CosmoVAE: Variational Autoencoder for CMB Image Inpainting*

1st Kai Yi
School of Mathematics and Statistics
The University of New South Wales
Sydney, Australia
ykyrie822@gmail.com

2nd Yi Guo
School of Mathematics and Statistics
The University of New South Wales
Sydney, Australia
dennis.guo.china@gmail.com

3rd Yanan Fan
School of Mathematics and Statistics
The University of New South Wales
Sydney, Australia
y.fan@unsw.edu.au

4th Jan Hamann
School of Physics
The University of New South Wales
Sydney, Australia
jan.hamann@unsw.edu.au

5th Yu Guang Wang
School of Mathematics and Statistics
The University of New South Wales
Sydney, Australia
yuguang.wang@unsw.edu.au

***Abstract*—Cosmic microwave background radiation (CMB) is critical to the understanding of the early universe and precise estimation of cosmological constants. Due to the contamination of thermal dust noise in the galaxy, the CMB map that is an image on the two-dimensional sphere has missing observations, mainly concentrated on the equatorial region. The noise of the CMB map has a significant impact on the estimation precision for cosmological parameters. Inpainting the CMB map can effectively reduce the uncertainty of parametric estimation. In this paper, we propose a deep learning-based variational autoencoder — CosmoVAE, to restoring the missing observations of the CMB map. The input and output of CosmoVAE are square images. To generate training, validation, and test data sets, we segment the full-sky CMB map into many small images by Cartesian projection. CosmoVAE assigns physical quantities to the parameters of the VAE network by using Fourier coefficients, which are sampled by the angular power spectrum of the Gaussian random field as latent variables. CosmoVAE adopts a new loss function to improve the learning performance of the model, which consists of ℓ_1 reconstruction loss, Kullback-Leibler divergence between the posterior distribution of encoder network and the prior distribution of latent variables, perceptual loss, and total-variation regularizer. The proposed model achieves state of the art performance for Planck `Commander` 2018 CMB map inpainting.**

***Index Terms*—variational autoencoder, cosmic microwave background, inpainting, deep learning, convolutional neural networks, uncertainty quantification, KL-divergence regularization, perceptual loss, total variation, angular power spectrum, VGG-16, ImageNet.**

I. Introduction

Cosmic Microwave Background (CMB) map details the cooled remnant of the light or electromagnetic radiation caused by the Big Bang in the early stages of the universe, which one can still observe today [1]. The CMB is a valuable resource containing information about how the early universe was formed. It is at a uniform temperature with small fluctuations visible only with high precision telescopes. By measuring and understanding fluctuations, cosmologists can learn the origin of galaxies and explore the basic parameters of the Big Bang theory. The CMB map is produced by map-marker using component separation such as `Commander` [2], [3], `NILC` [4], `SEVEM` [5] and `SMICA` [6]. Due to contamination of thermal noise in the galaxy, the region near the equator of the map (here the equator corresponds to the galaxy) has missing observations for CMB. The left picture of Figure 1 shows the Planck 2018 `Commander` CMB map, where the noise near the equator is apparent. The right panel of the picture shows the zoomed-in image containing the missing observations at noisy pixels.

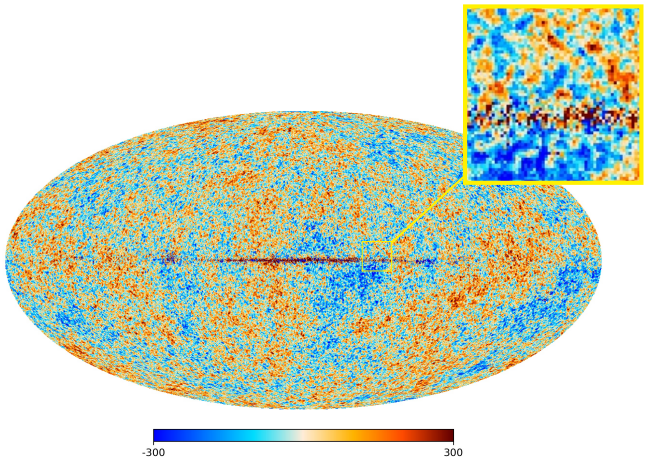


Fig. 1. Planck 2018 `Commander` CMB Map. The zoomed-in picture is a segmented image used as a test data in the CosmoVAE model, where the dark red indicates the missing pixels to be restored.

Estimating, or inpainting missing observations is a largely unsolved problem in CMB research [7]. In this work, we

Yanan Fan is grateful for the support of ARC center of excellence for mathematical and statistical frontiers (ACEMS). Yu Guang Wang acknowledges support from the Australian Research Council under Discovery Project DP180100506.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

propose a new inpainting method for restoring the missing observations — CosmoVAE, stemming from a Bayesian deep learning model. CosmoVAE is a variational autoencoder that takes deep convolutional neural networks as encoder and decoder and the whole model under the Bayesian inferential framework. The CMB field can be considered as a realization of the Gaussian random field on the two-dimensional unit sphere [1], [8]. The CMB map in high resolution is a big data set (for example, in resolution $N_{\mathrm{Side}} = 2048$, the data is stored at more than 50 million HEALPix points [9]). We show that this Bayes based deep learning model provides an efficient approach for inpainting big CMB data.

II. Literature Review

CMB maps have been produced using the following four approaches. `NILC` [4] is a linear combination which works in the needlet domain. `SEVEM` [5] is a foreground template-cleaning approach that works in the pixel domain to component separation. `SMICA` [6] is a non-parametric approach using spherical harmonics as basis. `Commander` [2], [3] is an MCMC based map maker. In `Commander`, the CMB map is viewed as a realization of a Gaussian random field which can be generated by using the angular power spectrum C_ℓ. The posterior density function of the power spectrum C_ℓ is estimated by using Gibbs sampling. Also, one can use Blackwell-Rao approximation for estimating the angular power spectrum [10]. The produced CMB map, however, has missing observations. Most missing observations (mask regions) are concentrated around the equatorial region due to thermal dust in the galaxy. Inpainting the CMB map is expected to reduce the uncertainty of the estimates for the cosmological parameters.

There are several existing methods for CMB map inpainting. For example, Planck consortium [11] uses Gaussian constrained realization to replace the high-foreground regions. One can also use Gaussian process regression for CMB image inpainting, which estimates the covariance function for the image pixels and then interpolates the missing pixels [12], [13]. Although traditional statistical methods are computational easy, they are computationally expensive and difficult to scale up for the large-sized data set of CMB. According to Gruetjen et al. [14], inpainting is an alternative way to construct accurate cut-sky CMB estimators. Gruetjen showed that one could apply inpainting to the problem of unbiased estimation of the power spectrum, which utilizes the linearity of inpainting to construct analytically debiased power spectrum estimates from inpainted maps.

Recently, with deep learning, inpainting methods have significantly improved reconstruction results by learning semantics from large scale data sets. These methods typically use different kinds of convolutional neural networks as mapping functions from masked images to inpainted images end-to-end. Context encoder [15] is the first algorithm that uses a deep learning approach to reconstruct masked images. It utilizes the auto-encoder architecture and convolutional neural network with reconstruction and adversarial loss for inpainting. It can achieve surprisingly good performance for restoring an image with a square mask region. Yang et al. [16] takes the result from context encoders as input and then propagates the texture information from non-mask regions to fill the mask regions as post-processing. Yu et al. [17] proposed an end-to-end image inpainting model with global and local discriminators to ensure the color and texture consistency of generated regions with surroundings. This method has no limitation on the location of mask regions, but the mask shape needs to be rectangular. Since the real CMB mask region is irregular, this method is not suitable. To achieve better inpainting performance for irregular masks, partial convolution [18] was proposed by Liu et al., where the convolution operation can skip the missing pixels and only use valid pixels. This specified convolution operation can appropriately process irregular masks and would not lead to artifacts such as color discrepancy and blurriness. With the combination of reconstruction loss, perceptual loss [19], and total variation loss as penalty term [20], the model achieves state-of-the-art image inpainting results on large data sets such as human faces and landscapes.

Researchers have proposed many generative probabilistic models based on neural networks in the past decade. Variational autoencoder (VAE) [21] is one of the most popular approaches. With a well-trained VAE model, we can generate various kinds of images by the sampling latent variables with a specific distribution (e.g., Gaussian distribution). In many cases, one is interested in training the generative models conditional on the image features such as labels and characteristics of the human face. Sohn et al. [22] proposed conditional variational autoencoder whose input observations modulate the prior on Gaussian latent variables, which then generate the outputs by the decoder. After training, it produces output images by controlling the latent variables. Ivanov et al. [23] modified the conditional VAE and proposed a variational autoencoder with arbitrary conditioning (VAEAC) model. VAEAC can learn the features from non-missing pixels and predict the missing pixels values. Ivanov et al. used this method to inpaint four different data sets, which achieved state of the art performance.

Our network architecture adopts the auto-encoder architecture, which is widely used in representation learning and image inpainting. We use variational Bayesian approximation to obtain the evidence lower bound (ELBO) of the likelihood of the reconstructed image. The ELBO will be used to form our loss function. In addition, we use skip-connection to build a sufficiently deep network and add perceptual loss to the loss function. We also replace the partial convolution layer [18] with the vanilla layer, which is more appropriate for CMB image inpainting tasks.

III. CosmoVAE for CMB

Our proposed model, as illustrated in Figure 2, is based on the variational autoencoders (VAE) [21], where the encoder and decoder combine the convolutional neural networks (CNNs) and multilayer perceptron (MLP). This modified VAE also uses skip connection between the encoder and decoder, which builds a U-Net-like architecture [18] in order to guarantee optimal transfer of spatial information from input to the output image. The basic

autoencoder compresses the high-dimensional input x (i.e., the segmented image of CMB map) to a low-dimensional latent variable z, and then decompresses z back to high-dimensional output y, and the input x and output y should be the same. In the CosmoVAE, the encoder takes the image with a missing region and produces a latent feature representation; the latent features are used by the decoder to produce the missing image. In the training stage, the generated image is compared with the ground truth, where the loss function is composed of the negative variational lower bound, perceptual loss, and a total variation regularizer. A well-trained model can rebuild the mask regions of the CMB map.

A. Statistical Interpretation in VAE

Let us consider the joint probability distributions of three random variables $((\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \in \mathcal{X} \times \mathcal{Z} \times \mathcal{Y})$, where $\mathbf{X} = \left\{\mathbf{x}^{(i)}\right\}_{i=1}^{N}$ is the input masked CMB maps, $\mathbf{Z}$ is the vector of latent variables and $\mathbf{Y}$ is the vector corresponding to the reconstructed CMB map. We use neural networks for probabilistic encoder (Q_ϕ) and decoder (P_θ). To be precise, the probabilistic encoder is defined as $q_\phi(\mathbf{z}|\mathbf{x})$ where $p_\phi(z) = \int_{\mathcal{X}} q_\phi(z|x)p(x)dx$ for all $z \in \mathcal{Z}$, ϕ denotes the parameters of the neural network. And the probabilistic decoder is given by $p_\theta(\mathbf{y}|\mathbf{z})$, with θ the parameters of the decoder network, and

$$p_\theta(y) := \int_{\mathcal{Z}} p_\theta(y|z)p(z)dz \quad \forall y \in \mathcal{Y}.$$

The marginal log-likelihood of output y is given by $\log p_\theta(y) = \sum_{i=1}^{N} \log p_\theta(y^{(i)})$, for $i = 1, \ldots, N$ training samples, which can be expressed as variational lower bound and used as a surrogate objective function where:

$$\begin{aligned} \log p_\theta\left(y^{(i)}\right) &\geq \mathbf{E}_z\left[\log p_\theta\left(y^{(i)}|z\right)\right] \\ &\quad - D_{\mathrm{KL}}\left(q_\phi\left(\cdot|x^{(i)}\right) \| p(\cdot)\right) \\ &= \mathcal{L}\left(y^{(i)}, x^{(i)}, \theta, \phi\right). \end{aligned} \tag{1}$$

The variational lower bound consists of negative reconstruction loss and Kullback-Leibler divergence D_{KL} between the approximated posterior $q_\phi(z|x)$ of the encoder network and the prior $p(z)$ of the latent variable.

By dual principle, maximizing log-likelihood function $\log p_\theta\left(y^{(i)}\right)$ is equivalent to minimizing the negative lower bound $\mathcal{L}\left(y^{(i)}, x^{(i)}, \theta, \phi\right)$ (with respect to the parameters θ and ϕ). We thus need to find the gradient of the marginal likelihood - this can be achieved via standard Monte Carlo approximation for the expectation term. In order to reduce the variance in the gradient estimator, we use the reparameterization trick [21].

B. Prior Specification for the Latent Variables

The KL-divergence term in the variational lower bound in (1) can be interpreted as regularization for the parameters θ and ϕ, encouraging the approximate posterior $q_\phi(z|x)$ to be close to the prior $p(z)$. The posterior distribution $q_\phi(z|x)$ can be estimated by probabilistic encoder but the prior distribution remains to be determined.

As CMB is a Gaussian random field and the temperature quantity can be expressed in Fourier series:

$$T(\hat{p}) = T_{\mathrm{CMB}}[1 + \Theta(\hat{p})],$$

where $\Theta(\hat{p})$ is the temperature anisotropy in the direction $\hat{p}$ which can be expanded with Fourier coefficients $a_{\ell m}$:

$$\Theta(\hat{p}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{p}).$$

A Gaussian random field is fully determined by its mean and variance. The $a_{\ell m}$ follows a zero-mean Gaussian with variance C_ℓ, where C_ℓ is the CMB angular power spectrum. Here, we assume that our latent variable $\mathbf{Z}$ is given by the angular power spectrum C_ℓ. And the generative field is connected with the latent variable by $a_{\ell m} \to T \to \mathbf{X}$ and $a_{\ell m} \sim \mathcal{N}(0, C_\ell)$. The KL-divergence term $D_{\mathrm{KL}}\left(q_\phi\left(a_{\ell m}|x^{(i)}\right) \| \mathcal{N}(0, C_\ell)\right)$, is now with $a_{\ell m} \sim \mathcal{N}(0, C_\ell)$, the divergence between the posterior distribution $q_\phi(a_{\ell m}|x)$ and the prior distribution $\mathcal{N}(0, C_\ell)$. By doing this, we now assign physical meaning to the VAE model; that is, the latent variable is the angular power spectrum of the learned field, and the angular power spectrum samples the Fourier coefficients of the generative field.

C. Loss Function

The overall loss function to train the model is defined as:

$$\mathbb{L} = \lambda_1 \mathbb{L}_{\mathrm{rec}} + \lambda_2 \mathbb{D}_{\mathrm{KL}} + \lambda_3 \mathbb{L}_{\mathrm{perceptual}} + \lambda_4 \mathbb{L}_{\mathrm{TV}}.$$

with appropriate weights λ_i for each term. Here the weights λ_i are hyper-parameters, tuned artificially. The $\lambda_1 \mathbb{L}_{\mathrm{rec}} + \lambda_2 \mathbb{D}_{\mathrm{KL}}$ is the negative variational lower bound which consists of the reconstruction loss and the KL-divergence regularization, the $\mathbb{L}_{\mathrm{perceptual}}$ is the perceptual loss [19], and the $\mathbb{L}_{\mathrm{TV}}$ is the total variation loss [20].

a) Reconstruction Loss: We use a binary mask M which is 0 for pixel outside the masked region and 1 for pixel inside the masked region. For the network prediction $\hat{y}$ and the ground truth y, the reconstruction loss is then

$$\mathbb{L}_{\mathrm{rec}} = \frac{1}{N} \left\|(1 - M) \odot (\hat{y} - y)\right\|_1 + \frac{1}{N} \left\|M \odot (\hat{y} - y)\right\|_1,$$

where N denotes normalization constant (where $N = C{*}H{*}W$ and C, H, W are the channel size, and the height and width of image).

b) Regularization: The regularization term is the KL-divergence between posterior distribution $q_\phi(a_{\ell m}|x)$ and the prior distribution. The latent variable $a_{\ell m} \sim \mathcal{N}(0, C_{\ell m})$, then, $D_{\mathrm{KL}}\left(q_\phi\left(a_{\ell m}|x^{(i)}\right) \| \mathcal{N}(0, C_{\ell m})\right)$ can be solved analytically by

$$\mathbb{D}_{\mathrm{KL}} = \frac{1}{2} \sum_{i=1}^{N} -\log \sigma_{i,1}^2 - \log C_{\ell m}^2 - 1 + \frac{\sigma_{i,1}^2 + \mu_{i,1}^2}{C_{\ell m}^2},$$

where $q_\phi\left(a_{\ell m}|x\right) \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ is the approximated posterior distribution of the encoder network.

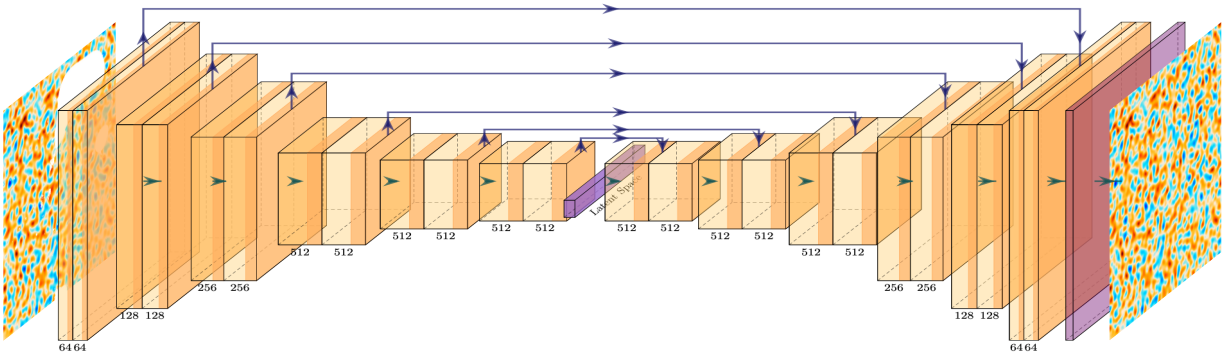


Fig. 2. Pictorial representation of the variational autoencoder model. The encoder and decoder, which are connected by channel-wise latent variables, have U-net architecture. The encoder and decoder are deep convolutional neural networks, each with six blocks and three fully-connected layers.

c) Perceptual Loss: Perceptual loss was first proposed in [19] to preserve image contents in style transfer and is now widely used for image inpainting. The perceptual loss computes the ℓ_1 loss of high-level feature maps between the predicted image and ground truth:

$$\mathbb{L}_{\text{perceptual}} = \sum_{n=0}^{N-1} \left\| \Psi_n^{\hat{\mathbf{y}}} - \Psi_n^{\mathbf{y}} \right\|_1,$$

where Ψ is the activation map of the pth selected layer which lies in a higher level feature space in ImageNet-pretrained VGG-16 [24]. We use Pool-1, Pool-2 and Pool-3 layers of VGG-16 for our loss.

d) Total Variation Loss: The final term for the loss is the total variation loss as a smoothing penalty:

$$\mathbb{L}_{\text{TV}} = \sum_{(i,j)\in P,\,(i,j+1)\in P} \frac{\left\| \hat{y}^{i,j+1} - \hat{y}^{i,j} \right\|_1}{N_{\text{hole}}} + \sum_{(i,j)\in P,\,(i+1,j)\in P} \frac{\left\| \hat{y}^{i+1,j} - \hat{y}^{i,j} \right\|_1}{N_{\text{hole}}},$$

where P is the region of 1-pixel dilation of the mask region and N_{hole} is the number of pixels in the mask region [20].

IV. Experimental Results

A. Generating Training and Test Data Sets

In the context of CMB map inpainting, preparing training sets has two challenges: the original map is not flat (it resides more naturally on a sphere), and there is only a single CMB map that we can use for training. We can solve the problems by projecting the spherical CMB map to the plane by *Cartesian Projection*. In order to have a sufficient number of data sets, we segment the whole map with around 50 million pixels into thousands of small images with 400×400 pixels. This is a reasonable approach as we assume the CMB field as an isotropic random field on the two-dimensional sphere. The resulting small images will be random fields on the plane, which can be assumed independent and identically distributed. We can then treat each small image of the CMB map independently as training data for the deep learning model.

a) Image data set: The data of CMB maps are stored at the HEALPix points on the unit sphere [9][1]. In the experiments, we use Planck 2018 `Commander` CMB map with $N_{\text{Side}} = 2048$ (and $50,331,648$ points), downloaded from Planck Legacy Archive[2]. To generate small images cropped from the original map, we project the original map to a large flat 2D image using *Cartesian Projection* of *healpy* (Python) package [25]. We equally space the points on the projected CMB map, and for each point which then becomes the center of the small image, we take the rectangle whose latitudinal and longitudinal angles ranging from -5° to 5° and -10° to 10° from the center. The spherical CMB full-sky map is then segmented into 4,042 small flat images, each with resolution 400× 400. The 772 among all small images contain missing regions and will constitute the test data set, and the remaining clean 3,320 images will be the training set.

b) Mask data set: The generative method of mask data set is similar with training data set. We use Planck 2018 Component Separation Inpainting Common mask in Intensity, downloaded from Planck Legacy Archive[2], see Figure 3. The spherical mask map is divided with the same size and same centres as the full-sky CMB map, and segmented into $4,042$ small flat mask images. Each mask image corresponds to a CMB image in the training data set, and will be used in masking the missing pixels in training and test. There are 772 masks having missing pixels, and we randomly select them during the training process.

B. Network Architecture and Training

Our proposed model is implemented in Keras[3]. The network architecture is a U-net-like network. The encoder and decoder have architectures that contain six blocks and three fully connected layers. The encoder network has architecture 64-128-256-512-512-512, and the decoder network architecture is 512-512-512-256-128-64. The latent variable is channel-wise with 2,507 component parameters, which corresponds to

[1] http://healpix.sf.net
[2] https://pla.esac.esa.int/#maps
[3] https://github.com/keras-team/keras

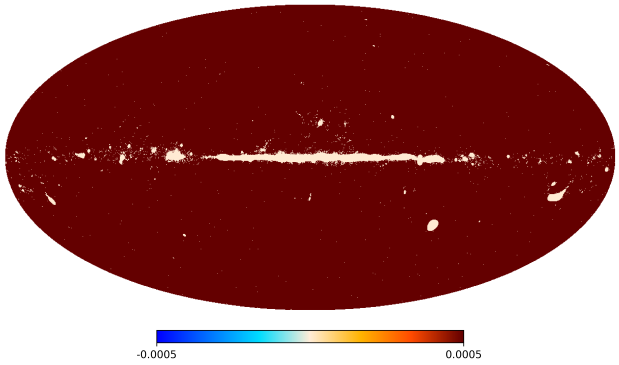


Fig. 3. Planck 2018 Component Separation Inpainting Common mask in Intensity. It is a binary map enciphered with 0 and 1 for clean and noisy pixels. The whole mask map is segmented into 4,042 small images in accord with the small images of CMB full-sky map, each with 400×400 pixels. They are used for masking the full-sky images in training and test.

angular power spectrum C_ℓ with ℓ up to 2,507. (There are thus the 6,290,064 Fourier coefficients which approximately represent the learned field.) The encoder output samples the latent variable. The whole network is trained using 3,320 400×400 images with batch size four and maximal epoch 1,000. The model is optimized using Adam optimizer with the parameters: learning rate 0.0002 and β_1 = 0.5. In the training stage, we use the best-fit ΛCDM CMB TT power spectra from the Planck PR3 baseline[4] as C_ℓ.

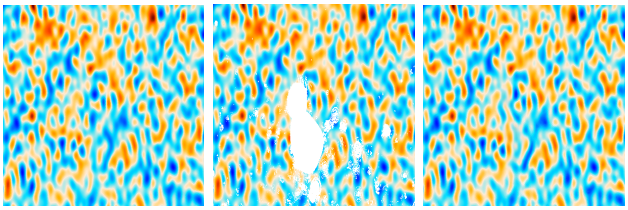
The experiment is carried out in Google Colab Nvidia Tesla K80 with 2496 CUDA cores, compute 3.7, 12GB GDDR5 VRAM. With no pre-trained weights, it roughly takes 6 hours to achieve convergence. Here the batch size is chosen to adapt to the memory allowed in Colab. If memory is sufficiently large, one can speed up the training by increasing the batch size.

C. Test Results

To evaluate our model's capacity for CMB image inpainting and evaluate the method's performance, we apply real masks on small CMB images with unknown pixel values in the mask. As illustrated in Figure 4, the result of CosmoVAE is visually close to the ground truth image. Quantitatively, the mean square error (MSE), mean absolute error (MAE), and peak signal-to-noise ratio (PSNR) over the masked pixels are 0.0055, 0.134 and 23.989 respectively, which reveals the excellent performance of the model.

Once we have trained the CosmoVAE, we can use it to predict the missing pixels of each small CMB image in the test data set. Figure 5 shows five examples of the predicted results by CosmoVAE. We compare our inpainted results with Planck 2018 results [11]. The left-most plot shows the inpainted CMB image of Planck 2018 results [11]. The second column shows the original (un-inpainted) CMB image with an irregular missing region, which is the actual input of the trained CosmoVAE. The third column panel shows the corresponding

[4]https://pla.esac.esa.int/#cosmology



(a) Ground Truth (b) Masked Image (c) CosmoVAE Predicted

Fig. 4. Comparison of predicted results of the proposed models with ground truth. Our models can leverage the surrounding textures and structures and consequently generate lifelike images with no blurriness in the masked area. The mean square error (MSE), mean absolute error (MAE), and peak signal-to-noise ratio (PSNR) are around 0.0055, 0.0134, and 23.989.

predicted images, where the network restores the missing region. As we can observe, the trained CosmoVAE can inpaint CMB images with irregular mask regions, even if the mask area is big, and when there are multiple mask regions. The predicted results are evident as compared with the Planck 2018 results. CosmoVAE thus provides a useful inpainting model for the CMB map.

D. Uncertainty Quantification

One can interpret the CosmoVAE as a probability model in a similar way to that of AEVB [21]. More concretely, denote the encoder neural networks model mapping input x to a stochastic latent variable z (see Figure 2) by the conditional probability $p_\theta(z|x)$, where θ denotes the parameters of the encoder network. Similarly, denote the decoder neural networks model mapping z to the output x, as $p_\phi(x|z)$ and ϕ denote the weights and biases of the decoder network.

In the Bayesian inferential context, $p(z|x)$ is the posterior distribution obtained from the prior and likelihood combination where $p(z|x) \propto p_\phi(x|z)p(z)$, and $p_\phi(x|z)$ is the likelihood (the "generative" model or, the decoder), and $p(z)$ is the prior. In the variational inference framework, a distribution $q_\lambda(z|x)$ can be used to approximate this intractable posterior $p(z|x)$.

If we take the latent variable z as a $\mathcal{N}(0,1)$ variate and the relationship between x and z is given by the encoder network with parameter θ. Then taking $q_\lambda(z|x)$ to be a Normal distribution with parameters $\lambda = (\mu(\theta), \sigma(\theta))$, minimising the Kullback-Leibler divergence between $q_\lambda(z|x)$ and the true posterior $p(z|x)$ is now the same as minimising the loss functions with respect to θ and ϕ, where different loss functions $\mathbb{L}(x)$ broadly correspond to the noise distribution we assume for x. In the CosmoVAE, the parameters θ depend on the input x_1 and ϕ depend on x_2, where the components of $x = (x_1, x_2) \equiv ((1-M)\odot x, M \odot x)$ are respectively the pixels outside and inside the masked regions.

Having a posterior distribution allows us to obtain uncertainty estimation. To do this, consider the posterior predictive distribution of x_2, where

$$p(x_2|x_1) = \int p(x_2|z)p(z|x_1)dz,$$

where $p(z|x_1) = q_\lambda(z|x)$ is simply the posterior distribution obtained from the variational approximation, using training

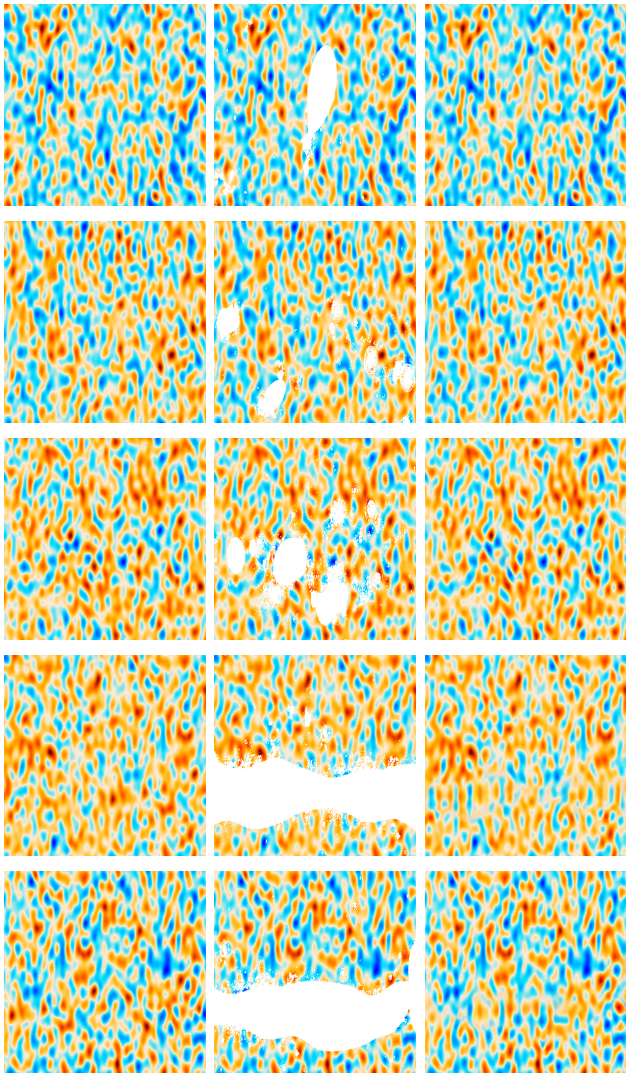


Fig. 5. Comparison of the results of our CosmoVAE model and Planck 2018 `Commander`. The left image is Planck 2018 `Commander` results. The middle is the contaminated image after cropping. The right image is the inpainted image. The CosmoVAE has good performance in predicting the missing observations in various mask regions.

data x, and $p(x_2|z)$ is the generative model, and when a discriminator model is added to $p(x_2|z)$, we can sample x_2 and retain images for which the discriminator has computed as true. The variability in the pixels of the "true" images provides us with the uncertainty measure.

Figure 6 shows the uncertainty in the CosmoVAE predictor. The second column is the CMB image with the missing region masked. The fourth column shows the standard deviation of the inpainted images of the trained-CosmoVAE at each missing pixel over 100 different realizations of the latent variable. Compared with the Planck 2018 result in the first column for the same image segment from the full-sky CMB map, the CosmoVAE inpainted image has the same image quality. The Std Dev. for each image is very small, and the location where the Std Dev. is significant in the square image is a fractional part of the mask region. Thus, the uncertainty of CosmoVAE

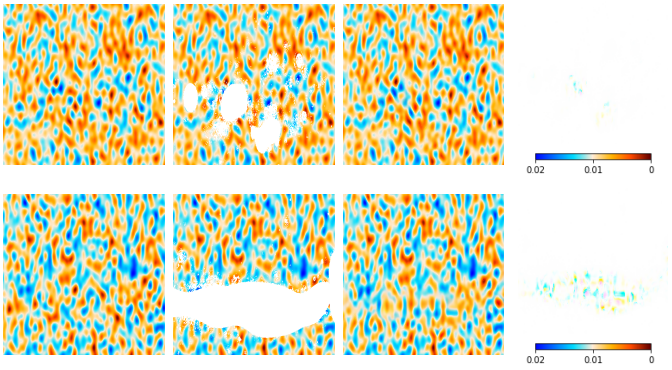


Fig. 6. A closer look at the test result for two images of the Planck 2018 `Commander` CMB map. The left column plots the inpainted image from Planck 2018 results. The second column plots the original image with the irregular missing region. The third column plot shows the predicted image by CosmoVAE. The right-most column plots the standard deviation of the test outputs for the same sample using 100 trained models.

is controllable and has little effect on the predicted image.

V. Conclusion and Future Plan

Statistical challenges to processing the CMB data is one of the biggest challenges in the analysis of CMB data. In this work, we reconstruct the cosmic microwave background radiation (CMB) map by using a modified variational autoencoder (VAE) as our baseline model. We cut the full-sky CMB map into many small images in order to generate our image and mask datasets, and then in training to inpaint the missing area with arbitrary mask shapes. To enhance the performance, we combine our neural network with the angular power spectrum, which can generate the Fourier coefficients of the Gaussian random field. Furthermore, we modified the original VAE loss function by adding in the perceptual loss and the total-variation regularizer. This new VAE model assigns cosmological meaning to the parameters of the network and achieves a state of the art performance for CMB map inpainting.

To better complete image inpainting task and for cosmology study, one needs to reconstruct the full-sky CMB map from all small inpainted CMB images. We can use the inpainted full-sky CMB map to estimate cosmological parameters such as the angular power spectrum C_ℓ, which can be computed directly by the healpy package. The inpainting of the CMB map will help reduce the uncertainty in the parametric estimation. By Olivier et al. [26], any method which is based on the marginal log-likelihood, including VAE, necessarily leads to the blurriness of output due to the gap between true negative log-likelihood and the upper bound (ELBO). We can further modify our loss function to improve the quality of reconstructed images by replacing the KL-divergence with GAN or WGAN (more stable) as regularizer. Our model can also be a baseline model for the reconstruction of other random fields (besides the CMB field). We will probe these problems in our future work.

Acknowledgments: Some of the results in this paper have been derived using the HEALPix [9] package.

REFERENCES

[1] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak *et al.*, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," *arXiv preprint arXiv:1807.06205*, 2018.

[2] H. K. Eriksen, C. Dickinson, C. R. Lawrence, C. Baccigalupi, A. J. Banday, K. M. Górski, F. K. Hansen, P. B. Lilje, E. Pierpaoli, M. D. Seiffert, K. M. Smith, and K. Vanderlinde, "Cosmic microwave background component separation by parameter estimation," *ApJ*, vol. 641, pp. 665–682, Apr. 2006.

[3] H. K. Eriksen, J. B. Jewell, C. Dickinson, A. J. Banday, K. M. Górski, and C. R. Lawrence, "Joint Bayesian component separation and CMB power spectrum estimation," *ApJ*, vol. 676, pp. 10–32, 2008.

[4] J. Delabrouille, J.-F. Cardoso, M. Le Jeune, M. Betoule, G. Fay, and F. Guilloux, "A full sky, low foreground, high resolution CMB map from WMAP," *A&A*, vol. 493, pp. 835–857, 2009.

[5] R. Fernández-Cobos, P. Vielva, R. B. Barreiro, and E. Martínez-González, "Multiresolution internal template cleaning: an application to the Wilkinson Microwave Anisotropy Probe 7-yr polarization data," *MNRAS*, vol. 420, pp. 2162–2169, Mar. 2012.

[6] J.-F. Cardoso, M. Le Jeune, J. Delabrouille, M. Betoule, and G. Patanchon, "Component separation with flexible models—application to multichannel astrophysical observations," *IEEE J. Sel. Top. Signal Process.*, vol. 2, pp. 735–746, Nov. 2008.

[7] P. Cabella, D. Marinucci *et al.*, "Statistical challenges in the analysis of cosmic microwave background radiation," *Ann. Appl. Stat.*, vol. 3, no. 1, pp. 61–95, 2009.

[8] D. Marinucci and G. Peccati, *Random fields on the sphere: representation, limit theorems and cosmological applications*. Cambridge University Press, 2011, vol. 389.

[9] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, "HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere," *ApJ*, vol. 622, pp. 759–771, Apr. 2005.

[10] M. Chu, H. Eriksen, L. Knox, K. Górski, J. Jewell, D. Larson, I. ODwyer, and B. Wandelt, "Cosmological parameter constraints as derived from the Wilkinson Microwave Anisotropy Probe data via Gibbs sampling and the Blackwell-Rao estimator," *Phys. Rev. D*, vol. 71, no. 10, p. 103002, 2005.

[11] Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak, K. Benabed *et al.*, "Planck 2018 results. IV. Diffuse component separation," *A&A*, 2019.

[12] A. Aghamousa, J. Hamann, and A. Shafieloo, "A non-parametric consistency test of the λCDM model with planck CMB data," *J. Cosmol. Astropart. Phys.*, vol. 2017, no. 09, pp. 031–031, sep 2017.

[13] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[14] H. Gruetjen, J. R. Fergusson, M. Liguori, and E. Shellard, "Using inpainting to construct accurate cut-sky CMB estimators," *Phys. Rev. D*, vol. 95, no. 4, p. 043532, 2017.

[15] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *CVPR*, 2016, pp. 2536–2544.

[16] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *CVPR*, 2017, pp. 6721–6729.

[17] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *CVPR*, 2018, pp. 5505–5514.

[18] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *ECCV*, 2018, pp. 85–100.

[19] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.

[20] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016, pp. 694–711.

[21] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *ICLR*, 2014.

[22] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *NIPS*, 2015, pp. 3483–3491.

[23] O. Ivanov, M. Figurnov, and D. Vetrov, "Variational autoencoder with arbitrary conditioning," in *ICLR*, 2019.

[24] N. Sundaram, T. Brox, and K. Keutzer, "Dense point trajectories by GPU-accelerated large displacement optical flow," in *ECCV*, 2010, pp. 438–451.

[25] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, "Healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python," *J. Open Source Softw.*, vol. 4, p. 1298, 2019.

[26] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf, "Wasserstein auto-encoders," in *ICLR*, 2018.