U-NET Based Spatial Resolution Enhancement for Geospatial Data

Sudarshan Zunja
Dept. of Information Technology
National Institute of Technology Karnataka, Surathkal
sudarshanzunja.211it072@nitk.edu.in

Dr. Shrutilipi Bhattacharjee
Dept. of Information Technology
National Institute of Technology Karnataka, Surathkal
shrutilipi@nitk.edu.in

Spurthy Maria Pais
Dept. of Information Technology
National Institute of Technology Karnataka, Surathkal
spurthymariapais.207it006@nitk.edu.in

***Abstract*—Geospatial data collected and extracted from various satellites comes with different spatial and temporal resolutions due to factors like varying geographic, environmental and climatic variations. To perform extensive analysis, the data must be available at a higher resolution with clear and precise metrics for various features. The objective of this research is to downscale lower-resolution TROPOMI-SIF data to higher resolution utilizing auxiliary MODIS data. Direct Convolution over features is not sufficient, hence Local Binary Pattern (LBP) is applied to capture local spatial features, overlaying these over SIF data for more accurate results. These features are then passed to the U-Net Model, a type of convolution neural network with encoder-decoder architecture, which performs convolution operations in the encoding-decoding form to learn and predict the target variable at a higher resolution.**

***Index Terms*—Geospatial data, spatial resolution, downscaling, TROPOMI-SIF, MODIS, Local Binary Pattern (LBP), U-Net, convolutional neural network, satellite data analysis, and data standardization.**

I. Introduction

Scientific data collected through various geo-satellites are in varying spatial and temporal resolutions. Different factors such as geographic, and demographic variations including different types of land masses, altitude variations, water bodies, and vegetative areas. For effective understanding and analysis of geographic metrics and data collected through these satellites, the data must be converted into a standard form for consistent analysis and comparison. This helps in retrieving complex patterns and the nature of geospatial data which further assists in achieving desired results.

One such aspect is the spatial resolution of geospatial data. Spatial resolution can determine how precise the collected data is in terms of spatial sense. Hence it is necessary to collect data at higher spatial resolution to achieve accurate results during the analysis of data. This project focuses on downscaling lower-resolution data to higher resolution. However, we cannot just convert them through simple regression tasks.

For such cases, we use an auxiliary dataset as a spatial reference for the target resolution to map and find necessary patterns. We need to choose such auxiliary data that is exactly equivalent to your target resolution. This project aims to utilize this auxiliary data along with input data with lower resolution to produce the output at a higher target resolution.

Here we aim to downscale TROPOMI-SIF from its original resolution 3.5x7 km to a resolution more precise. MODIS is used as the auxiliary data for the same which has a spatial resolution of 1x1 km. The study region is the entirety of the state of Minnesota along with a few parts of North and South Dakota, USA To capture local features and patterns, we implement localization using 'Local Binary Pattern'. LBP will overlap over local areas over SIF to capture more precise details.

After all the features are collected and curated for model input, we finalize our U-NET model. U-NET is a type of Convolutional Neural Network that encompasses an Encoder and Decoder structure to learn and perform Convolutional tasks over the input matrices. The model then learns and adapts to predict target SIF at a higher resolution.

II. Related Work

The paper "A Convolutional Neural Network for Spatial Downscaling of Satellite-Based Solar-Induced Chlorophyll Fluorescence (SIFnet)" by Gao et al. (2022) proposes a convolutional neural network-based methodology for downscaling of spatial resolution for satellite-based solar-induced chlorophyll fluorescence (SIF). This literature review examines the significance of spatial downscaling for SIF data, existing downscaling methods, and the role of CNNs in improving spatial resolution. The data retrieved from Sentiel5P has a coarse spatial resolution which limits its utility for analytical purposes as well as is unable to present more information for more precise spatial regions. Low resolution restricts the relaying of precise information related to vegetation health and photosynthesis. Hence, it is necessary to downscale it to a higher spatial resolution as it lifts these restrictions to produce utilizable data.

Traditional spatial downscaling methods, such as the nearest neighbour, bilinear interpolation, and cubic convolution, are straightforward but often result in the loss of important information and the introduction of artefacts. These methods may not fully capture the complex spatial patterns and relationships inherent in high-resolution datasets. Recent advances have seen machine learning techniques, particularly artificial neural networks like CNNs, emerge as powerful alternatives. These methods can learn intricate relationships between data, leading to more accurate and refined downscaled products.

CNNs have demonstrated significant success in the spatial downscaling of various remote sensing data products, including land cover maps and vegetation indices. By learning spatial features from high-resolution auxiliary data, such as land cover maps and vegetation indices, CNNs can enhance the quality of downscaled SIF products. Their ability to capture and utilize complex spatial relationships makes them well-suited for this task, surpassing the capabilities of traditional downscaling methods.

The SIFnet model proposed by Gao et al. (2022) represents a significant advancement in CNN-based spatial downscaling for SIF data. Its strengths include the capacity to capture complex spatial relationships and deliver superior downscaling performance compared to traditional methods. However, the model has some limitations, such as the requirement for substantial amounts of training data and the potential for black-box behaviour, where the internal mechanisms of the network are not easily interpretable. These aspects may affect the model's overall usability and understanding.

Building upon the work of Gao et al. (2022), several areas for improvement can be explored. Investigating alternative CNN architectures or fine-tuning hyperparameters could potentially enhance the downscaling performance of SIFnet. Additionally, incorporating additional auxiliary data sources might improve the accuracy of the downscaled SIF product. To address the black-box nature of CNNs, techniques to enhance model interpretability should be considered, enabling a clearer understanding of how the model performs downscaling. Finally, evaluating SIFnet across different types of vegetation or environmental conditions can help assess its generalizability and robustness. By addressing these aspects, future research can further advance the spatial downscaling of SIF data, building on the foundational work of Gao et al. (2022) and potentially achieving even greater accuracy and applicability in various environmental and agricultural contexts.

III. Proposed Methodology

The study leverages Sentinel-5P's TROPOMI-SIF data, which provides sun-induced fluorescence (SIF) data crucial for vegetation analysis, and MODIS Terra Vegetation Indices, which serve as auxiliary data for downscaling SIF datasets. SIF data is aggregated at a resolution of 3.5x7 km for 2019 and synthesized using QGIS for spatial and temporal aggregation at 1x1 km resolution. MODIS data is similarly processed, and both datasets are integrated grid-wise for model training. Features are extracted by correlating SIF743 with MODIS indices and employing Local Binary Patterns (LBP) for texture extraction. The final dataset is preprocessed to generate spatial matrices, which are then cropped into smaller segments for input into a multi-channel U-NET model. The U-NET model, comprising 5 Encoder Blocks, a Bottleneck, and 5 Decoder Blocks, is used to upscale the spatial resolution of SIF data, with the architecture tuned for optimal performance.

A. Data Collection

1) Sentinel-5P TROPOMI SIF: Sentinel-5P's TROPOMI-SIF data is the primary dataset used to train a model to produce higher spatial resolution samples from a lower one. Sentinel-5P's mission is to perform atmospheric measurements with high spatiotemporal resolution.

The sun-induced fluorescence (SIF) signal emitted by the chlorophyll-a of terrestrial vegetation is a closer indicator of vegetation functioning than other variables traditionally derived from optical remote sensing data. TROPOMI can provide reflectance data in the 675-775 nm spectral window.

The SIF data used is aggregated product at a resolution of 3.5x7km collected over each day for the year of 2019. The data can be collected on Noveltis web.

2) MODIS Terra Vegetation Indices: MODIS vegetation indices, produced on 16-day intervals and at multiple spatial resolutions, provide consistent spatial and temporal comparisons of vegetation canopy greenness, a composite property of leaf area, chlorophyll and canopy structure.

MODIS is used as auxiliary data for downscaling the SIF dataset. 1x1km grid MODIS vegetation indices are collected for 16 days over the year 2019. With a 1x1 km grid, this data can be coarsed to various resolutions to overlap and look for patterns over upscaled data. MODIS data for required study can be collected from sources like NASA EarthData search.

B. Data Extraction

TROPOMO-SIF data is avaialble in standard NETCDF4 format. Different products for present for varying spectral ranges. The model is trained over the SIF743 product for the study region of the state of Minnesota. Similarly, MODIS is available in standard HDF format. These data must be converted to valid CSV or standard format to be processed further.

C. QGIS and Data Synthesis

QGIS (Quantum Geographic Information System) is a free, open-source software that allows users to create, edit, visualize, analyze, and publish geospatial information. The final data is to be synthesised using tools available through QGIS.

Constant CRS must be chosen and maintained throughout the project based on the geographical aspects of the chosen study region. Proper consideration must be taken for ellipsoidal and cartesian metrics during the usage of QGIS.

1) SIF Data Synthesis: Firstly, SIF data for each day for a set of equivalent 16-day intervals of MODIS time window must be imported in QGIS and visualized in a grid manner throughout the study region. Begin by clipping the data by study region. It it necessary to generate a coarse-resolution dataset for our SIF data. The coarse resolution could be any factor of original resolution based on your study region, land cover variations and availability of computational resources.

Create a grid at coarse resolution aggregate all SIF values within and then assign those for the original set. Now for this entire study region, plot grids of 1x1 km, ensure proper coverage of the study region without any overlapping. Once this, assign the closest/overlapping SIF values to each of these grids of 1x1 km. This will generate coarse SIF data for each 16 days at 1x1 km scale. This is Spatial Aggregation. Now aggregated SIF values by location for all 16-day time interval windows. If there are less than 16 days of SIF data for a particular MODIS window, continue with whichever single-day SIF data is available for the existing time window. This is Temporal Aggregation. Repeat this spatial and temporal for all 16 days based on the MODIS time window interval. For 2 MODIS datasets, we need 16+16 = 32 single-day SIF data and so on. Here, the SIF data is ready once repeated at all available time intervals.

2) MODIS Data Synthesis: MODIS is available every 16 days in a 1x1 SIN grid. This data is in sinusoidal raster form. Begin by clipping the necessary data for the study region. If necessary convert this to a valid Vector Layer for efficient augmentation tasks further. Plot 1x1 km grid covering the study region. Assign the nearest coarse MODIS data points to each of these grids. This generates a MODIS dataset at a 1x1 km grid.

3) SIF-MODIS Integration: Once both datasets are ready, perform grid-wise integration of SIF-MODIS. Include SIF743 at both original and coarse resolution. The data produced here is considered as a single instance, which is the spatial and temporal aggregation of 16-day intervals. Perform this for a lot of 16-day intervals to generate enough instances to be sufficient for model training.

D. Feature Extraction

Once these samples are ready, we extract features by calculating the correlation between SIF743 and MODIS Indices (do the same for any other auxiliary data used in combination). The most correlated auxiliary features to SIF743 are to be extracted as new data along with SIF743 and coarse SIF743(which we will refer to as target SIF further). Multiple approaches can be employed for feature extraction based on the number of features in the auxiliary data. We add another additional feature known as the SIF difference, which is nothing but the difference between SIF743 and target SIF743.

E. Matrix Generation

Once the final dataset is ready, preprocess it for NaN values. Generate a spatial matrix using latitude and longitude coordinates to create a grid-like region. The matrix should represent the accurate spatial arrangement of these data points. Generate a new matrix for each single input feature to be utilized. Stack all features Repeat for all input samples to generate a stack of feature matrices.

F. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is an efficient texture operator that labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number. It extracts features locally, that is, by considering only adjacent pixels to maximise its effectiveness in texture feature extraction. Apply LBP to SIF743 to extract localized information and patterns. This maintains uniformity within the SIF values during predictions. Adjust the intensity of the LBP feature matrix based on the necessity.

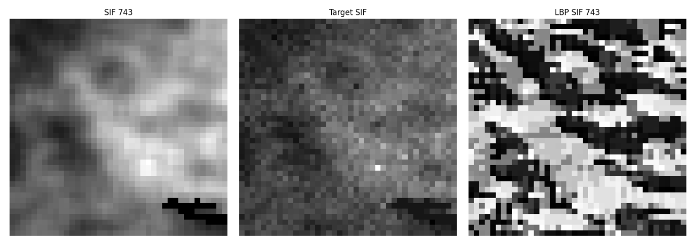

Fig. 1. Texture Extraction using Local Binary Pattern (LBP)

G. Matrix Cropping and Generation of Input Data

Once all sample matrices are generated, crop these matrices by MxM size to generate a larger input dataset. This also complements the capturing of local scale matching patterns along with learning information about neighbouring pixels. Cropping could either be done by random selection of matrices or using the sliding window technique where you slide to crop each matrix in every pixel column. These stacked matrices can be passed through a multi-channelled U-NET model for training.

H. U-NET Model Training and Testing

Pass the input stacked matrices in multi-channel format to this model. Adjust the number of Encoders and Decoders necessary for better performance. Simulate results with different kernel sizes and convolutional block configurations.

The model used for training for this research consists of 5 Encoder Blocks, a Bottleneck, and 5 Decoder Blocks with a final convolutional layer.

Each Encoder Block includes two convolutional layers with ReLU activation, progressively increasing the number of filters from 16-256. The bottleneck then employs 512 filters.

Each Decoder Block comprises an upscaling operation followed by concatenation with respective corresponding feature maps from Encoder Block layers, followed by additional convolutional layers. Filters decrease from 256-16 during upscaling in this block. The final convolutional layers output a single channel, resulting in prediction output.

IV. Results and Analysis

A. U-NET Model Performance

U-NET model manages to downscale the input SIF743 efficiently. The following results were obtained after the model was trained:

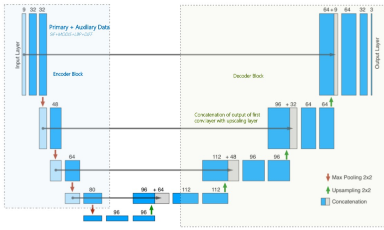

Fig. 2. U-NET Model Architecture

TABLE I
MODEL CONFIGURATION DETAILS

Configuration	Details
Encoder	
enc1	2 Conv. layers, ReLU, 16 filters
enc2	2 Conv. layers, ReLU, 64 filters
enc3	2 Conv. layers, ReLU, 128 filters
enc4	2 Conv. layers, ReLU, 256 filters
Bottleneck	1 Block with 2 Conv. layers, 512 filters
Decoder	
deco1	Upscaling, Concatenation with enc4, 256 filters
deco2	Upscaling, Concatenation with enc3, 128 filters
deco3	Upscaling, Concatenation with enc2, 64 filters
Output	Final Conv. layer, single channel output

TABLE II
U-NET MODEL PERFORMANCE

Metric	Value
R^2 Score	0.9692
MAE	0.0711
RMSE	0.0956

B. Comparison with CNN Model

The CNN-based model, SIFnet, used in the corresponding Johannes paper uses a direct approach of SIF along with auxiliary MODIS passed to a multi-channelled CNN model. This model however is not accurate in terms of capturing local features over input SIF thus not efficient in terms of local accuracy. Here are the performance results for the approach used to train SIFNet in Johannes paper against U-NET:

Localization techniques like Local Binary Pattern extract more localized information from data as compared to SIFnet which does not focus on local feature patterns. Supplementary to the above, U-Net concatenates encoded layers over down-scaled features during the decoding process. This enables the

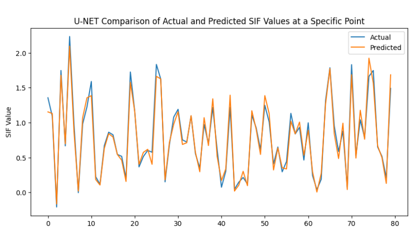

Fig. 3. U-NET Model Performance against Predictions

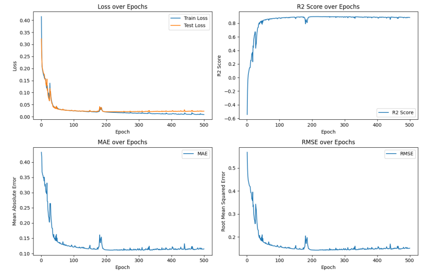

Fig. 4. U-NET Model Performance over Epochs

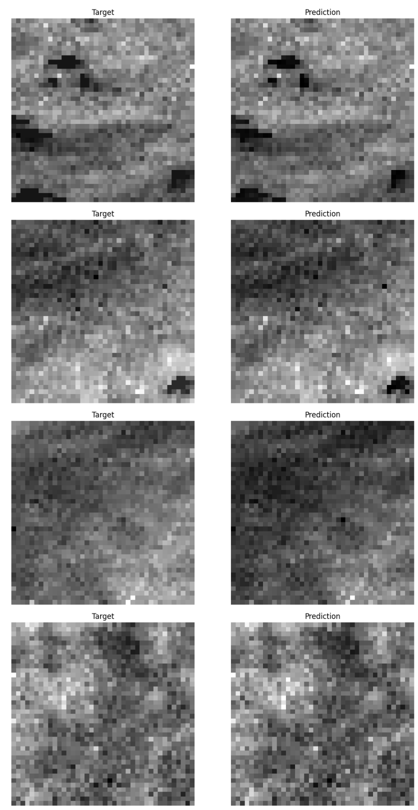

Fig. 5. U-NET: Sample Prediction

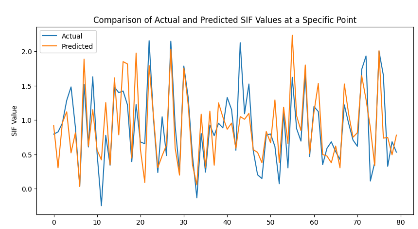

Fig. 6. SIFnet Model Performance against Predictions

TABLE III
MODEL PERFORMANCE COMPARISON

Metric	SIFnet-CNN Model	U-NET Model
R^2 Score	0.5300	**0.9692**
MAE	0.2850	**0.0711**
RMSE	0.3731	**0.0956**

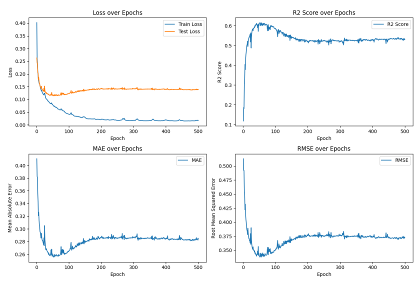

Fig. 7. SIFnet Model Performance over Epochs

model to gather more contextual information from encoder blocks assisting in capturing more regional features.

V. Conclusion and Future Work

This project effectively demonstrates a more efficient, precise methodology for downscaling spatial resolution using a U-NET-based model against SIFnet's CNN model. It also brings to light the effectiveness of localization and texture extraction of the Local Binary Pattern (LBP) technique.

Along with localization, it also distinguishes the efficient performance of the Encoder-Decoder assembled model. Concatenation of layers of these blocks along with convolutions achieves more precise results than just plain convolutions.

This work could be expanded to a multitude of applications like gap filling at higher resolution by passing the output of U-NET into dense networks for prediction tasks.

VI. References

1) Tian, F., Brandt, M., Liu, Y. Y., Rasmussen, K., & Fensholt, R. (2019). Enhanced vegetation index (EVI)-based downscaling of solar-induced chlorophyll fluorescence: Theoretical basis and validation. *Remote Sensing of Environment*, 223, 50-60.
2) Song, Z., Wang, H., Liu, D., & Huang, N. (2020). Spatial downscaling of MODIS LAI based on machine learning methods. *Remote Sensing*, 12(9), 1476.
3) Gao, F., Zhang, M., Liang, S., Sun, J., Liu, Y., & Zhang, L. (2022). A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet). *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-13.
4) Yu, K., Liu, Z., Yang, H., Liu, Z., & Liu, W. (2018). Spatial downscaling of satellite precipitation products using geographically weighted regression. *Remote Sensing*, 10(9), 1460.
5) Qu, X., Cao, J., Zhu, L., Li, M., & Wang, P. (2022). High-resolution land surface temperature estimation with CNN-based downscaling technique: A case study of metropolitan areas in China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, 215-228.
6) Ferg, A., Malakar, N., Hook, S. J., Li, Z., & Hulley, G. C. (2020). Deep learning-based downscaling of GOES-16 land surface temperature over the Conterminous United States. *Remote Sensing*, 12(22), 3732.
7) Liu, F., Wang, Y., Zeng, Y., Su, Y., Yuan, W., & Dong, W. (2021). Multi-source data fusion for high-resolution mapping of global solar-induced chlorophyll fluorescence using a deep neural network. *Remote Sensing of Environment*, 254, 112244.
8) Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *In Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234-241.
9) Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *In Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 424-432.
10) Isensee, F., Kickingereder, P., Schwede, M., & Maier-Hein, K. H. (2018). Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. *In Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 287-297.
11) Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700-4708.
12) Zheng, Y., Zheng, Y., & Li, W. (2020). A review on deep learning-based image segmentation with U-Net. *Journal of Computer Vision and Image Understanding*, 198, 102988.
13) Ali, M. A., Awais, M., & Ali, A. (2020). U-Net-based deep learning framework for semantic segmentation of agricultural fields. *Computers and Electronics in Agriculture*, 175, 105596.
14) Zhang, Z., Song, X., Xu, C., & Zhang, L. (2020). Deep U-Net with attention mechanism for end-to-end segmentation of biomedical images. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2354-2362.
15) Ojala, M., Pietikäinen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(7), 971-987.
16) Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face recognition with local binary patterns. *In European Conference on Computer Vision (ECCV)*, 469-481.
17) Zhang, X., Wu, X., & Wang, J. (2010). Local binary pattern (LBP) based image retrieval. *Journal of Visual*

Communication and Image Representation, 21(8), 740-749.
18) Liao, P. S., Shen, C., & Zhao, Y. (2007). Efficient texture classification and segmentation using local binary patterns. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 741-748.
19) Tan, X., Wang, Y., & Liu, Y. (2008). Enhanced local binary patterns for face recognition. *In IEEE International Conference on Image Processing (ICIP)*, 2641-2644.
20) Yu, H., Li, Y., & Zhang, Y. (2015). An improved local binary pattern method for facial emotion recognition. *Journal of Visual Communication and Image Representation*, 29, 27-35.
21) Zhang, Z., Liu, J., & Cheng, H. (2016). Local binary pattern with combined features for texture classification. *Pattern Recognition Letters*, 81, 94-101.