

DIFFUSION MODEL FOR DAILY HIGH-RESOLUTION URBAN LAND SURFACE TEMPERATURE

ERIC BRUNE & YIFANG BAN

DIVISION OF GEOINFORMATICS KTH ROYAL INSTITUTE OF TECHNOLOGY

EBrune@kth.se, Yifang@kth.se

ABSTRACT

The Urban Heat Island (UHI) phenomenon, where metropolitan areas are significantly warmer than their rural surroundings, poses a substantial threat to public health and environmental quality. Accurate and timely monitoring of Land Surface Temperature (LST) is essential for understanding and mitigating UHI effects. Satellite remote sensing provides an important tool for this purpose, but existing systems present a fundamental trade-off. Sensors like the Visible Infrared Imaging Radiometer Suite (VIIRS) offer daily global coverage, enabling high-frequency temporal monitoring, but their coarse spatial resolution (375–750 m) is insufficient to resolve the fine-scale thermal variations within complex urban landscapes [1]. Conversely, the Sustainable Development Science Satellite 1 (SDGSAT-1) Thermal Infrared Spectrometer (TIS) provides high-resolution thermal data (30 m) capable of distinguishing temperature differences between individual city blocks, parks, and infrastructure, but its revisit time precludes daily observation. This spatiotemporal gap hinders the development of effective, dynamic UHI monitoring and response strategies [2]. This research addresses this challenge by developing a deep learning framework to generate daily, high-resolution (30 m) LST maps. We propose a conditional diffusion model, a type of generative model capable of generating high-fidelity imagery. The model learns to fuse frequent, low-resolution thermal data with static, high-resolution contextual information. Specifically, it uses the daily VIIRS I5 thermal band (375 m) and the daily VIIRS Land Surface Temperature product (VNP21A1D, 1 km) as daily inputs. An LST map derived from the most recent, high-resolution SDGSAT-1 TIS acquisition provides a static spatial context. The model is trained to generate a daily 30 m LST map that is consistent with the thermal conditions observed by VIIRS. Our study area comprises eleven European cities: Athens, Barcelona, Heraklion, Lisbon, Madrid, Milan, Palermo, Rome, Stockholm, Thessaloniki, and Toulouse. We trained the model on a time series of paired VIIRS and SDGSAT-1 data from these locations. We tested the model's performance using a hold-out dataset consisting of dates with coincident VIIRS and SDGSAT-1 TIS acquisitions not seen during training. The model successfully generates realistic 30 m LST maps, with metrics indicating an agreement with ground truth SDGSAT-1 TIS LST. The results demonstrate that this fusion approach can effectively bridge the spatiotemporal resolution gap. The method showcases a potential for operational, daily UHI monitoring at a scale relevant for urban planning and public health interventions.

INTRODUCTION

Cities act as thermal amplifiers, storing daytime heat in asphalt, concrete, and rooftops and releasing it slowly at night. This Urban Heat Island (UHI) effect raises temperatures by several Kelvin, worsening heat-wave mortality, stressing energy grids and causing danger to vulnerable groups. Land-surface temperature (LST) products derived from satellites are indispensable for diagnosing and mitigating UHIs, yet they suffer from a fundamental sensing dilemma: instruments that revisit every day (e.g. VIIRS) measure at coarse resolution, while imagers that resolve urban environments (e.g. SDGSAT-1 TIS, Landsat, ECOSTRESS) revisit only weekly to monthly. Conventional downscaling approaches like physical energy-balance models, or simple machine-learning regressions capture large-scale gradients but blur fine thermal textures such as parks, waterways, and narrow streets that are critical for urban-climate action. Recent diffusion-based generative models have shown that they can predict realistic high-frequency content while remaining faithful to coarse observations. Building on SR3, we fuse the daily VIIRS I5 radiance and LST with the most recent 30 m SDGSAT-1 context, producing daily 30 m LST mosaics that bridge the spatiotemporal gap and enable operational city-scale UHI analytics.

OBJECTIVE

Produce daily 30 m urban LST grids by fusing:

- VIIRS I5 (11.45 μm) radiance (375 m)
- VIIRS Land Surface Temperature (1000 m)
- Latest SDGSAT-1 TIS channels (8–12.5 μm , 30 m) converted to LST

Eleven large European cities were used for training and testing, shown in Figure 1. Each urban footprint is clipped to its city boundaries and projected to its local CRS.



Figure 1: Study area used for training and testing.

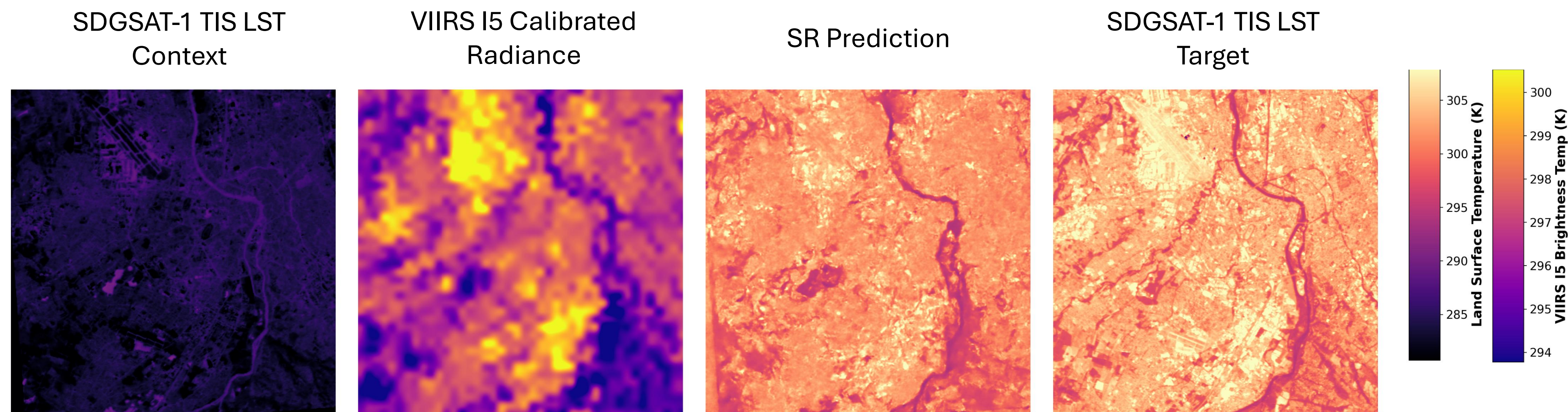


Figure 3: Model prediction of LST over Toulouse. From left to right: (a) Static LST context from a previous SDGSAT-1 acquisition. (b) Coarse daily VIIRS I5 calibrated radiance. (c) The model's 30 m super-resolved LST prediction. (d) The ground truth 30 m LST target from SDGSAT-1 for the same day. The left color bar applies to (a), (c), and (d), while the right color bar applies to (b).

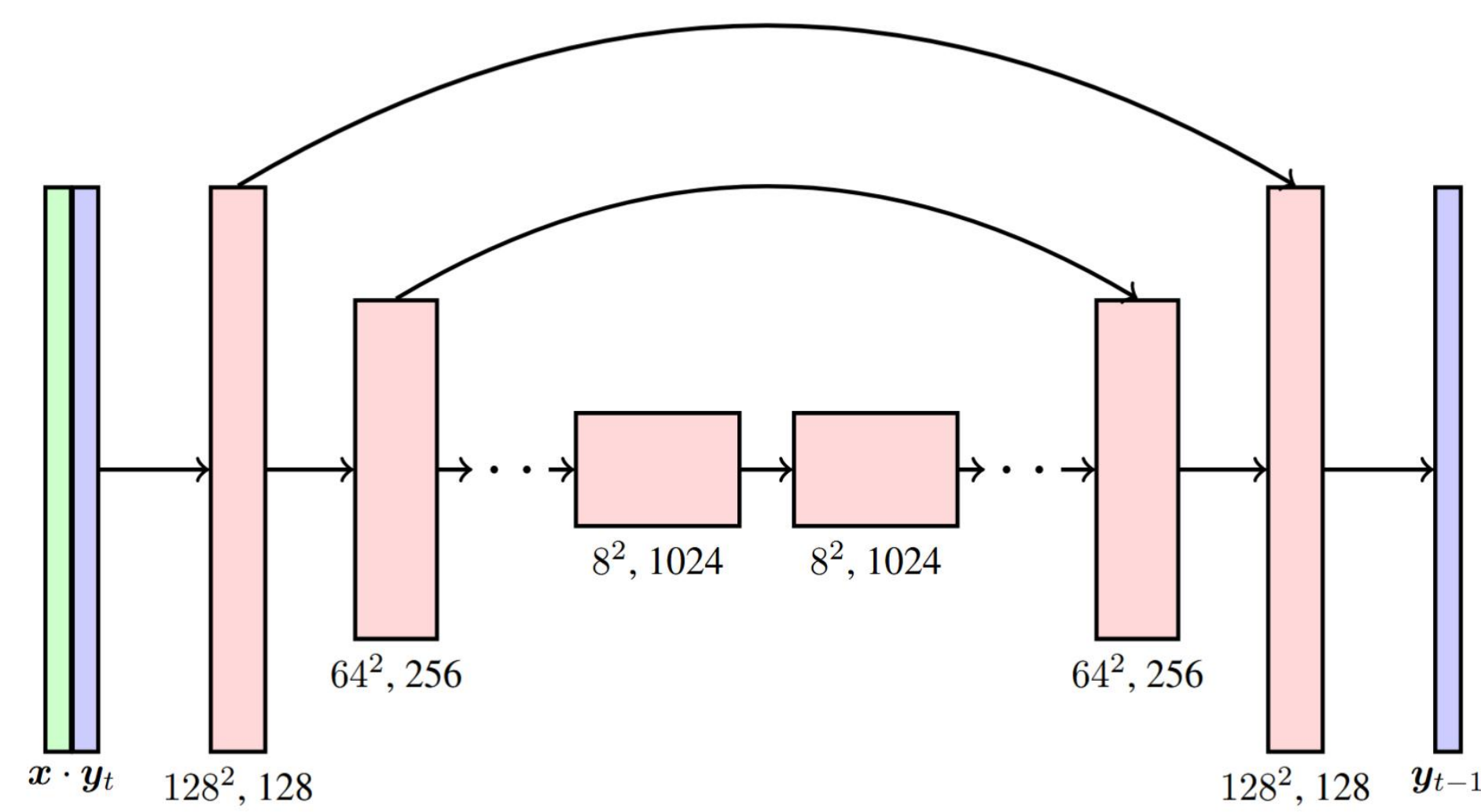


Figure 2: U-Net backbone of the SR3 diffusion model.

METHODS

We assembled 108 cloud-free SDGSAT-1 TIS scenes covering eleven European cities (sample counts: Athens 22, Barcelona 12, Heraklion 12, Lisbon 9, Madrid 7, Milan 3, Palermo 8, Rome 7, Stockholm 4, Thessaloniki 11, Toulouse 10). For every city we held out the most recent date as a test sample and the immediately previous date for temporal context. We sourced two daily VIIRS products: I-band 5 (I5) radiance from the VNP02IMG product using the LAADS DAAC and the Day Land Surface Temperature product (VNP21A1D) from Google Earth Engine [3]. For the static high-resolution context, we processed raw SDGSAT-1 TIS scenes to LST using a split-window algorithm with ERA5-derived total column water vapour for atmospheric correction [4]. The target data for training was the similarly processed SDGSAT-1 TIS LST on dates with coincident with VIIRS. All input layers were reprojected to each city's local UTM CRS. The VIIRS I5 (375 m) and VNP21A1D (1 km) were resampled to a common SDGSAT-1 TIS-native 30 m grid using cubic interpolation. A conditional diffusion model based on SR3 with a U-Net backbone, shown in Figure 2, using a three-channel tensor as conditioning: the upsampled daily VIIRS I5 radiance, the upsampled daily VNP21A1D LST, and the static SDGSAT-1 TIS LST context [5]. The forward diffusion used 1 000 time-steps with a cosine noise schedule; the reverse process was trained by minimising the L2 error between true and predicted noise. Training patches of 128×128 pixels were drawn with random cropping, and each patch was subjected to horizontal flipping with 50% probability to augment spatial diversity. We adopted the Lion optimiser with a learning rate of 10^{-5} and trained for 120 k iterations across four RTX 3080 GPUs using a batch size of 8. At inference time the model used 50 DDIM steps. The model was trained to directly predict the 30 m LST of the target date. Validation employed cloud-free SDGSAT-1 TIS LST scenes from dates withheld from the training set. We calculated root mean-squared error (RMSE) and mean absolute error (MAE) against the ground truth SDGSAT-1 TIS LST. The schematic U-Net illustrates the backbone that performs a single reverse-diffusion step, where the network outputs the mean μ of a Gaussian to move from y_t to y_{t-1} . The three equations below define the reverse diffusion process: (i) a Markov factorisation of $p_\theta(y_{0:T}|x)$ that denoises one step at a time given the conditioning tensor x (upsampled VIIRS I5 radiance, upsampled VIIRS LST and static SDGSAT-1 context); (ii) an isotropic standard-normal prior $p(\gamma_T)$ matching the fully-noised data distribution; and (iii) a Gaussian transition $p_\theta(y_{t-1}|y_t, x)$ whose mean is predicted by a U-Net from (x, y_t, γ_t) and whose variance follows the cosine noise schedule, trained via an L2 loss on the added noise:

$$\begin{aligned} p_\theta(y_{0:T}|x) &= p(y_T) \prod_{t=1}^T p_\theta(y_{t-1}|y_t, x) \\ p(y_T) &= \mathcal{N}(y_T | \mathbf{0}, I) \\ p_\theta(y_{t-1}|y_t, x) &= \mathcal{N}(y_{t-1} | \mu_\theta(x, y_t, \gamma_t), \sigma_t^2 I) \end{aligned}$$

RESULTS

Table 1: Mean Absolute Error and Root Mean Squared Error on the testing images.

City	MAE (K)	RMSE (K)
Athens	1.35	1.63
Barcelona	5.45	6.14
Heraklion	3.7	3.77
Lisbon	5.62	5.89
Madrid	5.96	6.53
Milan	6.9	7.79
Palermo	4.34	4.91
Rome	1.51	1.87
Stockholm	2.56	3.33
Thessaloniki	5.19	5.37
Toulouse	2.57	3.38
Mean	4.10	4.60

DISCUSSION

Diffusion-based super-resolution reconstructs intra-urban thermal structure at 30 m with relatively high accuracy, shown in Table 1. The resulting errors vary, with cities like Rome and Athens showing the highest accuracy. Milan exhibits the highest error, reflecting the single acquisition available for training. Figure 3 show that the diffusion model reconstructs sharp river corridors and urban hotspots that are indistinguishable in the original 375 m data. The model benefits from the static, high-resolution LST map provided by SDGSAT-1 TIS, which guides the downscaling of the daily VIIRS thermal patterns. Long skip connections propagate coarse context, while conditional inputs of VIIRS and SDGSAT-1 TIS guide the synthesis of fine-grained textures. During inference, only 50 DDIM steps are necessary, making daily heat-wave surveillance and near-real-time urban-climate analytics computationally feasible.

CONCLUSIONS

The SR3-based diffusion model is able to generate daily 30 m LST mosaics for European cities. Testing on a temporal split indicates below 8 K RMSE and 7 K MAE for all eleven cities. The method enhances heat-wave surveillance and urban-climate studies with a product unattainable from any single sensor. Future work will extend the training set, refine the preprocessing of input data, and integrate novel architecture and sampling methods for faster inference.

MAJOR REFERENCES

- Liu Y., Yu P., Wang H., Peng J., Yu Y., “Ten Years of VIIRS Land Surface Temperature Product Validation,” Remote Sensing, 2022.
- Guo H., Dou C., Chen H., Liu J., Fu B., Li X., Zou Z., Liang D., “SDGSAT-1: The World’s First Scientific Satellite for Sustainable Development Goals,” Science Bulletin, 2023.
- Joint Polar Satellite System (JPSS), VIIRS Land-Surface Temperature (VNP21) Algorithm Technical Basis Document, Version 2.0, NOAA/NESDIS, 2024.
- Teng Y., Ren H., Hu Y., Dou C., “Land-Surface Temperature Retrieval from SDGSAT-1 Thermal Infrared Spectrometer Images: Algorithm and Validation,” Remote Sensing of Environment, 2024.
- Saharia C., Ho J., Chan W., Salimans T., Fleet D. J., Norouzi M., “Image Super-Resolution via Iterative Refinement,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.

ACKNOWLEDGMENTS

This work was conducted as part of the Dragon 6 cooperation programme between the European Space Agency (ESA) and the Ministry of Science and Technology (MOST) of China. The authors acknowledge the support and coordination provided by the Dragon Programme Office