Convolutional Neural Processes for Inpainting Satellite Images

Alexander Pondaven[*]   **Märt Bakler**[*]   **Donghu Guo**[†]
Hamzah Hashim[†]   **Martin Ignatov**   **Harrison Zhu**

Imperial College London
{ap2619,mb1221,dg321,hh2019,mgi18,hbz15}@ic.ac.uk

Abstract

The widespread availability of satellite images has allowed researchers to model complex systems such as disease dynamics. However, many satellite images have missing values due to measurement defects, which render them unusable without data imputation. For example, the scanline corrector for the LANDSAT 7 satellite broke down in 2003, resulting in a loss of around 20% of its data. Inpainting involves predicting what is missing based on the known pixels and is an old problem in image processing, classically based on PDEs or interpolation methods, but recent deep learning approaches have shown promise. However, many of these methods do not explicitly take into account the inherent spatiotemporal structure of satellite images. In this work, we cast satellite image inpainting as a natural meta-learning problem, and propose using convolutional neural processes (ConvNPs) where we frame each satellite image as its own task or 2D regression problem. We show ConvNPs can outperform classical methods and state-of-the-art deep learning inpainting models on a scanline inpainting problem for LANDSAT 7 satellite images, assessed on a variety of in and out-of-distribution images.

1 Introduction

Satellite images are a valuable resource to research communities. With the surge of computational methods using remote sensing data, satellite images have been widely used for instance in epidemiology (Weiss et al., 2019), social sciences (Yeh et al., 2020) and crop yield modelling (Zhu et al., 2021). One of the satellites that is commonly used is the LANDSAT 7 (USGS) satellite due to its long temporal coverage and high spatial resolution. However, due to a mechanical fault in the satellite's scanline corrector (SLC), satellite imagery post 31st May 2003 suffers from lines of missing and corrupted pixels (Figure 1). As they occupy a significant area of the satellite images (about 20% of the data), the images obtained from LANDSAT 7 lost much of their research use as the scanlines significantly impair the performance of computational methods using the images.

To 'repair' the images, inpainting techniques must be used. Image inpainting, also known as gap-filling, aims to fill the corrupted pixels in an image with values that resemble the original pixel values as closely as possible. Inpainting for corrupted image data is an active area of research and many methods have been introduced. For instance, many deterministic methods have been proposed, that use higher order differential equations (Burger et al., 2009; Bertalmio et al., 2001; Bertozzi & Schönlieb, 2011). Moreover, recent advances in deep learning have shown promising results in image inpainting. Ronneberger et al. (2015) introduced the celebrated U-Net, which was originally used for biomedical image segmentation, but can be trained to perform image inpainting, and Liu et al. (2018)

---

[*]Equal contribution. † Order decided via a coin toss.

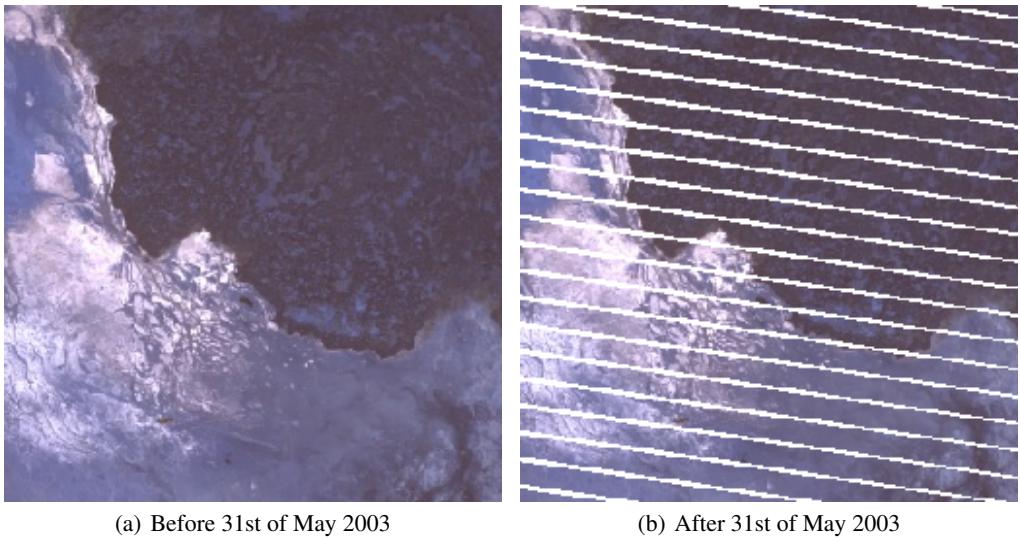(a) Before 31st of May 2003   (b) After 31st of May 2003

Figure 1: LANDSAT 7 images before and after the scanline corrector failure.

introduced Partial Convolutions (PartialConv), which is a modification to the classical convolutional layer to make it suitable for inpainting.

One drawback of the traditional deep learning methods is that they treat all of the images as a single task. They approximate a single function f_θ that is learned by the network and it is used during inference to get the predictive values $\tilde{y} = f_\theta(x)$. This approach does not take into account the spatiotemporal differences between different images, where different predictive functions f_{θ_m} could better suit different tasks m. This kind of problem may be better suited to meta-learning methods, which learn task-specific representations and are better at capturing the differences between various inputs.

In meta-learning, we learn a task representation that can determine the exact function for each distinct task, $\tilde{y} = f_{\theta_m}(x)$, where m is the current task. Garnelo et al. (2018a) introduced a meta-learning approach called Conditional Neural Processes (CNPs) that uses an encoder-decoder architecture to obtain a predictive distribution $p_{\theta_m}(y|x)$, which is equivalent to obtaining a distribution over predictive functions f_{θ_m}. Gordon et al. (2020) and Foong et al. (2020) introduced Convolutional Conditional Neural Processes (ConvCNPs) and Convolutional Latent Neural Processes (ConvLNPs) respectively, which are better suited for image inpainting tasks due to their translational equivariance property. These Convolutional Neural Processes (ConvNPs) are shown to exhibit very good few-shot and zero-shot learning capabilities as well, which has been demonstrated for inpainting weather data (Foong et al., 2020; Markou et al., 2022).

In this paper, we show that ConvNPs can be used for satellite image inpainting, in particular, to correct the scanlines of LANDSAT 7 images. We use ConvCNPs and ConvLNPs with an MS-SSIM similarity score loss function (Wang et al., 2004) that is better for generating sharp images. We show that our ConvNPs can outperform many state-of-the-art image inpainting models and accurate inpainting models can be trained with a relatively small dataset (training set of 800 images with dimensions 128x128 or 64x64). ConvNP models also show good performance for both in-distribution (inpainting of similar images as they are trained on) and out-of-distribution (OOD) satellite images (zero-shot tasks, images of different regions than the model was trained on). The consequence of the latter is that we are then able to construct a **global inpainter** despite only training with images from a **small subset** of spatiotemporal locations. In addition, via a downstream synthetic CNN regression problem, with inputs being the imputed images, we show that using the ConvLNP imputed images yield the closest results to regression over the original clean images.
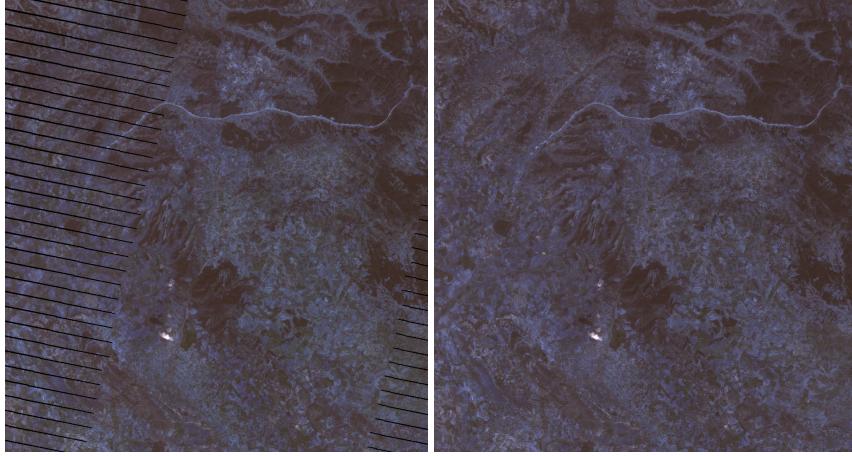
Figure 2: Kenya 1024x1024 by predicting on 64x64 patches with ConvCNP. (Left) Original image (Right) Inpainted image.

2 Related work

Classical approaches: Single source methods involve inpainting based on the undamaged pixels in the image. Traditional methods include interpolation and PDE or convolution based methods like Bertalmio's, Telea's and Oliveira's algorithms (Bertalmio et al., 2001; Telea, 2004; Richard & Chang, 2001), but these methods only inpaint by looking at neighbouring pixels within each image. Exemplar-based inpainting matches patches in a specific order (Shroff & Bombaywala, 2019), but these methods use defined low-level patterns that do not capture semantic structure in the images. The Navier-Stokes (Bertalmio et al., 2001) inpainting algorithm uses ideas from classical fluid dynamics and is used as a baseline for our problem. It is deterministic and does not require training, but the method does not utilise or learn any prior knowledge of the data distribution. In addition, Scaramuzza & Barsi (2005) introduced an inpainting algorithm specifically for LANDSAT 7, which compared pre- and post-2003 images for scanline filling in multiple phases, and is adopted for use as one of the official LANDSAT 7 products. However, the data-engineering required for this method is extremely complicated and we are unable to reproduce it.

Deep learning approaches: Deep learning approaches to inpainting are more globally consistent and achieve better local details. Many of these take the form of encoder-decoder CNNs, such as U-Net (Ronneberger et al., 2015) or GAN-based algorithms Pathak et al. (2016). U-Net consists of a U-shaped CNN that first downsamples the input image using convolutions and pooling layers, and then upsamples using transposed convolutions and skip-connections. Neural Path Synthesis Yang et al. (2016) enhances the GAN-based approach with a texture network inspired by style transfer to transfer the style of the context pixels to improve local details. Multi-scale discriminators have been employed as well to capture global and local details Iizuka et al. (2017); Demir & Ünal (2018).

Another approach is to mask the missing pixels and enhance the network with custom operations on the masked pixels. Inpainting irregular holes has seen success using this approach by using Partial Convolutions Liu et al. (2018), where the proposed architecture introduces novel partial convolution operations and blocks with custom loss functions, which are specially designed to take into account missing data. VAE networks like HI-VAEs (Nazabal et al., 2020) have been shown to be suitable for problems with missing data and can infer masked values. However, these models require large datasets and long training regimes, which make them difficult to generalise to new spatiotemporal locations.

Recent developments such as Dupont et al. (2021, 2022) use a continuous representation for images and pixels by modelling them as functions rather than discrete values, which can also be used for imputing missing values. Lugmayr et al. (2022) introduced RePaint, which uses denoising diffusion probabilistic models and produces high-quality inpainted images. Finally, Foong et al. (2020) and Markou et al. (2022) both perform image inpainting for Earth observation data, specifically gridded weather data, using ConvNPs. This problem is very similar to inpainting satellite images, although

our aim is to inpaint scanlines after the SLC failure of the LANDSAT 7 satellite and we propose to use a more suitable likelihood/loss function.

3   Data

LANDSAT 7 images are downloaded using the Google Earth Engine (GEE) API Gorelick et al. (2017). For this paper, we focus solely on the visible RGB bands of the LANDSAT 7 satellite (B3, B2, B1) with spatial resolutions of 30 meters. The images are sampled from a uniform spatial grid with a grid spacing of 0.4 degrees longitude and latitude. They are downloaded with a dimension of 256x256 pixels, corresponding to a land area of approximately 59km^2. Due to computational limitations, these images are cropped to 64x64 and 128x128, and model results are reported using these sizes. Note that smaller satellite images could be patched together to perform a larger inpainting task as seen in Figure 2 and Appendix A. Satellite images from Kenya are used for training. Data from UK, Norway, Brazil and Nepal are also collected to test the model's capabilities on out-of-distribution (unseen, location-wise) images.

The images are extracted from specific dates and locations, and are divided into pre- and post-2003 (when the SLC broke). All pre-2003 data is from between 1999 to 2003. Post-2003 data is collected from between 2003 to 2004. Images with missing pixels (alpha channel is present) are filtered out for pre-2003 images. UK images are sometimes completely white due to the presence of clouds, which resulted in 'better' inpainting results across all models. To create more challenging out-of-distribution tasks, Norway, Brazil and Nepal images are filtered by only taking images where the middle 64x64 section had less than 90% white pixels (so the cropped 64x64 dimension dataset also had less clouds). Post-2003 data is just used to extract a set of 100 scanline bit masks from Kenya data to apply to un-corrupted images during training. Some images had large sections of missing pixels, so the post-2003 images are filtered to have $< 20\%$ missing pixels, but also at least 100 missing pixels (set arbitrarily) to ensure the presence of scanlines.

4   Methodology

We cast satellite inpainting as a meta-learning problem. The pixel locations on the grid and RGB pixel values at those locations are denoted as $x \in \mathbb{R}^2$ and $y \in \mathbb{R}^3$ respectively. We denote the context set of pixels as $(x_C, y_C) := \{x_i, y_i\}_{i=1}^{N_C}$, and the target set as $(x_T, y_T) := \{\bar{x}_i, \bar{y}_i\}_{i=1}^{N_T}$. The union of the context and target set represents the task $D := \{C, T\}$, where $C = \{x_C, y_C\}$ and $T = \{x_T, y_T\}$. Each task D_m corresponds to an image, which could also be viewed as a 2D function (Dupont et al., 2021, 2022). At prediction time, x_C, x_T and y_C are observed but y_T is not.

Classical inpainting methods would assume that we learn a global function f_θ that gives a prediction $y_T \approx f_\theta(x_T)$. Similarly, U-Net and PartialConv would not explicitly, but **implicitly**, distinguish between different tasks and require enormous training sets with data augmentation in order to learn network weights such that $f_\theta(x_{C_m}, y_{C_m}, x_{T_m}) \approx f_{\theta_m}(x_{T_m})$. We argue that taking the meta-learning viewpoint allows us to **explicitly** take into account the spatiotemporal variations for each task and thus promote efficient learning.

Meta-learning methods (Thrun & Pratt, 2012; Finn et al., 2017) aim to solve the problem of using a distinct function at inference time to predict target set values. In our setting, during training, we learn a global parameter θ, which, given contexts of a task, could also output a **task-specific** representation R_m. The global objective function is given by $\mathbb{E}_{m \sim \mathcal{M}}[\mathcal{L}(D_\eta(E_\xi(x_{C_m}, y_{C_m}))(x_{T_m}), y_T)]$, with $D_\eta(E_\xi(x_{C_m}, y_{C_m}))(x_{T_m}) \approx f_{\theta_m}(x_{T_m})$, where $\theta = (\eta, \xi)$, E_ξ encodes the context set (x_C, y_C) to a task-specific representation, D_η decodes the task-specific representation and target location to the output, and $\mathcal{L}$ is a loss function.

This results in a model that adjusts the predictor function f depending on the context set of the task. One advantage of meta learning is that the predictive functions use both the information from the current context set of the task as well as the information that is shared across tasks, making the method well-suited for OOD tasks. This allows modelling of heterogeneous function distributions and is a beneficial property for satellite image inpainting as they have multiple zero-shot tasks for different spatial locations and times, that are not seen during training.

Neural Processes: Conditional (Garnelo et al., 2018a) and latent neural processes (Garnelo et al., 2018b), as a wider family of Neural Processes (NPs), employ a general encoder-decoder neural network architecture that enables meta-learning of functions or stochastic processes. The encoder E_ξ takes as input the context points C and outputs a task representation $R = E_\xi(C)$, which gets passed to the decoder $D_\eta(R, \cdot)$ to give a task-specific output function distribution. This makes it suitable for satellite image inpainting since the context points C are the pixels that are not missing, and the remaining target points T are pixels covered by the scanlines.

Convolutional Conditional Neural Processes: However, a shortfall of standard Conditional NPs (Garnelo et al., 2018a) for image inpainting is the lack of translation equivariance, which is an important property to have for image data. Gordon et al. (2020) introduced translational equivariance to the NP family through ConvCNPs. For our purposes, we are solely interested in the ConvCNPs for on-the-grid data (e.g. images). With the same notation, we denote the original image as I and the context mask M_C, for which $[M_C]_{i,j} = 1$ if the pixel at location (i, j) is in the context set, and 0 otherwise. Our masked context set is thus given by $Z_C = M_C \odot I$. Concatenating the context mask and the masked context point, we thus get $\phi = [M_C, Z_C]^\intercal$. Applying a convolution to ϕ, we obtain the functional representation $R = \text{Conv}_\theta([M_C, Z_C]^\intercal)$, where Conv_θ is the 2D convolution operator with positively-constrained kernel parameters θ. We then apply the normalisation $R^{(1:C)} = R^{(1:C)}/R^0$. This step is known as **SetConv** (when not evaluated at the target points). We can decode R using a CNN, which includes an absorbed MLP to map the output of the CNN at each location (i, j) to $\mathbb{R}^2$ and gives $\boldsymbol{\mu}$, the image prediction.

Convolutional Latent Neural Processes: Foong et al. (2020) presents the Convolutional Neural Processes (ConvNPs, in this paper referred to as Convolutional Latent Neural process or ConvLNP) that utilise a latent variable to capture information from the context set. It is similar in architecture to the conditional neural process with the encoder-decoder architecture, but in the ConvLNP the encoder outputs a distribution over the latent variable $\mathbf{z}$ with the SetConv representations: $\mathbf{z} \sim p(\mathbf{z}|R)$. This enables ConvCNPs to learn 'richer joint predictive distributions' (Foong et al., 2020) and handle multimodalities. ConvLNP can be straighforwardly implemented on top of ConvCNPs by using the SetConv representations to parameterise the Gaussian latent variable distribution, and then decode the latent variable samples using a CNN. The full computational graphs for ConvCNP and ConvLNP are described in Figure 3.

Training objective: Following Foong et al. (2020), we use the maximum likelihood training objective for the ConvNPs: $\mathcal{L} = \log p(y_T|x_T, C)$, which measures the predictive performance of the models. For the log-likelihood function, $\log p(y_T|x_T, C)$, we instead use the MS-SSIM metric (Multi Scale Structural Similarity, Wang et al. (2003)) between the mean predictions and ground truth images. MS-SSIM is a structural similarity metric for images, and is widely used in the field of signal processing, having shown empirically to increase sharpness of final prediction images. In the ConvLNP training objective, the maximum likelihood approach uses sample estimates to approximate the likelihood of the predictions: $\mathcal{L} \approx \frac{1}{L} \sum_{l=1}^{L} \log p(y_T|\mathbf{z}, x_T, C)$.

5   Experiments

We study the performance of ConvCNPs and ConvLNPs for the task of inpainting LANDSAT 7 scanlines. We compare the results to the baseline models consisting of Navier-Stokes (NS) inpainting algorithm, U-Net and PartialConv, for which the latter two are vastly popular and yield state-of-the-art results on a variety of image inpainting problems. We train the ConvCNP, ConvLNP and U-Net using the MS-SSIM loss function, and for PartialConv we use the loss proposed in Liu et al. (2018). We conduct extensive experiments to measure the in-distribution performance of each model by inpainting satellite images of the same country, as well as OOD performance on a set of different countries (zero-shot prediction over unseen spatial locations). The inpainted results are also evaluated on a synthetic downstream regression task. The extracted scanline masks are randomly applied to the datasets of pre-2003 images so that both the clean and corrupted images are available during training. For ConvNP models, we treat the pixels outside of the scanline mask as the context set. To evaluate the performance of all the models, we compute the MS-SSIM between the predictions and the ground truth images. For ConvCNP, U-Net and PartialConv, the prediction is the forward pass of the networks. For NS, the prediction is the gap-filled image. For ConvLNP, we take the mean of the
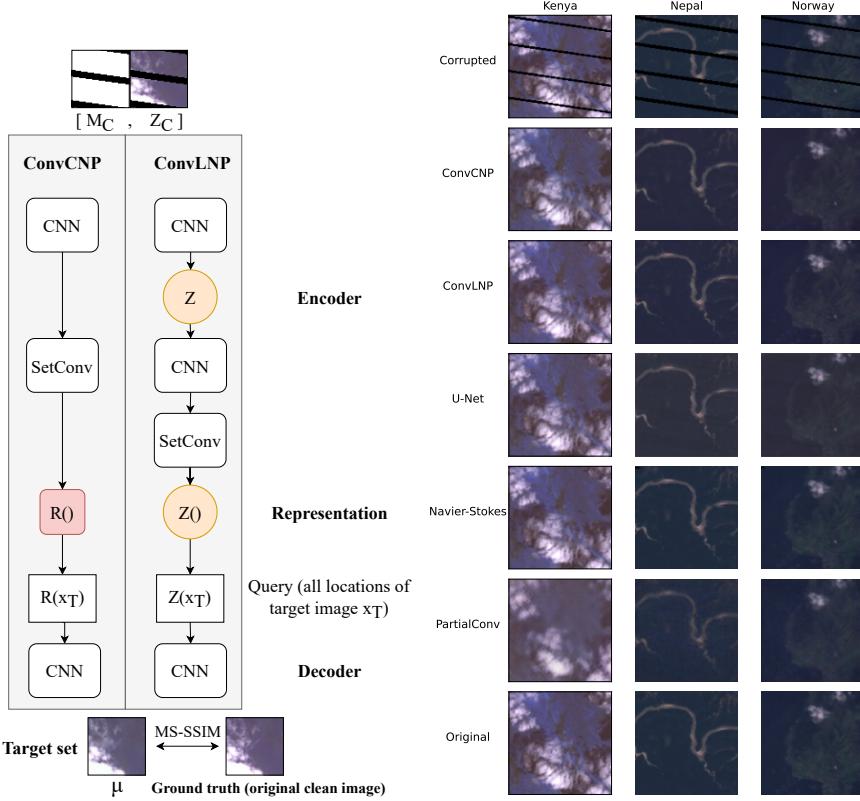
Figure 3: (Left) ConvCNP and ConvLNP on-the-grid architecture. (Right) Inpainting predictions for all models on 128x128 images (all region predictions in Figure 7).

model outputs over many samples from the latent variable distribution. The MS-SSIM is bounded in $[0, 1]$, where values closer to 1 show that the images are more similar.

5.1 Inpainting LANDSAT 7 Images

5.1.1 Data

The training images are acquired from the LANDSAT 7 Satellite before the scanlines are present in the images. The training set consists of 1000 images from Kenya with dimensions 128x128 and 64x64. The scanlines are acquired from 100 Kenya images post-SLC failure. We perform 5-fold cross validation with a 80%-20% train-test ratio for each split. During training, a scanline is applied to each image as a mask chosen randomly within the 100 scanlines extracted. To test the model's performance on the in-distribution test set, each model is tested on the respective Kenya test set of that split. To test the models' zero-shot capabilities, 1000 images of UK, Nepal, Brazil and Norway are collected, for which the model had not seen during training. For each location, a clean image and a "corrupted" image is created, where the corrupted image is created by applying one of the randomly chosen extracted scanlines as a mask to a clean image.

5.1.2 ConvNP training:

Our implementation follows Dubois et al. (2020). Both ConvCNPs and ConvLNPs use Resnet blocks in the encoder and linear MLPs in the decoder. The ConvCNP has a 10-layer ResNet encoder with a representation size of 128 channels and the decoder MLP has 4 layers. It is trained for 400 epochs, with batch size 8 and learning rate 10^{-4}, which decays exponentially by a factor of 5. The ConvLNP model for 128x128 images is trained for 200 epochs with a batch size of 4 (low batch size due to computational limitations) and during training, 4 samples are obtained of the latent variable while during evaluation, 8 samples are used. For 64x64, the latent samples are increased for ConvLNP,
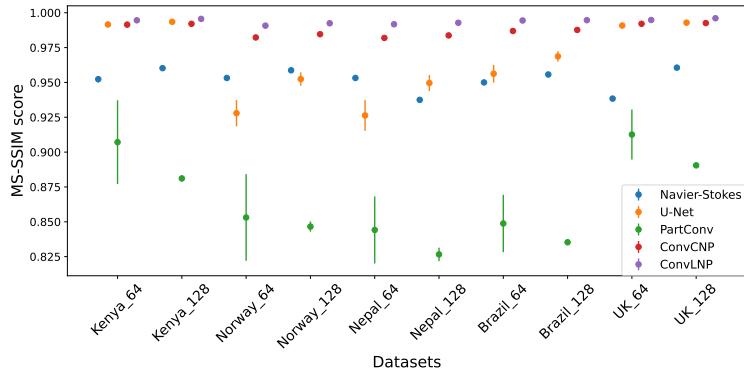
Figure 4: Mean and standard error of the MS-SSIM scores over 5-fold cross validation for predicting over Kenya and OOD datasets. Note that standard errors lower than 0.01 have not been visualised. namely 16 latent samples are used for training and 32 are used during inference. Both Resnets used in ConvLNP have 8 layers.

5.1.3 Results

We first examine the MS-SSIM scores of the Neural Process models and baselines in the in-distribution setting - trained on images of Kenya and inference in a held out test-set of Kenya images. The MS-SSIM score is calculated for 200 test images by comparing the clean image and inpainted image for all the algorithms. The results are seen in Figure 4, where average results across 5 splits are reported. One caveat is that we only ran PartialConv for 3 splits, due to a software issue. The HI-VAE model is also tested, but is not able to produce consistent images in our small dataset and low training time setting and hence has been omitted from the analysis.

As can be seen both empirically and by the MS-SSIM scores in Figure 4, ConvNPs and U-Net achieve very good results for the Kenya test dataset. PartialConv is not able to output as good inpainting results and there is a noticeable difference between original and inpainted images in terms of image sharpness and quality of detail. The model seems to be able to successfully remove the scanlines but the resulting images are blurry compared to the original image which leads to lower MS-SSIM scores. One of the reasons could be that the model requires a longer training time; in the original paper Liu et al. (2018) the model is trained for significantly longer with a significantly larger dataset. Navier-Stokes fails in particular scenarios where the target pixels are at the border between two differently coloured regions as the predicted pixel values are an average over the two regions, resulting in noticeably wrong estimations. However, ConvNPs and U-Net produce outputs with high MS-SSIM scores. The predictions are good in terms of consistency and sharpness, and scanlines are not visible.

For the second experiment, the zero-shot or OOD capabilities of the inpainting models are examined. The models trained on the 800-image Kenya test split are evaluated on different countries including UK, Nepal, Brazil and Norway, where the model has not seen any of these locations during training. The MS-SSIM score between predictions and original images is again calculated and averaged across the 5 splits. The Navier-Stokes algorithm here cannot be considered as solving a zero-shot task as it does not involve training, but the results are shown for comparison. The average MS-SSIM scores can again be seen in Figure 4.

In the OOD setting, the ConvNP models outperform all the baseline models. For both U-Net and PartialConv, the MS-SSIM scores are considerably lower and scanlines are quite visible in the predictions compared to the in-distribution results. The ConvNP models, however, generalise quite well for the OOD task and predictions have high MS-SSIM scores. The scanlines are fairly well inpainted in comparison to the original pixel values. U-Net performs surprisingly well for UK images, however this is related to the high cloud presence in UK images similarly to Kenya images. The difference in generalisation performance is due to the architecture of classical deep learning networks, which are treating all images for inpainting as a single task and are inherently approximating a single function from the input to the output. However, the image locations, times and scanline locations are different, and this approach does not model the differences between satellite images very well, which results in lower generalisation performance of the models. However, the meta-learning
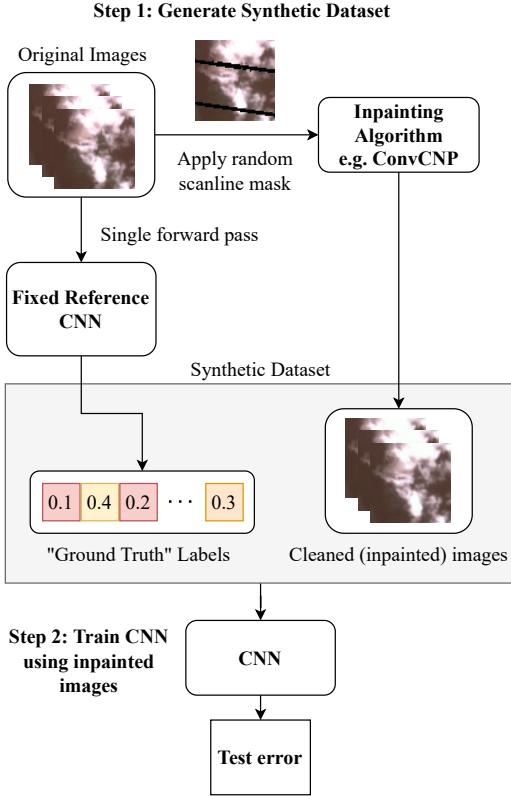
7

Figure 5: Synthetic downstream task workflow.

approach treats the input images as different tasks and hence the variability between images and their characteristics is better accounted. Therefore, we see an improvement in quality of predictions for new and unseen images. One considerable advantage, however, of U-Net is its speed and training stability. The training is considerably faster for U-Net. For PartialConv, training speed is comparable to the ConvNP training times and the results are considerably worse, as it converges a lot slower than ConvNPs or U-Net.

5.2 Synthetic downstream task

5.2.1 Data

Inpainted results from each model are evaluated on a synthetic downstream task (Figure 5). A synthetic dataset is formed by passing non-corrupted images X through a small randomly initialised CNN model, f, and scaling the images by a factor of $a = 10$ to create the ground truth output for each image, $f(a * X)$. This scaling is done to make the downstream task more sensitive to different values within the scanlines. Another CNN, g, with the same architecture but a different random initialisation is then trained using the inpainted images $\tilde{X}$ to evaluate how well it can perform in a regression task as a substitute to the clean image. The inpainted images are created by adding a random scanline and inpainting it using the models discussed. This generates a synthetic training set: $(\tilde{X}, f(a * X))$. As a reference, the training performance on the original images X and on the images with the scanline applied is also evaluated.

5.2.2 Training

All images are normalised to the range [0,1]. The CNN architecture includes two convolutional layers with kernel size 3. The output is then fed through a fully connected layer to produce a single scalar output. Training used the MSE loss and a batch size of 8. Training is done with 5-fold cross-validation and each run lasted for 300 epochs. We used a learning rate of 0.001 with a reduction of a factor
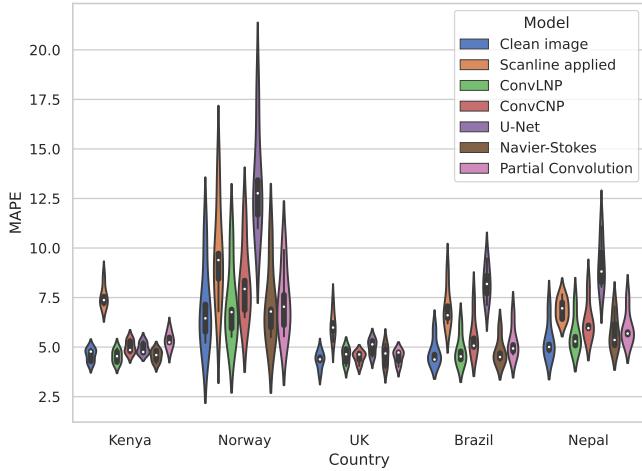
Figure 6: MAPE test errors over 5-fold cross-validation on downstream regression tasks over Kenya and OOD datasets with image dimension 64x64.

of 0.1 if it plateaus with a patience of 3 epochs. Early stopping is also used with patience 8 and threshold 0.0001.

5.2.3 Results

The results for 64x64 satellite image predictions on the downstream task are presented in Figure 6. The predictions for 128x128 dimensional images are unstable with the current training settings, so are omitted from the report. As expected, the clean image baseline results show the best performance as they are used to generate the synthetic dataset. For Kenya (in-distribution), ConvLNP outperform the others, performing very similarly to the clean images in the downstream task. As expected, the scanline results have the worst performance as they are not inpaint. U-Net performs similarly to ConvCNP, but for out-of-distribution tasks, U-Net performs the worst out of all models, even performing worse than running the downstream task with the scanline applied images. As U-Net only approximates a single function during inpainting, it fails at generalising to these different tasks.

Overall, ConvLNP achieves among the best MAPE errors in both in-distribution and out-of-distribution tasks. ConvCNP has slightly higher test MAPE. The ConvLNP can learn more complex data distributions than ConvCNPs by using latent variables, allowing for improved prediction results more similar to the clean image. Navier-Stokes does get very similar MAPE, although this is likely due to most of the pixels being the exact same as the clean image. Given only the scanlines have changed, this is not an appropriate measure of the quality of inpainting predictions made using this algorithm. Surprisingly, Partial Convolution test MAPE consistently performs slightly better than ConvCNP, despite inpainting results being blurry and of worse quality. Norway satellite images seem to also be the most challenging to learn given this approach. This can be seen by the large error even with the clean image baseline.

6 Limitations

There are several improvements to our approach that could be done in future work. Firstly, the downstream regression task is synthetic and does not reflect the performance of imputed satellite data in real-life tasks. A more suitable approach is to evaluate predictions on an epidemiology downstream regression task such as Malaria prevalence mapping. This could involve inpainting LANDSAT 7 maps inside regions of interest to predict Malaria cases, and using models such as DeepSets (Zaheer et al., 2017), Set Transformers (Lee et al., 2018), and distribution regression methods (Zhu et al., 2021).

9

Another limitation is that each original satellite image, of around 6000×6000 pixels wide, has scanlines that are increasing in thickness. Our training set only uses relatively thinner scanlines after data processing, resulting in poor performance on thicker scanlines, so further exploration of performance of ConvNPs when trained on larger scanlines could be done, such as augmenting the set of training scanlines so that we include thicker ones. Finally, this work can also be readily extended to the task of cloud removal.

7 Conclusion and Discussion

We find that ConvNPs are successful at inpainting LANDSAT 7 satellite images corrupted by scanlines in both in-distribution and out-of-distribution tasks, outperforming classic and state-of-the-art inpainting methods. We also observe that ConvLNPs perform the best out of these models in a synthetic downstream regression task. These ConvNP models are able to take advantage of the spatiotemporal nature of satellite images to understand the underlying structure of the data. The direct consequence of this is that with ConvNPs, we may be able to obtain a **global inpainter** for LANDSAT 7, by only training on a **small subset** of spatiotemporal locations, which is computationally tractable compared to training U-Net or PartialConv over training images taken from all over Earth. Future work could involve making use of recent advances in ConvNPs to improve expressiveness (Bruinsma et al., 2020; Markou et al., 2021, 2022) and to more explicitly account for space-time (Singh et al., 2019).

8 Acknowledgements and Funding Disclosure

References

Bertalmio, M., Bertozzi, A. L., and Sapiro, G. Navier-stokes, fluid dynamics, and image and video inpainting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pp. I–I. IEEE, 2001.

Bertozzi, A. and Schönlieb, C.-B. Unconditionally stable schemes for higher order inpainting. *Communications in Mathematical Sciences*, 9(2):413–457, 2011.

Bruinsma, W. P., Requeima, J., Foong, A. Y., Gordon, J., and Turner, R. E. The gaussian neural process. *3rd Symposium on Advances in Approximate Bayesian Inference*, 2020.

Burger, M., He, L., and Schönlieb, C.-B. Cahn–hilliard inpainting and a generalization for grayvalue images. *SIAM Journal on Imaging Sciences*, 2(4):1129–1167, 2009.

Demir, U. and Ünal, G. B. Patch-based image inpainting with generative adversarial networks. *CoRR*, abs/1803.07422, 2018. URL http://arxiv.org/abs/1803.07422.

Dubois, Y., Gordon, J., and Foong, A. Y. Neural process family. http://yanndubs.github.io/Neural-Process-Family/, September 2020.

Dupont, E., Teh, Y. W., and Doucet, A. Generative models as distributions of functions. *NeurIPS*, 2021.

Dupont, E., Kim, H., Eslami, S. M. A., Rezende, D. J., and Rosenbaum, D. From data to functa: Your data point is a function and you should treat it like one. *CoRR*, abs/2201.12204, 2022. URL https://arxiv.org/abs/2201.12204.

Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017.

Foong, A., Bruinsma, W., Gordon, J., Dubois, Y., Requeima, J., and Turner, R. Meta-learning stationary stochastic process prediction with convolutional neural processes. *Advances in Neural Information Processing Systems*, 33:8284–8295, 2020.

Garnelo, M., Rosenbaum, D., Maddison, C., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y. W., Rezende, D., and Eslami, S. A. Conditional neural processes. In *International Conference on Machine Learning*, pp. 1704–1713. PMLR, 2018a.

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S., and Teh, Y. W. Neural processes. *ICML Workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018b.

Gordon, J., Bruinsma, W. P., Foong, A. Y., Requeima, J., Dubois, Y., and Turner, R. E. Convolutional conditional neural processes. *ICLR*, 2020.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017. doi: 10.1016/j.rse.2017.06.031. URL https://doi.org/10.1016/j.rse.2017.06.031.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. Globally and locally consistent image completion. *ACM Trans. Graph.*, 36(4), jul 2017. ISSN 0730-0301. doi: 10.1145/3072959.3073659. URL https://doi.org/10.1145/3072959.3073659.

Lee, J., Lee, Y., Kim, J., Kosiorek, A. R., Choi, S., and Teh, Y. W. Set transformer: A framework for attention-based permutation-invariant neural networks, 2018. URL https://arxiv.org/abs/1810.00825.

Liu, G., Reda, F. A., Shih, K. J., Wang, T., Tao, A., and Catanzaro, B. Image inpainting for irregular holes using partial convolutions. *CoRR*, abs/1804.07723, 2018. URL http://arxiv.org/abs/1804.07723.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L. Repaint: Inpainting using denoising diffusion probabilistic models. *arXiv preprint arXiv:2201.09865*, 2022.

Markou, S., Requeima, J., Bruinsma, W., and Turner, R. Efficient gaussian neural processes for regression. *ICML 2021 Workshop on Uncertainty and Robust- ness in Deep Learning*, 2021.

Markou, S., Requeima, J., Bruinsma, W. P., Vaughan, A., and Turner, R. E. Practical conditional neural processes via tractable dependent predictions. *ICLR*, 2022.

Nazabal, A., Olmos, P. M., Ghahramani, Z., and Valera, I. Handling incomplete heterogeneous data using vaes. *Pattern Recognition*, 107:107501, 2020.

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., and Efros, A. A. Context encoders: Feature learning by inpainting. *CoRR*, abs/1604.07379, 2016. URL http://arxiv.org/abs/1604.07379.

Richard, M. and Chang, M. Y.-S. Fast digital image inpainting. In *Appeared in the Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001), Marbella, Spain*, pp. 106–107, 2001.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Scaramuzza, P. and Barsi, J. Landsat 7 scan line corrector-off gap-filled product development. In *Proceeding of Pecora*, volume 16, pp. 23–27, 2005.

Shroff, M. and Bombaywala, M. S. R. A qualitative study of exemplar based image inpainting. *SN Applied Sciences*, 1(12):1730, Nov 2019. ISSN 2523-3971. doi: 10.1007/s42452-019-1775-7. URL https://doi.org/10.1007/s42452-019-1775-7.

Singh, G., Yoon, J., Son, Y., and Ahn, S. Sequential neural processes. *Advances in Neural Information Processing Systems*, 32, 2019.

Telea, A. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004.

Thrun, S. and Pratt, L. *Learning to learn.* Springer Science & Business Media, 2012.

USGS. Landsat 7 courtesy of the u.s. geological survey.

Wang, Z., Simoncelli, E. P., and Bovik, A. C. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pp. 1398–1402. Ieee, 2003.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Weiss, D. J., Lucas, T. C., Nguyen, M., Nandi, A. K., Bisanzio, D., Battle, K. E., Cameron, E., Twohig, K. A., Pfeffer, D. A., Rozier, J. A., et al. Mapping the global prevalence, incidence, and mortality of plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *The Lancet*, 394(10195):322–331, 2019.

Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., and Li, H. High-resolution image inpainting using multi-scale neural patch synthesis. *CoRR*, abs/1611.09969, 2016. URL `http://arxiv.org/abs/1611.09969`.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., and Burke, M. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, 11(1):1–11, 2020.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., and Smola, A. Deep sets, 2017. URL `https://arxiv.org/abs/1703.06114`.

Zhu, H., Howes, A., van Eer, O., Rischard, M., Li, Y., Sejdinovic, D., and Flaxman, S. Aggregated gaussian processes with multiresolution earth observation covariates, 2021. URL `https://arxiv.org/abs/2105.01460`.

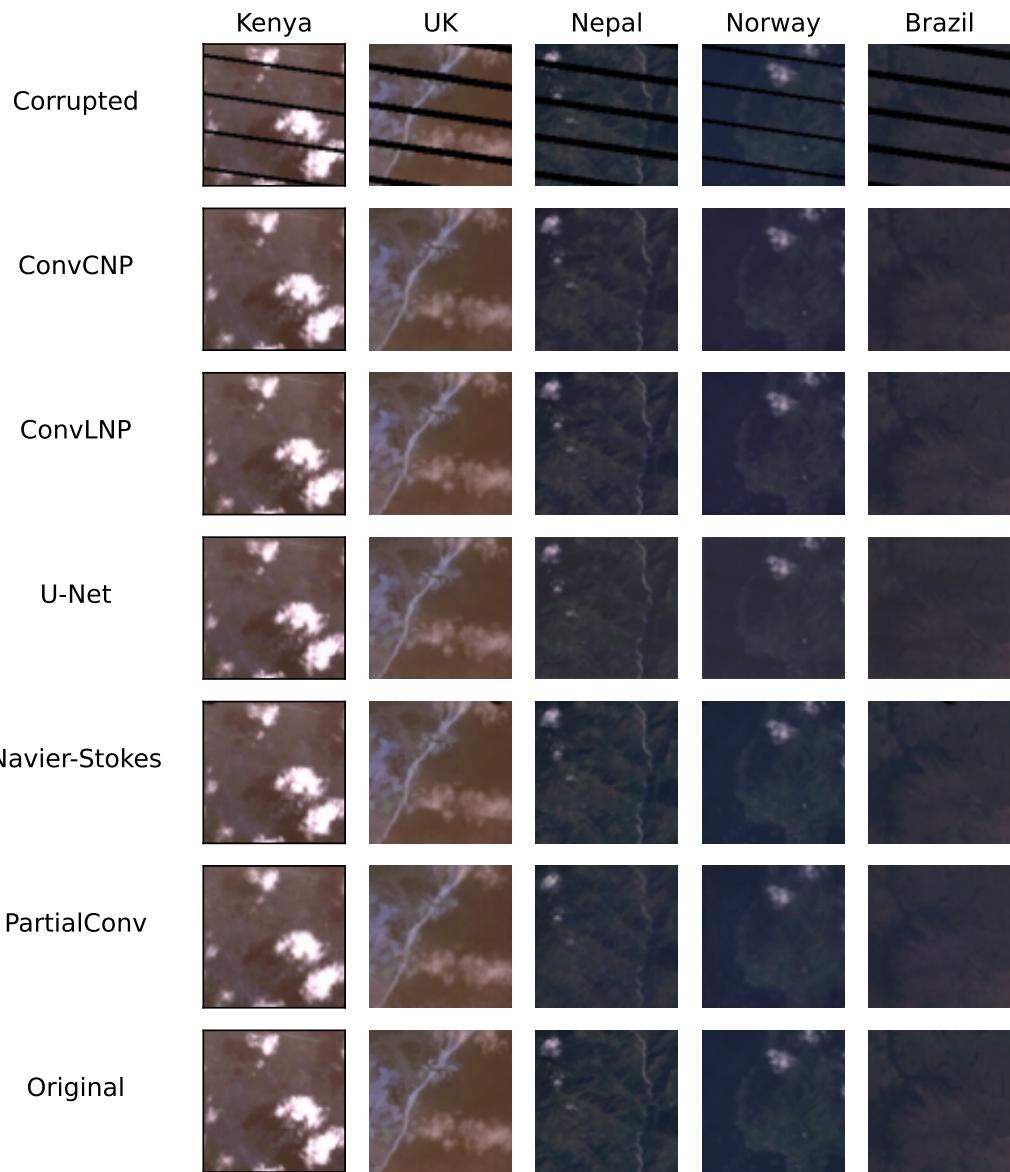# A Additional Experimental Results

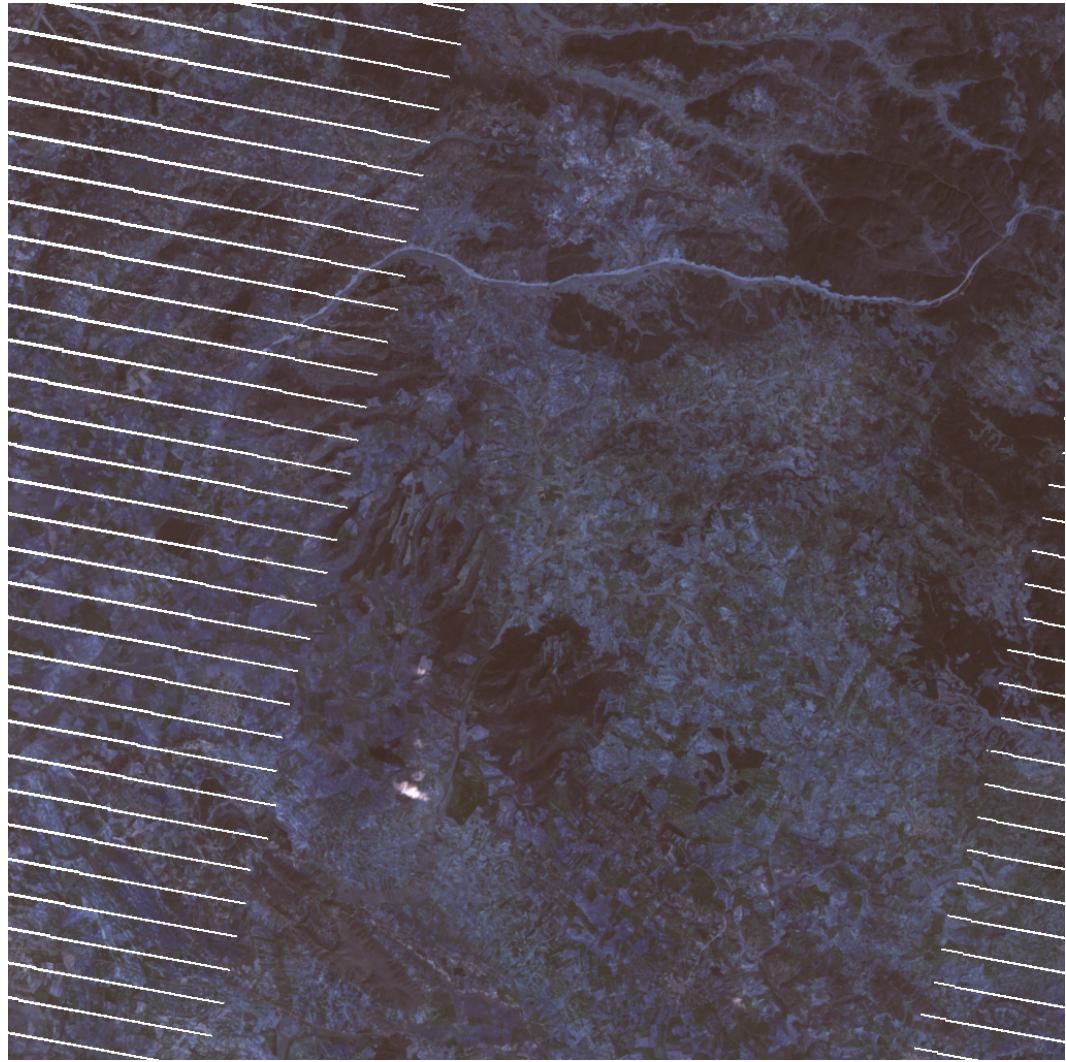Figure 7: Inpainting results on 128x128 images for all models over multiple regions.

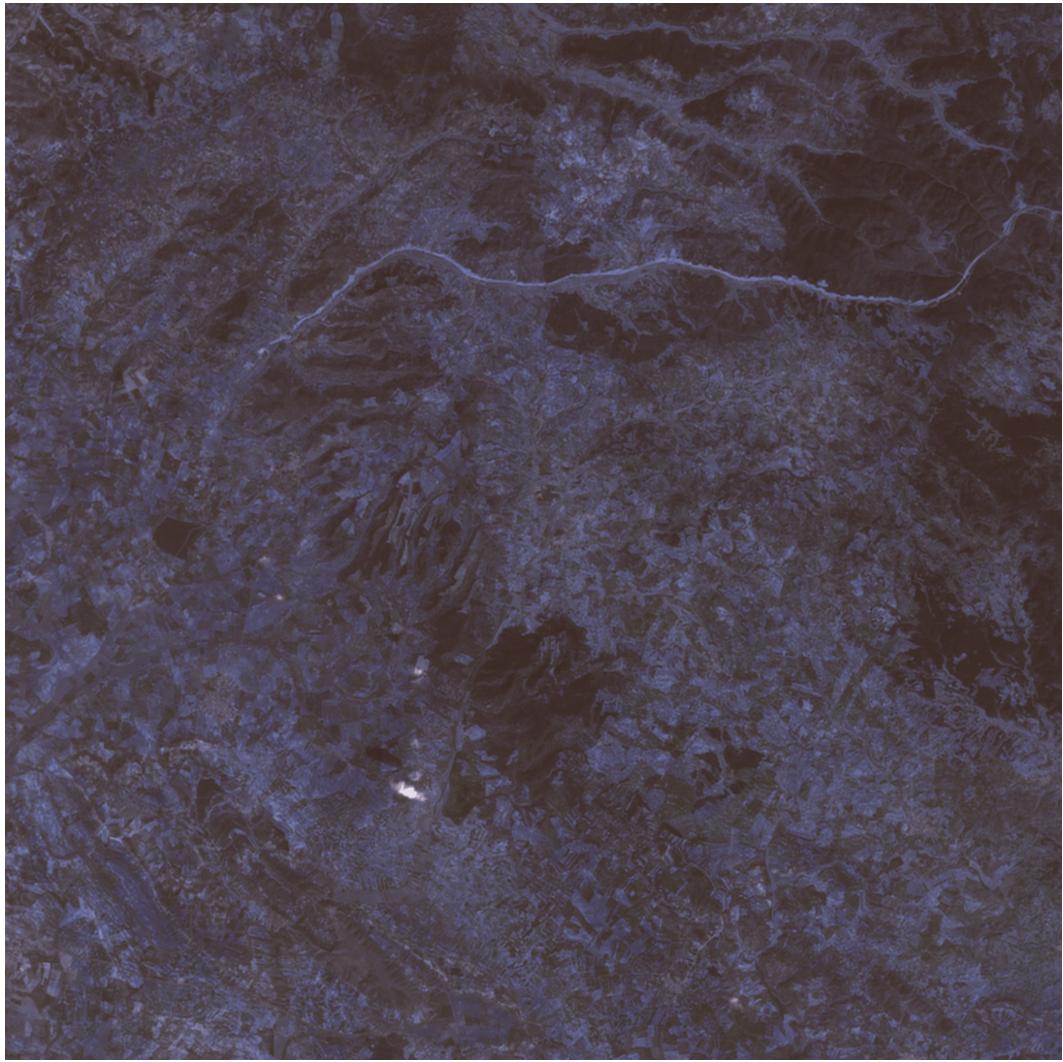Figure 8: Corrupted 1024x1024 Kenya image used for inpainting of figures in this section.

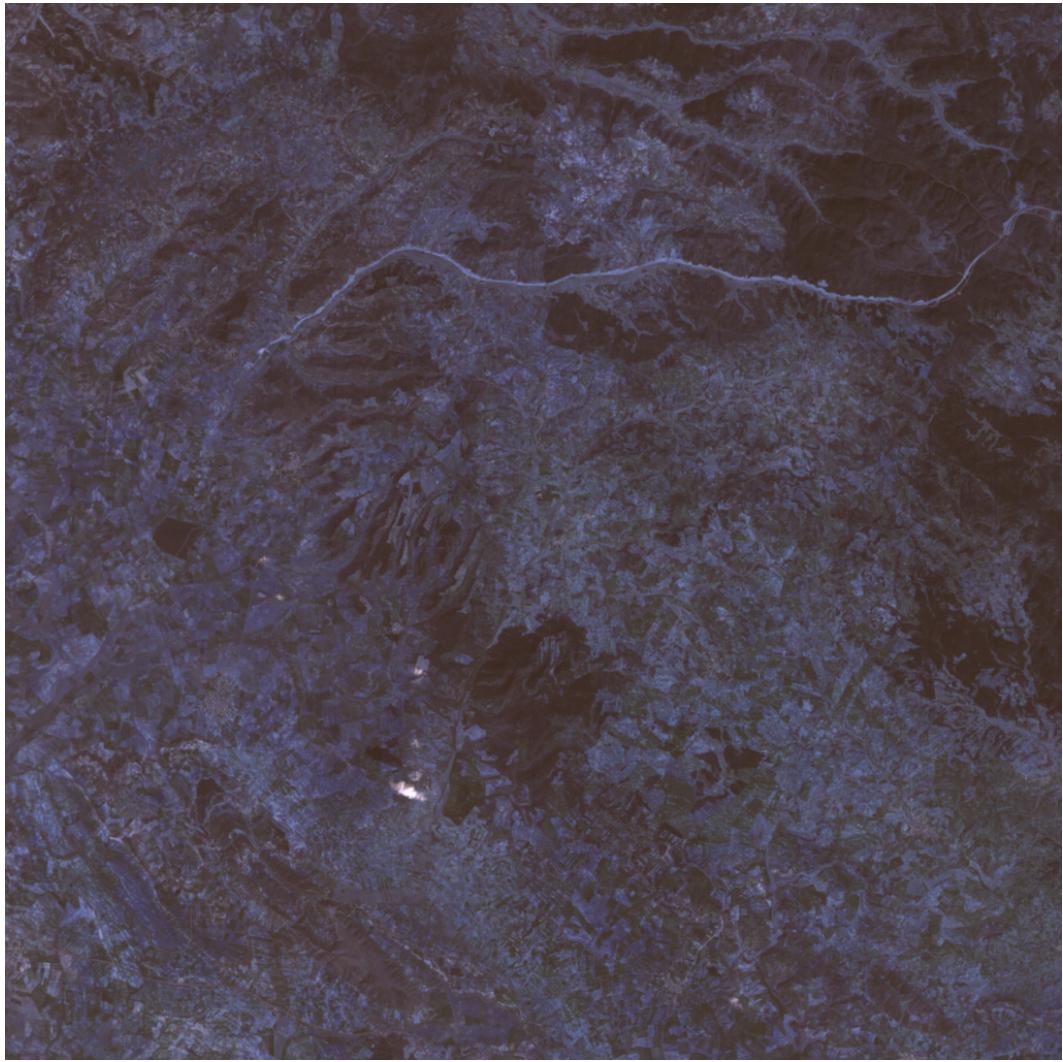Figure 9: ConvCNP inpainting of 1024x1024 Kenya image using 64x64 patches.

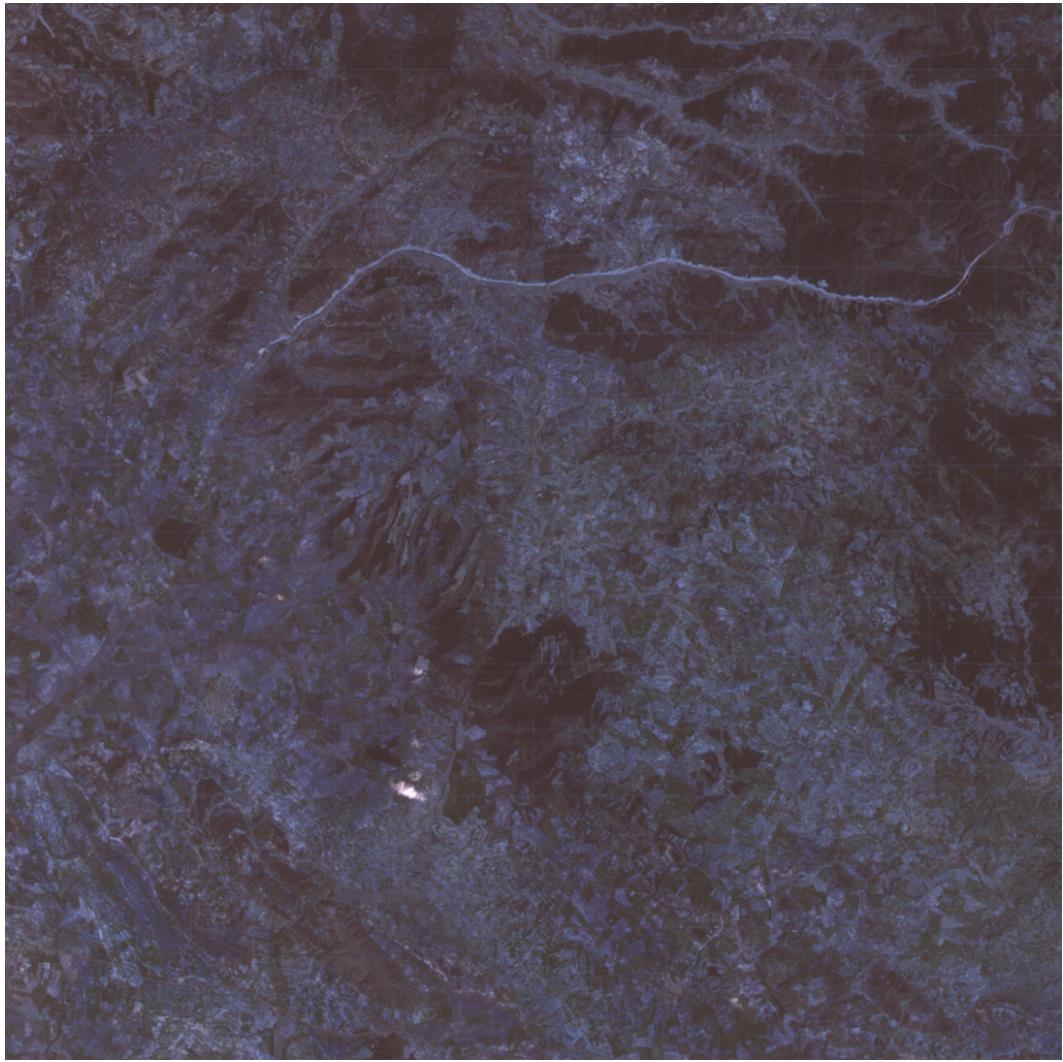Figure 10: ConvLNP inpainting of 1024x1024 Kenya image using 64x64 patches.

Figure 11: U-Net inpainting of 1024x1024 Kenya image using 64x64 patches.
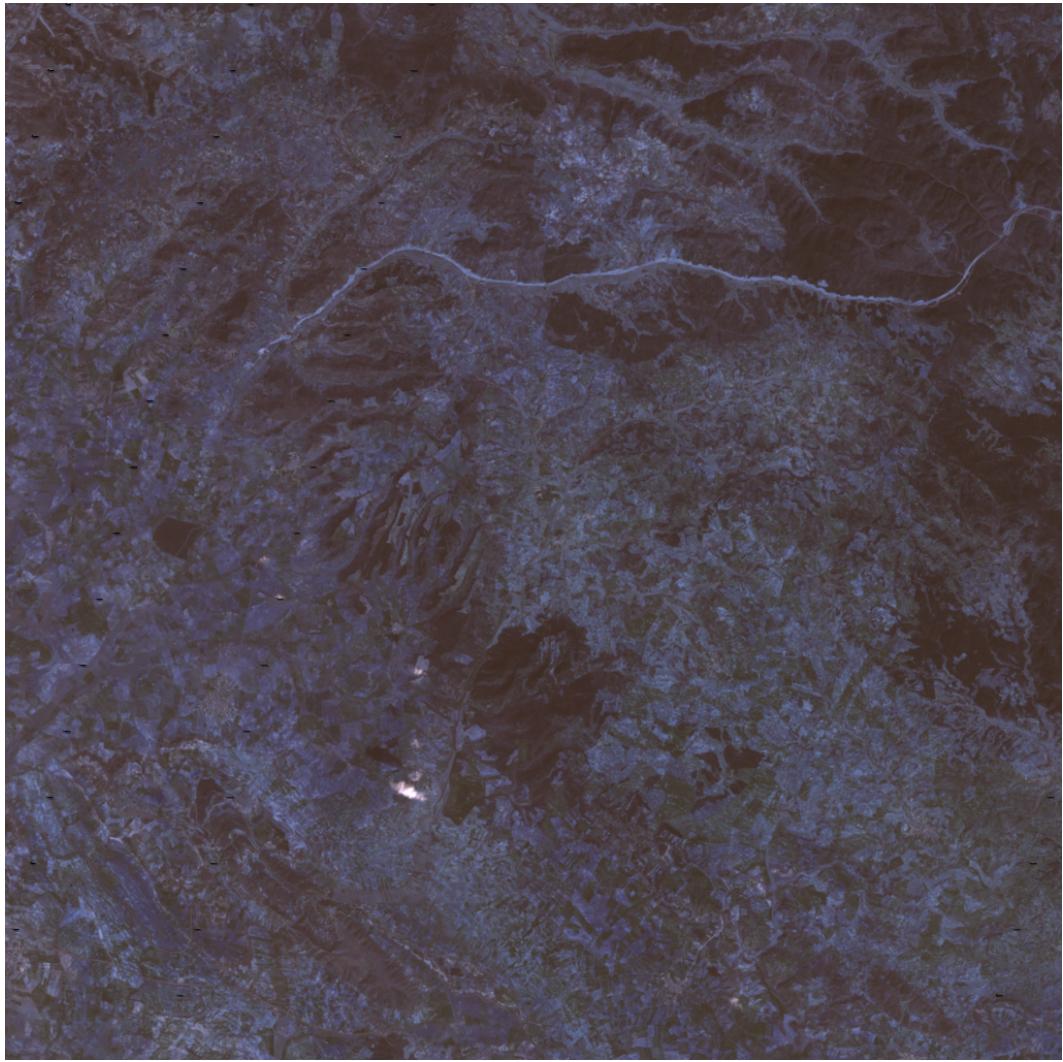
Figure 12: Navier-Stokes inpainting of 1024x1024 Kenya image using 64x64 patches.
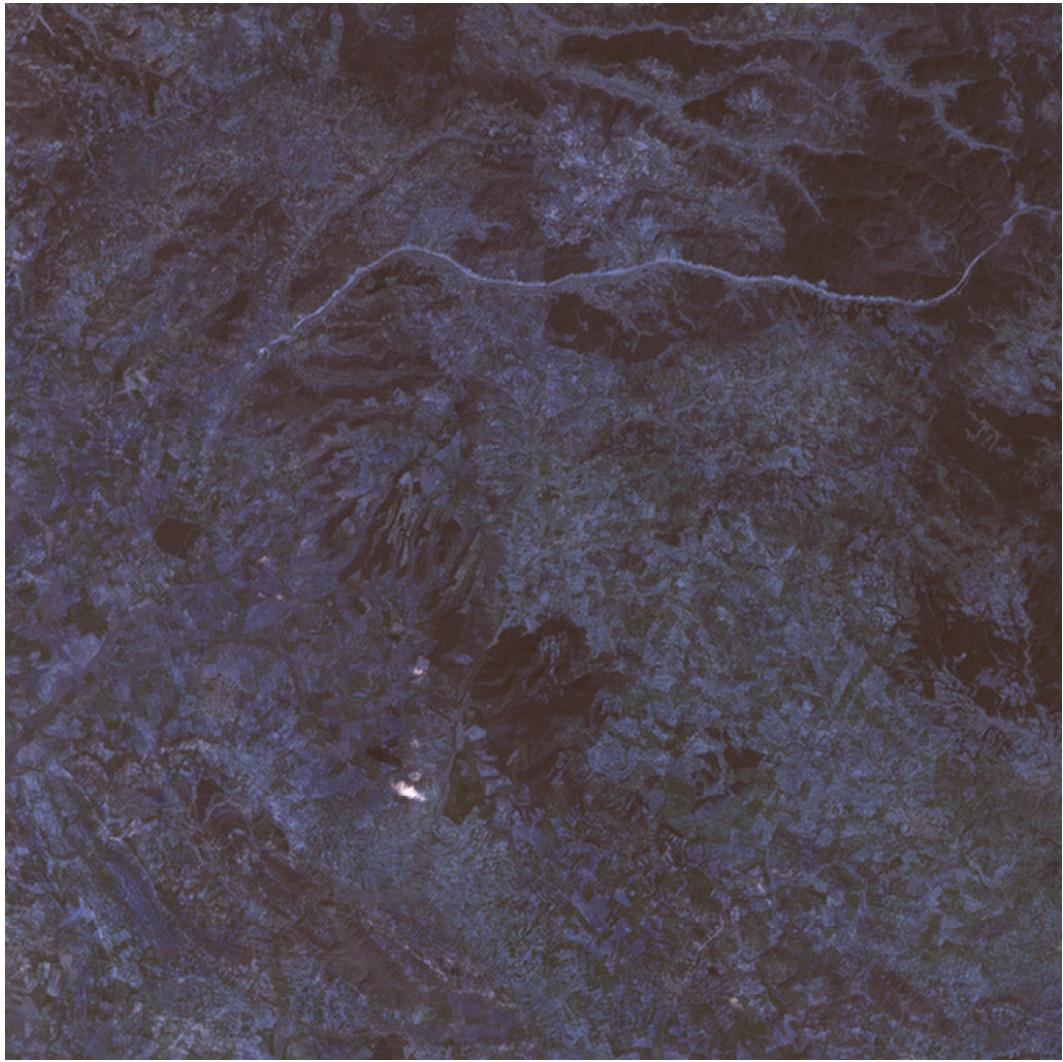
Figure 13: PartialConv inpainting of 1024x1024 Kenya image using 64x64 patches.