



HAL
open science

A Multimodal Deep Learning Approach for High-Resolution Land Surface Temperature Estimation

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

► To cite this version:

Issam Khedher, Jean-Marie Favreau, Serge Miguet, Gilles Gesquière. A Multimodal Deep Learning Approach for High-Resolution Land Surface Temperature Estimation. 7th International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS'23), PES University, Dec 2023, Bangalore, India. <10.1007/978-981-97-2004-0_26>. <hal-04460000>

HAL Id: hal-04460000

<https://hal.science/hal-04460000v1>

Submitted on 15 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A Multimodal Deep Learning Approach for High-Resolution Land Surface Temperature Estimation

Issam Khedher¹, Jean-Marie Favreau², Serge Miguet¹, and Gilles Gesquière¹

¹ Univ Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, Centrale Lyon, LIRIS,
UMR5205, F-69676 Bron, France,

`issam.khedher@univ-lyon2.fr`,

² Université Clermont Auvergne, CNRS, Mines de Saint-Étienne, LIMOS, F-63000
Clermont-Ferrand, France

Abstract. Urban Heat Islands (UHI), characterized by elevated temperatures, present important challenges to sustainability. This study introduces a novel multimodal approach for high-resolution Land Surface Temperature (LST) estimation, a critical component in addressing UHI. The methodology initially employs RGB orthophotography for LST estimation and progressively integrates additional relevant variables correlated with LST, including elevation and land cover. Leveraging conditional Generative Adversarial Networks (cGANs), LST maps are generated, enabling informed urban planning. Experimental results highlight the potential of this multimodal approach, emphasizing that the combination of all data variables yields the most favorable outcomes. These findings advance UHI research and support data-driven urban climate management.

Keywords: conditional Generative Adversarial Networks (cGANs), Multimodal Approach, Land Surface Temperature (LST) Estimation, Urban Heat Islands (UHI)

1 Introduction

UHI persist as a prominent concern in the realms of urban planning and environmental science. These localized zones, characterized by elevated temperatures within urban environments compared to their surroundings, exert a profound influence on energy consumption, air quality, and public health [4] [10] [28]. In light of these challenges, the accurate estimation of LST assumes paramount importance, given its central role in UHI characterization and urban climate studies [14] [16] [18].

LST estimation relies on various key parameters that influence its spatial distribution. Among these, RGB imagery offers valuable insights into urban characteristics, such as buildings, roads, and green spaces. This visual data aids in understanding how these features influence LST patterns. For example, densely built-up areas with extensive impervious surfaces tend to absorb and re-emit

heat, contributing to elevated LST values, while water bodies and vegetation act as cooling agents by reducing surface temperatures through evaporation and shading [22]. Furthermore, RGB channels, combined with other satellite channels, are pivotal for deriving and quantifying specific biophysical indices closely linked to LST, providing deeper insights into the relationships between urban characteristics and temperature variations [13]. These biophysical indices like the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Built-up Index (NDBI) are instrumental in assessing the impact of vegetation density and urbanization on LST. Higher NDVI values indicate denser vegetation, associated with cooler temperatures due to shading and transpiration, whereas elevated NDBI values signify intensified urban heat effects [6] [13]. Furthermore, authors in [12] [20] confirm that Digital Elevation Models (referred to as (D) in this paper) provide topographic information, as elevation significantly impacts temperature patterns; higher altitudes generally exhibit cooler temperatures, while slope steepness affects heat dispersion and accumulation. Among the key parameters that influence the spatial distribution of LST estimation, land cover types are critical determinants, encompassing a range of surface materials, such as impervious urban areas, vegetation, water bodies, and barren [8] [11] [31]. These land cover categories exhibit distinct thermal properties, resulting in significant LST variations [23] [24] [29]. These interrelated parameters emphasize the importance of integrating multi-source data, including RGB, D, and land cover, to enhance the accuracy of LST estimation models, a crucial aspect addressed in this study.

Various methods have been developed to estimate LST using satellite data. [25] involves the use of radiative transfer models, such as the Planck-Blackbody equation, which leverages thermal infrared (TIR) bands from multispectral sensors to calculate LST. Additionally, in [17], temperature-emissivity separation (TES) methods are employed to separate surface temperature from emissivity, offering more accurate results. Furthermore, mono-window algorithm (MWA), split window algorithm (SWA), and single-channel (SC) make use of different TIR bands to correct for atmospheric effects [15] [26]. Advanced methods incorporate ancillary data like (D) to correct for topographic influences, and machine learning algorithms are increasingly applied for their ability to capture complex spatial and spectral relationships in multispectral imagery [27] [30]. While these methods provide valuable insights into surface temperature dynamics, they are not without limitations. These LST estimation methods often face challenges. They may struggle to account for fine-scale surface heterogeneity, leading to inaccuracies in areas with complex land cover patterns. They also rely on detailed knowledge of sensor calibration and atmospheric parameters, which can be a source of uncertainty. Furthermore, different satellite platforms exhibit trade-offs between data frequency and spatial resolution. Some satellites provide frequent data but at a moderate spatial resolution, while others offer higher spatial resolution with less frequent data. Moreover, these satellites often encounter challenges such as missing data. In such cases, a robust model bases

its imputation of missing LST data on available information, thereby addressing the limitations associated with satellite data characteristics.

GANs [5] empower the acquisition of profound insights by implicitly modeling complex, high-dimensional data distributions using two competitive submodels: the generator and the discriminator. This capability extends to conditional Generative Adversarial Networks (cGANs), wherein the generator and discriminator are conditioned on external information, enabling precise image generation. The emergence of cGANs, specially pix2pix [9], has seen them excel in numerous image-to-image translation tasks, including RGB/(NDVI and NDRE) translation [3], RGB/NIR translation [2], and Areal/Map transformations [7]. This impressive track record naturally prompts the question: given a sufficiently extensive training dataset, can cGANs proficiently translate multimodal data into LST maps?

The paper is structured into four sections that (at least partially) answers the question. Section 2 will delve into the intricacies of the methodology, providing a breakdown of the data pipeline and system architecture. In section 3, the findings will be presented, offering both qualitative and quantitative insights. Section 4 will be dedicated to the in-depth discussion of the results. Finally, in section 5, we will draw this paper to a close, summarizing the conclusions and outlining potential avenues for future research.

2 Methodology

In this section, we provide a detailed description of the methodology followed to estimate LST from diverse data modalities and sources, including raster RGB imagery, raster Digital Elevation Model (D) data, vector Corine Land Cover (C) data, and additional vector Land Cover (L) data. The approach is based on a four-step pipeline designed to maximize the utilization of both vector and raster information from these heterogeneous data sources while ensuring spatial consistency. An overview of this pipeline is presented below (see Fig. 1).

2.1 Data collection

LST Imagery. Incorporated within the dataset is the LST surface imagery sourced from [1], depicted through a gradient of colors that represent varying degrees of warmth and coolness. This raster dataset is monochannel, with each pixel representing a 30-meter section of terrain, corresponding to the spatial resolution of the source data from Landsat 8. The snapshot was captured on July 4, 2015, at 10:20 UTC (12:20 local time). The dataset was subjected to correction procedures to mitigate atmospheric, topographic, and emissivity effects, ensuring the accuracy of the temperature readings. This LST dataset provides insights into the warmest surface areas as observed from the sky during daylight hours. In densely populated areas, the LST primarily reflects the temperature of rooftops, while in wider thoroughfares, it encompasses temperature information from roadways and other surfaces. The LST imagery can be visualized in Fig. 2.

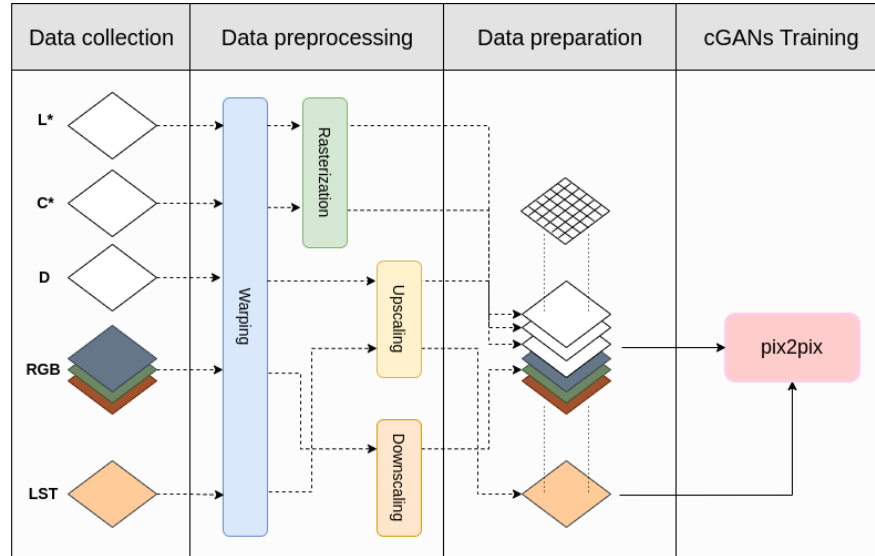


Fig. 1: Overall pipeline. LST: Land surface temperature; R: Red; G: Green; B: Blue; D: Digital Elevation Model; C: Corine Land Cover; L: Land Cover Data.
*Vector data

RGB Imagery. The RGB imagery used is the 'Orthophotography 2015 of the Lyon Metropolis' dataset sourced from [1]. This raster dataset covers an extensive area of 1399 km², including the Lyon region and its surroundings. The assembly table within the data source enabled the precise localization of 64 tiles, each covering 25 km², according to the RGF93 CC46 grid, with a pixel resolution of 0.08 m. The 'Orthophotography 2015 of the Lyon Metropolis' dataset are acquired through aerial imaging operations using a digital camera in May 2015. It is georeferenced using the EPSG:4171 coordinate system, ensuring accurate spatial alignment for subsequent analysis. It's noteworthy that RGB orthophotography data, exemplified by the 'Orthophotography 2015 of the Lyon Metropolis' dataset, is readily accessible, providing a solid rationale for its incorporation into this study.

Digital Elevation Model (D). The (D) dataset employed is provided in the Raster TIFF format in [1] and is generated from raw data sources, including slope break lines and ground-level reference points. The pixel size of this dataset is 10 meters, offering a detailed representation of the terrain's elevation characteristics. The accessibility and the availability of elevation information provides a robust foundation for the integration of such data into our analysis.

Corine Land Cover dataset (C). Corine Land Cover (C) data, available as vector data in [1], provides comprehensive land cover information for the

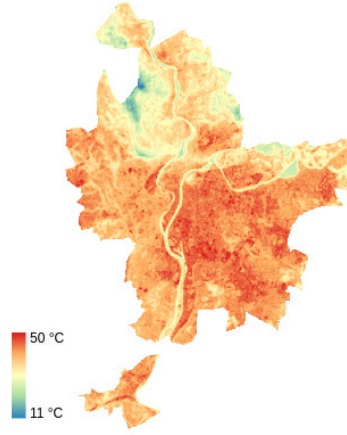


Fig. 2: False-Color LST Visualization of the Lyon Metropolitan Area in July 2015

metropolitan area, categorized into 25 main classes aligned with Corine Land Cover level 1. These categories include artificial land, agricultural land, forests, semi-natural areas, wetlands, and water bodies. C data enriches the understanding of vegetation in urban and peri-urban environments. Data processing involves photo-interpretation using 2015 orthophotography with an 0.08 m pixel resolution, infrared imagery, and 2015 LIDAR data. C data is also integrated with selected vector data, such as Grand-Lyon cadastre (buildings and parcels), to enhance urban context understanding. Notably, C data achieves a spatial precision of 1 hectare (locally lower for continuity preservation), compared to the 25-hectare granularity of the standard Corine Land Cover dataset. It is noteworthy that the widespread availability of Corine Land Cover data, as exemplified by its presence beyond specific regional datasets such as the one for Lyon Metropolis in this study, underscores its importance. It is noteworthy that this ubiquity makes Corine Land Cover data a valuable asset for integration into diverse analytical contexts.

Land Cover data (L). The Land Cover Data (L), available as vector data in [1], offers a comprehensive perspective on the land use, classifying it into 45 distinct classes. This fine-grained data serve as a complementary source of information to enhance the understanding of land use patterns in urban and peri-urban environments. It's worth noting that the Land Cover Data achieves a spatial precision level, with a Minimum Mapping Unit (MMU) of 100 m² and a Minimum Mapping Width (MMW) of 5 meters, maintaining continuity of at least 1 meter.

2.2 Data preprocessing

The dataset employed in this study comprises a diverse array of geospatial information, encompassing both vector and raster data with distinct characteristics. These datasets exhibit variations in coordinate systems, projection methods, and spatial resolutions. To harness their full potential for analysis and modeling, a comprehensive data preprocessing stage is imperative. This section elucidates the preprocessing procedures undertaken to harmonize, reproject, and optimize the disparate datasets, rendering them compatible and ready for subsequent analysis and model training.

Warping. The first step in this preprocessing involves transforming the data into a uniform Coordinate Reference System (CRS). The chosen CRS for this purpose is the EPSG:3946 - RGF93 / CC46, characterized by its utilization of the Lambert Conformal Conic projection method. This transformation harmonizes the spatial references across all datasets.

Scaling for Resolution Uniformity. In order to ensure consistent data handling, a preprocessing step involves both downscaling and upscaling of the various datasets to achieve a uniform spatial resolution of 1 meter. 1 meter is considered appropriate as it aligns with the resolution interval present in our dataset. The resolution of LST data is enhanced. Simultaneously, the spatial resolution of the RGB imagery is reduced to the same 1-meter resolution, while the (D) underwent an upscaling process. These scaling operations were performed using the GDAL library, specifically the `gdal translate` tool, with cubic spline interpolation to maintain spatial fidelity.

Rasterization. To enhance their utilization within the cGANs model, rasterization of vector datasets: (C) and (L) are conducted. Both datasets were transformed into raster format with a uniform resolution of 1 meter. During this process, each class from (C) (25 categories) and (L) (45 distinct classes) is mapped to pixel values within the $[0, 255]$ range. The rasterization procedure is executed using the `gdal rasterize` tool. This conversion ensures that the vector data can be employed for further processing and training.

2.3 Data preparation

Data is standardized to share common characteristics, encompassing a raster data type, uniform spatial projection, and an identical spatial resolution. These measures have set the stage for the subsequent Data Preparation phase, where we focus on preparing the datasets specifically for the cGANs-based training of the model.

Step 1: Regular Grid Generation. This step involves the application of a consistent grid to all data layers. This process results in the creation of monochannel tiles, each with dimensions of (1, 256, 256). These monochannel tiles collectively cover the entire study area without overlap, forming a systematic grid. The retiling phase are executed using gdal retile tool.

Step 2: Data Combination. The monochromatic tiles are joined to obtain multichannel tiles of size (H, 256, 256) using Python. H varies from 3 to 6, depending on which data layers being integrated. Data combinations explored in this study encompass RGB, DCL, RGBD, RGBC, RGBL, RGBDC, RGBDL, RGBCL, RGBDCL. These diverse data combinations enable to investigate how the integration of different information sources impacts the accuracy of the pix2pix cGANs model for LST estimation.

2.4 cGANs Training

Architecture overview. The pix2pix model employs a conditional Generative Adversarial Network (cGANs) architecture. The generator is based on the U-Net architecture, which takes combined tiles as input and generates LST output images. The discriminator, composed of conventional layers, assesses pairs of combined tiles and LST images to distinguish plausible transformations. Through adversarial training, the generator aims to produce LST images that are indistinguishable from real ones, enhancing the overall image generation quality.

Areas of experiments: Urban zone and extended zone. These studies are conducted in two zones: an urban zone and an extended zone encompassing the urban environment and its surroundings. Beginning in the urban zone allows a progressive complexity assessment, ensuring model reliability before extending it. It also aids in addressing urban-specific challenges and implementing data augmentation for urban environments, enhancing overall performance.

Training configuration. In the urban zone, the data is composed of 400 tiles, with 80% (320 tiles) allocation for training and the remaining 20% (80 tiles) for testing. In the extended zone, we follow a similar 80% (6,500 tiles) training and 20% (1,623 tiles) testing split, totaling 8,123 tiles. A consistency is maintained in hyperparameters. The Adam optimizer, known for handling complex neural networks, is employed. The learning rate is set at 0.0002 and a batch size of 1 is used. The training loss function combines mean absolute error and cross-entropy loss, capturing both generation and classification aspects, following the same approach as the original pix2pix paper. Over 50 epochs, our model iteratively learns and refines from the data.

Evaluation Metrics. Root Mean Square Error (RMSE) is the evaluation metric employed to assess the performance of the model.

RMSE quantifies the average difference between the predicted LST images and their corresponding ground truth LST images. Specifically, the RMSE is calculated between every pair of predicted and ground truth images and then compute the mean RMSE across all pairs. RMSE is calculated using the following formula:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - G_i)^2} \quad (1)$$

where N represents the total number of pixels in the image, (P_i) denotes the pixel value of the predicted LST image, (G_i) denotes the pixel value of the corresponding ground truth LST image.

3 Results

Fig. 3 and Fig. 4 depict the evolution of mean RMSE over epochs. During the training phase, the model is saved after each epoch, employing these saved models for test image inference in test set. RMSE is calculated between the generated and ground truth images and the mean RMSE is computed for each model. Fig. 5 and Fig. 6 display the multimodal data to LST translation results for six example sites using nine different data combinations for both the urban and extended zones, respectively.

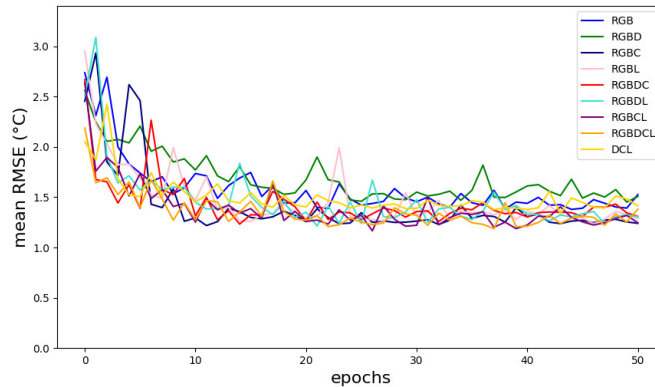


Fig. 3: mean RMSE in the urban zone

3.1 Urban zone

The initial input modality to the model is RGB alone, given its wide availability and applicability to various regions. The mean RMSE hovers around 1.5°C. Visual comparison with ground-truth LST imagery reveals low performance when

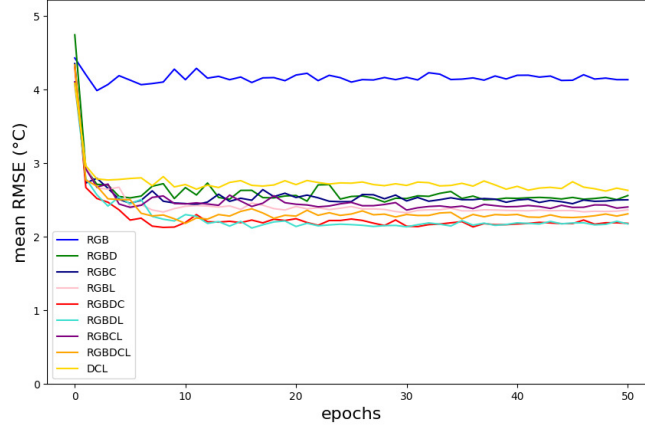


Fig. 4: mean RMSE in the extended zone

using only RGB data, as the generated LST imagery does not align well with the ground truth. Subsequently, (D) modality is introduced due to its common availability. Unfortunately, (D) has an adverse effect when merged with RGB. The RGBD combination yields the poorest performance. Notably, the inclusion of (D) in the majority of combinations has a negative impact on performance. This can be explained by the urban zone’s somewhat constrained terrain elevation variability. This will be further discussed in section 4. As a next step, land cover data is incorporated, which provides additional information relevant to LST. Here, performance improvements are observed, and the model exhibits proficiency in generating LST imagery. Both qualitative and quantitative analyses reveal that RGBCL and RGBDCL produce equivalent results, demonstrating the best performance, with a slight advantage leaning toward RGBCL.

3.2 Extended zone

In the extended zone analysis, the objective is to augment the dataset and introduce increased complexity by incorporating additional vegetation, slopes, and terrain variations. This expansion aims to provide more challenging conditions for the system, thus further assessing its performance. Following a similar approach as in the urban zone, the model is initially trained exclusively on RGB data, resulting in unsatisfactory performance characterized by the highest RMSE and numerous errors in the generated tiles, as evidenced by the visual examples in the figure. To enhance the model’s performance, the (D) modality is introduced to provide additional information. This inclusion yields significant improvements in results, as clearly observed in the third example, where a substantial disparity is evident between the LST generated solely from RGB data and that generated

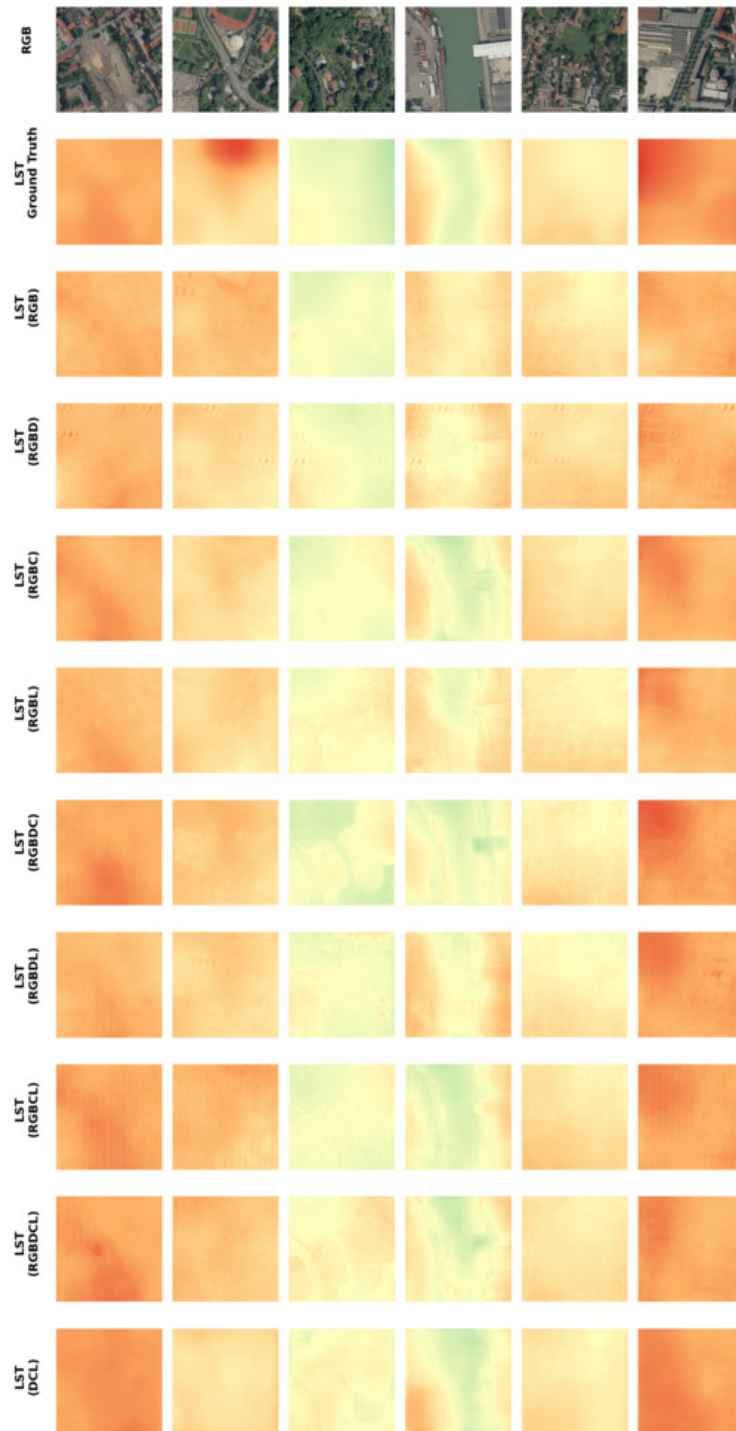


Fig. 5: Translation results in the urban zone

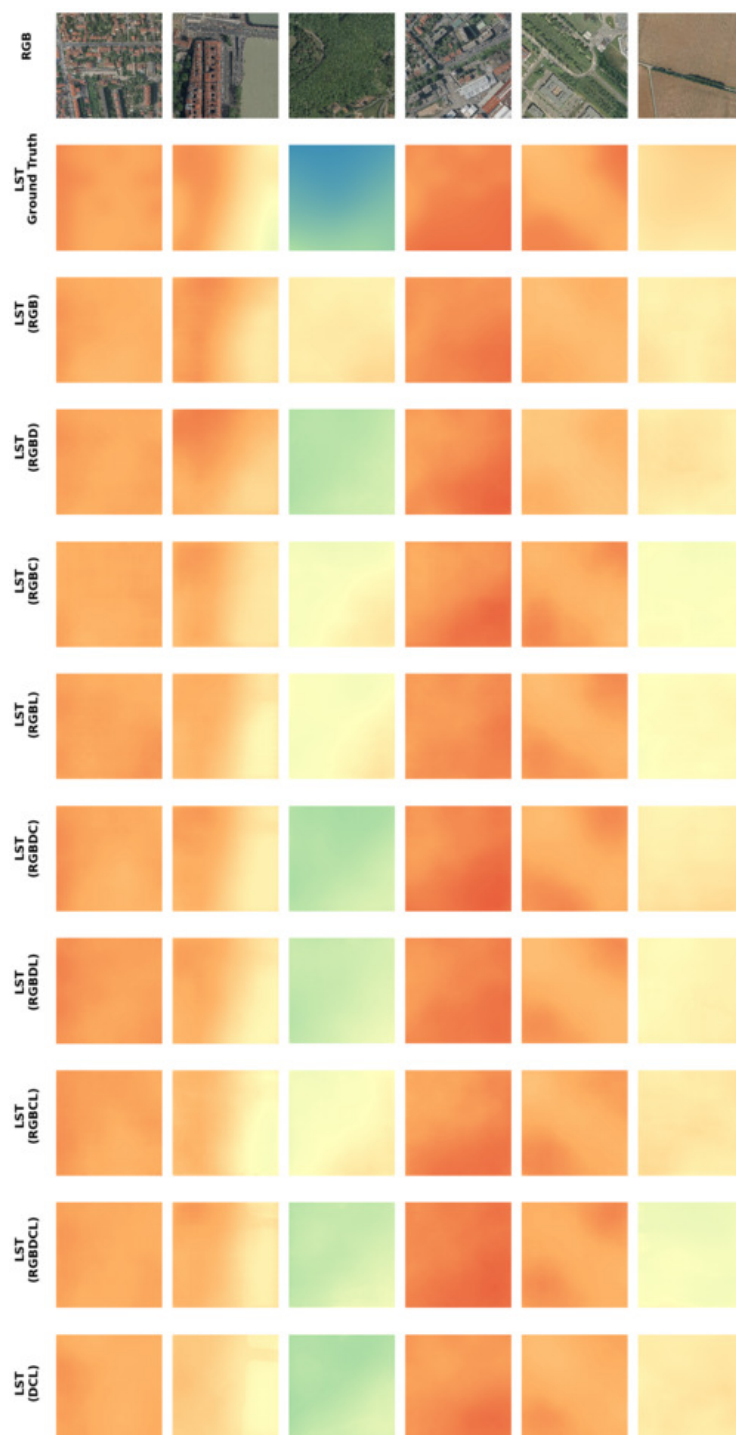


Fig. 6: Translation results in the extended zone

from RGBD. Subsequently, land cover data, (C) and (L), is integrated leading to a notable enhancement in system performance. This improvement is exemplified in the fifth example in Fig. 6, where the model accurately identified urban areas in the lower-left and upper-right corners among other vegetated regions along the diagonal, a capability not achieved with RGBD and RGB alone. Among all combinations, RGBDL and RGBDC consistently yield the best results where all three types of information are included (orthophotography, elevation and land cover).

4 Discussion

The model demonstrates robust performance in both geographic zones we have delineated. However, it is noteworthy that its performance in the urban zone surpasses that in the extended zone. This discrepancy can be attributed to several factors that warrant further elucidation. Firstly, the urban zone inherently includes less complex and varied data compared to the extended zone. Within the extended zone, we encounter instances such as non-cultivated fields, which exhibit higher temperatures, yet the model tends to estimate them as cooler. Similarly, forested areas and mountainous terrain, typically characterized by lower temperatures, may not be accurately captured by the model due to a relative scarcity of training data for such high-elevation forested regions.

The models based only on RGB data do not exhibit substantial performance, which can be attributed to the limitation of utilizing orthophotography data alone for estimating surface temperatures. An illustrative case encountered in our study is the similarity in RGB color between natural and synthetic football fields, despite a notable difference in surface temperatures.

In the urban observations, a negative impact of the elevation (D) is noticed. This can be attributed to the urban zone’s relatively limited variability in terrain elevation. Incorporating (D) to the input data, for this area, does not bring any useful information for temperature estimation, makes learning only more challenging, and contributes to diminishing the model’s ability to discern relevant patterns.

The transition to the extended zone allows to leverage the positive effect of D more effectively. In this scenario, the dataset featured a broader spectrum of elevation variations, enabling the model to better utilize the D information and understand its significance in the context of LST estimation.

The inclusion of land cover data proves to be beneficial to the model, regardless of whether it is combined with RGBD or RGB. This further underscores the importance of land cover information in the model. Consequently, when estimating surface temperature, integrating such land cover data can significantly enhance the accuracy of LST estimations. Notably, this integration allows for the differentiation between synthetic sports fields and other natural terrains, as well as distinguishing various types of vegetation and other land cover types.

Collectively, these observations underscore that the system relies on the synergy of all types of data (RGB, elevation and land cover data) to achieve accurate LST estimations.

Amongst related works employing pix2pix in various image-to-image translation tasks [19] [21] [32] and using RMSE as an evaluation metric, our study demonstrates a comparable level of performance, as evidenced by the achieved low RMSE, despite the complexity of the task.

While the study produces promising results, it is crucial to acknowledge certain limitations that may influence outcomes. The impact of data quality and resolution on results cannot be overlooked, and adaptations may be necessary when applying the model to diverse datasets. It is noteworthy that the training set consistently demonstrates superior performance compared to the test set, although occasional artifacts were observed between adjacent generated tiles. In future iterations of this work, addressing these challenges will involve implementing data augmentation methods, such as overlapping tiles, to mitigate artifacts and enhance the model’s performance across both training and test sets. Despite the correlation between RGB and other modalities with LST, it’s crucial to note that these inputs don’t provide the exact LST value. This underscores the need for a thorough understanding of the model’s capabilities and limitations, aligning with our roadmap for LST estimation. Despite these challenges, the study presents significant potential for innovative applications in LST remote sensing and temperature monitoring. Emphasizing the importance of context-awareness and the utilization of high-quality data is paramount for achieving optimal results in such endeavors.

5 Conclusion

This study investigates the utilization of conditional Generative Adversarial Networks (cGANs) to transform RGB data into LST imagery. The results firmly establish the feasibility of this data conversion task, showcasing significant improvements when integrating elevation and land cover data, thereby aligning the generated imagery more closely with ground-truth values. The uniqueness of this work lies in its reliance on RGB (and the other modalities) to generate LST images, a novel approach not explored in previous research based on our findings. Additionally, our work highlights the capability of the pix2pix network to manipulate and produce this type of imagery, a facet that has not been thoroughly tested before. Demonstrating its proficiency in this task, despite its complexity and dependency on various factors, showcases the adaptability of the pix2pix model. This contributes to the originality of our research, offering a valuable perspective on the broader applicability and versatility of pix2pix network. The implications of this research extend far beyond the confines of the study itself. They hold the potential to revolutionize various aspects of LST remote sensing applications, offering precise LST estimations and creating opportunities for applications spanning UHI analysis, climate change monitoring, and thermal anomaly detection.

6 Acknowledgments

This work was supported by the *IATO Aura* project of the *Auvergne Rhône-Alpes* Region, laureate of the *Défis pour l'IA* call. It was 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)*

References

1. Data grand lyon. <https://data.grandlyon.com>
2. Akagic, A., Buza, E., Horvat, M.: Mapping rgb-to-nir with pix2pix image-to-image translation for fire detection applications. In: 34th Central European Conference on Information and Intelligent Systems (2023)
3. Davidson, C., Jaganathan, V., Sivakumar, A.N., Czarnecki, J.M.P., Chowdhary, G.: Ndvi/ndre prediction from standard rgb aerial imagery using deep learning. *Computers and Electronics in Agriculture* **203**, 107,396 (2022)
4. European Environment Agency: Healthy environment, healthy lives: how the environment influences health and well-being in Europe (2020)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. *Advances in Neural Information Processing Systems* **3** (2014)
6. Guechi, I., Gherraz, H., Alkama, D.: Correlation analysis between biophysical indices and lst using remote sensing and gis in guelma city (algeria). *Bulletin de la Societe Royale des Sciences de Liege* **90**, 158–180 (2021)
7. Henry, J., Natalie, T., Madsen, D.: Pix2pix gan for image-to-image translation (2021)
8. Igun, E., Williams, M.: Impact of urban land cover change on lst. *Global Journal of Environmental Science and Management* (2018)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976 (2017)
10. Jabbar, H., Hamoodi, M., Al-Hameedawi, A.: Urban heat islands: a review of contributing factors, effects and data. *IOP Conference Series: Earth and Environmental Science* **1129**, 012,038 (2023)
11. Kesgin Atak, B.: Analysing the relationships between land use/land cover and urban land surface temperature using regression tree in İzmir. *International Journal of Geography and Geography Education* (41), 280 – 291 (2020)
12. Khandelwal, S., Goyal, R., Kaul, N., Mathew, A.: Assessment of land surface temperature variation due to change in elevation of area surrounding jaipur, india. *The Egyptian Journal of Remote Sensing and Space Science* **21**(1), 87–94 (2018)
13. Kim, M., Kim, D., Kim, G.: 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* **19**(23) (2022)
14. Kumar, A., Agarwal, V., Pal, L., Chandniha, S., Mishra, V.: Effect of lst on urban heat island in varanasi city, india. *J* **4**, 420–429 (2021)
15. Käfer, P., Rolim, S., Ribeiro Diaz, L., Souza da Rocha, N., Iglesias, M., Rex, F.: Comparative analysis of split-window and single-channel algorithms for land surface temperature retrieval of a pseudo-invariant target. *Boletim de Ciências Geodésicas* **26** (2020)

16. Labib, M., Wibowo, A., Shidiq, I.: Lst-based threshold method for detecting uhi in a complex urban landscape. *IOP Conference Series: Earth and Environmental Science* **986**(1) (2022)
17. Liu, W., Shi, J., Liang, S., Zhou, S., Cheng, J.: Simultaneous retrieval of land surface temperature and emissivity from the fengyun-4a advanced geosynchronous radiation imager. *International Journal of Digital Earth* **15**(1), 198–225 (2022)
18. Mas’uddin, Karlinasari, L., Pertiwi, S., 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* **24**(11), 91–107 (2023)
19. Mustafa, H.: Generating synthetic x-rays using generative adversarial networks (2020)
20. Peng, X., Wu, W., Zheng, Y., Sun, J., Hu, T., Wang, P.: Correlation analysis of land surface temperature and topographic elements in hangzhou, china. *Scientific Reports* **10**(1), 10,451 (2020)
21. Rahadiani, L., Azizah, A.Y., Deborah, H.: Evaluation of the quality indicators in dehazed images: Color, contrast, naturalness, and visual pleasingness. *Heliyon* **7**(9), e08,038 (2021)
22. Rahman, M.N., et al.: Impact of urbanization on urban heat island intensity in major districts of bangladesh using remote sensing and geo-spatial tools. *Climate* (2022)
23. Safariah, R., Majid, M.R., Rusli, N.: The contribution of housing area to lst in bandung. *Journal of Sustainability Science and Management* (2022)
24. Schwaab, J., Meier, R., Mussetti, G., Seneviratne, S., Bürgi, C., Davin, E.L.: The role of urban trees in reducing land surface temperatures in european cities. *Nature Communications* **12**(1), 6763 (2021)
25. Sekertekin, A.: 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* **196**, 105,161 (2019)
26. Wang, L., Lu, Y., Yao, Y.: Comparison of three algorithms for the retrieval of lst from landsat 8 images. *Sensors* (2019)
27. Wang, X., Zhong, L., Ma, Y.: 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)
28. World Bank: Analysis of heat waves and urban heat island effects in central european cities and implications for urban planning. Washington, D.C. (2020)
29. Xian, G., Shi, H., Auch, R., Gallo, K., Zhou, Q., Wu, Z., Kolian, M.: The effects of urban land cover dynamics on urban heat island intensity and temporal trends. *GIScience & Remote Sensing* **58**(4), 501–515 (2021)
30. Xu, S., Zhao, Q., Yin, K., He, G., Zhang, Z., Wang, G., Wen, M., Zhang, N.: Spatial downscaling of land surface temperature based on a multi-factor geographically weighted machine learning model. *Remote Sensing* **13**(6) (2021)
31. Yin, S., Liu, J., Han, Z.: Relationship between urban morphology and land surface temperature—a case study of nanjing city. *PLOS ONE* **17**(2), 1–17 (2022)
32. Zhou, S., Wang, Y., Jia, W., Wang, M., Wu, Y., Qiao, R., Wu, Z.: Automatic responsive-generation of 3d urban morphology coupled with local climate zones using generative adversarial network. *Building and Environment* **245**, 110,855 (2023)