

Optically Enhanced Super-Resolution of Sea Surface Temperature Using Deep Learning

David T. Lloyd^{ID}, Aaron Abela, *Member, IEEE*, Reuben A. Farrugia^{ID}, *Senior Member, IEEE*, Anthony Galea^{ID}, and Gianluca Valentino^{ID}, *Senior Member, IEEE*

Abstract—Sea surface temperature (SST) can be measured from space using infrared sensors on Earth-observing satellites. However, the tradeoff between spatial resolution and swath size (and hence revisit time) means that SST products derived from remote sensing measurements commonly only have a moderate resolution (>1 km). In this article, we adapt the design of a super-resolution neural network architecture [specifically very deep super-resolution (VDSR)] to enhance the resolution of both top-of-atmosphere thermal images of sea regions and bottom-of-atmosphere SST images by a factor of 5. When tested on an unseen dataset, the trained neural network yields thermal images that have an RMSE 2–3x smaller than interpolation, with a 6–9 dB improvement in PSNR. A major contribution of the proposed neural network architecture is that it fuses optical and thermal images to propagate the high-resolution information present in the optical image to the restored thermal image. To illustrate the potential benefits of using super-resolution (SR) in the context of oceanography, we present super-resolved SST images of a gyre and an ocean front, revealing details and features otherwise poorly resolved by moderate resolution satellite images.

Index Terms—Data fusion, deep learning, gyre, Landsat 8, ocean front, sea surface temperature (SST), Sentinel 3, super-resolution (SR), thermal infrared.

I. INTRODUCTION

IMAGES provided by earth-observing satellites are a vital information source for recording and understanding the global environment. Satellite sensors that record radiance in the far-infrared region have proven to be crucial for monitoring long-term changes in oceans [1] and the impact of human habitation on land regions [2], among many other applications. For long-timescale, synoptic studies, such as sea temperature records [3] or meteorology [4], instruments providing moderate spatial resolution (≈ 1 km) images with high revisit time are commonly used, for instance, MODIS on board satellites Terra and Aqua, and SLSTR on board Sentinel 3. Many oceanographic phenomena, such as fronts [5],

Manuscript received March 29, 2021; revised May 17, 2021; accepted June 18, 2021. Date of publication July 15, 2021; date of current version January 14, 2022. This work was supported by the Project SATellite data Fusion and Imaging Resolution Enhancement for coastal areas (SAT-FIRE) financed by the Malta Council for Science and Technology, for and on behalf of the Foundation for Science and Technology, through the Space Research Fund under Grant SRF-2018-1S1. (Corresponding author: Gianluca Valentino.)

David T. Lloyd, Aaron Abela, Reuben A. Farrugia, and Gianluca Valentino are with the Department of Communications and Computer Engineering, University of Malta, Msida MSD 2080, Malta (e-mail: gianluca.valentino@um.edu.mt).

Anthony Galea is with the Physical Oceanography Research Group, Department of Geosciences, University of Malta, Msida MSD 2080, Malta.

Digital Object Identifier 10.1109/TGRS.2021.3094117

filaments [6], eddies [7], and meanders [8], can have features and structures across a broad range of spatial scales, including scales much smaller than what is resolvable with moderate resolution sensors. A single remote sensing instrument must tradeoff spatial resolution with coverage (also known as swath), such that high (low) resolution instruments sample a small (large) surface area. Larger coverage can be achieved with constellations of identical satellites, such as the case of Sentinel 3 of the Copernicus constellation [9], but with the drawback of increased cost. Consequently, there is ongoing interest in enhancing the resolution of remote sensing images as a post-processing step aimed at producing comprehensive datasets with both high spatial resolution (< 1 km) and high revisit time, in order to accurately record simultaneously both large scale and small-scale phenomena [10], [11].

Super-resolution (SR) is the term used to describe the process of computationally enhancing the resolution of a digital image. The problem is ill-posed, since multiple different high resolution (HR) images can be consistent with a single low resolution (LR) image. Therefore, the task of SR is to constrain the problem such that a unique mapping can be found between an LR image and its HR counterpart, in such a way that the LR image is downsampled with high fidelity compared to the HR image ground truth. The difference in spatial resolution between LR and HR image is referred to here as “scale factor,” which is the term that is used throughout this article.

Initial work in the area of SR focused on improving the resolution of purposely degraded natural images. Among early efforts was the example-based approach for single image SR of Freeman *et al.* [12], who used a k-nearest neighbors method to find closest matching examples from pairs of LR and HR images. Further improvements in performance were obtained by using “neighborhood embedding” where the LR and HR patches are recast as feature vectors, with the SR patch derived from combining local neighboring vectors [13]. Subsequent developments focused on efficient encapsulation of the LR–HR mapping function by creating a dictionary and using sparse coding [14], [15]. In the context of thermal images from remote sensing, there has been sustained development of scene-specific methods, such as thermal-sharpening (TsHARP), where combinations of HR optical images are used to downscale LR thermal images acquired over land regions [16], [17]. Collectively, older approaches to SR that do not utilize deep learning are referred to as “classical” methods.

Starting with the work of Dong *et al.* [18], methods based on deep learning are nowadays widely studied for the problem of SR, owing to their superior performance over “classical” methods. In the case of Dong *et al.*’s work, referred to as SRCNN, a three-layer convolutional network was trained with sets of LR and HR patch pairs, producing a more accurate downscaling of “unseen” LR patches when compared to classical methods such as sparse coding. Subsequently, Kim *et al.* [19] showed that a deeper network, composed of up to 20 identical convolution layers exceeded the performance of SRCNN. Kim *et al.*’s network architecture is referred to as very deep super-resolution (VDSR) and uses gradient clipping with residual learning to reduce the time required to train the more complex network. This architecture has proved to be a popular basis for subsequent developments and extensions beyond SR of natural images [20], [21]. Ledig *et al.* [22] adopted a different approach using a generative adversarial network (GAN) for SR. GANs are composed of two networks: a generator and a discriminator network. The former acts to produce new candidate HR patches while the latter acts as a judge of the generative network output. Subsequent refinements of the GAN concept focused on modifying the architecture, the discriminative network, and perceptual loss in order to reduce the presence of noise and artifacts, and improve performance [23]. Further improvements in SR can be obtained with complex network structures, such as that proposed by Wang *et al.* [24], which they named Progressive Super-resolution (ProSr). In this case the scale factor was increased with depth within the network, yielding high SR performance for large scale factors ($\times 8$ in their case). A detailed review of deep learning for SR can be found elsewhere [25].

Among the first researchers to explore deep-learning and SR with remote sensing images were Liebel and Körner [26], who utilized the SRCNN architecture to enhance the resolution of Sentinel 2 optical images. They showed that training the network with multiple Sentinel 2 bands simultaneously leads to improved performance compared to the case of a single band input. To address the shortcoming of SRCNN, which offers diminishing benefits over interpolation for large-scale factors, Lei *et al.* [27] developed a network architecture composed of multiple forks. This innovation allows activation maps in the first few layers to be used deeper in the network, in a method they named local-global combined network (LGCNET). In doing so, their network captured representations of remote sensing optical images at multiple length scales simultaneously, improving the accuracy of downscaling.

The process of training a neural network requires a large number of LR and HR image patch pairs, with larger training datasets generally leading to better performance. Commonly, LR image patches are created by upscaling and blurring corresponding HR images [28]. However, this process is sensitive to the exact downscaling process employed and may omit other factors, such as noise. Alternatively, other studies have utilized LR images and HR images recorded by the same satellite, albeit in different bands, creating deep-learning analogs of pan-sharpening without the need for a specific panchromatic band [29], [30]. Studies investigating the use

of LR and HR images from different satellite sources are rarer, due to factors such as sensor calibration differences, viewing angle, and irregularity of coincidence acquisitions. Nevertheless, Tan *et al.* [31] tackled the problem of fusing data from two satellite sources with their work utilizing deep-learning for SR of optical images provided by Landsat 8 and MODIS. The authors noted the importance of selecting high-quality data, opting to use level 3, eight-day average MODIS products (which can cause issues as oceanographic features evolve during that time) instead of lower quality, daily level 2 product that better matched the properties of the Landsat 8 data.

In order to retrieve information about surface conditions, e.g., sea surface temperature (SST), atmospheric effects (i.e., radiation absorption and emission) need to be accounted for. Before SST retrieval, the satellite products are labeled as top of atmosphere (TOA); after SST retrieval they are labeled as the bottom of atmosphere (BOA).

There has recently been growing interest in using deep-learning to enhance the resolution of environmental parameters, with notable examples including SST [32], atmospheric properties [33]–[35] and local chlorophyll content [36]. Separately, denoising methods based upon convolutional neural networks are emerging as effective tools for improving the radiometric accuracy of the satellite images themselves [37], [38]. Previous work shows that there is clear potential for deep-learning to be utilized for enhancing the quality and utility of products derived from satellite data. However, there are very few examples of deep-learning being used to convert TOA satellite images into BOA environmental parameters in an end-to-end arrangement. This process has the potential to streamline and unify remote sensing processing pipelines, improving the timeliness and accuracy of derived products.

In this article, we explore the potential of deep learning to enhance the resolution of both TOA thermal images of sea regions and (BOA) SST images. To achieve this goal, we modify an established convolution neural network architecture so that it can be trained using multiple bands and with image patches taken from two different satellite sources. We extract LR patches from images recorded by the SLSTR instrument on board the Sentinel 3 satellite, while the HR patches come from images recorded by the TIRS instrument on board Landsat 8. Our results show that as follows.

- 1) Deep-learning can be used to downscale the resolution of Sentinel 3 TOA thermal images by a scale factor of 5, with far better accuracy than can be achieved with interpolation.
- 2) Deep learning can also be used to simultaneously downscale and perform SST retrieval on Sentinel 3 images, again with better accuracy and quality than interpolation.
- 3) A neural network can be used to extract HR information from optical bands, which can be fused with the thermal band inputs to boost the performance of SR of thermal images.

The trained networks enable the synthesis of thermal image datasets that simultaneously possess the high revisit time of Sentinel 3, with the enhanced resolution of super-resolved

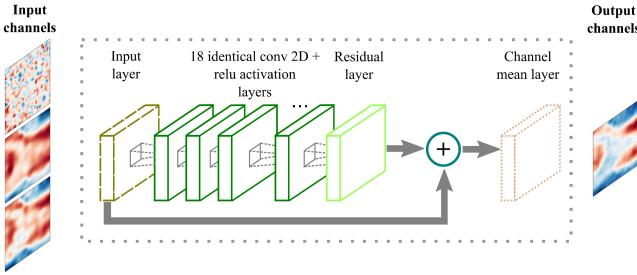


Fig. 1. Diagram depicting the modified VDSR network. Multiple possible LR input channels are represented by the three images on the left, with the single HR output channel represented on the right of the figure. The 18 2-D convolution layers common with VDSR are represented by dark green rectangles. The light green rectangle represents a residual layer, while the pink rectangle represents a layer added for our work which outputs the mean along the channel data dimension.

images, making a significant step toward fulfilling the wishes of SST product users [39].

This article is structured as follows. First, the proposed very deep convolutional neural network for fusing optical and thermal data to estimate a high-resolution thermal image is presented. Then the steps taken to create a training dataset for the neural network are described. The network performance after training on different band combinations for generation of HR brightness temperature and SST images is then reported. Finally, two case studies, a gyre and an ocean front, are investigated using super-resolved SST images, to illustrate the benefits of using SR to assist in the study of oceanographic phenomena.

II. VERY DEEP SUPER RESOLUTION ARCHITECTURE

This work is inspired by the very deep SR network first described by Kim *et al.* [19]. VDSR has the benefit of being comparatively quick to train, thanks to the use of residual learning, allowing rapid evaluation of different input parameters and conditions. Our modified version of VDSR is shown in Fig. 1. While the core of our network matches that first described by Kim *et al.*, there is one notable difference.

Since satellites typically record images in multiple bands simultaneously, and multiple bands are often combined to generate specific indices and physical parameters, it is desirable for the network to be able to use multiple LR images as an input. However, owing to the use of residual learning, the VDSR architecture was originally developed for single-channel HR and LR images. To extend the applicability of VDSR to multiple channel inputs, while maintaining a single channel output, we include an additional layer at the end of the network which outputs the mean calculated over the channel dimension of the tensor. The channel mean layer does not have any programmable parameters and was found to perform better than the case where an additional, dimension reducing, convolution layer was used instead. We attribute this to the common observation that adding more layers to a network does not always lead to improved performance [40], due to the increased risk of over-fitting. A quantitative comparison

TABLE I
SENTINEL 3 AND LANDSAT 8 THERMAL BANDS SPECTRAL COVERAGE AND IMAGE RESOLUTION. BAND S7 OF SENTINEL 3 IS OMITTED SINCE THERE IS NO EQUIVALENT BAND FOR LANDSAT 8

	Center wavelength (μm)	Bandwidth (μm)	Resolution (m)
Landsat 8 band 10	10.9	0.6	100
Sentinel 3 band S8	10.9	0.8	1000
Landsat 8 band 11	12.0	1.0	100
Sentinel 3 band S9	12.0	0.9	1000

of the two modified, multi-channel network architectures can be found in Section V.

III. TRAINING DATASET CREATION

The selection, processing, and volume of data used to train a neural network are just as important as the choices made in designing the network architecture. In this section, the process for creating training and testing datasets is described, including the criteria employed to ensure the dataset is composed of high-quality image patches. In our case, a training/testing dataset is not just a set of LR and HR patch pairs, but rather a set of patches taken from all Sentinel 3 SLSTR bands with corresponding patches from Landsat 8 bands 10 and 11, as well as an SST product derived from the two Landsat bands. In all cases, the Landsat 8 patches match the location and acquisition time (within approximately 1 h) of the Sentinel 3 patches.

Landsat 8 and Sentinel 3 occupy sun-synchronous orbits, with a local time of descending node of 10:00 AM for both satellites. This combination of sun-synchronicity and common local time means that scenes recorded by each satellite can be treated as being acquired near-simultaneously. Furthermore, Sentinel 3 and Landsat 8 record thermal infrared images with near identical spectral coverage, as summarized in Table I. Thus, with the exception of any differences in sensor calibration and quality, Sentinel 3 and Landsat 8 are expected to record similar thermal images, with the main difference being the respective image resolutions, as well as spectral responses and the zenith angle of the two satellites.

We limited the selection of patches to the region covered by four adjacent Landsat 8 tiles (rows 34 and 35 from paths 188 and 189), covering the Mediterranean sea adjacent to Sicily and Malta, and for the time period spanning 1st January 2017–31st October 2019. Owing to our interest in optical bands, daytime acquisitions are exclusively used in this study. The total volume of Landsat 8 and Sentinel 3 SLSTR data, including all bands, for this limited location and time period, amount to ≈ 550 GB of disk space, meaning that storage and processing of the data are possible on a desktop computer. The selected tiles had the benefit of each containing a landmass, therefore higher quality image registration was employed by USGS compared to tiles where no landmass is present at all. Patches of size $12 \text{ km} \times 12 \text{ km}$ (400×400 pixels) were extracted from randomly selected sea regions. The patches themselves are allowed to overlap and have a density (patches per km^2) proportional to the fraction of cloud-free, water

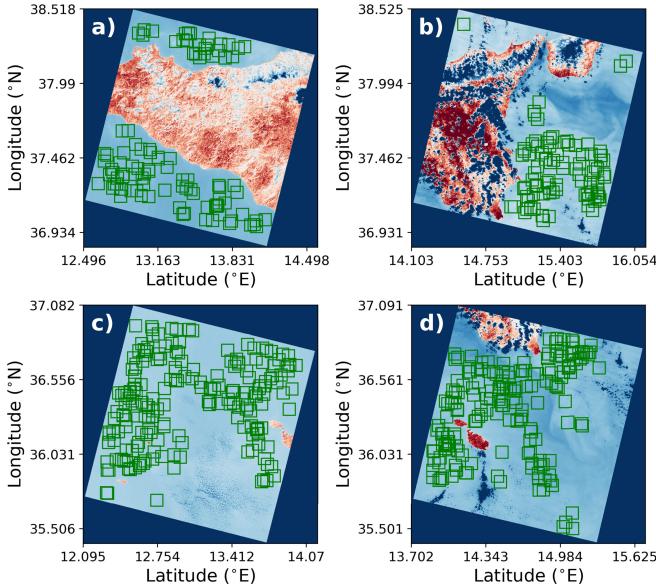


Fig. 2. Patch locations taken from four example Landsat 8 tiles. The 12km × 12km patches are denoted by green squares in the four images. (a) Path 189, row 34. (b) Path 188, row 34. (c) Path 189, row 35. (d) Path 188, row 35.

pixels present in the scene. An example of candidate patch locations selected from four representative Landsat 8 tiles is shown in Fig. 2. For each location, patches from bands 10 and 11 were saved.

The resolution of thermal images produced by Landsat 8 is officially stated as 100 m [41]. However, quantification of the edge response function from images containing coastlines has indicated that the true, on-orbit, resolution may be closer to 180–200 m [42]. As part of the data pre-processing routine, all Landsat 8 images used in this work are upscaled from a pixel size of 30 to 200 m using bicubic rescaling. In this study, both TOA and BOA images are produced from TOA Sentinel 3 images by different neural networks trained separately for the two tasks. While the process of producing gap-free level 4 SST products is complex, involving the ingestion of multiple sources of remote sensing and *in situ* measurements, level 2 SST products are generated from a single satellite source alone. In the case of Landsat 8, recent studies have explored using a split window (SW) method for converting TOA brightness temperatures into BOA SST fields [43]–[45]. We opt to replicate the approach presented by Vanhellemont [43] for SST retrieval, who in turn used coefficients first derived by Du *et al.* [46]. The method takes two TOA thermal bands as an input (in the case of Landsat 8, bands 10 and 11) and produces an SST image as an output. The column water vapor is an optional input parameter, which is specified through a choice of coefficient values used in the SW expression. SW SST retrieval has the benefit of being simple to implement and reasonably accurate for a variety of sea conditions.

Prior to application of the SW SST retrieval algorithm, the Landsat 8 TOA patches for both bands 10 and 11 undergo a series of pre-processing steps are following.

1) Digital Number to Brightness Temperature Conversion:

The pixel values in the downloaded Landsat 8 products

are converted from digital number to radiance and then from radiance to brightness temperature, using expressions and coefficient values provided in the Landsat 8 metadata. The expression for converting a digital number to brightness temperature is as follows:

$$B_T = C_4 \left[\log \left(\frac{C_3}{C_1 \times D_N + C_2} \right) \right]^{-1} \quad (1)$$

where B_T is brightness temperature in kelvin, D_N is pixel digital number, and C_{1-4} are constants supplied in the Landsat 8 product metadata.

- 2) *Destriping*: A customized, oriented, spectral filter is used to remove stripe noise from the Landsat 8 patches [47]. The filter is shaped such as a super-Gaussian function, allowing strong suppression of high-frequency stripe noise while leaving genuine, low-frequency features in the patches unmodified. The exact method used for destripping is not crucial and other methods tested on thermal sea images, such as the total variation minimization approach of Bouali and Ignatov [48], also suffice.
- 3) *Upscaling*: The patches are upscaled using bicubic rescaling to a pixel size of 200 m.
- 4) *Denoising*: The noise of thermal infrared sensors can be parameterized with the noise equivalent differential temperature ($NE\Delta T$). The pre-launch ($NE\Delta T$) for the Landsat 8 TIRS bands is ≈ 360 mK [41]. This is large compared to SLSTR ($NE\Delta T \approx 30$ mK [49]). To address this difference a standard denoising algorithm based on total-variation is used to denoise the Landsat 8 patches, while preserving the edges in the image.
- 5) *SST Retrieval*: The SW method described above is used to convert the two processed bands into a single SST image. The processed TOA brightness temperature images are also retained. The equation used is the following:

$$\begin{aligned} SST = b_0 + & \left(b_1 + b_2 \frac{1 - \bar{\epsilon}}{\bar{\epsilon}} + b_3 \frac{\Delta\epsilon}{\bar{\epsilon}^2} \right) \frac{BT_{10} + BT_{11}}{2} \\ & + \left(b_4 + b_5 \frac{1 - \bar{\epsilon}}{\bar{\epsilon}} + b_6 \frac{\Delta\epsilon}{\bar{\epsilon}^2} \right) \frac{BT_{10} - BT_{11}}{2} \\ & + b_7(BT_{10} - BT_{11})^2 \end{aligned} \quad (2)$$

where BT_{10} and BT_{11} are the Landsat 8 B10 and B11 brightness temperature bands, respectively, $\bar{\epsilon}$ is the average emissivity of the two bands, $\Delta\epsilon$ is the difference in emissivity between the two bands, and b_0-b_7 are the coefficients established by [46] for a broad atmospheric column water vapor (CWV) range shown in Table II.

The Landsat 8 pre-processing steps are depicted for example band 10 brightness temperature patch in Fig. 3.

The geo-location and date of acquisition of each Landsat 8 patch are used to then extract matching Sentinel 3 patches. Sentinel 3 SLSTR level 1-B products were downloaded in bulk from Copernicus Open Access Hub,¹ from which bands S8 and S9 were reprojected to the WGS84 projection. The grid size was fixed to a size of 1 km × 1 km

¹<https://scihub.copernicus.eu>

TABLE II

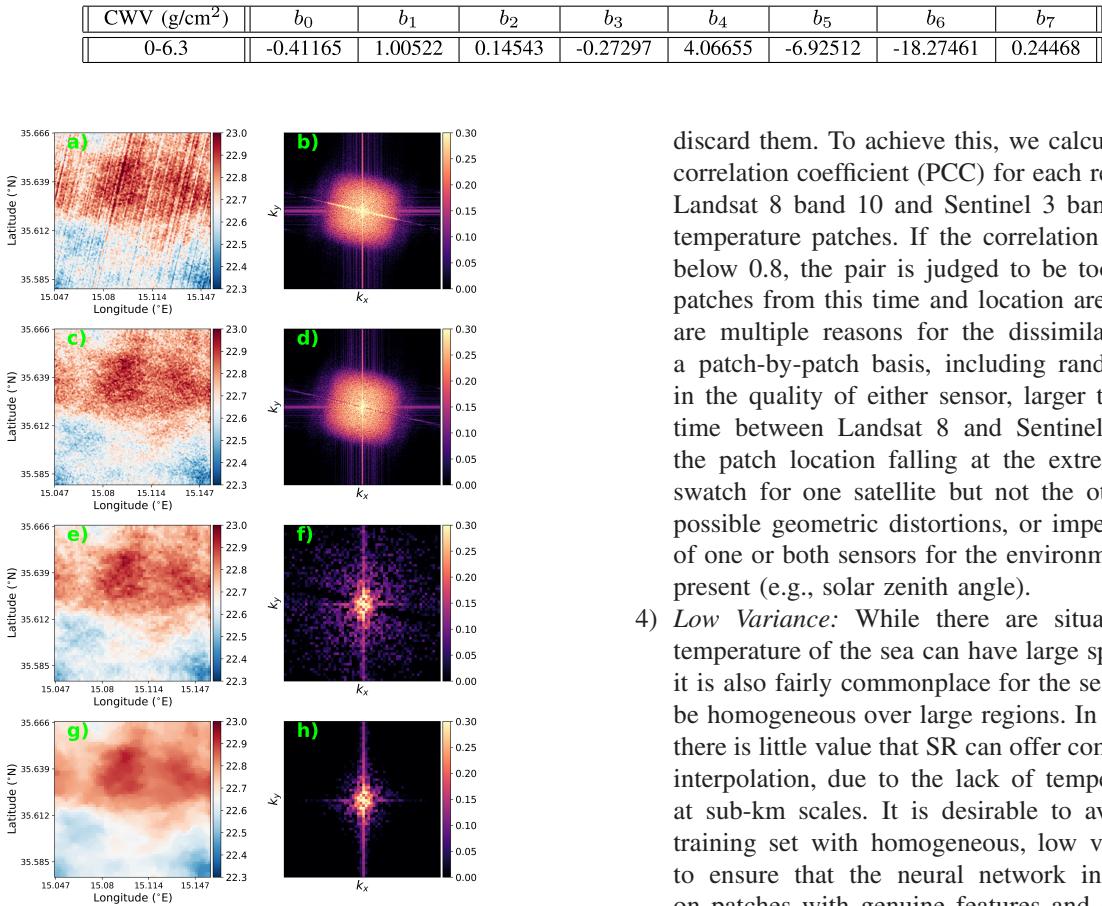
COEFFICIENTS b_k ($k = 0, 1, 2, \dots, 7$) FOR THE BROAD ATMOSPHERIC CWV RANGE USED TO RETRIEVE THE SST

Fig. 3. Landsat 8 pre-processing steps for an example band 10 patch. The patches correspond to a 12km \times 12km region in the Mediterranean sea with brightness temperature shown in $^{\circ}\text{C}$. (b), (d), (f), and (h) Correspond to the normalized magnitude of the Fourier transform of (a) patch showing clear detector striping, (c) patch desstripped using an oriented super-Gaussian filter, (e) desstripped patch upscaled by a factor of 5, and (g) desstriped, upsampled patch denoised using a total variation method.

throughout the processing. To ensure that the neural network is trained on high-quality data, several tests are performed on the matching patches from the two satellites. This process of data curation is common in deep learning tasks where data can be incomplete or vary in quality significantly. The four tests are as follows.

- 1) *Cloud Rejection*: Cloud masks provided by Landsat 8 and Sentinel 3 are both used to determine if either the Landsat 8 or Sentinel 3 patches have clouds present. If clouds are detected the patches are discarded. This step ensures that the structures and textures in the training patches are the results of surface features alone.
- 2) *Incomplete Swath Overlap*: The Landsat 8 patch location can sometimes lie partially outside of the Sentinel 3 swath. In this case the patches are discarded.
- 3) *Dissimilarity*: Since we wish to treat the Landsat 8 and Sentinel 3 thermal images as HR and LR representations of the same object, it is desirable to identify any patches where the two thermal images are dissimilar and

discard them. To achieve this, we calculate the Pearson correlation coefficient (PCC) for each remaining pair of Landsat 8 band 10 and Sentinel 3 band S8 brightness temperature patches. If the correlation coefficient falls below 0.8, the pair is judged to be too dissimilar and patches from this time and location are rejected. There are multiple reasons for the dissimilarity to vary on a patch-by-patch basis, including random degradation in the quality of either sensor, larger than the average time between Landsat 8 and Sentinel 3 acquisitions, the patch location falling at the extreme edge of the swatch for one satellite but not the other—leading to possible geometric distortions, or imperfect calibration of one or both sensors for the environmental conditions present (e.g., solar zenith angle).

- 4) *Low Variance*: While there are situations when the temperature of the sea can have large spatial variability, it is also fairly commonplace for the sea temperature to be homogeneous over large regions. In these situations, there is little value that SR can offer compared to simple interpolation, due to the lack of temperature variation at sub-km scales. It is desirable to avoid biasing the training set with homogeneous, low variance patches, to ensure that the neural network instead is trained on patches with genuine features and structures at the sub-km scale. To achieve this, we use apply a threshold based on the variance of each remaining Landsat 8 band 10 patch: if the variance is less than 150 mK, patches from that time and location are rejected.

The net result of applying the four rejection criteria described above is the reduction of candidate patch locations from 14 709 to 2076. If higher thresholds for dissimilarity and variance conditions are used, fewer patches are retained, leading to a smaller training set and likely poorer performance of the trained network. The values of 0.8 and 150 mK, for PCC and variance, respectively, therefore offer a good compromise, yielding a moderately large dataset whilst maintaining high quality.

The remaining patches (from all bands and both satellite sources) then undergo three final pre-processing steps as follows.

- 1) *Padding*: The combination of a deep network with convolutional filters can give rise to unrealistic artifacts at the edges of super-resolved images. In addition to the zero-padding layers in the standard VDSR architecture, we opt to pad all patches prior to their use for training. While this additional padding step is not described in the original VDSR publication, subsequent studies have explored image padding as a pre-processing step for VDSR [20]. In our case, each patch is padded by duplicating the outermost pixel values 10 times, such that a 60×60 pixel patch is padded to a size of 80×80 .

The padding is removed before evaluation using the testing dataset.

- 2) *Normalization*: Level 1 Sentinel 3 thermal images are provided as brightness temperature whereas optical images represent measured radiance. It is advisable to normalize training data to have zero mean and unit variance, in order to improve the accuracy and reliability of training. To satisfy this requirement while preserving the relative absolute differences in bands, we adopt a normalization scheme that adjusts the input pixel values from each band to have “close to zero” mean and “close to unit” variance. This is achieved as follows: the mean value of all Sentinel 3 band S8 patches is subtracted from all thermal images (Sentinel 3 bands S8 and S9, Landsat 8 bands 10, 11, and SST), the thermal images are then divided by the variance of all Sentinel 3 band S8 patches. This leads to the set of Sentinel 3 band S8 patches having zero mean and unit variance, while the other thermal images have “near-zero” mean and “near-unit” variance. By using the same values to normalize all bands, any relative difference between images and bands is preserved (i.e., brightness temperature patches will still be colder than SST patches from a given time and place in the new normalized units as they are before normalization). For the optical bands investigated the same process is employed but the mean and variance of Sentinel 3 band S1 are used instead. The choice of band used for normalization of thermal and optical images is not important, provided the normalization is applied consistently, as described above.
- 3) *Augmentation*: Similar to the original VDSR work, the patches are augmented, creating duplicate patches that are rotated by 90°, 180°, and 270° and flipped horizontally and vertically. This increases the number of patches by a factor of 6, from 2076 to 12456.

The 12456 preprocessed patches were randomly split, with 80% (9959 patches) allocated to the training dataset, while 20% (2485 patches) allocated to a testing dataset.

IV. BAND CORRELATION ANALYSIS

The redesigned VDSR network, described in Section II, allows an additional user choice over what combinations of bands can be used as an input for SR. In the case of the SLSTR instrument, there are 11 different bands covering the optical and infrared spectral regions, leading to many possible combinations. One of the hypotheses explored in this work is whether optical bands can provide complimentary information that can be used to improve the SR of thermal bands. The optical bands on the SLSTR instrument have a higher resolution (500 m) than the thermal bands (1 km), opening the possibility of incorporating HR features and structures in the optical image with LR temperature information in the infrared image. The strength of deep learning for this task is that a formal relationship between optical and thermal images does not need to be found. Rather, the process of training the network can be used to exploit any beneficial relations between the two different types of input images.

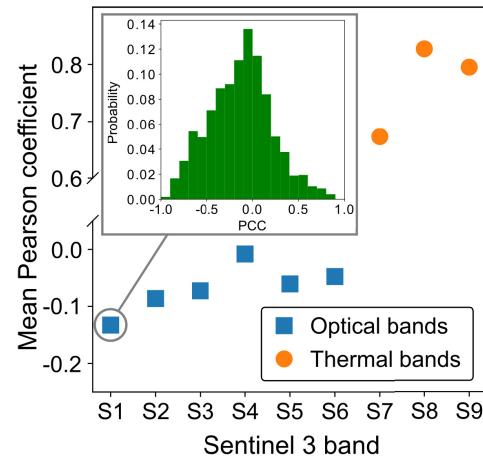


Fig. 4. Mean PCC calculated between TOA images from nine Sentinel 3 SLSTR bands and corresponding Landsat 8 SW SST images. The training set described in Section III is used for the comparison. For the optical bands, the SST image is upscaled to a pixel size of 500 m, while for the thermal bands the upscaling is to a pixel size of 1 km. In both cases, bicubic interpolation is used to perform the upscaling. The inset shows a histogram of the PCC values for the case of correlations between Sentinel 3 band S1 and Landsat 8 SST.

As a first step toward exploring such a relationship between optical and thermal images, we calculate the PCC between patches from each Sentinel 3 SLSTR band and the corresponding Landsat 8 SST patch, with the latter upscaled to match the resolution of the Sentinel 3 images. For reference, a large, positive PCC corresponds to high correlation (images have high and low valued pixel values in the same locations), whereas a large, negative PCC corresponds to a high anti-correlation (where one image has large values the other has low). A PCC of zero indicates no correlation is present. The PCC was calculated on patches taken from the training dataset, described in more detail in Section III.

In Fig. 4, the mean PCC is plotted for nine Sentinel 3 SLSTR bands. The thermal bands S7–S9 unsurprisingly have a high correlation with the Landsat 8 SST since TOA brightness temperatures are heavily influenced by the local surface temperature represented by the Landsat 8 SST patch. In this work, the task of SR Band S7 from Sentinel-3 is excluded due to the lower correlation as a result of the different wavelength compared to Landsat 8 SST. Band S4 has a near-zero correlation coefficient because it has a very poor signal-to-noise ratio owing to high atmospheric absorption, leading to few features which correlate with SST. Intriguingly, band S1 and to a lesser extent S2, have a small, negative PCC indicating weak anti-correlation between these optical bands and the SST. A histogram of the PCC between S1 TOA radiance and Landsat 8 SST images is shown in the inset of Fig. 4. For this band, of the 9959 patch pairs tested, approximately a third (3118) have a $PCC < -0.3$, suggesting that warm sea regions have a nonnegligible tendency in this data sample to have a low optical radiance, and vice versa. The exact relationship between optical radiance and SST does not need to be known in order for it to be exploited for SR, as demonstrated in the proceeding sections.

TABLE III

BRIGHTNESS TEMPERATURE SR RESULTS. THE GROUND-TRUTH IMAGES WERE PROCESSED LANDSAT 8 TOA BRIGHTNESS TEMPERATURE PATCHES FROM TIRS BAND 10. IN THE FIRST COLUMN, THERMAL BANDS ARE HIGHLIGHTED WITH RED TEXT, WHEREAS OPTICAL BANDS ARE HIGHLIGHTED IN BLUE TEXT. CNN DENOTES AN SR USING A NEURAL NETWORK TRAINED USING THE IMAGES FROM THE SENTINEL 3 BANDS STATED IN THE LEFTMOST COLUMN. BOLD FACE HIGHLIGHTS THE BEST PERFORMING METHOD FOR THE THREE METRICS

Input Bands	Downscaling method	RMSE (K)	SSIM	PSNR (dB)
S8	Bicubic interp.	0.59	0.862	21.69
S8	Lanczos interp.	0.59	0.862	21.69
S8	CNN	0.336	0.917	26.61
S8 & S9	CNN	0.272	0.957	28.03
S1 & S8	CNN	0.227	0.965	29.14
S1, S2 & S8	CNN	0.200	0.966	30.04
S2, S8 & S9	CNN	0.188	0.969	30.58
S1, S8 & S9	CNN	0.175	0.970	31.08
S1, S2, S8 & S9	CNN	0.175	0.971	31.05

V. SUPER-RESOLUTION RESULTS

A. Performance Metrics

The super-resolved images are compared to matching, HR “ground truth” Landsat 8 images using three performance metrics: root mean square error (RMSE), structural SIMilarity index (SSIM), and peak signal to noise ratio (PSNR). A low value of RMSE and high values of SSIM and PSNR are indicative of good performance. The metrics are evaluated on the testing dataset, described in Section III.

B. Downscaling Brightness Temperature

The VDSR architecture described in Section II is trained using the training dataset with the aim of producing TOA brightness temperature images that best match Landsat 8 band 10, using combinations of Sentinel 3 TOA images as an input. The resultant, downsampled pixel size is 200 m. For each combination of input bands, the neural network is trained from a random initialization 5 separate times. The performance metric values presented in Table III are the mean results after the 5 training instances, averaged also over all patches in the testing set. The network hyper-parameters are fixed throughout, with a learning rate of 0.001 and 80 epochs of training with a batch size of 64. For this combination of hyper-parameters the network loss is found to stagnate with increasing epochs for nearly all cases tested. In addition to tests using the neural network, interpolation is also investigated for downscaling Sentinel 3 brightness temperature images. Two interpolation kernels are assessed: bicubic (over a 4×4 pixel neighborhood) and a Lanczos kernel (over an 8×8 pixel neighborhood).

Several conclusions can be drawn from the results shown in Table III. In terms of the three performance metrics, there is no difference between interpolation using a bicubic or Lanczos kernel. Also, all of the neural networks tested, each trained with different band combinations, outperform interpolation by a significant margin, according to all three metrics. The results in Table III show that it is preferable to include multiple bands for SR, over just band S8 alone. This is likely because additional complementary information is provided by the additional bands, potentially providing better resilience to noise while introducing extra constraints useful

for solving the ill-posed SR problem. The results suggest that the optical bands offer the best complementarity: the combination of S1 and S8 (optical + thermal) is better than S8 and S9 (thermal + thermal). In addition, the highest performing combination overall is S1, S8, and S9. A similar combination of S2, S8, and S9 performs slightly worse. In Fig. 4, S2 has a weaker anti-correlation with SST compared to S1 with SST, suggesting that the PCC may be a good indicator for assessing complimentary information for SR. Finally, the combination of S1, S2, S8, and S9 performed no better than S1, S8, and S9, implying that the full benefit of using an optical image to enhance thermal image SR is not improved with the use of additional images from different optical bands. The explanation for the better performance of the optical + thermal as opposed to thermal + thermal is the following: when training a neural network, it is ideal if there is a correlation between the input and the output, however, if the input features (bands) are correlated between themselves, this results in redundancy which does not give an improvement in performance. On the other hand, adding complementary information from, e.g., optical bands (even though they are less correlated with the output than the thermal bands) can provide an improvement in performance.

An example Landsat 8 B10 brightness temperature patch from the testing set is shown in Fig. 5(e) along with the corresponding Sentinel 3 band S8 patch interpolated to a pixel size of 200 m in Fig. 5(g). The native resolution Sentinel-3 band S8 and Landsat 8 B10 brightness temperature patches are shown in Fig. 5(a) and (c). In Fig. 5(h) the histograms of the two patches are compared, demonstrating a small offset in brightness temperature between the two. Super-resolved patches produced by a selection of four of the neural networks are also shown in Fig. 5. It can be seen that when an optical band is included the result more closely qualitatively resembles the ground-truth image [see Fig. 5(m)], compared to the case where just a single thermal band was used [see Fig 5(i)]. With respect to the histograms of the super-resolved patches, when a single Sentinel 3 thermal band (i.e., band S8) is used to train the network, a difference in the distribution between super-resolved and ground-truth histograms is observed [see Fig. 5(j)]. The addition of an optical band improves the histogram similarity, as shown for the cases of bands S1 and S8

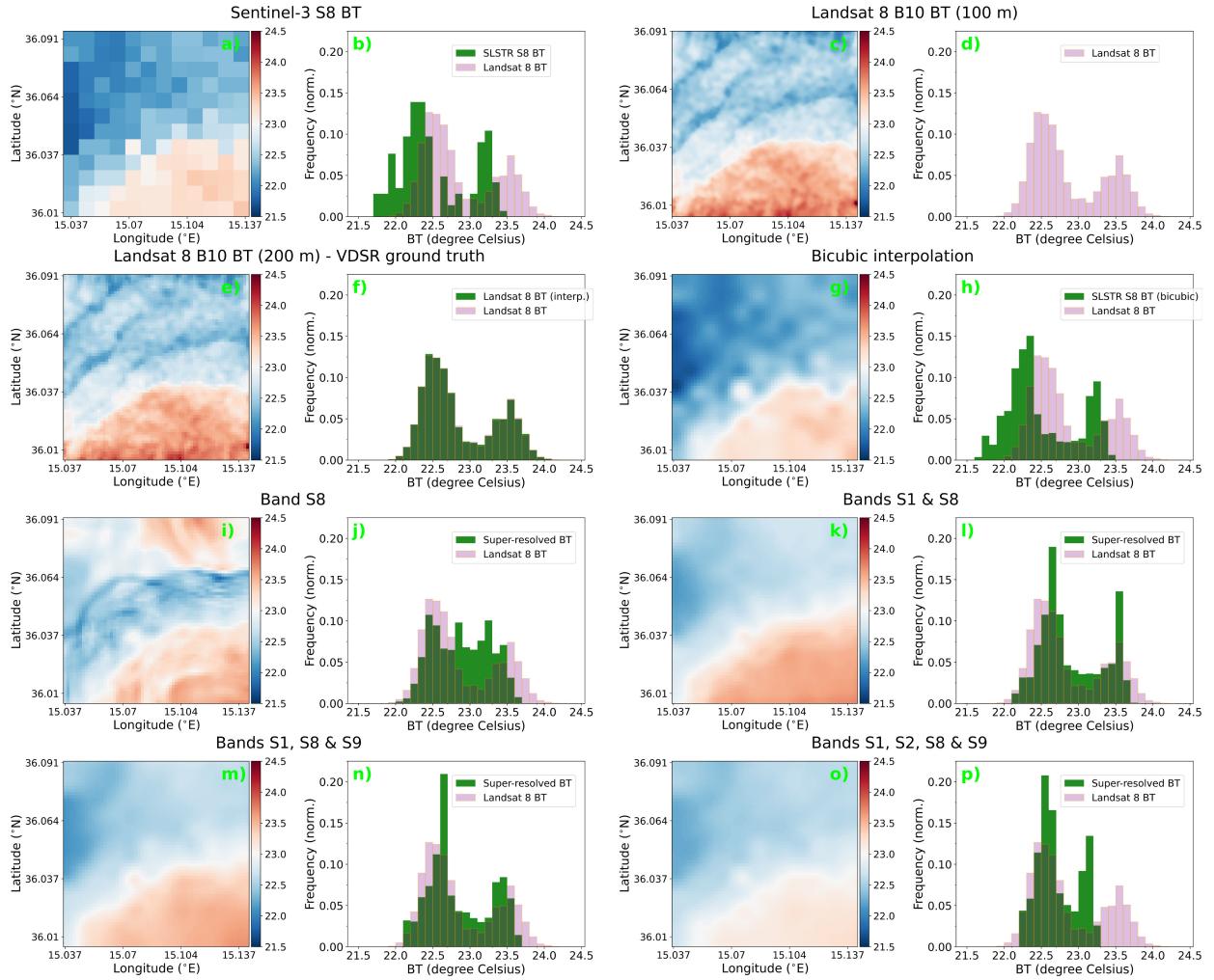


Fig. 5. Example brightness temperature patches and histograms of brightness temperature, for (a) and (b) Sentinel-3 BT data (S8), (c) and (d) Landsat 8 BT data (B10), (e) and (f) upscaled Landsat 8 B10 ground truth used to train the VDSR network, (g) and (h) bicubic interpolation, (i) and (j) neural network trained with Sentinel 3 band S8, (k) and (l) trained with Sentinel 3 bands S1 and S8, and (m) and (n) bands S1, S8, and S9, or (o) and (p) finally bands S1, S2, S8, and S9. (f) Pink histograms correspond to the ground-truth patch.

[see Fig. 5(l)]; S1, S8, and S9 [see Fig. 5(n)]; and to a lesser extent S1, S2, S8, and S9 [see Fig. 5(p)]. Together this suggests that the optical band not only introduces higher resolution features to the network but also assists in ensuring the correct brightness temperature values are reconstructed.

It should be noted that a similar experiment can be carried out for the case where images from Landsat 8 band 11 are used as ground truth, however, these results are not shown here for the sake of brevity.

C. Super-Resolved SST

The same VDSR architecture used for super-resolving brightness temperature can be re-purposed to investigate the possibility of mapping LR Sentinel 3 TOA bands to HR SST. In this situation, the network performs two functions simultaneously: downscaling (via SR) and SST retrieval (specifically, accounting for surface emissivity and the effect of the atmosphere). For network training, the same hyper-parameters are used as before. By way of comparison, HR SST patches are also produced by first interpolating (separately) images

from Sentinel 3 bands S8 and S9, then using the SW method to obtain an SST patch. While the SW SST retrieval was developed for use with Landsat 8, the close spectral similarity between Landsat 8 and Sentinel 3 thermal bands means that it is appropriate for use with Sentinel 3 bands S8 and S9 as well.

The results of these tests are summarized in Table IV. Again, the performance metric values represent the mean value, averaged over 5 random initiations of the network weights. The same combination of Sentinel 3 bands used in the brightness temperature downscaling experiments is also employed here. In absolute terms, overall performance is worse compared to brightness temperature SR, due to the more complex process being modeled by the neural network. There is again no performance difference comparing the two interpolation kernels, however, interpolation is not the worst performing overall. Using just Sentinel 3 band S8 as an input yields poorer performance than interpolation, possibly reflecting that SW methods are generally better than single-channel methods for SST retrieval [50], in part because multiple bands can enable better estimation of radiation transfer in

TABLE IV

SST SR RESULTS. IN THE FIRST COLUMN, THERMAL BANDS ARE HIGHLIGHTED WITH RED TEXT, WHEREAS OPTICAL BANDS ARE HIGHLIGHTED IN BLUE TEXT. CNN STANDS FOR RESULTS OBTAINED BY TRAINING THE VDSR CONVOLUTIONAL NEURAL NETWORK WITH THE S3 INPUT BANDS SPECIFIED. SW STANDS FOR SW, INDICATING THE TWO DOWNSCALED BRIGHTNESS TEMPERATURE BANDS WAS CONVERTED TO SST USING THE SW EXPRESSION

Input Bands	HR SST Retrieval	RMSE (K)	SSIM	PSNR (dB)
S8 & S9	Bicubic + SW	0.61	0.789	20.84
S8 & S9	Lanczos + SW	0.61	0.789	20.84
S1, S8 & S9	CNN + SW	0.419	0.861	24.02
S8	CNN	0.735	0.767	19.26
S8 & S9	CNN	0.350	0.880	25.20
S1 & S8	CNN	0.515	0.844	22.36
S1, S2 & S8	CNN	0.401	0.863	24.18
S2, S8 & S9	CNN	0.303	0.899	26.33
S1, S8 & S9	CNN	0.282	0.903	26.87
S1, S2, S8 & S9	CNN	0.305	0.898	26.29

the atmosphere. In support of this, the results obtained when two thermal bands are used (i.e., S8 and S9) are significantly better than all cases where only one thermal band is used for training. Similar to the brightness temperature downscaling results, the highest performing combination comes from using optical and thermal together; bands S1, S8, and S9 is also the best combination for simultaneous downscaling and SST retrieval. This suggests that the benefits to brightness temperature SR enjoyed by this band combination are transferred to the task of producing HR SST images. The comparatively poor performance of the combination S1, S2, and S8 indicates that in this instance two thermal bands are necessary for accurate SST estimation. Similar to the brightness temperature downscaling tests, swapping bands S1 with S2, leads to a fall in performance, and the combination S1, S2, S8, and S9 offers no benefits compared to S1, S8, and S9.

One additional comparison is offered in Table IV. The results corresponding to “CNN + SW” are produced by training two separate networks: one to produce HR TOA brightness temperatures matching Landsat 8 band 10, and the other network to produce patches with the same specification but matching band 11. For each network, the same input bands are used: S1, S8, and S9. The two downscaled brightness temperature patches that this process produces are then converted to SST using the SW method. While this approach is better than using interpolation, it is inferior to using the same architecture to convert TOA brightness temperature to SST, demonstrating that the end-to-end process of transforming LR TOA input images to HR SST images achieved by the neural network better encapsulates the task than splitting SR and SST retrieval into separate processes.

In Fig. 6, an example SST patch from the testing set is shown. The ground-truth image and histogram of SST values are shown in Fig. 6(e) and (f), respectively. The patch downsampled using bicubic interpolation is shown in Fig. 6(g), displaying uneven texture typical of interpolated images. The native resolution Sentinel-3 band S8 and Landsat 8 B10 brightness temperature patches are shown in Fig. 6(a) and (c). With respect to using a neural network for simultaneous downscaling and SST retrieval, three cases, each corresponding to a different combination of input bands, are shown. When the network was trained on just band S8 alone [see Fig. 6(k)], the output has similar texture to the case where interpolation

was used, and there is a difference when the retrieved SST values are compared to the ground truth [see Fig. 6(l)]. The situation is slightly improved with the addition of band S9 to the input [see Fig. 6(m)], especially with respect to the SST values [see Fig. 6(n)]. The best performing band combination is S1, S8, and S9. The network output for this combination is shown in Fig. 6(o), displaying a similar texture and structure compared to the ground truth, with equally similar SST values, as seen in the comparison of histograms shown in Fig. 6(p). The SST patch recovered by first downscaling bands S8 and S9 brightness temperature images, separately with different networks, and then applying SW SST retrieval is shown in fig. 6(i), displaying good similarity to the ground truth. The histograms plotted in Fig. 6(j) show that the SST values retrieved this way are similar as well. However, as highlighted earlier, the quantitative performance of this particular method averaged over the whole testing dataset is poorer than when a network is trained to retrieve HR SST patches directly, with bands S1, S8, and S9 used as inputs.

D. Data Fusion Layer Performance

In Section II, it was claimed that adding a channel mean layer to the standard VDSR network architecture provides superior performance compared to the case where an additional convolution layer is used instead. Here, the performances of the two network architectures are compared. To achieve a reduction in the number of channels the convolution layer must have one filter with a kernel size equal to the number of input channels. The same hyper-parameters used in Section V were also used here for both networks since no significant difference was observed in the behavior of the loss function for the network with the additional convolution layer.

In Table V, we show results comparing the two cases evaluated on the testing training set. The channel mean layer outperforms the convolution layer according to all three metrics, justifying the use of the channel mean layer in the rest of this study.

VI. SUPER-RESOLVED SST STRUCTURES

One potential application of super-resolved thermal images is for the study of sea surface features and phenomena in

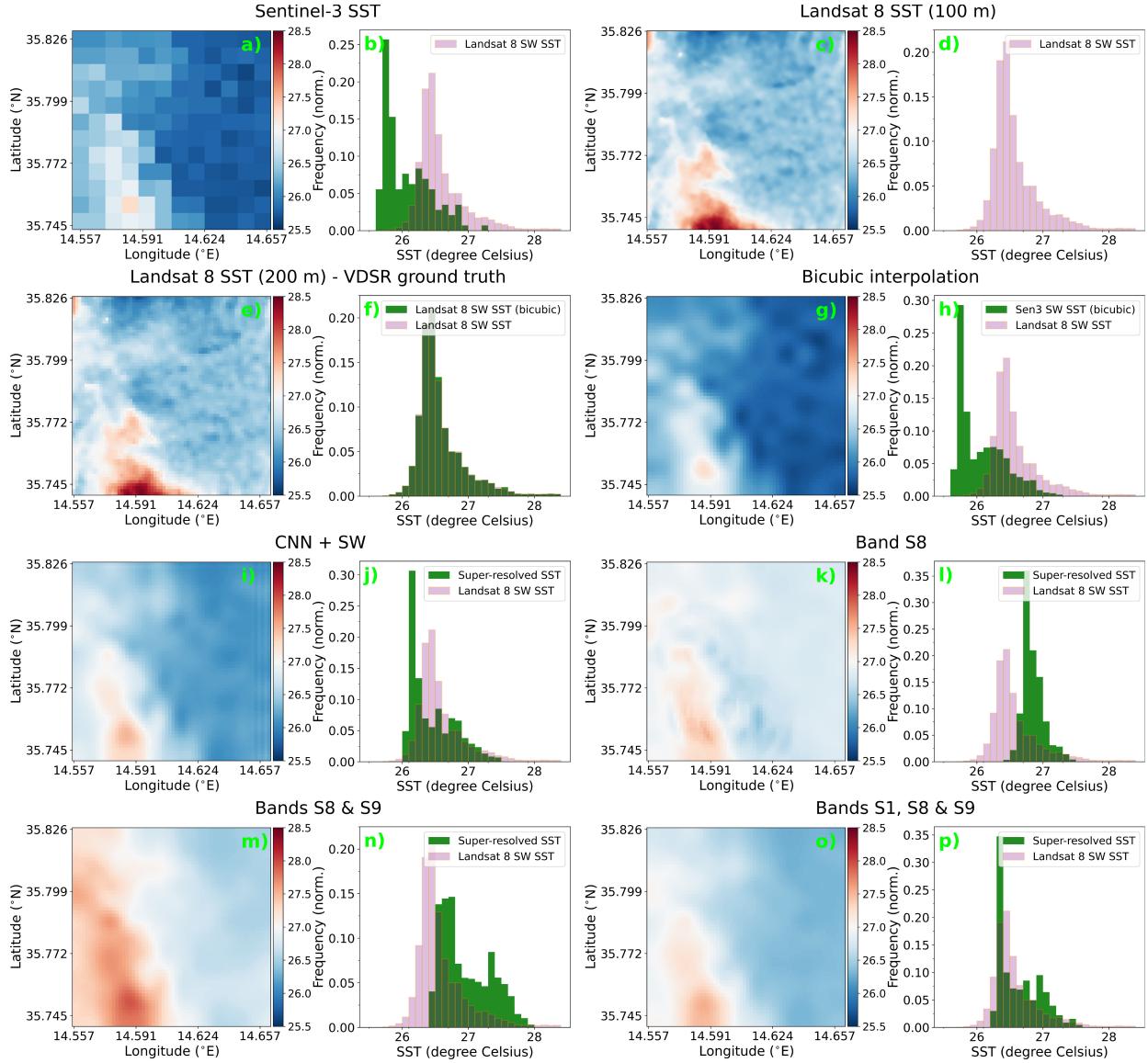


Fig. 6. Example SST patches and histograms of surface temperature, for (a) and (b) Sentinel-3 SST data; (c) and (d) Landsat 8 SST data computed using the SW technique at original resolution; (e) and (f) upscaled Landsat 8 ground truth used to train the VDSR network; (g) and (h) bicubic interpolation; (i) and (j) two separate networks trained to downscale brightness temperature followed by separate SW SST retrieval; (k) and (l) trained with Sentinel 3 band S8; and (m) and (n) bands S8 and S9; or (o) and (p) bands S1, S8, and S9. (f) Pink histograms correspond to the ground-truth patch.

TABLE V

COMPARING A MEAN LAYER WITH A CONVOLUTION LAYER FOR CHANNEL REDUCTION IN A VDSR NETWORK ARCHITECTURE.
PERFORMANCE IS COMPARED ON THE SAME TRAINING
DATASET USED THROUGHOUT THIS WORK

VDSR Modification	RMSE (K)	SSIM	PSNR (dB)
Channel mean	0.282	0.903	26.87
Convolution layer	0.328	0.882	25.59

greater detail than currently possible. Here, we consider two test cases, an ocean gyre and an ocean front, from two different geographical regions, and use a neural network to produce HR SST images from LR Sentinel 3 input data.

A. Ocean Gyre

The Malta Channel is the marine area between the islands of Sicily and Malta, linking the Eastern and Western

Mediterranean basins. The area is known for variable sea conditions, with influences including the complex local bathymetry [51] and the Mediterranean overturning circulation [52]. The Malta-Sicily Gyre (MSG) is an anticyclonic surface feature that occurs in the Malta channel [53], which has previously been identified using ground-based high-frequency radar in combination with satellite-based altimetry and SST products with 4.6-km resolution. In this section, we present SST images of a gyre-type surface structure in the Malta channel at a much higher resolution than previously examined.

Fig. 7(a) shows a Landsat 8 SW SST image of sea location in the Malta channel taken on July 20, 2018, corresponding to a surface area of 30km × 30km. The contributing Landsat 8 bands were first destriped and denoised before the SW SST retrieval was applied. The SST structure at the center of the image resembles a gyre. A histogram of the temperatures is shown in Fig. 7(b), displaying a bimodal

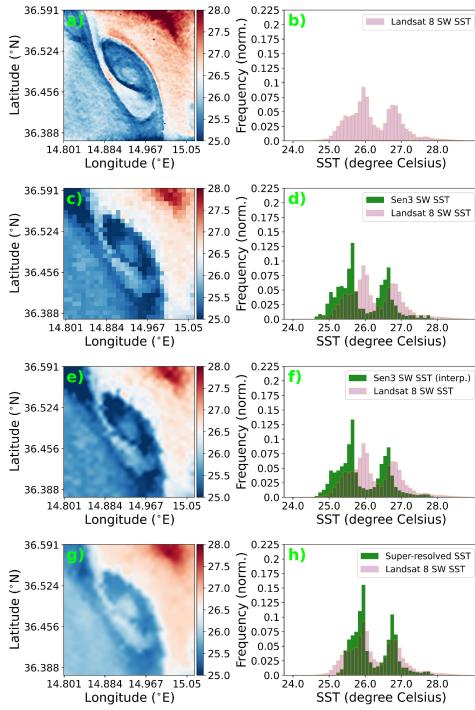


Fig. 7. SST images and histograms of a possible gyre in the Malta channel, recorded on July 20, 2018 for (a) and (b) Landsat 8 SW SST (100–200-m resolution); (c) and (d) Sentinel 3 SW SST (1-km resolution); (e) and (f) S3 SW SST, downsampled using bicubic interpolation (200-m pixel size); and (g) and (h) Super-resolved SST (200-m pixel size), for the case where Sentinel 3 TOA images from bands S1, S8, and S9 are used as input. (b) Pink histograms correspond to the ground-truth patch.

distribution consistent with the presence of a gyre. Fig. 7(c) shows an SST image derived using Sentinel 3 bands 8 and 9 and the SW retrieval method. The pixel size corresponds to 1 km. The same data, but with an interpolated pixel size of 200 m is shown in Fig. 7(e). Note the “staircasing” effect in the boundary between warm and cold regions typical of interpolation. Histograms of the Sentinel 3 SST patches are shown in Fig. 7(d) and (f), for native resolution and interpolated images, respectively. A small temperature offset between the Sentinel 3 and Landsat 8 SST images is apparent. Finally, the super-resolved SST image derived from Sentinel 3 bands S1, S8, and S9 is shown in Fig. 7(g). The texture of the image is smoother than the interpolated image and more in common with the Landsat 8 SST image. Similarly, the histogram of the super-resolved SST values closely resembles the Landsat 8 SST histogram, as shown in Fig. 7(h). It should be noted that owing to the time and location of the data presented in Fig. 7, patches used to train the network may overlap partially with the area in question.

As highlighted by Reyes Suarez *et al.* [53] in their work, the MSG is not a widely studied phenomenon and has sometimes been confused with similar features of the central Mediterranean, such as the Malta Channel Crest and Ionian Shelf Break Vortex. The ability to generate accurate, daily, high-resolution SST images of the Malta channel would allow further study of the MSG and similar phenomena.

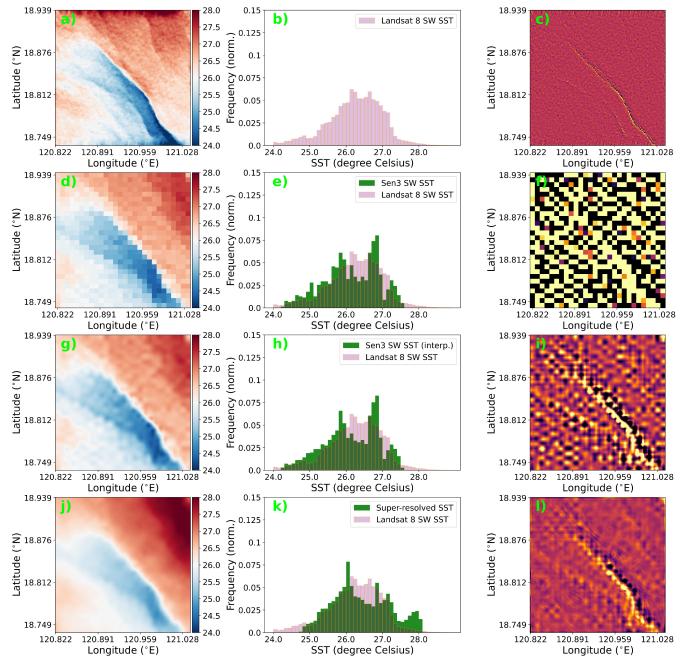


Fig. 8. SST images of a possible ocean front in the South China Sea recorded on March 18, 2020. (a) Landsat 8 SW SST (100–200-m resolution). (b) Histogram of the temperatures shown in (a). (c) SST gradient of the image shown in (a). (d) Sentinel 3 SW SST (1-km resolution). (e) Histogram of the temperatures shown in (d). (f) SST gradient of the image shown in (d). (g) S3 SW SST downsampled using bicubic interpolation (200-m resolution). (h) Histogram of the temperatures shown in (g). (i) SST gradient of the image shown in (h). (j) Super-resolved SST (200-m resolution), for the case where Sentinel 3 TOA images from bands S1, S8, and S9 are used as input. (k) Histogram of the temperatures shown in (j). (l) SST gradient of the image shown in (j).

B. Ocean Front

The South China Sea is a widely studied and monitored region [54], [55], owing in part to the presence of numerous commercial shipping lanes. A combination of monsoon winds and coastal dynamics, among other factors, lead to the creation of ocean fronts in this region. A front is a common oceanographic phenomenon found at the boundary of two different water masses, characterized by abnormally large horizontal gradients in surface quantities such as temperature and salinity. Recently, the seasonal variation of ocean fronts in the South China Sea was the topic of study by Wang *et al.* [28]. Satellite observations made by the MODIS instrument on-board the Aqua satellite were used to detect ocean fronts using an improved, gradient method [56]. The resolution of the data products used was 4.5 km.

In Fig. 8, satellite-derived SST images are shown of a likely ocean front identified in the South China Sea on March 18, 2020, corresponding to a region 28.5km × 28.5km in size. The SST derived using the SW method is shown for data from Landsat 8 [see Fig. 8(a)] and Sentinel 3 [see Fig. 8(d)]. For the Landsat 8 SST image, corresponding patches from bands 10 and 11 were first destriped and denoised before the SW SST retrieval was applied. In Fig. 8(g) the SST produced by first interpolating Sentinel 3 bands S8 and S9 and then using the SW SST retrieval is shown. In Fig. 8(j) the super-resolved

SST derived using a network trained with bands S1, S8 and S9 from Sentinel 3 is shown. Qualitatively, the super-resolved image more faithfully replicates the uniformity of the front feature, as well as the overall texture of the Landsat 8 SST patch. Turning to the histograms of patch temperatures, the super-resolved image histogram [see Fig. 8(k)] is a close match to the Landsat 8 SST histogram.

It should be noted that the neural network used to produce the SST patch shown in Fig. 8(g) is the same as that used to produce the best performing results shown in Table IV. Even though the network was trained on patches taken from the central Mediterranean sea, the example shown in Fig. 8(g) demonstrates that the network is able to successfully generate an HR SST patch in a geographic region far from where the training data were selected. This suggests that the network has generalized well and is not over-fit with respect to the training dataset.

Evaluating the performance of ocean front detection algorithms on super-resolved images is beyond the scope of this article. However, a qualitative appraisal of the possibility of using super-resolved SST images for enhanced ocean front detection can be made by considering the SST image gradients. In Fig. 8(c) the gradient of the Landsat 8 SW SST image is shown. The gradient was calculated numerically using the Laplacian operator. The ocean front is clearly evident in the gradient image against the otherwise smooth background. The front is not as apparent for the native resolution and interpolated Sentinel 3 SW SST gradient images, shown in Fig. 8(f) and (i) respectively. In Fig. 8(l), the SST gradient is shown for the super-resolved SST image. While the absolute value of the gradient is marginally smaller than the case of interpolated Sentinel 3 data, the ocean front location is more apparent and better defined in the case of super-resolved image. The possibility of producing “cleaner” SST gradients offered by SR of thermal images could improve the accuracy of front detection and benefit studies similar to that of Wang *et al.* [28]. In a similar context, SR has been shown to be valuable for object detection in HR optical images [57], it is, therefore, reasonable to expect similar benefits for object detection in the context of thermal sea images as well.

VII. DISCUSSION

The SR results presented in this article represent a proof-of-principle study into what is possible with modest modification to existing neural network architecture and sensible curation of remote sensing data. The Python implementation of our model is available at [58]. The overall high performance of the trained networks on the “unseen” testing dataset, alongside the qualitative results presented in Section VI-B suggests that the network has generalized well and is not over-fit. Nevertheless, it is worth highlighting the potential avenues that could lead to further improvements in the accuracy of the super-resolved images. The inclusion of training data patches from more varied locations, including deep ocean and polar regions, may further improve the performance and generalization. Equally, while the flexibility of the VDSR architecture was important for our work, especially for the investigation of optical bands for thermal image SR, the possibility remains

that a more complex network architecture, that is proven to be superior to VDSR for natural image SR (e.g., ProSR [24]), may yield better performance.

The agreement between super-resolved SST values and *in situ* measurements can only be as accurate as the agreement between the HR SST training patches and *in situ* measurements. While the use of analytic SW expressions with Landsat 8 bands has previously been shown to be reasonably well-matched to *in situ* data, there are some outstanding concerns about the size of an SST bias temperature [43] and accuracy under different satellite zenith angles [44], [45]. It should be noted, however, that if there are subsequent improvements in the accuracy and quality of HR SST images, then these new images could be used to improve the agreement between super-resolved SST images and *in situ* measurements. This could be achieved either via transfer learning or through retraining of the same network from scratch with the new data.

Our demonstration that higher resolution optical bands can be used to improve the resolution of lower resolution thermal images is consistent with the results of prior work utilizing optical indices for sharpening thermal land scenes [16], [17]. Our focus on sea scenes and the use of a neural network to explore and exploit the correlation between thermal and optical bands sets our work apart from prior efforts. Indeed, the results presented in our work further support the decision to install both optical and far-infrared instruments on future remote sensing satellite missions [59]. What is more, the methodology described here can be extended to any earth-observing satellites with sun-synchronous orbits and similar local times, enabling the investigation of hitherto hidden relationships between optical; near-, mid-, and far-infrared; and even radar bands.

VIII. CONCLUSION

The results presented in this article show that SR of thermal infrared images is a viable route for producing datasets of both brightness temperature and SST images with HR. Such datasets are desired by the user community but are not currently available from extant remote sensing sources. Indeed, until large improvements in infrared detectors and instrument design are made, or the launch of large constellations of satellites carrying HR infrared instruments, the only route toward producing surface temperature products with daily revisit time and high spatial resolution is via processing of existing remote sensing data. Deep learning offers a flexible and efficient way to extract new and useful information from current datasets. This process is clearly evidenced in our work through the demonstration enhancing thermal image SR with complimentary images from optical bands.

Overall, the tools of deep learning are well-matched to the data produced by earth-observing satellites and our results stand alongside a growing body of work exploring the possibility of enhancing environment datasets using artificial intelligence. With further study and careful *in situ* validation, super-resolved SST images could be an invaluable data source for ingestion into new HR SST products, enabling environmental and oceanographic studies with unprecedented detail.

REFERENCES

- [1] H. Hurlbert *et al.*, "High-resolution global and basin-scale ocean analyses and forecasts," *Oceanography*, vol. 22, no. 3, pp. 110–127, Sep. 2009.
- [2] P. Zhang, M. L. Imhoff, R. E. Wolfe, and L. Bounoua, "Characterizing urban heat islands of global settlements using MODIS and nighttime lights products," *Can. J. Remote Sens.*, vol. 36, no. 3, pp. 185–196, Jan. 2010.
- [3] R. W. Reynolds, "A real-time global sea surface temperature analysis," *J. Climate*, vol. 1, no. 1, pp. 75–87, Jan. 1988.
- [4] S. W. Seemann, J. Li, W. P. Menzel, and L. E. Gumley, "Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances," *J. Appl. Meteorol.*, vol. 42, no. 8, pp. 1072–1091, Aug. 2003.
- [5] R. Lohmann and I. M. Belkin, "Organic pollutants and ocean fronts across the Atlantic Ocean: A review," *Prog. Oceanogr.*, vol. 128, pp. 172–184, Nov. 2014.
- [6] J. C. McWilliams, "Submesoscale currents in the ocean," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 472, no. 2189, May 2016, Art. no. 20160117.
- [7] F. Nencioli, F. d'Ovidio, A. M. Doglioli, and A. A. Petrenko, "In situ estimates of submesoscale horizontal eddy diffusivity across an ocean front," *J. Geophys. Res. Oceans*, vol. 118, no. 12, pp. 7066–7080, Dec. 2013.
- [8] D. K. Savidge, "Gulf stream meander propagation past cape hatteras," *J. Phys. Oceanogr.*, vol. 34, no. 9, pp. 2073–2085, Sep. 2004.
- [9] J. Aschbacher, "ESA's Earth observation strategy and copernicus," in *Satellite Earth Observations and Their Impact on Society and Policy*, O. Masami and O. R. Young, Eds. Singapore: Springer, 2017, pp. 81–86.
- [10] S. G. George, "Satellite measurement of ocean turbulence," Ph.D. dissertation, Dept. Aeronaut., Astronaut. Comput. Eng., Univ. Southampton, Southampton, U.K., 2014.
- [11] S. Castro, W. Emery, G. Wick, and W. Tandy, "Submesoscale sea surface temperature variability from UAV and satellite measurements," *Remote Sens.*, vol. 9, no. 11, p. 1089, Oct. 2017.
- [12] T. William Freeman, R. Thouis Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar. 2002.
- [13] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jul. 2004.
- [14] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [15] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
- [16] W. P. Kustas, J. M. Norman, M. C. Anderson, and A. N. French, "Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship," *Remote Sens. Environ.*, vol. 85, no. 4, pp. 429–440, Jun. 2003.
- [17] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.
- [18] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Jun. 2016.
- [19] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
- [20] Y. Luo, L. Zhou, S. Wang, and Z. Wang, "Video satellite imagery super resolution via convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2398–2402, Dec. 2017.
- [21] N. Huang, Y. Yang, J. Liu, X. Gu, and H. Cai, "Single-image super-resolution for remote sensing data using deep residual-learning neural network," in *Neural Information Processing*, D. Liu, S. Xie, Y. Li, D. Zhao, El-S. M. El-Alfy, Eds. Cham, Switzerland: Springer, 2017, pp. 622–630.
- [22] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
- [23] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Comput. Vis.-ECCV Workshops*, L. Leal-Taixé and S. Roth, Eds. Cham, Switzerland: Springer, 2019, pp. 63–79.
- [24] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 864–873.
- [25] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 23, 2020, doi: 10.1109/TPAMI.2020.2982166.
- [26] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLI-B3, pp. 883–890, Jun. 2016.
- [27] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global combined network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, Aug. 2017.
- [28] Y. Wang, Y. Yu, Y. Zhang, H.-R. Zhang, and F. Chai, "Distribution and variability of sea surface temperature fronts in the south China sea," *Estuarine, Coastal Shelf Sci.*, vol. 240, Aug. 2020, Art. no. 106793.
- [29] N. Brodu, "Super-resolving multiresolution images with band-independent geometry of multispectral pixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4610–4617, Aug. 2017.
- [30] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 305–319, Dec. 2018.
- [31] Z. Tan, P. Yue, L. Di, and J. Tang, "Deriving high spatiotemporal remote sensing images using deep convolutional network," *Remote Sens.*, vol. 10, no. 7, pp. 1–16, 2018.
- [32] A. Ducournau and R. Fablet, "Deep learning for ocean remote sensing: An application of convolutional neural networks for super-resolution on satellite-derived SST data," in *Proc. 9th IAPR Workshop Pattern Recognition Remote Sens. (PRRS)*, Dec. 2016, pp. 1–6.
- [33] S. Mouatadid, S. Easterbrook, and A. R. Erler, "A machine learning approach to non-uniform spatial downscaling of climate variables," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 332–341.
- [34] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Generating high resolution climate change projections through single image super-resolution: An abridged version," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 5389–5393.
- [35] J. Leinonen, D. Nerini, and A. Berne, "Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, early access, Nov. 20, 2020, doi: 10.1109/TGRS.2020.3032790.
- [36] M. Zhang *et al.*, "Super-resolution enhancement of Sentinel-2 image for retrieving LAI and chlorophyll content of summer corn," *Eur. J. Agronomy*, vol. 111, Nov. 2019, Art. no. 125938.
- [37] J. Guan, R. Lai, and A. Xiong, "Wavelet deep neural network for stripe noise removal," *IEEE Access*, vol. 7, pp. 44544–44554, 2019.
- [38] Y. Chang, M. Chen, L. Yan, X.-L. Zhao, Y. Li, and S. Zhong, "Toward universal stripe removal via wavelet-based deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2880–2897, Apr. 2020.
- [39] G. Anne O'Carroll *et al.*, "Observational needs of sea surface temperature," *Frontiers Mar. Sci.*, vol. 6, p. 420, Aug. 2019.
- [40] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [41] D. Reuter *et al.*, "The thermal infrared sensor (TIRS) on Landsat 8: Design overview and pre-launch characterization," *Remote Sens.*, vol. 7, no. 1, pp. 1135–1153, Jan. 2015.
- [42] B. Wenny, D. Helder, J. Hong, L. Leigh, K. Thome, and D. Reuter, "Pre-and post-launch spatial quality of the Landsat 8 thermal infrared sensor," *Remote Sens.*, vol. 7, no. 2, pp. 1962–1980, Feb. 2015.
- [43] Q. Vanhellemont, "Automated water surface temperature retrieval from Landsat 8/TIRS," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111518.
- [44] J.-C. Jang and K.-A. Park, "High-resolution sea surface temperature retrieval from Landsat 8 OLI/TIRS data at coastal regions," *Remote Sens.*, vol. 11, no. 22, p. 2687, Nov. 2019.
- [45] K.-A. Park and J.-C. Jang, "High resolution sea surface temperature retrieval from Landsat 8 oli/tirs data at coastal regions," in *Proc. 21st GHRSST Int. Sci. Team Meeting (G-XXI)*, 2020.

- [46] C. Du, H. Ren, Q. Qin, J. Meng, and S. Zhao, "A practical split-window algorithm for estimating Land Surface Temperature from Landsat 8 data," *Remote Sens.*, vol. 7, no. 1, pp. 647–665, Jan. 2015.
- [47] T. David Lloyd, M. Bouali, A. Abela, R. Farrugia, and G. Valentino, "Efficient destriping of remote sensing images using an oriented super-Gaussian filter," *Proc. SPIE*, vol. 11533, pp. 19–37, Sep. 2020.
- [48] M. Bouali and A. Ignatov, "Adaptive reduction of striping for improved sea surface temperature imagery from Suomi National Polar-Orbiting Partnership (S-NPP) Visible Infrared Imaging Radiometer Suite (VIIRS)," *J. Atmos. Ocean. Technol.*, vol. 31, no. 1, pp. 150–163, Jan. 2014.
- [49] P. Coppo *et al.*, "SLSTR: A high accuracy dual scan temperature radiometer for sea and land surface monitoring from space," *J. Mod. Opt.*, vol. 57, no. 18, pp. 1815–1830, Oct. 2010.
- [50] P. J. Minnett *et al.*, "Half a century of satellite remote sensing of sea-surface temperature," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111366.
- [51] M. Würtz and M. Rovere, "Atlas of the Mediterranean seamounts and seamount-like structures," IUCN Gland, Switzerland, Tech. Rep. 45816, 2015.
- [52] R. Waldman, N. Brüggemann, A. Bosse, M. Spall, S. Somot, and F. Sevault, "Overturning the Mediterranean thermohaline circulation," *Geophys. Res. Lett.*, vol. 45, no. 16, pp. 8407–8415, Aug. 2018.
- [53] N. C. R. Suarez, G. S. Cook, M. Gačić, J. D. Paduan, A. Drago, and V. Cardin, "Sea surface circulation structures in the malta-sicily channel from remote sensing data," *Water*, vol. 11, no. 8, p. 1589, Jul. 2019.
- [54] B. Wang, F. Huang, Z. Wu, J. Yang, X. Fu, and K. Kikuchi, "Multi-scale climate variability of the south China sea monsoon: A review," *Dyn. Atmos. Oceans*, vol. 47, nos. 1–3, pp. 15–37, Jun. 2009.
- [55] Q. Han, X. Huang, Q. Xing, and P. Shi, "A review of environment problems in the coastal sea of south China," *Aquatic Ecosystem Health Manage.*, vol. 15, no. 2, pp. 108–117, Apr. 2012.
- [56] J.-F. Cayula and P. Cornillon, "Edge detection algorithm for SST images," *J. Atmos. Ocean. Technol.*, vol. 9, no. 1, pp. 67–80, Feb. 1992.
- [57] J. Shermeyer and A. Van Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1432–1441.
- [58] Python Super-Resolution Code GIT Repository. Accessed: May 17, 2021. [Online]. Available: <https://dsrg-ict.research.um.edu.mt/gianluca/super-resolution-of-sst-imagery>
- [59] B. Koetz *et al.*, "High spatio-temporal resolution land surface temperature mission—A copernicus candidate mission in support of agricultural monitoring," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 8160–8162.



David T. Lloyd received the M.Sc. degree in physics from Imperial College London, London, U.K., in 2010, and the D.Phil. degree in physics from the University of Oxford, Oxford, U.K., in 2015.

He held Post-Doctoral Researcher position at the University of Oxford from 2014 to 2018. He also held Post-Doctoral Researcher position at the University of Malta, Msida, Malta, from 2019 to 2020. He is a Senior Data Scientist with STMicroelectronics, Malta. His research interests include experimental optics, machine learning, and remote sensing.



Aaron Abela (Member, IEEE) received the B.Sc. degree (Hons.) in computer engineering, the M.Sc. degree in information and communications technology (ICT) (in telecommunications engineering), and the M.Res. degree in image processing from the University of Malta, Msida, Malta, in 2016, 2018, and 2021, respectively.

He held Research Support Officer position at the University of Malta from 2019 to 2020. He is a Systems Administrator and provides aid to research projects on due diligence at Aquibix Ltd., Naxxar, Malta. His research interests include telecommunications, image processing, and remote sensing.



Reuben A. Farrugia (Senior Member, IEEE) received the Ph.D. degree from the University of Malta, Msida, Malta, in 2009.

He has been working in the fields of image processing and pattern recognition with the University of Malta since 2004 with a specific focus on image processing and computer vision. He is a Senior Lecturer with the University of Malta. He has published more than 80 international peer-reviewed publications, and has been involved in technical and organizational committees of several national and international conferences.

Dr. Farrugia is an Area Editor of the *Signal Processing: Image Communications* journal (Elsevier). In September 2013, he was appointed as the National Contact Point on the European Association of Biometrics (EAB).



Anthony Galea received the B.Sc. degree (Hons.) in maths and physics and the M.Sc. by research degree in physical oceanography from the University of Malta (UM), Msida, Malta, in 2008 and 2010, respectively, and the Ph.D. degree in environmental and industrial fluid mechanics from the University of Trieste, Trieste, Italy, in 2014.

He forms part of the Physical Oceanography Research Group, UM, where the group carries out research in coastal meteorology, hydrography, and physical oceanography with an emphasis on the study of the hydrodynamics of the sea around the Maltese Islands. He is a Senior Lecturer with the Department of Geosciences, Faculty of Science, UM. His research interests include computational fluid mechanics, numerical modeling, physical oceanography, and science communication. He is involved in a number of national and European research projects.



Gianluca Valentino (Senior Member, IEEE) received the B.Sc. degree (Hons.) in computer engineering from the University of Malta, Msida, Malta, in 2010, and the Ph.D. degree from University of Malta in 2013 for developing a fast automatic technique for aligning the Large Hadron Collider collimators using machine learning.

From 2013 to 2016, he held a Marie Curie Post-Doctoral Fellowship with the Beams Department, CERN, Geneva, Switzerland. He is a Lecturer with the Department of Communications and Computer Engineering, Faculty of Information & Communication Technology (ICT), University of Malta, and a Visiting Scientist with CERN. His research interests include the application of machine learning techniques, in particular to high-energy physics, computer vision, and remote sensing. He has been awarded funding for several projects in these above areas. He is the Principal Investigator of the SAT-FIRE Project.