ENHANCED SUPER RESOLUTION FOR REMOTE SENSING IMAGERIES

R.G.C.J. Kapilaratne [1, *], S. Kakuta[1], S. Kaneta [1]

[1] Asia Air Survey Co. Ltd., Asao Ward, Kawasaki City, Kanagawa, Japan (jie.kapilaratne, stm.kakuta, sni.kaneta)@ajiko.co.jp

KEY WORDS: Remote Sensing**,** Super Resolution**,** ESRGAN, Deep Learning, WorldView-3, Sentinel-2.

ABSTRACT:

Single image super resolution (SISR) technology has been attracted much attention from remote sensing community due to its proven potentials in remote sensing applications. Existing SISR techniques varying from conventional interpolation methods to different network architectures. Generative adversarial networks (GANs) are one of the latest network architectures proven a greater potential as a SISR method whereas least attention has been given by the remote sensing community. Several studies have already been carried out on this context. However, yet there is no generalized GAN based approach to super resolve remote sensing imageries. Therefore, this study investigated the potentials of enhanced super resolution generative adversarial (ESRGAN) model to super resolve very high to medium resolution images from high to coarse resolution images for remote sensing applications. Two models were trained and Worldview-3 (WV3) images used as for very high resolution images. Whereas, down sampled WV3 and Sentinel-2(S2) were used as low resolution counterparts. Model performances were qualitatively and quantitatively analysed using standard metrics such as PSNR, SSIM, UIQI, CC, SAM, SID. Evaluation results emphasised super resolved images were preserved the original quality of the satellite images to a greater extent while improving its ground resolution.

1. INTRODUCTION

Remote sensing images, specifically the images taken from satellite and airborne platforms are crucial for applications such as environmental surveillance, hazard monitoring, traffic mapping, agriculture monitoring, oceanography, hydrology etc. due to their capability to monitor wider area within a shorter period of time (Drusch et al., 2012). However, the specific application of the remote sensing images is determined by its spatial resolution. For instance, high resolution (HR) images contains more descriptive information and are crucial for the applications such as disaster damage detection in hazard monitoring. With the recent advancements in satellite sensors, images with very high ground resolution or spatial resolution covering all parts of the earth is widely available. However, the use of such HR satellite images for larger area and multi temporal analysis is mostly impractical due to the cost constraint. Moreover at the time of a disaster, images are rarely taken at nadir and mostly the sensor will be tilted to rapid grasp of a wider area. Such oblique images (Figure 1) hinder its usability for damage extractions hence a further resolution improvement techniques are required.
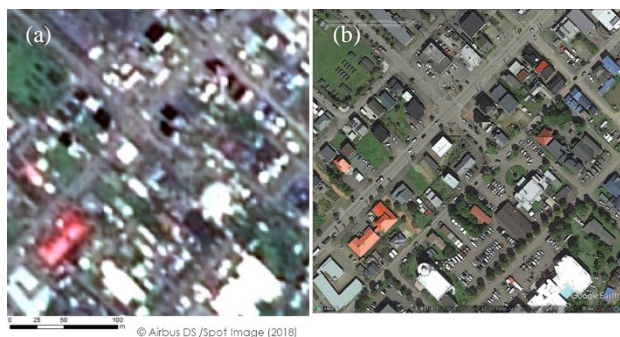
Figure 1. (a) Image taken on 2018/09/11 with ground resolution (GR) 1.5m (b) respective aerial image taken at same location on 2018/09/11 with 0.1m GR.

In contrary, regardless of the fact that the earth observation data sets covering all parts of the earth at 5 days temporal resolution is freely provided by the Sentinel-2 missions of European Space Agency (ESA) its popularity for post processing applications such as damage detection or feature extraction is quite low (Ma et., al, 2019) due to its coarse spatial resolution (Figure 2). Therefore, a technology which allow to improve original resolution of the satellite images would be desirable for comprehensive use of remote sensing images in the contexts of applications as well as for a broader user community.
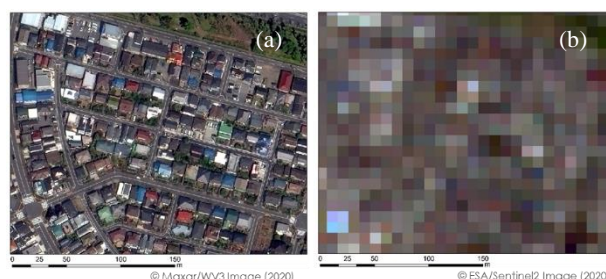
Figure 2. (a) WV3 image taken on 2020/04/02 with ground resolution (GR) 0.3m (b) respective S2 image taken at same location on 2020/04/02 with 10.0m GR.

1.1 Single Image Super Resolution

Since 2012, deep learning became a prominent tool for computer vision and image processing tasks such as object detection, segmentation etc. (Galar et al., 2019). Less than a decade, its applicability has grown beyond the standard applications. Remote sensing is a key field which benefitted immensely from the recent advancements in deep learning.

Image fusion was the only option for the resolution improvements in remote sensing images. One of the major draw backs of image fusion technology was the requirement of panchromatic band along with the multispectral band as some of

*Corresponding author

the satellite sensors such as Copernicus mission of S2 only provides the multispectral imageries.

Single image super resolution (SISR) is one of recent advancements which attracted much attention of remote sensing community due to its greater potential to add extra value to remote sensing images without using additional information (Yang et al., 2018, Pan et al., 2013, Yang et al., 2008).

In literature, several methods (Lim et al., 2017, Kim et al., 2019, Zhang et al., 2014, Dong et al., 2014) have been tested for SISR ranging from conventional interpolation methods (e.g. linear, bi-cubic etc.) to different network architectures from standard CNNs to generative adversarial networks (GANs). As of authors' knowledge about existing literature, least attention has been given to GANs and underestimated its potential in SISR by remote sensing community. Most prominent concern is the possible spectral discrepancies in generated images.

Mehmood A., (2020), Galar et al., (2018, 2019) and Romero et al., (2020) are some of the recent studies attempted to apply GAN models for resolution improvements in images for remote sensing applications. However, authors have learnt from existing literature that the adequate attention has not been given to the alleviation in spectral quality of the generated images by some studies. Further, several studies are carried out exclusively with selected data sets (Galar et al., 2019) therefore, its generality is unclear along with its applicability for the worst case scenarios such as oblique images.

Therefore, this study is an initial communication of a series of studies to develop generalized SISR technique based on GAN models for resolution improvements without degrading the quality and value of the satellite imageries. The objective of this study is to investigate the potentials and applicability of enhanced super resolution generative adversarial networks (ESRGANs) as a general method to super resolve remote sensing images at different resolution levels. Two experiments were designed and the objective of the first experiment (EXP1) to evaluate the model performance for resolve the original resolution of the satellite images from degraded images.

And from the experiment 2 (EXP2) a model was trained to resolve S2 images by four folds. Through these experiments authors expected to cater two prevailing concerns with remote sensing data usage. The 1st and the foremost is to develop a generalized methodology to improve the image quality of satellite images captured even at severe conditions without extra information. And the 2nd concern is to support a broader remote sensing user community for comprehensive analysis using largest earth observation satellite data sets available at no cost.

2. MATERIALS AND METHODS

2.1 Network Architecture

GANs were proposed by Goodfellow et al., (2016) and have been widely used in super resolution tasks due to their ability to generate more photo realistic outputs with rich texture and quality (Romero et al., 2020). As shown in the Figure 3, both generator network and discriminator network train simultaneously with training data samples with an aim to train the generator network which can generate fake images in such a way that the discriminator could not be able to distinguish it as a fake image.

This study has adopted one of the state of art network ESRGAN (Wang et al., 2018) due to its novelty, proven success and the usability. ESRGAN model was developed to further enhance the visual quality of the resolved image by improving three key components of SRGAN model, namely network architecture, adversarial loss and perceptual loss function (Xintao et al., 2018).

Figure 3. Overview of the generative adversarial network.

ESRGAN model was introduced the Residual in Residual Dense Block (RRDB) (Figure 4) without batch normalization to the original SRGAN model. With the aforementioned advancements ESRGAN model generated much realistic and natural textures eliminating the prominent blurring effects found at SRGAN outputs (Xintao et al., 2018).
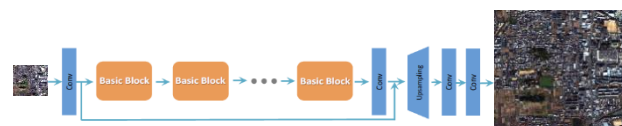

**Figure 4**. Basic network architecture of SRResNet model.

2.2 Datasets and Study Area

As explained in the introduction section, this study carried out two experiments. HR and LR datasets used for each experiment along with imaging data are included in Table 1. All datasets used in this study are captured in Japan (Figure 5). EXP1 was designed to evaluate the model capability to recreate original resolution of WV3 images from down-sampled WV3 images. The test image selected for EXP1 was taken at relatively high nadir angle and low ground resolution in comparison to training data set with a motive to test the capabilities of ESRGAN model on images taken at non-ideal conditions.

Experiment	HR data	LR data
EXP1	WV3 [0.3m]	WV3[1.2m]
	Itabashi (2020/04/04)	
	Joban (2021/02/14)	
	Mashiki (2016/11/01)	
	Mitaka(2020/08/15)	
	Nara(2020/11/18)	
	Nihonmatsu[0.5m](2021/02/14)	
EXP2	WV3 [2.5m]	S2[10.0m]
	Itabashi (2020/04/02)	
	Joban (2021/02/13)	
	Mashiki (2016/12/28)	
	Mitaka(2020/08/15)	
	Nara(2020/11/26)	
	Nihonmatsu (2021/03/21) : S2 image	
	Ichihara (2020/01/03): S2 image	

Table 1. Summary of the data sets used in respective experiment (image names written in bold italic corresponds to test data sets).

The motive of the EXP2 was to resolve S2 images by four folds. Down-sampled WV3 images into 2.5m used as HR images for corresponding LR S2 images. As shown in Table 1 Nihonmatsu image along with Ichihara S2 images are used as test data sets. Longest time lag observed for Nihonmatsu image among the used data set. S2 Ichihara image is specifically selected investigate the model's performance in in urban landscapes as Nihonmatsu image covers mountainous region. However, Ichihara image has been included only for qualitative analysis due to unavailability of the corresponding HR image for a quantitative analysis.
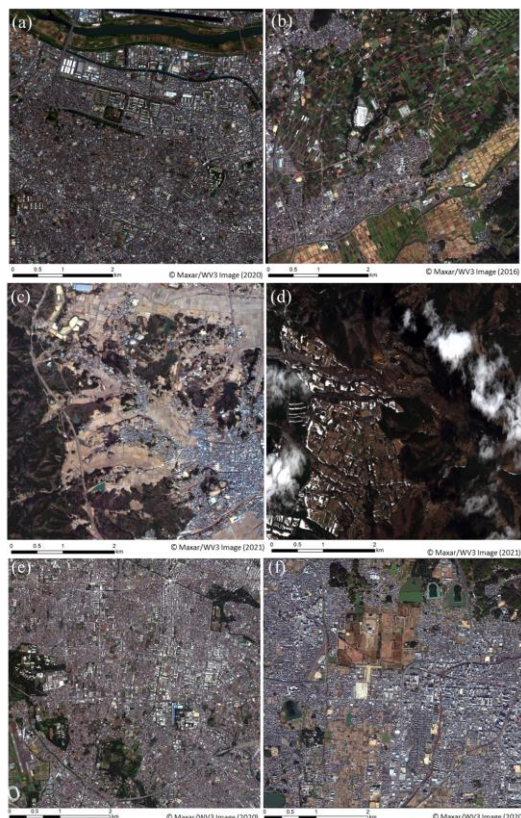
Figure 5. WV3 images used for EXP1 and 2 (a) Itabashi (b) Mashiki (c) Joban (d) Nihonmatsu (e) Mitaka (f) Nara.
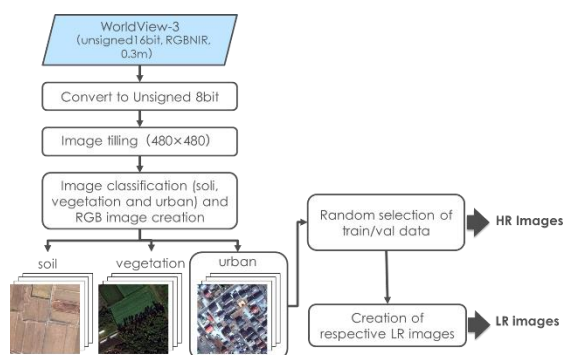
Figure 6. Overview of dataset preparation for EXP1.

2.2.1 Dataset Preparation EXP1: As shown in Figure 6, WV3 images were converted to 8bit from 16bit and then created the image tiles before classify them into soil, vegetation and urban. Thereafter, image samples exceeding threshold value of 0.5 for NDVI (normalized differential vegetation index) and NDSI (normalized differential soil index) indices are excluded from data set expecting better training performance on urban landscapes. Subsequently, training and validation image samples were randomly selected based on 8:2 ratio and created respective LR images.

2.2.2 Dataset Preparation EXP2: There are three main differences in dataset preparation of EXP1 with EXP2. First and the foremost is the downgrade the original resolution of WV3 from 0.3m to 10.0m. Resampling was performed in stepwise by 2x to minimize the image detail loss at direct downsampling of the satellite images. WV3 image samples at 2.5m resolution are used as HR whereas, corresponding 10.0m resolution samples as LR during model pre-training phase. The second difference is the during the actual model training with WV3-S2 data sets, resampled WV3 image tiles(2.5 m) created with 50% overlap are used as HR and respective S2 image tiles are used as LR image samples (Figure 7). The third difference is that the no image tile classification has performed due to less amount of image tiles. Number of training and validation samples used for each experiments is summarized at Table 2.

Experiment	training	validation	test
EXP1	2259	565	441
EXP2	389	97	124

Table 2. Number of image pairs used for each experiment.

2.3 Network Training

ESRGAN model was implemented as per the guideline given in the BasicSR library (Xintao et al., 2018) developed under PyTorch (Paszke et al., 2017) framework. Authors found that the pertained networks are not necessary for EXP1 based on the studies carried out by Lanaras et al., (2018), Galar et al., (2019) and Romero et al., (2021) due to the fact that common data source used for both LR and HR image tiles. Subsequently, model pre-training has been carried out in EXP2 with only WV3 images due to different data sources are used in HR and LR.

Random flips, rotation and crops were used as data augmentation option at the network training. Models were trained with a learning rate of 4×10^{-5} and batch size of one as it has been demonstrated as effective in image tasks (Ulyanov et al., 2016).
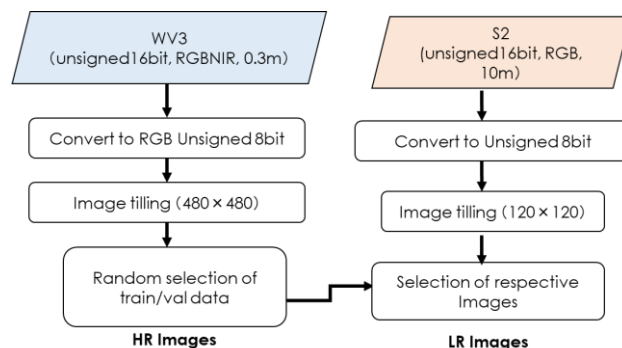
Figure 7. Overview of dataset preparation for EXP2.

2.4 Evaluation Methodology

Model performances were evaluated qualitatively and quantitatively. At the validation phase widely used matrices for super resolution such as Peak signal to noise ratio (PSNR) and structural similarity index (SSIM) introduced by Zhou et al., (2004) were applied to quantify the image quality degradation along with the other indices mentioned in the text.

The higher the index value the better the quality of generated image with respect to ground truth image. An extended evaluation was carried out both at the validation as well as test phase with special emphasis to evaluate the deterioration of spectral quality of the generated image with corresponding ground truth image. Consequently, following indices which estimate pixel wise error for all bands of the images were also used.

2.4.1 Spectral Angle Mapper (SAM):
SAM attempts to obtain the angles formed between reference spectrum and the image spectrum treating them as vectors in a space with a dimensionality equal to the number of bands ((Kruse et al.,1993, Boardman, 1992). A value of cos (SAM) equal to one denotes none existence of spectral deviations with its ground truth image (Cetin et al, 2009). Thus, SAM measures how realistic is the spectral distribution of a reconstructed pixel regardless of the absolute brightness (Lanaras et al., 2018). However, Carvalho and Meneses (2000) argued that even though there is an apparent difference in original and tested image, the cos (SAM) shows a high correlation (close to 1) that does not reflect the truth. In such cases Pearson Correlation is tend to be more accurate as it ranges from -1 to 1.

$$SAM(X,Y) = \arccos\left(\frac{X.Y}{|X|_2|Y|_2}\right) \qquad (1)$$

Where, X and Y are generated image and corresponding ground truth respectively.

2.4.2 Universal Image quality Index (UIQI):
UIQI (Wang and Bovik, (2002)) estimates the difference between two images based on three factors namely, loss of correlation, luminance and contrast. The index is used as a measure of spectral quality of the output image, the higher the value closer to one the better the image quality (Cetin and Musaoglu, 2009). UIQI defined as

$$UIQI = \frac{\sigma_{xy}}{\sigma_x\sigma_y}\frac{2\bar{x}\bar{y}}{(\bar{x})^2+(\bar{y})^2}\frac{2\sigma_x\sigma_y}{\sigma_x{}^2+\sigma_y{}^2} \qquad (2)$$

Where,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i \ , \bar{y} = \frac{1}{N}\sum_{i=1}^{N}y_i$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2, \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}).$$

2.4.3 Spectral Information Divergence (SID):
SID views each pixel spectrum as a random variable, and measures the band-band variability. Thus, Chang, (1999) argued that SID is relatively effective index for determining the similarity and variability in the context of spectral information at a pixel than SAM. Zero SID value indicates the original image spectral quality is fully preserved. SID at a pixel (x, y) is defined as follows.

$$SID(x,y) = \sum_{i=1}^{l}\left[\frac{I_X(x,y,\gamma_i)}{\sum_{j=1}^{l}I_X(x,y,\gamma_j)} - \frac{I_X(x,y,\gamma_i)}{\sum_{j=1}^{l}I_X(x,y,\gamma_j)}\right]\left[log\frac{I_X(x,y,\gamma_i)}{\sum_{j=1}^{l}I_X(x,y,\gamma_j)} - log\frac{I_X(x,y,\gamma_i)}{\sum_{j=1}^{l}I_X(x,y,\gamma_j)}\right] \qquad (3)$$

Where,

I_X: generate image
I_Y : ground truth image
$I(x,y,\gamma_i)$: spectral reflectance at pixel of image of band i
$m*n*l$: image dimensions
$x = 1,...,m$
$y = 1,...,n$
$\gamma_i (i= 1,...,l)$

2.4.4 Correlation Coefficient (CC):
CC computes the linear correlation of each pixel of the generated image with its ground truth image averaged over all 03 bands. Positive one indicates the positive correlation between generated and original image.

As for the comprehensive comparison of spectral values, histograms and scatter plots of generated and corresponding ground truth images are also included where it's necessary. Index values are estimated for all three bands (red, green, blue) used in the true colour composites.

3. RESULTS AND DISCUSSION

This section presents the model performance at validation and test phase. Analysis has carried out and results are summarised qualitatively and quantitatively as per the availability of ground truth data. For EXP1 quantitative results are given in terms of mean and standard deviation of the evaluation indices presented in sub section 2.4. Further, a validation of reflectance values of super resolved images were carried out for selected image samples from test data sets of EXP1 and EXP2 by comparing histograms and scatterplots of their corresponding ground truth samples.

3.1 Validation Results

This sub section discusses the qualitative and quantitative results achieved at validation phase of each experiment. Results are presented in the forms of figures as well as tables for a thorough analysis. Qualitative analysis followed by a quantitative analysis.
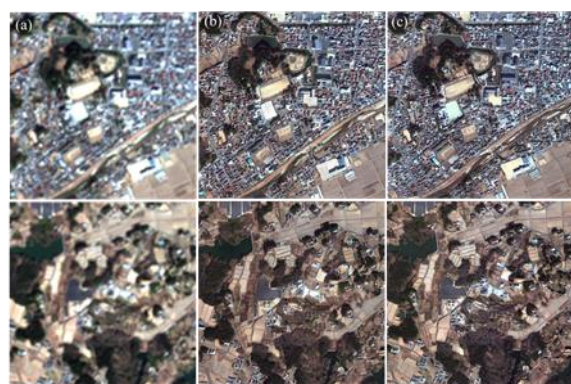
3.1.1 Qualitative Analysis:
The performance of trained model at each experiments was assessed by visually comparing the super resolve image output of the model with HR ground truth images along with LR input image. All images used in EXP1 and HR images for EXP2 were created through a stepwise down-sampling of original WV3 images with bi-cubic interpolation method by a factor of 2.

Figure 8 shows validation results of results of EXP1. It is clearly evident from the visual inspection that results shown in Figure 8(b), separation of original image and generated image is quite unlikely at first sight. From the close inspection it's clear the objects of the inference output is sharper and the original texture is preserved to greater extent. Moreover, model's capability to a recreate the finer image details from road signs to vehicles is comprehensive.

Figure 9 further emphasized the model's potential to super resolve images by generating sharper object details and realistic texture information with a negligible difference with its ground truth even from a coarse resolution images.

Figure 8. Validation results of EXP1 (a) WV3 LR image (1.2 m) (b) Inference output (0.3m) (c) WV3 Ground truth image (0.3m).

Figure 9. Validation results of EXP2 (a). S2 LR image (10.0 m) (b). Inference output (2.5m) (c). WV3 Ground truth image (2.5m).

3.1.2 Quantitative Analysis: In the quantitative analysis, a discussion on model performance evaluated with respect to corresponding ground truth sample based on the six widely used image evaluation scales discussed in the above sections is included. Table 3 summarizes quality matrix results for all experiments. Quality matrix results in Table 3 provides a generalized summary of all image samples used at validation phase. UIQI and CC obtained a relatively higher values for EXP1 and EXP2 where as other indices deviated much from the expected value as per the apparent quality at visual inspection. Index values tend to get lower values for the image tiles with low NDVI and NDSI values. Thus, a much deeper analysis along the spectrum such as histogram comparison is required.

Index		EXP1	EXP2
PSNR	mean	29.6489	25.7144
	std	2.5896	1.9144
SSIM	mean	0.5362	0.4799
	std	0.1425	0.1023
UIQI	mean	0.9621	0.9207
	std	0.0698	0.0769
SAM	mean	0.2897	0.2954
	std	0.0782	0.0961
SID	mean	0.1986	0.2075
	std	0.0986	0.0956
CC	mean	0.9862	0.9687
	std	0.0765	0.0985

**Table 3.** Quality matrix results of three experiments at validation phase.

3.2 Test Results

EXP1 of this study investigated the potentials of ESRGAN models for super resolving downgraded WV3 images to its original resolution. And in EXP2 resolving S2 images by 4x with WV3 spectral characteristics was examined. Ground resolution of satellite image used for the testing phase in EXP1 was 0.5 m. Therefore, image was down sampled into 2m for the test data set creation. Further, a performance compassion to super resolve S2 images with models trained at EXP1 and EXP2 also included to emphasize the need of a model trained with original resolution of S2 images for a better performance in super resolving coarser images. Histogram analysis is also included along with the scatter plots in the test phase for a thorough analysis on spectral quality of the generated image with corresponding HR image.

3.2.1 Qualitative Analysis: Model performances at test phase were assessed by visually comparing the inference results with their respective ground truths. Test results of EXP1 (Figure 10) and EXP2 revealed that both models were successful in resolving 4x images from down sampled WV3 and S2 images at small scale comparisons. However, at large scale comparison in Figure 11 demonstrated that the model has confused in resolving snowy image patches as a consequence of their absence during the training phase. Except that, the model performance at test phase was relatively well regardless of the fact that model was originally trained for super resolve 1.2 m images and images taken at ideal conditions than the image used at test phase. Test has been carried out at worst case scenario as authors' intention was to thoroughly investigate the potentials of ESRGAN model for super resolving satellite data taken at non-ideal situations such as emergency image captures with higher tilt angels during a disaster. A point to noteworthy in EXP2 is that even though the model was trained with extremely less amount of training samples, ESRGAN model has learnt extensively to achieve results which are unattainable with conventional resampling methods (Figure 12).
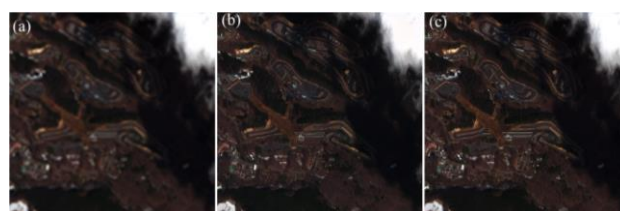
Figure 10. Test results of EXP1 (a) LR image (2.0 m) (b) Output (0.5m) (c) Ground truth image (0.5m).
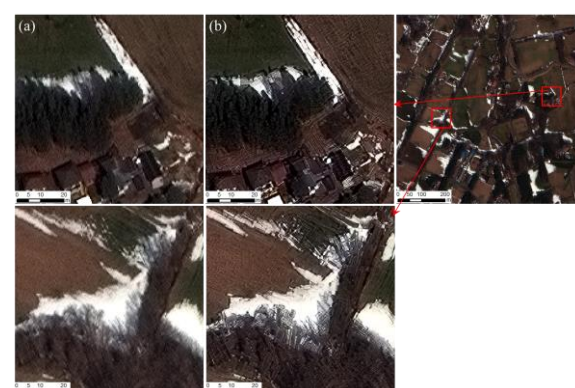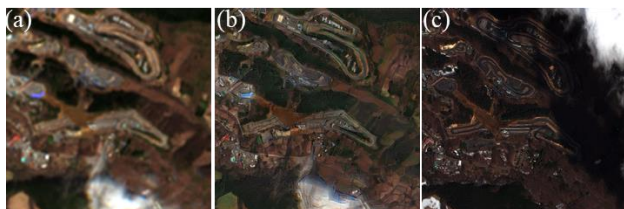

Figure 11. Recreation of snowy image patches of EXP1 (a) ground truth image (0.5 m) (b) Output (0.5m).

Figure 12. Comparison of interpolated (bicubic interpolation) and model output (a) interpolated image (0.5m) (b) EXP1 model output (0.5m) (c) interpolated image (2.5m) (b) EXP2 model output (2.5m).

Figure 13. Test results of EXP2 for Nihonmatsu image (a) LR S2 image (10.0 m) (b) Output (2.5m) (c) Ground truth image (2.5m).

Test results of EXP2 with S2 Nihonmatsu image emphasize the negative impact of relatively longer time lag between LR images with the HR image. Moreover, seasonal bias of the training data set has also affected to the performance of ESRGAN model in super resolving image sample captured during early spring. Further, the absence of snow and cloud covered image samples during training phase and the mountainous landscapes in the test samples might have contributed adversely to the model performance. Aforementioned facts are further confirmed by the comprehensive model performance with Ichihara S2 image tiles (Figure 14) where the image is taken at ideal conditions at which the model is well trained for.

Figure 14. Test results of EXP2 for Ichihara image (a) LR S2 image (10.0 m) (b) Output (2.5m) (c) Google Earth image (18/12/2020).

Figure 15. Comparison EXP1&2 test results for S2 (a) GR 10m input (b) EXP1 model output (c) EXP2 model output and (d) Google Earth HR image taken at 2015/10/8.

Relatively high naturalistic appearance of Figure 15(c) than Figure 15(b) emphasize the necessity of a model trained at original resolution of S2 for coarser resolution images regardless of the fact that the EXP1 model was trained with about 6-7 times higher number of samples than EXP2. However, observed relatively sharper object boundaries at Figure 15(b) than the features on Figure 15(c) highlighted the requirement of

qualitatively and quantitatively improved training data set addressing the issues such as seasonal bias, time lag etc. for better performance in EXP2. In general, test results of EXP2 with S2 data revealed that there is a great potential to generate high resolution naturalistic images from coarser images with tested SISR technology. Yet, the suitability of the generated images for remote sensing applications need to be further investigated

3.2.2 Quantitative Analysis: In this section, comprehensive analysis with specific emphasis on spectral quality of the generated image was carried out. Histogram comparisons of original and generated images for selected image patches also included for a thorough analysis. Quantitative analysis of EXP2 only carried out with WV3 image of Nihonmatsu. In all other cases quantitative values are given in terms of mean and standard deviation of the indexes mentioned in the text.

Index		EXP1	EXP2
PSNR	mean	26.5317	19.5152
	std	2.1362	3.0005
SSIM	mean	0.3964	0.2998
	std	0.2145	0.1935
UIQI	mean	0.9489	0.5622
	std	0.0978	0.1036
SAM	mean	0.2941	0.3144
	std	0.0896	0.1214
SID	mean	0.2013	0.3998
	std	0.1003	0.1461
CC	mean	0.9768	0.5153
	std	0.0980	0.1572

**Table 4.** Quality matrix results of EXP1 and EXP2 at test phase.

From Table 4 results it is obvious that the model performance at test phase is quite alleviated from its validation phase. Such performance much likely for the EXP2 than EXP1 due to larger time lag between tested LR S2 images with its HR counterpart. Authors assume such lower index values as a consequence of the snow, cloud conditions and insufficient model training for the land use type found in the Nihonmatsu image. However, higher SAM and SID values even at the validation phase for EXP1 in comparison to similar works carried out by Romero et al., (2020) might greatly affected by the absence of NIR band in this study.

Relatively weaker performance of trained model in EXP2 both in validation and the test phase probably caused by the insufficient amount of training data. Similar works carried out by Romero et al., (2020) and Galar et al., (2019) have used about 7-8 times larger data sets which led to obtain a comparatively better results. Time difference in HR and LR image samples and presence of cloud and snow conditions at Nihonmatsu image have further worsen the test phase results of EXP2.

One of the main concerns of GANs applications in super resolving remote sensing images is spectral quality of the generated images. Quantitative analysis based on pixel-wise comparisons are always supportive on this regard. Therefore, histogram and scatter plot comparisons have also been incorporated in quantitative analysis to provide a broader overview of the spectral quality of the generated images. The image tiles were selected based on closeness of their quality matrix values to respective mean values summarized in Table 3. Histograms in Figure 16 corresponds to image tile used in Figure 10. Figure 16 shows that original histograms are not altered drastically even though the model was trained selectively for urban landscapes using approximately one third of the total data

set. Observed discrepancies in lower part of histograms and greater coincidence of the upper part of the histogram agreed with exclusion of larger portion of vegetative samples (Song et al., 2015) and bias towards urban landscapes (Sun et al., 2017) during training phase.

In contrary, through the slight lag between histograms of ground truth and generated output in Figure 17, explicitly demonstrated aforementioned performance deficiencies at test phase of EXP2. Further, it implicated the necessity of a qualitatively and quantitatively improved dataset during training process of the respective model as the central wavelength of the WV3 and S2 has a considerable deviation for all 03 bands.
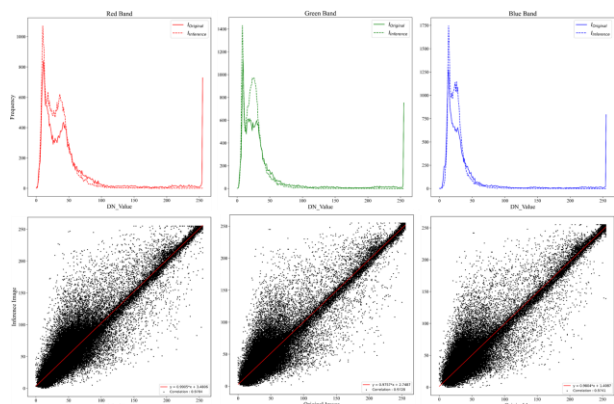
Figure 16. Comparison of histograms (top) and scatterplots (bottom) between ground truth and model generated image of a selected test data tile from EXP1.
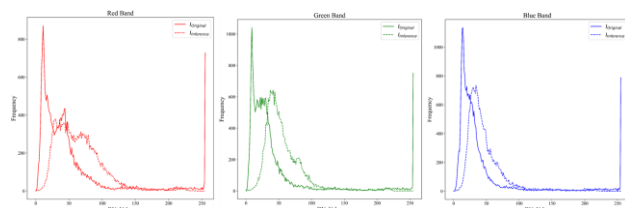
Figure 17. Comparison of histograms between ground truth (up sampled S2) and model generated image of a selected test data tile from EXP2 (correspond to image tile used in Figure 13 (b) and (c)).

Overall, authors would like to mention that the model training carried out in this study was not at ideal conditions with stricter filtering process of data as per the objective of this initial communication was to investigate the extent of ESRGAN model's potential to super resolve images for remote sensing applications. Finally, qualitative and quantitative analysis results conveyed the strong potential of GANs for super resolving images for remote sensing applications with preserving original spectral quality to an adequate extent by a comprehensive amount of training.

4. CONCUSIONS AND FUTURE WORK

This study was investigated and evaluated the potentials of ESRGAN model's application for remote sensing images in two different levels of resolutions. Two individual experiments were designed for super resolving down-sampled WV3 images and S2 images by four folds. Two models were trained at none-ideal conditions by excluding stricter data filtering rules due to the fact that the overall objective of the study was to develop a

generalized method to super resolve satellite images taken at any conditions. EXP1 model was trained only for an urban landscapes. Thus, the image tiles with majority vegetation and soil were removed in EXP1 based on 0.5 threshold values for NDVI and NDSI.

Overall, trained model performed relatively well for the tested urban landscapes even at non-ideal conditions regardless of the fact that the data set preparation did not adhere to the general practise used by the other super resolution studies carried out on remote sensing context. Moreover, from the EXP2 learning capacity of the tested ESRGAN model was greatly demonstrated which implicated the strong potential in super resolving capability of S2 images preserving the original spectral quality by a training process with a qualitatively and quantitatively enhanced data set.

However, as mentioned in the text this study is an initial communication of a series of studies carried out to develop a generalized technology to super resolve remote sensing images to further extend their applicability. During this study authors have found that generating realistic images while preserving original spectral quality is a challenging even at the validation phase. Further, authors have identified the necessity of spectral profiling to be incorporated along with widely used quantification methods like indexing to provide a broader overview of spectral quality of the generated images for remote sensing images.

Further, authors are expecting Spectral Correlation Mapper (SCM) method which is a derivative of Pearsonian Correlation Coefficient that eliminates the assumption of positive and the negative correlations have an equal value of SAM while maintains the SAM characteristic varies from −1 to 1 has greater potential to provide a comprehensive evaluation of generated images for their suitability for remote sensing applications. Consequently, authors are expected to extend the robustness of the model through improvements in data sets along with refinements in model itself for multichannel usage and incorporating a comprehensive image quality analyser for improved performance in ESRGAN model as SISR technology to generate images for remote sensing applications.

REFERENCES

Boardman, 1992, SIPS User's Guide Spectral Image Processing System, Version 1.2, Center for the Study of Earth from Space, Boulder, CO. 88 pp.
Cetin, M., Musaoglu, N., 2009. Merging hyperspectral and panchromatic image data: qualitative and quantitative analysis. International Journal of Remote Sensing, 30(7), pp.1779-1804.

Chang, C., 1999. Spectral information divergence for hyperspectral image analysis, Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1, 509-511.

De Carvalho, O.A., Meneses, P.R., 2000. Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM). In Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18 (Vol. 9). Pasadena, CA, USA: JPL publication.

Dong, C., Loy, C.C., He, K., Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In European

Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany. pp. 184–199.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment.

Galar, M., Sesma, R., Ayla, C., Aranda, C., 2019. Super Resolution for Sentinel-2 Images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* XLII-2/W16, 95-102.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press. http://www.deeplearningbook.org.

Kruse, F.A., Lefkoff, A. B., Boardman, J. W., Heiedbrecht, K. B., Shapiro, A. T., Barloon, P. J., Goetz, A. F. H., 1993, the Spectral Image Processing System (SIPS) – Interactive Visualization and Analysis of Imaging Spectrometer Data. REMOTE SENS. ENVIRON. 44:145-163.

Kim, J., Kwon L. J., Mu L. K., 2016. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA. pp. 1646–1654.

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E. and Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. ISPRS Journal of Photogrammetry and Remote Sensing, 146, pp.305-319.

Lim, B., Son, S., Kim, H., Nah, S., Lee, K. M., 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017-July, 1132–1140.

Ma, W., Pan, Z., Yuan, F., Lei, B., 2019. Super Resolution of Remote Sensing Images via a Dense Residual Generative Adversarial Network. *Remote Sensing.11, 2578; doi: 10.3390/rs11212578.*

Mehmood, A., 2019, May. Deep learning based super resolution of aerial and satellite imagery. In Automatic Target Recognition XXIX (Vol. 10988, p. 109880V). International Society for Optics and Photonics.

Pan, Z., Yu, J., Huang, H., Hu, S., Zhang, A., Ma, H., Sun, W., 2013. Super-resolution based on compressive sensing and structural self-similarity for remote sensing images. IEEE Trans. Geosci. Remote Sens. 51, 4864–4876.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A., 2017. Automatic differentiation in pytorch.

Salgueiro, L., Marcello, J., Vilaplana, V., 2020. Super-resolution of sentinel-2 imagery using generative adversarial networks. Remote Sensing, 12(15), p.2424.

Song, W., Mu, X., Yan, G., Huang, S., 2015. Extracting the green fractional vegetation cover from digital images using a shadow-resistant algorithm (SHAR-LABFVC). Remote Sensing, 7(8), pp.10425-10443.

Sun, Z., Wang, C., Guo, H., Shang, R., 2017. A modified normalized difference impervious surface index (MNDISI) for automatic urban mapping from Landsat imagery. Remote Sensing, 9(9), p.942.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2016. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change L., 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).

Wang, Z. and Bovik, A.C., 2002. A universal image quality index. IEEE signal processing letters, 9(3), pp.81-84.

Xintao, W., Ke, Y., Kelvin C.K., Chan, C., Chen, L., 2018. BasicSR: Open Source Image and Video Restoration Toolbox. https://github.com/xinntao/BasicSR.

Yang, J., Wright, J., Huang, T., Ma, Y., 2008. Image super-resolution as sparse representation of raw image patches. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June. pp. 1–8.

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., 2018. Deep Learning for Single Image Super-Resolution: A Brief Review. arXiv, 1–17.

Yuhas, R.H., Goetz, A.F., Boardman, J.W., 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In Proceedings of the Summaries of the Third Annual JPL Airborne Geoscience Workshop, Pasadena, CA, USA. pp. 147–149.

Zhang, Y., Du, Y., Ling, F., Fang, S., Li, X., 2014. Example-based super-resolution land cover mapping using support vector regression. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7, 1271–1283.

Zhou, W., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612.