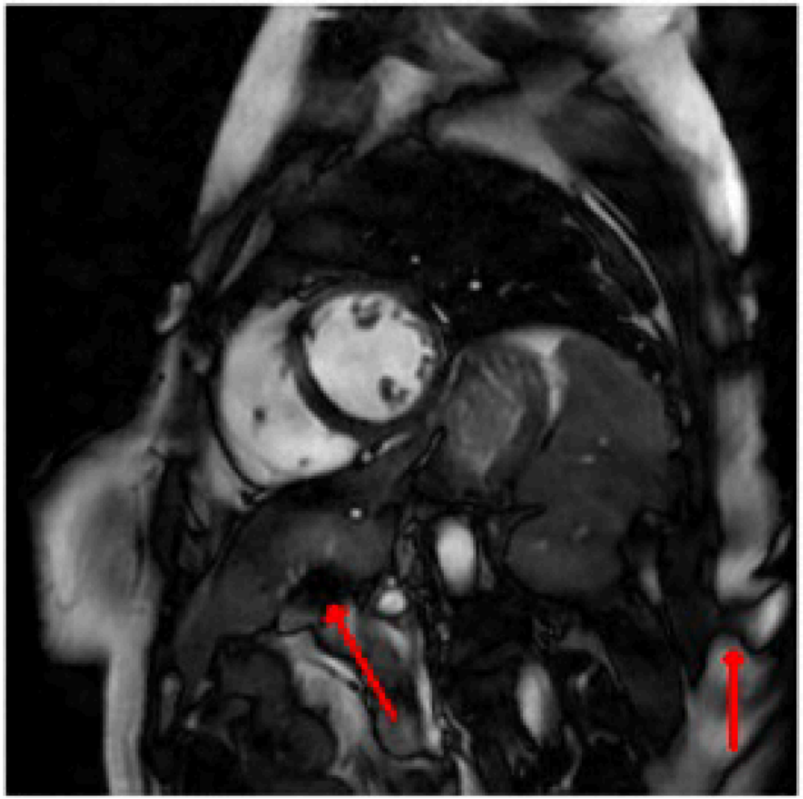
- **False negative:** Structure absent in the reconstructed image $\hat{\mathbf{f}}$ but present in the true object $\mathbf{f}$

Example: False positive structure

True object

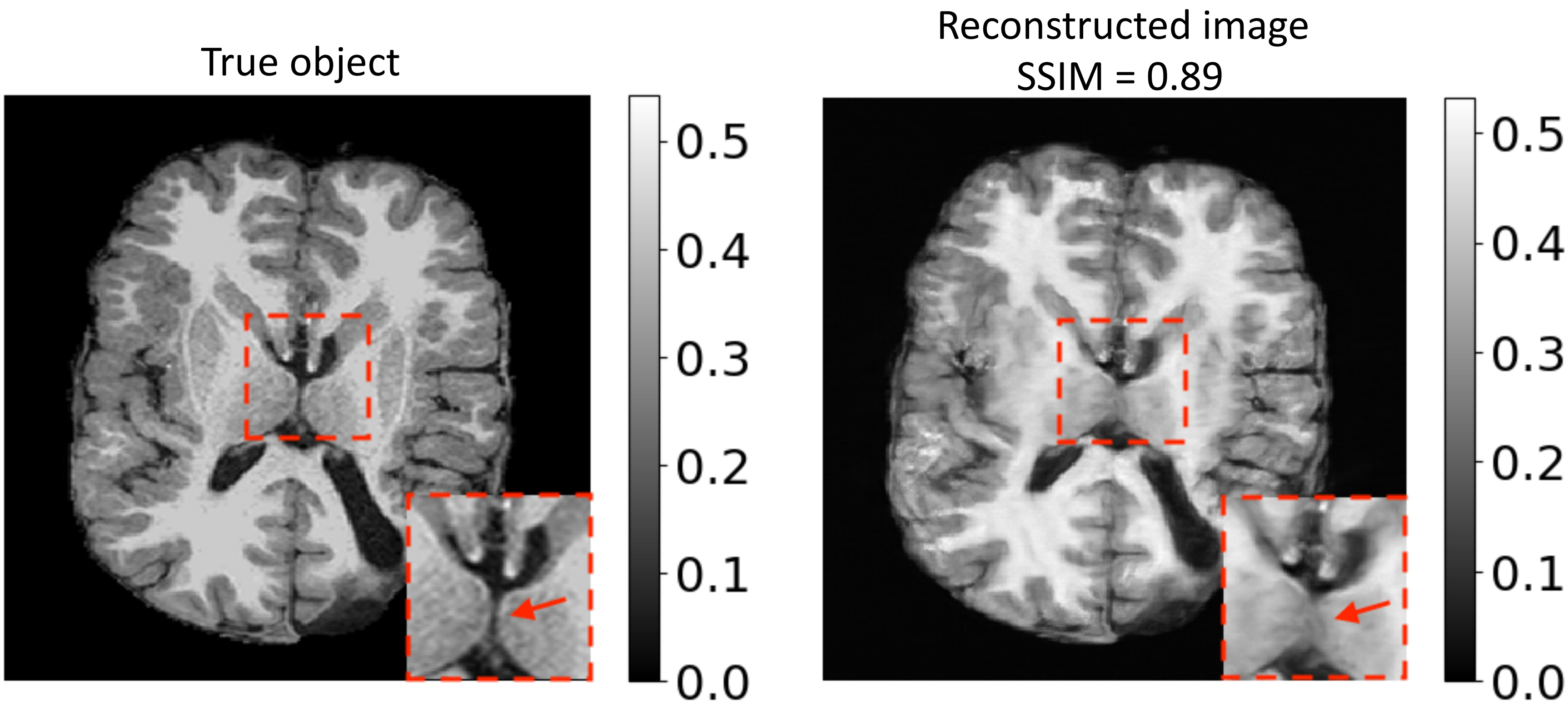


Reconstructed image



Antun *et al.* *"On instabilities of deep learning in image reconstruction – Does AI come at a cost?"* (PNAS 2020)

Example: False negative structure

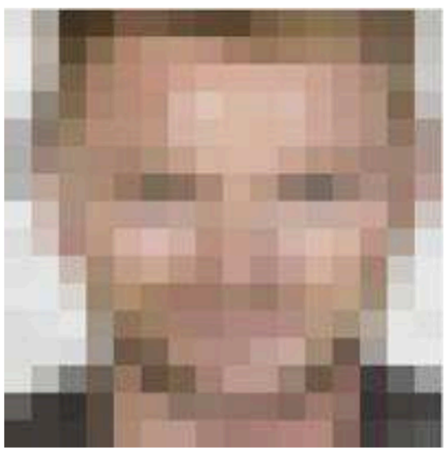


Example of a false negative structure in a reconstructed image from incomplete brain MR k-space data. SSIM is high but the third ventricle is almost missing.

Definition of hallucinations

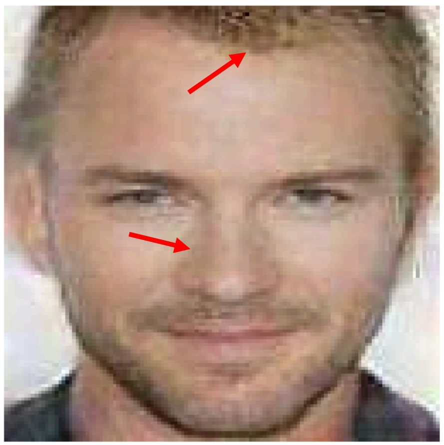
- **Definition:** False structures that occur due to the prior and cannot be produced from the measurements
- Previously introduced in image super-resolution problems

Low-resolution image



High-resolution ground truth

Super-resolved image



Xin *et al. "Super-resolving very low-resolution face images with supplementary attributes"* (CVPR 2018)

Hallucinations in tomographic imaging

- Hallucinations can exist in image reconstruction ➡ depends on the imaging operator null space characteristics

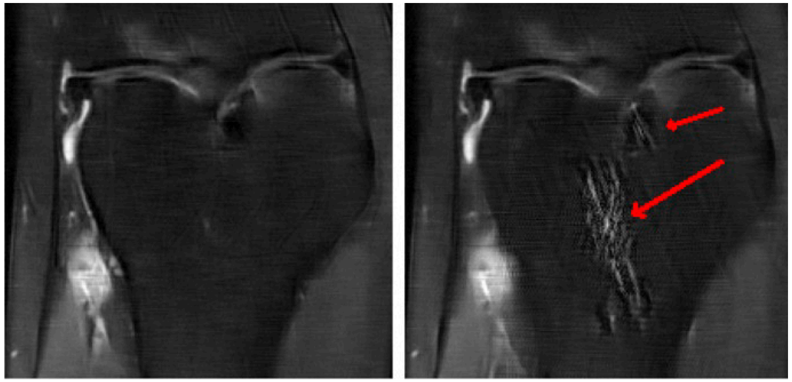
- Analyzing the presence and source of hallucinations is important for understanding the effect of the imposed prior

- Need for formal definition of hallucinations in general linear inverse problems such as image reconstruction

Hallucinations in deep learning-based methods

- **Need to analyze hallucinations when applying deep learning-based reconstruction methods:**
  - Often treated like a "black box" due to limited understanding of the mechanism of deep neural networks (*interpretability*)
  - Lack of guarantees for generalization to measurements that lie outside the training distribution (*generalizability*)
  - Small perturbations in the measurement data may cause large artifacts in the reconstructed estimate (*stability*)

$\hat{\mathbf{f}}$ from $\mathbf{g}$ $\hat{\mathbf{f}}$ from $\mathbf{g} + \boldsymbol{\delta}$



Antun *et al.* "On instabilities of deep learning in image reconstruction – Does AI come at a cost?" (PNAS 2020)

Analysis of hallucinations

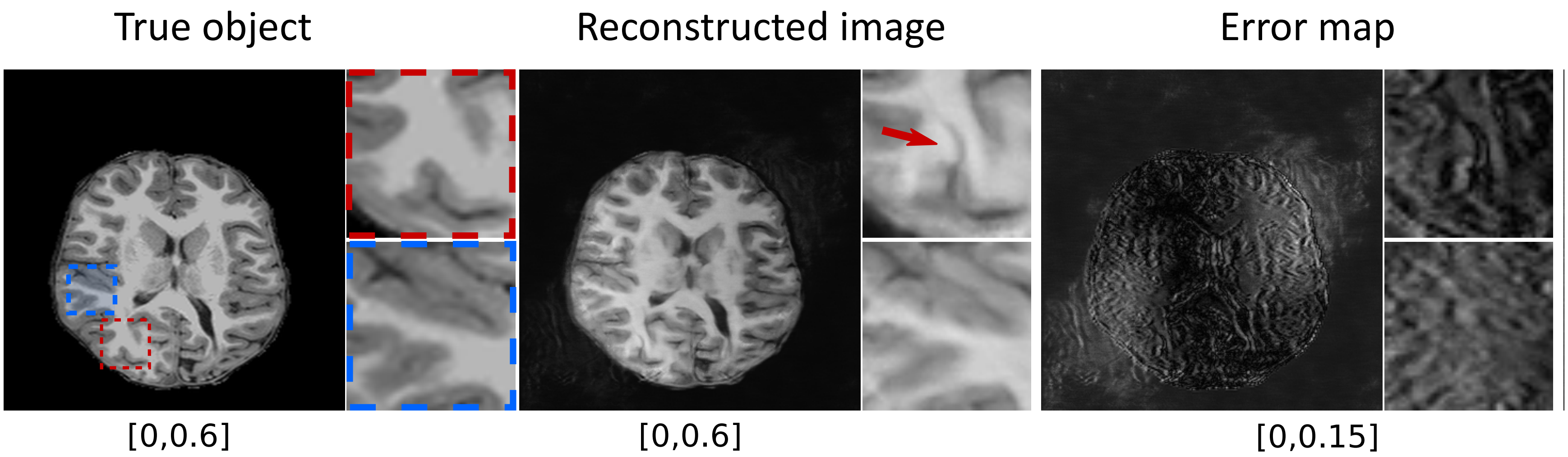
- We can compute the error map between the object and the reconstructed estimate to detect presence of false structures

- Error map contains contribution from errors due to
  — Imperfect prior
  — Data inconsistency (due to measurement noise, part of $\mathbf{g}$ may lie outside $\mathcal{R}(H)$)
  — Model mismatch

- Does not provide information about which errors represent hallucinations induced by the assumed prior

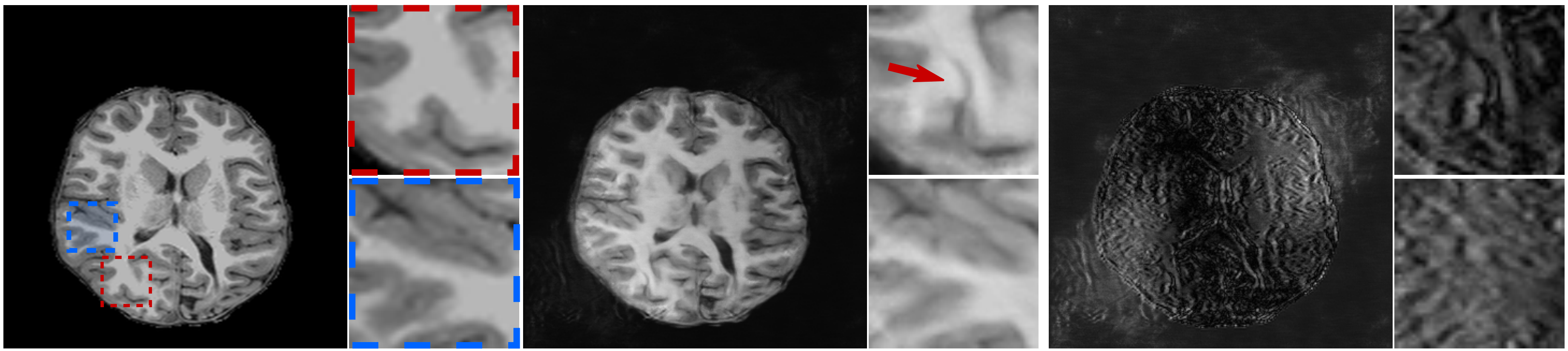
Analysis of hallucinations

- From the error map alone, it is not possible to determine whether any error structure is a hallucination



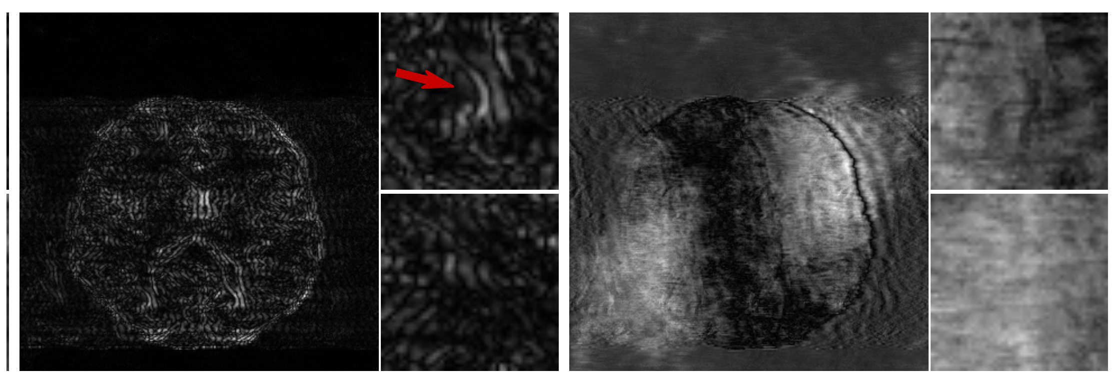
Analysis of hallucinations

True object | Reconstructed image | Error map



[0,0.6] | [0,0.6] | [0,0.15]

Null component error | **Measurement component error**



[0,0.12] | [0,0.07]

Hallucination maps

- *Hallucination maps* formally define hallucinations in linear inverse problems such as image reconstruction

- Hallucination maps are defined separately in the generalized measurement space and null space

Hallucination maps

- **Measurement space hallucination map** ($\hat{\mathbf{f}}_{meas}^{HM}$):

$$\hat{\mathbf{f}}_{meas}^{HM} \equiv \hat{\mathbf{f}}_{meas} - \hat{\mathbf{f}}_{tp}$$

— $\hat{\mathbf{f}}_{meas}$ is the generalized measurement component of the reconstructed estimate

- Measures discrepancies in $\hat{\mathbf{f}}_{meas}$ with respect to the stable estimate $\hat{\mathbf{f}}_{tp}$ in the generalized measurement space due to the imposed regularization

Hallucination maps

- **<u>Null space hallucination map</u>** ($\hat{\mathbf{f}}_{null}^{HM}$):

$$\hat{\mathbf{f}}_{null}^{HM} \equiv \mathbb{1}(\hat{\mathbf{f}}_{null}) \odot (\hat{\mathbf{f}}_{null} - \mathbf{f}_{null})$$

— $\hat{\mathbf{f}}_{null}$ is the generalized null component of the reconstructed estimate

— $\mathbb{1}: \mathbb{R}^N \to \mathbb{R}^N$ is a pixelwise indicator function such that for any $\vartheta \in \mathbb{R}^N$

$$[\mathbb{1}(\vartheta)]_n = \begin{cases} 1, & if\ [\vartheta]_n \neq 0 \\ 0, & if\ [\vartheta]_n = 0 \end{cases}$$

- Captures deviations due to *only the imposed prior.*

Hallucination maps

- Definitions maintain that $\hat{\mathbf{f}}_{tp}$ does not have any null space hallucinations since no prior employed

- Measurement space hallucinations can be mitigated with the measurement component fixed as $\hat{\mathbf{f}}_{tp}$:

$$\hat{\mathbf{f}}' \equiv \mathbf{H}_P^{+}\mathbf{g} + \mathcal{P}_{null}\hat{\mathbf{f}}$$

- **No such remedy for null space hallucinations!**

Specific hallucination maps (SHM)

- Task-specific information need to be identified from hallucination maps

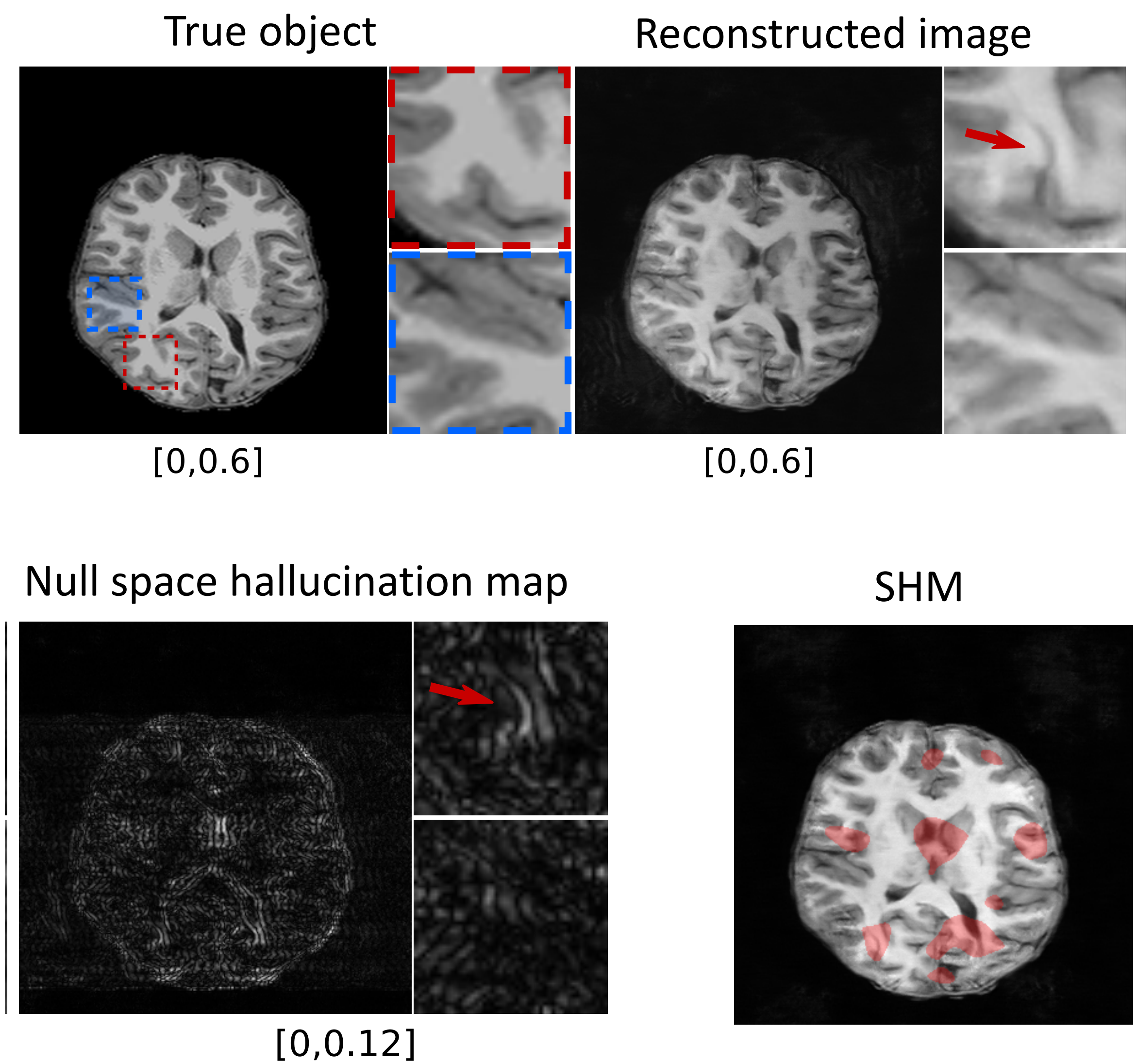
- Apply image processing transformation T *after* computing hallucination maps to localize only task-relevant features or textures:

$$\hat{\mathbf{f}}_{null}^{SHM} = T\hat{\mathbf{f}}_{null}^{HM}$$

- Design of transformation T is application-dependent

- The observer characteristics (human or computational procedure) will influence design of T

Specific hallucination maps (SHM)



Example: stylized single-coil 2D MRI

- Forward operation:

$$\mathbf{g} = S_{mask} \odot \mathrm{FT}(\mathbf{f})$$

  - FT: Fourier transform
  - S_{mask} : sampling mask
  - $[S_{mask}]_n = \begin{cases} 1, & \text{if } [\mathrm{FT}(\mathbf{f})]_{\mathrm{n}} \text{ sampled} \\ 0, & \text{if } [\mathrm{FT}(\mathbf{f})]_{\mathrm{n}} \text{ not sampled} \end{cases}$

import numpy.fft as fft

```
# Forward operator
def forward(f,mask):
    H_f = mask*fft.fftshift(fft.fft2(fft.ifftshift(f),norm='ortho'))
    return H_f
```

Example: stylized single-coil 2D MRI

- **Step 1**: Compute the truncated pseudoinverse solution $\hat{\mathbf{f}}_{tp} \equiv \mathbf{H}_t^+ \mathbf{g}$

```
# Pseudoinverse operator
def pinv(g):
    pinv_g = fft.fftshift(fft.ifft2(fft.ifftshift(g),norm='ortho'))
    return pinv_g
```

(No truncation required since singular values for Fourier Transform operator are either 0 or 1)

Homework

- Can you prove that the IFFT is the pseudoinverse operator for subsampled zero-filled k-space?

Hint : Represent applying mask for subsampling and zero-filling after FT as a matrix operation and use the definition of pseudoinverse

Example: stylized single-coil 2D MRI

- **Step 2**: Compute the generalized measurement component of $\hat{\mathbf{f}}$:

$$\hat{\mathbf{f}}_{meas} = \mathcal{P}_{meas}\hat{\mathbf{f}} = \mathbf{H}_t^{+}\mathbf{H}\hat{\mathbf{f}}$$

```
# Function for computing measurement and null components of an object
def f_meas_null(f,mask):
    H_f = forward(f,mask)
    f_meas = pinv(H_f)
    f_null = f - f_meas
    return f_meas, f_null
```

Example: stylized single-coil 2D MRI

- **Step 3:** Compute the measurement space hallucination map:

$$\hat{\mathbf{f}}_{meas}^{HM} \equiv \hat{\mathbf{f}}_{meas} - \hat{\mathbf{f}}_{tp}$$

```python
# Function for computing the measurement space hallucination map
def meas_hm(recon,pinv_g,mask):
    recon_meas,_ = f_meas_null(recon,mask)
    h_map = recon_meas - pinv_g
    return h_map
```

Example: stylized single-coil 2D MRI

- **Step 4**: Compute the generalized null components of $\mathbf{f}$ and $\hat{\mathbf{f}}$:

$$\mathbf{f}_{null} = \mathcal{P}_{null}\mathbf{f} = [\mathbf{I}_N - \mathbf{H}_t^{+}\mathbf{H}]\mathbf{f}$$
$$\hat{\mathbf{f}}_{null} = \mathcal{P}_{null}\hat{\mathbf{f}} = [\mathbf{I}_N - \mathbf{H}_t^{+}\mathbf{H}]\hat{\mathbf{f}}$$

```
# Function for computing measurement and null components of an object
def f_meas_null(f,mask):
    H_f = forward(f,mask)
    f_meas = pinv(H_f)
    f_null = f - f_meas
    return f_meas, f_null
```

Example: stylized single-coil 2D MRI

- **Step 5**: Compute the null space hallucination map $\hat{\mathbf{f}}_{null}^{HM} \equiv \mathbb{1}(\hat{\mathbf{f}}_{null}) \odot (\hat{\mathbf{f}}_{null} - \mathbf{f}_{null})$

```
# Function for computing the null space hallucination map
def null_hm(recon,gt,mask):
    _, recon_null = f_meas_null(recon,mask)
    _, gt_null = f_meas_null(gt,mask)
    h_map = recon_null - gt_null
    h_map[recon_null==0]=0
    return h_map
```

Example: stylized single-coil 2D MRI

- **Step 6**: Compute specific hallucination map based on the task concerned:

$$\hat{\mathbf{f}}_{null}^{SHM} = T\hat{\mathbf{f}}_{null}^{HM}$$

Numerical studies [Bhadra *et al.* (2019)]

- Stylized single-coil 2D MRI with Cartesian sampling mask and undersampling factor of 3

- **Noise model:** Independent and identically distributed (i.i.d) complex Gaussian noise[1]

- Random uniform phase noise was also added[2]

$$\widetilde{\mathbf{H}} = \mathbf{H}\exp(j\delta)$$

1. Aja-Fernandez and Tristan-Vega (2013). *A review on statistical noise models for magnetic resonance imaging*
2. Xiaoyu *et al.* (2017). *Compressed sensing MRI with phase noise disturbance based on adaptive tight frame and total variation*

Numerical studies [Bhadra *et al.* (2019)]

- **Reconstruction methods:**
  - **Non-data-driven:**
    - PLS-TV
    - DIP-TV [Ulyanov et al. (2018), Liu et al. (2019)]
  - **Data-driven:** U-Net based reconstruction [Jin et al. (2017)]
- **PLS-TV:**

— Regularization with a sparsity-promoting penalty :

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{f}\|_{TV}$$

— $\|\mathbf{f}\|_{TV}$: total variation semi-norm

— λ : controls trade-off between data fidelity and regularization

Numerical studies [Bhadra *et al.* (2019)]

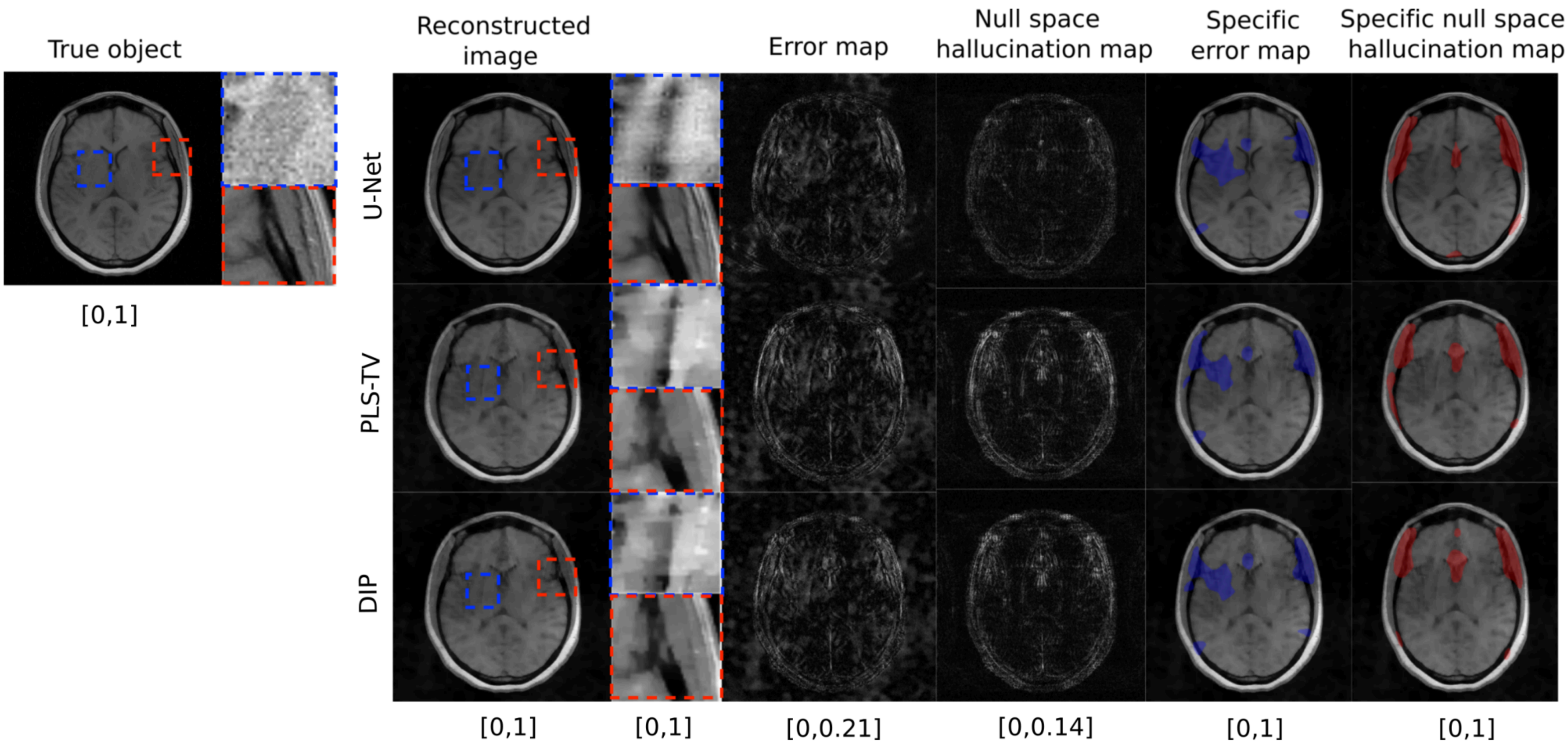
- **Datasets:** Trained U-Net with 3,000 ground truth 2D axial adult brain MRI images and simulated k-space data from NYU fastMRI dataset

- Reconstructed images at test time obtained from unseen in-distribution (IND) and simulated out-of-distribution (OOD) measurements

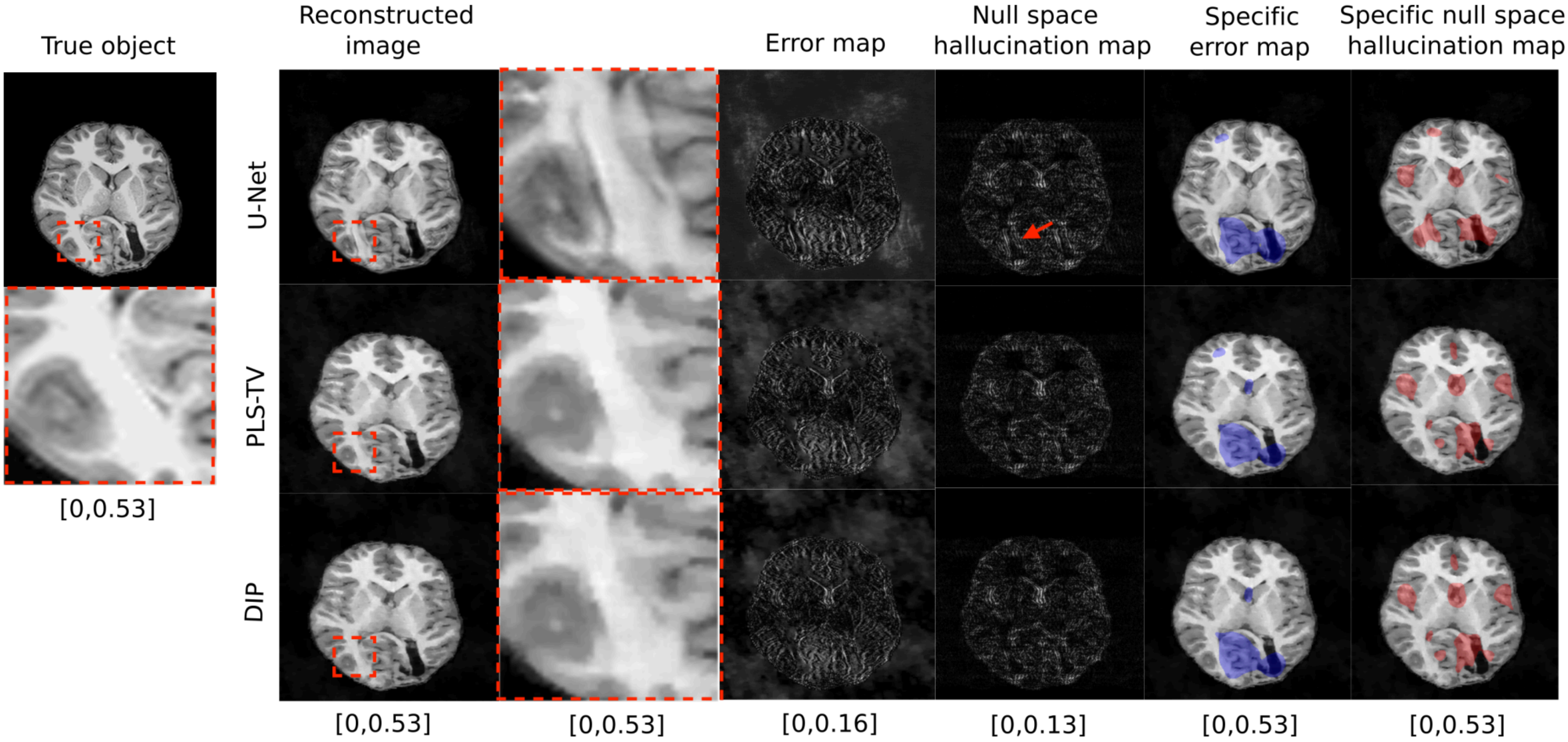
- OOD ground truth images obtained from a publicly available pediatric epilepsy resection dataset

Maallo *et al.* *"Effects of unilateral cortical resection of the visual cortex on bilateral human white matter"* (NeuroImage 2020)

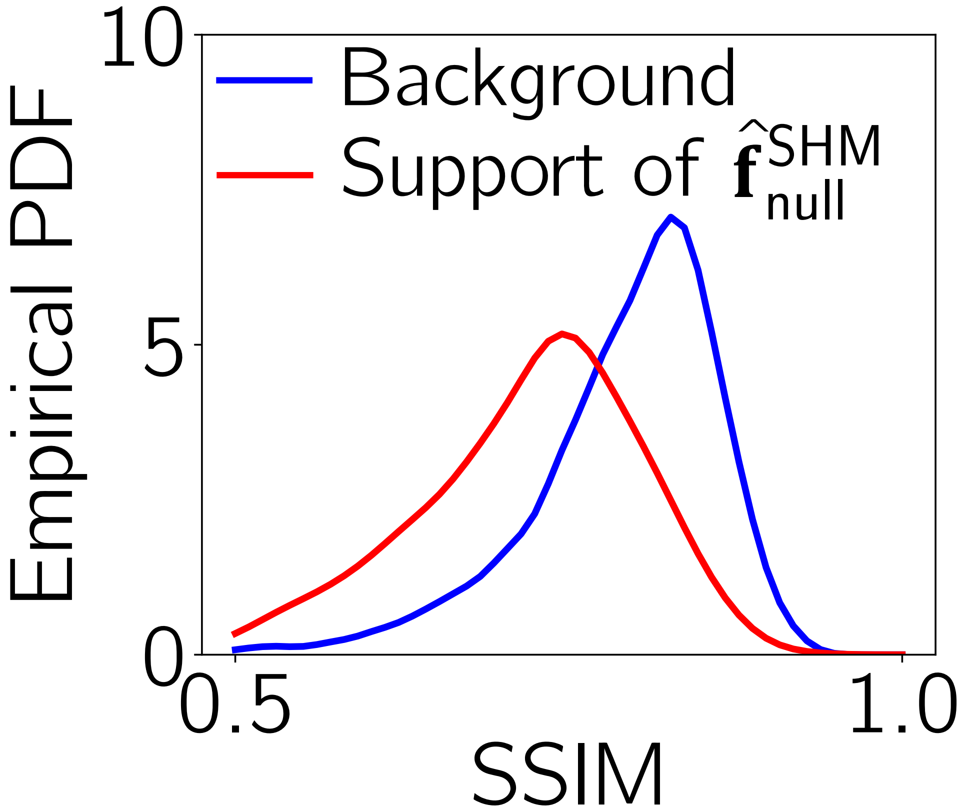
Results from IND test data



Results from OOD test data

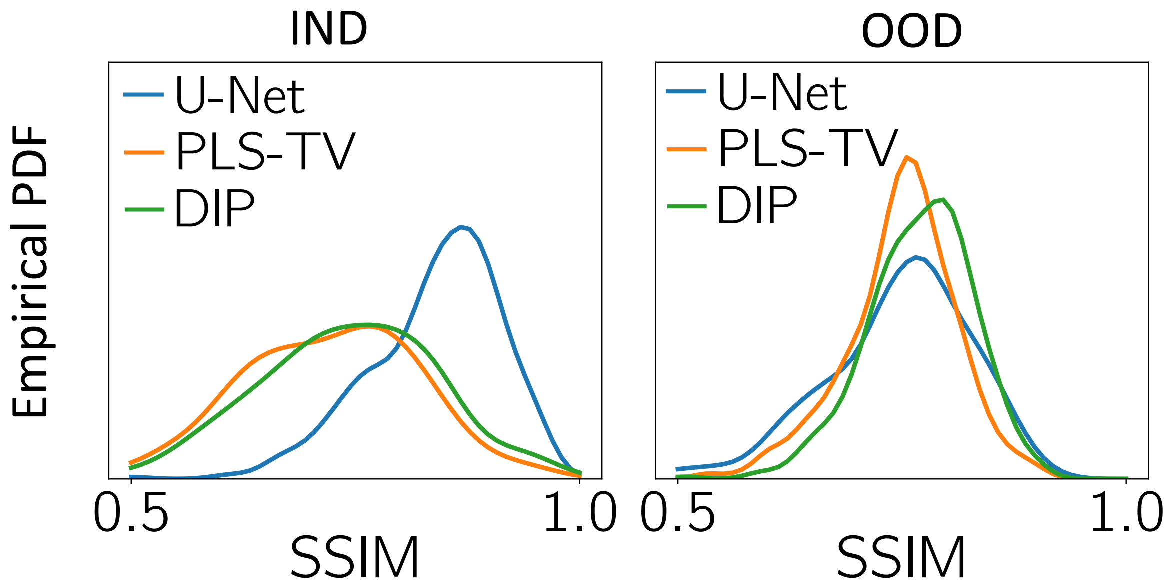


Impact of null space hallucinations



(for U-Net reconstruction with OOD data)

Impact of null space hallucinations



Median of SSIM in support region of $\hat{\mathbf{f}}_{null}^{SHM}$

Data distribution	U-Net	PLS-TV	DIP
IND	**0.84**	0.71	0.73
OOD	0.75	0.73	**0.76**

Generality of hallucination maps

- The proposed method is general – not specific to any imaging system or reconstruction method
- Different data acquisition strategies (changing sampling pattern etc.) will lead to different null space characteristics and corresponding hallucination maps

Extensions

- A figure-of-merit (FOM) can summarize the task performance and help to optimize imaging systems
- Can we devise meaningful FOMs for realistic applications, e.g. detection task?
- How frequently can we expect to see impactful hallucinations for a given type of object, imaging system and reconstruction method?

Caveat

- Computation of projection operators via SVD may be infeasible for large scale problems

- Ongoing research by Joseph and Albert in our lab

Summary

- We formally defined hallucinations in tomographic imaging from the perspective of linear operator theory

- Hallucination maps can be used to identify false structures in reconstructed images due to inaccuracies in the imposed prior

- Certain data-driven methods may result in structured hallucinations in the reconstructed image when there is a shift in the data distribution