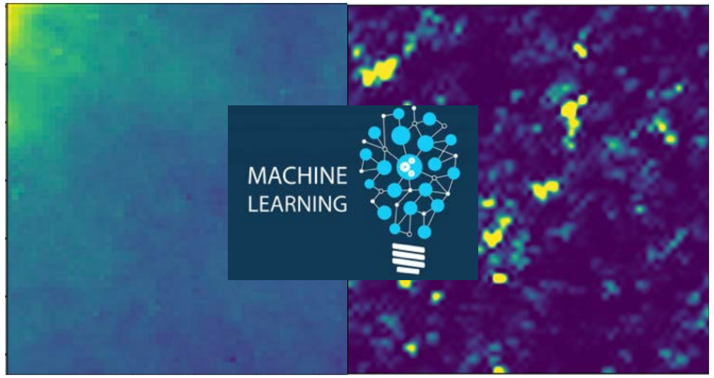
An AI method for 21-cm foreground removal based on frequency-band difference



Feng Shi （史峰）

with Haoxiang Chang, Le Zhang, Huanyuan Shan, Jiajun Zhang

21cm workshop & Tianlai meeting, Shenyang

2023/07/19

西安电子科技大学
XIDIAN UNIVERSITY
空间科学与技术学院
School of Aerospace Science And Technology

Motivation

- Biggest challenge in 21cm intensity mapping: foreground!
- Traditional blind method, such as PCA,
    - Assume a strong frequency correlation
    - Reduce statistically foregrounds into a few main components
    - Advantage: little prior knowledge of the foregrounds
    - Disadvantage: foreground residuals and signal loss

Motivation

- Biggest challenge in 21cm intensity mapping: foreground!

- Traditional blind method, such as PCA,
  - Assume a strong frequency correlation
  - Reduce statistically foregrounds into a few main components
  - Advantage: little prior knowledge of the foregrounds
  - Disadvantage: foreground residuals and signal loss

- New method based on UNet (Deep21-Makinen et al 2021, Ni et al 2022)
  - Perform effectively with a necessary PCA preprocessing
  - PCA for removing the smooth foregrounds; UNet for the non-smooth contaminations
  - However, PCA has HI signal loss, which is not expected to be recovered by UNet

Motivation

- Biggest challenge in 21cm intensity mapping: foreground!
- Traditional blind method, such as PCA,
  - Assume a strong frequency correlation
  - Reduce statistically foregrounds into a few main components
  - Advantage: little prior knowledge of the foregrounds
  - Disadvantage: foreground residuals and signal loss
- New method based on UNet (Deep21-Makinen et al 2021, Ni et al 2022)
  - Perform effectively with a necessary PCA preprocessing
  - PCA for removing the smooth foregrounds; UNet for the non-smooth contaminations
  - However, PCA has HI signal loss, which is not expected to be recovered by UNet
- Our solution is still based on the UNet, but expect to avoid the HI signal loss

21cm simulations

➢ CRIME:

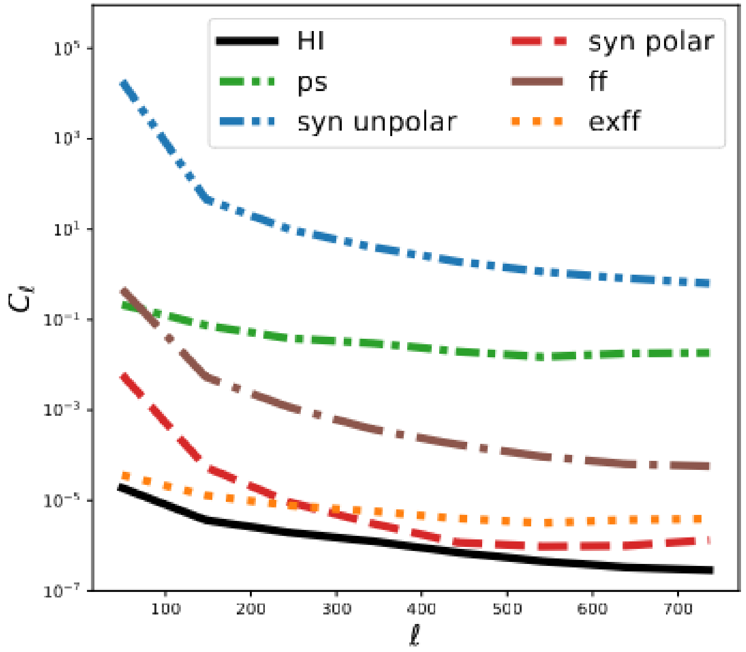
- HI cosmological signal: generated from linear density and velocity field

$$T_{\mathrm{HI}}(\hat{\boldsymbol{n}}, z) = \bar{T}_{\mathrm{HI}}(z)(1 + \delta_{\mathrm{HI}}(\hat{\boldsymbol{n}}, z))$$

$$\bar{T}_{\mathrm{HI}}(z) = (190.55\mathrm{mK})\frac{\Omega_{\mathrm{b}} h(1+z)^2 x_{\mathrm{HI}}(z)}{\sqrt{\Omega_{\mathrm{m}}(1+z)^3 + \Omega_\Lambda}}$$

$$x_{\mathrm{HI}}(z) = 0.008(1+z)$$

$$\Delta z_{\mathrm{RSD}} = (1+z)v_{\mathrm{r}}/c$$



- Foregrounds:

Galactic synchrotron anisotropic structure:

$$T_{\mathrm{syn},0}(\nu, \hat{\boldsymbol{n}}) = T_{\mathrm{Haslam}}(\hat{\boldsymbol{n}})\left(\frac{408\mathrm{MHz}}{\nu}\right)^{\beta(\hat{\boldsymbol{n}})}$$

Isotropic structure by Gaussian realization:

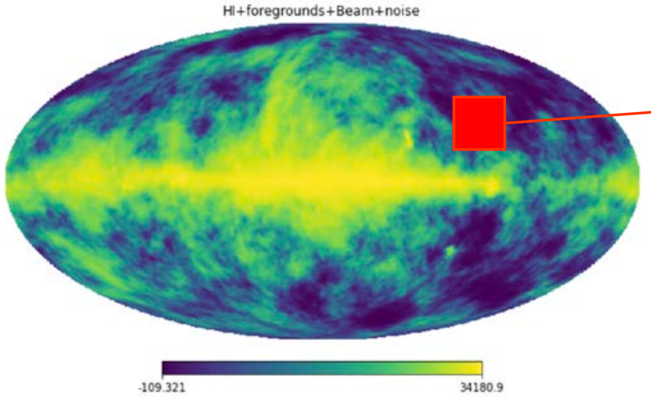
$$C_\ell(v_1, v_2) = A\left(\frac{\ell_{\mathrm{ref}}}{\ell}\right)^{\beta}\left(\frac{v_{\mathrm{ref}}^2}{v_1 v_2}\right)^{\alpha}\exp\left(-\frac{\log^2(v_1/v_2)}{2\xi^2}\right)$$

Foreground	$A(mK^2)$	β	α	ξ
Galactic synchrotron	700	2.4	2.80	4.0
Point sources	57	1.1	2.07	1.0
Galactic free-free	0.088	3.0	2.15	35
Extragalactic free-free	0.014	1.0	2.10	35

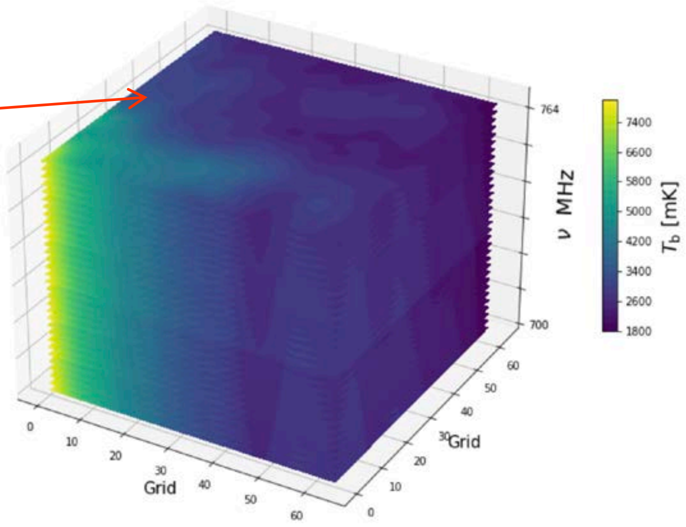
Santos et al (2005)

21cm simulations

CRIME output



Samples used in our analysis

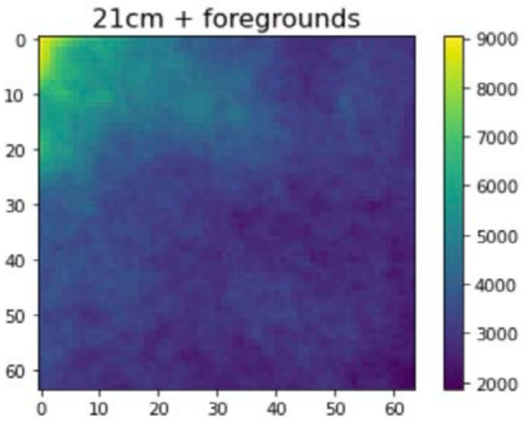


- 700~764 MHz
- 0.859<z<1.029

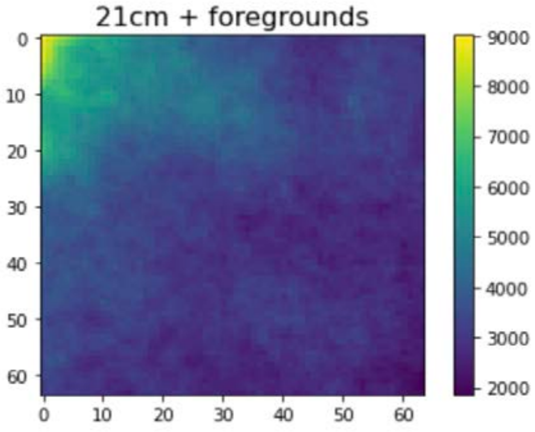
- 214.85 deg^2
- volume ~460 h^{-1}Mpc
- Grid: 64 × 64 × 64
- 192 sky patches

Temperature difference between neighboring frequency bands

Frequency ν_0

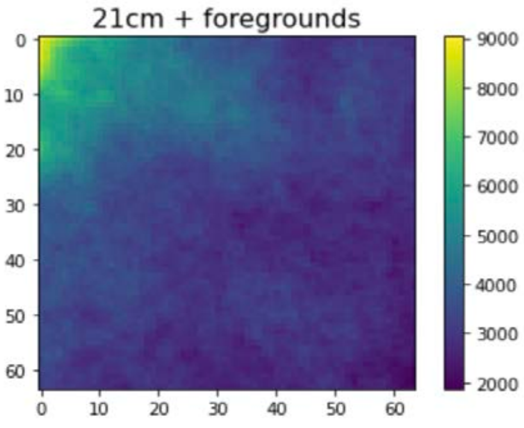


Frequency ν_1



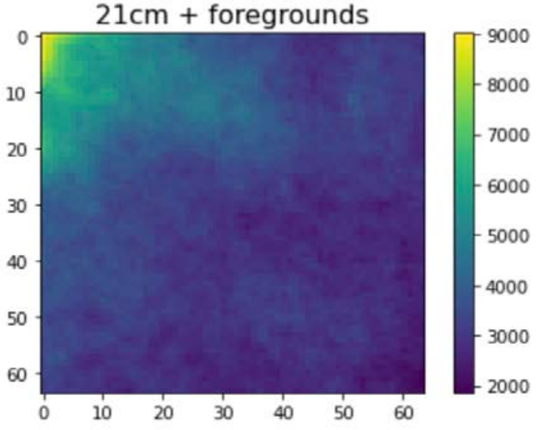
Temperature difference between neighboring frequency bands

Frequency ν_0

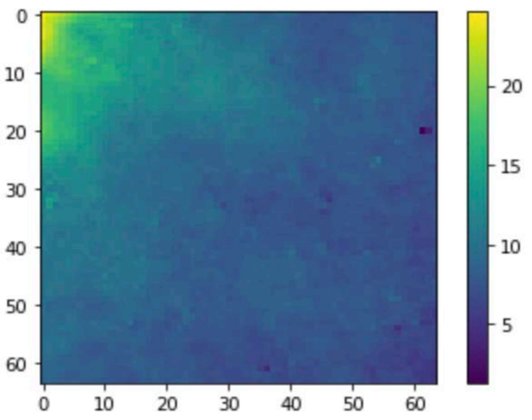


−

Frequency ν_1



=

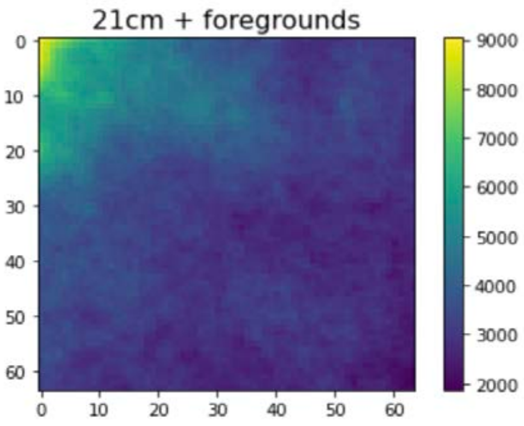


HI : $\nu_0 + \nu_1$

foregrounds: $\nu_0 - \nu_1$

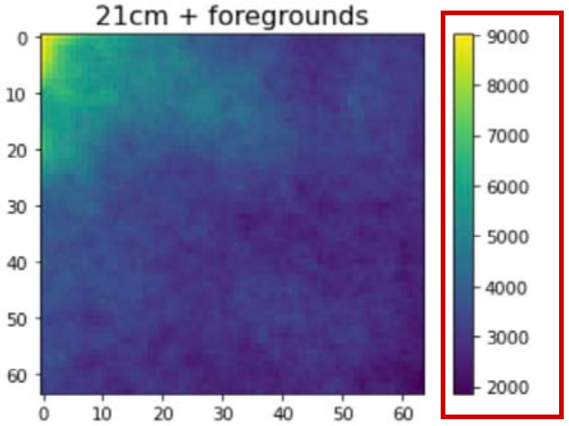
Temperature difference between neighboring frequency bands

Frequency ν_0

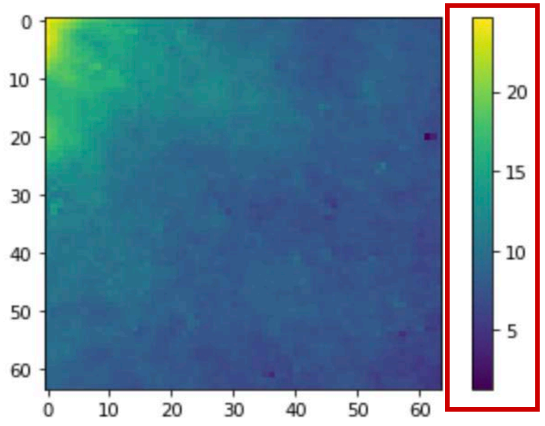


$-$

Frequency ν_1



$=$

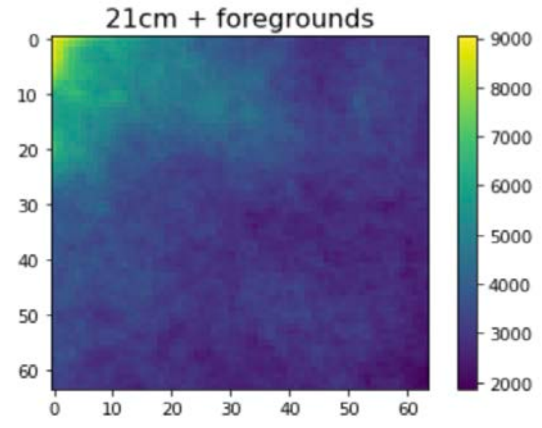


HI : $\nu_0 + \nu_1$

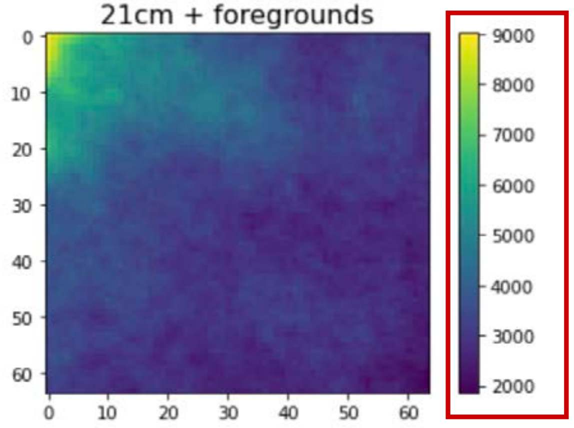
foregrounds: $\nu_0 - \nu_1$

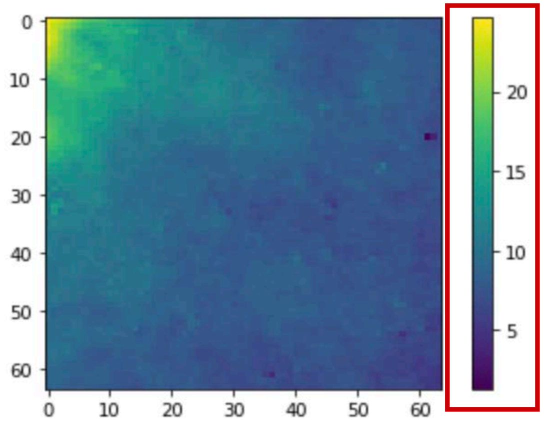
Temperature difference between neighboring frequency bands

Frequency ν_0



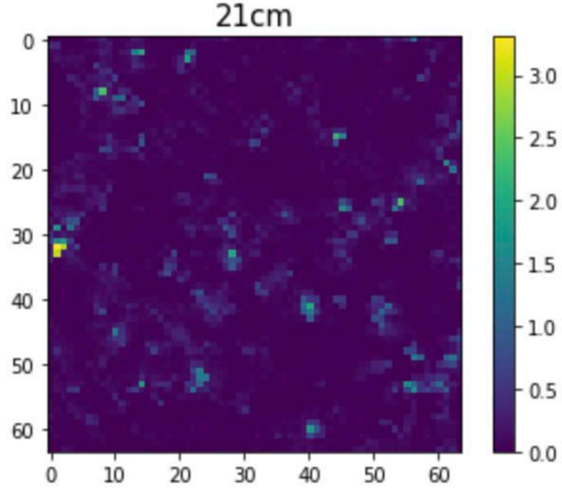
Frequency ν_1





HI : $\nu_0 + \nu_1$

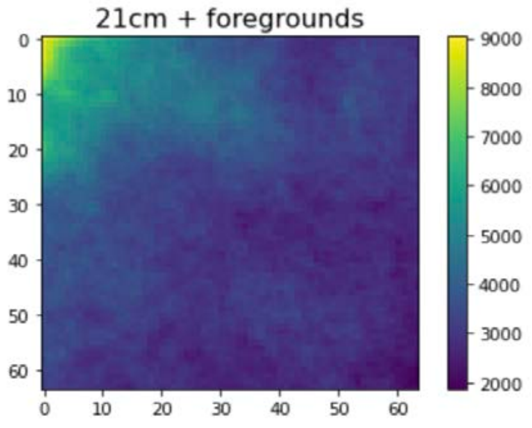
foregrounds: $\nu_0 - \nu_1$



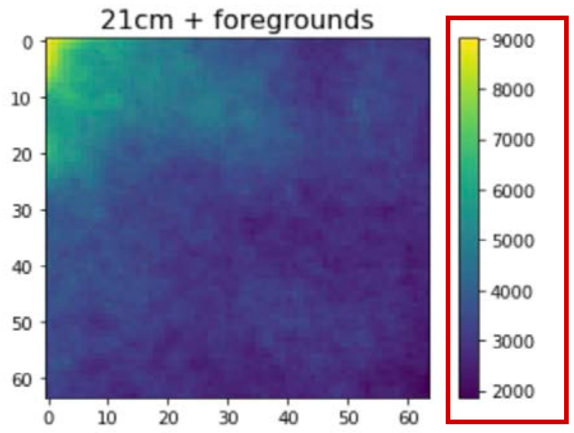
HI: ν_0

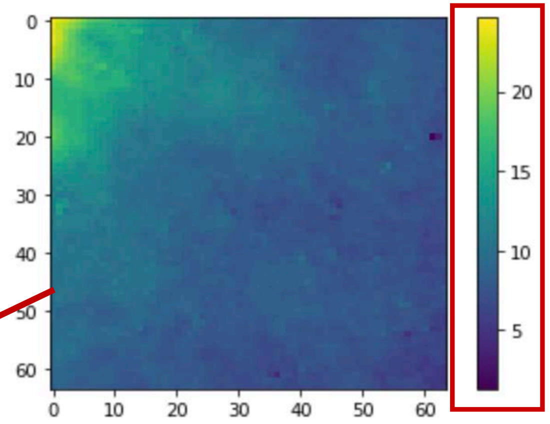
Temperature difference between neighboring frequency bands

Frequency ν_0



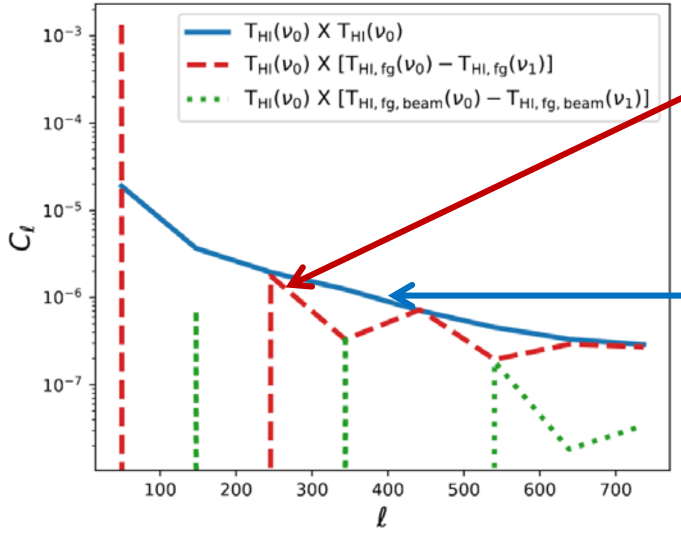
Frequency ν_1

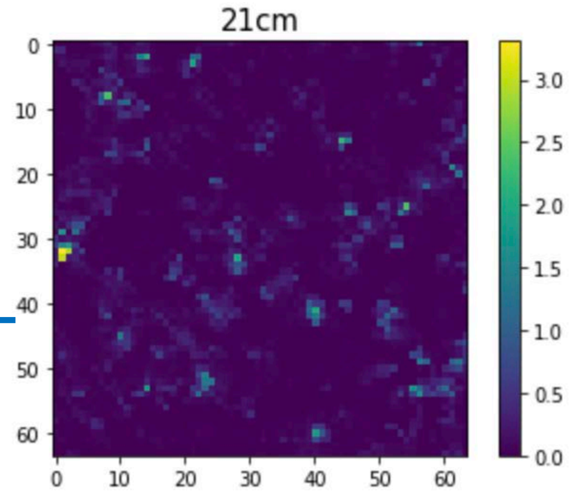




HI : $\nu_0 + \nu_1$

foregrounds: ν_0 - ν_1

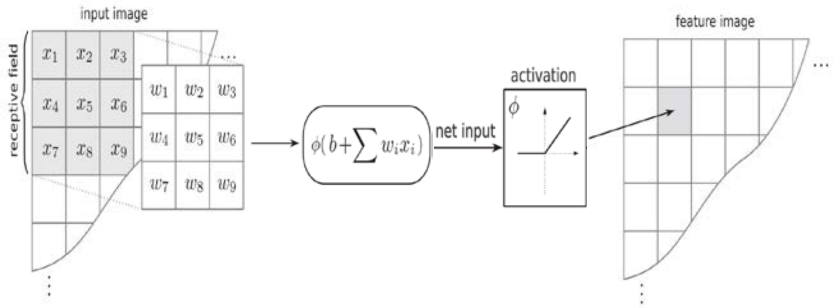


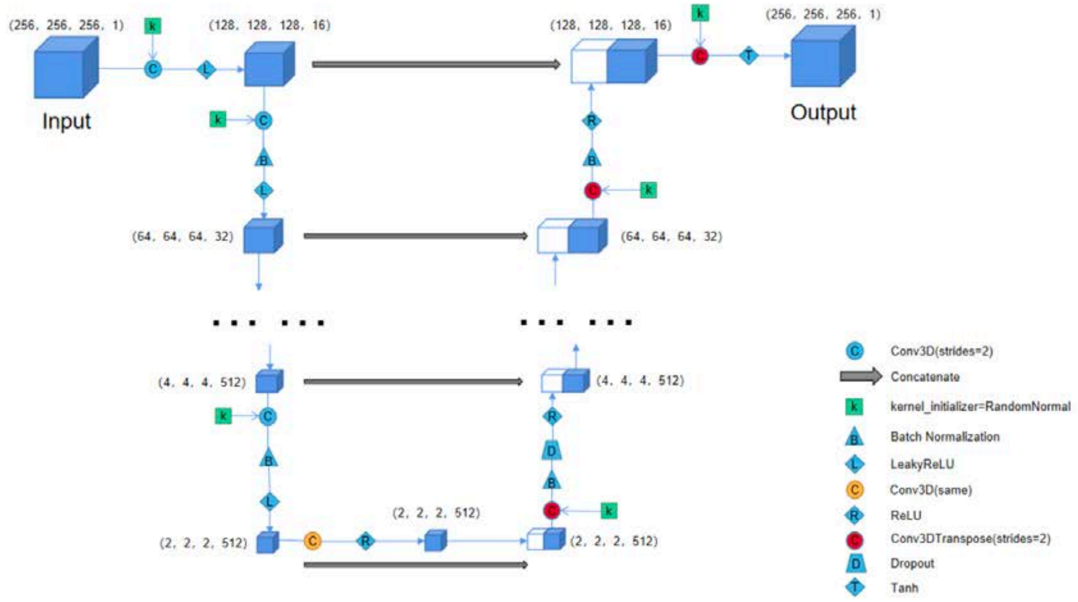


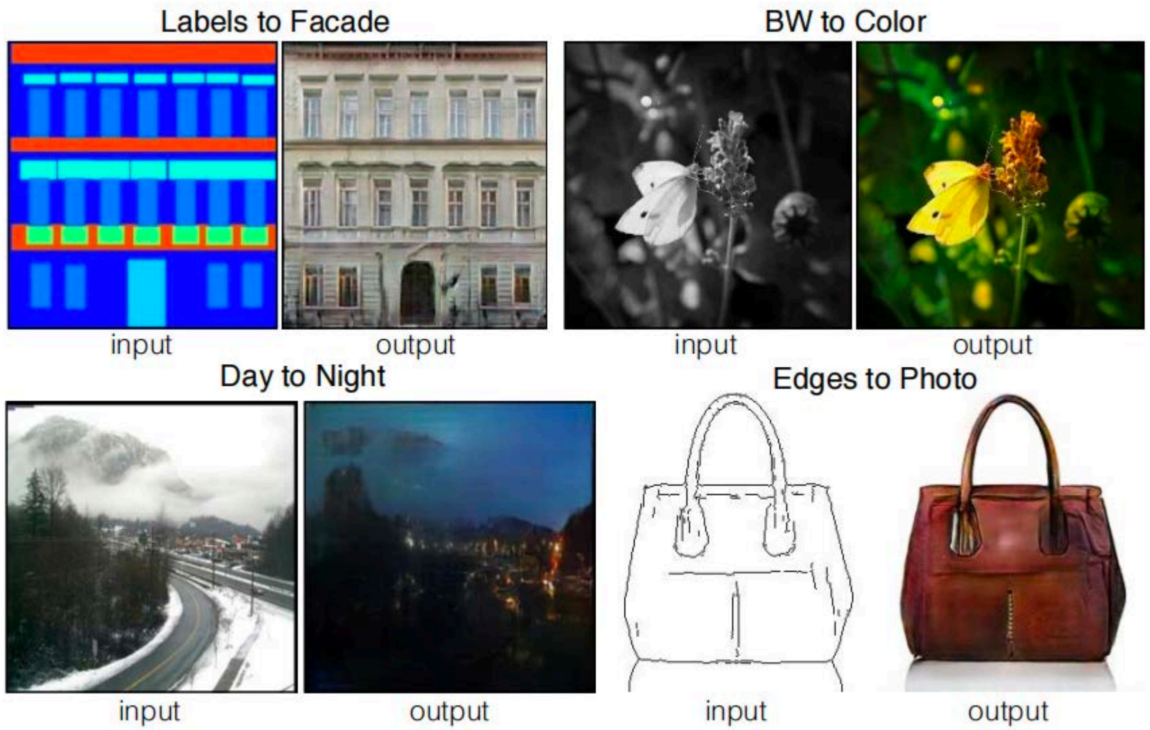
HI: ν_0

Deeping learning method: UNet model

- Convolutional neural network
- Encoder-decoder with skip connections
- General model for image-to-image translation

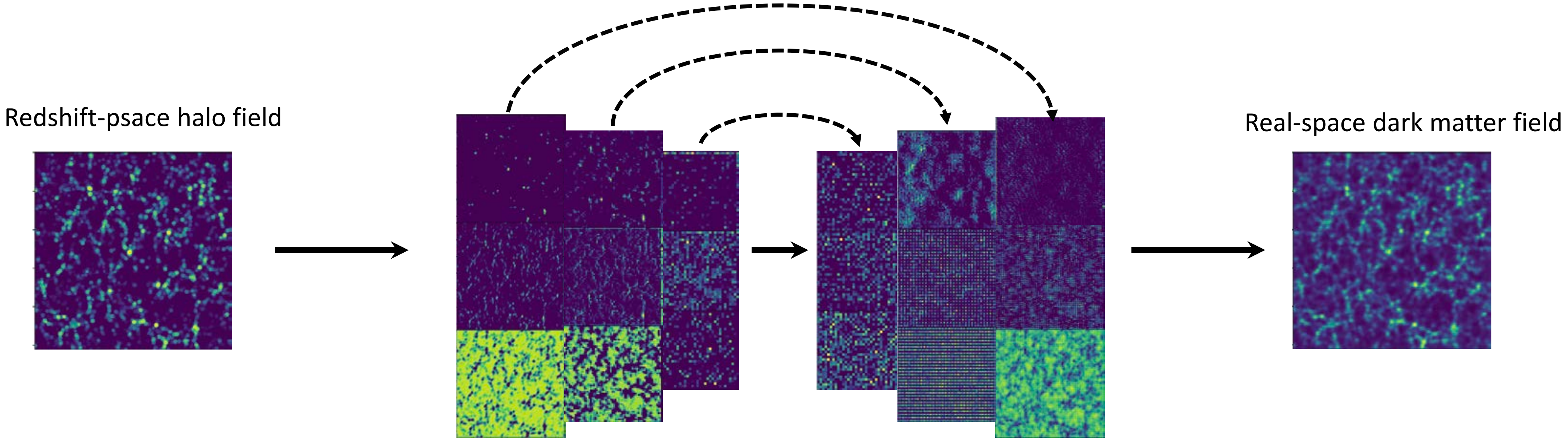






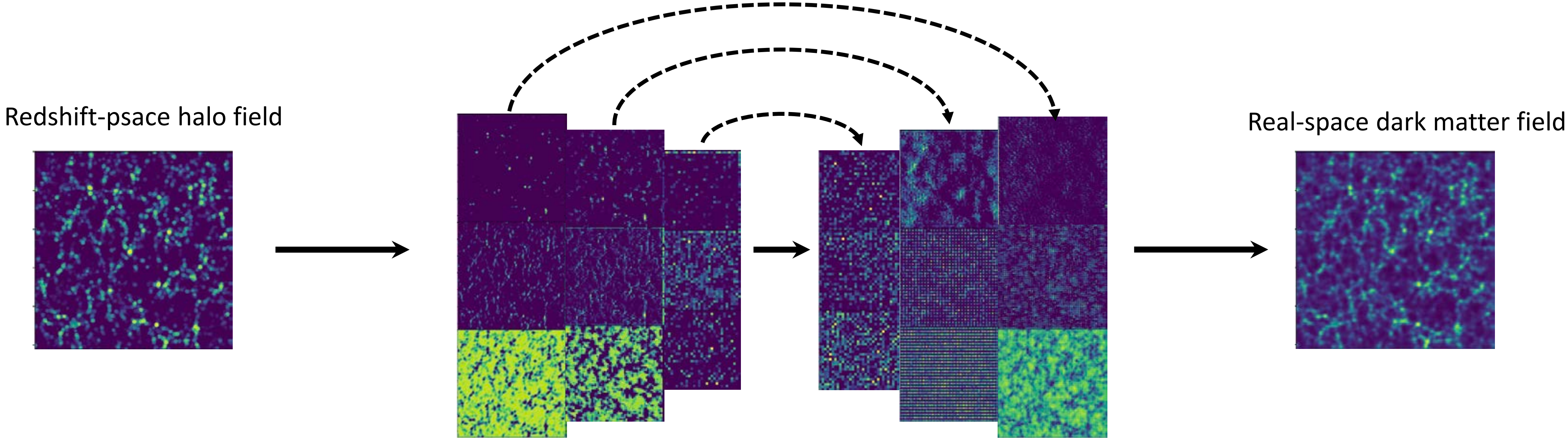
Deeping learning method: UNet model

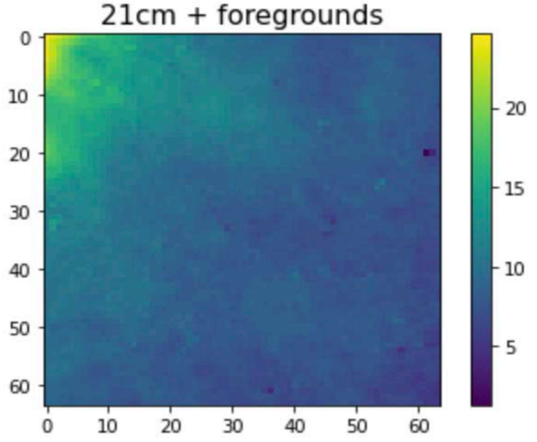
DarkAI: Reconstructing the large-scale density field of dark matter using AI
(Wang, Shi, et al 2023, arXiv:2305.11431)



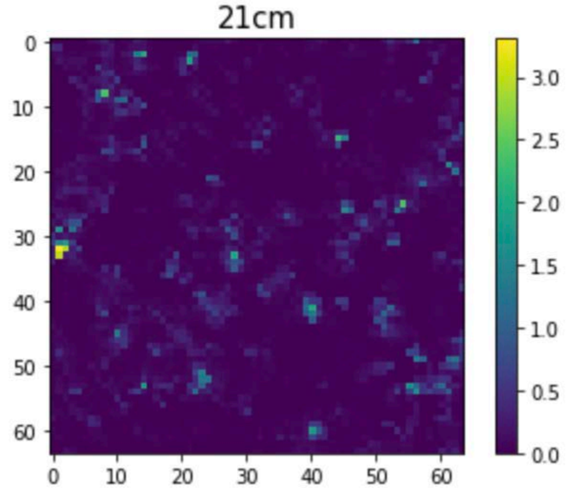
Deeping learning method: UNet model

DarkAI: Reconstructing the large-scale density field of dark matter using AI
(Wang, Shi, et al 2023, arXiv:2305.11431)





UNet removing noise?



Toy model : testing UNet ability for cleaning foreground

Toy model : testing UNet ability for cleaning foreground

- We run the UNet by inputting $T_{HI} + T_{fg}$, where $T_{fg}=\alpha T_{fg}^{CRIME}$

$$\beta_{\mathrm{fg}} = \sqrt{\frac{\langle (T_{\mathrm{fg}} - \overline{T}_{\mathrm{fg}})^2 \rangle}{\langle (T_{\mathrm{HI}} - \overline{T}_{\mathrm{HI}})^2 \rangle}}$$

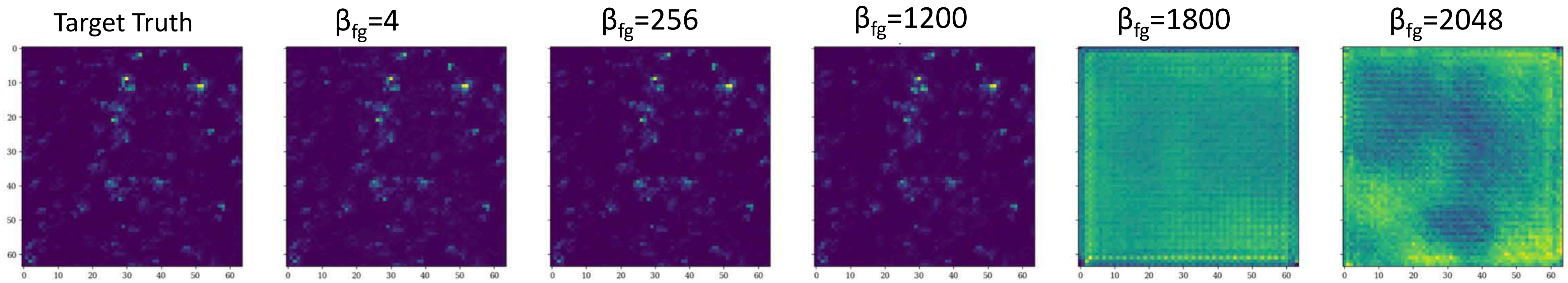
UNet Output：

Toy model : testing UNet ability for cleaning foreground

- We run the UNet by inputting $T_{HI} + T_{fg}$, where $T_{fg}=\alpha T_{fg}^{CRIME}$

$$\beta_{\mathrm{fg}} = \sqrt{\frac{\langle (T_{\mathrm{fg}} - \overline{T}_{\mathrm{fg}})^2 \rangle}{\langle (T_{\mathrm{HI}} - \overline{T}_{\mathrm{HI}})^2 \rangle}}$$

UNet Output：



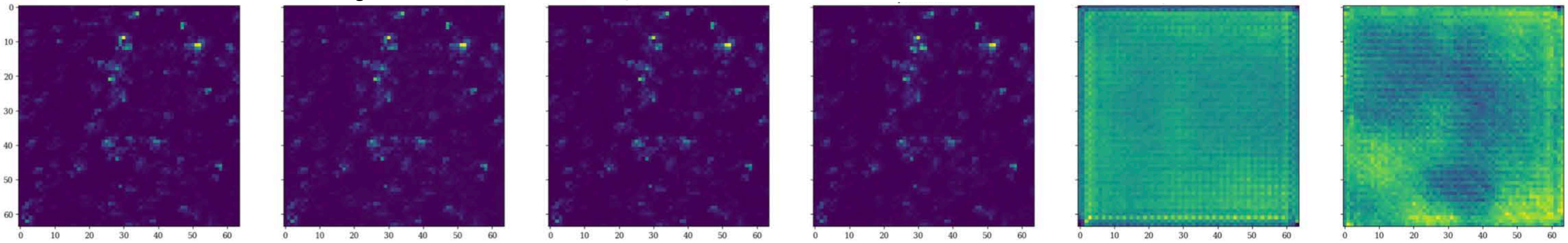
Toy model : testing UNet ability for cleaning foreground

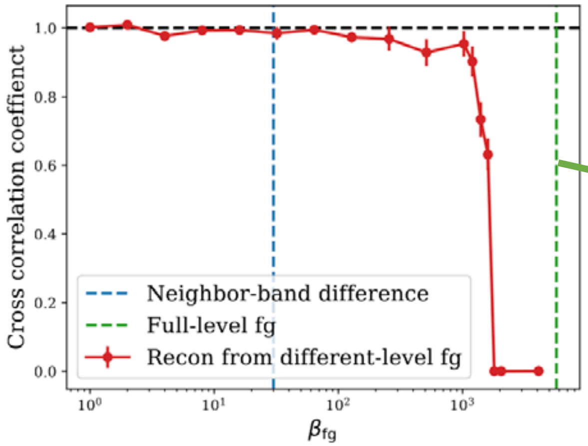
- We run the UNet by inputting $T_{HI} + T_{fg}$, where $T_{fg}=\alpha T_{fg}^{CRIME}$

$$\beta_{\rm fg} = \sqrt{\frac{\langle (T_{\rm fg} - \overline{T}_{\rm fg})^2 \rangle}{\langle (T_{\rm HI} - \overline{T}_{\rm HI})^2 \rangle}}$$

UNet Output：

Target Truth | β_{fg}=4 | β_{fg}=256 | β_{fg}=1200 | β_{fg}=1800 | β_{fg}=2048





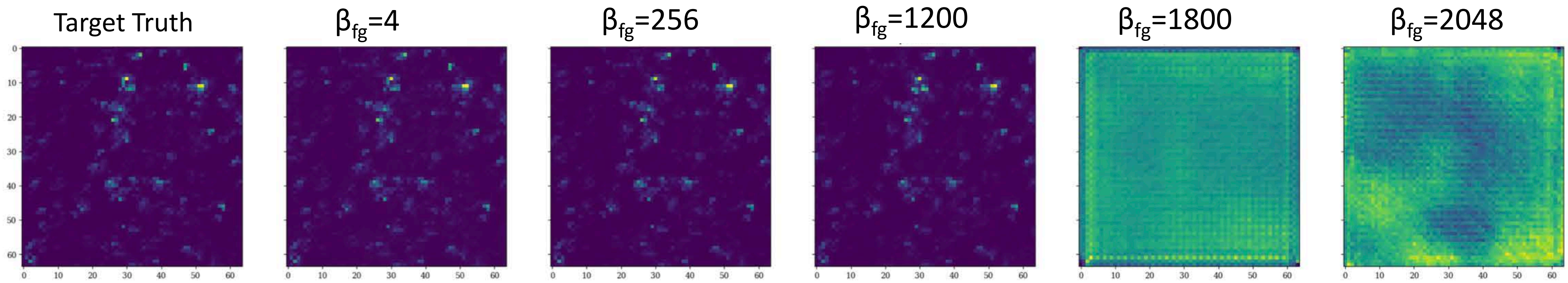
CRIME simulated foreground

Toy model : testing UNet ability for cleaning foreground

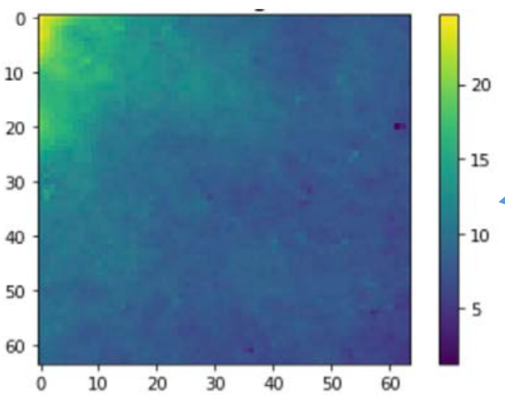
- We run the UNet by inputting $T_{HI} + T_{fg}$, where $T_{fg}=\alpha T_{fg}^{CRIME}$

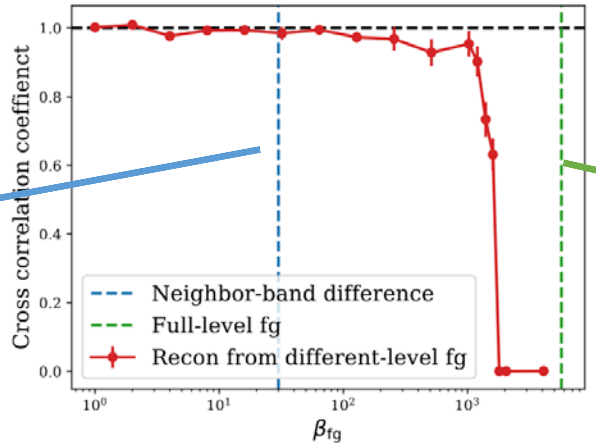
$$\beta_{\mathrm{fg}} = \sqrt{\frac{\langle (T_{\mathrm{fg}} - \overline{T}_{\mathrm{fg}})^2 \rangle}{\langle (T_{\mathrm{HI}} - \overline{T}_{\mathrm{HI}})^2 \rangle}}$$

UNet Output：



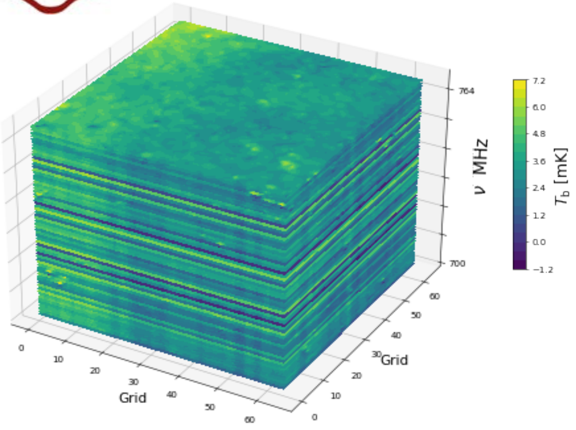
ν_0-ν_1

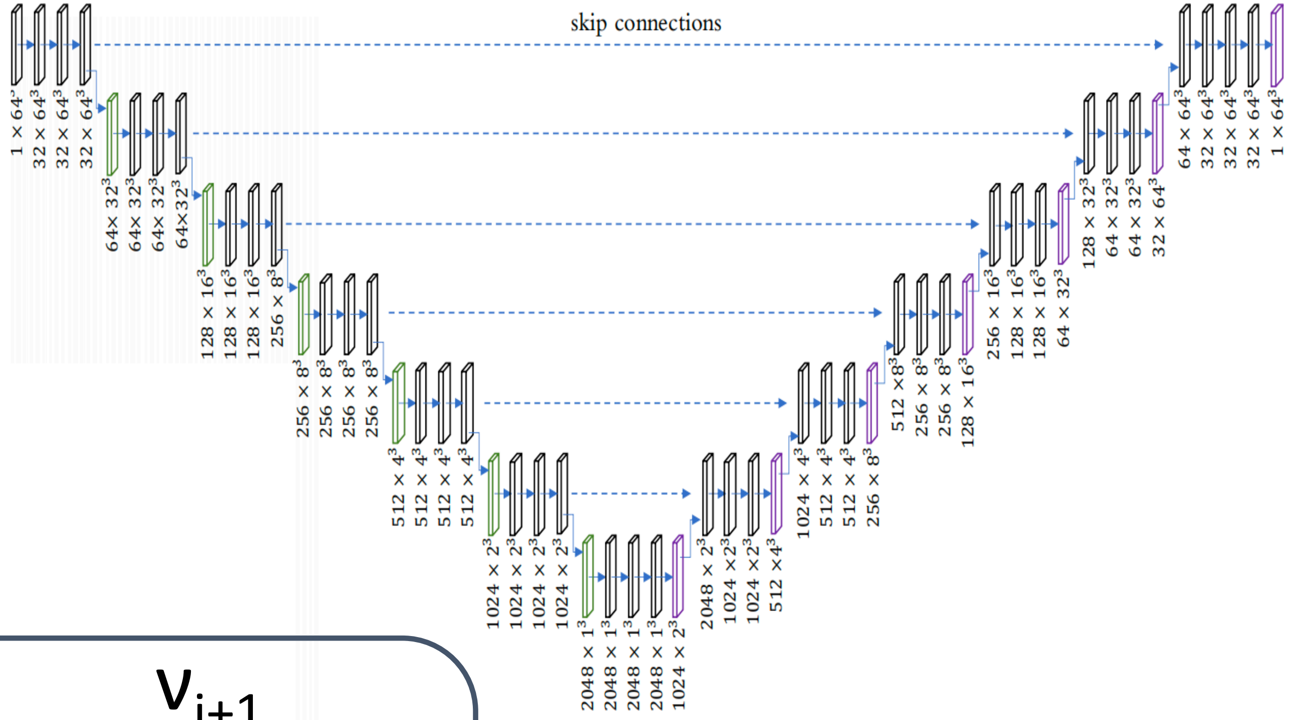


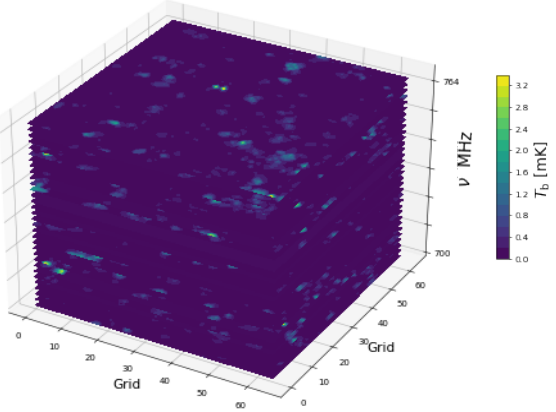


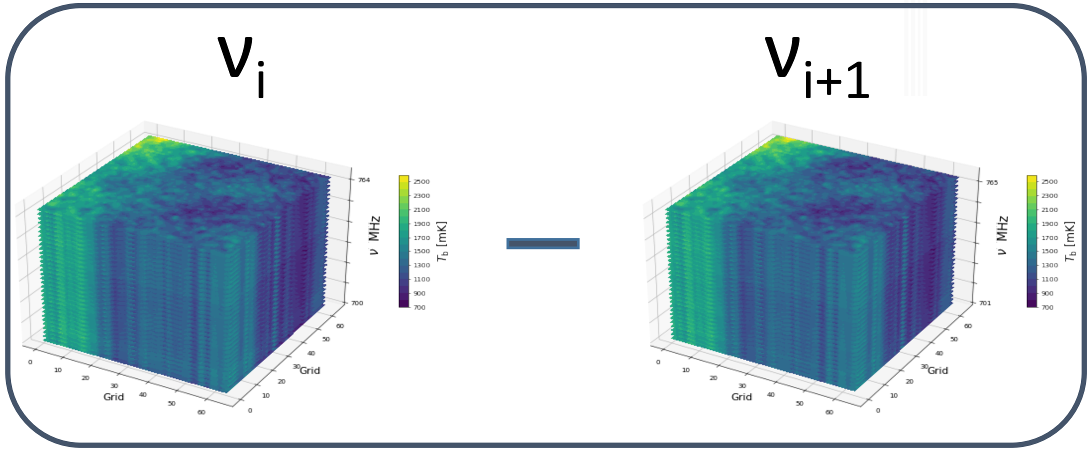
CRIME simulated foreground

Cleaning model: Neighbor-band difference + UNet









UNet-fd

Training samples: 140

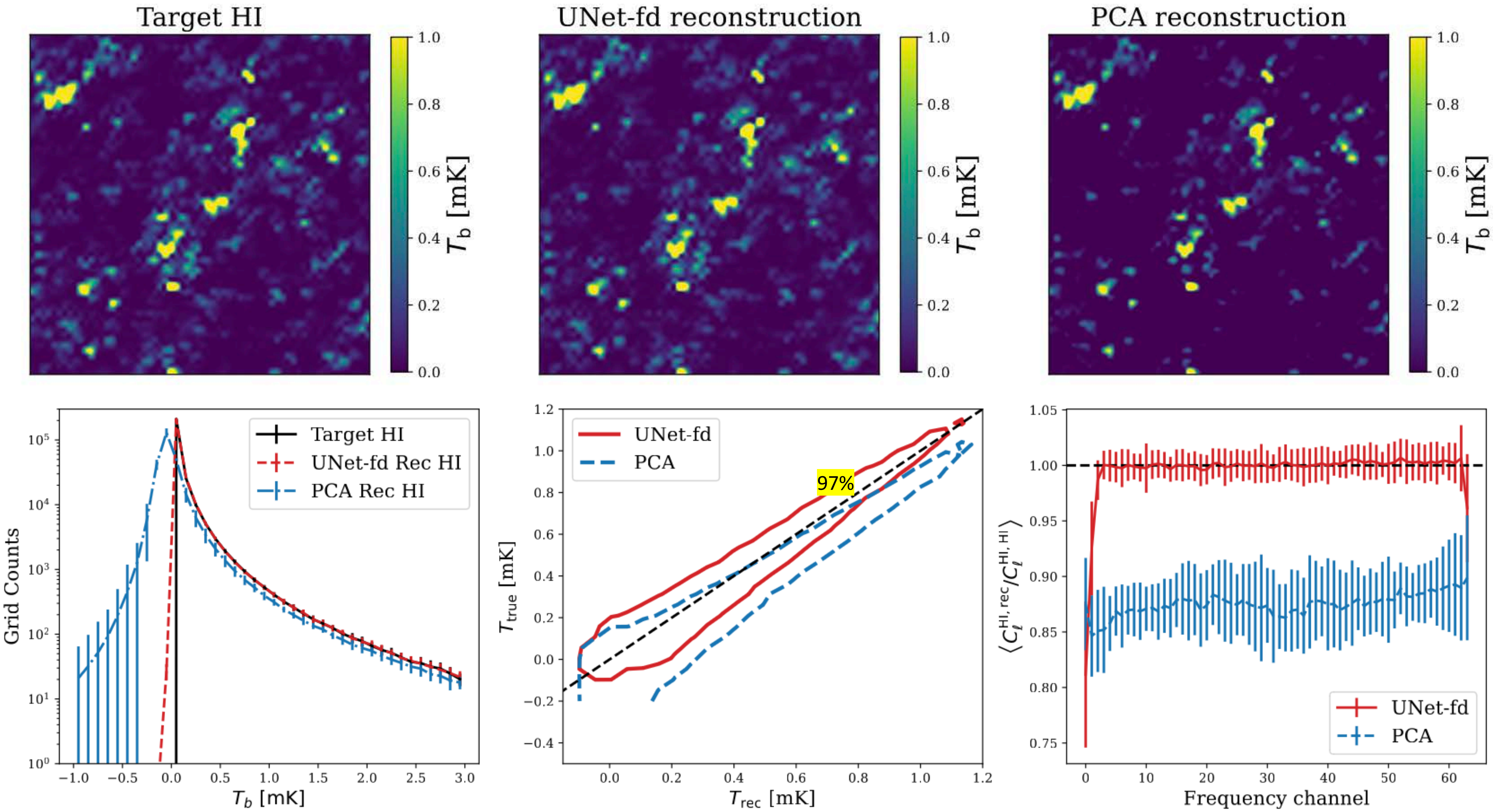
Validation samples: 12

Test samples: 40

Test I:

- 21cm + foreground
- No beam
- No thermal noise

Test I : 21cm+foreground



Test I : 21cm+foreground

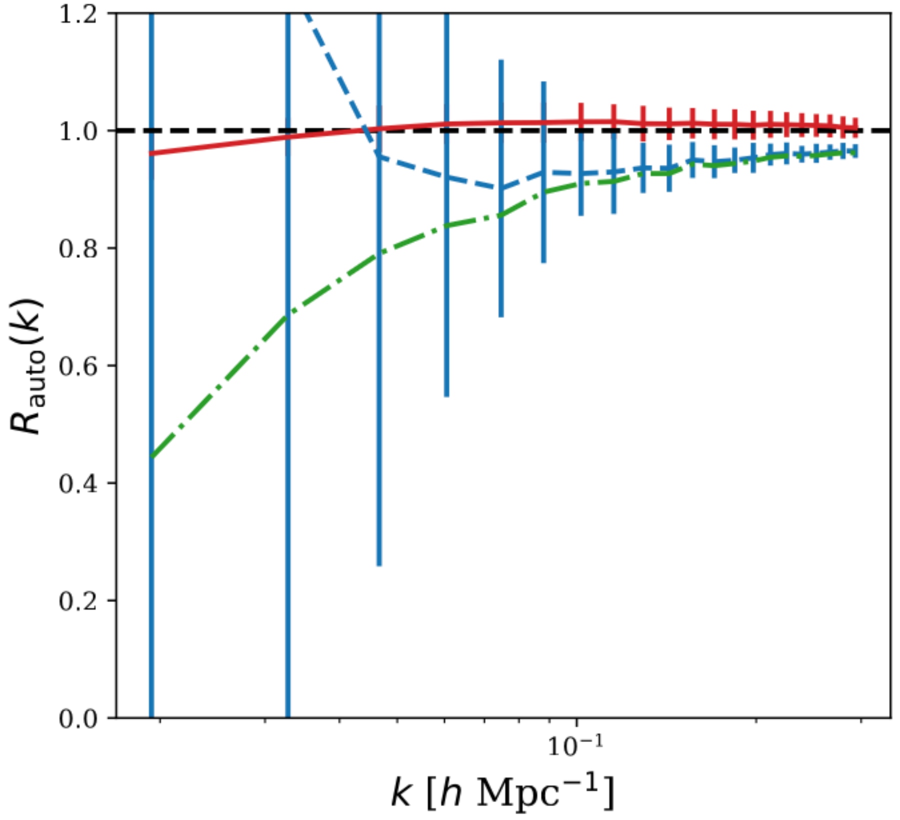
$$R_{\text{auto}}(k) = \frac{P_{\text{rec,rec}}(k)}{P_{\text{HI,HI}}(k)}$$

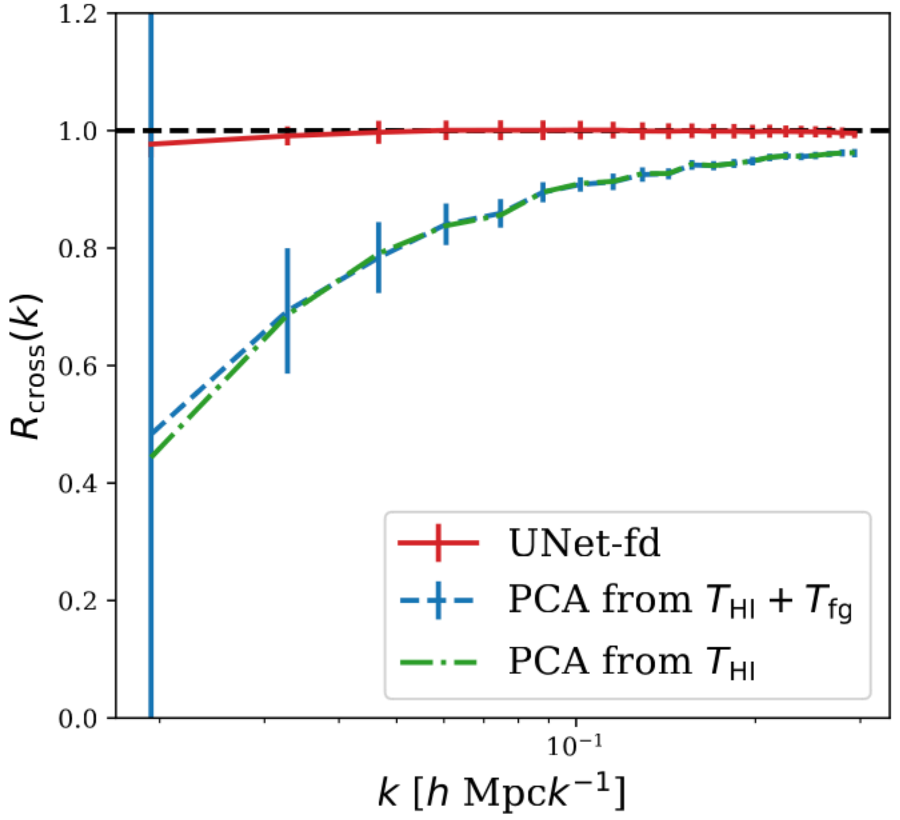
$$R_{\text{cross}}(k) = \frac{P_{\text{rec,HI}}(k)}{P_{\text{HI,HI}}(k)}$$

Test I : 21cm+foreground

$$R_{\mathrm{auto}}(k) = \frac{P_{\mathrm{rec,rec}}(k)}{P_{\mathrm{HI,HI}}(k)}$$

$$R_{\mathrm{cross}}(k) = \frac{P_{\mathrm{rec,HI}}(k)}{P_{\mathrm{HI,HI}}(k)}$$

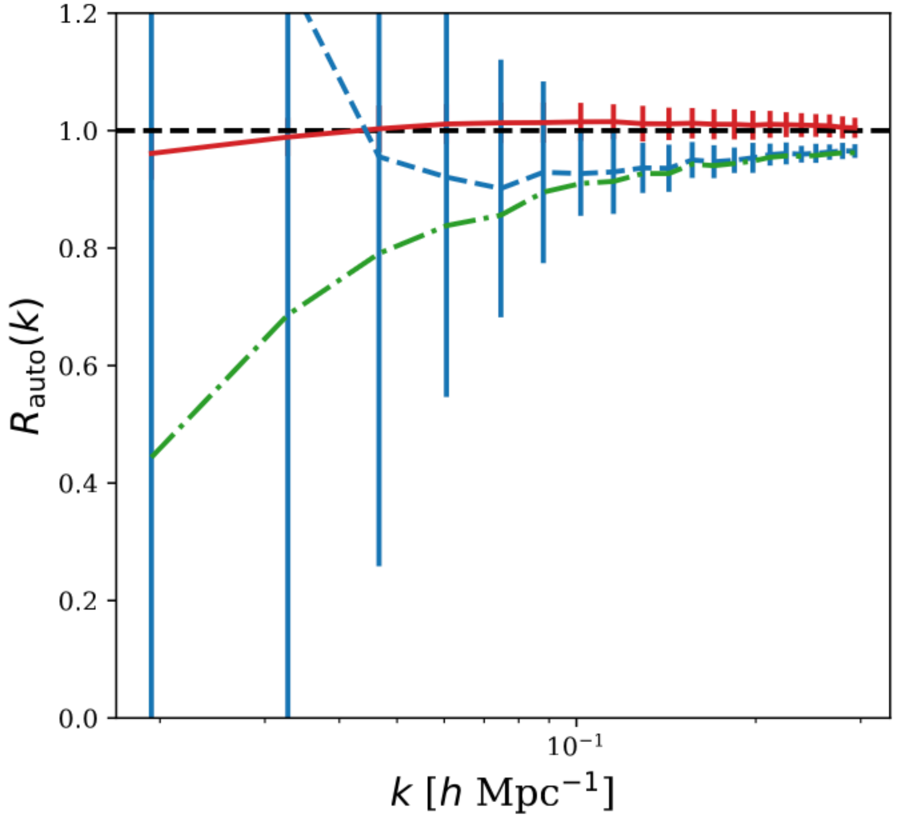


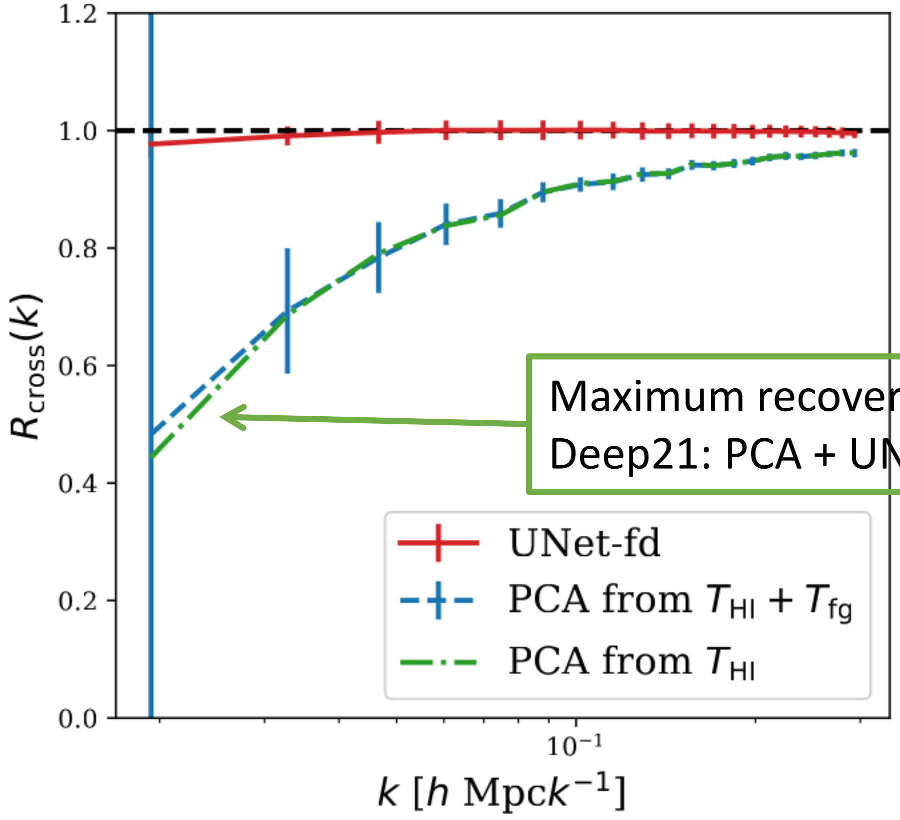


Test I : 21cm+foreground

$$R_{\mathrm{auto}}(k) = \frac{P_{\mathrm{rec,rec}}(k)}{P_{\mathrm{HI,HI}}(k)}$$

$$R_{\mathrm{cross}}(k) = \frac{P_{\mathrm{rec,HI}}(k)}{P_{\mathrm{HI,HI}}(k)}$$

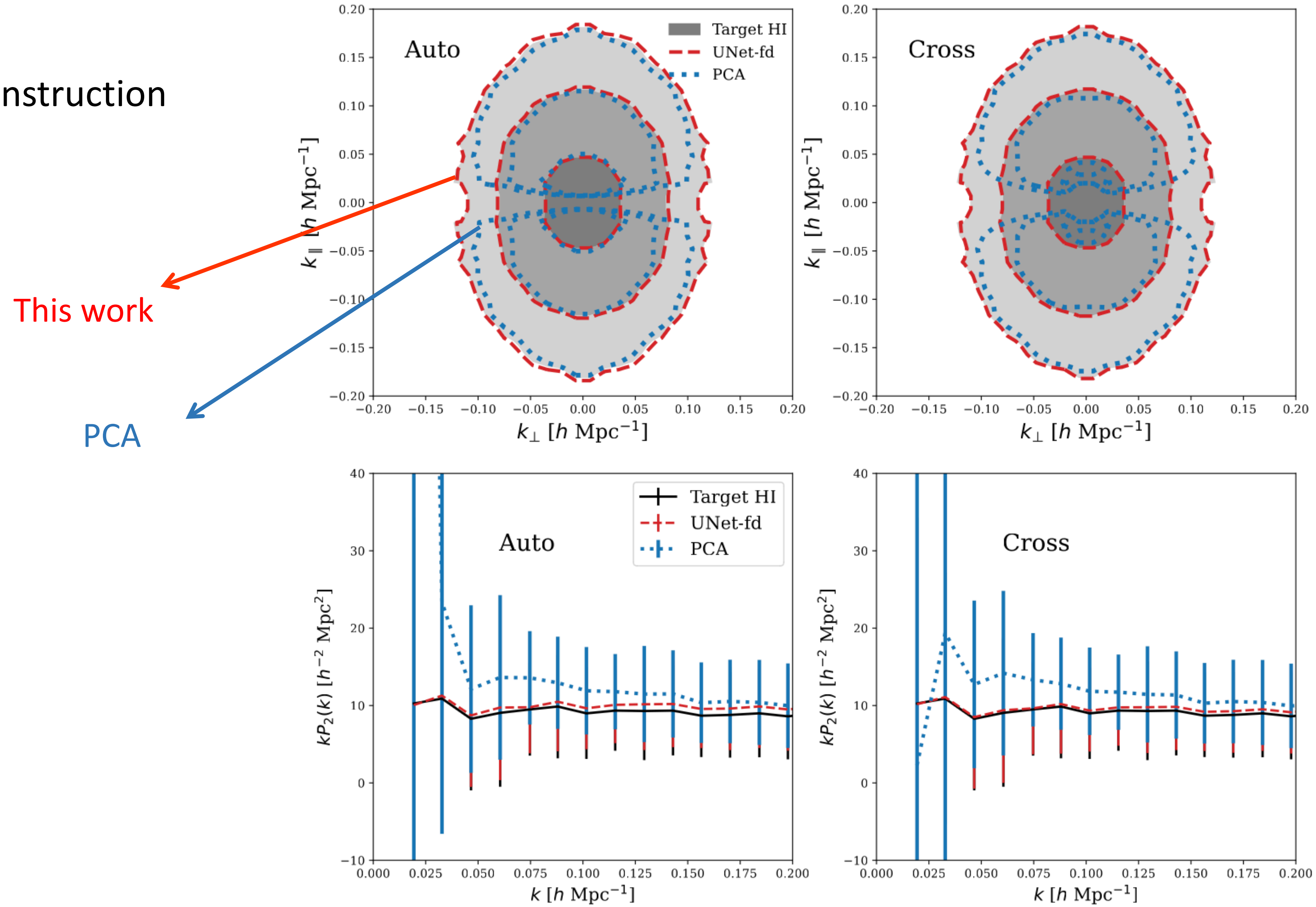




Maximum recovery level of Deep21: PCA + UNet

Test I : 21cm+foreground

RSD reconstruction

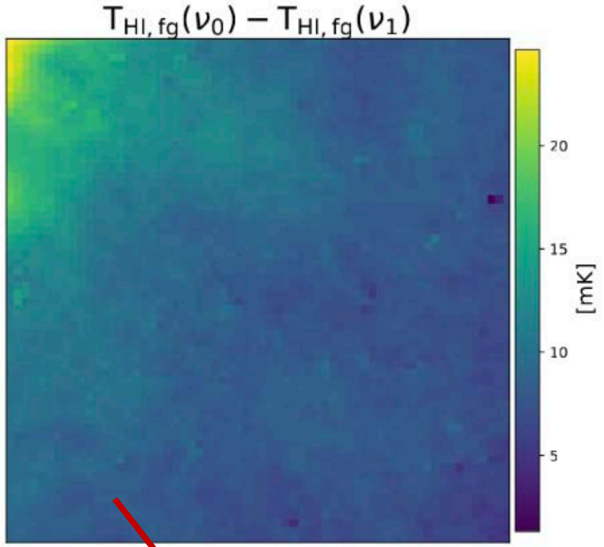


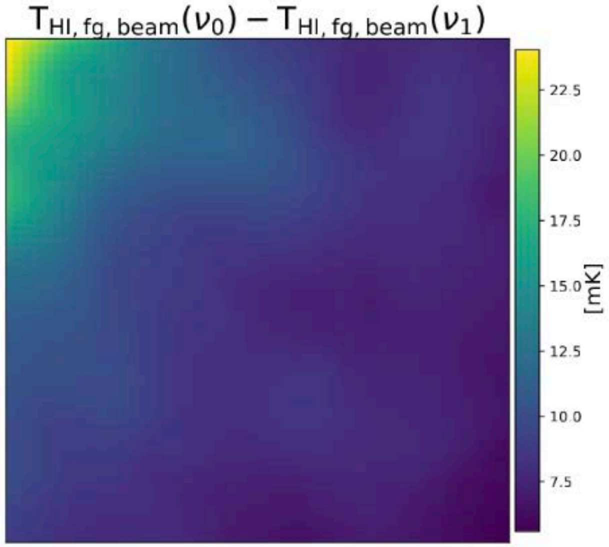
Test II:

- 21cm + forgrounds
- Beam effect
- No thermal noise

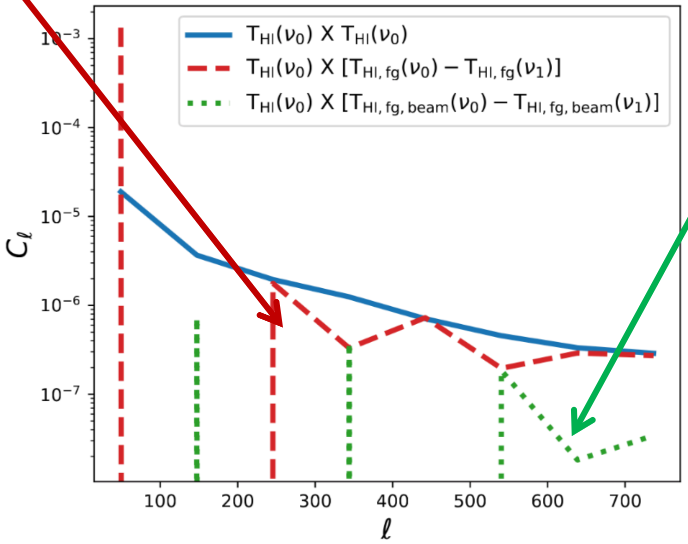
Test II : Adding beam effect

No Beam



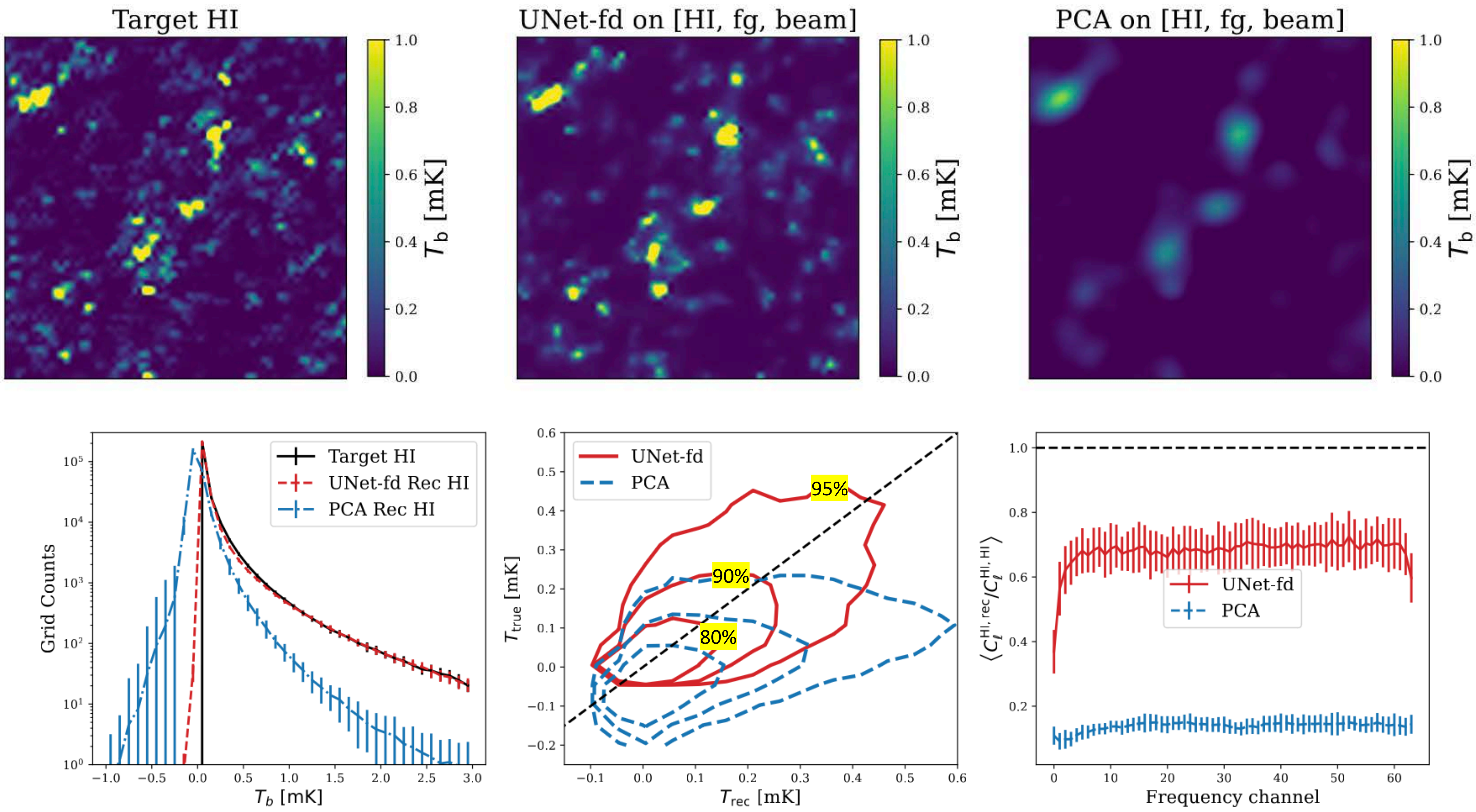


Beam included

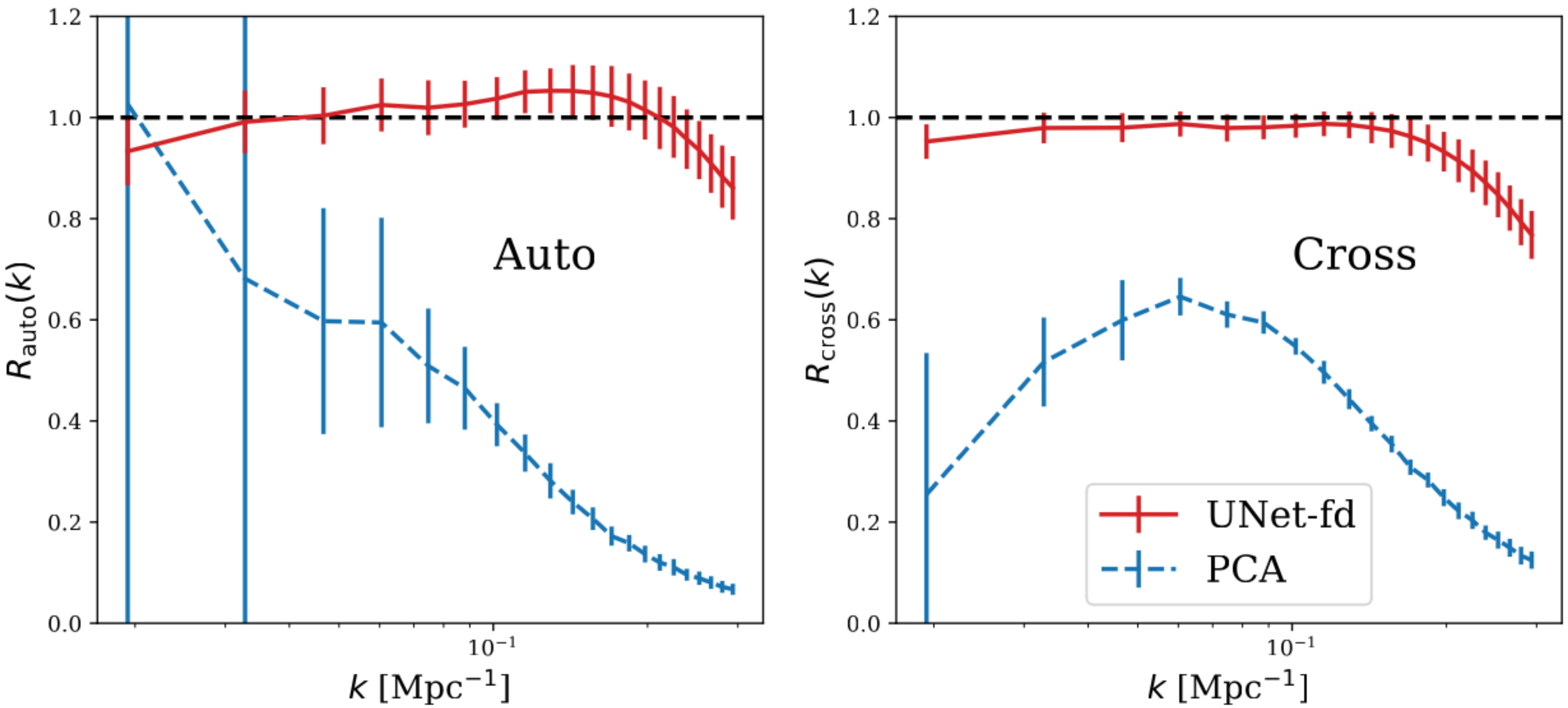


$$B(\nu,\theta) = \exp\left[-4\ln 2\left(\frac{\theta}{\theta_{\mathrm{FWHM}}(\nu)}\right)^2\right]$$

Test II : Adding beam effect



Test II : Adding beam effect



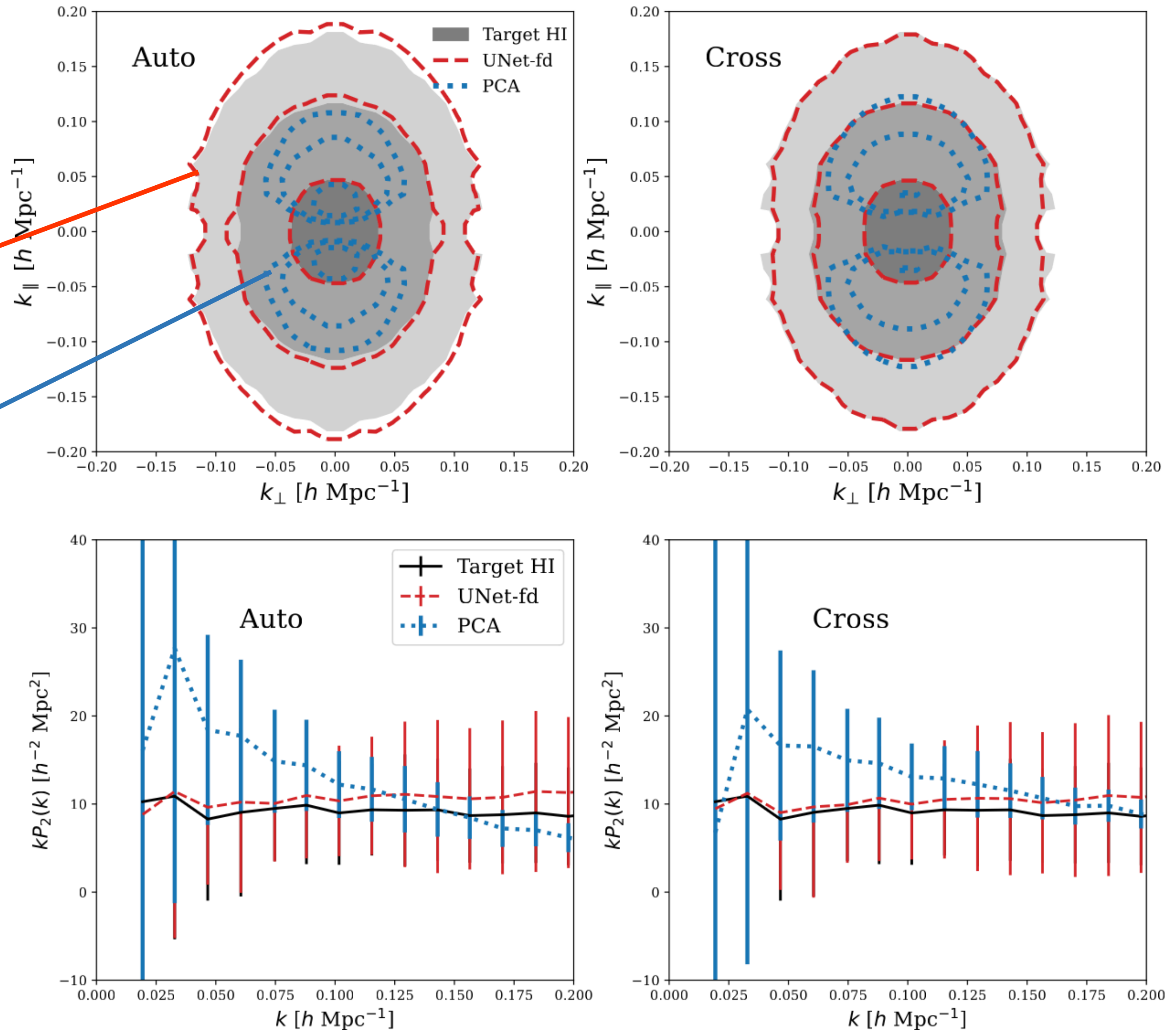
- UNet-fd reconstruction has auto-correlation and cross-correlation P(k) ratios consistently at the 1σ level over the scales k < 0.1 h Mpc^{-1}, with only 10% reduction of the cross-correlation power spectrum at k = 0.2 h Mpc^{-1}
- Our method outperforms the PCA method, whose cross-correlation ratios are underestimated by about 60%

Test II : Adding beam effect

RSD reconstruction

This work

PCA



Test III:

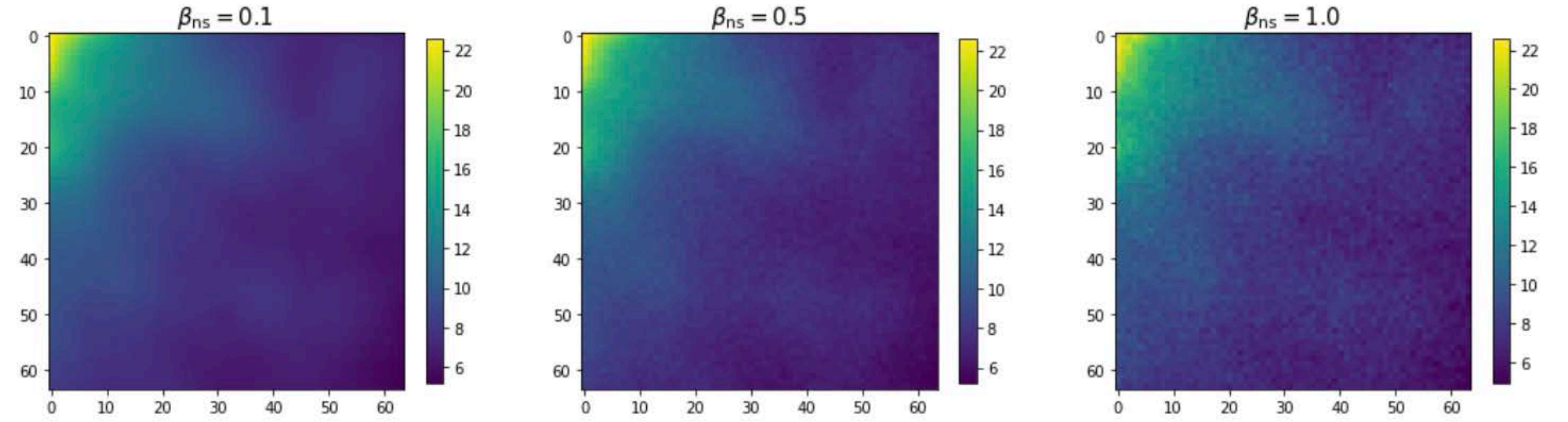
- 21cm + foregrounds
- Beam
- Thermal noise

Test III : Varying thermal noise

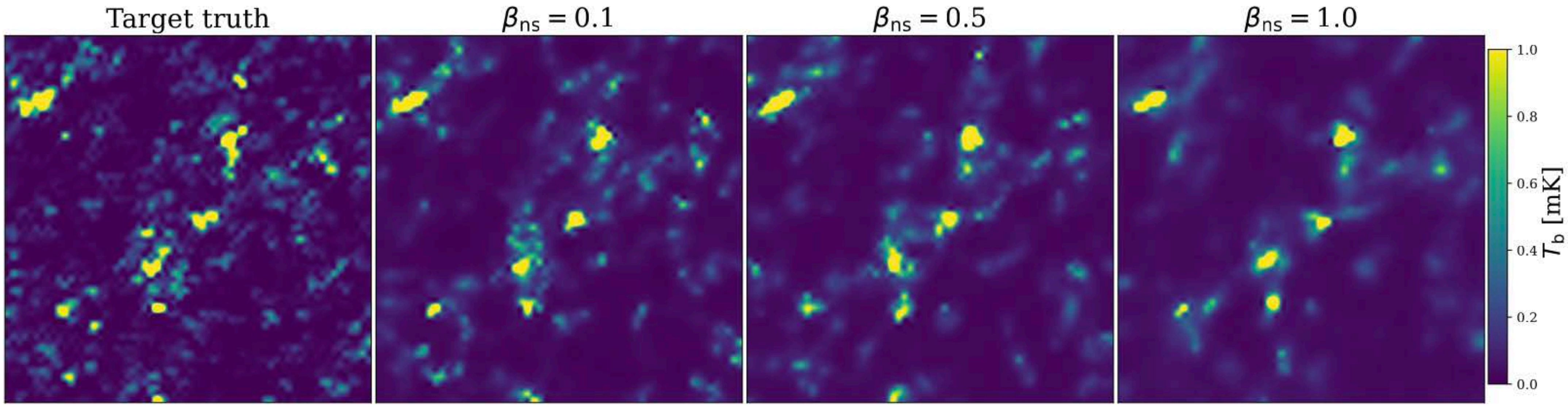
Input:

$$\sigma_{\rm ns} = T_{\rm sys}\sqrt{\frac{4\pi f_{\rm sky}}{\Omega_{\rm beam} N_{\rm dish} t_{\rm obs} \Delta\nu}}$$

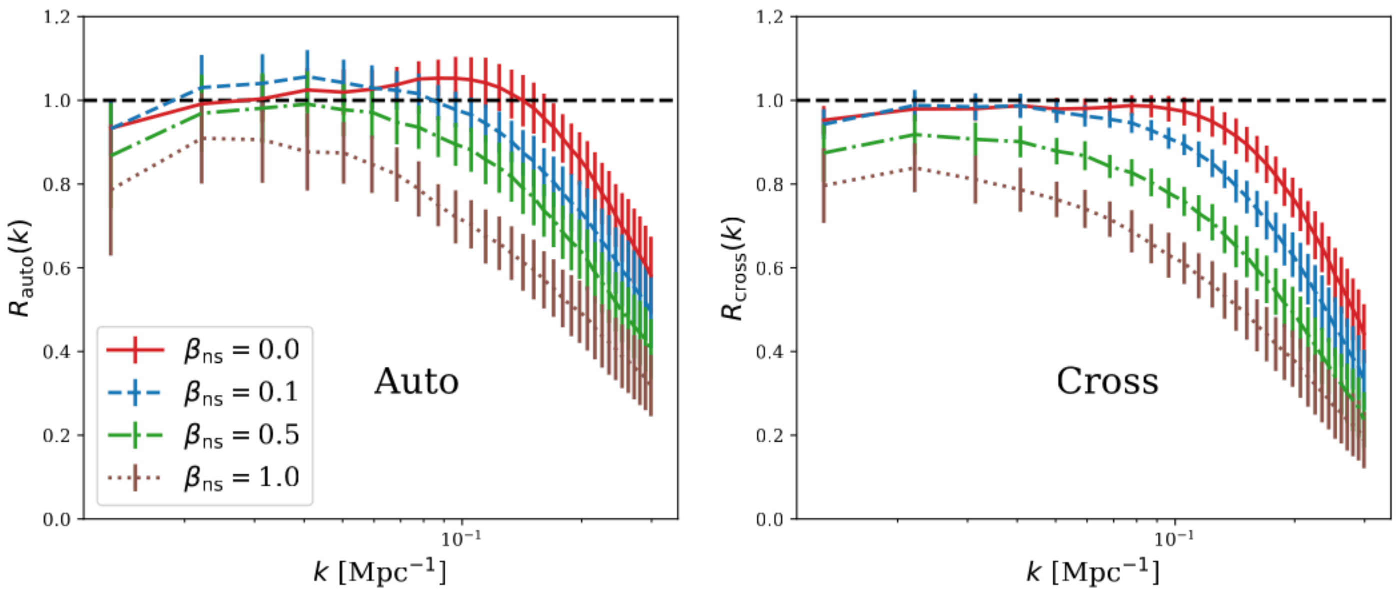
$$\beta_{\rm ns} = \frac{\sigma_{\rm ns}}{\sigma_{\rm HI}}$$



Output



Test III : Varying thermal noise



- For $\beta_{ns} = 0.1$, both $R_{auto}(k)$ and $R_{cross}(k)$ are near to unity at 1σ level over scales $k < 0.06$ h Mpc^{-1}, but have 10% and 30% reduction at $k = 0.1$ h Mpc^{-1} and 0.2 h Mpc^{-1}, respectively.

- For $\beta_{ns} = 0.5$ and 1.0, the auto- and cross-correlation power are underestimated by about 10% and 20%, respectively ($k < 0.06$ h Mpc^{-1}).

Summary & Conlusions

1) Method: Cleaning foreground using U-Net + neighbor-band difference

2) Testing:

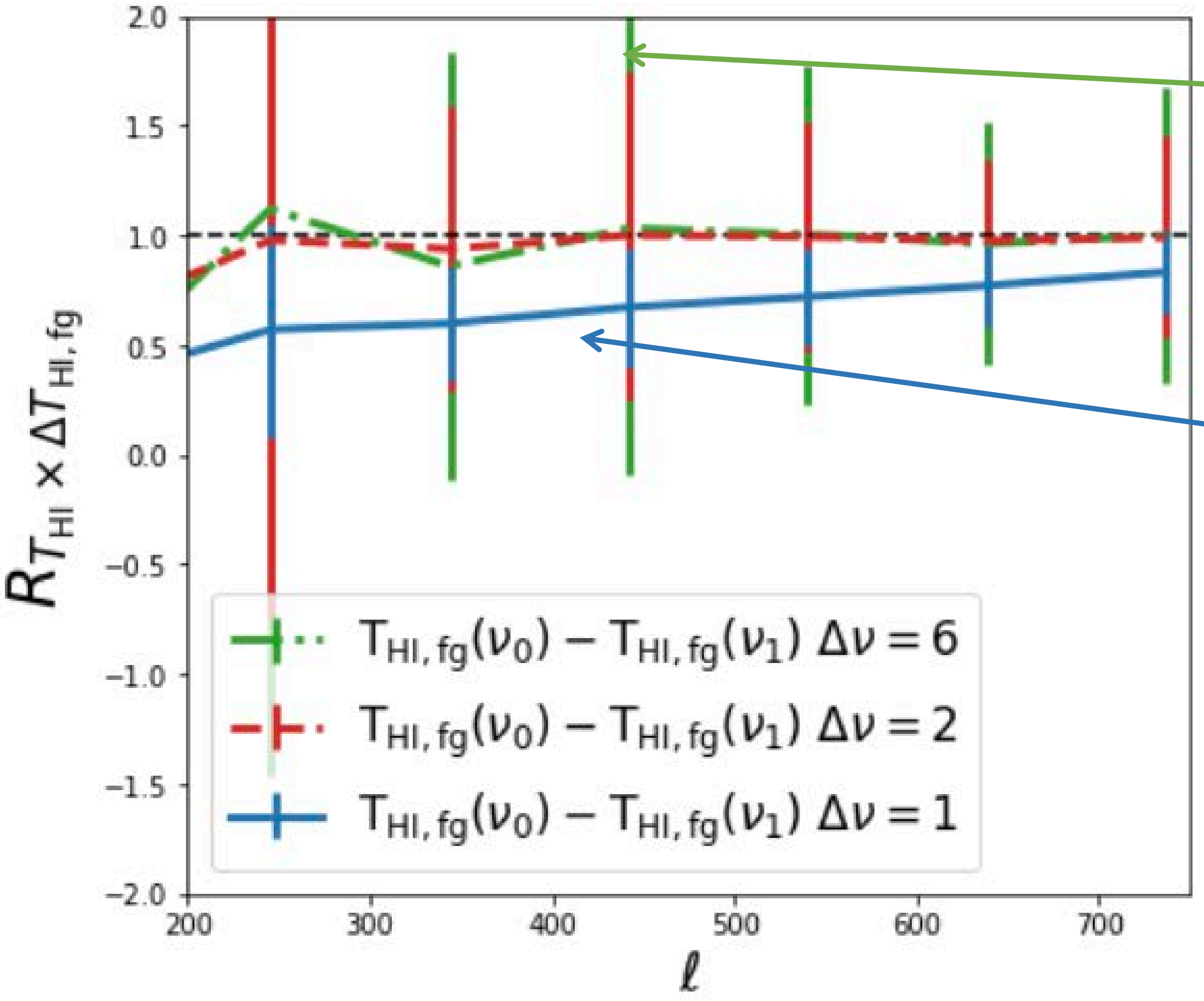
- Foregrounds, beam and noise
- Statistics: image, temperature distribution, 1D & 2D P(k)

3) Conlusions:

- Frequency-band difference can significantly improve network performance by reducing the amplitude range of the smooth foreground components and helping in the prevention of HI loss.
- The HI is recovered consistently at the 1σ level over the scales $k < 0.1\ hMpc^{-1}$
- RSD is also reconstructed successfull
- Our method outperforms outperforms PCA , whose cross-correlation ratios are underestimated by about 75%

Temperature difference between neighboring frequency bands

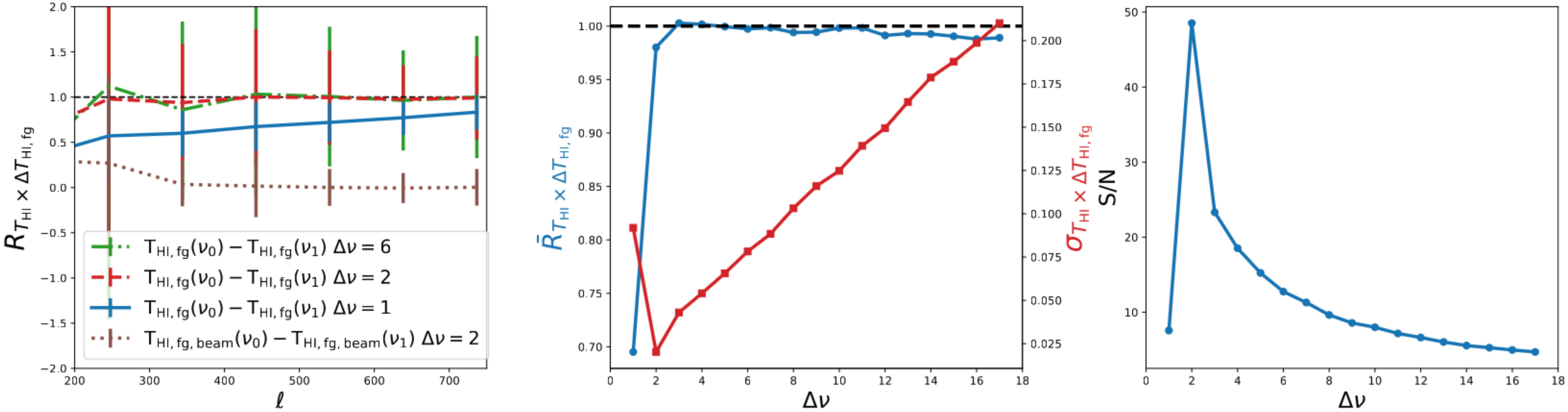
Cross-correlation Cι between pure HI and neighboring band difference/auto-correlation Cι of the pure HI



foregrounds' less correlation at wider bands

large-scale correlations of the HI between narrow bands

Temperature difference between neighboring frequency bands



To show how the neighbor-band difference correlated to the pure HI in different frequency width $\Delta\nu = \nu_0 - \nu_1$

$\Delta\nu = 1$: signal loss $\boxed{\leftarrow}$ large-scale correlations of the HI between narrow bands

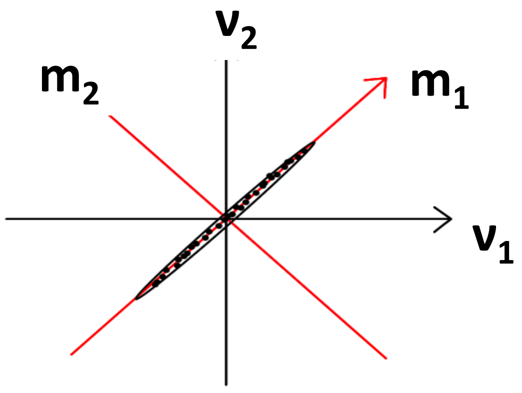
$\Delta\nu > 2$: significantly greater errors $\boxed{\leftarrow}$ foregrounds' less correlation at wider bands

$\Delta\nu$: minimize the HI correlation while maximizing the foreground correlation between neighboring bands.

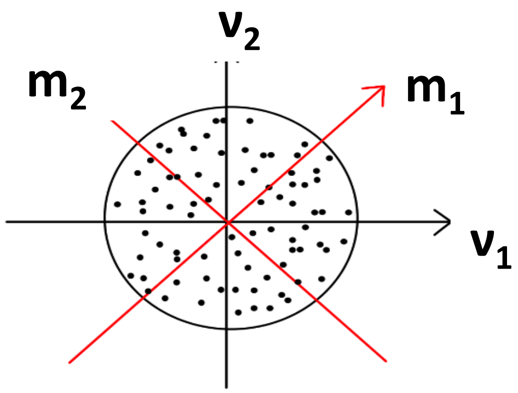
$$\mathrm{S/N}(\Delta\nu) = \sqrt{\frac{\bar{R}_{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}}}{\sigma_{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}}}}.$$

Traditional foreground cleaning

- Biggest challenge in detecting 21cm: foreground!
- Traditional blind techniques, such as PCA,
  - Assume a strong frequency correlation
  - Reduce statistically foregrounds into a few main components
  - Advantage: little prior knowledge of the foregrounds
  - Disadvantage: foreground residuals and signal loss



Foreground

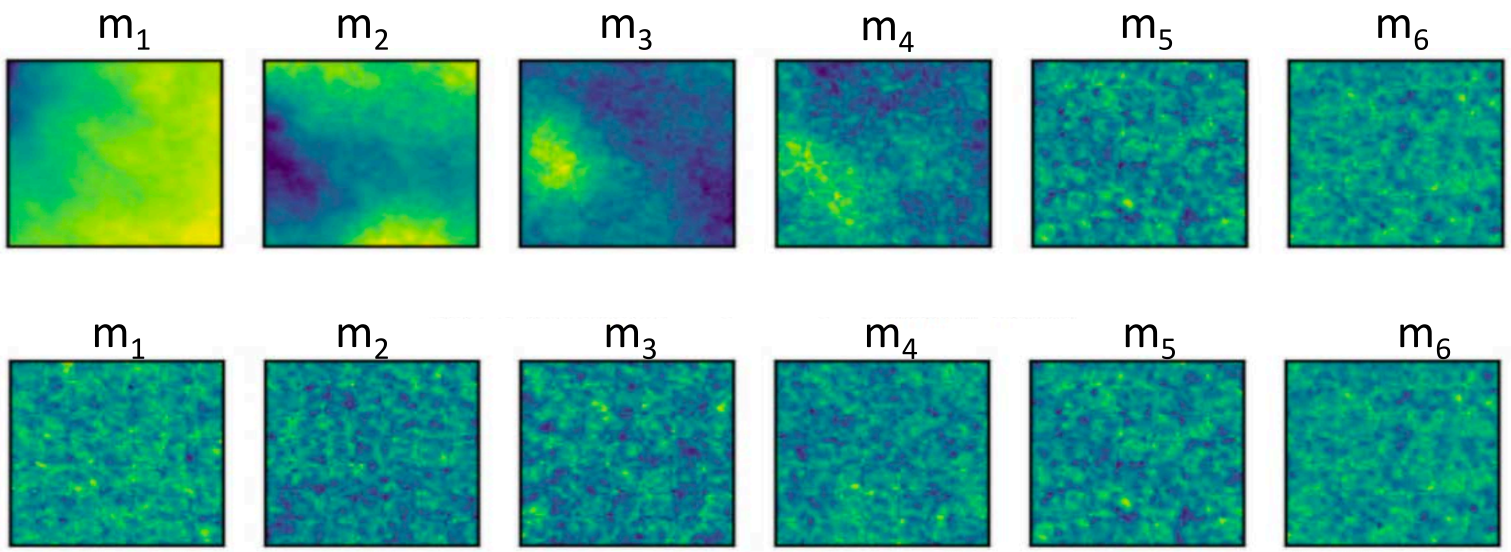


HI

The first 6 principal component maps

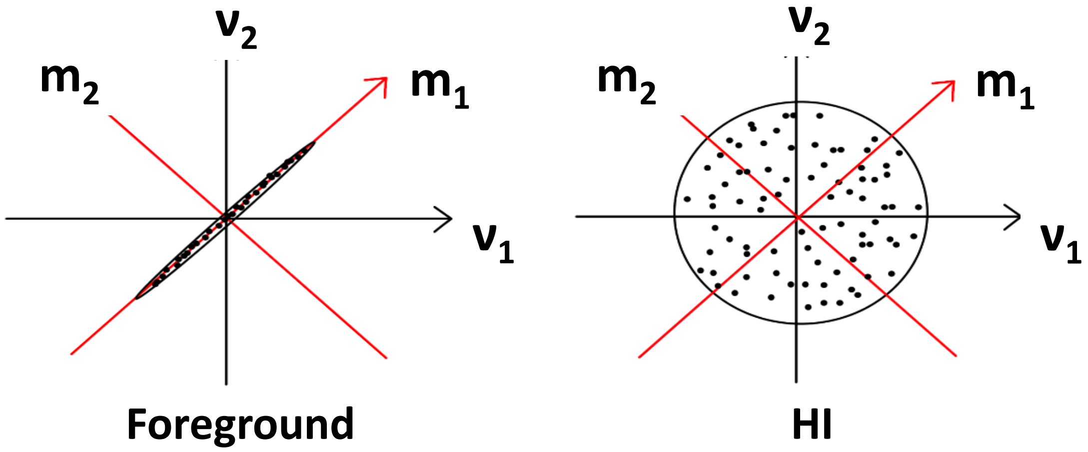
Foreground

HI

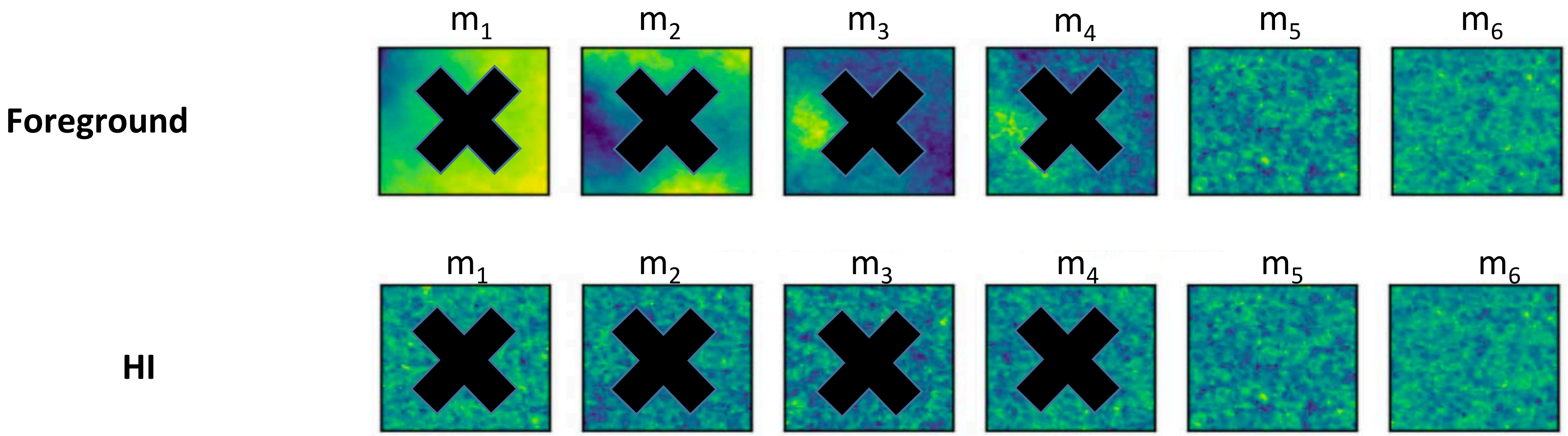


Traditional foreground cleaning

- Biggest challenge in detecting 21cm: foreground!
- Traditional blind techniques, such as PCA,
  - Assume a strong frequency correlation
  - Reduce statistically foregrounds into a few main components
  - Advantage: little prior knowledge of the foregrounds
  - Disadvantage: foreground residuals and signal loss



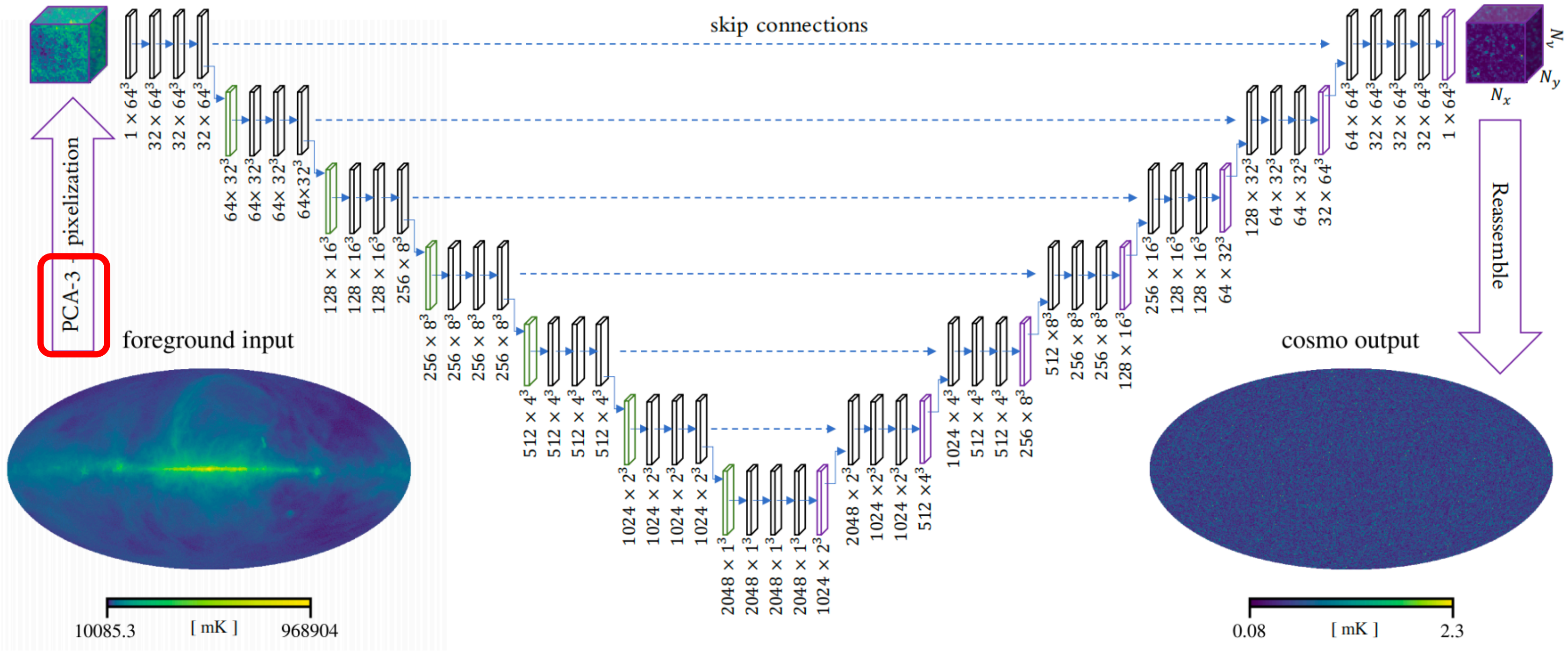
The first 6 principal component maps



New technique based on AI

Deep21: PCA+UNet

(Makinen et al 2021, Ni et al 2022)

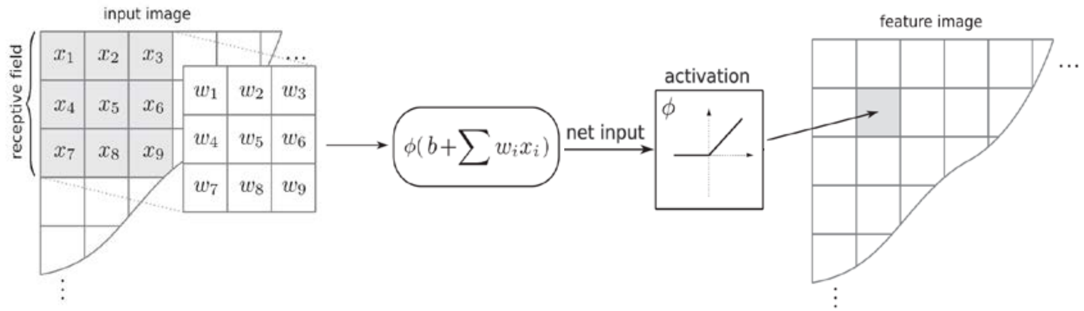


- Perform effectively with a necessary PCA preprocessing
- PCA for removing the smooth foregrounds; UNet for the non-smooth foreground residual
- However, PCA has HI signal loss, which is not expected to be recovered in the UNet

➢ Our solution is still based on the UNet, but expect to avoid the HI signal loss

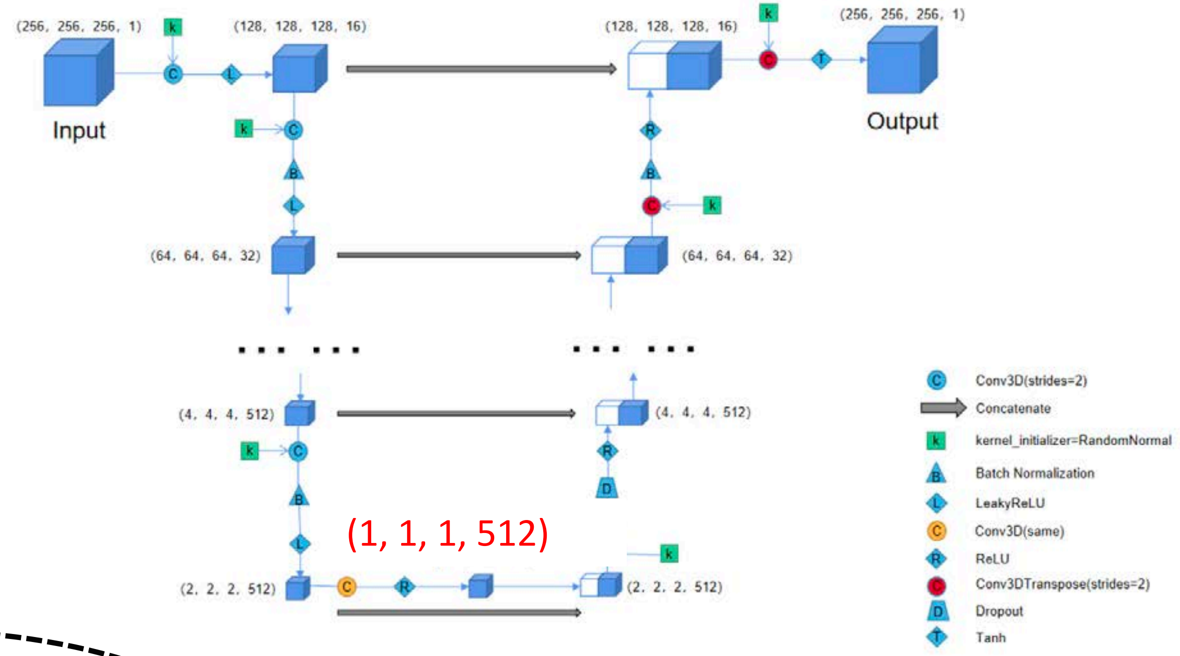
Deeping learning method: UNet model

DarkAI: Reconstructing the large-scale density field of dark matter using AI
(Wang, Shi, et al 2023, arXiv:2305.11431)

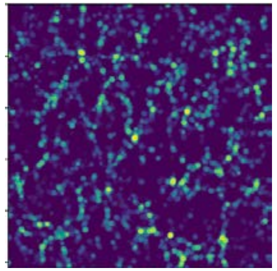


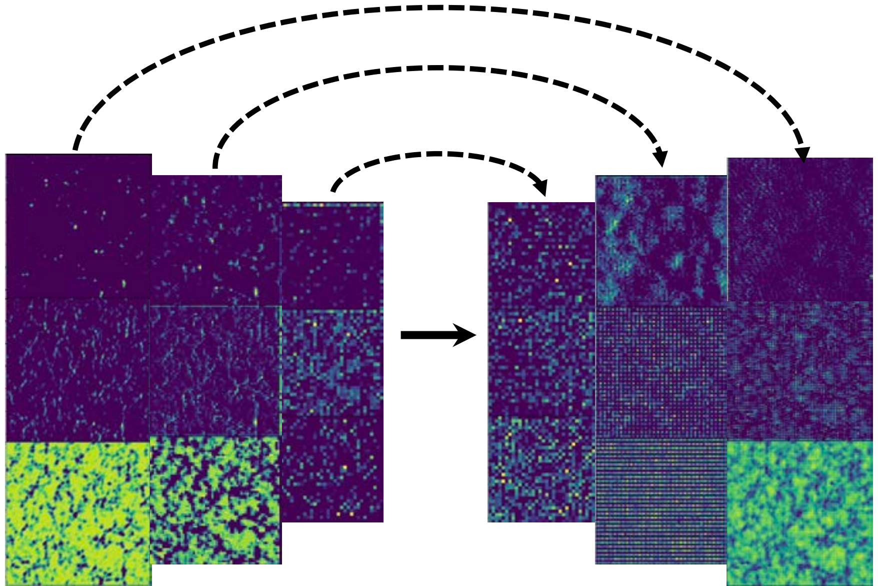
Convolutional neural network (CNN)

Encoder-decoder with skip connections:

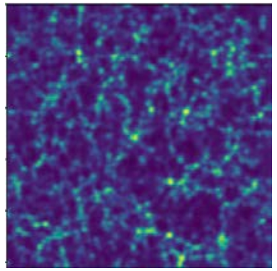


Halo field



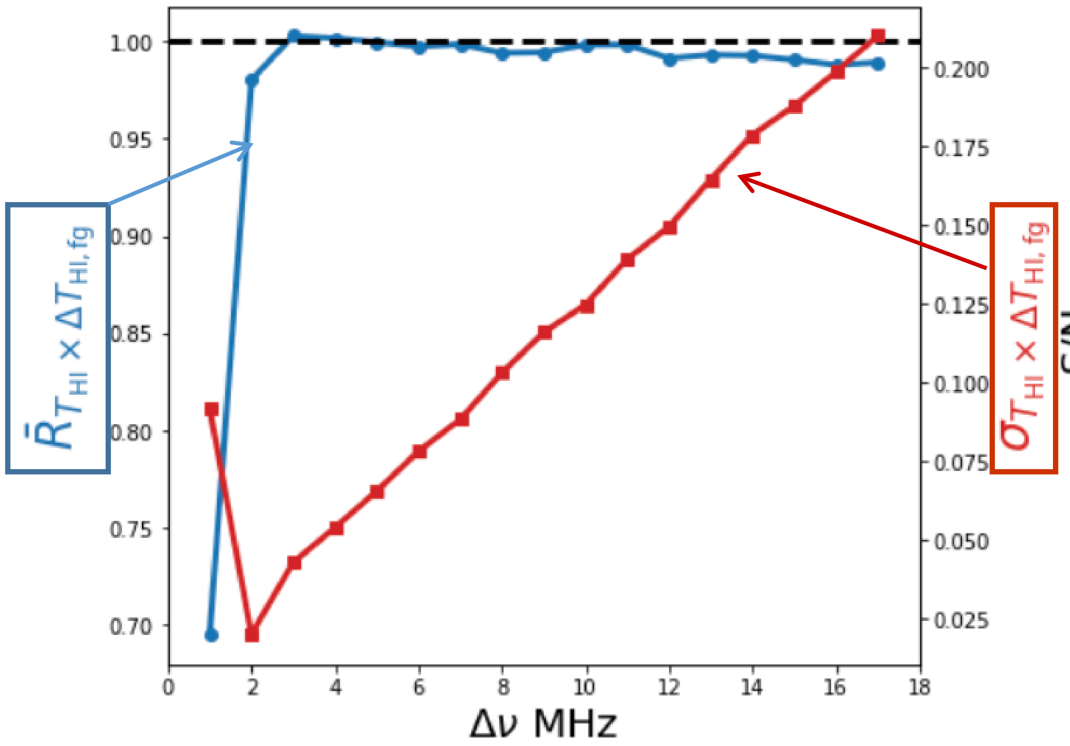


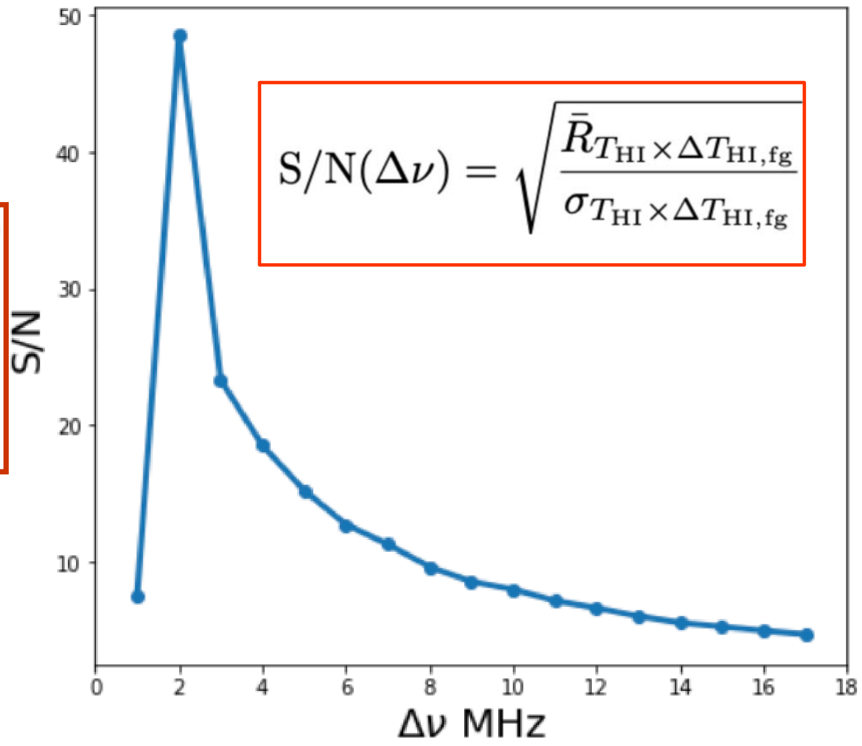
Dark matter field



Temperature difference between neighboring frequency bands

To show how the neighbor-band difference correlated to the pure HI in different frequency width Δν





Δν = 1: signal loss

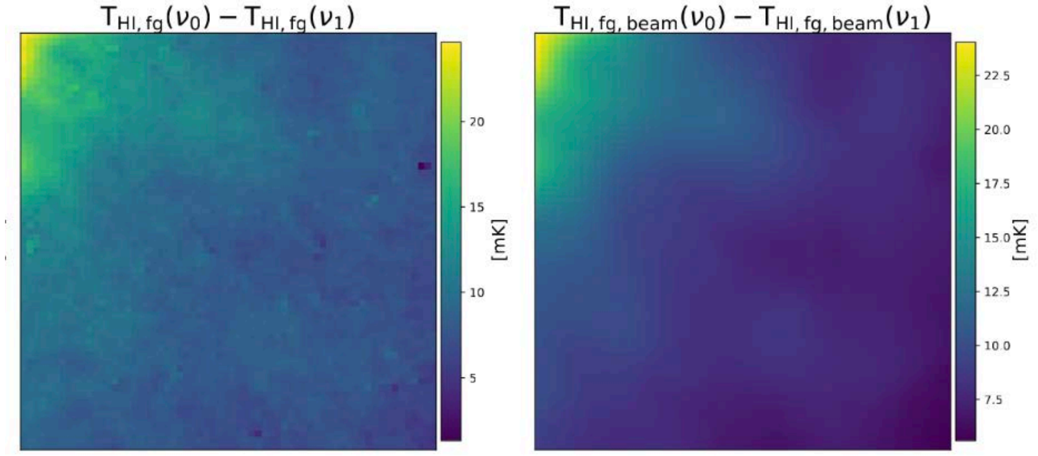
Δν >2 : significantly greater errors

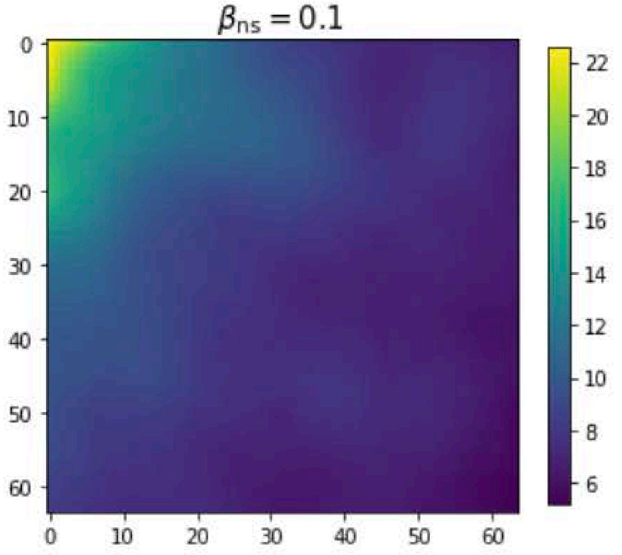
Δν: minimize the HI correlation while maximizing the foreground correlation between neighboring bands.

Test III : Varying thermal noise

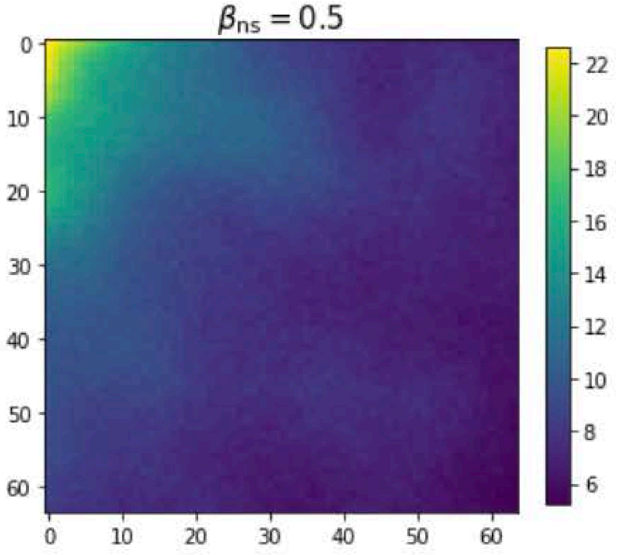
$$\sigma_{\rm ns} = T_{\rm sys}\sqrt{\frac{4\pi f_{\rm sky}}{\Omega_{\rm beam} N_{\rm dish} t_{\rm obs} \Delta\nu}}$$

$$\beta_{\rm ns} = \frac{\sigma_{\rm ns}}{\sigma_{\rm HI}}$$

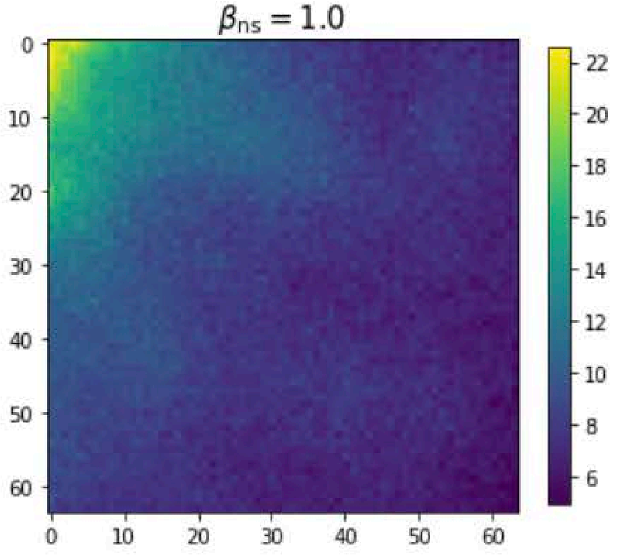




$t_{\rm obs}$= 2000 hours



$t_{\rm obs}$=1000 hours



$t_{\rm obs}$=200 hours

Correlation matrix