Draft version April 1, 2020
Typeset using LaTeX **twocolumn** style in AASTeX63

Inpainting Galactic Foreground Intensity and Polarization maps using Convolutional Neural Networks

Giuseppe Puglisi[1,2,3] and Xiran Bai[1,2]

[1] *Department of Physics, Stanford University, Stanford, CA 94305, USA*
[2] *Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*
[3] *Computational Cosmology Center, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

(Received June 1, 2019; Revised January 10, 2019; Accepted April 1, 2020)

Submitted to ApJ

ABSTRACT

Deep convolutional neural networks have been a popular tool for image generation and restoration. The performance of these networks is related to the capability of learning realistic features from a large dataset. In this work, we applied the problem of inpainting non-Gaussian signal, in the context of Galactic diffuse emissions at the millimetric and sub-millimetric regimes, specifically Synchrotron and Thermal Dust emission. Both of them are affected by contamination at small angular scales due to extra-galactic radio sources (the former) and to dusty star-forming galaxies (the latter). We consider the performances of a nearest-neighbors inpainting technique and compare it with two novels methodologies relying on generative Neural Networks. We show that the generative network is able to reproduce the statistical properties of the ground truth signal more consistently with high confidence level. The Python Inpainter for Cosmological and AStrophysical SOurces (PICASSO) is a package encoding a suite of inpainting methods described in this work and has been made publicly available.

Keywords: convolutional networks, generative adversarial networks, cosmic microwave background, galactic dust simulations, galactic synchrotron simulations, radio sources

1. INTRODUCTION

Over the last few years, the use of machine learning techniques has become increasingly popular in analyzing scientific data. In particular, several features of Deep Convolutional Neural Networks (DCNNs) has opened a wide range of interesting applications (Farsian et al. 2020; Caldeira et al. 2019; Aylor et al. 2019; Krachmalnicoff & Tomasi 2019; Perraudin et al. 2019; Rodríguez et al. 2018; Mustafa et al. 2017).

In this work, we investigate how DCNNs can be used to estimate and reconstruct missing or masked regions of observations. This technique has been widely used for image restoration and face completion ,and it may also be used to reconstruct images of astrophysical signals. The task of *inpainting* masked regions, for example, can be reduced to generate semantically contents for missing pixels based on DCNNs.

gpuglisi@lbl.gov

In particular, we focus on the case of reconstructing polarized signal emitted in the radio and sub-mm regimes: i) at $\nu \lesssim 60\,$GHz where the emission is mostly dominated by Galactic synchrotron, described by a power law $\beta_{synch} \sim -3$ (Krachmalnicoff et al. 2018), ii) at $\nu \gtrsim 150\,$GHz where most of the polarization is due to the thermal Galactic dust grains aligning with the Galactic magnetic field, described by a *modified black-body* law (Planck Collaboration et al. 2018).

At $80 < \nu < 110\,$GHz, the Cosmic Microwave Background (CMB) polarization has a non-negligible contribution especially at high Galactic latitudes. A reliable assessment of both synchrotron and dust polarized emissions in the two regimes i) and ii) is critical to separate the Galactic contamination in CMB measurements and further detect the divergence-less pattern in the CMB polarization called B-mode. CMB B-modes, at degree angular scales, are directly related to the imprint of a stochastic background of gravitational waves produced during the inflationary phase of our universe, commonly referred as *tensorial* anisotropies. To date, *primordial* B-modes have not yet been detected and the latest upper limits have been provided by Sayre et al. (2019); Adachi et al. (2019); BICEP2 Collaboration et al. (2018). Future experiments aim at better characterizing diffuse polarized emission from our own Galaxy with high-sensitivity measurements (Carlstrom et al. 2019; The Simons Observatory Collaboration et al. 2019).

At the arcminute angular scales, B-modes are sourced by the gravitational lensing of large scale structures which deflect the CMB *scalar* polarization anisotropies into the so-called *lensing* B-modes, (see latest constraints in Sayre et al. (2019); POLARBEAR Collaboration et al. (2017); Louis et al. (2017)). At these scales, extra-galactic radio sources and star-forming galaxies are the major polarized contaminants. The majority of these contaminants mostly appears as bright and unresolved point-sources in a typical CMB map (latest measurements can be found in Gupta et al. (2019); Datta et al. (2019)). Puglisi et al. (2018) have shown that hundreds of polarized sources will be detected by the forthcoming experiments given the expected nominal sensitivity and the observation sky fraction ($\sim 10-30\%$). Hence an aggressive masking may be applied on maps surveyed by the forthcoming CMB experiments, preventing a high-resolution Galactic foreground template as well as a reliable analysis involving high-order estimators beyond the two-point correlation function. Reconstructing signals in the masking area to fill the missing data is done to ameliorate these issues, a procedure sometimes referred to as *inpainting* (used in e.g. Starck et al. (2013)).

In this work, three different methodologies are tested to *inpaint* maps at the locations of extra-galactic point-sources. Two of the inpainting techniques involve *generative* DCNNs. We compare the DCNNs inpainting performances with the standard diffusive inpainting approach used in Bucher et al. (2016), which is simply filling the missing pixel with the average value of its nearest-neighbours.

We organize the paper as follows: In Sect.2, we present the three inpainting methodologies adopted in this work. Sect.3 describes the data used for training and validation purposes. Finally, Sect.4 includes the results achieved by inpainting on simulations (Subsect.4.1) and on more realistic data-sets (Subsect.4.2). Finally, we apply our inpainting method to the the map of the S-band Polarization All Sky Survey (SPASS, Carretti et al. (2019)) at several source locations (Sect.4.3) and demonstrate that we robustly recover the background signal.

2. METHODS OF INPAINTING

Inpainting algorithms can be divided into two main groups: i) diffusive-based methods and ii) learning-based methods that rely on training DCNNs to fill the missing pixels with the predictions learned from a training data-set. We choose three inpainting techniques from both groups: Sect.2.1 describes a diffusive-based method from group i), and Sect. 2.2 and 2.3 present methods from group ii).

2.1. *Nearest-Neighbors*

One of the simplest inpainting methods is the *diffusive inpainting* described in Bucher et al. (2016), which has been adopted in Ade et al. (2014, 2016). In this method, each masked pixel is iteratively filled with the mean value of its nearest-neighbor pixels, being often referred to as the Nearest-Neighbors (NN) in the image reconstruction algorithm.

The iterative procedure can be performed in two ways: i) *Gauss-Seidel* method, which computes the average of neighbors at the current iteration. As a consequence, the pixels near the boundary are updated in earlier iterations while pixels near the center of the inpainting regions require several iterations. ii) *Jacobi* method, which estimates the average value from a buffer of pixel values at the previous iteration. Bucher et al. (2016) found that $\sim \mathcal{O}(10^3)$ iterations were needed to inpaint ~ 10 arcmin areas on a map with ~ 2 arcmin pixel size. Although Bucher et al. (2016) found that both methods did not impact the quality of the inpainted results, the Gauss-Seidel method achieves faster convergence than the Jacobi method. We therefore adopted the former method as suggested by Bucher et al. (2016).

2.2. *Deep-Prior*

Deep-Prior (DP; Ulyanov et al. 2017) relies on *untrained* DCNNs. As opposed to the common approach of training a network on a large data-set, DP fits a generator network to a single image and uses the network weights to parametrize the reconstructed image. The weights are fitted to maximize a likelihood which can be specific to the task (e.g. super-resolution, inpainting, de-noising). Ulyanov et al. (2017) showed that the information required to restore an image is encoded in the single degraded input image and in the network architecture used for the reconstruction.

Inpainting can be formalized as an *image generating* procedure via training an encoder-decoder network, $x = f_\theta(z)$. The network maps a random vector z, known as the *the prior*, to an image x, through a parametrization f_θ that involves convolutions, non-linear activations with Leaky Rectified Linear Units (LeakyReLUs) and upsampling.

For the task of inpainting, we thus follow the prescription given in Ulyanov et al. (2017)[1], and build an U-Net type architecture without skip-connections, summarized in Table A2. Given an image x_0 with width W and height H, and with missing pixels correspondent to a binary mask m, with zero values in the masked region, the goal is to minimize the following loss function:

$$E(x^*, x_0) = \| (f_{\theta^*}(z) - x_0) \odot m \|^2, \tag{1}$$

with $\odot$ being the Hadamard's product, $x^* = f_{\theta^*}(z)$ being the output of the reconstruction and θ^* the optimal set of weights obtained by minimizing eq. (1) with gradient descent. Notice that this energy function does not depend on the values of the missing pixels. Furthermore, Ulyanov et al. (2017) showed that this parametrization presents a high impedance to noise allowing convergence to *naturally-looking images* based on the features outside the mask. Details of the DP architecture can be found in the Appendix Tab.A2 and A3.

2.3. *Generative Adversarial Networks with Contextual Attention*

Generative Adversarial Networks (GAN; Goodfellow et al. 2014) is a popular machine learning approach used in image restoration problems. The GAN framework consists of a generator, trained to generate indistinguishable samples from a training set, and a discriminator, trained to evaluate whether a sample generated by the generator is distinguishable from the simulated data. Generally in a GAN, both the generator and the discriminator networks are simultaneously optimized by means of gradient descent.

As a result, this kind of networks seek for coherency between generated and existing pixels by jointly training an adversarial network with a convolutional encoder-decoder. One of the biggest advantages of using GAN based methods is that they have been proven to be less affected by problems observed in simple convolutional networks (as variational auto-encoders), such as boundary artifacts and blurry textures that are inconsistent with the surrounding regions Iizuka et al. (2017); Yu & Koltun (2015).

Yu et al. (2018) presented a novel generative inpainting network with a *contextual attention* branch. The proposed network accounts for two stages: a *coarse reconstruction* stage which employs a simple dilated DCNN trained to inpaint the missing region, and a *refinement* stage aimed at improving the inpainted features both locally and globally with contextual attention. The general idea of contextual attention is to exploit the features in the known area of the image as convolutional filters that will be consecutively applied to the reconstructed regions. The overall architecture used in generative inpainting is sketched in Yu et al. (2018, Fig.2, 3 and 4)

The coarse stage represents the first stage and it is essentially based on an encoder-decoder network trained with reconstruction loss to quickly approximate the content in the missing region. The second stage is also an encoder-decoder network. In this case, the encoder part of the network is organized into two parallel convolutional pathways both fed with the output of the coarse stage, i. e. the full image with an approximated content in the missing region. One branch aims at hallucinating novel contents in the missing region in order to better refine the content inside the mask by injecting smaller scale features. The other branch encodes the contextual attention and it is meant to enhance spatial coherency of the local features inside the masked area with the global features. To better visualize how the contextual attention works we report in Fig.B1 an example of inpainting with GAN and the attention map related to this case. The attention map helps to identify which regions of the input image are attended by contextual-attention to refine the corrupted image.

The output of the two branches is then combined and fed to a single decoder for the final inpainting output. Both the coarse and refinement stages are trained *end-to-end* with reconstruction losses for the generator network and two Wasserstein GAN (WGAN) adversarial losses (Arjovsky et al. 2017) devoted to serve as global and local image critics for the discriminator.

To apply GAN in the problem of inpainting, we follow the prescriptions in Yu et al. (2018) and we adopted the reconstruction loss as a weighted sum of pixel-wise ℓ_1 loss (instead of the commonly adopted mean-squared-error) to train explicitly the coarse and the refinement networks. Furthermore, the training for the latter, is performed by combining together the reconstruction and WGAN losses, so that the encoder network in the refinement stage has access to a more complete scene than the one in the coarse stage and better learn how to represent features.

The generative network G takes the image with missing pixels (x_0), and the corresponding binary mask (m), as inputs to generate a image with the missing pixels

[1] Further details can be found in supplementary material, https://dmitryulyanov.github.io/deep_image_prior.

filled. The final output $x' = G(x_0, m)$ is then the result of pasting the generated masked region on top of the uncorrupted one, i.e.

$$\tilde{x} = x_0 \odot m + x' \odot (1 - m). \quad (2)$$

The network architectures adopted for the GAN is summarized in Table A1 and the hyper-parameters network setting is chosen to be the same as the one provided in Yu et al. (2018). Each convolution layer is implemented by using mirror padding, without batch normalization and Exponential Linear Units (ELUs, Clevert et al. (2015)) activation functions.

3. DATA AND SIMULATIONS

In this study, we mainly focus on inpainting maps of polarized emissions in two microwave regimes: namely synchrotron and thermal dust. These two foreground components contaminate the CMB polarization measurements and need to be modeled both at large and small angular scales for foreground removal. Since the statistical properties of Galactic foregrounds are highly non-Gaussian, an interesting application of DCNNs is to reconstruct images with complex, non-Gaussian features. In contrast, traditional inpainting methods used in CMB studies such as the Gaussian Constrained Inpainting methods (Hoffman & Ribak 1991; Bucher & Louis 2012) assume that the background is a Gaussian field, making it incapable of capturing the non-Gaussianity in the foregrounds.

Both Galactic foregrounds, unpolarized and polarized emissions data (respectively encoded in the brightness temperature, T and *Stokes* parameters Q and U maps) are simulated using the PySM package (Thorne et al. 2017) and will be described below in Subsections 3.1 and 3.2.

3.1. *Galactic Synchrotron*

For the synchrotron data, we consider SPASS Carretti et al. (2019) which observed the Southern sky ($\delta < -1°$) at 2.3 GHz with an 8.9 arcmin full width at half maximum (FWHM). The methodology used to generate the intensity and polarization maps are described in Carretti et al. 2019. 98.6% of the pixels in the Q and U SPASS maps have signal-to-noise ratio (SNR) $>$ 3, making these maps a promising synchrotron polarization template (Krachmalnicoff et al. 2018). Lamee et al. (2016) cross-matched the extra-galactic radio quasars (mostly steep spectrum sources) detected by SPASS with the ones detected by NRAO/VLA Sky Survey, (NVSS, Condon et al. (1998)), at 1.4 GHz and released a polarization catalogue with 533 bright sources in the overlapping area of the two surveys. However, because the SPASS T, Q and U maps are filled with radio sources, an assessment on the map level to the smallest angular scales is essentially compromised by the point-source bias. Therefore, masking and inpainting these sources provide an overall benefit in fully exploiting the angular scales probed by SPASS.

In order to inpaint the SPASS map, we firstly create a simulated training data-set which will be used for training the GAN (training set) and for evaluating the quality of the reconstructions (testing set). We therefore simulate SPASS synchrotron-only TQU maps at the SPASS frequency with `s1` PySM model. This model is one of the most representative since it parametrizes the synchrotron power-law emission with a spatially varying spectral index. Maps are pixellized on a HEALPix [2](Górski et al. 2005; Zonca et al. 2019) `nside=2048` grid convolved with a 8.9 arcmin FWHM beam.

3.2. *Thermal Dust*

We use the thermal dust maps at 353 and 857 GHz from the third *Planck* public release[3]. Both frequency maps are dominated by the thermal dust emission emitted by our own Galaxy and encode contribution from Cosmic Infrared Background (CIB). We choose the 353 GHz frequency channel in order to test the inpainting techniques on both dust temperature and polarization maps. Because the 857 GHz channel is not polarization sensitive, we use this channel to assess how the different SNR and different CIB contribution affect inpainting reconstructions in total intensity. At these frequencies, the emission is dominated by star forming galaxies, and blazars are expected to have minor contribution to the total intensity.

We build the training set by simulating TQU Thermal dust maps at 353 GHz with the `d1` PySM model, which describes the modified blackbody emission law with a spectral index and a temperature, both spatially varying. Maps are simulated on a `nside=2048` grid and convolved with a 5 arcmin FWHM beam, similarly to the one in the *Planck* 353 GHz observations.

Furthermore, since the pixel values of the maps are rescaled during the training and inpainting processes, the GAN reconstruction is not affected by the overall amplitude of a signal. We will show in Sec.4.2 that inpainting performances do not change as a function of frequency as long as the brightest signal in the map coincides with the one used in the training. This independence of frequency is the reason why we trained the

[2] https://healpix.sourceforge.io

[3] https://pla.esac.esa.int

GAN for thermal dust emission using PySM signal-only simulations with single frequency at 353 GHz.

3.3. *The Training Data-set*

Both the training and testing data-sets are made from the PySM simulated maps. We forecast with PS4C package [4] (Puglisi et al. 2018), the number of sources, $N_{\rm src}$ whose density fluxes will be detected at 5σ significance above the sensitivity flux. For a generic large aperture forthcoming CMB experiment (e. g. The Simons Observatory Collaboration et al. (2019)), the forecasted detections is about $N_{\rm src} \sim 30\,000$. We then generate a point-source mask by randomly extracting N_{src} locations following a *Poisson* distribution. We mask the sources with circular holes centered at the source locations and with a radius three times larger than the beam FWHM size (namely 26.7 and 15 arcmin respectively for synchrotron and dust maps).

Both masked and unmasked maps are then split into $3\times3\,\deg^2$ square tiles, composed of 128×128 pixels with resolution closer to the HEALPix one (i. e. ~ 1.5 arcmin at `nside=2048`). We finally build the training set for the GAN network by combining 45 000 images from square patches extracted equally from T, Q and U maps. The remaining 5000 images are used for validation and 500 for testing.

4. RESULTS

Figures 1, 2 and B3 show examples of maps extracted from the test set and reconstructed with the three methods outlined in Section 2. We estimate the minimum and maximum values of each ground-truth image to rescale it (together with the respective inpainted ones) to $[0, 1]$ with the `MinMax` normalization. This rescaling forces the generated maps to have the same range as the test ones, so the differences between the ground-truth and reconstructed map can be spotted more easily. Notice that we further zoom in dust maps ($1.5\times1.5\,\deg^2$ crops) to better inspect the inpainted region.

As expected, the inpainting performed with NN algorithm is smooth and lacks of finer details, making them distinguishable from the original map.

On the contrary, for the inpainting performed with DP and GAN, it is harder to point out which one is the ground-truth and which is the reconstructed maps. Moreover, we note that DP and GAN are able to reproduce the large scale features, the most correlated angular scales, as well as the typical Q/U pattern for the polarization maps.

[4] https://gitlab.com/giuse.puglisi/PS4C

4.1. *Evaluation of fidelity*

We employ several methods used in the literature to assess the quality of the reconstructed maps quantitatively. First, we follow the approach from Mustafa et al. (2017) and Aylor et al. (2019), focusing on evaluating the ability to replicate the summary statistics of the underlying signal that needs to be reconstructed. Those statistics are based on i) the pixel intensity distribution, ii) the angular power spectra of the two point correlation function, and iii) the first three Minkowski functionals. We, therefore, consider the inpainting as *successful* if it passes these three statistical tests.

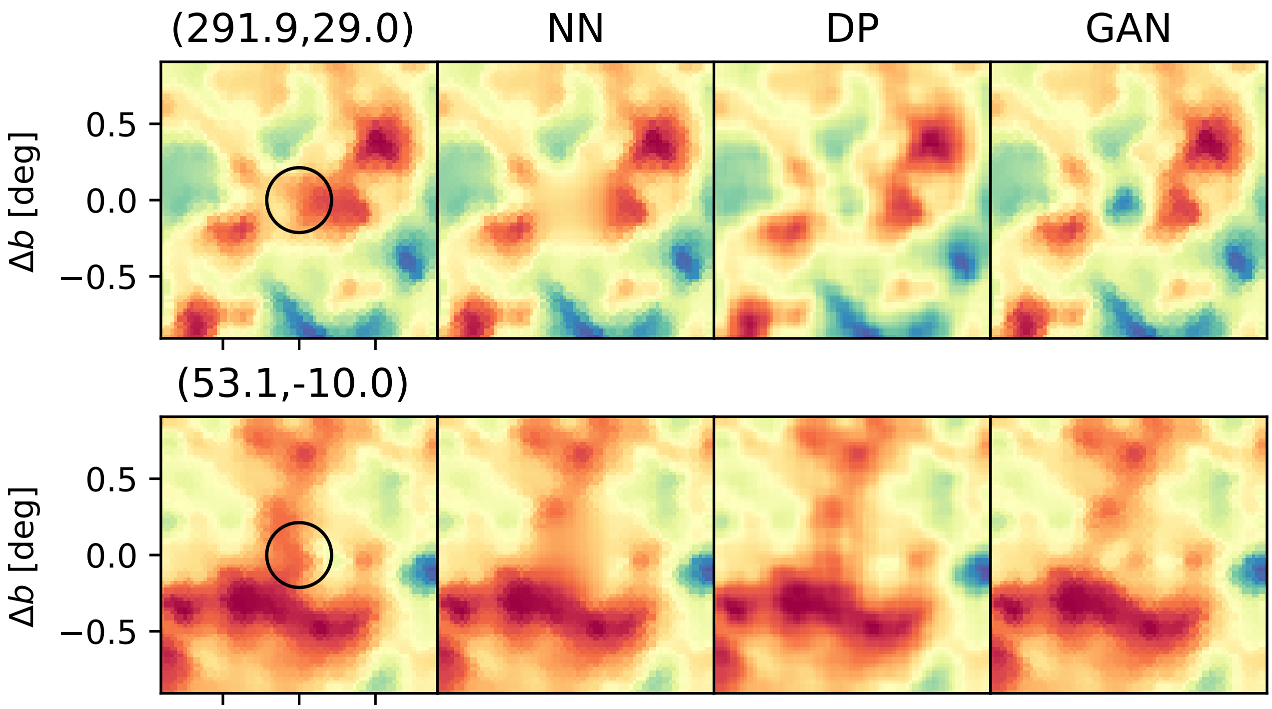
	Synchrotron	Thermal Dust
NN	> 0.997	> 0.999
DP	> 0.963	> 0.999
GAN	> 0.861	> 0.997

Table 1. p-value of KS test performed on the pixel intensity distribution of Q maps shown in Fig.3.

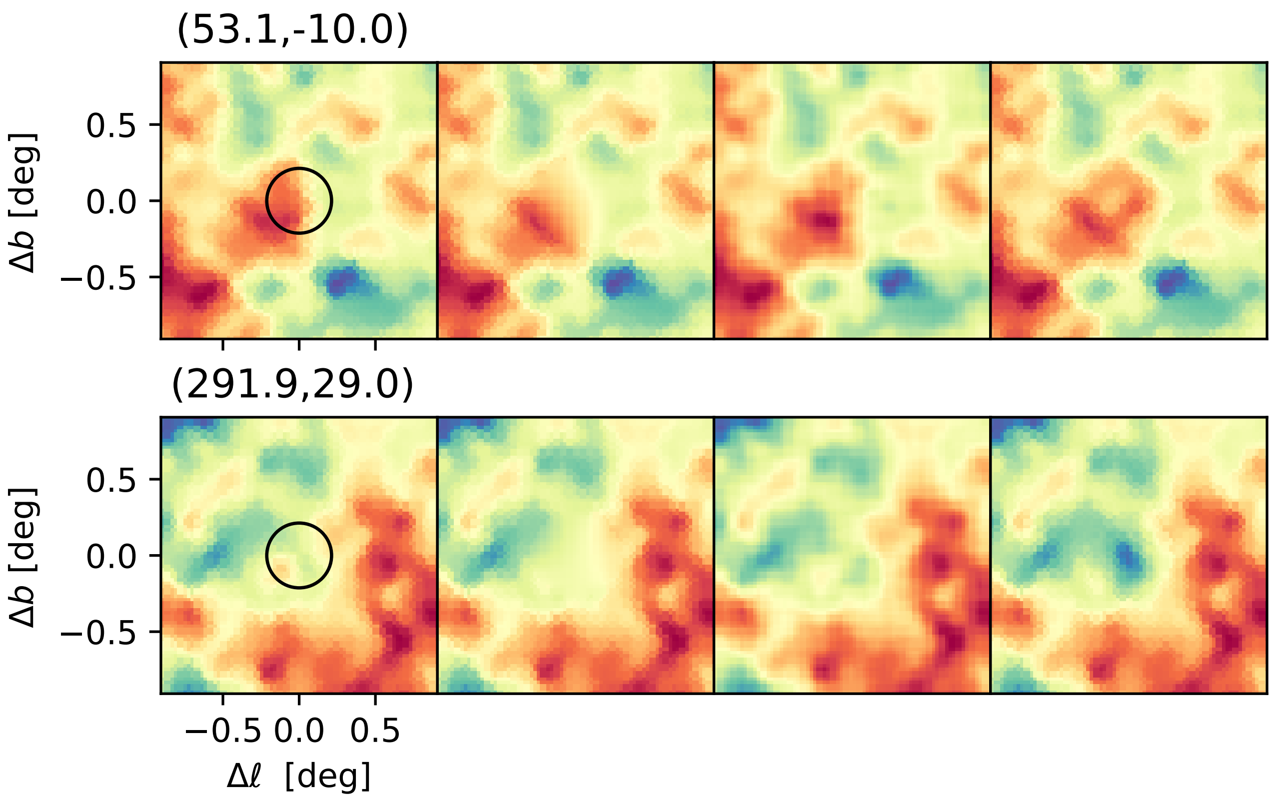
The distribution of pixel intensity provides information about whether the range of pixel values of the ground-truth maps are reproduced in the generated maps. Fig. 3 shows the histogram of the pixel intensities of 500 generated maps compared with the corresponding ground-truth maps from the test set. For the types of foregrounds we consider in this analysis, the amplitude is strongly directional dependant. Therefore, we scale the sample maps with `MinMax` rescaling to $[-1, 1]$ to reduce the patch-to-patch variation. Although the differences between the pixel distributions are nearly negligible, we run a Kolmogorov-Smirnov (KS) two-sample test on each case and assess how likely the distribution of inpainted images is drawn from the ground-truth distribution. We thus estimate the empirical distribution function on the pixel samples from the ground-truth images and from images inpainted with three methods introduced in section 2, and then derive the KS two sample statistics to test the null hypothesis. From the KS p-values summarized in Tab.1, we find that $p > 0.86$ (0.997) for synchrotron (thermal dust) maps so that we can not reject the null hypothesis.

We additionally test that the Fourier modes in the ground-truth are reproduced in the inpainted maps. We evaluate the power spectra of intensity (TT), and polarization (EE and BB) maps[5] in each flat square map

[5] Following the decomposition of Q and U maps proposed in Seljak & Zaldarriaga (1997); Hu & White (1997).

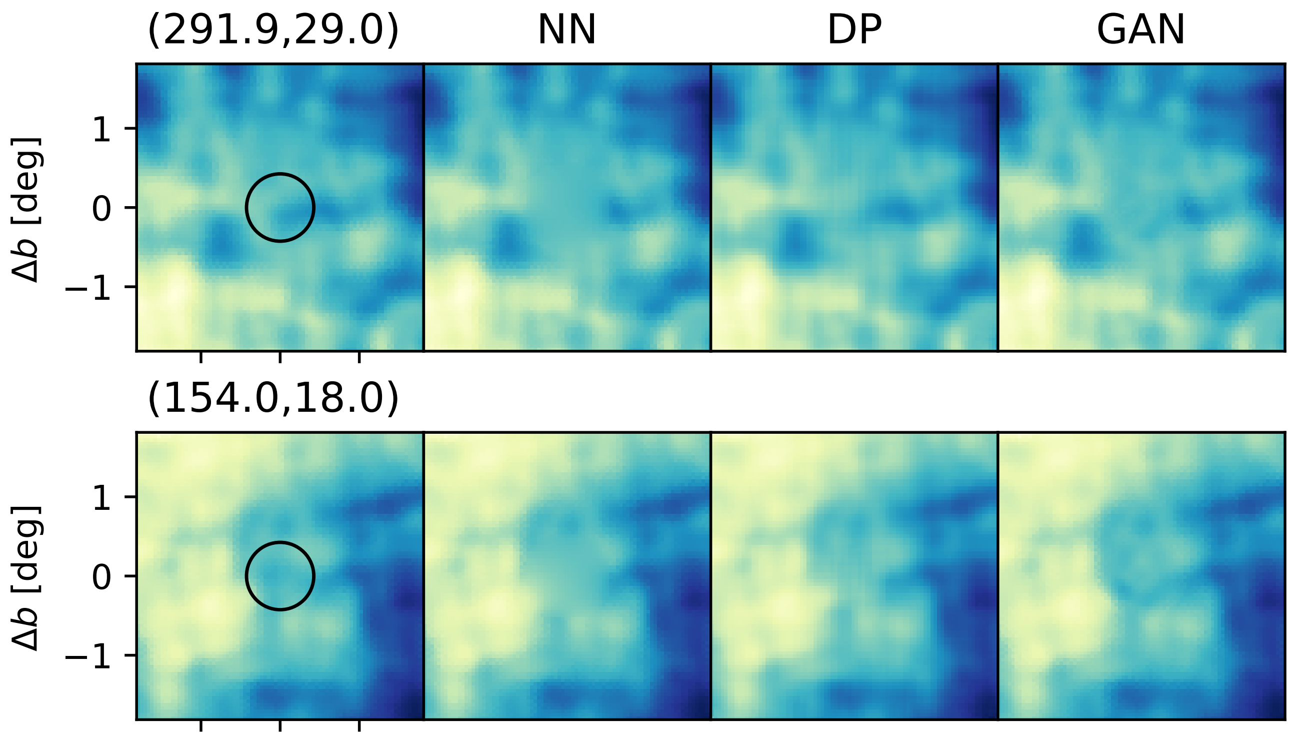


(a)

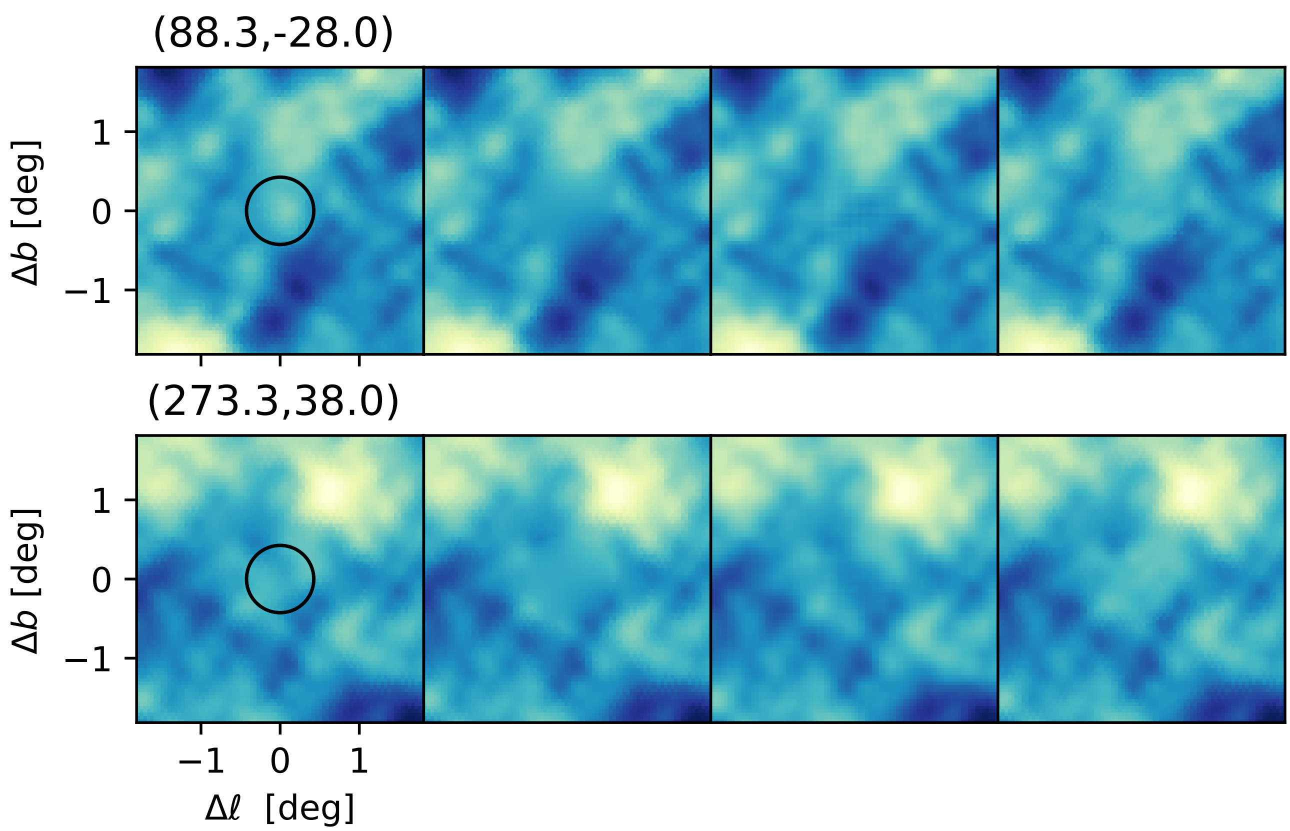


(b)

Figure 1. Thumbnail $1.5 \times 1.5 \deg^2$ crops for (a) Q and (c) U maps predictions of thermal dust. The radius of the reconstructed area (black circle) is 15 arcmin. Columns from left to right show ground truth maps from the test set, predictions obtained with DP, NN and GAN respectively. The colorbar is set to be the same in each row by `MinMax` rescaling all the images with the same *min* and *max* values of the ground-truth ones. Notice that we further zoom in the dust maps to better inspect of the inpainted region (maps are originally $3 \times 3 \deg^2$. Temperature maps are shown in Fig.B3 of the Appendix.



(a)



(b)

Figure 2. Thumbnail $3 \times 3\,\mathrm{deg}^2$ crops for (a) Q and (c) U maps predictions of synchrotron emission. The arrangement and the colorbar setting are the same as those in Fig.1. The radius of the reconstructed area (black circle) is 30 arcmin. Temperature maps are shown in Fig.B3.

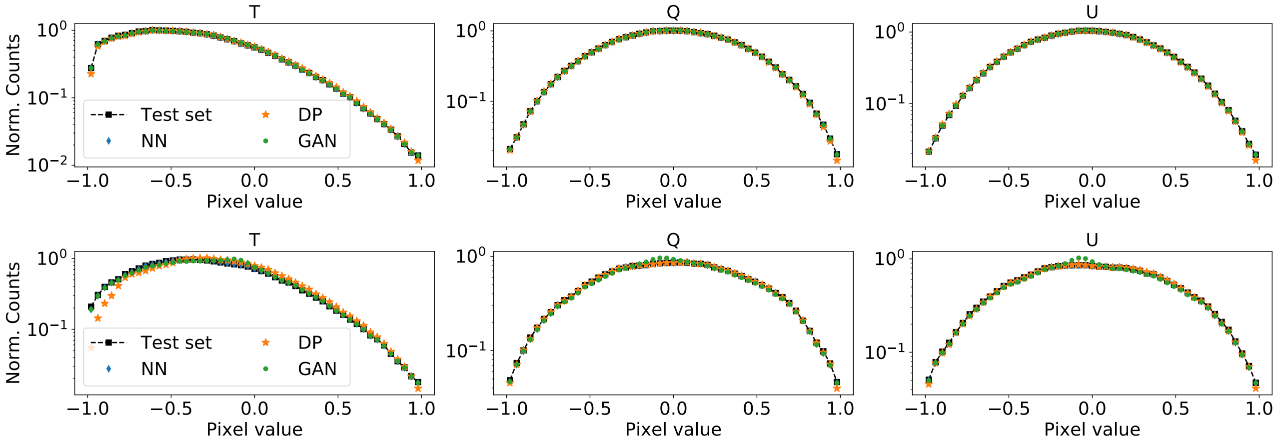


Figure 3. Top: Thermal dust pixel intensity distribution of 500 generated maps with (diamonds) NN, (stars) DP,(circles) GAN compared to 500 maps from test set (black squares). Bottom: Synchrotron pixel intensity distribution. The KS test statistics and the p-values listed in Tab.1 indicates that the distribution of inpainted images are likely to be drawn from the distribution of the test set.

from the test set and the ones inpainted with the three methods.

Each spectra shown in Fig. 4, is estimated with NAMASTER[6] (Alonso et al. 2019), binned into equally spaced multipoles with $\Delta\ell = 450$. The maximum multipole is chosen accordingly to the beam FWHM with whom the signal is convolved, i. e. $\ell_{max} = 4000$ for dust and $\ell_{max} = 2000$ for synchrotron. The median of the binned power spectra is plotted at each multipole estimated from test set including 500 ground truth maps and corresponding inpainted maps.

The shaded-gray area represents 95 per cent of the power spectra estimated from the test set and its vertical width indicates how much the amplitude of the signal can vary at different locations of the sky (as much as 2 orders of magnitude). The median power spectra estimated from inpainting the test set are shown as points, and the area within the dashed lines corresponds to the 95 per cent of the power spectra. Notice that DP power spectra for thermal dust tend to systematically depart from the ones estimated with the test set spectra at $\ell > 2500$ scales. On the contrary, GAN and NN are overall consistent with the spectra from the ground-truth maps.

To assess more quantitatively that the power spectrum at a given ℓ bin is correctly reproduced, we consider 3 different multipole bins. We bootstrap resample 5000 times the distribution in each bin of 500 spectra and perform the KS test on the resampled distribution.

Fig. 5 shows the distribution of EE spectra for thermal dust (top) and synchrotron (bottom) and favors GAN as the method that better resembles the ground-truth distribution, with KS p-value > 0.978 (> 0.808) for dust (synchrotron) spectra. On the other hand, NN spectra present the lowest p-values of > 0.808 for dust and > 0.538 for synchrotron.

	Synchrotron	Thermal Dust
NN	> 0.306 (2 bins)	> 0.537 (4 bins)
	> 0.792 (13 bins)	> 0.801 (23 bins)
DP	> 0.538 (5 bins)	> 0.153 (4 bins)
	> 0.792 (10 bins)	> 0.788 (23 bins)
GAN	> 0.538 (5 bins)	> 0.153 (3 bins)
	> 0.792 (10 bins)	> 0.788 (24 bins)

Table 2. p-value of KS test performed on each multipole bin. We combined together TT, EE, BB spectra, total 27 (15) bins for dust (synchrotron).

In Tab.2, we summarize the KS test p-values estimated on each multipole bin and after having bootstrapped resampled the TT, EE, BB spectra 5000 times in each bin. In total, we account for (3×9) 27 KS p-values for dust and (3×5) 15 KS p-values for synchrotron spectra. Although there are few bins where the p-value is as low as 0.153, the KS statistics for those bins is small enough that the null hypothesis cannot be rejected at a significance level of $\alpha > 5\%$. We therefore conclude that all the three methodologies are able to reproduce angular correlations of the underlying signal coherently.

Since the Galactic emission is highly non-Gaussian, it is essential to evaluate how well the methodologies described here are able to reproduce the non-Gaussian fea-

[6] https://github.com/LSSTDESC/NaMaster

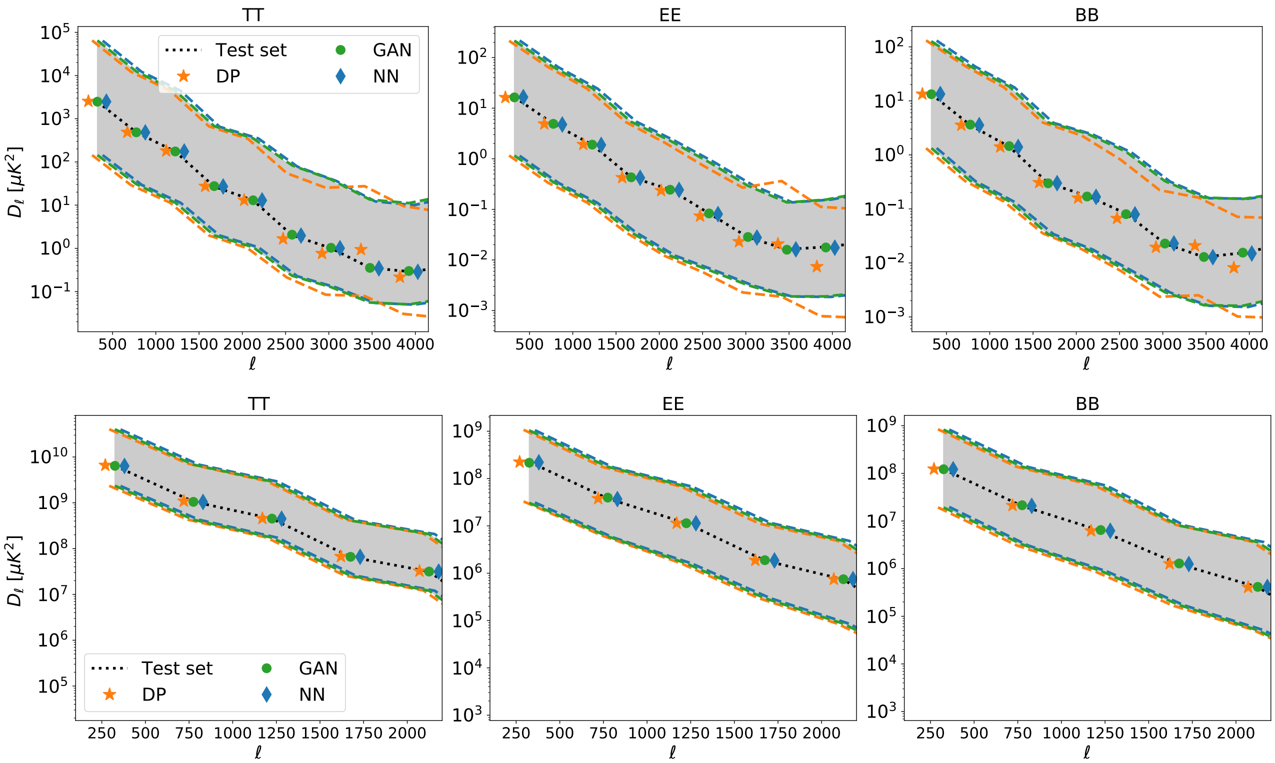


Figure 4. From left to right, TT, EE, BB power spectra estimated from a sample of 500 thermal dust (top) and synchrotron (bottom) maps. Median power spectra are shown as (black dotted) for the test set, (blue diamonds) for NN inpainting, (orange stars) for DP inpainting, (green circles) for GAN inpainting. The gray shaded area corresponds to the 95% of test set spectra to be compared with the 95 % of the set of power spectra inpainted with NN, DP and GAN respectively shown as (dashed blue), (dashed orange) and (dashed green). The spectra are uniformly binned with $\Delta\ell = 450$.

tures. We thus evaluate the three Minkowski functionals V_0, V_1, V_2 for each rescaled map (ranging in $[-1, 1]$). The first three Minkowski functionals are related respectively to the area, the perimeter and the connectivity in an image as a function of a threshold ρ. The dotted black lines in Fig. 6 and 7 show the median for the three Minkowski functionals (calculated using (Mantz et al. 2008)) estimated from 500 samples and evaluated at 10 equally spaced thresholds. Minkowski functionals for inpainted images are shown in Fig. 6 and 7 with the same color scheme as in Fig.4. We notice that both GAN and NN are able to fully reproduce the non-Gaussianity of both the unpolarized and the polarized Galactic emission.

Similarly to what is stated above, the Minkowski functionals estimated on maps inpainted with DP clearly depart from the ground-truth especially at intermediate thresholds, $(-0.5 < \rho < 0.5)$ for the V_1 and V_2 functionals. This can point to further investigations for the DP inpaintings especially to better characterize the non-Gaussian features and the small angular scales $\ell > 2500$ of both dust and synchrotron which are not fully captured by the DP network.

4.2. *Validation on real data*

To further test and validate our methodologies, we run tests on real data, cropping 500 images at random locations from the *Planck* maps at 353 and 857 GHz.

Fig.8 shows the $1.5 \times 1.5\,\mathrm{deg}^2$ *Planck* maps at (a) 353 and (b) 857 GHz. Notice that we deliberately choose two locations to highlight reconstruction performances at two different SNR regimes. Closer to the Galactic midplane, the dust emission is stronger so that SNR is high both at 353 and 857 GHz, e. g. see top panels of Fig.8(a) and (b). On the contrary, at high Galactic latitudes and at 353 GHz, the SNR can be lower and noise contribution is clearly noticeable in the maps, e. g. as in the bottom panel of Fig.8(a). In this case, the inpainting performances with DP and GAN can be affected by noise, and reconstruction area can be visually distinguished in the square patch.

For a more quantitative assessment, we thus estimate the summary statistics shown in Subsect.4.1 and we show the results in Fig.9.

A clear indication that the thermal dust emission data is contaminated by instrumental noise can be inferred by comparing the shapes of Minkowski functionals estimated for the maps at 353 GHz, with the ones from

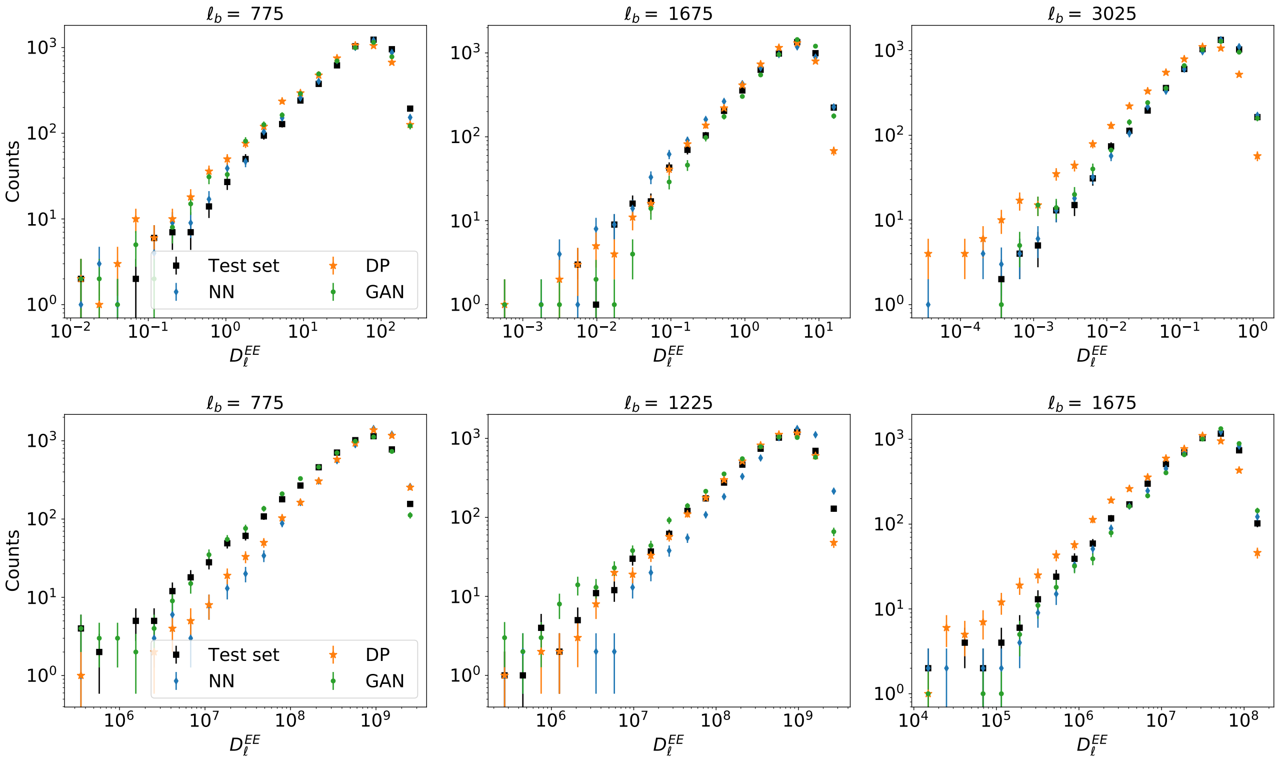


Figure 5. Distributions of EE power spectra of thermal dust(top panel) and synchrotron (bottom panel) at 3 different multipole bins resampled from the power spectra of 500 images inpainted with (diamonds) NN, (stars) DP,(circles) GAN. We compare them with the resampled EE spectra of test maps (black squares). Error bars are estimated as the squared root of number of elements in each bin after bootstrap resampling.

signal-only simulations (fig.1) or from signal-dominated maps (fig.9 (b)). The morphology of the Minkowski functionals of the former largely resembles the functionals estimated from a Gaussian signal. Two possible candidates for the Gaussian component are: i) *Planck* instrumental noise which can be approximated as white within $\sim 9\,\mathrm{deg}^2$ patches and ii) CMB residual emission.

DP inpaintings are the most affected ones in the presence of the noise at 353 GHz, e. g. notice how the pixel distribution (top left panel) of Fig.9(a) noticeably departs from the ground truth one. However, the KS test p-value performed on the DP and ground truth samples is $p > 0.536$, i. e. not significantly low enough to reject the null hypothesis that the two samples are different. On the other hand, inpainting with GAN and NN does not show any dependence with SNR, as the performances with these methodologies are essentially similar compared with the ones observed with the signal-only simulated maps (e. g. Fig.3, 4 and 6).

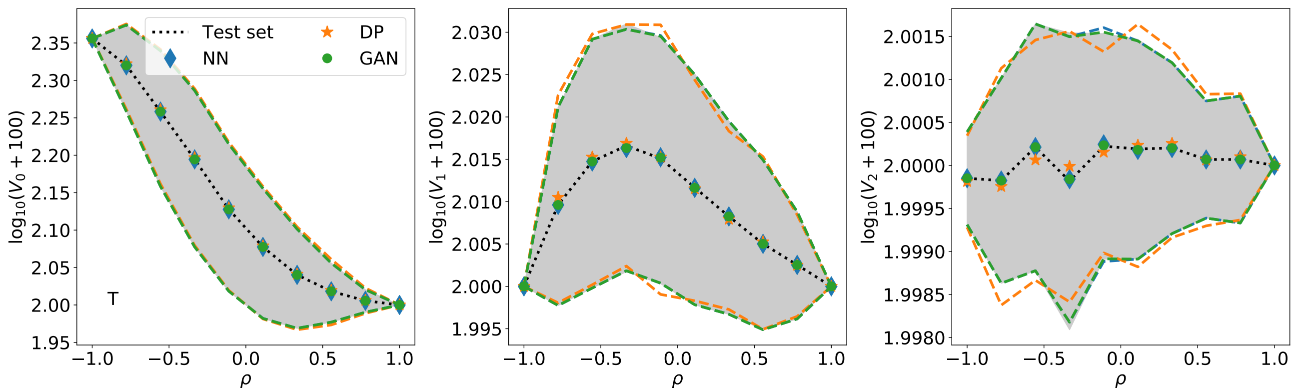
Finally, we would like to point out that for the case of inpainting with GAN, we used the weights which have been derived from the training set composed of signal-only dust TQU simulated maps. By looking at Fig.9 we notice that GAN is able to statistically reproduce at 353 GHz the features composed by signal and noise, remarkably indicating that the network has correctly learned the features related to the intrinsic signal in presence of noise , and injects signal plus noise features statistically coherent with the ones outside the masked area. However, when the noise is higly dominating in the patch, we can clearly distinguish smooth artifacts in some cases inpainted with GAN (see Fig.8(a, bottom)). This is somewhat expected being GAN trained on signal-only simulated images. Further investigations are needed (training GAN with noisy data) and we will address them in a future work.

4.3. *Inpainting maps with Point-Sources*

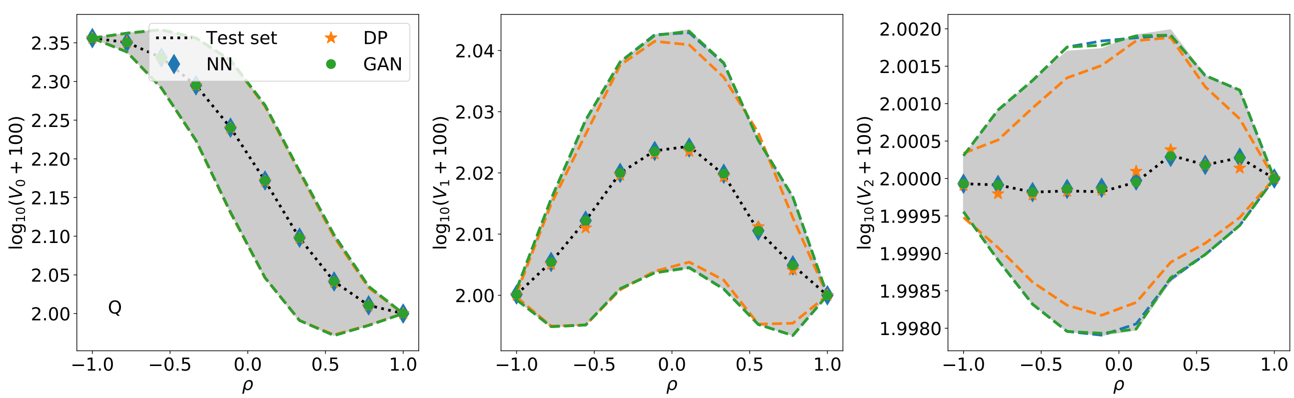
In this section, we aim at showing real world applications of the three reconstruction methodologies by inpainting areas in real maps with detected point-sources.

We consider the TQU SPASS synchrotron map at 2.3 GHz[7] and we run a *Matched Filter* (Marriage et al. 2011), to detect the brightest polarized sources in the map. We considered 7σ as a threshold for a point-

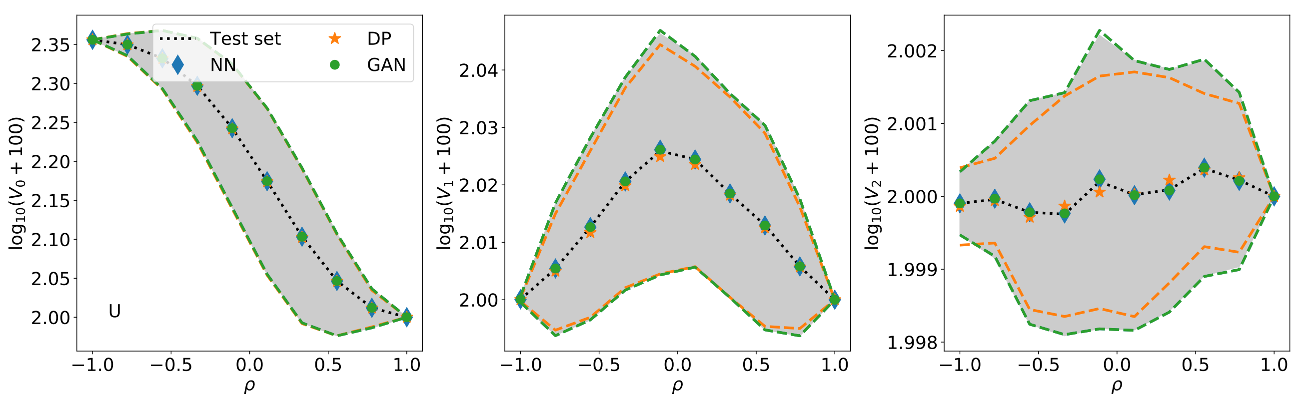
[7] Maps are available online at https://sites.google.com/inaf.it/spass.



(a)

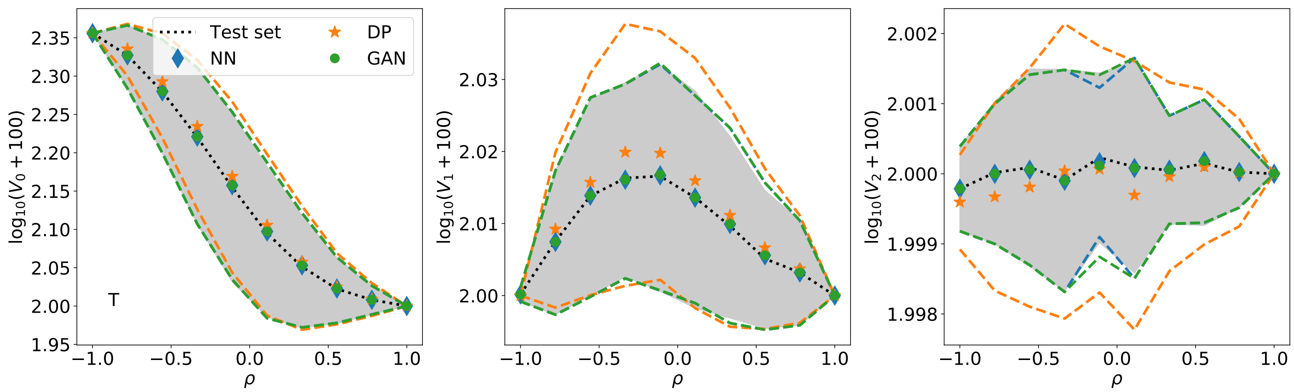


(b)

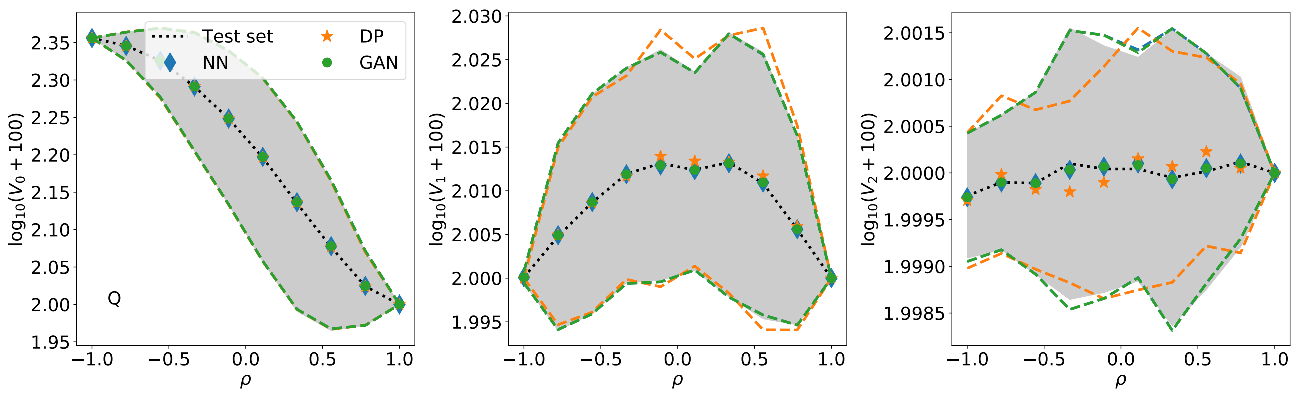


(c)

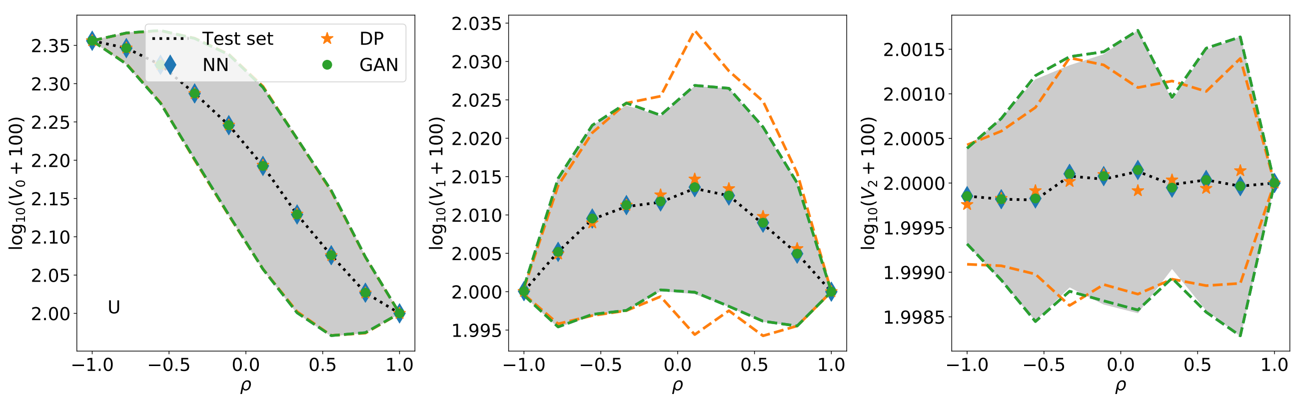
Figure 6. From left to right, V_0, V_1, V_2 Minkowski functionals estimated from the test set of 500 thermal dust T,Q and U maps respectively in (a), (b), (c). We use the same coloring scheme as in Fig. 4: (black dotted) median of the functionals estimated from the test set, 95 % of the functionals is shown as a gray shaded area . Points and dashed lines refer to medians and 95 % interval of the functionals estimated from the sets of inpainted maps with (blue diamonds and blue dashed) NN, (orange stars and orange dashed) DP and (green circles and green dashed) GAN.



(a)

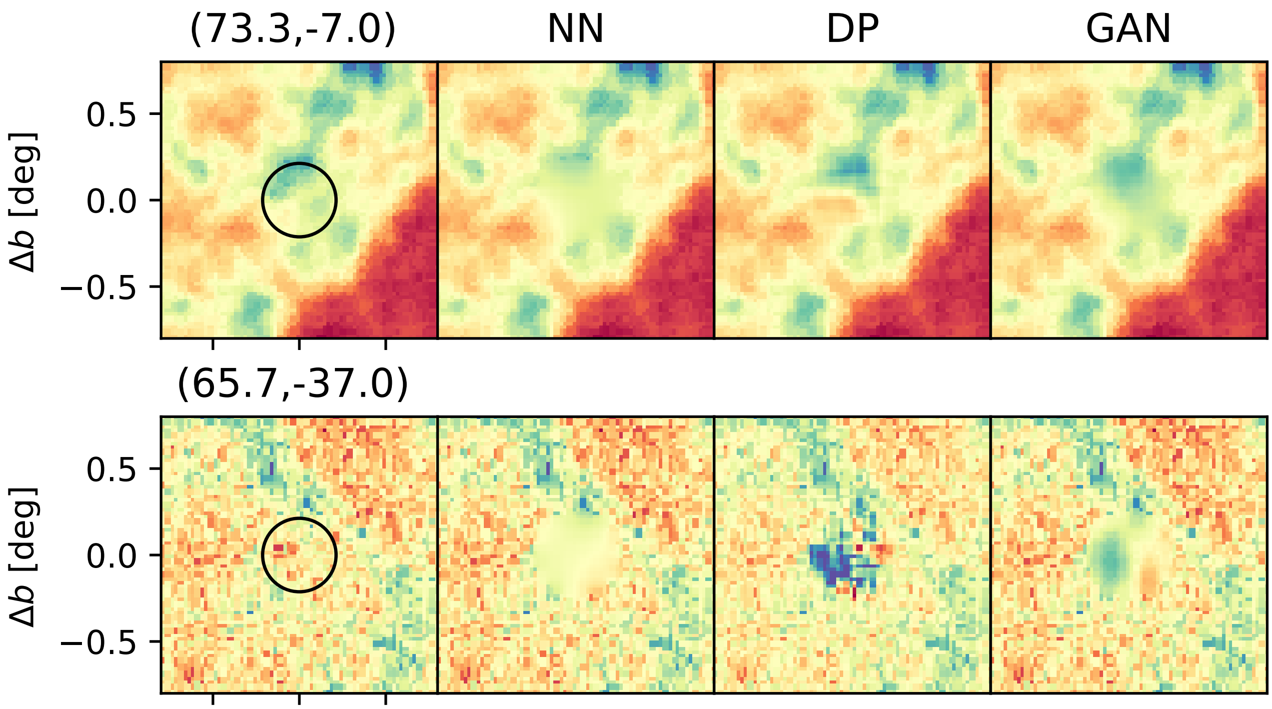


(b)

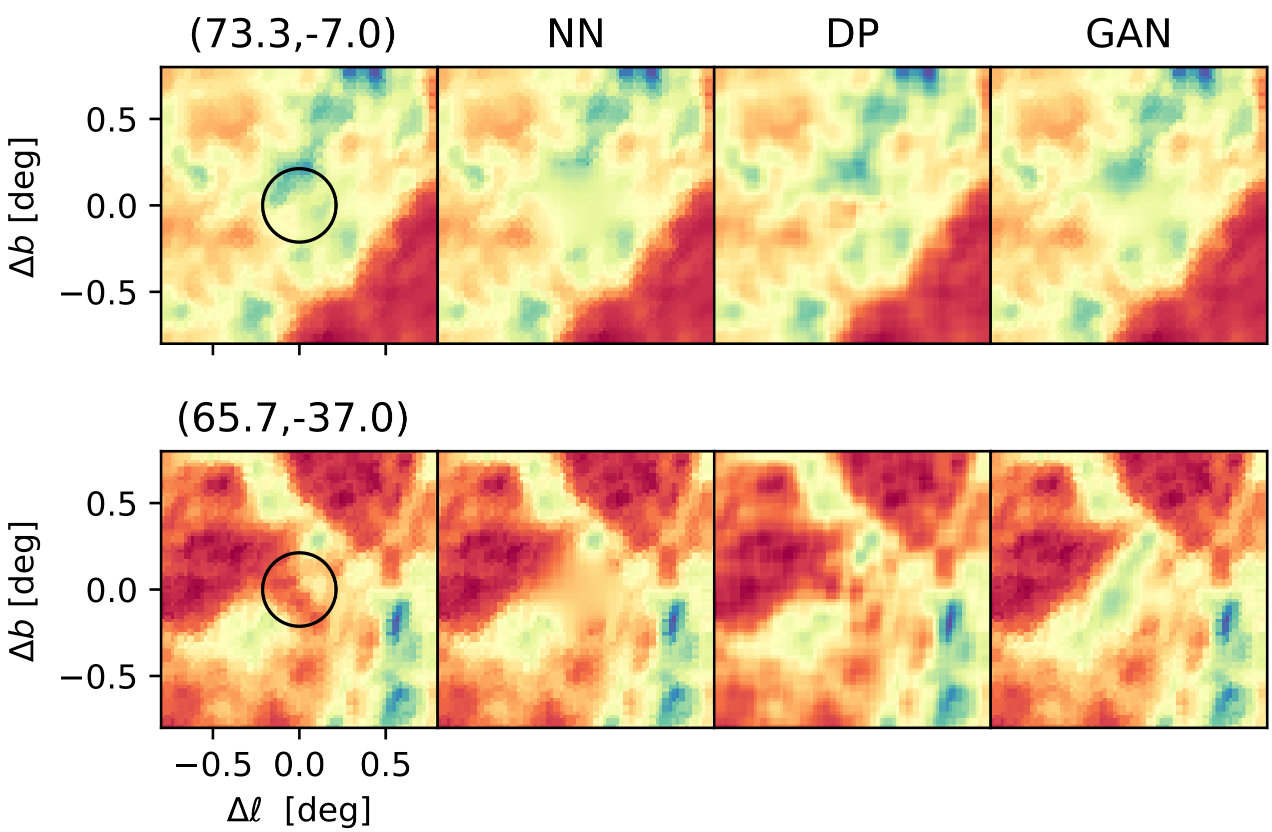


(c)

Figure 7. From left to right, V_0, V_1, V_2 Minkowski functionals estimated from the test set of 500 synchrotron TQU maps respectively in (a), (b), (c). We use the same coloring scheme as in Fig. 4: (black dotted) median of the functionals estimated from the test set, 95 % of the functionals is shown as a gray shaded area . Points and dashed lines refer to medians and 95 % interval of the functionals estimated from the sets of inpainted maps with (blue diamonds and blue dashed) NN, (orange stars and orange dashed) DP and (green circles and green dashed) GAN.

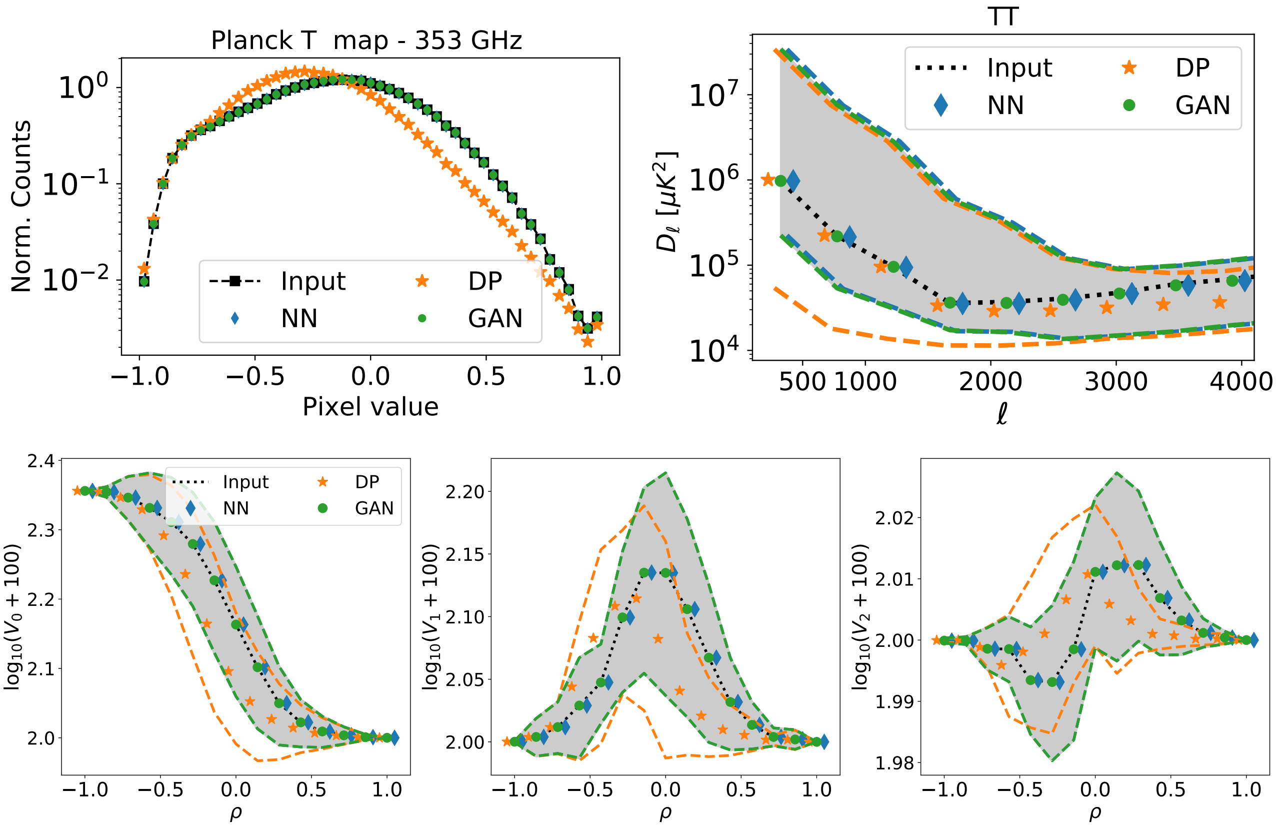


(a)

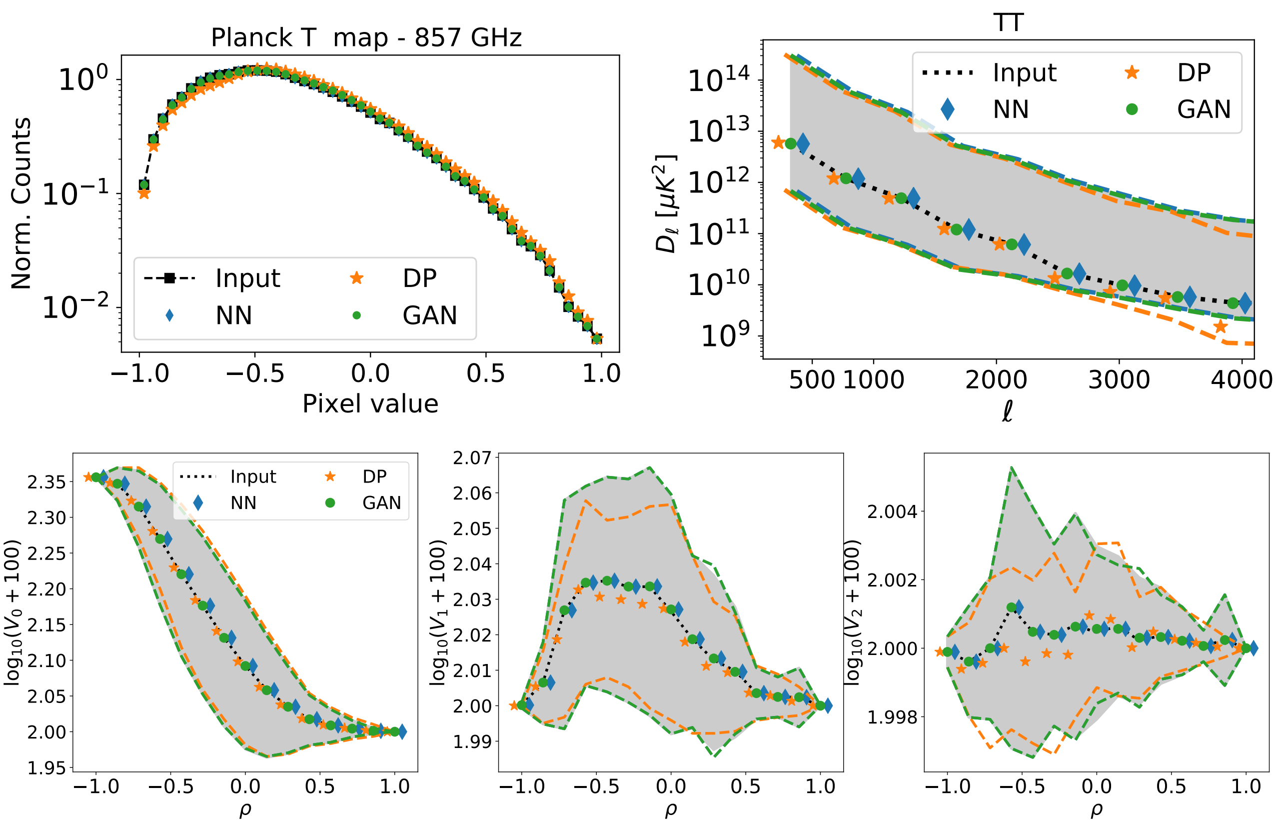


(b)

Figure 8. Thumbnail images of *Planck* temperature maps at (a) 353 and (b) 857 GHz. The radius of the reconstructed area (black circle) is 15 $'$. Two locations are chosen to highlight different inpainting performances with high and low SNRs of *Planck* 353 map, respectively at high Galactic latitude, i. e. $(l, b) = (65.7°, -7°)$ and at low Galactic latitude,i. e. $(l, b) = (73.3°, -37°)$.



(a)



(b)

Figure 9. Summary statistics estimated on *Planck* dust T maps at (a) 353 and (b) 857 GHz.

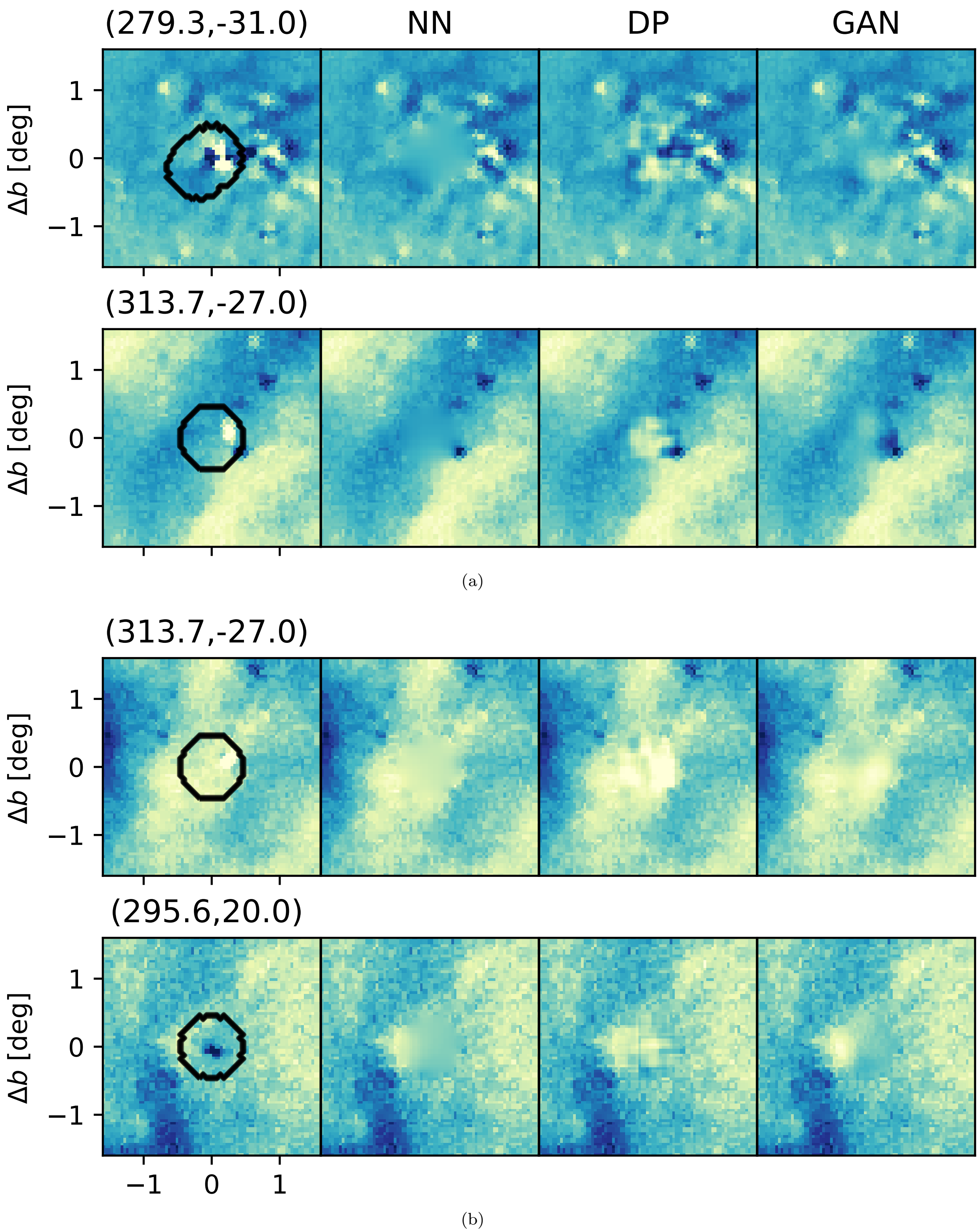


Figure 10. From left to right, Input SPASS, DP, NN and GAN polarization maps inpainted in 30 arcmin region surrounding the detected point-source (black circle). The range of the input map is chosen to be the same as the one of the inpainted ones. Temperature maps are shown in Fig.B2.

source detection. In particular, we focus on unresolved point-sources (i. e. sources whose projected solid angle is smaller than the SPASS beam solid angle) detected at intermediate Galactic latitudes $|b| > 20$ deg. As a result, 45 polarized sources are detected in the SPASS map, which is very close to 60 the number forecasted by PS4C with the adopted SPASS specifications.

Fig.10 shows a selection of images extracted from the SPASS TQU maps and centered at the coordinates of the detected sources. The shape and size of the region to be inpainted are chosen proportionally to the flux of each source (see the black circle in fig.10). However, we expect the inpainting not to be affected by different shapes and/or sizes of the masked area as it has been already demonstrated in Yu et al. (2018) and Ulyanov et al. (2017). Moreover, we set the color scale of the input SPASS map to be the same as the one in the inpainted maps so that we can better highlight the consistency with reconstructed maps. As expected, images inpainted with GANs are visually injecting more coherent features and less artifacts with respect to DP and NN.

Given the presence of the point-source at the center of the patch biasing the evaluation of fidelity with the pixel distribution and the Minkowski functionals, we estimate the power spectra from all the sets of maps in order to assess more quantitatively the quality of the generated maps. We masked the source with a circular mask with 30 $'$ radius and estimate the spectra in the area outside the mask. On the other hand, we did not apply the point-source mask for the power spectra estimated from the inpainted maps.

We estimate the power spectra as described in the previous sections and in Fig.11 we show the TT, EE and BB power spectra, which shows all methodologies essentially are able to reproduce consistently the power spectrum at all angular scales of the input SPASS masked maps. Moreover, on smaller angular scales, the power spectra from maps inpainted with GAN show a lower amplitude tail, possibly implying that Poissonian bias from undetected point-sources is further reduced (a factor of ~ 4 for EE and BB spectra).

5. CODE RELEASE

The three inpainting methods have been collected into a `python` package, the Python Inpainter for Cosmological and AStrophysical SOurces (PICASSO[8]). It has been made publicly available together with a documentation web page[9].

[8] github.com/giuspugl/picasso

[9] http://giuspugl.github.io/picasso/

We trained GAN separately on dust and synchrotron map sets[10]. The training process for each set took ~ 12 hours on a GPU node of SHERLOCK Cluster of Stanford Supercomputing Center with 4 interconnected NVIDIA Tesla P40 GPUs [11].

Finally, we measure the inpainting speed for each of the three methods. We performed this benchmark within a GPU node at NERSC equipped with 4 interconnected NVIDIA Tesla -V100 GPUs[12]. GAN is the fastest method since fast-forwarding the trained weights is very quick and it takes $\sim 4 \div 6$ s. Although NN does not involve any DCNN in the map reconstruction, it iterates over the pixels in the missing region and it takes $\sim$ 15 s per image. A single inpainting with DP takes ~ 30 s, since this is the time spent to minimize the loss function in eq.1 over 3000 - 5000 epochs with gradient descent.

6. SUMMARY AND CONCLUSIONS

In this work, we demonstrate that three inpainting methodologies can reproduce an underlying non-Gaussian signal without modifying the overall summary statistics of the signal itself.

The first method (NN) has been already used in the literature and it is based on diffusing the pixels in the masked area with the average of nearest-neighbour pixels. We further adopted two novel techniques relying on DCNN, namely DP and GAN. They have been firstly used to inpaint natural images by the deep learning community.

We validated the three techniques on simulated data, tested them on data-set with a wide range of SNRs. We show a real-life application, by inpainting a map in regions where a bright point-sources is detected. To evaluate the quality of inpainted results, we adopted three summary statistics based on the pixel distribution, the angular power spectra and the first three Minkowski functionals. We find that all techniques are able to reproduce the overall summary statistics when applied to signal-only data.

Inside the masked regions inpainted with NN, the results are smooth, and lacks of finer details, which makes them perceptually less similar to the original map. Generally, a map inpainted with NN sharply transitions from the area outside the mask with sub-structures and noisy pixels to a very smooth one encoding only long and smooth modes inside the masked area. However, we did not notice any clear effect or bias due to NN in-

[10] Training weights can be downloaded from https://bit.ly/2TI6x4o

[11] For more details, see https://www.sherlock.stanford.edu.

[12] https://docs-dev.nersc.gov/cgpu

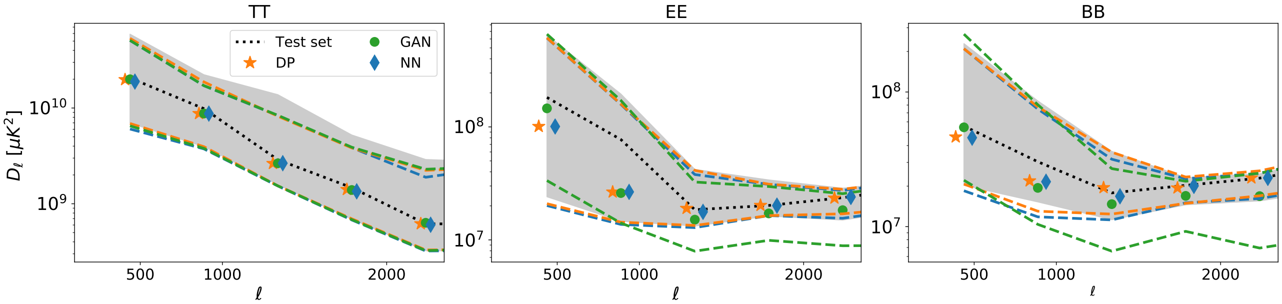


Figure 11. TT, EE, BB power spectra estimated on the SPASS map. In this case, the input map encodes a point-source located at the center and is masked out before computing the power spectra. Vice versa, spectra estimated on inpainted maps do not have the mask applied.

painting on the statistical tests we adopted in this work. On the other hand, DP reconstructions on images extracted from signal and noise maps present a different Minkowski functionals with respect to the ground-truth ones. As pointed out at the end of Sect.4.1, this failure case of DP needs to be further investigated by means of a better tuning of hyper-parameters.

Conversely, GAN has been demonstrated to be very promising it generates images visually indistinguishable from the ground-truth on signal dominated maps. However, we have identified cases where it fails to produce high fidelity images in noise dominated maps. We planned to further investigated this in a future work by training GAN with more realistic dataset including several levels of SNR.

To our knowledge, this is the first time that GAN have been used to successfully generate high resolution intensity and polarization maps of Galactic foreground polarization maps. This promising approach opens up many possibilities of generating foreground maps using adversarial networks, which overcome the limitations of existing templates.

In conclusion, we focused this work on Galactic foreground emission motivated by the challenges in inpainting non-Gaussian signal and in dealing with Galactic (and Extra-galactic) foregrounds in CMB B-mode polarization studies. For the future work, we plan to apply similar techniques to different non-Gaussian signals spanning from galaxy weak lensing to HI data, which are highly affected by foreground emission as well.

Acknowledgements. The authors are very thankful to Yuuki Omori and Ben Thorne for carefully reading the manuscript. They also thank Alexandre Refregier, Nicoletta Krachmalnicoff, Ioannis Liodakis and Warren Morningstar for the fruitful discussions and useful comments shared during the development of this work. Some of the results in this paper have been derived using the HEALPIX (Górski et al. 2005) package.

REFERENCES

Adachi, S., Aguilar Faúndez, M. A. O., Arnold, K., et al. 2019, arXiv e-prints, arXiv:1910.02608. https://arxiv.org/abs/1910.02608

Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., et al. 2014, A&A, 571, A24, doi: 10.1051/0004-6361/201321554

Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A17, doi: 10.1051/0004-6361/201525836

Alonso, D., Sanchez, J., Slosar, A., & LSST Dark Energy Science Collaboration. 2019, MNRAS, 484, 4127, doi: 10.1093/mnras/stz093

Arjovsky, M., Chintala, S., & Bottou, L. 2017, arXiv e-prints, arXiv:1701.07875. https://arxiv.org/abs/1701.07875

Aylor, K., Haq, M., Knox, L., Hezaveh, Y., & Perreault-Levasseur, L. 2019. https://arxiv.org/abs/1909.06467

BICEP2 Collaboration, Keck Array Collaboration, Ade, P. A. R., et al. 2018, PhRvL, 121, 221301, doi: 10.1103/PhysRevLett.121.221301

Bucher, M., & Louis, T. 2012, Monthly Notices of the Royal Astronomical Society, 424, 16941713, doi: 10.1111/j.1365-2966.2012.21138.x

Bucher, M., Racine, B., & van Tent, B. 2016, JCAP, 2016, 055, doi: 10.1088/1475-7516/2016/05/055

Caldeira, J., Wu, W., Nord, B., et al. 2019, Astronomy and Computing, 28, 100307, doi: https://doi.org/10.1016/j.ascom.2019.100307

Carlstrom, J., Abazajian, K., Addison, G., et al. 2019, in BAAS, Vol. 51, 209. https://arxiv.org/abs/1908.01062

Carretti, E., Haverkorn, M., Staveley-Smith, L., et al. 2019, MNRAS, 489, 2330, doi: 10.1093/mnras/stz806

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. 2015, Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). https://arxiv.org/abs/1511.07289

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693, doi: 10.1086/300337

Datta, R., Aiola, S., Choi, S. K., et al. 2019, MNRAS, 486, 5239, doi: 10.1093/mnras/sty2934

Farsian, F., Krachmalnicoff, N., & Baccigalupi, C. 2020, Foreground model recognition through Neural Networks for CMB B-mode observations. https://arxiv.org/abs/2003.02278

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014. https://arxiv.org/abs/1406.2661

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759, doi: 10.1086/427976

Gupta, N., Reichardt, C. L., Ade, P. A. R., et al. 2019, Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/stz2905

Hoffman, Y., & Ribak, E. 1991, ApJL, 380, L5, doi: 10.1086/186160

Hu, W., & White, M. 1997, Physical Review D, 56, 596615, doi: 10.1103/physrevd.56.596

Iizuka, S., Simo-Serra, E., & Ishikawa, H. 2017, ACM Transactions on Graphics (Proc. of SIGGRAPH 2017), 36, 107:1

Krachmalnicoff, N., & Tomasi, M. 2019, A&A, 628, 129, doi: 10.1051/0004-6361/201935211

Krachmalnicoff, N., Carretti, E., Baccigalupi, C., et al. 2018, A&A, 618, A166, doi: 10.1051/0004-6361/201832768

Lamee, M., Rudnick, L., Farnes, J. S., et al. 2016, ApJ, 829, 5, doi: 10.3847/0004-637X/829/1/5

Louis, T., Grace, E., Hasselfield, M., et al. 2017, JCAP, 2017, 031, doi: 10.1088/1475-7516/2017/06/031

Mantz, H., Jacobs, K., & Mecke, K. 2008, Journal of Statistical Mechanics: Theory and Experiment, 2008, P12015, doi: 10.1088/1742-5468/2008/12/P12015

Marriage, T., Juin, J., & Lin, Y. 2011, The Astrophysical, 731, 1, doi: 10.1088/0004-637X/731/2/100

Mustafa, M., Bard, D., Bhimji, W., et al. 2017, CosmoGAN: creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks, Tech. rep.

Perraudin, N., Srivastava, A., Lucchi, A., et al. 2019, Cosmological N-body simulations: a challenge for scalable generative models. https://arxiv.org/abs/1908.05519

Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2018, arXiv e-prints, arXiv:1807.06208. https://arxiv.org/abs/1807.06208

POLARBEAR Collaboration, Ade, P. A. R., Aguilar, M., et al. 2017, ApJ, 848, 121, doi: 10.3847/1538-4357/aa8e9f

Puglisi, G., Galluzzi, V., Bonavera, L., et al. 2018, ApJ, 858, 85, doi: 10.3847/1538-4357/aab3c7

Rodríguez, A. C., Kacprzak, T., Lucchi, A., et al. 2018, Computational Astrophysics and Cosmology, 5, 4, doi: 10.1186/s40668-018-0026-4

Sayre, J. T., Reichardt, C. L., Henning, J. W., et al. 2019, arXiv e-prints, arXiv:1910.05748. https://arxiv.org/abs/1910.05748

Seljak, U., & Zaldarriaga, M. 1997, Physical Review Letters, 78, 20542057, doi: 10.1103/physrevlett.78.2054

Starck, J.-L., Fadili, M. J., & Rassat, A. 2013, Astronomy and Astrophysics, 550, A15, doi: 10.1051/0004-6361/201220332

The Simons Observatory Collaboration, Ade, P., Aguirre, J., et al. 2019, JCAP, 2019, 056, doi: 10.1088/1475-7516/2019/02/056

The Simons Observatory Collaboration, Ade, P., Aguirre, J., Ahmed, Z., et al. 2019, Journal of Cosmology and Astroparticle Physics, 2019, 056056, doi: 10.1088/1475-7516/2019/02/056

Thorne, B., Dunkley, J., Alonso, D., & Næss, S. 2017, MNRAS, 469, 2821, doi: 10.1093/mnras/stx949

Ulyanov, D., Vedaldi, A., & Lempitsky, V. 2017, arXiv:1711.10925

Yu, F., & Koltun, V. 2015, Multi-Scale Context Aggregation by Dilated Convolutions. https://arxiv.org/abs/1511.07122

Yu, J., Lin, Z., Yang, J., et al. 2018, arXiv preprint arXiv:1801.07892

Zonca, A., Singer, L., Lenz, D., et al. 2019, The Journal of Open Source Software, 4, 1298, doi: 10.21105/joss.01298

APPENDIX

A. NEURAL NETWORK ARCHITECTURES

Networks	S	D	Layers	Networks	S	D	Layers
Inpainting Network	1		5 x 5 x 32 x 1	Contextual Attention	1		5 x 5 x 32 x 1
	2		3 x 3 x 64 x 1		2		3 x 3 x 64 x 1
	1		3 x 3 x 64 x 1		2		3 x 3 x 128 x 1
	2		3 x 3 x 128 x 1		1		3 x 3 x 128 x 2
	1		3 x 3 x 128 x 2				Contextual Attention
	1	2	3 x 3 x 128 x 1		1		3 x 3 x 128 x 2
	1	4	3 x 3 x 128 x 1				Concatenate
	1	8	3 x 3 x 128 x 1				
	1	16	3 x 3 x 128 x 1				
	1		3 x 3 x 128 x 2				
	1		3 x 3 x 64 x 2				
	1		3 x 3 x 32 x 1				
	1		3 x 3 x 16 x 1				
	1		3 x 3 x 3 x 1				
Global WGAN-GP	2		5 x 5 x 64 x 1	Local WGAN-GP	2		5 x 5 x 64 x 1
	2		5 x 5 x 128 x 1		2		5 x 5 x 128 x 1
	2		5 x 5 x 256 x 2		2		5 x 5 x 256 x 1
			Fully Connected				Fully Connected

Table A1. The architecture of GAN with contextual-attention: S (stride size), D (dilution), and Layers (kernel size × kernel size × channel × number of layers). For further details see Yu et al. (2018).

Network	Hyperparameters		
n_d	`[16, 32, 64, 128, 128, 128]`	n_u	`[128, 128, 128, 64, 32, 16 ]`
k_d	`[3, 3, 3, 3, 3, 3]`	k_u	`[5, 5, 5, 5, 5, 5]`

Table A2. Architecture for DP hole inpainting as described in Ulyanov et al. (2017). n_u, n_d correspond to respectively the number of upsampling and downsampling filters, k_d, k_u correspond to the kernel sizes.

Encoder Block	S	Parameters	Decoder Block	S	Parameters
`Conv2D`	1	$k_d[i] \times k_d[i] \times n_d[i] \times 1$	`Conv2D`	1	$k_u[i] \times k_u[i] \times n_u[i] \times 1$
`Conv2D`	2	$k_d[i] \times k_d[i] \times n_d[i] \times 1$	`LeakyRELU`		$\alpha_{LR} = 0.1$
`LeakyRELU`		$\alpha_{LR} = 0.1$	`Conv2D`	1	$k_u[i] \times k_u[i] \times n_u[i] \times 1$
`Conv2D`	1	$k_d[i] \times k_d[i] \times n_d[i] \times 1$	`LeakyRELU`		$\alpha_{LR} = 0.1$
`LeakyRELU`		$\alpha_{LR} = 0.1$	`UpSampling2D`		`size` $= (2, 2)$

Table A3. Sequence of layers encoded in the i-th downsampling (encoder) and upsampling (decoder) block. S refers to the stride size.

B. SUPPLEMENTARY PLOTS

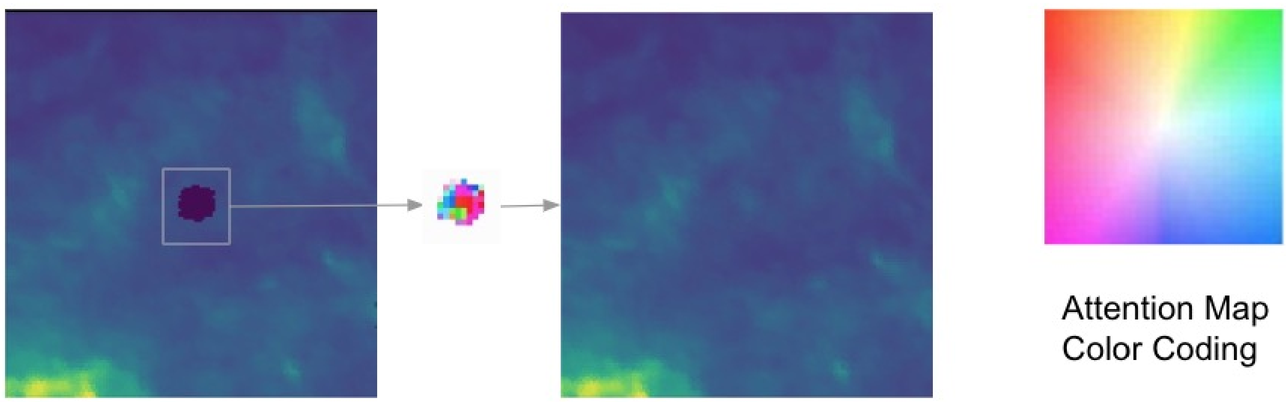


Figure B1. Visualization of an inpainting contextual-attention map. Colors indicate the portion of the whole image the neural network has focused on in order to inpaint a given pixel location in the masked area. In this particular example, most of the pixels have been inpainted by GAN by looking at the left upper region of the image.

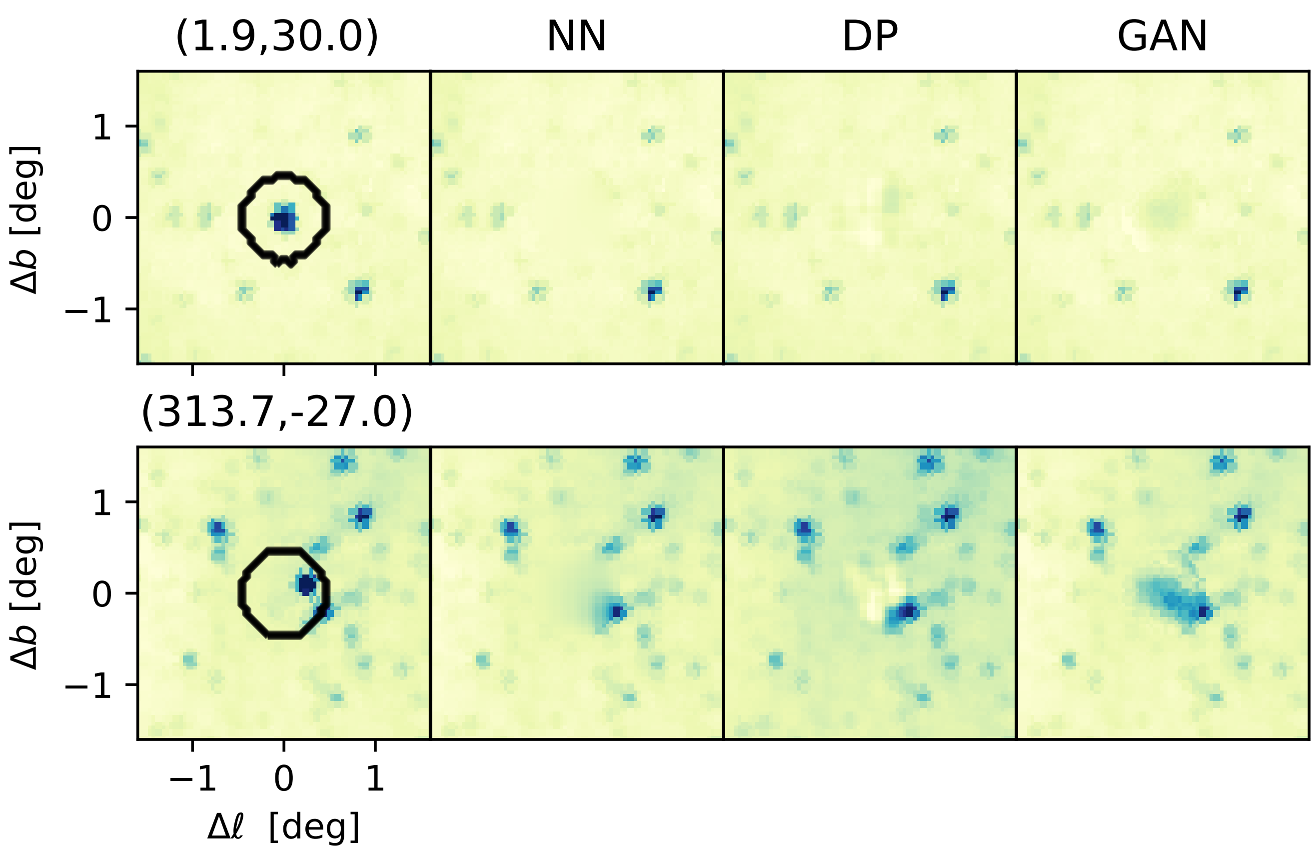
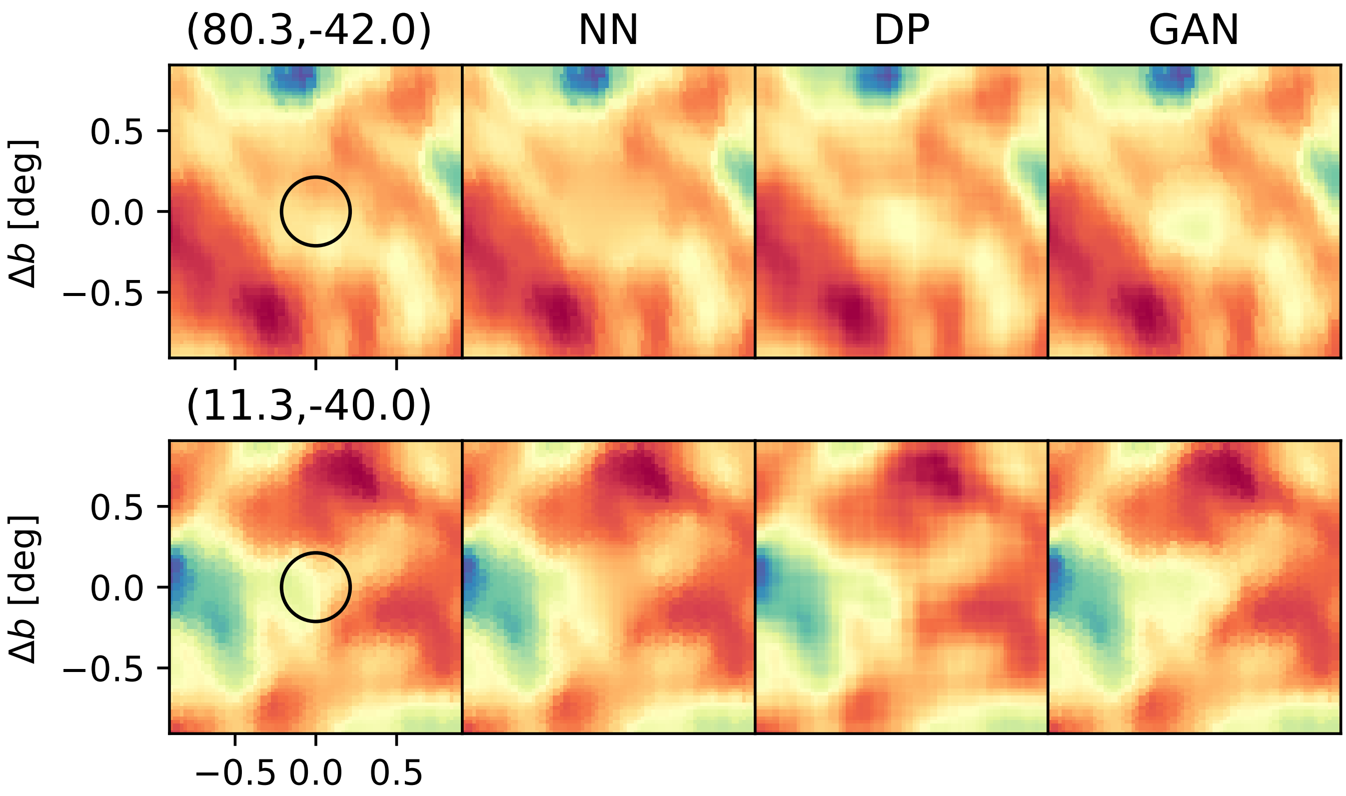
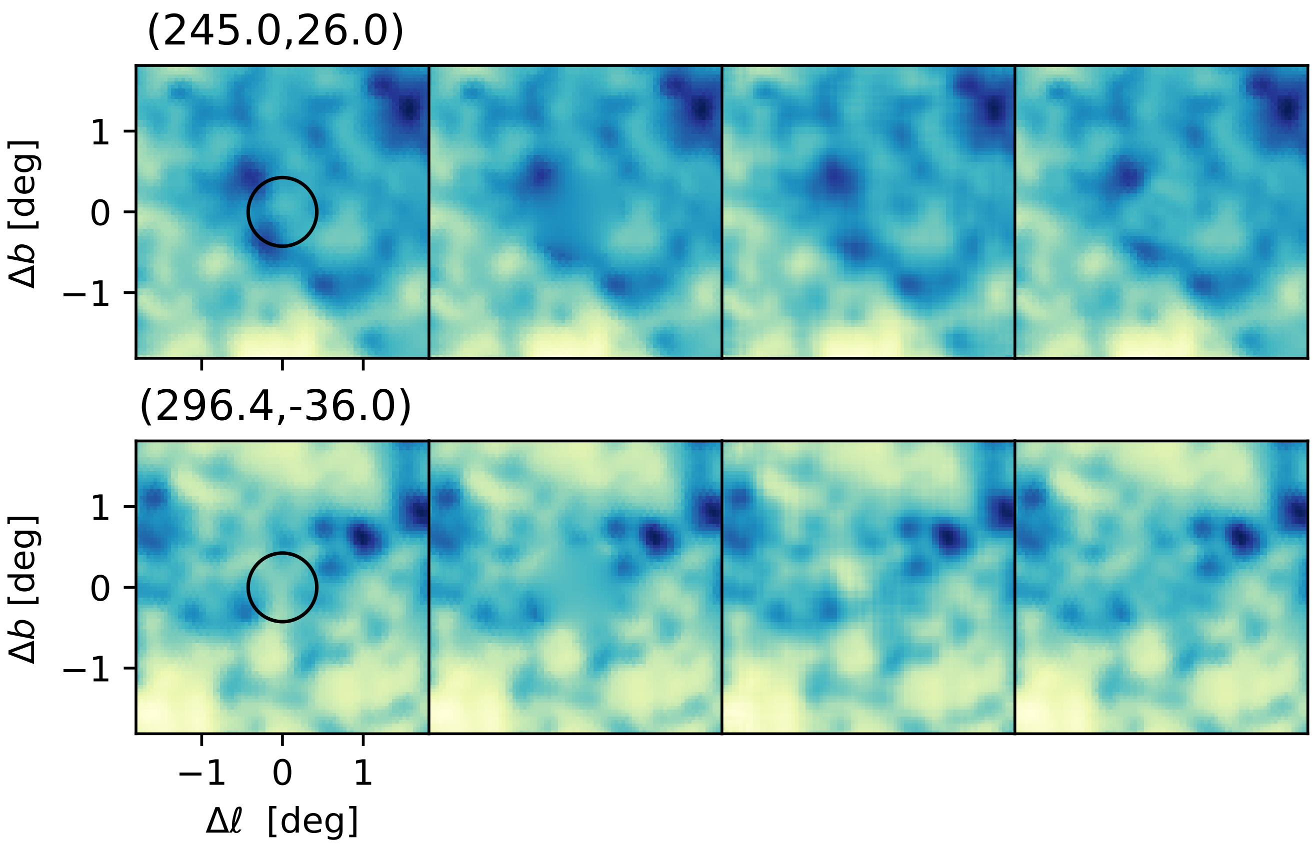


Figure B2. SPASS temperature maps inpainted in 30 arcmin region surrounding the detected point-source (black circle). The range of the input map is chosen to be the same as the one of the inpainted ones.



(a)



(b)

Figure B3. Thumbnail images for (a)dust, (b) synchrotron T maps predictions from the testing set simulated with PySM. The radius of the reconstructed area is shown as grey circle.