Spatiotemporal estimation of TROPOMI NO2 column with depthwise partial convolutional neural network

Yannic Lops1 · Masoud Ghahtemanloo1 · Arman Pousaei1 · Yunsoo Choi1 · Jia Jung1 · Seyedali Mousavinezhad1 · Ahmed Khan Salman1 · Davyda Hammond2

Received: 22 November 2021 / Accepted: 29 March 2023 / Published online: 17 April 2023© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract

Satellite-derived measurements are negatively impacted by cloud cover and surface reflectivity. These biases must be discarded and significantly increase the amount of missing data within remote sensing images. This paper expands the application of a partial convolutional neural network (PCNN) to incorporate depthwise convolution layers, conferring temporal dimensionality to the imputation process. The addition of a temporal dimension to the imputation process adds a state of successive existence within the dataset, which spatial imputation cannot capture. The depthwise convolution process enables the PCNN to independently convolve the data for each channel. The deep learning system is trained with the Community Multiscale Air Quality model-simulated tropospheric column density of Nitrogen Dioxide (TCDNO2) to impute TROPOspheric Monitoring Instrument TCDNO2. The depthwise PCNN model achieves an index of agreement of 0.88 and outperforms the default PCNN models, with and without temporal dimensionality of data, and conventional data imputation methods such as inverse distance weighting by 4–7 and 10–15% in the index of agreement and correlation, respectively. The model demonstrates more consistency in the reconstruction of TROPOspheric Monitoring Instrument tropospheric column density of NO2 images. The model has also demonstrated the accurate imputation of remote sensing images with over 95% of the data missing. PCNN enables the accurate imputation of remote sensing data with large regions of missing data and will benefit future researchers conducting data assimilation for numerical models, emission studies, and human health impact analyses from air pollution.

Keywords Partial convolution · Depthwise · TROPOMI · Kriging · Spatiotemporal imputation · CMAQ

1 Introduction

Nitrogen oxides (NOX = NO + NO2) are some of the major pollutants [1] resulting from human activity [2]. NOX sources include anthropogenic and natural origins, such as the combustion of fossil fuels [3, 4], burning of biomass [5], soil microbial activity [6], and lightning [3]. Nitrogen dioxide (NO2) has been associated with adverse negative health conditions such as cardiovascular diseases [7] and respiratory-related ailments [8, 9].

Remote sensing measures the characteristics of an area by utilizing reflected and emitted radiation at a distance. Satellite remote sensing instruments have contributed essential data pertaining to the global distribution [10–12], evolution [13], and the transport of atmospheric pollutants [14–16]. Unfortunately, remote sensing has limitations such as low spatial and temporal resolutions [17] and measurement issues caused by the impact of cloud cover contamination, false reflectance, and significant bias within the data [18, 19]. Furthermore, the system can also experience sensor errors that corrupt or lead to failed data measurements [20, 21], thus limiting the comprehensive application of remote sensing data for forecasting and data assimilation techniques for chemical transport models [2, 4, 22, 23].

To non-temporally impute missing data within remote sensing images, studies have applied several methods such as geostatistical approaches [24, 25], linear regression
models [20], inpainting algorithms [26, 27], and deep learning algorithms [28–30]. Deep learning algorithms [31] have shown significant promise in addressing the limitations of missing data, for they model high-level abstractions within datasets [32, 33]. Among the various deep learning algorithms, convolutional neural networks (CNNs) [34] have been among the most successful and widely used approaches [35, 36] for various purposes such as forecasting [37–40], classification [41, 42], speech recognition [43, 44], and imputation [29–45]. Nevertheless, a number of models and methods still have difficulty in imputing remote sensing data that are missing a significant percentage of the data or contain large gaps within datasets [45, 46]. Advanced methods of imputation use temporal dimensionality to enhance the accuracy of the imputation process [46, 47].

Although convolution models have been used to impute missing remote sensing data with a temporal dimension within the dataset [45], they require data with a low frequency of missing pixels. This paper expands the application of a partial convolutional neural network (PCNN) [48] to imputing missing remote sensing data [30] by adding temporal dimensionality within the model. The PCNN model performs well at imputing images with a significant amount of missing data and spatial distances, and its performance is further enhanced by the addition of the temporal dimension of the model input. The temporal component of the PCNN is applied through the implementation of depthwise convolutions [49], in which the convolution process is independently performed for each channel. The depthwise partial convolutional neural network (DW-PCNN) aims to address the limitations of the regular PCNN model [30] by incorporating the temporal component for imputation, improving the sharpness of the imputed image, and enhancing the accuracy over that of the regular PCNN.

2 Methods

2.1 Data preparation

The TROPOspheric Monitoring Instrument (TROPOMI) is a key instrument aboard the Copernicus Sentinel-5 Pre-cursor (SSP) satellite. The instrument obtains data of key atmospheric constituents such as ozone (O₃), NO₂, formaldehyde (CH₂O), and aerosol through ten spectra bands of ultraviolet (UV), visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) [50]. The system is a near-polar, sun-synchronous orbit that provides daily global coverage at high spatial resolution (7 x 3.5 km at the nadir) [51, 52]. We performed an initial data filtering process to exclude pixels failing the initial quality assurance (QA) value threshold (i.e., QA < 0.5) that presents an error flag or solar zenith angle exceeding 70°, cloud cover, and air mass factor below 0.1 [53]. In addition, we filtered each image to exclude isolated pixel clusters of four or fewer pixels within a defined filter grid to ensure the exclusion of outliers.

We implemented the United States Environmental Protection Agency (US EPA) Community Multiscale Air Quality (CMAQ v5.2) model [54]. The domain of the model has a 12-km grid horizontal spacing with 27 vertical layers reaching 100 hPa, which we used to estimate and predict the tropospheric column density of NO₂ (TCDNO₂) over the contiguous United States (CONUS). The system uses CB6 and AERO6 chemical mechanisms for the gas-phase and aerosol chemical processes. We also used the 2017 U.S. EPA National Emission Inventory (NEI) [55, 56] with parameterized lightning-induced emissions, biogenic emissions computed by using the Biogenic Emission Inventory System, and biomass burning emissions used by the Fire Inventory from the National Center for Atmospheric Research model, version 1.5 [57–59]. The CMAQ model received simulated meteorological variables from the Weather Research and Forecasting model version 4.0 from the National Centers for Environmental Prediction. For initial and boundary conditions, the North Americans Regional Reanalysis (NARR) data were utilized. Furthermore, we incorporated indirect soil moisture and the temperature nudging technique [60, 61], as well as a four-dimensional data assimilation option for the temperature, the water vapor mixing ratio, and wind components [62] to enhance the model performance in simulating meteorological fields and performed simulations for the months of February-June, 2019, and 2020.

The simulated NO₂ column by CMAQ acts as the basis for preparing training data for the PCNN model. This process ensures that the training data contain images without any missing data. The pixel size of the original CMAQ output images was 299 x 459, and we extracted ten images with a resolution of 256 x 256. To ensure the availability of enough training samples for the partial CNN model, we applied several image augmentation processes [63]. The first image augmentation step of the process was to apply a random noise function with Gaussian smoothing to the CMAQ images to replicate the pixel variations observed in the TROPOMI images. The second step was to apply basic image augmentation to the updated CMAQ image through the random selection of rotation, flipping, or combination. The final augmentation phase involved the random selection, rotation, flipping, and application of TROPOMI masks to the updated CMAQ images. The augmentation phase ensured that the partial convolution had enough training data to become robust at imputing missing data in various images and to ensure that the...
system did not produce extreme variations or outliers within the missing data imputation process.

The domains of TROPOMI and CMAQ within the CONUS are shown in Fig. 1.

### 2.2 DW-PCNN structure

We used a deep convolutional neural network (deep CNN) [34] based on U-net architecture [64] that replaces conventional convolutional layers with partial convolution layers [48]. CNNs process data by convolving the data at each layer over multiple image channels, assigning weights and biases at various aspects within the data, and differentiating between them. The benefits of CNN models are their capability to reduce data into a more manageable form for processing (without losing key features) and extract high-level features of the input data through the use of kernels during the convolving phase [65].

Single-channel images are defined as grayscale images made of one of the primary colors (red, green, and blue) from the three channels. Most common digital images use three channels to collectively form a colored image. As the PCNN model can process one or three channels from an image at a specific instance that the image represents, no temporal dimension is considered. The utilization of recurrent neural networks (RNNs) was considered, but because of the size and complexity of the PCNN model, the utilization of RNNs would have significantly increased the training time of the model without sufficient improvements [37]. Thus, we applied the temporal dimensionality of the PCNN model by including grayscale images of the TROPOMI images for each channel within the digital image format.

The original partial convolution padding process gradually reduces the significance of the missing data mask at each encoding phase of the PCNN model. Compared to the regular convolution process, the partial convolution process depends only on valid pixels, and normalization is adjusted to only a fraction of missing data. During this process, the convolution padding of the mask is applied in unison across all channels. Unfortunately, using different masks for each channel with the conventional convolution kernel causes the convolution kernel to process all the channels at the same time as one unified mask. This process does not reduce the significance of the mask at a gradual pace but at a much faster rate than expected, losing individual features of the mask and reducing the potential performance of the model. To address this limitation, we replaced the conventional 2D convolution layer (within the partial convolution) with depthwise convolutions [49] within the encoding phase of the PCNN model (see Fig. 2). A comparison of regular convolution padding and depthwise padding appears in Fig. S1. Since the significance of the mask is already removed during the final encoding process of the model, the decoding phase of the U-net architecture remains unchanged. We implemented the PCNN algorithm in the Keras and TensorFlow environments [66, 67].

We trained the PCNN model on CMAQ TCDNO$_2$ images of model runs from 2018 to 2019 and extracted

![Fig. 1 Map of the contiguous United States (CONUS) with domains of the CMAQ model (green) and TROPOMI measurement (blue) datasets used for the study (Color figure online)](image-url)
To perform a partial normalization process of the CMAQ TROPOMI images, we divided the entire dataset by a set value of $1 \times 10^{17}$ based on a value slightly above the maximum TCDNO$_2$ within the TROPOMI 2019 dataset. This ensured that proper distribution or regularization of the data would improve the model performance by reducing the significance of rare outliers within the dataset [39, 68]. To further increase the number of training samples and enhance the robustness of the model, we implemented two phases of image augmentation (transforming data into modified samples) [69] of the CMAQ images to generate more training images (see Table S1). The first phase involved applying a modified form of white noise within the CMAQ images to more accurately represent the pixel variations of the original TROPOMI images (see Fig. S2). The second phase involved the basic rotation and flipping of the image as well as a randomized linear function that added a positive or negative value shift to the image. TROPOMI TCDNO$_2$ missing data masks also underwent basic augmentation (randomized flipping and rotation) and were overlaid on each CMAQ image.

The partial convolution model structure consists of a total of 16 layers comprised of one input layer, seven depthwise partial convolution encoding layers, seven partial convolution decoding layers, and an output layer. Each encoding layer contains a pooling layer, a depthwise partial convolution layer with batch normalization, and an output layer. Each decoding layer consists of an upsampling layer, concatenated with a respective layer in the encoding layer. We processed the upsampling layer through a partial convolution layer with batch normalization and leaky ReLU activation function (negative slope coefficient = 0.5). Leaky ReLU prevents information loss and allows the negative parts of features within the convolution to activate [70]. We then processed the final decoding layer through a regular convolutional layer with the hyperbolic tangent (tanh) activation function, which provided the final output image (see Fig. 2) for the schematic of the system. We used the same loss function for the model training like that in Liu et al. [48]:

$$L_{\text{total}} = L_{\text{valid}} + 6L_{\text{hole}} + 0.05L_{\text{perceptual}} + 120 \left( L_{\text{styleout}} + L_{\text{stylecomp}} \right) + 0.1L_{\text{tv}}$$

which is comprised of pixel hole loss ($L_{\text{hole}}$), pixel valid loss ($L_{\text{valid}}$), perceptual loss ($L_{\text{perceptual}}$), raw style output ($L_{\text{styleout}}$), composited output ($L_{\text{stylecomp}}$), and total variation loss ($L_{\text{tv}}$).

The training of the models consisted of three phases: (i) the first training phase with batch normalization for 400 epochs with a learning rate of 0.001; (ii) the second training phase without batch normalization for 800 epochs with a learning rate of 0.001, and (iii) the final training phase with a reduced initialized learning rate of 0.0001 for 800 epochs. Batch normalization is performed during the first training phase to improve the initial speed of the model training by more effectively reducing the loss [71]. The second and third phases exclude batch normalization to optimize loss and reduce the potential bias of the model for imputation. Based on internal tests, the three-phase training improved loss optimization and reduced the overall training time. The optimizer used for the model was the adaptive moment estimation (Adam) [72] stochastic gradient descent method, which adaptively estimates the first- and second-order moments. Checkpoints were enabled by saving the model with the lowest validation error from each training phase.
2.3 Model comparisons

To compare the performance of the DW-PCNN, we used non-temporal-based imputation methods such as inverse distance weighting (IDW) [73] and the regular PCNN without depthwise convolutions. The IDW interpolation method assumes that pixels close to each other are likely to have similar values and that the local influence of available points on predictions diminishes with distance [74]. These two models showed the best performance in a previous study for the non-temporal imputation of Geostationary Ocean Color Imager aerosol optical depth images [30].

Because of the size of the dataset, which led to significant processing time that required more memory than the high-performance computing system could allocate, implementing the spatiotemporal kriging method [75] as a direct comparison to the DW-PCNN was not possible; therefore, to fill in any remaining missing data, we used IDW for the weekly mean TROPOMI images and integrated the results with CoKriging [75]. The IDW-CoKriging coupled system used the IDW imputed weekly mean (as a substitute to the temporal mean of the dataset) and fed it to the CoKriging process as a co-variable. CoKriging takes advantage of the covariance of the potential relationship of regionalized variables (the weekly mean within filled missing datasets by IDW) during the imputation process. Kriging, based on Gaussian process regression, assumes that spatial variation in a phenomenon is statistically homogeneous throughout a surface based on available data from nearby locations [76]. Both Kriging and IDW (weighting power = 5) models were based on the gstat package [77].

2.4 Evaluation

We evaluated the models based on various datasets and methods. Since we trained the PCNN model on CMAQ data, we did not conduct an evaluation based on these data. All daily images had measurements of TCDNO$_2$ values within the 2019 and 2020 study periods. We evaluated the imputation models based on TROPOMI NO$_2$ images by processing the daily TROPOMI images into a weekly moving average of TROPOMI images, and we observed a strong temporal correlation (0.96 $r$ for 2019 and 2020) in the weekly averages between daily variations (see Fig. S3); thus, we performed a weekly shift format (0.69 and 0.68 $r$ for 2019 and 2020, respectively) as input for the spatiotemporal imputation models. We evaluated TROPOMI TCDNO$_2$ by applying TROPOMI TCDNO$_2$ daily missing masks on the weekly mean TCDNO$_2$ images. Then we expected the models to accurately impute the TCDNO$_2$ images. In addition, we applied an estimated distance mask based on the distance to the nearest available data point within each image mask. The purpose of this process was to evaluate the imputation bias of the models to the distance of the nearest data variable.

3 Results and discussion

3.1 Imputation of the TROPOMI images

The evaluation of TROPOMI NO$_2$ imputation performances utilizes the index of agreement (IOA) [78], the correlation coefficient ($r$) [79], and the root-mean-square error (RMSE) [80] methods. The extraction of evaluation variables is based on the differences between the weekly mean mask and the daily masks for their respective days. The resulting extraction compiles the data into a one-dimensional format to be evaluated by the statistical evaluation methods for each daily dataset. The statistical results are separated between 2019 and 2020 data time periods. Due to IOA incorporating both correlation and bias performance, we used IOA to evaluate the model imputation performance based on the percentage of missing data within the image. The TROPOMI dataset for 2019 was composed of images missing 1 to 20% of data with $\sim 17\%$ of images from 2019 and 30% from 2020 with 10% or more pixels missing from an image (see Fig. S5).

The Taylor Diagram [81] statistical results of the various models and algorithms for the TROPOMI 2019 and TROPOMI 2020 cases are shown in Figs. 3 and 4, respectively. The mean statistical results are shown in Table 1. For both 2019 and 2020 TROPOMI TCDNO$_2$ images, the DW-PCNN model achieved the best overall performance in the IOA (0.85 for 2019 and 0.88 for 2020) and $r$ (0.76 for 2019 and 0.79 for 2020). The default PCNN model with spatiotemporal data (PCNN-ST) achieved the lowest MAE ($5.86 \times 10^{14}$ molecules/cm$^2$) and RMSE ($9.17 \times 10^{14}$ molecules/cm$^2$) statistical results for 2019, while the DW-PCNN had the lowest MAE ($5.39 \times 10^{14}$ molecules/cm$^2$) and RMSE ($8.07 \times 10^{14}$ molecules/cm$^2$) scores for 2020. Despite the lower performance of both IDW and IDW-CoKriging imputation methods in correlation and RMSE, both methods achieved a minor improvement in matching the TROPOMI standard deviation over the DW-PCNN model. Overall, the statistical comparisons for 2019 showed minimal differences between IDW and the coupled IDW-CoKriging imputation model (0.07 and 0.13% for the IOA and $r$, respectively). As a result of the reduced processing resources required, we focused on the IDW model for the 2020 comparisons. Based on the percentage of missing values within the evaluated pixels, the DW-PCNN model outperformed all models. For 2019, it outperformed the other methods in 6 to 34% of images with more than 10% missing data, 10 to 14% of images with
between 5 and 10% missing data, and 4 to 7% of images with less than 5% missing data. For 2020, the DW-PCNN also outperformed the other methods in 8 to 11% of images with more than 10% missing data, 4 to 11% of images with between 5 and 10% missing data, and 2 to 11% of images with less than 5% missing data. In contrast, the IDW-CoKriging model and the default PCNN models had mixed results with no clear indicator of which model was more accurate overall (see Table 1). Furthermore, both DW-PCNN and PCNN-ST were able to impute a TROPOMI image with over 95% missing data significantly more accurately than the other models (see Fig. S6 in Supplementary Document). The advantages and disadvantages of the models based on the results of imputing TROPOMI NO2 are provided in Table S2.

3.2 Pixel distance evaluation

The secondary evaluation method is the comparison of bias in relation to the distance in kilometers (km) from the nearest available data point within the TROPOMI image. Distance is calculated by applying a Euclidean process within the spatial fields and adjusting for pixel grid size in relation to the spatial coverage of TROPOMI for each missing data mask with the output rounded to the nearest integer. The bias of the imputation methods is performed by the subtraction of the measured pixel grid to the imputed pixel grid, which is then categorized based on the distance value of the respective pixels from the distance mask. Figure S4 in the supplementary document demonstrates how pixel distance is extracted for evaluation purposes. These distance biases within the evaluation mask of each TCDNO2 are then sorted into distance intervals by up to 1000 km within the TROPOMI data. The plots present the variances of the biases of the imputed TCDNO2 values to those of the available TROPOMI TCDNO2 value. We evaluated DW-PCNN, IDW-CoKriging (for 2019 only), IDW, PCNN, and PCNN-ST models. The distance plot (see Fig. 5) shows that the imputation models had bias variances related to the concentration range of the TROPOMI dataset in relation to the distance. The base PCNN model was the only model with significant bias and variance in imputing TCDNO2 compared to the other models and increase in bias as distance from the nearest available data point within the spatial data. The addition of temporal data into the base PCNN model (PCNN-ST) and DW-PCNN has shown a significant improvement in max bias variance and consistency over a large distance from the nearest data point. IDW-CoKriging in the mean bias comparison had minimal difference over the default IDW model with IDW-CoKriging overlapping the IDW plots. IDW-CoKriging and IDW had low mean biases in medium distances.
(100–600 km), but output a negative bias as the distance increased in relation to the spatiotemporal PCNN models (PCNN-ST and DW-PCNN). Both spatiotemporal PCNN models had similar performance across most distances with PCNN-ST performing slightly better mean bias at shorter distances over the DW-PCNN. In contrast, DW-PCNN had a narrower bias range across various distances over the PCNN-ST, indicating more consistency and stability in the imputation of the DW-PCNN as reflected in the statistical comparisons.

For TROPOMI 2020, the biases of the model imputations were lower than the 2019 biases over the same distance (Fig. 6). The PCNN presented the largest mean bias across the various distances with general mean positive bias in short to medium distances and mean negative bias in long distances (> 800 km) from the nearest available data point. IDW also achieved relatively stable biases over the distance ranges. Despite the slightly larger bias range compared to the DW-PCNN and the PCNN-ST, the IDW, compared to all of the models, had the most consistent standard deviation (SD). The right figure indicates the index of agreement (IOA) performance of the models, based on the percentage of missing data split into three categories. The main section of the boxplot presents an interquartile range between the 25 and 75th percentiles. The whiskers (vertical lines) of the boxplot represent the variability outside the interquartile range. The blue and yellow horizontal lines represent the mean and median of the dataset, respectively.

Table 1 The statistical evaluation results of imputing TROPOMI images by the inverse distance weighting (IDW), coupled IDW with CoKriging (IDW-CoKriging), regular Partial Convolutional Neural Network (Base PCNN), Base PCNN with Spatio- and temporal datasets (ST), and Depthwise PCNN (DW-PCNN) models. The evaluations are based on the index of agreement (IOA), correlation (r), mean absolute error (MAE), and root-mean-squared error (RMSE).

| Models            | 2019 IOA | r   | MAE  | RMSE  | 2020 IOA | r   | MAE  | RMSE  |
|-------------------|----------|-----|------|-------|----------|-----|------|-------|
| IDW               | 0.82  | 0.69 | 6.27 × 10^{14} | 1.04 × 10^{15} | 0.82  | 0.69 | 6.07 × 10^{14} | 9.58 × 10^{14} |
| IDW-CoKriging     | 0.82  | 0.69 | 6.20 × 10^{14} | 1.03 × 10^{15} | ––    | ––     | ––    | –– |
| PCNN              | 0.75  | 0.60 | 8.10 × 10^{14} | 1.32 × 10^{15} | 0.84  | 0.72 | 6.30 × 10^{14} | 9.63 × 10^{14} |
| PCNN-ST           | 0.80  | 0.67 | 5.86 × 10^{14} | 9.17 × 10^{14} | 0.81  | 0.70 | 5.84 × 10^{14} | 8.85 × 10^{14} |
| DW-PCNN           | **0.85** | **0.76** | **5.95 × 10^{14}** | **9.72 × 10^{14}** | **0.88** | **0.79** | **5.39 × 10^{14}** | **8.07 × 10^{14}** |

Bold represents the best case for each metric.
mean bias without a positive or negative trend below 700 km distances. As with 2019, the DW-PCNN and PCNN-ST models showed the narrowest bias range of the models across all the distances for the 2020 TROPOMI dataset. Beyond 700 km distances, the IDW, PCNN-ST, and DW-PCNN models had similar positive mean biases in relation to TROPOMI NO\textsubscript{2} concentrations.

### 4 Conclusion

This research demonstrated the improved capability of the depthwise partial convolutional neural network in the application of spatiotemporal imputation of missing remote sensing data. Both the 2019 and 2020 TROPOMI TCDNO\textsubscript{2} imputation results demonstrated that the DW-PCNN most consistently and accurately imputed missing TROPOMI TCDNO\textsubscript{2} data. With the addition of spatiotemporal data to the PCNN model, it showed a significant improvement over the regular PCNN model (without temporal data) with datasets containing large percentages of missing data and at extended distances. Despite the improvements of adding temporal dimensionality within the input of the PCNN model, the mask padding of the regular convolution process shows limitations, which has led to some bias. The implementation of depthwise convolutions, in which the masks at each image channel are padded separately, showed further improvement, demonstrating the importance of maintaining the individual channel masks and gradual feature reduction over the conventional method. Furthermore, the DW-PCNN was the only model that maintained a mean IOA above 0.85 and mean correlation above 0.76 in all of the statistical comparisons for both the TROPOMI 2019 and 2020 TCDNO\textsubscript{2} datasets. The current limitation of the DW-PCNN is the constrained number of channels (only three) that the model can process, thus...
restricting the temporal samples and the addition of covariables that enhance the accuracy of imputation.

While the bias-distance comparison showed that the PCNN-ST performed as well as the DW-PCNN, the statistical comparisons showed that the PCNN-ST had a lower correlation when reconstructing the TROPOMI TCDNO₂ dataset than the default PCNN model. This finding can be explained by the output of the PCNN-ST model, which showed smoother transitions than the PCNN or DW-PCNN models, a slight under-prediction of high TCDNO₂ concentrations, and an over-prediction of low TCDNO₂ column concentrations. Although this phenomenon may minimally impact the evaluation of bias, it impacts the correlation and IOA scores to a greater extent.

For a spatial imputation algorithm without temporal dimensionality, IDW performed consistently compared to the PCNN and PCNN-ST models. Although IDW was not able to surpass the PCNN models in different metrics, it did not perform the worst in the respective metrics. The major limitation of IDW is the computational cost of such large datasets, especially when required to take all available samples within the TROPOMI image. In fact, IDW and other interpolation-based algorithms (e.g., Kriging) require exponentially more computation power and resources as the dataset size increases spatially and when temporal dimensionality is added. For smaller dataset sizes, however, these algorithms show performances similar to that of the PCNN model (refer to [30]). As datasets increase in size, dimensionality, and missing data, the benefits of deep learning algorithms for imputation purposes also increase.

Once trained, not only does the DW-PCNN model impute large remote sensing datasets in significantly less processing time than interpolation-based algorithms [30], but it is also significantly more accurate than the default PCNN models with and without the temporal dimensionality of datasets. For the accurate imputation of such
and Texas Advanced Computing Center (TACC) at the University of Houston. Guber is available at https://github.com/MathiasGruber/PConv-Keras.

References

Conflict of interest  The authors declare that they have no conflict of interest.

Declarations

Data availability  TROPOMI NO2 data can be freely accessed from the European Space Copernicus Open Access Hub or the NASA EarthData Portal (https://doi.org/10.5270/S5P-9bnp8q8). The repository for the prepared datasets, model codes, and their respective weights is accessible in https://doi.org/10.5281/zenodo.7770693. The original Partial Convolutional Neural Network code from Mathias Guber is available at https://github.com/MathiasGruber/PConv-Keras.

Supplementary Information  The online version contains supplementary material available at https://doi.org/10.1007/s00521-023-08558-1.

Acknowledgements  This study was supported by funding from the Oak Ridge Associated Universities (ORAU) Directed Research and Development Program. The authors acknowledge the Research Computing Data Core (RCDC) at the University of Houston (https://uh.edu/rcdc/) and Texas Advanced Computing Center (TACC) at the University of Texas at Austin (http://www.tacc.utexas.edu) for providing high-performance computing resources that have contributed to the research results reported within this paper. We would like to thank Mathias Gruber for reconstructing the partial CNN code for the TensorFlow implementation, which we have updated and modified.

Data availability  TROPOMI NO2 data can be freely accessed from the European Space Copernicus Open Access Hub or the NASA EarthData Portal (https://doi.org/10.5270/S5P-9bnp8q8). The repository for the prepared datasets, model codes, and their respective weights is accessible in https://doi.org/10.5281/zenodo.7770693. The original Partial Convolutional Neural Network code from Mathias Guber is available at https://github.com/MathiasGruber/PConv-Keras.

Supplementary Information  The online version contains supplementary material available at https://doi.org/10.1007/s00521-023-08558-1.

Acknowledgements  This study was supported by funding from the Oak Ridge Associated Universities (ORAU) Directed Research and Development Program. The authors acknowledge the Research Computing Data Core (RCDC) at the University of Houston (https://uh.edu/rcdc/) and Texas Advanced Computing Center (TACC) at the University of Texas at Austin (http://www.tacc.utexas.edu) for providing high-performance computing resources that have contributed to the research results reported within this paper. We would like to thank Mathias Gruber for reconstructing the partial CNN code for the TensorFlow implementation, which we have updated and modified.

Data availability  TROPOMI NO2 data can be freely accessed from the European Space Copernicus Open Access Hub or the NASA EarthData Portal (https://doi.org/10.5270/S5P-9bnp8q8). The repository for the prepared datasets, model codes, and their respective weights is accessible in https://doi.org/10.5281/zenodo.7770693. The original Partial Convolutional Neural Network code from Mathias Guber is available at https://github.com/MathiasGruber/PConv-Keras.

Supplementary Information  The online version contains supplementary material available at https://doi.org/10.1007/s00521-023-08558-1.

Acknowledgements  This study was supported by funding from the Oak Ridge Associated Universities (ORAU) Directed Research and Development Program. The authors acknowledge the Research Computing Data Core (RCDC) at the University of Houston (https://uh.edu/rcdc/) and Texas Advanced Computing Center (TACC) at the University of Texas at Austin (http://www.tacc.utexas.edu) for providing high-performance computing resources that have contributed to the research results reported within this paper. We would like to thank Mathias Gruber for reconstructing the partial CNN code for the TensorFlow implementation, which we have updated and modified.

Data availability  TROPOMI NO2 data can be freely accessed from the European Space Copernicus Open Access Hub or the NASA EarthData Portal (https://doi.org/10.5270/S5P-9bnp8q8). The repository for the prepared datasets, model codes, and their respective weights is accessible in https://doi.org/10.5281/zenodo.7770693. The original Partial Convolutional Neural Network code from Mathias Guber is available at https://github.com/MathiasGruber/PConv-Keras.
22. Bocquet M, Elbern H, Eskes H, Hirtl M, Žabkar R, Carmichael GR, Saide PE (2015) Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. Atmos Chem Phys 15(10):5325–5358
23. Jung J, Souri AH, Wong DC, Lee S, Jeon W, Kim J, Choi Y (2019) The impact of the direct effect of aerosols on meteorology and air quality using aerosol optical depth assimilation during the KORUS-AQ campaign. J Geophys Res: Atmos 124(14):8303–8319
24. Zhang C, Li W, Travis D (2007) Gaps-fill of SLC-off Landsat ETM+ satellite image using a geostatistical approach. Int J Remote Sens 28(22):5103–5122
25. Yu C, Chen L, Su L, Fan M, Li S (2011) Krigeing interpolation method and its application in retrieval of MODIS aerosol optical depth. In: 2011 19th international conference on geoinformatics, IEEE, pp 1–6
26. Bertalmío M, Sapiro G, Caselles V, Ballester C (2000) Image inpainting. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp 417–424
27. Bugeau A, Bertalmio M, Caselles V, Sapiro G (2010) A comprehensive framework for image inpainting. IEEE Trans Image Process 19(10):2634–2645
28. Ghahremanloo M, Choi Y, Sayeed A, Salman AK, Pan S, Amani M (2021) Estimating daily high-resolution PM2.5 concentrations over Texas: machine learning approach. Atmos Environ 247:78–118209
29. Li Y, Xie W, Li H (2017) Hyperspectral image reconstruction by deep convolutional neural network for classification. Pattern Recogn 63:371–383
30. Lops Y, Pouyaei A, Choi Y, Jung J, Salman AK, Sayeed A (2021) Application of a partial convolutional neural network for estimating geostationary aerosol optical depth data. Geophys Res Lett 48(15):e2021GL093096
31. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
32. Bengio Y (2009) Learning deep architectures for AI. Now Publishers Inc, Netherlands
33. Deng L, Yu D (2014) Deep learning: methods and applications. Found Trends Signal Process 7(3–4):197–387
34. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
35. LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series, Handbook Brain Theory Neural Netw 512(555):436–444
36. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:65–117
37. Lops Y, Choi Y, Eslami E, Sayeed A (2019) Real-time 7-day forecast of pollen counts using a deep convolutional neural network. Neural Comput Appl 32:1–10
38. Sayeed A, Choi Y, Eslami E, Lops Y, Roy A, Jung J (2020) Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. Neural Netw 121:396–408
39. Sayeed A, Lops Y, Choi Y, Jung J, Salman AK, Sayeed A (2021) Bias correction and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. Atmos Environ 253:118376
40. Yeo I, Choi Y, Lops Y, Sayeed A (2021) Efficient PM2.5 forecasting using geographical correlation based on integrated deep learning algorithms. Neural Comput Appl 33(22):15073–15089
41. Anthimopoulos M, Christoudoulis S, Ebnner L, Christe A, Mougakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35(5):1207–1216
42. Lee H, Kwon H (2017) Going deeper with contextual CNN for hyperspectral image classification. IEEE Trans Image Process 26(10):4843–4855
43. Mikolov T, Kombrink S, Burget L, Černocký J, Khudanpur S (2011) Extensions of recurrent neural network language model. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 5528–5531
44. Mukherjee A, Agrawal M (2017) World air particulate matter: sources, distribution and health effects. Environ Chem Lett 15(2):283–309
45. Park S, Jeong Y, Kim HS (2017) Multiresolution CNN for reverberant speech recognition. In: 2017 20th conference of the oriental chapter of the international coordinating committee on speech databases and speech I/O systems and assessment (OCOCOSDA), IEEE, pp 1–4
46. Zhang Q, Yuan Q, Zeng C, Li X, Wei Y (2018) Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. IEEE Trans Geosci Remote Sens 56(8):4274–4288
47. Gerber F, de Jong R, Schaepman ME, Schaepman-Strub G, Furrr R (2018) Predicting missing values in spatio-temporal remote sensing data. IEEE Trans Geosci Remote Sens 56(5):2841–2853
48. Carvalho JP, Nakai AM, Monteiro JE (2016) Spatio-Temporal modeling of data imputation for daily rainfall series in Homogeneous Zones. Rev Bras de Meteorol 31:196–201
49. Liu G, Reda FA, Shih KJ, Wang TC, Tao A, Catanzerio B (2018) Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European conference on computer vision (ECCV), pp 85–100
50. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
51. Veekind JP, Aben I, McMullan K, Förster H, De Vries J, Otter G, Level P (2012) TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sens Environ 120:70–83
52. Guanter L, Aben I, Tol P, Krijger JM, Hollstein A, Köhler P, Landgraf J (2015) Potential of the TROPOSpheric monitoring instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. Atmos Meas Tech 8(3):1337–1352
53. Ludewig A (2021) SSP mission performance centre level 1b Readme. Reference: SSP-MPC-KNMI-PRF-L1B, (2.0), 3.0.0
54. Wiedinmyer C, Akagi SK, Yokelson RJ, Emmons LK, Al-Saadi JA, Orlando JJ, Soja AJ (2011) The Fire InVentory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning. Geosci Model Develop 4(3):625–641
59. Wiedinmyer C, Quayle B, Geron C, Belote A, McKenzie D, Zhang X, Wynne KK (2006) Estimating emissions from fires in North America for air quality modeling. Atmos Environ 40(19):3419–3432
60. Wiedinmyer C, Yokelson RJ, Gullett BK (2014) Global emissions of trace gases, particulate matter, and hazardous air pollutants from open burning of domestic waste. Environ Sci Technol 48(16):9523–9530
61. Pleim JE, Gilliam R (2009) An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land surface model. J Appl Meteorol Climatol 48(7):1362–1376
62. Pleim JE, Xi A (2003) Development of a land surface model. Part II: data assimilation. J Appl Meteorol 42(12):1811–1822
63. Hogrefe C, Poulidou G, Wong D, Torian A, Roselle S, Pleim J, Mathur R (2015) Annual application and evaluation of the online coupled WRF–CMAQ system over North America under AQMEII phase 2. Atmos Environ 115:683–694
64. Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In: 2018 international interdisciplinary PhD workshop (IIPhDW), IEEE, pp 117–122
65. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, Springer, Cham, pp 234–241
66. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113
67. Chollet F (2018) Keras: The python deep learning library. ascl, ascl-1806
68. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Zheng X (2016) Tensorflow: a system for large-scale machine learning. In: 12th [USENIX] symposium on operating systems design and implementation (OSDI) 16, pp 265–283
69. Eslami E, Salman AK, Choi Y, Sayeed A, Lops Y (2019) A data ensemble approach for real-time air quality forecasting using extremely randomized trees and deep neural networks. Neural Comput Appl 32:1–17
70. Fawzi A, Samulowitz H, Turaga D, Frossard P (2016) Adaptive data augmentation for image classification. In: 2016 IEEE international conference on image processing (ICIP), IEEE, pp 3688–3692
71. Zhang X, Zou Y, Shi W (2017) Dilated convolution neural network with LeakyReLU for environmental sound classification. In: 2017 22nd international conference on digital signal processing (DSP), IEEE, pp 1–5
72. Bjorck J, Gomes C, Selman B, Weinberger KQ (2018) Understanding batch normalization. arXiv preprint arXiv:1806.02375
73. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980
74. Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. Comput Geosci 34(9):1044–1055
75. Fisher NI, Lewis T, Embleton BJ (1993) Statistical analysis of spherical data. Cambridge University Press, Cambridge
76. Kyriakidis PC, Journel AG (1999) Geostatistical space–time models: a review. Math Geol 31(6):651–684
77. Cressie N (1990) The origins of kriging. Math Geol 22(3):239–252
78. Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30(7):683–691
79. Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, Rowe CM (1985) Statistics for the evaluation and comparison of models. J Geophys Res: Oceans 90(C5):8995–9005
80. Benesty J, Chen J, Huang Y, Cohen I (2009) Pearson correlation coefficient. In: Cohen I, Huang Y, Chen J, Benest J (eds) Noise reduction in speech processing. Springer, Berlin, Heidelberg, pp 1–4
81. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. Geosci Model Develop 7(3):1247–1250