Article

Towards Streamlined Single-Image Super-Resolution: Demonstration with 10 m Sentinel-2 Colour and 10–60 m Multi-Spectral VNIR and SWIR Bands

Yu Tao [1,*] **, Siting Xiong** [2] **, Rui Song** [1] **and Jan-Peter Muller** [1]

[1] Imaging Group, Mullard Space Science Laboratory, Department of Space and Climate Physics, University College London, Holmbury St Mary, Surrey RH5 6NT, UK; j.muller@ucl.ac.uk (J.-P.M.); rui.song@ucl.ac.uk (R.S.)

[2] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China; xiongsiting@szu.edu.cn

[*] Correspondence: yu.tao@ucl.ac.uk

Abstract: Higher spatial resolution imaging data are considered desirable in many Earth observation applications. In this work, we propose and demonstrate the TARSGAN (learning Terrestrial image deblurring using Adaptive weighted dense Residual Super-resolution Generative Adversarial Network) system for Super-resolution Restoration (SRR) of 10 m/pixel Sentinel-2 "true" colour images as well as all the other multispectral bands. In parallel, the ELF (automated image Edge detection and measurements of edge spread function, Line spread function, and Full width at half maximum) system is proposed to achieve automated and precise assessments of the effective resolutions of the input and SRR images. Subsequent ELF measurements of the TARSGAN SRR results suggest an averaged effective resolution enhancement factor of about 2.91 times (equivalent to ~3.44 m/pixel for the 10 m/pixel bands) given a nominal SRR upscaling factor of 4 times. Several examples are provided for different types of scenes from urban landscapes to agricultural scenes and sea-ice floes.

Keywords: super-resolution restoration; SR; SRR; Sentinel-2; Sentinel; multispectral; sharpness; effective resolution assessment; image quality

1. Introduction

Very high spatial resolution imaging data play an important role in many fields of Earth Observation (EO) applications, such as precision agriculture, forestry, urban planning, city intelligence, cartography, geology, oceanography, and energy and utility maintenance. Although there are very high spatial resolution imaging sensors, e.g., the 31 cm/pixel Digital Globe® WorldView-3 images, the cost of such very high spatial resolution images is generally high, especially when and where large spatial-temporal volumes are required. On the other hand, while improvements in the spatial resolution are gaining priority in the design of new optical-electronic sensors onboard EO satellites, we still need to trade-off spatial resolution against spectral resolution, swath-width, signal-to-noise ratio of the sensor, launch mass, and requested telecommunications bandwidth. Subsequently, using super-resolution restoration (SRR/SR) to enhance existing EO data, especially those open access data, such as the European Space Agency's Copernicus Sentinel systems, is becoming an increasingly attractive alternative, especially if the resultant products can be employed to derive higher spatial resolution products like reflectance and derivatives of reflectance.

SRR refers to the process of enhancing (or increasing) the spatial resolution of images (or video frames) by exploiting non-redundant information from a set of repeat observations, or through a deep learning-based training and inference process. The growing technology interest in SRR, over the past 20 years, has led to the development and subsequent applications of many new algorithms, networks, and/or optimisations [1–4].

Classically, SRR was based on the idea of combining non-redundant information from multiple overlapping lower resolution (LR) images to produce the best estimation of a higher resolution (HR) image. This process was done either via image sub-pixel stacking [5,6], exploring the shifting and aliasing properties of the frequency domain [7–9], image degradation modelling [10–12], or multi-angle view modelling [13–15].

Over the past ten years, deep learning techniques have been very successful in the field of SRR due to their performance in terms of processing speed and flexibility over different input data. A variety of deep networks have been proposed over this time period to address the SRR problem. This includes the use of residual networks [16–18], recursive networks [19,20], selective attention networks [21,22], and Generative Adversarial Networks (GANs) [23–27]. Among these, the most recent works include Wide Activation Deep Residual SR network (WDSR) [17], the Residual Channel Attention Network (RCAN) [22], and the Multi-scale Adaptive weighted dense Residual SR GAN (MARSGAN) [27].

In particular, WDSR [17] improves the Enhanced Deep residual SR network (EDSR) [16], using slim residual blocks that have wider channels (2 to 9 times), while keeping the same parameter complexity. WDSR uses linear low-rank convolutions that factorise large convolutional kernels into two low-rank convolutional kernels, and use weight normalisation, to tackle the issues of slimmed layer pathway and training of very deep networks, respectively. RCAN [22] employs a residual-in-residual architecture and the Residual Channel Attention Blocks (RCABs) as its basic residual blocks, in order to rescale features adaptively by considering interdependencies between feature channels. MARSGAN [27] employs a very deep, densely connected, and adaptively weighted residual-in-residual architecture to further improve network capacity and information flow on top of the SR GAN network (SRGAN) [23] and the Enhanced SR GAN (ESRGAN) [25] network.

While many different ideas have been proposed to optimise the existing SRR networks, improvements in SRR performance have become more and more marginal if the modifications are purely based on the network architecture. Therefore, recent studies have been focused on either exploring different loss functions (e.g., exploring the perceptual-pleasing solutions [23–25]), or exploring the effect of using more realistic training datasets [28,29].

In particular, the Content Adaptive Re-sampler (CAR) based SRR model [28] employs a separately learned content-adaptive image downscaling model, which produces LR images that could keep the key information for best reproducing the HR images. The authors achieved state-of-the-art SRR performance in 2019/2020 using training datasets produced through CAR with an existing EDSR network [16]. Moreover, the authors in [29] constructed a "real-world" SRR training dataset (called RealSR), where paired LR and HR images are captured by adjusting the focal length of a digital camera, to replace the traditional synthetic training LR images, i.e., the bicubic down-sampled HR. RealSR achieved state-of-the-art SRR performance in 2020/2021 using their "real-world" training dataset with the existing RCAN network as well as a newly proposed Laplacian Pyramid-based Kernel Prediction Network (LP-KPN) [29].

In this paper, we further explore our in-house MARSGAN [27] model that was previously developed for Mars applications, using the Sentinel-2 10 m/pixel colour images and 10–60 m/pixel multi-spectral images. Inspired by [28,29], we propose practical modifications of the loss function, training dataset, and network architecture of MARSGAN, which we call learning Terrestrial image deblurring with Adaptive weighted dense Residual SR GAN (TARSGAN).

We show TARSGAN SRR results using the 10 m/pixel Sentinel-2 "true" colour images over a wide range of different types of natural and artificial surface features. These include SRR over urban sites (buildings, roads, cars, ships, airports), forestry and agriculture sites, and natural sites (mountains, deserts, the sea, the snow, and the sea-ice). Figure 1 gives an example of the 10 m/pixel Sentinel-2 "true" colour image and the 3.44 m/pixel TARSGAN SRR results over a geo-calibration site at Baotou, China.
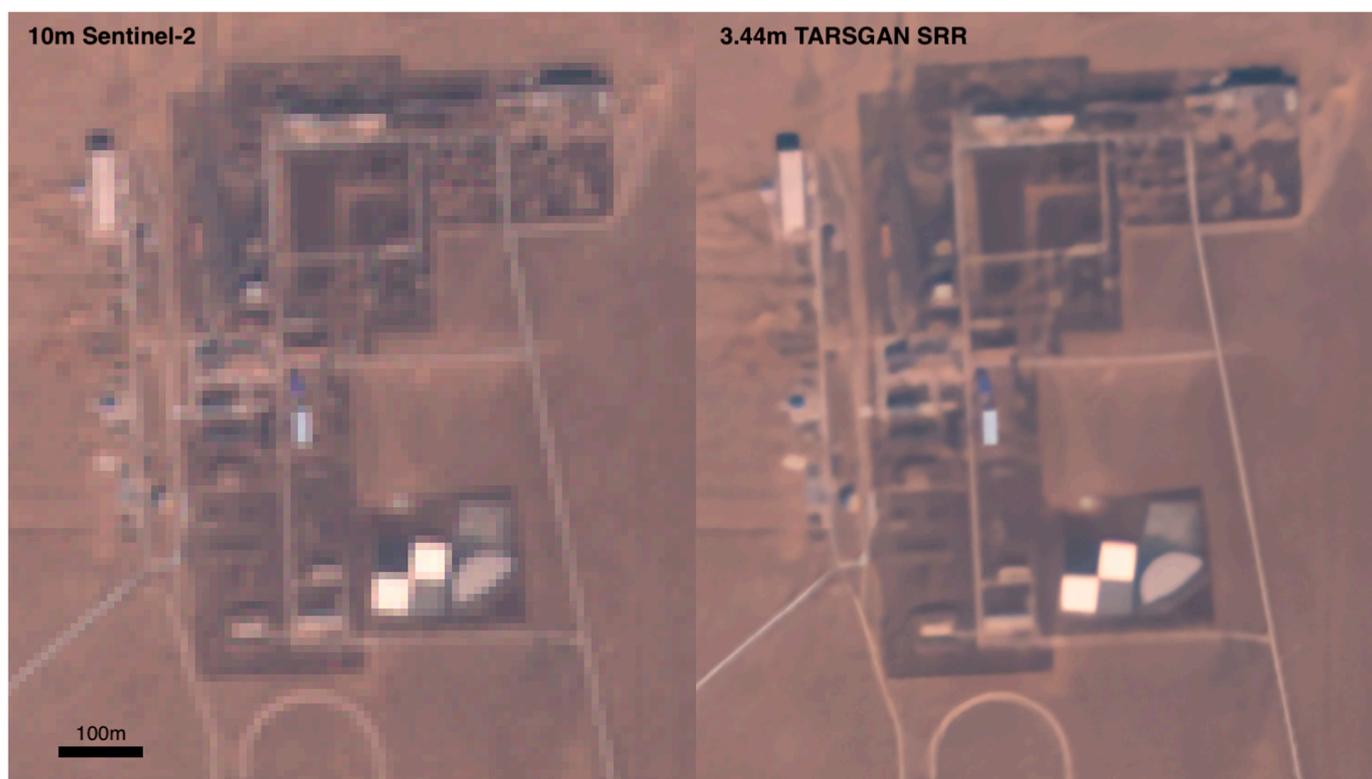
Figure 1. An example of the 10 m/pixel Sentinel-2 "true" colour image and the 3.44 m/pixel TARSGAN SRR results over a geo-calibration site at Baotou, China (Sentinel-2 image ID: S2A_MSIL1C_20171031T032851_N0206_R018 T49TCF_20171031T032851_TCI).

Moreover, we compare the spectral reflectance of the multispectral SRR product against the original Sentinel-2 multispectral product on all available spectral bands to demonstrate spectral invariance of the proposed TARSGAN system.

Furthermore, we propose an automated image effective resolution assessment system, using automated image Edge detection and filtering, and automated measurements of Edge Spread Function (ESF), Line Spread Function (LSF), and Full Width at Half Maximum (FWHM) – for brevity, this system is referred to hereafter as ELF. ELF is considered essential to building a streamlined and on-demand SRR processing system, in the future, which would require automated SRR algorithm selection and performance evaluation.

The ELF measurements suggest a factor of 2.91 times of effective resolution improvement on top of the 10 m/pixel Sentinel-2 "true" colour images using the proposed TARSGAN SRR system. This suggests our Sentinel-2 TARSGAN SRR results have an averaged effective spatial resolution of about 3.44 m/pixel. More importantly, in contrast to other generative SRR networks, TARSGAN does not introduce synthetic textures and artefacts.

With the proposed TARSGAN SRR and ELF effective resolution assessment system, we believe Sentinel-2 global 10–60 m/pixel multispectral images can be transformed into 3–20 m/pixel multispectral SRR images, fully automatically in the near future (readers should refer to the conceptual implementation of a streamlined SRR processing system in Section 4.3), allowing better analytics to be performed in a transformative way.

The layout of this paper is as follows. In Section 2.1, we introduce the training and test dataset. In Sections 2.2 and 2.3, we introduce technical details of the TARSGAN SRR system. In Section 2.4, we describe the ELF image effective resolution assessment system. Experimental results of the ELF system, TARSGAN SRR of 10 m/pixel Sentinel-2 "true" colour images, and TARSGAN SRR of the 10–60 m/pixel multispectral images, are demonstrated in Section 3.1, Section 3.2, and Section 3.3, respectively. In Section 4, we

discuss key issues, potential improvements, and future work before drawing conclusions in Section 5.

2. Materials and Methods

2.1. Datasets for Testing and Training

Our test dataset in this work consists of Sentinel-2 images. The Copernicus Sentinel-2 mission comprises a constellation of two identical polar-orbiting satellites (Sentinel-2A/2B), providing multi-spectral moderate spatial resolutions from 10 m/pixel to 60 m/pixel for its visible and near-infrared (VNIR) bands and short-wave infrared (SWIR) bands. Details of the Sentinel-2A/2B image spatial resolutions and spectral information can be found in Table 1. Sentinel-2A and Sentinel-2B are phased at 180° to each other, placed in the same sun-synchronous orbit. Sentinel-2A and Sentinel-2B have a wide swath width of 290km and together provide a high revisit frequency (5 days at equator and 2–3 days at mid-latitudes). Sentinel-2 data are accessible through the Copernicus open access hub (previously known as Sentinel scientific data hub; https://scihub.copernicus.eu/; accessed on 2 July 2021).

Table 1. Sentinel-2 spatial resolution, central wavelength, and bandwidth for each of the VNIR and SWIR bands.

Band name	Band No.	Spatial Resolution (m)	Central Wavelength (nm) 2A/2B	Bandwidth (nm) 2A/2B
VNIR	B01	60	442.7/442.2	20/21
	B02	10	492.4/492.1	66/66
	B03	10	559.8/559.0	36/36
	B04	10	664.6/664.9	31/31
	B05	20	704.1/703.8	15/16
	B06	20	740.5/739.1	15/15
	B07	20	782.8/779.7	20/20
	B08	10	832.8/832.9	106/106
	B8A	20	864.7/864.0	21/22
	B09	60	945.1/943.2	20/21
SWIR	B10	60	1373.5/1376.9	31/30
	B11	20	1613.7/1610.4	91/94
	B12	20	2202.4/2185.7	175/185

In this work, we perform SRR testing with a wide range of Sentinel-2 images over 6 different testing sites covering different types of natural and artificial surface feature/targets (see Table 2). The 6 test sites are located over Baotou/China, Dubai/ United Arab Emirates, Hainich/Germany, London/UK, Desert Rock/US, and Lincoln Sea/Greenland. These surface features/targets include artificial structures, residential buildings, industrial buildings, farms, countryside roads, highway roads, tower buildings, ships, artificial islands, airports, airplanes, forest, isolated trees, hills, mountains, train stations, urban building blocks, urban landmarks, bridges, deserts, river, sea-ice, leads, open water, and snow-covered surfaces. In this work, we mainly demonstrate SRR results with the level 1 (L1C) images, however, for 2 of the 6 sites (Site-3 Hainich and Site-4 London), where atmosphere clarity is low, we also show SRR results from the level 2 (L2A) images. These form a total number of 8 test Sentinel-2 images.

Table 2. Testing site, input Sentinel-2 image IDs (10 m/pixel TCI colour product), and containing features/targets.

Site Name	Image ID (Product Level)	Type of Features or Targets			
		Area-1	Area-2	Area-3	Area-4
Site-1 Baotou, China	S2A_MSIL1C_20171031T032851_N0206_R018_ T49TCF_20171031T070327 (L1C)	Geo-calibration targets and buildings	Buildings and roads	Farms and roads	Industrial building blocks
Site-2 Dubai, AE	S2B_MSIL1C_20210528T064629_N0300_R020_ T40RCN_20210528T084809 (L1C)	Tower buildings	Ships and sea waves	Artificial island	Airport, airplane, and roads
Site-3 Hainich, Germany	S2A_MSIL1C_20200921T103031_N0209_R108_ T32UNB_20200921T142406 (L1C)	Farms, houses, and roads	Forest	Farms with structures	Farms and hills
	S2B_MSIL2A_20210531T101559_N0300_R065_ T32UNB_20210531T140040 (L2A)	Agriculture site	Agriculture site	Agriculture site and village	Agriculture site
Site-4 London, UK	S2B_MSIL1C_20201217T111359_N0209_R137_ T30UXC_20201217T132006 (L1C)	Train stations and buildings	Small building blocks and bridges	Building blocks	Bridges and ships
	S2A_MSIL2A_20210309T105901_N0214_R094_ T30UXC_20210309T135358 (L2A)	London millennium wheel (with thin clouds)	London Stadium	Canary Wharf	Kensington Gardens (with thin clouds)
Site-5 Desert Rock, U.S.	S2A_MSIL1C_20210507T180921_N0300_R084_ T12SUD_20210507T215833 (L1C)	Mountain and road	Trees in desert	Desert and river	Desert and trees
Site-6 Lincoln Sea, Greenland	S2A_MSIL1C_20200729T190921_N0209_R056_ T21XWM_20200729T222945 (L1C)	Sea-ice and leads	Isolated sea-ice	Snow on mountain surface	Sea-ice and open water

Our training dataset is formed with Deimos Imaging, S.L. Deimos-2 images. Deimos-2 is a high-resolution Earth observation satellite, owned and operated by Deimos Imaging, S.L. Deimos-2 collects 0.75 m/pixel panchromatic (PAN) band and 4 m/pixel Multi-Spectral (MS) band images with a swath width of 12km (at nadir) from an orbit at ~600 km. The MS capability includes 4 channels in the visible: Red, Green, Blue bands and near-infrared (NIR) band. In this work, our training dataset is conducted with 102 non-repeat and cloud-free Deimos-2 PAN band images (sampled at 1 m/pixel), which consists of 300,512 pairs of LR and HR training samples. Instead of simply performing the "standard" bicubic down-sampling processing of the HR images (Deimos-2 PAN) to produce their LR counterparts, we use bicubic down-sampling followed by an average up-sampling and Gaussian blurring operations, to form the degraded LR images at the same scale as the HR images (1 m/pixel).

2.2. Key Modifications of MARSGAN

In contrast to "photo-enhancing" SRR tasks, the desired SRR outputs of remote sensing applications are fundamentally different. In remote sensing applications, higher signal-to-noise ratio (SNR), minimised artefacts, sharper edges and object outlines, and ultimately, the higher image effective resolution, are much more desirable in comparison to "re-creating" high-frequency textures and/or objects. The original design of SRGAN [23] and ESRGAN [25] are based on the idea that human vision does not care if the generated high-frequency textures are not strictly correlated with the ground truth as long as they look realistic. Such generated high-frequency textures can significantly improve the "perceptual sharpness" but are considered artefacts in remote sensing or scientific applications. For example, satellite image users probably do not want a synthetic map even it looks extremely real. Therefore, we consider perceptual quality-driven SRR techniques are not suitable to be used directly in any remote sensing applications. In the original work of MARSGAN [27], experiments were made to reduce the weights of perceptual loss terms, but consequently, the edge sharpness is also lowered as a trade-off of reducing the high-frequency artefacts.

In this work, we base our model on the MARSGAN architecture [27] but abandon the idea of training the model with weighted perceptual loss. Instead, we use a structural

similarity loss (see Section 2.3 for details) to reconstruct sharper outlines, without re-creating any synthetic textures/objects.

Moreover, the authors in [28,29], demonstrated that the information contained within the LR image actually plays an important role in successive SRR restoration on top of purely improving a network architecture. Inspired by this, we propose to bring the LR image into the same passive resolution as the HR image and apply a blurring operation at the HR scale, to better model the fuzzy appearance of an LR image after being upscaled or unsuccessfully super-resolved. This is based on the observation that even high-frequency components can be effectively learned in the LR space with upscaling convolutions in the end (as discovered since some early works of [30,31]), the information of the blurring effect is not well preserved in the LR space. For example, an oversmoothed edge of an image could be seen as a sharp zigzag edge after the down-sampling of the image. Figure 2 shows an example of the TARSGAN training LR image that was created via the proposed down-sampling, up-sampling, and Gaussian blur operations (refer to Section 2.1) of the 1 m/pixel Deimos-2 PAN band image, in comparison to the "standard" training LR image created from a simple bicubic down-sampling operation as used in general in SRR works.
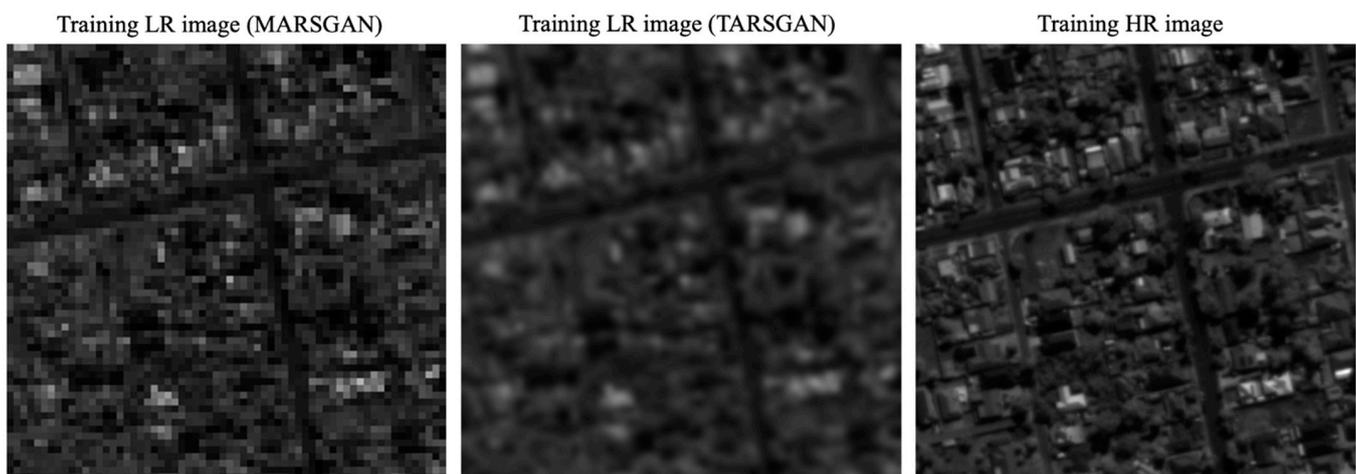

Figure 2. An example of the MARSGAN training LR image (4 times down-sampled version of the training HR image–similar to using the 4m MS band image), the TARSGAN training LR image (created via 4 times down-sampling, 4 times up-sampling, and Gaussian blurring), and the training HR image (same for both MARSGAN and TARSGAN). Image dimensions: 256 m × 256 m. Deimos-2 image courtesy of Deimos Imaging, S.L. 2021.

In summary, our goal in this work is to limit the room for the SRR network to learn "synthetic SRR" and encourage the network to learn "deblurring SRR". We focus on training a deblurring-oriented SRR network, i.e., TARSGAN, that fits the goal of remote sensing SRR applications, by modifying the loss function and removing the up-scaling process of the original MARSGAN system [27] and constructing a new training dataset that preserves the blurring information.

2.3. The TARSGAN System

The backbone of our proposed TARSGAN system is the MARSGAN model [27], which itself is based on a GAN framework [32–34]. GAN provides an efficient framework for learning generative tasks like SRR. Described in the fundamental work of [32,33], GAN trains a generative model for SRR, whilst in parallel, it trains a discriminator model to distinguish the predicted SRR image from ground-truth HR. Through alternative updates of the two adversarial networks, the generative model is trained to produce SRR images that are barely distinguishable from the HR images. For TARSGAN, we apply two practical modifications to the original MARSGAN system as follows.

Firstly, the adaptive weighted multi-scale reconstruction block is removed as the training LR and testing LR images are pre-upsampled at the same passive resolution as the HR images as described previously. Denoting the SRR image as I_{SRR}, the LR image as I_{LR}, and the HR ground truth as I_{HR}. The TARSGAN generator can be simplified from the MARSGAN generator [27] as N ($N = 16$ in this work) layers of Adaptive-weighted Residual-in-Residual Dense Blocks (AWRRDBs), where layers exclude the first and the last layers. The first layer has 64 filters of size 3×3 for initial feature extraction, denoted as f_{ext}, and the last layer has a single filter of size $3 \times 3 \times 64$ for SRR image reconstruction, denoted as f_{rec}. Denoting the n-th ($n \in N$) AWRRDB unit as f^n_{AWRRDB}, the TARSGAN generator can be expressed as

$$I_{SRR} = f_{rec}\left(f^N_{AWRRDB}\left(f^{N-1}_{AWRRDB}\left(\cdots f^1_{AWRRDB}(f_{ext}(I_{LR}))\cdots\right)\right)\right) + I_{LR} \tag{1}$$

It should be noted that we use fewer AWRRDB layers in TARSGAN ($N = 16$) in comparison to MARSGAN ($N = 23$), as needed to reduce the computation cost when having LR image at the same scale as the HR image. However, we empirically found that because TARSGAN is not trained with a perceptual loss term, the improvement from stacking more AWRRDB layers is marginal. For AWRRDBs, we use the same architecture as described in MARSGAN [27], i.e., each AWRRDB contains 3 dense blocks, and each dense block contains 5 convolutional layers (3×3 kernels, 32 feature maps, stride 1) and 4 Leaky Rectified Linear Unit (LReLU) activation with a negative slope of 0.2. The generator network architecture of TARSGAN is shown in Figure 3. For a detailed description of the AWRRDB blocks, please refer to [27].
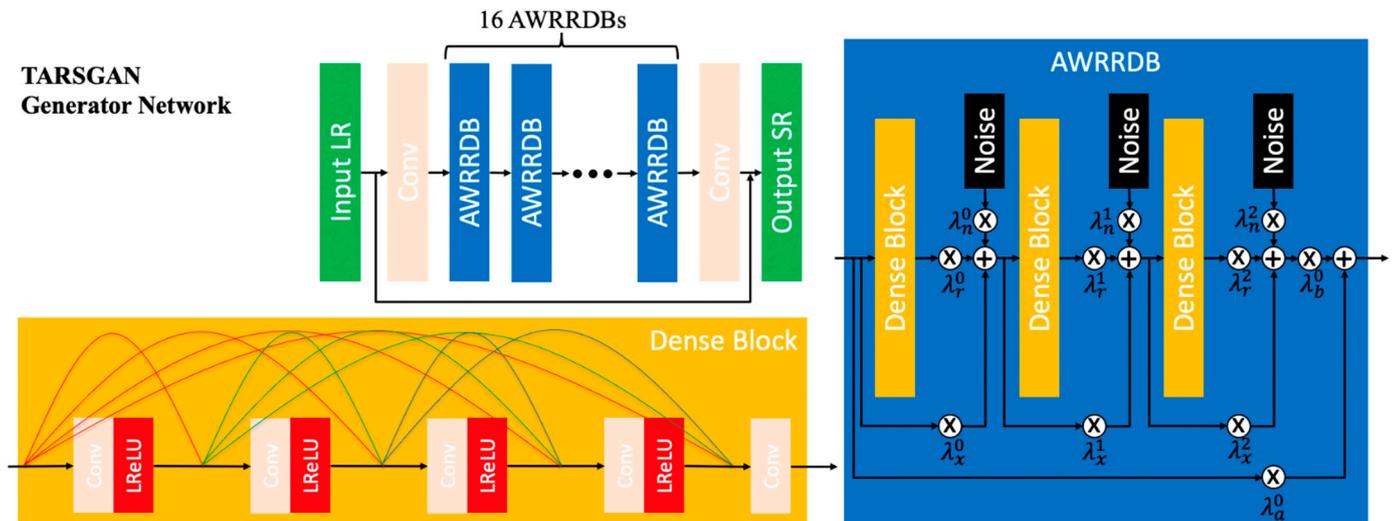

Figure 3. Network architecture of the TARSGAN generator.

Secondly, we redefine the total loss function, denoted as l_{total}, as a weighted sum of the Mean Squared Error (MSE) loss, denoted as l_{MSE}, the adversarial loss, denoted as l_{gen}, and the Structural Similarity (SSIM) [35] loss, denoted as l_{SSIM}. SSIM is a commonly used metric in image reconstruction tasks; in particular, it has been widely used in unsupervised image depth estimation tasks to quantify the differences between a back-projected image and the reference image (e.g., [36,37]), and as well as being an evaluation metric in many SRR works representing the retrieval quality of structural features. SSIM is derived using patterns of pixel intensities among neighbouring pixels with normalised brightness and contrast as introduced in [35]. For the generated target image I_{SRR} and the reference truth I_{HR}, $SSIM(I_{SRR}, I_{HR})$ can be formulated as

$$SSIM(I_{SRR}, I_{HR}) = \frac{\left(2\mu_{I_{SRR}}\mu_{I_{HR}} + C_1\right)\left(2\sigma_{I_{SRR},I_{HR}} + C_2\right)}{\left(\mu_{I_{SRR}}^2 + \mu_{I_{HR}}^2 + C_1\right)\left(\sigma_{I_{SRR}}^2 + \sigma_{I_{HR}}^2 + C_2\right)} \tag{2}$$

where $\mu_{I_{SRR}}$, $\mu_{I_{HR}}$, $\sigma_{I_{SRR}}$, $\sigma_{I_{HR}}$, and $\sigma_{I_{SRR},I_{HR}}$ are the local means, standard deviations, and cross-covariance of I_{SRR} and I_{HR}, respectively. C_1 and C_2 are constants based on the dynamic range of pixel values. As SSIM has an upper bound of 1, l_{SSIM} can be defined as

$$l_{SSIM} = \frac{1 - SSIM(I_{SRR}, I_{HR})}{2} \tag{3}$$

The other two loss terms, i.e., l_{MSE} and l_{gen}, are the same as the ones described in [27]. The total loss l_{total} of TARSGAN can be expressed as

$$l_{total} = \gamma l_{MSE} + \lambda l_{gen} + \eta l_{SSIM} \tag{4}$$

where γ, λ and η are the weights to balance the pixel-wise MSE loss, adversarial loss of the discriminator, and the SSIM loss. In practice, TARSGAN is initialised with $\gamma = 1$, $\lambda = 5 \times 10^{-3}$, $\eta = 0$, then refined with $\gamma = 0.5, \lambda = 5 \times 10^{-3}$, $\eta = 1$. The initial learning rate is 10^{-4}, and standard Adam optimisation [38] is used with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. Training and testing are achieved on the latest Nvidia® RTX 3090 GPU (Graphic Processing Unit).

As discussed in [23,27], the potential texture details from an SRR network, are typically synthetic textures (if not absent) and therefore cannot be "pixel-to-pixel" matched with the ground truth HR, thus leading to a smoother solution that averages all potential synthetic solutions when an MSE loss is used. Optimising an SRR network with the MSE loss generally results in a smoothed reconstruction, however, with fewer synthetic artefacts. In TARSGAN, we initialise the network first towards a smoother solution with respect to the HR image to resolve large-scale and intermediate-scale features. Then the network is refined towards better structural similarity measurement with respect to the HR image to resolve shaper edges and shape/outline of small objects that are visible (but blurred) in the LR image. For small objects or textures that are fundamentally not visible from the LR image, we do not try to re-create them with TARSGAN.

2.4. The ELF System

In parallel to the TARSGAN SRR system, we also propose the ELF automated image-effective resolution assessment system. The design of ELF is based on the Imatest® (see https://www.imatest.com/; accessed on 2 July 2021) slanted-edge method and previous collaborative work within the UK Space Agency funded SuperRes-EO project using FWHM to assess the image effective resolutions. ELF measures the averaged FWHM of all detectable slanted edges within an SRR image and compared against the averaged FWHM of the same edges within the corresponding LR image. The overall workflow of ELF is shown in Figure 4.

ELF takes the SRR image and the reference LR image, which is up-sampled to the same scale as the SRR image, as inputs, and follows 9 processing steps that are briefly described below.

(1) Create a binary image from the input SRR image using the Otsu adaptive thresholding method [39].
(2) Use a Canny edge detector [40] to extract all potential edges.
(3) Use a Hough transform [41] to detect potential lines from the output of (2) and filter for the given thresholds of lengths, gaps, and intersections.
(4) Crop for any number of regions of interest (ROIs) centred on the filtered lines and apply the same cropping for the same areas with the same sizes using the corresponding LR image.

(5) Perform image normalisation within each crop for both the crops from SRR and crops from LR.
(6) Calculate and plot the ESF for each slanted edge within each normalised crop from (5).
(7) Filter each continuous ESF and only leave the peak ESF for each slanted edge.
(8) Calculate and plot the LSF for each ESF from (7).
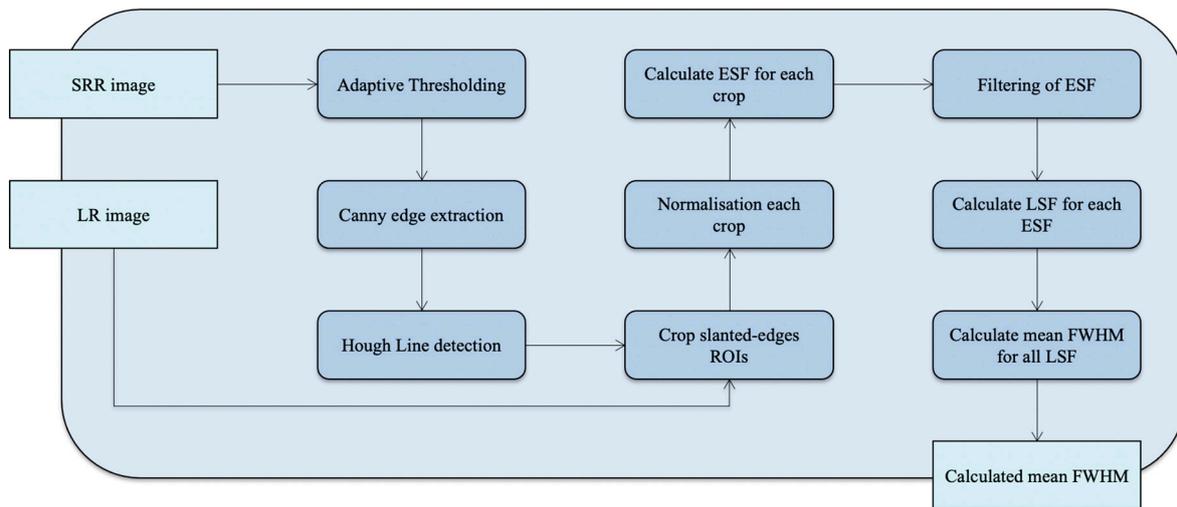(9) Calculate FWHM for each LSF from (8) and calculate the mean FWHM for the SRR and LR images.

Figure 4. Flow diagram of the proposed ELF automated image effective resolution assessment system.

The final calculated mean FWHM (MFWHM) over all detected slanted edges is used to assess the image effective resolution with respect to the given native resolution of the LR image. The final effective resolution of the Sentinel-2 SRR image can be estimated by calculating the ratio, denoted as β, of the $MFWHM(I_{LR})$ and $MFWHM(I_{SRR})$, which is proportional to the resolution enhancement factor, denoted as α. This is expressed as

$$\alpha = \frac{Res(I_{LR})}{Res(I_{HR})} \propto \beta = \frac{MFWHM(I_{LR})}{MFWHM(I_{HR})} \tag{5}$$

where $Res(I_{LR})$ and $Res(I_{HR})$ represent the native image resolution of I_{LR} and I_{HR}, respectively, and $MFWHM(I_{LR})$ and $MFWHM(I_{HR})$ represent the averaged FWHM of all detected slant edges of the same areas of I_{LR} and I_{HR}, respectively. The relationship between α and β can be explored and validation performed using the adjacent bands of the 10 m/pixel, 20 m/pixel, and 60 m/pixel Sentinel-2 images. In Section 3.1, we initially calculate β using the Sentinel-2 20 m/pixel B05 and 10 m/pixel B04 images (for $\alpha = 2$), and then calculate β (for $\alpha = 2, 3, 6$) using the 10 m/pixel B08, 20 m/pixel B8A, and 60 m/pixel B09 images to ensure that the inter-comparison is done with spectrally close channels.

3. Results

3.1. Estimation of Image Effective Resolution through ELF

In order to estimate the image effective resolution through the ELF measurements, we perform two experiments based on the original Sentinel-2 10 m/pixel, 20 m/pixel, and 60 m/pixel images. The first experiment (Exp-1) is based on the 10 m/pixel B04 image and 20 m/pixel B05 image. The second experiment (Exp-2) is based on the 10 m/pixel B08 image, 20 m/pixel B8A image, and 60 m/pixel B09 image.

We calculate $MFWHM(I_{HR})$ and $MFWHM(I_{LR})$ in Equation (5) using all detectable edges from I_{HR} and the same slanted edges at the same locations from I_{LR}, respectively, thus ensuring that the value of β can be computed for a specific image crop. The purpose

of Exp-1 (using B05/B04) is to validate against Exp-2 (using B8A/B08), to check if there is any significant difference in the calculated β, when α is fixed ($\alpha = 2$). As the computed α is demonstrated similar for using B05/B04 and using B8A/B08, we therefore repeat Exp-2 with 4 cropped images of size 8 km $\times$ 8 km to calculate the mean values of β for 2 times, 3times, and 6 times resolution differences, using B08 and B8A ($\alpha = 2$), B8A and B09 ($\alpha = 3$), and B08 and B09 ($\alpha = 6$), respectively.

In Figure 5, we demonstrate two examples (two detected slanted edges) within one 8 km $\times$ 8 km crop (at Site-1) using 10 m/pixel B04 image (I_{HR}) and 20 m/pixel B05 image (I_{LR}). For all measurement records of all detected slanted edges, please refer to the Supplementary Material. There are 29 valid FWHM measurements out of 77 detected slanted edges for this image crop. The calculated $MFWHM(I_{HR})$ is 3.19 pixels and $MFWHM(I_{LR})$ is 3.31 pixels, which suggests a β of 1.037 for an α of 2. In Figure 6, we demonstrate with two other examples (two detected slanted edges) within the same 8 km $\times$ 8 km crop (at Site-1) but using three bands, i.e., 10 m/pixel B08 (I_{HR}), 20 m/pixel B8A (I_{LR}), and 60 m/pixel B09 (I_{LR}). The FWHM measurements of the B08 image are compared against the FWHM measurements of the B8A image for $\alpha = 2$, and also against the FWHM measurements of the B09 image for $\alpha = 6$. There are 41 and 16 valid FWHM measurements out of 105 detected slanted edges for B08 and B8A and for B08 and B09, respectively. The calculated $MFWHM(I_{HR})$ and $MFWHM(I_{LR})$ for B08 and B8A are 3.14 pixels and 3.42 pixels (averaged from 41 FWHM measurements), respectively. The calculated $MFWHM(I_{HR})$ and $MFWHM(I_{LR})$ for B08 and B09 are 3.33 pixels and 4.89 pixels (averaged from 16 FWHM measurements), respectively. These suggest an average β of 1.057 for $\alpha = 2$, which is close to the measurements in Exp-1 ($\beta = 1.037$), and a β of 4.086 for $\alpha = 6$.

Performing Exp-2 for three more image crops (8 km $\times$ 8 km) for B08 and B8A ($\alpha = 2$), B8A and B09 ($\alpha = 3$), and B08 and B09 ($\alpha = 6$), we can obtain the mean β of 1.055 for $\alpha = 2$, mean β of 1.386 for $\alpha = 3$, and mean β of 1.517 for $\alpha = 6$. This is summarised in Table 3. For full measurement details, please refer to the Supplementary Material.

Table 3. ELF measurements of three cropped regions for B08 and B8A ($\alpha = 2$), B8A and B09 ($\alpha = 3$), and B08 and B09 ($\alpha = 6$).

Crops	$\alpha = 2$				$\alpha = 3$				$\alpha = 6$			
	Edges	*MFWHM* B08	*MFWHM* B8A	β	Edges	*MFWHM* B8A	*MFWHM* B09	β	Edges	*MFWHM* B08	*MFWHM* B09	β
1	75/166	2.76	2.85	1.033	20/104	3.47	4.15	1.196	12/166	3.24	4.53	1.398
2	22/43	2.92	3.14	1.075	2/35	3.20	4.40	1.375	9/43	3.36	5.44	1.619
3	32/65	2.74	2.90	1.058	7/49	2.94	4.67	1.588	10/65	3.30	5.06	1.533
Avg.	-	-	-	1.055	-	-	-	1.386	-	-	-	1.517

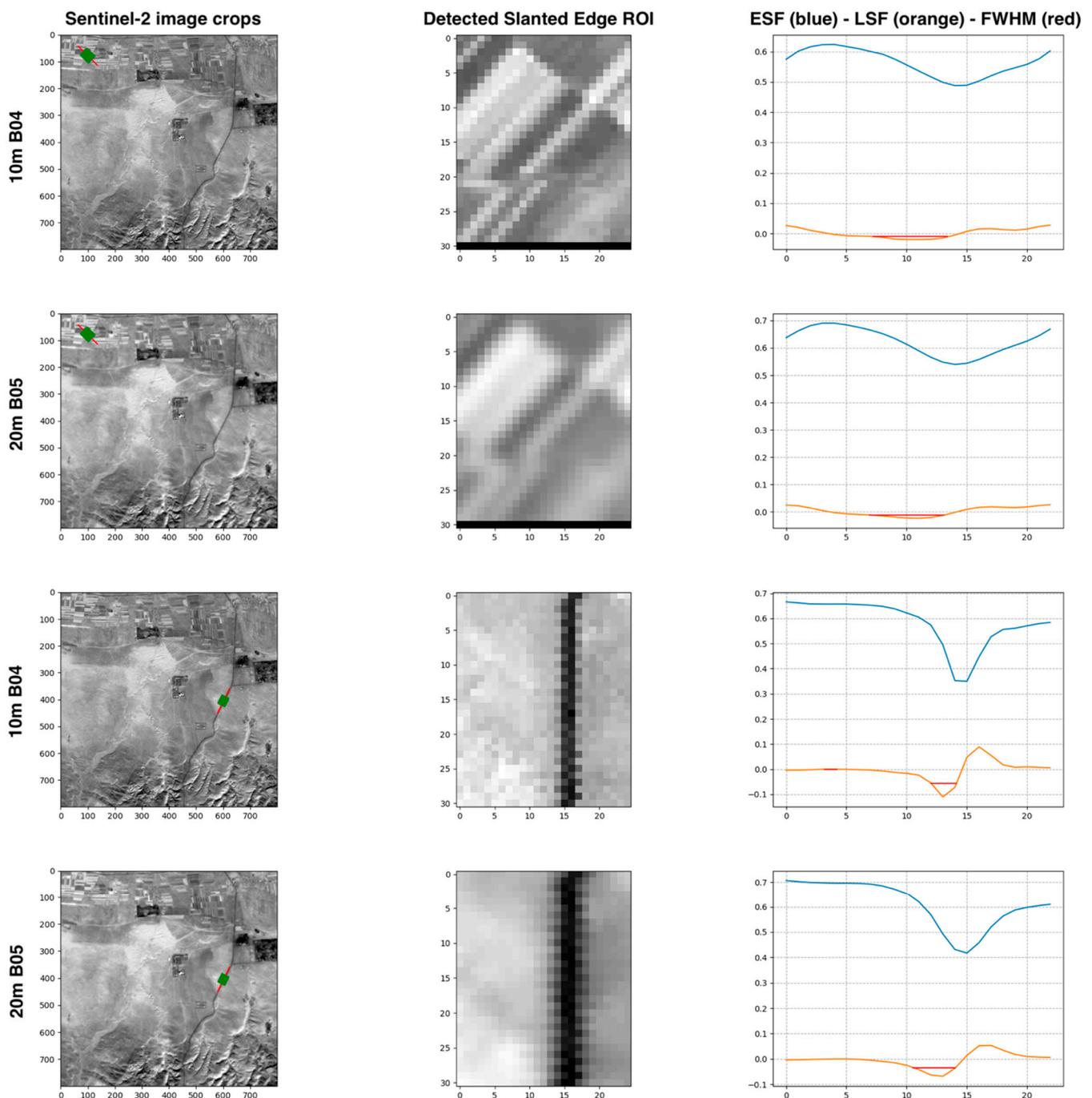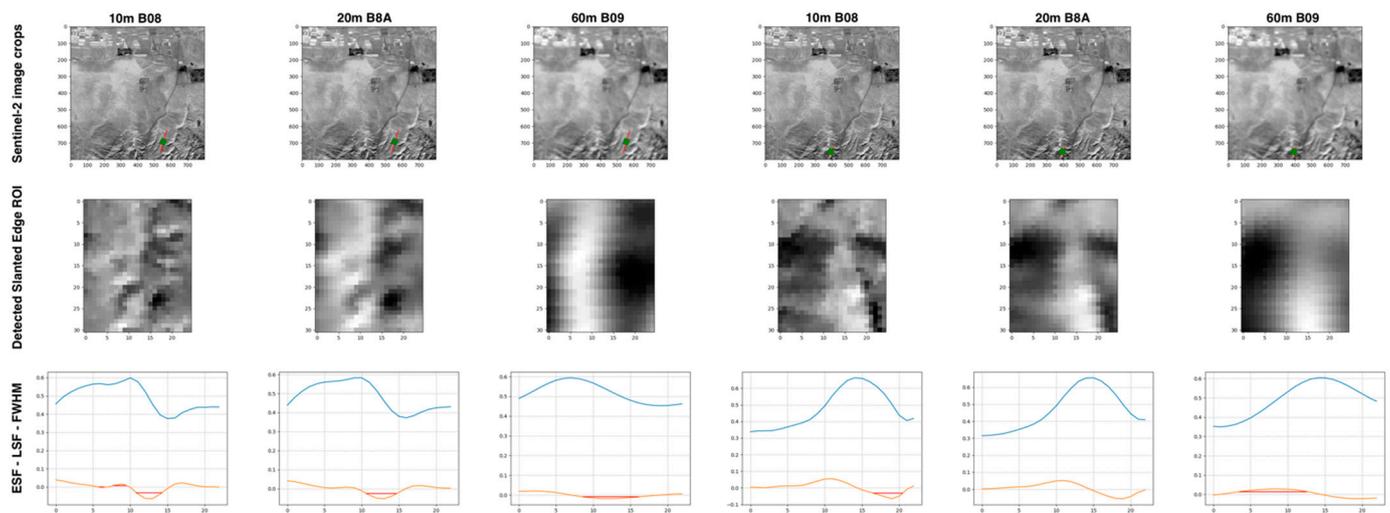**Figure 5.** Examples of the Exp-1 ELF measurements for two detected slanted edge ROIs of a 8 km × 8 km image crop at Site-1. 1st column: Sentinel-2 image crops of the 10 m/pixel B04 band and 20 m/pixel B05 band images (pre-upsampled to 10 m/pixel for comparison) showing two examples of the detected slanted edges in the green box; 2nd column: zoom-in views of the examplar slanted edges within the automatically extracted ROIs; 3rd column: plots of ESFs (blue curve), LSFs (orange curve) and FWHMs (red line) for the examplar slanted edges. For all Exp-1 ELF measurements of all detected slanted edges within the 8 km × 8 km image crop at Site-1, please refer to Supplementary Material. N.B. Units of the x and y axes of the 1st column and the 2nd column, and x axes of the 3rd column figures are "pixels"; units of the y axes of the 3rd column figures are normalised intensity values—[0, 1] for ESF and [−0.1, 0.1] for LSF. 1st and 2nd columns show images at different sizes of 8 km × 8 km and 250 m × 300 m, respectively.

Figure 6. Examples of the Exp-2 ELF measurements for two detected slanted edge ROIs of a 8 km × 8 km image crop at Site-1. 1st row: Sentinel-2 image crops of the 10 m/pixel B08 band, 20 m/pixel B8A band, and 60 m/pixel B09 band images, showing two examples of the detected slanted edges in green box; 2nd row: zoom-in views of the examplar slanted edges within the automatically extracted ROIs; 3rd row: plots of ESFs (blue curve), LSFs (orange curve) and FWHMs (red line) for the examplar slanted edges. For all Exp-2 ELF measurements of all detected slanted edges within the 8 km × 8 km image crop at Site-1, please refer to Supplementary Material. N.B. Units of the x and y axes of the 1st row and the 2nd row, and x axes of the 3rd row figures are "pixels"; units of the y axes of the 3rd row figures are normalised intensity values—[0, 1] for ESF and [−0.1, 0.1] for LSF. 1st and 2nd rows show images at different sizes of 8 km × 8 km and 250 m × 300 m, respectively.

3.2. Demonstration of TARSGAN SRR Results and Subsequent ELF Assessment

In order to demonstrate SRR performance over different features and targets, we show the results of six test sites with eight images, including six Sentinel-2 L1C images and two L2A images. For each Sentinel-2 image, we show four small crops (250 × 250 pixels each and with a nominal spatial resolution of 2.5 m/pixel) that cover a variety of different features and targets of interest that are summarised in Table 2. N.B. to look into more details, please refer to the original full-size SRR images provided in the Supplementary Material.

Figure 7 shows cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour image (L1C) and TARSGAN SRR result of Site-1, which is located over Baotou, Inner Mongolia, China. It is part of the CEOS-WGCV (Committee on Earth Observation Satellites Working Group on Calibration and Validation) geometric calibration site described in [42,43]. Area-1 shows the artificial geo-calibration targets together with a few buildings and roads. Area-2 shows a dome-shaped building in the centre with gardens and roads surrounded. Area-3 shows farms with linear farm roads. Area-4 shows industrial building blocks. We can observe from the SRR image that the black and white geo-calibration targets in Area-1 and the farms in Area-3 were brought out with clearer outlines. The buildings and roads in Area 1,2 and 4 can be identified more easily from the SRR image in comparison to the original Sentinel-2 image. There is no artefact found in the SRR image for the four areas of Site-1.

Figure 8 shows cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour image (L1C) and TARSGAN SRR result of Site-2, which is located over Dubai, United Arab Emirates. Area-1 shows many tower buildings in the city centre. Area-2 shows two ships sailing on the sea that is close to the beach. Area-3 shows an artificial island over the nearby beach with flat buildings and roads. Area-4 shows an airport with parked airplanes and viaducts. We can observe from the SRR that the flat and tower buildings are much clearer, and the roads were well resolved in Area-1, 3 and 4. The ships and airplanes in Area-2 and 4 can be better identified from the SRR image. Especially for Area-2, fine-scale

waves are revealed in the SRR image. There is no synthetic artefact (e.g., SRR generated objects/textures) found in the SRR image for the four areas of Site-2.
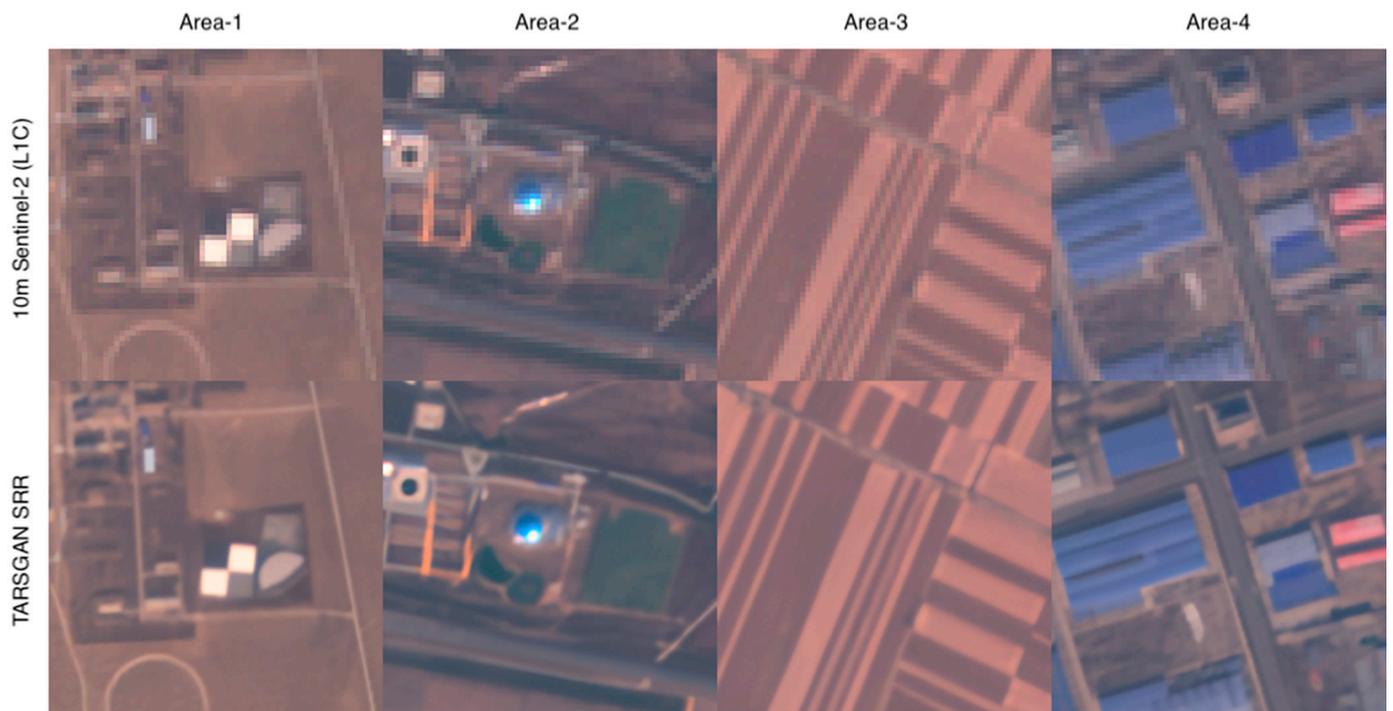

Figure 7. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L1C) and TARSGAN SRR result over Baotou, China (Site-1). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2A_MSIL1C_20171031T032851_N0206_R018 T49TCF_20171031T070327).
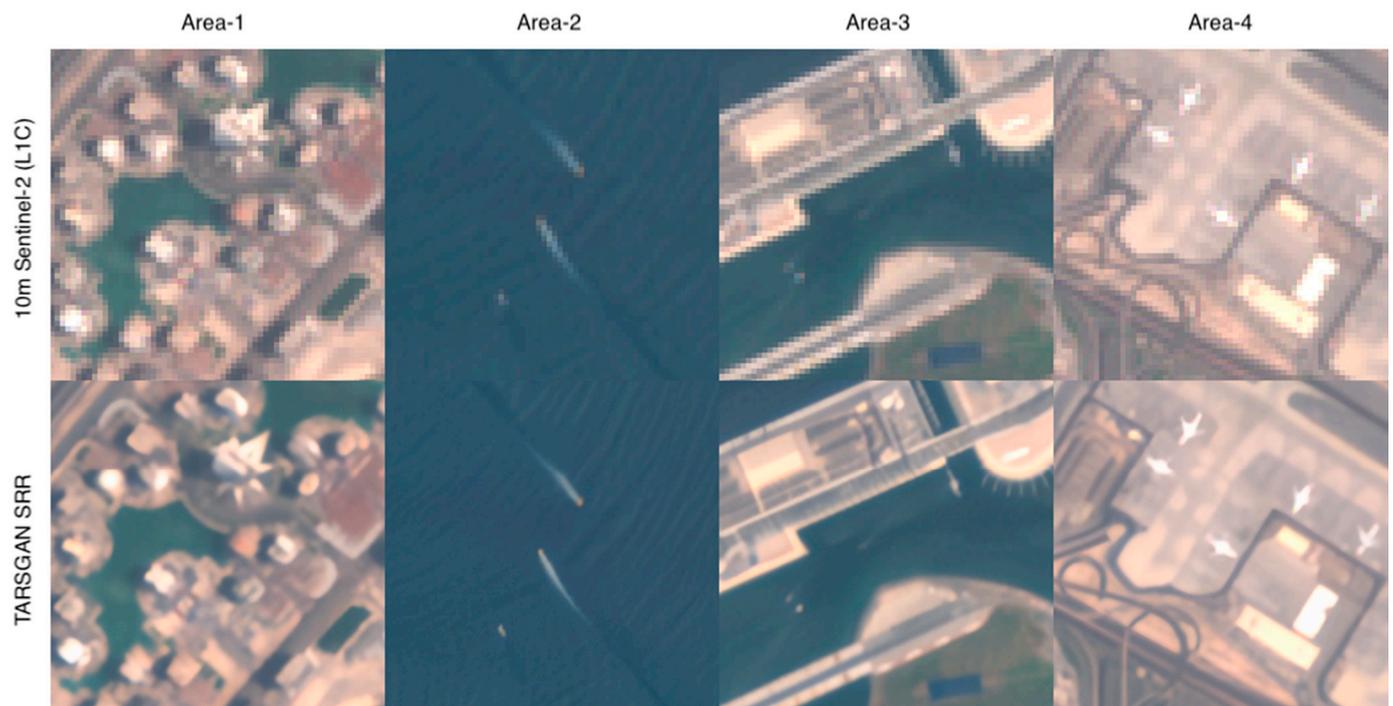

Figure 8. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L1C) and TARSGAN SRR result over Dubai, United Arab Emirates (Site-2). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2B_MSIL1C_20210528T064629_N0300_R020 T40RCN_20210528T084809).
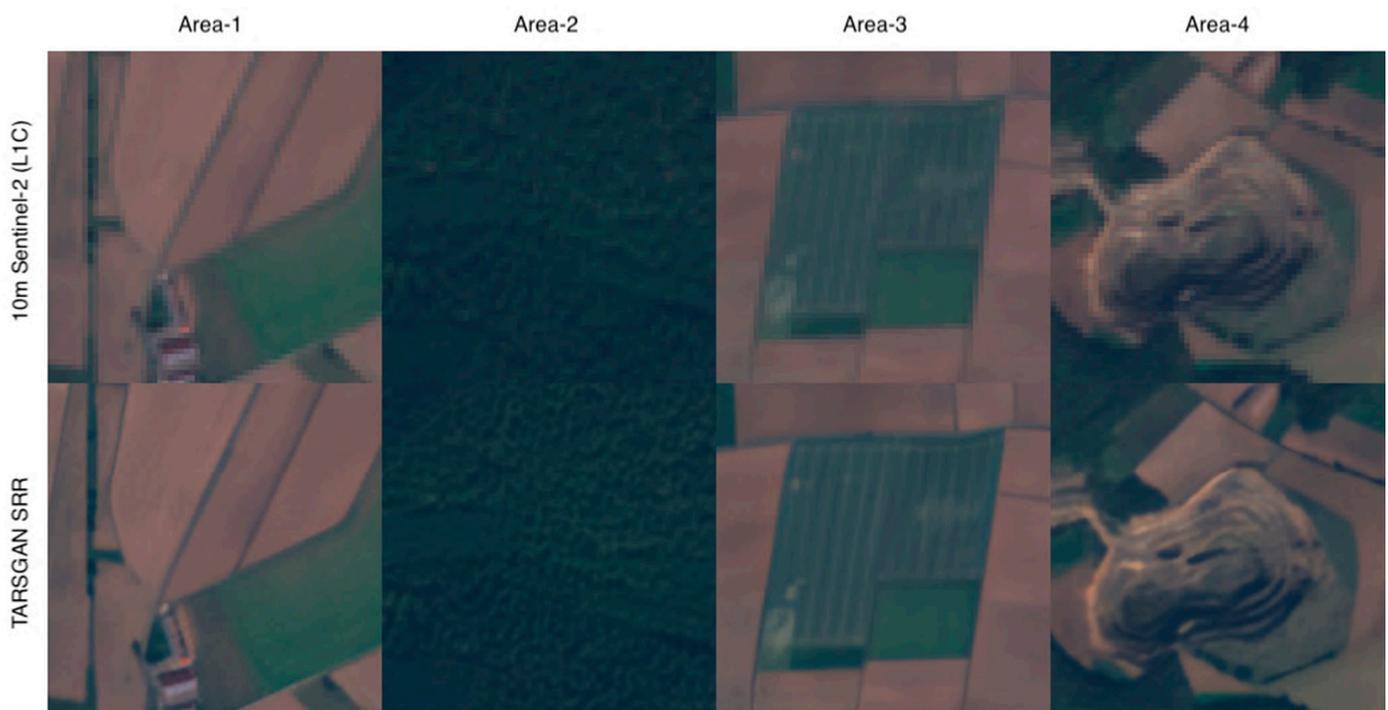
Figure 9 shows cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour image (L1C) and TARSGAN SRR results of Site-3, which is located over a forestry and agriculture area near Hainich, Germany. Area-1 shows yellow (possibly rapeseed) and green coloured farms and a farmhouse in the centre. Area-2 shows an area of forest. Area-3 shows gridded farms under some thin clouds. Area-4 shows a terraced field with farms on the ground. We can observe from the SRR image that the boundaries of the farms can be clearly identified in Area-1, 2 and 4. The farmhouse in Area-1 can be seen with clear outlines. Although the atmospheric clarity is very low in the L1C image, some of the individual trees in Area-2 can still be identified from the SRR image. In order to compare the SRR results from the atmospherically corrected L2A images, we show different cropped areas for the same site in Figure 10, focusing on the agriculture fields. In Figure 10, Area-1 shows some dark green coloured farms with farm roads. Area-2 shows a mixture of light green and dark green coloured farms with farm roads. Area-3 shows a small village surrounded by farms. Area-4 shows light green coloured farms with a road in the middle. We can observe from the SRR image that the boundaries of the farms were clearly brought out and the narrow farm roads are much more visible in comparison to the original Sentinel-2 image. There are no artefacts found from the SRR image for the eight areas of Site-3.

Figure 9. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L1C) and TARSGAN SRR result over Hainich, Germany (Site-3). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2A_MSIL1C_20200921T103031_N0209_R108 T32UNB_20200921T142406).

Figure 11 shows cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour image (L1C) and TARSGAN SRR result of Site-4, which is located over London, UK. Area-1 shows the London Bridge train station and small buildings. Area-2 shows very dense building blocks and bridges crossing the Thames river. Area-3 shows more urban building blocks. Area-4 shows bridges and ships on the Thames river. We can observe from the SRR image that the building blocks in Area-2 and 3 look more realistic than the original Sentinel-2 image. The outlines of the bridges in Area-2 and 4 are clearer and cars on the bridges in Area-4 are identifiable in the SRR image. Although image quality is lowered by haze in the L1C image, edges of the different objects in all four areas were significantly improved with SRR. In comparison, we show an L2A image and its corresponding SRR

result in Figure 12 for London over different landmarks. Although the L2A image has shown better contrast and more vivid colours, some stretching (intensity clipping) issues caused overexposures over very bright objects. Area-1 shows the millennium wheel ("London Eye") under some thin clouds. Area-2 shows the London Stadium. Area-3 shows the London commercial centre at Canary Wharf with many tower buildings. Area-4 shows Kensington Gardens under some thin clouds. We can observe from the SRR image that the super-resolved landmarks can be more easily identified from the SRR. In particular, the millennium wheel and the paths on the grass in Area-1 and the garden path of the Kensington park in Area-4 can all be better identified from the SRR image. Some buildings, in Area-2 and 3, are overexposed, but the shapes and outlines are clearer in SRR. There is no artefact found from the SRR image for the eight areas of Site-3.
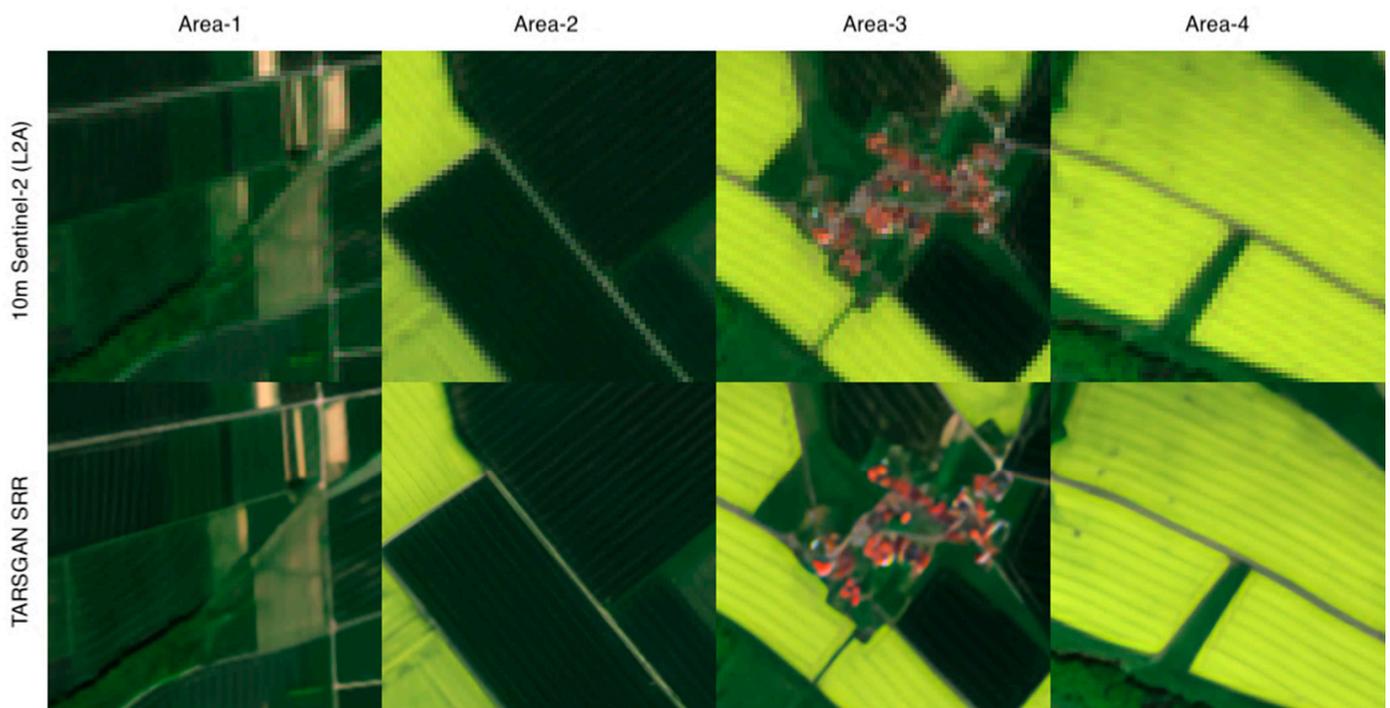
Figure 10. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L2A) and TARSGAN SRR result over Hainich, Germany (Site-3). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2B_MSIL2A_20210531T101559_N0300_R065 T32UNB_20210531T140040).

Figure 13 shows cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour image (L1C) and TARSGAN SRR result of Site-5, which is in a rural area over Desert Rock, near Sedona and Flagstaff, Arizona, U.S.A. Area-1 shows a mountain peak and a segment of the hillside road. Area-2 shows a mixture of desert and trees. Area-3 shows a yellow river in the desert. Area-4 shows a rural desert surface. Probably affected by the atmosphere clarity and lack of patterns, the quality of the SRR image for this site is lower than other sites. However, the shape of the mountain peak in Area-1, outlines of the individual trees in Area-2 and 4, and outline of the river in Area-3, in the SRR image have shown significant improvements over the original Sentinel-2 image.

Figure 11. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L1C) and TARSGAN SRR result over London, UK (Site-4). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2B_MSIL1C_20201217T111359_N0209_R137 T30UXC_20201217T132006).
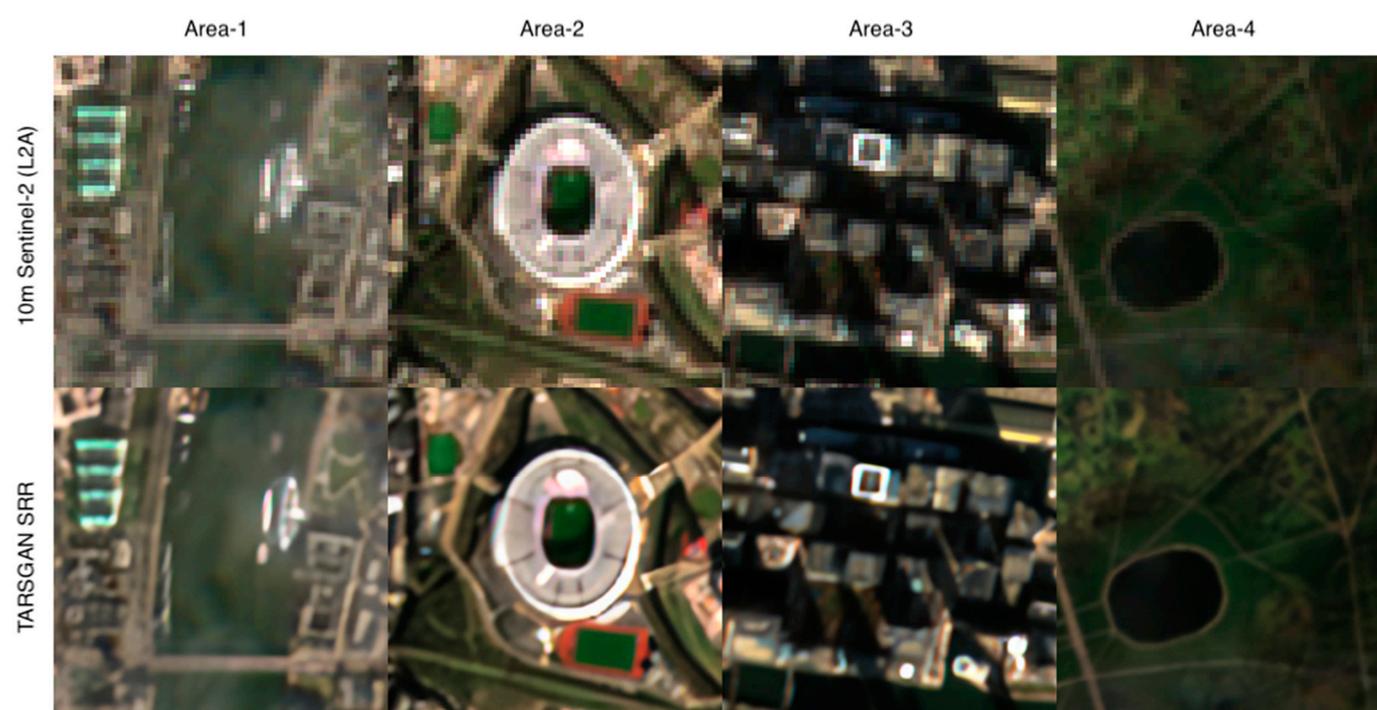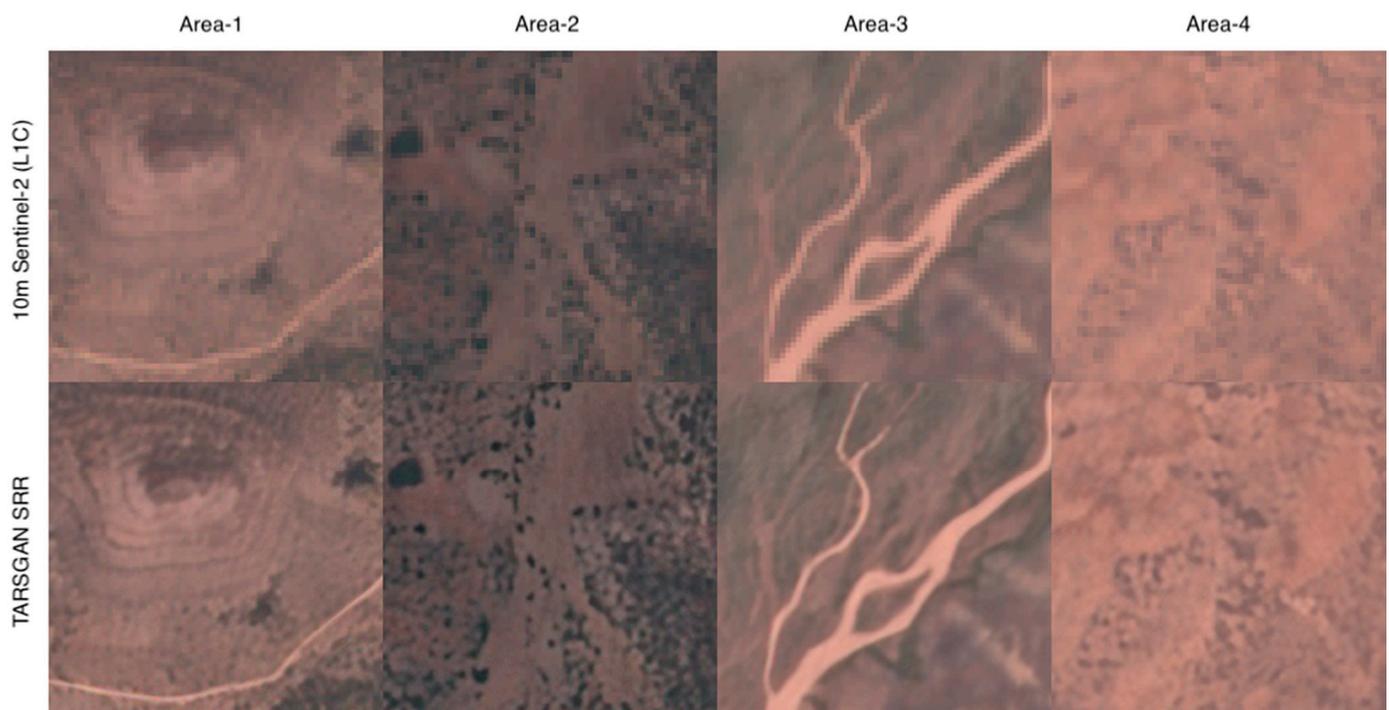

Figure 12. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L2A) and TARSGAN SRR result over London, UK (Site-4). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2A_MSIL2A_20210309T105901_N0214_R094 T30UXC_20210309T135358).

Figure 13. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L1C) and TARSGAN SRR result over Desert Rock, near Flagstaff, AZ, U.S. (Site-5). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2A_MSIL1C_20210507T180921_N0300_R084 T12SUD_20210507T215833).

Figure 14 shows cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour image (L1C) and TARSGAN SRR result of Site-6, which is located in the Lincoln Sea of Greenland. Area-1 shows sea ice floes with various cracks and leads. Area-2 shows small pieces of ice floating on the water. Area-3 shows a snow-covered mountain surface. Area-4 shows sea-ice, leads and open waters. We can observe from the SRR image that the outlines of sea-ice and the narrow path of the leads and cracks were clearly brought out in comparison to the original Sentinel-2 image.

In order to calculate the effective resolution of the TARSGAN SRR results, we perform ELF measurements for all afore-demonstrated examples. The measured $MFWHM(I_{LR})$, $MFWHM(I_{SRR})$, and the calculated ratio of $MFWHM(I_{LR})/MFWHM(I_{SRR})$, i.e., β, for all six sites (32 image crops), are summarised in Table 4. The effective resolution enhancement factor (last column of Table 4—Avg. α) are calculated using the averaged value of β for each site (second last column of Table 4—Avg. β) and the calibrated values of α and β ($\alpha \propto \beta$) are shown in Table 3. It should be noted that the exemplar image crops are very small (250 × 250 pixels), for the forest, sea, and desert crops, and as well as some crops that were affected by severe haze or under thin clouds, it is impossible to obtain any valid slanted edge ROIs from ELF (these missing values are marked as "-" in Table 4). The total averaged effective resolution enhancement factor (Avg. α) of 2.91 times suggests the TARSGAN SRR results have an averaged effective resolution of 3.44 m/pixel in comparison to the 10 m/pixel Sentinel-2 inputs.
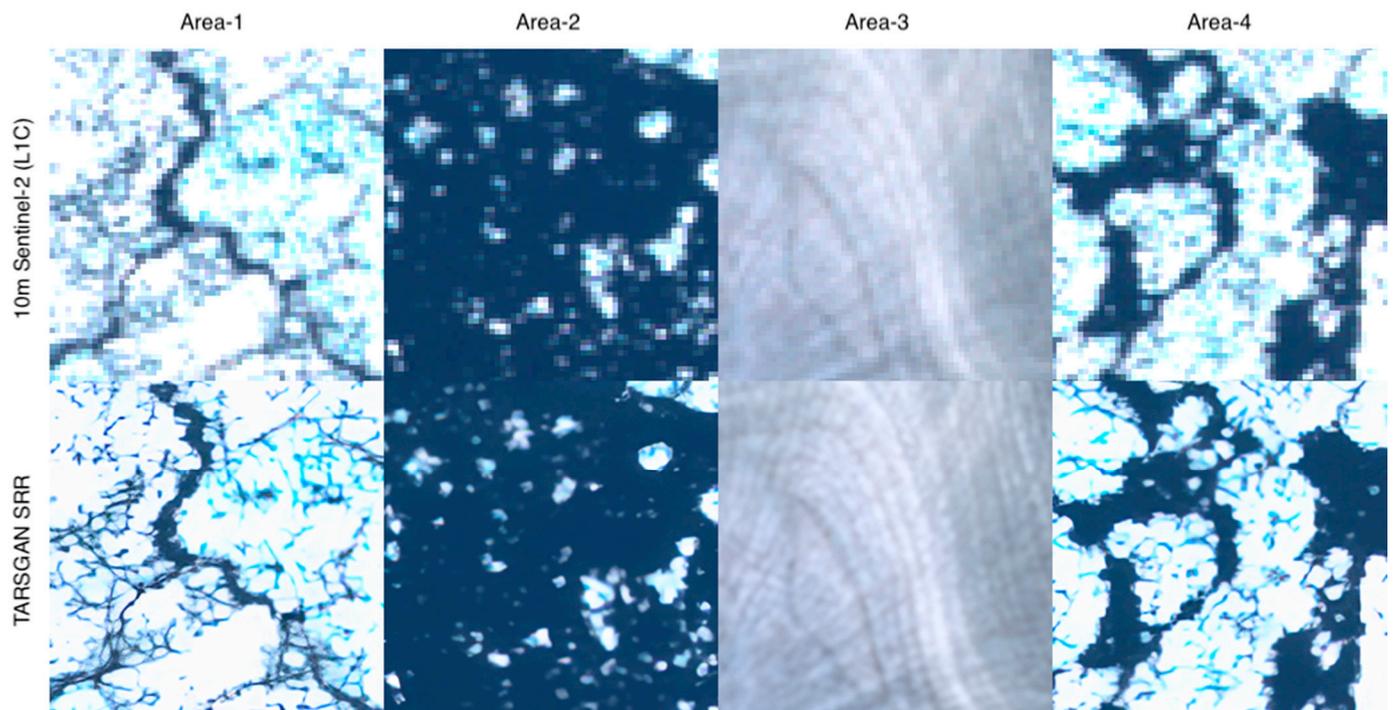
Figure 14. Cropped examples (625 m × 625 m) of the 10 m/pixel Sentinel-2 "true" colour images (L1C) and TARSGAN SRR result over Lincoln Sea, Greenland (Site-6). Please refer to Supplementary Material for full-size SRR (produced from Sentinel-2 S2A_MSIL1C_20200729T190921_N0209_R056 T21XWM_20200729T222945).

Table 4. Summary of the ELF measurements of all demonstrated examples and estimated effective resolution improvement factors.

Site#	$MFWHM(I_{LR})/MFWHM(I_{SRR}) = \beta$				Avg. β	Avg. α
	Area-1	**Area-2**	**Area-3**	**Area-4**		
Site-1	4.37/2.73 = 1.60	-	4.24/3.49 = 1.21	3.63/2.83 = 1.28	1.363	2.695 times
Site-2	4.7/3.95 = 1.19	-	3.80/3.08 = 1.23	4.30/3.40 = 1.26	1.227	2.520 times
Site-3 (L1C)	4.67/3.53 = 1.32	-	-	5.00/3.30 = 1.52	1.42	3.779 times
Site-3 (L2A)	3.35/2.60 = 1.29	4.27/2.93 = 1.46	5.08/3.78 = 1.34	4.15/2.80 = 1.48	1.393	3.160 times
Site-4 (L1C)	-	-	5.27/4.24 = 1.24	4.07/3.58 = 1.14	1.19	2.408 times
Site-4 (L2A)		-	4.75/3.86 = 1.23	3.79/3.44 = 1.10	1.165	2.332 times
Site-5	-	4.58/3.22 = 1.42	4.27/3.93 = 1.09	-	1.255	2.604 times
Site-6	3.68/3.12 = 1.18	-	-	4.88/2.94 = 1.66	1.42	3.779 times
Total Avg.			-		1.304	2.910 times

3.3. Results from Multispectral Bands

The proposed TARSGAN model can be used to improve the image effective resolution of any different multispectral bands without changing their spectral property. In order to demonstrate spectral invariance of TARSGAN SRR, we test on each individual band over two test images (both cropped to a size of 30 km × 30 km)—one from Site-2 (covering different surface features of urban, desert, and water; image ID: S2B_MSIL2A_20210528T064629_N0300_R020_T40RCN_20210528T091914) and the other one from Site-3 (covering forestry and agriculture features; image ID: S2B_MSIL2A_20210531T101559_N0300_R065_T32UNB_20210531T140040).
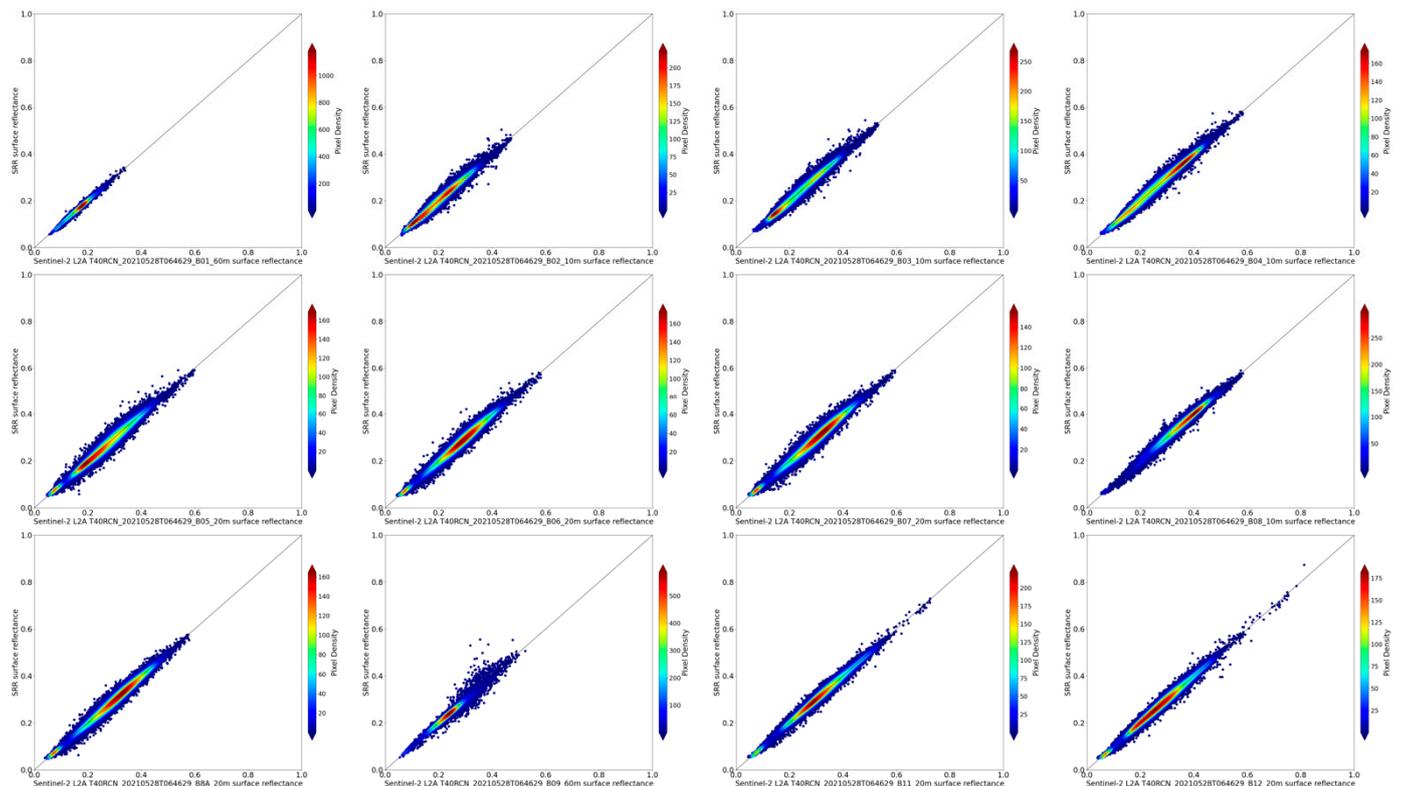
Figure 15 shows for Site-2 intercomparisons of all spectral bands of the SRR product against the original Sentinel-2 L2A surface reflectance product. It should be noted that the SRR images are down-sampled to the same scale as the L2A product in order to achieve the comparison. For 60 m bands, 100×100 pixels are plotted; for 20 m bands, 300×300 pixels are plotted for the same area, for 10 m bands, 300×300 pixels are plotted for a smaller area. For a detailed comparison of the full area of Site-2, please refer to the multispectral SRR product and corresponding Sentinel-2 L2A product, provided in the Supplementary Material.
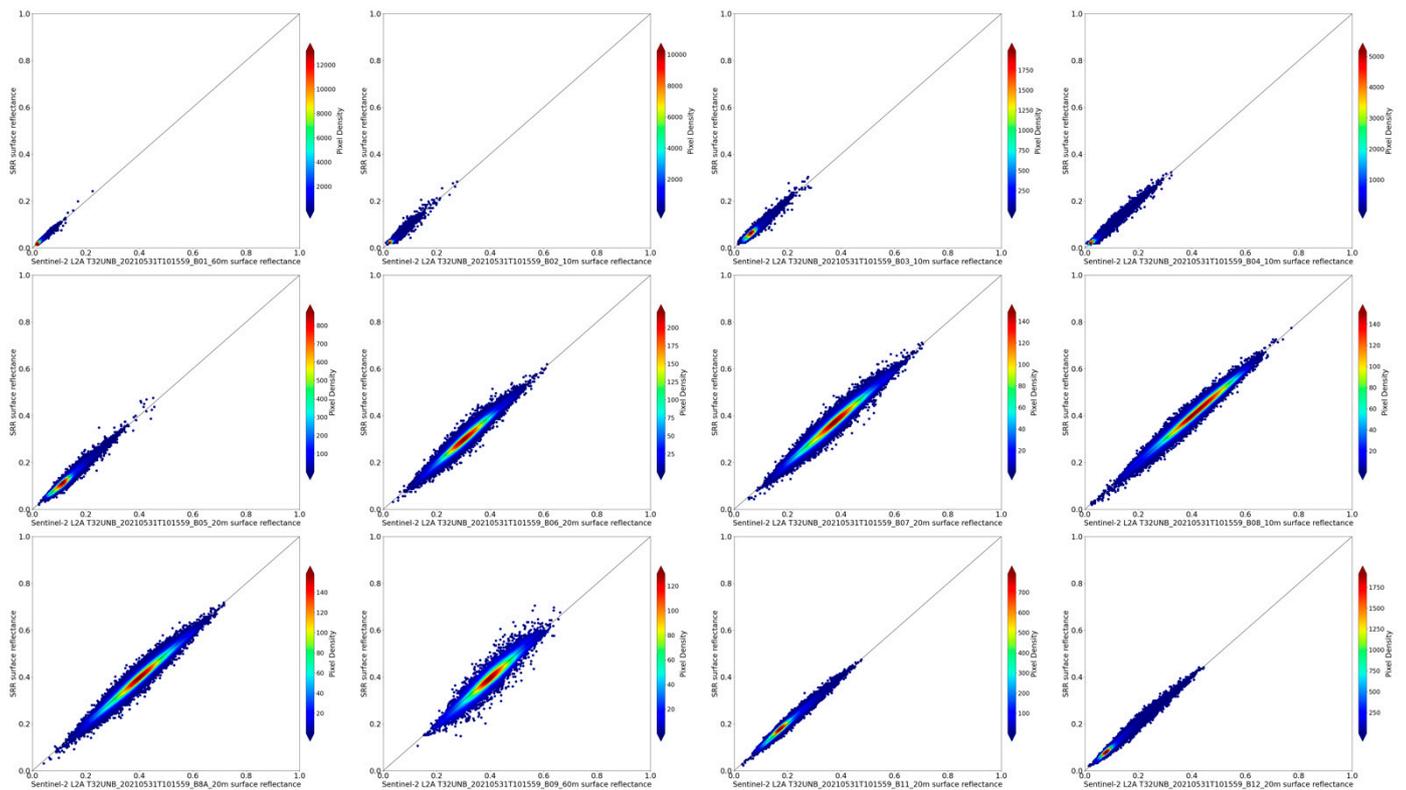

Figure 15. Site-2 intercomparisons of all spectral bands of the SRR product against the original Sentinel-2 L2A surface reflectance product (S2B_MSIL2A_20210528T064629_N0300_R020_T40RCN_20210528T091914).

Considering the resolution gap, the SSR surface reflectance values all show good correlations against the original Sentinel-2 L2A surface reflectance values—with the majority of the pixels lying on the 1:1 line, which can be observed from the individual scatter plots of Figure 15.

Figure 16 shows the Site-3 intercomparisons of all spectral bands of the SRR product against the original Sentinel-2 L2A surface reflectance product. The same sampling rates are used as Site-2 intercomparisons. Similarly, a good correlation between the multispectral SRR product and the original Sentinel-2 L2A product can be observed. For a detailed comparison of the full area of Site-3, please refer to the multispectral SRR product and corresponding Sentinel-2 L2A product, provided in the Supplementary Material.

Figure 16. Site-3 intercomparisons of all spectral bands of the SRR product against the original Sentinel-2 L2A surface reflectance product (S2B_MSIL2A_20210531T101559_N0300_R065_T32UNB_20210531T140040).

4. Discussion

4.1. From MARSGAN to TARSGAN

In this work, we explore the MARSGAN [27] model with the Sentinel-2 L1C and L2A images. Two practical modifications were applied over the original MARSGAN network to form a new model, which we call TARSGAN. The first modification is removing the adaptive weighted multi-scale reconstruction block and using pre-upsampled and blurred training LR images to achieve better network learning of the blurring effect of the LR images and thus better resolve small and burry objects. The second modification is replacing the perceptual loss used in MARSGAN with an SSIM loss to obtain better overall edge sharpness, whilst avoiding adding synthetic or stochastic textures or artefacts to the data. Figure 17 shows some examples from Site-1 demonstrating the impact of the two modifications to MARSGAN with the original 10 m/pixel Sentinel-2 image, the MARSGAN SRR as described in [27] (MARSGANv0), MARSGAN SRR trained with up-sampled and blurred LR dataset using perceptual loss (MARSGANv1), and the MARSGAN SRR trained with up-sampled and blurred LR dataset using the SSIM loss (TARSGAN). From Figure 17, we can observe the impact of using the proposed training strategy (MARSGANv1) that results in the sharpest edges among the three, but with synthetic appearance, and replacing the perceptual loss with SSIM loss, which eliminates the synthetic artefacts, while keeping compatible edge sharpness.
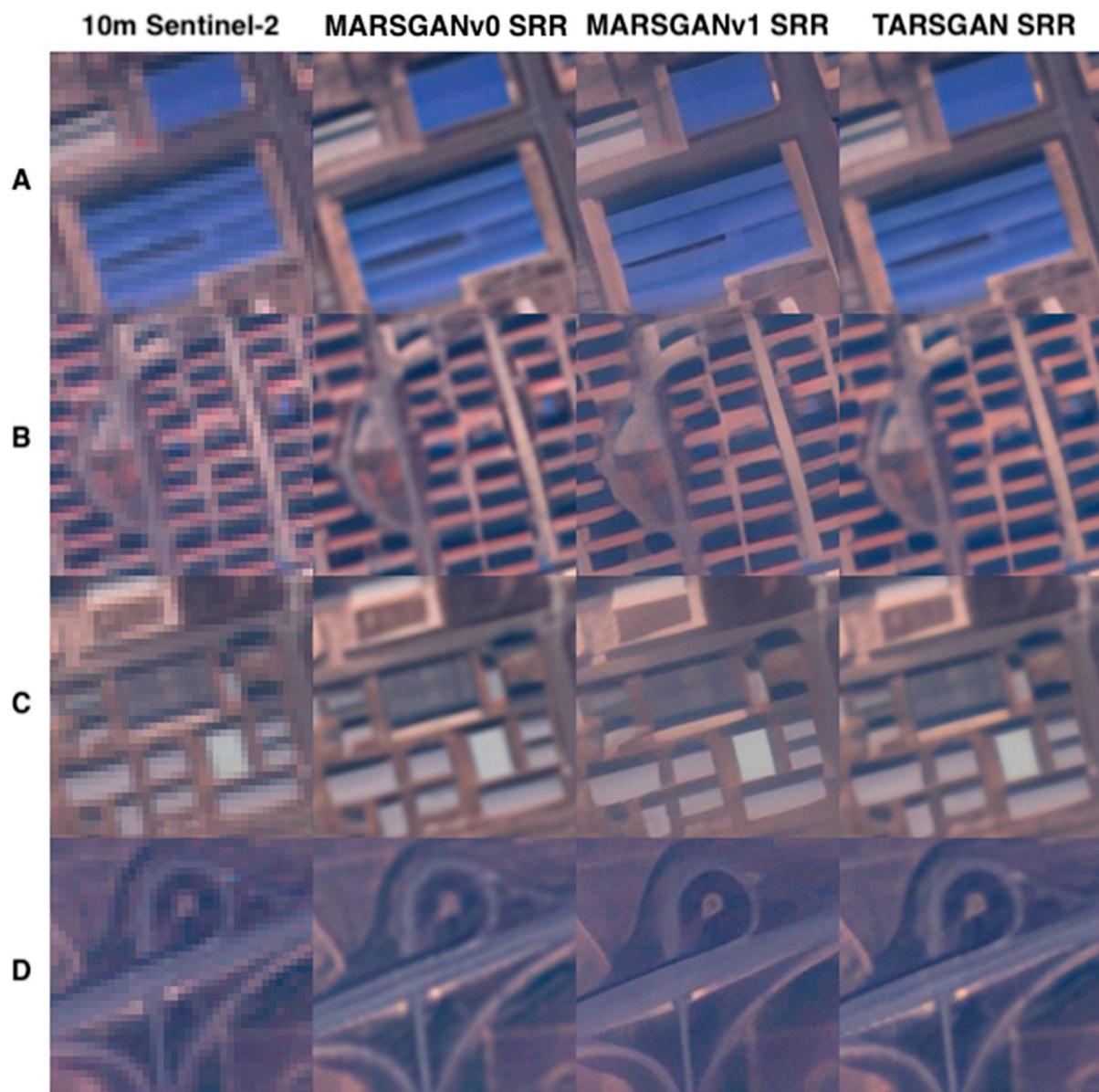
Figure 17. (**A**–**D**) Cropped examples of Site-1 of the original 10 m/pixel Sentinel-2 image, the MARSGAN SRR as described in [27] (MARSGANv0), MARSGAN SRR trained with upsampled and blurred LR using perceptual loss (MARSGANv1), and the MARSGAN SRR trained with upsampled and blurred LR using the SSIM loss (TARSGAN). All subfigures are self contrast stretched and have sizes of 625 m × 625 m.

4.2. Potential Improvements to TARSGAN and ELF

In this paper, we focused on the demonstration of the TARSGAN SRR results for Sentinel-2 "true" colour and multispectral images, alongside a demonstration of the automated image effective resolution assessment. The two proposed modifications of MARSGAN are considered light-touch to the original network architecture. In the future, we would like to test out more optimisation ideas to further improve the TARSGAN SRR results. For example, training the network in a cascaded manner, i.e., providing the intermediate 2× resolution HR as supervision data to better guide the network learning artefact-free 4× resolution enhancement. Or similar to [28], training a separate model to learn a comprehensive set of degradation effects of LR images, could also be helpful in constructing a better training dataset that not only focuses on the blurring effect of the LR images. Other work like separating the training datasets for different features (e.g.,

separating forestry scenes with urban scenes) may help improve the SRR performance for a particular category of the applications.

On the other hand, the proposed ELF system is capable of directly assessing the image effective resolution that is considered to be a better metric of representing the performance of an SRR algorithm, in comparison to the commonly used subjective quality metrics or perceptual indexes [44,45]. ELF does not need a reference HR image, which is generally difficult or expensive to obtain. However, ELF is not always applicable if test images are very small and do not contain sufficient edge features. In the future, we would like to extend the ELF system with automated circle feature detection and ring measurements to improve its robustness over scenes that have insufficient edge features. Moreover, combining the subjective quality metrics or perceptual indexes as well as automated detection of the smallest resolvable objects could also improve the performance of the image effective resolution assessment system.

4.3. A Future Streamlined SRR System

Over the last few years, deep learning-based techniques have achieved significant success in the field of SRR due to the richly available training datasets (as no labelling is required) and significantly faster processing speed in comparison to the traditional SRR approaches. However, some suboptimal restorations or occasional false predictions cannot be fully eliminated when using a single deep learning-based SRR network. Even with any state-of-the-art algorithms/models, the quality of the SRR results may still differ from dataset to dataset, from scene to scene, and from one area to another area. Therefore, we propose the concept of employing a streamlined SRR system on a GPU or GPU cloud server that is capable of achieving automated algorithms/networks selection and results assessments, which will provide the optimal solution for automated SRR processing of all sorts of input EO datasets. Figure 18 shows a conceptual implementation of this future streamlined SRR processing system using a processing scheduler with web-based data delivery, a combination of different SRR algorithms/networks as the "core processor", and a combination of the proposed ELF system and other image quality assessment metrics as the "quality assessor".
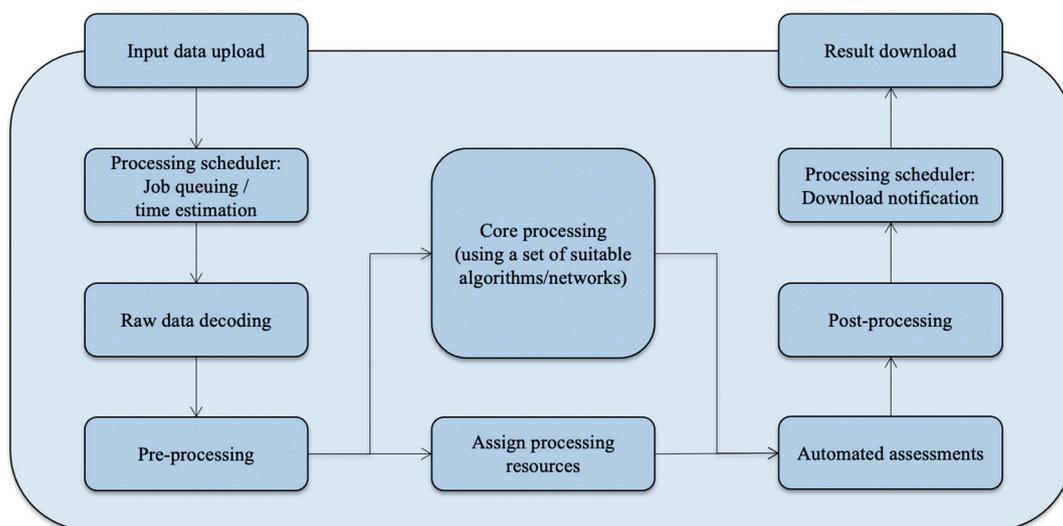

Figure 18. Proposed future streamlined SRR processing system based on automated SRR and quality assessments.

5. Conclusions

In this paper, we introduced the TARSGAN SRR model and the ELF image-effective resolution assessment system. We demonstrated TARSGAN SRR results using the 10 m/pixel Sentinel-2 "true" colour images over a wide range of different types of natural and artificial

surface features/targets. The ELF measurements show an averaged effective resolution enhancement factor of about 2.91 times over the passive SRR upscaling factor of 4 times. This suggests an effective resolution of ~3.44 m/pixel achieved with TARSGAN SRR over the 10 m/pixel bands. In addition, the multispectral properties of the TARSGAN SRR images were demonstrated to have good correlation, considering the resolution gap of 4 times, in comparison to the original Sentinel-2 images for all spectral bands. This suggests multispectral applications (e.g., calculation of multispectral indices indicative of crop health/stress and potential yield in precision agriculture) could be seamlessly applied using the TARSGAN super-resolved images but with better precision. We believe the demonstrated Sentinel-2 TARSGAN SRR system has potential for new applications in a variety of different fields, such as planning infrastructure, public services and monitoring of small urban targets in the field of urban intelligence; providing field scale mapping and boundary management at a global scale in the field of agriculture; achieving better detection and classification accuracy of different science targets, e.g., sea-ice leads and melt-ponds, in the field of oceanology and geology.

References

1. Van Ouwerkerk, J.D. Image super-resolution survey. *Image Vis. Comput.* **2006**, *24*, 1039–1052. [CrossRef]
2. Shah, A.J.; Gupta, S.B. Image super resolution-a survey. In Proceedings of the 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking 2012, Surat, India, 19–21 December 2012; pp. 1–6.
3. Ha, V.K.; Ren, J.; Xu, X.; Zhao, S.; Xie, G.; Vargas, V.M. Deep Learning Based Single Image Super-Resolution: A Survey. In Proceedings of the International Conference on Brain Inspired Cognitive Systems, Xi'an, China, 7–8 July 2018; Springer: Cham, Switzerland, 2018; pp. 106–119.

4. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef] [PubMed]

5. Tsai, R.Y.; Huang, T.S. Multipleframe Image Restoration and Registration. In *Advances in Computer Vision and Image Processing*; JAI Press Inc.: New York, NY, USA, 1984; pp. 317–339.

6. Keren, D.; Peleg, S.; Brada, R. Image sequence enhancement using subpixel displacements. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Ann Arbor, MI, USA, 5–9 June 1988; pp. 742–746.

7. Kim, S.P.; Bose, N.K.; Valenzuela, H.M. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1013–1027. [CrossRef]

8. Bose, N.K.; Kim, H.C.; Valenzuela, H.M. Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes. In Proceedings of the IEEE Conference Acoustics, Speech and Signal Processing, Minneapolis, MN, USA, 27–30 April 1993; Volume 5, pp. 269–272.

9. SRhee, H.; Kang, M.G. Discrete cosine transform based regularized high-resolution image reconstruction algorithm. *Opt. Eng.* **1999**, *38*, 1348–1356.

10. Hardie, R.C.; Barnard, K.J.; Armstrong, E.E. Joint MAP registration and high resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.* **1997**, *6*, 1621–1633. [CrossRef]

11. Farsiu, S.; Robinson, D.; Elad, M.; Milanfar, P. Fast and robust multi-frame super-resolution. *IEEE Trans. Image Process.* **2004**, *13*, 1327–1344. [CrossRef]

12. Yuan, Q.; Zhang, L.; Shen, H. Multiframe super-resolution employing a spatially weighted total variation model. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 379–392. [CrossRef]

13. Tao, Y.; Muller, J.-P. A novel method for surface exploration: Super-resolution restoration of Mars repeat-pass orbital imagery. *Planet. Space Sci.* **2016**, *121*, 103–114. [CrossRef]

14. Tao, Y.; Muller, J.-P. Super-Resolution Restoration of Spaceborne HD Videos Using the UCL MAGiGAN System. In *Image and Signal Processing for Remote Sensing XXV*; SPIE: Strasbourg, France, 2019; pp. 1115508-1–1115508-7.

15. Tao, Y.; Muller, J.-P. Super-resolution restoration of MISR images using the UCL MAGiGAN system. *Remote Sens.* **2019**, *11*, 52. [CrossRef]

16. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.

17. Yu, J.; Fan, Y.; Yang, J.; Xu, N.; Wang, Z.; Wang, X.; Huang, T. Wide activation for efficient and accurate image super-resolution. *arXiv* **2018**, arXiv:1808.08718.

18. Ahn, N.; Kang, B.; Sohn, K.A. Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European Conference on Computer Vision (ECCV) 2018, Munich, Germany, 8–14 September 2018; pp. 252–268.

19. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1637–1645.

20. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 3147–3155.

21. Wang, C.; Li, Z.; Shi, J. Lightweight image super-resolution with adaptive weighted learning network. *arXiv* **2019**, arXiv:1904.02358.

22. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. ; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV) 2018, Munich, Germany, 8–14 September 2018; pp. 286–301.

23. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.

24. Sajjadi, M.S.; Scholkopf, B.; Hirsch, M. EnhanceNet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision 2017, Venice, Italy, 22–29 October 2017; pp. 4491–4500.

25. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops 2018, Munich, Germany, 8–14 September 2018.

26. Rakotonirina, N.C.; Rasoanaivo, A. ESRGAN+: Further improving enhanced super-resolution generative adversarial network. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020, Barcelona, Spain, 4–8 May 2020; pp. 3637–3641.

27. Tao, Y.; Conway, S.J.; Muller, J.-P.; Putri, A.R.D.; Thomas, N.; Cremonese, G. Single Image Super-Resolution Restoration of TGO CaSSIS Colour Images: Demonstration with Perseverance Rover Landing Site and Mars Science Targets. *Remote Sens.* **2021**, *13*, 1777. [CrossRef]

28. Sun, W.; Chen, Z. Learned image downscaling for upscaling using content adaptive resampler. *IEEE Trans. Image Process.* **2020**, *29*, 4027–4040. [CrossRef] [PubMed]

29. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2019, Seoul, Korea, 27 October–2 November 2019; pp. 3086–3095.

30. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 391–407.

31. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1874–1883.

32. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661. [CrossRef]

33. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.

34. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.

35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

36. Godard, C.; Mac Aodha, O.; Firman, M.; Brostow, G.J. Digging into self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3828–3838.

37. Alhashim, I.; Wonka, P. High quality monocular depth estimation via transfer learning. *arXiv* **2018**, arXiv:1812.11941.

38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

39. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

40. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]

41. Hart, P.E.; Duda, R.O. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* **1972**, *15*, 11–15.

42. Li, C.R.; Tang, L.L.; Ma, L.L.; Zhou, Y.S.; Gao, C.X.; Wang, N.; Li, X.H.; Zhou, X.H. A comprehensive calibration and validation site for information remote sensing. *ISPRS-IAPRSIS* **2015**, *XL-7/W3*, 1233–1240. [CrossRef]

43. Zhou, Y.; Li, C.; Tang, L.; Wang, Q.; Liu, Q. A Permanent Bar Pattern Distributed Target for Microwave Image Resolution Analysis. *IEEE Geosci. Rem. Sens.* **2017**, *14*, 164–168. [CrossRef]

44. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]

45. Venkatanath, N.; Praneeth, D.; Chandrasekhar, B.M.; Channappayya, S.S.; Medasani, S.S. Blind Image Quality Evaluation Using Perception Based Features. In Proceedings of the 21st National Conference on Communications (NCC) 2015, Mumbai, India, 27 February–1 March 2015.