RGB2LST: Enhancing Deep Learning-Based Land Surface Temperature Estimation with Multi-Modality and Artifacts Removal

Issam Khedher, Jean-Marie Favreau, Serge Miguet, Gilles Gesquière

RGB2LST: Enhancing Deep Learning-Based Land Surface Temperature Estimation with Multi-Modality and Artifacts Removal

Issam Khedher*, Jean-Marie Favreau†, Serge Miguet*, and Gilles Gesquière*
*Univ Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, Centrale Lyon, LIRIS, UMR5205, F-69676 Bron, France
Email: {issam.khedher, serge.miguet, gilles.gesquiere}@univ-lyon2.fr
†Université Clermont Auvergne, CNRS, Mines de Saint-Étienne, LIMOS, F-63000 Clermont–Ferrand, France
Email: j-marie.favreau@uca.fr

***Abstract*—Accurate Land Surface Temperature (LST) estimation is crucial for understanding environmental dynamics and addressing diverse scientific and societal challenges. This study explores a novel approach for LST estimation using RGB data and integrating it with additional data modalities. By leveraging conditioned Generative Adversarial Networks (cGANs) for LST generation on adjacent tiles, certain artifacts are observed. To address this issue, we introduce a comprehensive processing pipeline and tiling strategy, evaluating fusion methods for artifact removal and improving LST generation accuracy. Our findings demonstrate the potential of integrating multi-modal data to enhance LST estimation, leading to promising advancements, particularly in LST data inference.**

***Index Terms*—Land Surface Temperature estimation, cGANs, fusion methods, multi-modality, tiling strategy, artifact removal**

I. Introduction

Land surface temperature (LST) estimation is crucial for understanding global and regional environmental conditions. LST is closely linked to critical Earth system variables such as water vapor content, soil moisture, and evapotranspiration [1]. Accurate estimation of LST is essential for assessing Urban Heat Island (UHI) effects, where urban areas experience elevated temperatures compared to rural surroundings due to human activities and land surface modifications. These effects have significant implications for urban planning, public health, and climate-resilient strategies [2]–[4]. Traditionally, LST estimation encompasses a spectrum of approaches, including field measurements, satellite observations, and model simulations [5]. Field measurements, also referred to as in-situ measurements, involve direct observations of LST at specific locations [6]. While providing accurate data, this method is limited by its spatial coverage and accessibility, particularly in remote or densely populated areas. Additionally, gaps in the data may occur due to factors such as equipment malfunctions or adverse weather conditions [7]. Satellite observations have become indispensable for their global coverage and consistent monitoring capabilities. Sensors onboard satellites, including EOS/MODIS, NOAA/AVHRR, and FY/VISSR, offer LST products with acceptable temporal and spatial resolutions [8]. However, these satellite-based observations confront challenges such as coarse spatial resolution, typically ranging from tens of meters to kilometers, which may limit their ability to capture fine-scale variations in LST. Furthermore, the estimation algorithms used, such as split-window algorithms and single-channel algorithms, often incorporate auxiliary variables like surface emissivity, altitude, latitude, longitude, and vegetation indices such as NDVI (Normalized Difference Vegetation Index), and other indices like EVI (Enhanced Vegetation Index) and SAVI (Soil Adjusted Vegetation Index), to enhance accuracy [9]–[14]. Despite these advancements, cloud cover in satellite images can lead to missing data and reduced accuracy, particularly in urban areas where surface temperature variations are complex and influenced by factors like land cover and human activity [5]. To address these limitations, method integrating RGB-based (Red-Green-Blue) data with land cover and elevation information should be emerged as a means to supplement satellite-based estimation especially in areas where satellite observations are hindered by adverse weather conditions. This fusion of different data modalities is a holistic approach to LST estimation, aiming to overcome missing data, ultimately enhancing our understanding of surface temperature on both global and regional scales. To develop this approach, a model that translates RGB (and other modalities) to LST is needed. Image-to-image translation methods in remote sensing utilize deep learning architectures to transform images between different domains. These techniques include traditional approaches like linear regression as well as convolutional neural networks (CNNs) and Generative Adversarial Networks (GANs) [15]. Among GAN-based models, pix2pix is notable for its ability to generate realistic images with pixel-level correspondence between input and output domains [16], [17]. In remote sensing, pix2pix has been applied for tasks like RGB to NIR [18], NIR to NDVI translation [19], offering accurate results while maintaining spatial and spectral fidelity.

Pix2pix requires tiles of size (256, 256), and when using grid tiling, artifacts between generated tiles are observed.

This work is supported by the IATOAURA project of the Auvergne Rhône-Alpes Region, laureate of the Défis pour l'IA call. It is also supported by the LabEx IMU (ANR-10-LABX-0088) of Université de Lyon, within the Plan France 2030 operated by the French National Research Agency (ANR)

Consequently, the use of overlapping tiles and the fusion of generated overlapping tiles are employed to gather more information for each meter of the area. Various fusion techniques have been explored for merging overlapping images, including classical methods like average arithmetic methods (AAW) and weighted methods such as exponential distance weighting [20]and Gaussian distribution-based methods [21]. Additionally, other fusion techniques such as machine learning-based fusion methods have been investigated [22].

This study extends our previous research [23], which initially focused on RGB to LST generation, by introducing a tiling strategy and by employing fusion methods for artifact removal in LST generated tiles, thus enhancing our deep learning based LST estimation. In section II, we detail the different methods at each stage of the pipeline. Section III presents the outcomes of these methods, while section IV delves into their implications. In section V, we conclude with an outlook on future directions for this research.

II. Methodology

A. Data Collection and Preprocessing

Data collection is conducted according to the standards established in our previous work [23]. Data is collected from various sources, including RGB satellite imagery, raster Digital Elevation Model (D) data and vector Land Cover (L) data. A comprehensive preprocessing is crucial to standardize, reproject, and optimize these datasets for analysis and modeling. The preprocessing begins by standardizing the data to a uniform Coordinate Reference System (CRS), ensuring spatial consistency across all datasets. Spatial resolution is then adjusted to 1 meter for uniformity, enhancing resolution for LST data while converting other datasets like RGB imagery and (D) to match. The vector datasets (L) are then rasterized to a 1-meter resolution, enabling seamless integration into further processing and training.

B. Tiling Strategy

A tiling strategy is employed to prepare the data for deep learning model training. The dataset is partionated into 256x256 pixel tiles, representing square areas of 256 meter per side, aligning with the requirements of the model architecture. Tiles are systematically generated to cover the entire study area, ensuring comprehensive spatial coverage and consistency in size. To augment the dataset, 1/2 overlapping is introduced between adjacent tiles. This overlapping not only ensures spatial continuity but also effectively increases the sample size for model training. Additionally, a concept of super-tiles is employed. A super-tile is a grid of 4.5 tiles. Its size is (1152,1152). Super-tiles are specifically utilized for the testing phase, with 90 strategically placed super-tiles spaced evenly across the study area to ensure robust model generalization. The remaining tiles are dedicated to the training phase. To maintain the independence between the training and testing datasets, overlapping tiles between super-tiles and the training area are excluded from both sets. The training dataset consists of 22,633 tiles, while the testing dataset comprises 5,760 tiles, maintaining an 80% /20% division between the two sets. The data combination phase involves the merging of tiles from different data types to create composite input tiles. The final dataset consisted of two main combinations: RGB tiles, containing only spectral information, and RGBDL tiles, integrating both spectral and terrain features. Notably, the RGB and RGBDL tiles have sizes of (3, 256, 256) and (5, 256, 256) respectively.

C. Model Training and Evaluation

1) Architecture: The architecture utilized is a conditional Generative Adversarial Network (cGANs), pix2pix. The generator component of the model is based on the U-Net architecture, renowned for its effectiveness in various image-to-image translation tasks. The U-Net architecture consists of an encoder-decoder structure with skip connections, allowing for efficient feature extraction and preservation of spatial information. The encoder part of the U-Net architecture comprises a series of convolutional layers followed by max-pooling layers, progressively reducing the spatial dimensions of the input image while extracting hierarchical features. The decoder part of the architecture consists of up-sampling layers combined with skip connections from the corresponding layers in the encoder, enabling localization and reconstruction of features. The discriminator component of the pix2pix model comprises conventional layers, evaluating pairs of combined tiles and LST images to discern plausible transformations. Through adversarial training, the generator aims to produce LST images that closely match real ones.

2) Configuration: The Adam optimizer is employed with a learning rate of 0.0002 and a batch size of 1. The training loss function combines mean absolute error and cross-entropy loss. Over 50 epochs, the model iteratively learns and refines from the data.

3) Evaluation Metrics: After the training phase, generated super-tiles are constructed and evaluation metrics are computed. Constructing the generated tiles involved multiple methods, including simply merging adjacent tiles with four possible combinations. The construction process begins by merging tiles in groups of four adjacent tiles and then completing the remaining pixels at the edges with the last generated tiles at the edges. The edges can be abbreviated as follows: Bottom Left (BL), Top Left (TL), Top Right (TR), and Bottom Right (BR). RMSE, PSNR, and SSIM are computed to evaluate the model's performance in the test partition. Specifically, these metrics are calculated between each pair of super-tiles, representing the generated super-tiles and the corresponding ground truth. Subsequently, the mean value of each metric across all pairs is calculated. Notably, the RMSE input values range from [0,1], while the SSIM values are normalized from [-1,1] to [0,1]. An SSIM value of 0 indicates no similarity between two images, whereas an SSIM value of 1 indicates perfect similarity.

D. Super-tile Fusion

Various fusion methods for the super-tiles fusion are employed, including average, exponential distance, Gaussian distribution, and a novel method specific for our case. Each

method will be tested and compared both qualitatively and quantitatively to evaluate their effectiveness. In this section, method is proposed for fusing super-tiles generated from overlapping tiles. The fusion process involves computing the pixel values of the super-tile at position (i, j) by summing the weighted contributions from individual tiles within the overlapping zone. A local coordinate system, $(O_1, \vec{O_1O_4}, \vec{O_1O_2})$, is assigned to the overlapping region (Fig. 1). The weight assigned to each tile is determined based on its relative position within the overlapping region $O_1O_2O_3O_4$. $V^{(i,j)}$ denote the value of the pixel P (Fig. 1) at position (i, j) in the the super-tile. The pixel value is computed as follows:

$$V^{(i,j)} = \sum_{T \in C^{(i,j)}} w_T \cdot V_T \tag{1}$$

where $C^{(i,j)}$ is the set of all overlapping tiles at position (i, j). Each point P within the super-tile may fall within an overlapping zone, where one tile (at the corners), two tiles (along the edges of the super-tile), or four tiles (in the remainder of the super-tile) are overlapping. $C^{(i,j)}$ is defined as:

$$C^{(i,j)} = \{T_k : k \in [1, 2, 3, 4], \text{ Tile centered on } O_k \text{ and overlap the position} (i, j)\} \tag{2}$$

V_T represents the pixel value of tile T, and w_T is the weight assigned to tile T. The weights w_T are determined using coefficients α_T and computed using the following formula:

$$w_T = \frac{\alpha_T}{\sum_{T' \in C^{(i,j)}} \alpha_{T'}} \tag{3}$$

where α_T represents the coefficient associated with tile T, and the sum in the denominator is over all tiles within the overlapping zone $C^{(i,j)}$. Additionally, the sum of weights w_T equals 1, ensuring that the weighted contributions are normalized:

$$\sum_{T \in C^{(i,j)}} w_T = 1 \tag{4}$$

The coefficients α_T are determined based on the relative position of the pixel within the overlapping zone, which is defined with respect to the local coordinate system. These coefficients correspond to the areas of the rectangles formed by point P with the the centers of overlapping tiles. Specifically, we define the following coefficients:

$$\alpha_{T_1} = (1 - a)(1 - b) \tag{5}$$

$$\alpha_{T_2} = b(1 - a) \tag{6}$$

$$\alpha_{T_3} = ab \tag{7}$$

$$\alpha_{T_4} = a(1 - b) \tag{8}$$

where a and b $(0 \leq a, b \leq 1)$ represent the coordinates of the pixel within the local coordinate system. The proposed super-tiles fusion method leverages coefficients α to compute weights w assigned to individual tiles based on their contributions to the super-tile pixel values. This approach ensures spatial integrity and minimizes artifacts.

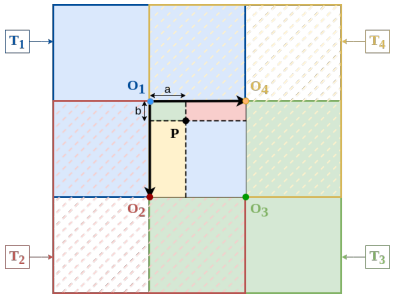


Fig. 1: Proposed fusion method: The final value of each pixel P in the overlapping zone $O_1O_2O_3O_4$ is a combination of values predicted by the four overlapping tiles, weighted by the corresponding coefficient.

III. Results

A. Generation and Construction of LST Super-tiles

Fig. 2 illustrates the construction of super-tiles using adjacent generated tiles, without fusion. The completion of super-tiles is achieved by filling in the missing pixels with the generated tiles at the edges, specifically at the bottom left (BL), bottom right (BR), top left (TL) and top right (TR) corners. Based solely on the RGB modality, the generated super-tiles exhibit a good construction across all construction methods, closely resembling the ground truth. This observation is supported by quantitative metrics (TABLE I), which indicate low values for RMSE and high values for PSNR and SSIM. However, noticeable artifacts are present in this construction. Upon incorporating terrain elevation data and land cover information, improvements are observed both quantitatively, especially in terms of RMSE, and visually, as depicted in the accompanying figure. Note in particular that the lower river temperature is better captured by the RGBDL-based prediction than by the RGB-only estimation. The addition of these data enhances the construction, resulting in super-tiles that more accurately reflect the real LST. Nonetheless, artifacts are still prominent between the generated tiles, prompting an investigation into the source of these artifacts and how errors propagate across all tiles.

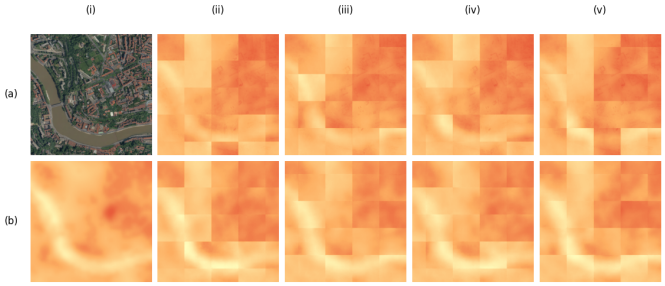


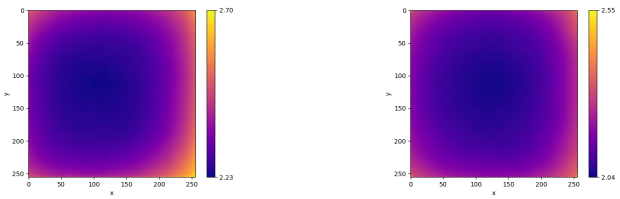
Fig. 2: Example of a constructed super-tile, (i) represents the corresponding RGB (a) and the expected target (b); (ii) to (v) are constructed super-tiles with BR, TL, BL and TR construction methods with respectively RGB (a) and RGBDL modalities (b).

TABLE I: Comparison of super-tiles construction methods.

Modalities	Method	RMSE	PSNR	SSIM
RGB	**BL**	0.071	**24.663**	**0.893**
	BR	0.071	24.619	**0.893**
	TL	0.071	24.624	0.892
	TR	0.071	24.602	0.892
RGBDL	**BL**	**0.065**	**25.052**	**0.896**
	BR	0.066	24.915	0.895
	TL	**0.065**	25.034	**0.896**
	TR	0.066	24.924	**0.896**

B. Error Localization

To analyze error localization, a tile of mean absolute errors is calculated, where each pixel represents the average absolute error between the generated pixel and its corresponding ground truth pixel. The resulting error tiles for both combinations are presented in Fig. 3, revealing that errors are not localized uniformly across the entire tile surface. Our models tend to make more errors at the edges of tiles compared to the center. This can be explained by the model losing information from adjacent tiles at the edges, while the center benefits from more surrounding information. This finding underscores the importance of constructing test super-tiles using all overlapping tiles but in a more context-appropriate manner. Traditional methods of fusing overlapping tiles, such as averaging, assign equal weights to all pixels at a given position, which may not adequately address error localization characteristics (i.e., from the center to the edges). Consequently, we plan to explore alternative fusion methods from the state of the art that assign weights to pixel values based on the proximity of their tile centers to better address error localization patterns.



(a) (b)

Fig. 3: Mean absolute error (in °C) across tiles for (a) RGB based model and (b) RGBDL based model

C. Super-tiles Fusion

Several fusion methods are employed to merge overlapping tiles. For each pixel in the overlapping region, its value is a combination of the pixel values from the overlapping tiles at that position. TABLE II illustrates the evaluation metrics. The fusion methods include the arithmetic average weighted method (AAW), exponential distance weighted method (EDW), truncated Gaussian function based method (TGF), and the proposed method described in the previous section. The comparison highlights the superiority of fusion techniques in improving image quality over (BL) construction method. Among fusion methods, those like AAW, EDW, TGF, and the proposed approach show enhanced fidelity to ground truth images, quantitatively. However, artifact occurrence varies, with AAW and EDW methods exhibiting artifacts, while TGF and the proposed method effectively suppress artifacts, leading to smoother image reconstructions (Fig. 4).

TABLE II: Comparison of super-tiles construction and fusion methods.

Modalities	Method	RMSE	PSNR	SSIM
RGB	**BL**	0.071	24.663	0.893
	AAW	0.068	25.187	0.895
	EDW	0.069	25.002	0.895
	TGF	**0.067**	25.192	0.897
	proposed	**0.067**	**25.22**	**0.898**
RGBDL	**BL**	0.065	25.052	0.896
	AAW	**0.062**	**25.578**	0.898
	EDW	0.063	25.35	0.899
	TGF	**0.062**	25.535	**0.901**
	proposed	**0.062**	25.565	**0.901**

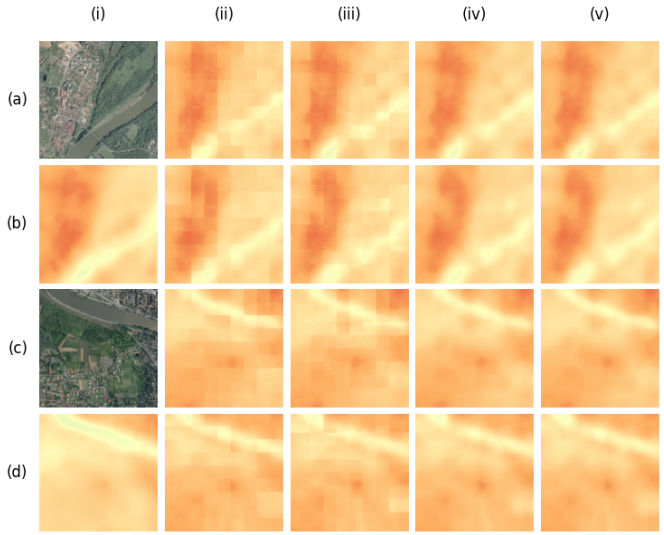


Fig. 4: Qualitative comparison of fusion methods on generated super-tiles, (i) are RGB ((a);(c)) and super-tiles groundtruth ((b);(d)), (ii) to (v) are super-tiles after fusion methods respectivly with AAW, EDW, TGF and the proposed method through RGB based model ((a);(c)) and RGBDL based model ((b);(d)).

IV. Discussion

The construction of the generated super-tiles using nonoverlapping tiles, with many tiles overlapping at the borders, resulted in a satisfactory level of fidelity to the ground truth super-tiles. However, clear artifacts are observed, especially when using RGB data alone, with some improvement when incorporating RGBDL information. These artifacts are attributed to the nonuniform error localization along the generated tiles, a known issue in U-Net networks due to kernel constraints and the loss of information at the tile edges compared to the center. To address this issue, a post-processing method involving the fusion of all overlapping generated tiles is employed. Fusion methods demonstrated an enhancement in similarity to the ground truth compared to construction methods relying solely on adjacent tiles. While the Average Arithmetic Weighting

(AAW) method achieved quantitative similarity, it does not yield qualitative similarity due to artifact retention. Results from the AAW method underscored the effectiveness of the Truncated Gaussian Fusion (TGF) method and the proposed method, which assigns greater weight to pixel values from the nearest tiles' centers. Utilizing RGBDL modalities with TGF and the proposed method proved to be the optimal choice for generating super-tiles that are both similar and faithful to real data. This is attributed to the need to model additional data beyond RGB, such as elevation and land cover, which are essential factors influencing real-world land surface temperature (LST). Furthermore, fusion methods leverage a greater number of generated tiles, providing more information to guide towards the true value, while assigning greater weight to the most influential tiles, thereby mitigating artifacts. Overall, these experimental iterations have led to improvements in LST estimation from RGB data alone and underscore the potential of leveraging additional data modalities, such as elevation and land cover, to further enhance results and facilitate meaningful LST data imputation for research and societal applications. While the estimation of LST using RGB and RGBDL data has shown promising results, there are avenues for further improvement. We aim to expand this work to incorporate temporal aspects, which are crucial for LST estimation. Additionally, increasing the dataset size to better align with the test and training sets, with a more frequent overlap of 1/2 tile, is another area we plan to explore to enhance our results. These considerations will contribute to refining our methodology and advancing the accuracy of LST estimation in future iterations of our research.

V. Conclusion

This study highlights the effectiveness of employing cGANs in generating Land Surface Temperature (LST) estimates using RGB data and its integration with additional modalities. By leveraging cGANs for LST generation on adjacent tiles, certain artifacts are observed, prompting the introduction of a comprehensive tiling strategy to address this issue. Through the evaluation of fusion methods for artifact removal and enhancing LST generation accuracy, our findings emphasize the potential of integrating multi-modal data to improve LST estimation. These advancements hold promise for addressing environmental challenges and advancing climate research. Future research directions may involve expanding temporal dimensions to further refine LST estimation techniques.

References

[1] H. Jabbar, M. Hamoodi, and A. Al-Hameedawi, "Urban heat islands: a review of contributing factors, effects and data," *IOP Conference Series: Earth and Environmental Science*, vol. 1129, p. 012038, 01 2023.

[2] A. Kumar, V. Agarwal, L. Pal, S. Chandniha, and V. Mishra, "Effect of lst on urban heat island in varanasi city, india," *J*, vol. 4, pp. 420–429, 2021.

[3] Mas'uddin, L. Karlinasari, S. Pertiwi, and Erizal, "Urban heat island index change detection based on land surface temperature, normalized difference vegetation index, normalized difference built-up index: A case study," *Journal of Ecological Engineering*, vol. 24, no. 11, pp. 91–107, 2023.

[4] M. Labib, A. Wibowo, and I. Shidiq, "Lst-based threshold method for detecting uhi in a complex urban landscape," *IOP Conference Series: Earth and Environmental Science*, vol. 986, no. 1, Feb. 2022.

[5] P. Yu, T. Zhao, J. Shi, Y. Ran, L. Jia, D. Ji, and H. Xue, "Global spatiotemporally continuous modis land surface temperature dataset," *Scientific Data*, vol. 9, 12 2022.

[6] J. Liu, D. F. T. Hagan, and Y. Liu, "Global land surface temperature change (2003–2017) and its relationship with climate drivers: Airs, modis, and era5-land based analysis," *Remote Sensing*, vol. 13, no. 1, p. 44, 2020.

[7] S. Shiff, D. Helman, and I. M. Lensky, "Worldwide continuous gap-filled modis land surface temperature dataset," *Scientific Data*, vol. 8, no. 1, p. 74, 2021.

[8] Z. Wan, Y. Zhang, Q. Zhang, and Z.-l. Li, "Validation of the land-surface temperature products retrieved from terra moderate resolution imaging spectroradiometer data," *Remote sensing of Environment*, vol. 83, no. 1-2, pp. 163–180, 2002.

[9] P. Käfer, S. Rolim, L. Ribeiro Diaz, N. Souza da Rocha, M. Iglesias, and F. Rex, "Comparative analysis of split-window and single-channel algorithms for land surface temperature retrieval of a pseudo-invariant target," *Boletim de Ciências Geodésicas*, vol. 26, 01 2020.

[10] L. Wang, Y. Lu, and Y. Yao, "Comparison of three algorithms for the retrieval of lst from landsat 8 images," *Sensors*, 2019.

[11] X. Wang, L. Zhong, and Y. Ma, "Estimation of 30m lsts over the entire tibetan plateau based on landsat-7 etm+ data and machine learning methods," *International Journal of Digital Earth*, 2022.

[12] M. Kim, D. Kim, and G. Kim, "Examining the relationship between land use/land cover (lulc) and land surface temperature (lst) using explainable artificial intelligence (xai) models: A case study of seoul, south korea," *International Journal of Environmental Research and Public Health*, vol. 19, no. 23, 2022.

[13] A. Sekertekin, "Validation of physical radiative transfer equation-based land surface temperature using landsat 8 satellite imagery and surfrad in-situ measurements," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 196, p. 105161, 2019.

[14] W. Liu, J. Shi, S. Liang, S. Zhou, and J. Cheng, "Simultaneous retrieval of land surface temperature and emissivity from the fengyun-4a advanced geosynchronous radiation imager," *International Journal of Digital Earth*, vol. 15, no. 1, pp. 198–225, 2022.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, 06 2014.

[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[17] J. Henry, T. Natalie, and D. Madsen, "Pix2pix gan for image-to-image translation," 08 2021.

[18] A. Akagic, E. Buza, and M. Horvat, "Mapping rgb-to-nir with pix2pix image-to-image translation for fire detection applications," in *34th Central European Conference on Information and Intelligent Systems*, 09 2023.

[19] C. Davidson, V. Jaganathan, A. N. Sivakumar, J. M. P. Czarnecki, and G. Chowdhary, "Ndvi/ndre prediction from standard rgb aerial imagery using deep learning," *Computers and Electronics in Agriculture*, vol. 203, p. 107396, 2022.

[20] L. Wu, S. Hu, and C. Liu, "Exponential-distance weights for reducing grid-like artifacts in patch-based medical image registration," *Sensors*, vol. 21, no. 21, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/21/7112

[21] Y. Xu, S. Hu, and Y. Du, "Research on optimization scheme for blocking artifacts after patch-based medical image reconstruction," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–17, 07 2022.

[22] D. Müller and F. Kramer, "Miscnn: a framework for medical image segmentation with convolutional neural networks and deep learning," *BMC Medical Imaging*, vol. 21, 01 2021.

[23] I. Khedher, J.-M. Favreau, S. Miguet, and G. Gesquière, "A Multimodal Deep Learning Approach for High-Resolution Land Surface Temperature Estimation," in *7th International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS'23)*, ser. Lecture Notes in Networks and Systems, PES University. Bangalore, India: Springer Verlag, Dec. 2023.