

ALERT: Adversarial Learning with Expert Regularization using Tikhonov Operator for Missing Band Reconstruction

Litu Rout[†]

Abstract—The earth observation using remote sensing is one of the most important technologies to assimilate key attributes about the earth’s surface. To achieve tangible consequence, the internal building blocks of such a complex system must operate flawlessly. However, due to dynamically changing environment, degradation in sensor electronics, and extreme weather condition remotely sensed images often miss essential information. As the sensors operate over several years in space the likelihood of sensor degradation persists. This results in commonly observed issues, such as stripe noise, missing partial data, and missing band. Various ground based solutions have been developed to address these technological bottlenecks individually. In this manuscript, we devise a method, which we call ALERT, to tackle missing band reconstruction. The proposed method reconstructs missing band with the sole supervision of spectral and spatial priors. We compare the proposed framework with state-of-the-art methods and show compelling improvement both qualitatively and quantitatively. We provide both theoretical and empirical evidence of better performance by regularized adversarial learning as compared to complete supervision. Further, we propose a new Residual-Dense-Block (RDB) module to preserve geometric fidelity and assist in efficient gradient flow. We show that ALERT captures essential features such that the spatial and spectral characteristics of reconstructed band remains preserved. To critically analyze the generalization we test the performance on two different satellite datasets: Resourcesat-2A and WorldView-2. As per our extensive experimentation, the proposed method achieves 20.72%, 13.81%, 1.05%, 15.91%, and 2.94% improvement in RMSE, SAM, SSIM, PSNR, and SRE respectively over state-of-the-art model.

Index Terms—Remote Sensing, Adversarial Learning, Missing Band Reconstruction, Expert Regularization, Tikhonov Operator.

I. INTRODUCTION

REMOTE Sensing is a complex composition of various subsystems which are intended to operate coherently for tangible outcome. In the process of launching a satellite to acquiring remotely sensed images, lot of things could occur in an adverse manner. The dynamics of atmosphere and degradation in the internal building blocks of these complex subsystems are often mirrored by poor image quality. Some of the commonly observed issues with such remote sensing images are stripe noise, dead detectors, missing partial data, haze, and missing band. These unavoidable issues on board encourage researchers to devise a ground based solution so as

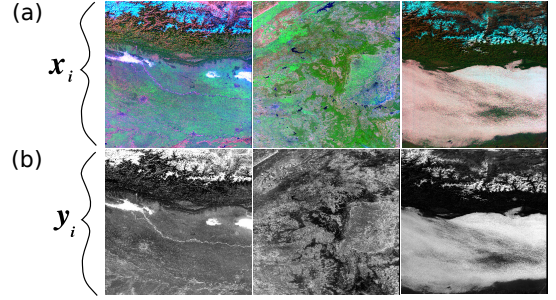


Fig. 1. Samples of paired concurrent data. (a) FCC of SWIR (R), NIR (G), and G (B). (b) Corresponding R band.

to improve the data usability. As per recent studies [1], [2], [3], [4], [5], there is a serious dependency on the ground based solutions to improve the usability of remotely sensed imagery.

In this manuscript, we focus on the missing band issue that may arise due to several reasons including degradation in sensor electronics of that particular band. Since Red (R) band possess a vital contribution in various remote sensing applications, we primarily focus on reconstructing this band with the sole supervision of existing concurrent multi-spectral bands: Short-Wave InfraRed (SWIR), Near InfraRed (NIR), and Green (G). Nevertheless, the proposed approach can also be extended to reconstruct other bands without loss of generality. In Fig. 1, we show paired concurrent data used for reconstructing missing band information. The developed model is intended to operate on the False Colour Composite (FCC) image, as shown in Fig. 1 and reconstruct the corresponding Red band by preserving both the spatial and spectral characteristics.

Over the years many data reconstruction techniques have been proposed taking into account several key aspects of remote sensing imagery. The notion of inserting additional information into an image based on cues derived from spatial, spectral, or temporal domain is a challenging problem in the field of computer vision. The spatial domain methods rely on the assumption that sufficient statistics about missing information is captured in the undamaged regions. Though these methods [6], [7], [8] address this issue to some extent, the reconstruction of large missing areas with required precision still remains challenging. Also, the assumption of similar statistics is not always satisfied in practical real world applications. Therefore, to articulate missing data reconstruction, recent studies [9], [10] propose to augment spectral information with

[†]Optical Data Processing Division, Signal and Image Processing Group, Space Applications Centre, ISRO, Ahmedabad, India - 380015. mail id: liturout1997@gmail.com, Final peer-reviewed version available at DOI: 10.1109/TGRS.2019.2963818

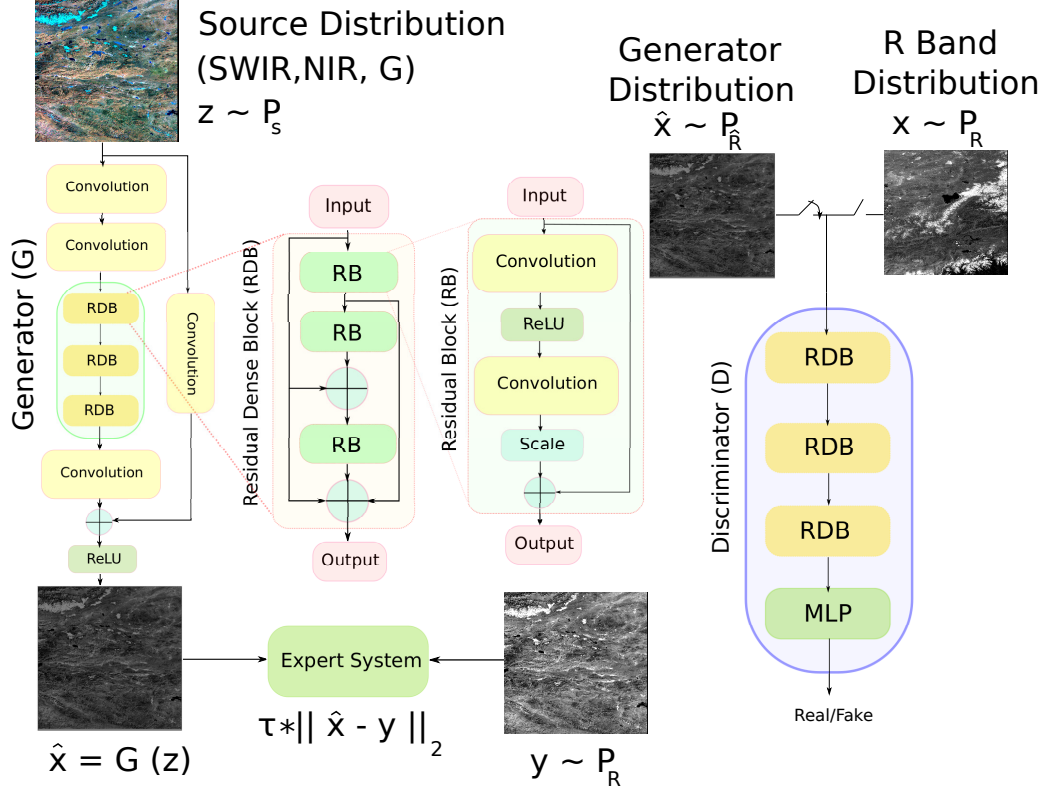


Fig. 2. Pipeline of the proposed framework. The generator learns to generate samples of red band distribution given corresponding samples of the source distribution. The expert regularization and discriminator assists the generator in the process of learning target distribution.

conditional spatial data along with temporal information which effectively captures rich texture and features of missing targets. But the registration artifacts and difficulty in getting strictly chronological data over same region reduces the efficiency of such methods [11]. Beyond that, a unified framework to capture the essence of spatial, spectral, and temporal domain is proved to be effective in addressing the issues of partial data reconstruction [2]. However, this method is also limited by the absence of well registered and strictly chronological data. In addition, such methods extract substantial cues from the available partial data in the same spatial domain, which is not available in case of complete loss of a certain band. This therefore increases the complexity of reconstructing a missing band entirely from priors. Our contributions can be summarized as following.

- One of the primary objectives of this study is to analyze the efficacy of direct supervision and supervision with augmented adversarial objective with respect to missing band reconstruction of satellite imagery.
- In the process, we develop an end to end framework, which we call ALERT and provide theoretical evidence on why the proposed augmented objective is considerably better than sole supervision.
- Further, we propose a new building block, namely Residual Dense Block (RDB) to enforce geometric fidelity while allowing better gradient flow.
- Also, we study the generalization of proposed method to a different satellite, WorldView-2 and isolated physical

location, Washington.

In the following section we briefly discuss about recent works attempting to address similar problems as discussed in this study.

II. RELATED WORK

Recently, a deep learning model: DeepSWIR proposed an alternative to synthesize a high resolution SWIR band by learning a dense fusion of features extracted through global and local residual feature extraction units [5]. This method exploits the spatial characteristics from existing high resolution bands (G, R, NIR) at 5 m and spectral characteristics from a coarse resolution SWIR band at 24 m to reconstruct the high resolution SWIR at 5 m. However, this method has one limitation as it relies on a coarser resolution version of the missing band to learn spectral characteristics. Thus, we can not directly apply this approach for missing band reconstruction. We therefore modify the architecture of DeepSWIR to address the missing band reconstruction of Advanced Wide Field Sensor (AWiFS), which we call DeepAWiFS and use it as our baseline for comparison. As per Fig. 1, the DeepAWiFS model takes FCC, x_i as input and learns to map it to the corresponding red band, y_i in a simulated environment using purely supervised learning. The internal building blocks of DeepAWiFS are exactly same as DeepSWIR other than the paired training dataset. The pretrained DeepAWiFS is applied on real missing band reconstruction problem. In a loose sense, our proposed adversarial approach: ALERT is built

upon the idea of reconstructing missing band entirely from priors similar to DeepAWiFS. The generator part of ALERT and DeepAWiFS share similar architectures apart from the proposed residual dense block which will be discussed in the subsequent sections.

Earlier, it was believed that reconstruction of a spectral band from other spectral bands is not feasible with sufficient degree of realism as the spectral bands have uncorrelated distributions. However, recent advancements in computer vision have shown promising results to counter this earlier belief. The ability of deep neural networks to learn unique representations of various target attributes drives the success of band reconstruction from uncorrelated spectral bands. Despite the fact that supervised learning driven deep neural networks have achieved remarkable success in many challenging vision problems, the designing of a suitable objective function still remains extremely difficult. In remote sensing, the hand-crafted objective functions, such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) etc. introduce unacceptable artifacts that pushes the reconstructed images away from realistic manifold of actual data.

To address this issue, Goodfellow et. al. [12] introduced an innovative learning framework, namely Generative Adversarial Nets (GANs) which is inspired by game theory. Instead of using any fixed hand-crafted objective function, it uses two neural networks: generator and discriminator to play against each other and converge to a minimally divergent solution. Unlike in supervised learning, where a target network receives gradient from a fixed objective throughout training, in adversarial learning the target network, namely generator receives gradient from a continually evolving objective, namely discriminator. As the training progresses, the discriminator becomes stronger and stronger which allows the generator to receive better gradients and become superior to its previous self. This, upon reaching at Nash equilibrium, enables the generator to push the data generating distribution close to the natural manifold of realistic data. The GANs provide two fold advantages. First, it can learn distributions confined to a low dimensional manifold, such as distribution of a missing spectral band. Second, it can directly generate samples from the intended target distribution, which is the missing band under current study.

Based on this adversarial learning, A. Rangnekar et.al. train conditional GANs to map coarser resolution RGB images into coarser resolution hyper-spectral images [13]. The trained model is then used to predict high resolution hyper-spectral images given high resolution RGB images. Since the objective here is to make the spectral resolution finer by resolving existing overlapping wider bandwidth signals into finer ranges, the task becomes relatively straightforward. In this case the RGB images already contain essential broader spectral signatures. However, the problem of missing band reconstruction does not have such conditional information covering entire wavelength spectrum, which makes it particularly different from super-resolution. Also, this method uses Jensen-Shannon Divergence (JSD) as primary objective function which has empirical evidence of mode-collapse, a commonly observed issue with such objectives. On the other hand, the proposed

ALERT method with Wasserstein objective is reported very often by independent researchers to have alleviated this adverse effect. Further, the authors in [14], [15], [16] propose various methods of extracting hyper-spectral bands from undersampled acquisitions in the spectral domain. It is worth mentioning that though both the problems are quite challenging in their own way, the missing band reconstruction is particularly less straightforward as it does not have access to as many cues as super-resolution.

Motivated by these findings, we build upon this adversarial framework and show that adversarial objectives can perform better than hand-crafted supervised objectives even in remote sensing imagery. Apart from that, we identify some key limitations of adversarial learning and propose solutions to address those issues with sufficient experimentation and theoretical backup. It is to be noted that we devise a method to address missing band reconstruction from distribution perspective using a guided adversarial framework, which we call ALERT that only requires concurrent multi-spectral data. To put more succinctly, the model receives gradients from an expert system and an adversary which penalizes the generator configuration that otherwise leads to divergence from real manifold. The expert regularization, as proposed in this paper, serves as an imitation learning objective and allows faster convergence of adversarial learning. In addition, we propose architectural innovations to improve quality, stability, and robustness of ALERT in reconstructing missing band from deep priors. The overall pipeline of ALERT is shown in Fig. 2.

The rest of the paper is organized as follows. In Section III, we briefly discuss the prerequisites and elaborate on the proposed methodology. In Section IV, we give details on the study area and implementation setup. In Section V, we analyze the results in both simulated and real experiments. Finally, we summarize our discussion in Section VI.

III. METHODOLOGY

GANs framework is widely used in various computer vision tasks. The high quality natural looking images generated by adversarial learning gives this method an edge over its counterparts. As per recent studies, though GANs are capable of producing high quality images, it sometimes excessively constructs fine granular structures which do not exist in practice. This is due to the fact that discriminator recognizes difficult-to-learn latent patterns that forces the generator to make the images perceptually realistic. Thus, the images generated by adversarial loss though achieve lesser PSNR than pixel loss, such as RMSE and MAE these images have better subjective image quality [17]. In remote sensing applications, however, synthesis of such artificial features, though appear to be realistic, is unacceptable as it reduces scientific fidelity. In addition, to generate high fidelity realistic images as in [18], it requires considerably large training time on high-end graphics cards. We believe this requirement of large training time is probably due to lack of supervision in GANs framework. To our knowledge, the training process of GANs is still an active area of research where stability, quality, and variation are of prime importance.

We, therefore, propose to address these issues by leveraging the efficacy of adversarial learning while guiding the gradient updates of generator via expert regularization. The generator receives gradients from discriminator as well as from an expert system which has access to the corresponding sample in the target distribution. As the proposed expert regularization is proportional to L2 distance between generated sample and original sample, it does not affect the sample quality when it lies within a confined manifold of realistic samples. In this case, the adversarial loss helps the generator improve perceptual quality of generated samples. Note that the earlier gradient updates receive significant contribution from the expert system which triggers faster convergence. Further, the expert regularization forces the generated sample to lie within a close proximity of realistic satellite imagery unlike adversarial loss. This impedes unwanted feature synthesis and improves the scientific accuracy needed for remote sensing applications. In this process, we also introduce a stable architecture specifically designed to cooperate with adversarial learning and thereby, improve the quality and variation of generated samples. In the following section, we make this hypothesis more concrete by providing mathematical backup.

A. Theoretical Framework

Let $z_1 \sim \mathbb{P}_{SWIR}$, $z_2 \sim \mathbb{P}_{NIR}$, $z_3 \sim \mathbb{P}_G$, and $x \sim \mathbb{P}_R$, where $\mathbb{P}_{SWIR}$, $\mathbb{P}_{NIR}$, $\mathbb{P}_G$, and $\mathbb{P}_R$ represent the distribution of SWIR, NIR, G, and R, respectively. Let $z \sim \mathbb{P}_S(z_1, z_2, z_3)$, where $z_i \in \mathbb{R}^{M \times N}$, $i = 1, 2, 3$, and $\mathbb{P}_S$ represents joint source distribution. Let $\hat{x} \sim \mathbb{P}_{\hat{R}}$, where $\hat{x} = G(z)$ and $\mathbb{P}_{\hat{R}}$ represents generator distribution. The corresponding sample of z in the target distribution is denoted by $y \sim \mathbb{P}_R$. The problem of missing band reconstruction is formulated such that the generator learns to sample from the distribution of missing band, $\mathbb{P}_R$ given the corresponding sample from existing concurrent joint distribution, $\mathbb{P}_S$.

In conventional GANs framework, the minimax value function between G and D can be optimized by,

$$\min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_R} [\log(D(x))] + \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{R}}} [\log(1 - D(\hat{x}))]. \quad (1)$$

To address the issues of mode collapse, stability, and performance in conventional GANs, Arjovsky et. al. introduced Wasserstein GANs (WGANs) [19]. In case of missing band reconstruction, the Earth Mover (EM) distance as minimized by WGANs is given by,

$$\min_G \max_{D \in \mathcal{D}} \mathbb{E}_{x \sim \mathbb{P}_R} [D(x)] - \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{R}}} [D(\hat{x})], \quad (2)$$

where $\mathcal{D}$ is the set of 1-Lipschitz functions of D . To enforce 1-Lipschitz on D , WGANs used weight clipping such that the weights of D lies within a compact space.

The value function that the generator tries to optimize is reported to be discontinuous with respect to its parameters. This in fact is one of the major reasons for undesired behaviour of virtually all generative models reported in classical adversarial learning literature. On the contrary, the EM distance based value function along with weight clipping, as described in [19], assures continuity everywhere and differentiability

almost everywhere. However, as reported in [20], the WGANs with weight clipping fail to capture higher order moments of data generating distributions. Since, weight clipping enforces the discriminator weights to move towards extremas of clipping values, it is limited by a small set of critic functions which may not be sufficient to capture difficult-to-learn latent patterns. Gulrajani et. al. [20] proposed Gradient Penalty (GP) term to address this issue and thereby, increase the set of probable critic functions. Especially, the large ground cover of satellite imagery with wide spectral band is difficult to be represented using lower order moments. We therefore propose to use WGANs+GP [20] so as to broaden the horizon of missing band generating distribution. In the current setting, WGANs+GP is given by,

$$\min_G \max_D \mathbb{E}_{x \sim \mathbb{P}_R} [D(x)] - \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{R}}} [D(\hat{x})] - \lambda \mathbb{E}_{\tilde{x} \sim \mathbb{P}_{\tilde{R}}} \left[(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2 \right], \quad (3)$$

where $\mathbb{P}_{\tilde{R}}$ represents a distribution that samples uniformly along the line connecting points sampled from $\mathbb{P}_R$ and $\mathbb{P}_{\hat{R}}$. Arjovsky et. al. discuss that enforcing gradient norm to be 1 along this line is a sufficient condition to maintain stability as it is intractable to enforce gradient norm to be 1 everywhere in the parametric space. In this setting, the generator does not receive a feedback signal from an expert system having access to corresponding sample in actual red band distribution ($\mathbb{P}_R$). This results in slower convergence and allows feature synthesis as a part of learning target distribution, which is unacceptable in satellite image processing. Probably, the main reason for not accepting these natural looking artificial images is that the generator may synthesize some additional features, such as building or vegetation, which do not exist on the physical location, just to make the image fall on realistic manifold of natural images. To address this issue, we propose to use expert regularization based on Tikhonov operator which will prevent this additional feature synthesis. This also guides the gradient update step of generator in the desired direction which will effectively increase the rate of convergence. To bring this insight into fruition, we modify the WGANs+GP objective by adding another constraint, which we call expert regularization. The objective of discriminator remains unchanged as given by equation (4).

$$\min_D \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{R}}} [D(\hat{x})] - \mathbb{E}_{x \sim \mathbb{P}_R} [D(x)] + \lambda \mathbb{E}_{\tilde{x} \sim \mathbb{P}_{\tilde{R}}} \left[(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2 \right]. \quad (4)$$

The new objective function of generator becomes,

$$\min_G \mathbb{E}_{x \sim \mathbb{P}_R} [D(x)] - \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{R}}} [D(\hat{x})] + \mathbb{E}_{y \sim \mathbb{P}_R} [\|\tau(\hat{x} - y)\|_2], \quad (5)$$

where $\tau = \alpha I$ denotes Tikhonov operator. Here, α denotes expert's contribution and I represents identity matrix of appropriate dimension. It is important to mention that we are not the only one to discover the idea of expert regularization in guiding gradient updates. LeCun et. al. [21] showed the efficacy of expert regularization in model predictive policy learning. Though it was experimented in a different problem

setting, the notion of expert trajectory was indeed beneficial to develop our theoretical framework.

B. Manifold Mapping Theorem

Definition: Let A be a subset of $\mathbb{R}^n$. We say A has measure zero in $\mathbb{R}^n$ if for every $\epsilon > 0$, there is a covering $Q_1, Q_2, Q_3, \dots$ of A by countably many rectangles such that

$$\sum_{i=1}^{\infty} v(Q_i) < \epsilon.$$

Here, $v(Q)$ represents volume of rectangle Q .

Theorem I: if A be a collection of images sampled from source ($\mathbb{P}_S$) and target ($\mathbb{P}_R$) distribution, then A has measure zero in $\mathbb{R}^{M \times N}$.

Proof: Let $x_i \in A, i = 1, 2, \dots$ and $\tilde{x}_j = \text{Vec}(x_i), j = 1, 2, \dots, MN$, where $\text{Vec}(\cdot)$ denotes vectorization operator of a matrix. For $\delta = \frac{\alpha}{i}$, where $\alpha > 0$, we can choose $Q_i = [\tilde{x}_1 - \delta, \tilde{x}_1 + \delta] \times \dots \times [\tilde{x}_j - \delta, \tilde{x}_j + \delta] \times \dots \times [\tilde{x}_{MN} - \delta, \tilde{x}_{MN} + \delta]$. Now,

$$\sum_{i=1}^{\infty} v(Q_i) = \sum_{i=1}^{\infty} \left(\frac{2\alpha}{i} \right)^{MN} = (2\alpha)^{MN} \sum_{i=1}^{\infty} \left(\frac{1}{i^{MN}} \right)$$

According to the definition, we can finish the proof by finding a suitable δ for every $\epsilon > 0$. So,

$$\begin{aligned} \sum_{i=1}^{\infty} v(Q_i) &< \epsilon \\ \Rightarrow (2\alpha)^{MN} \sum_{i=1}^{\infty} \left(\frac{1}{i^{MN}} \right) &< \epsilon \\ \Rightarrow (2\alpha)^{MN} &< \epsilon \left(\sum_{i=1}^{\infty} \left(\frac{1}{i^{MN}} \right) > 1 \right) \\ \Rightarrow \alpha &< \frac{\epsilon^{\frac{1}{MN}}}{2} \end{aligned}$$

Therefore, for every $\epsilon > 0$, there exists a $\delta = \frac{\epsilon^{\frac{1}{MN}}}{2i}$, such that $\sum_{i=1}^{\infty} v(Q_i) < \epsilon$.

Corollary I: A subset of a set of measure zero has measure zero.

Let $\mathcal{M}$ be a subspace of images sampled from $\mathbb{P}_S$ and $\mathcal{N}$ be a subspace of corresponding paired images sampled from $\mathbb{P}_R$. By **Corollary I**, $\mathcal{M}$ and $\mathcal{N}$ have measure zero in $\mathbb{R}^{M \times N}$. Since the generator (G) with expert regularization carries each element $p \in \mathcal{M}$ to $\mathcal{N}$ in a one-to-one fashion, it is easy to show that G forms a coordinate patch on MN -manifold $\mathcal{M}$ about p . Here, the discriminator is expected to distinguish between the real manifold $\mathcal{M}_R$ and manifold of generated samples $\mathcal{M}_{\hat{R}}$. However, if $\mathcal{M}_R$ and $\mathcal{M}_{\hat{R}}$ are inseparable, then discriminator provides no assistance to generator. We discuss this with more details in the following theorem.

C. Constant Discriminator Theorem

If two manifolds of real and generated bands match perfectly in a metric space, then no discriminator can separate them. Therefore, we can safely assume that any two manifolds under our current study never align perfectly. This

gives rise to the weak assumption that the support of real: $\text{supp } \mathbb{P}_R$ and generated: $\text{supp } \mathbb{P}_{\hat{R}}$ red band are contained in separated compact subsets, which we call $\mathbb{X}$ and $\mathbb{Y}$. Under the aforementioned condition, there exists a perfect discriminator D^* that has accuracy 1 $\forall x \in \mathbb{X} \cup \mathbb{Y}$.

Proof: Given $\mathbb{X} \cap \mathbb{Y} = \emptyset$ and $\bar{\mathbb{X}} \cap \mathbb{Y} = \emptyset$, where $\bar{\mathbb{X}}$ denotes closure of $\mathbb{X}$ and $\emptyset$ represents empty set. Let d be the metric defined in $\mathbb{R}^{M \times N}$ and $\mathbb{X}, \mathbb{Y} \subset \mathbb{R}^{M \times N}$. Let $d(\mathbb{X}, \mathbb{Y}) = \epsilon$. By ϵ -neighbourhood theorem $\mathbb{X} \subset \mathbb{R}^{M \times N}$ implies $N_{\epsilon/3}(\mathbb{X}) \subset \mathbb{R}^{M \times N}$. Analogously, $N_{\epsilon/3}(\mathbb{Y}) \subset \mathbb{R}^{M \times N}$.

ϵ -neighbourhood Theorem: Let $\mathbb{X}$ be a compact subspace of $\mathbb{R}^{M \times N}$; U be an open subset of $\mathbb{R}^{M \times N}$ containing $\mathbb{X}$. Then there exists an $\epsilon > 0$ such that the ϵ -neighbourhood of $\mathbb{X}$, $N_{\epsilon}(\mathbb{X})$ (in both euclidean and sup metric) is contained in U , i.e., $N_{\epsilon}(\mathbb{X}) \subset U$.

There is strong empirical and theoretical evidence to believe that $D^*(x) = 1 \forall x \in N_{\epsilon/3}(\mathbb{X})$ (real) and $D^*(y) = -1 \forall y \in N_{\epsilon/3}(\mathbb{Y})$ (fake) for sufficiently trained $D^*(\cdot)$ [19]. Therefore $\nabla_x D^*(x) = \lim_{\delta x \rightarrow 0} \frac{D^*(x+\delta x) - D^*(x)}{\delta x} = 0$ by choosing $\delta x < \epsilon/3$. Similarly, $\nabla_y D^*(y) = 0$ by choosing $\delta y < \epsilon/3$. This shows that there exist open balls $B(\mathbb{X}, \epsilon/3)$ and $B(\mathbb{Y}, \epsilon/3)$ where $D^*(\cdot)$ remains constant. Thus, we finish the proof of constant discriminator theorem.

Since discriminator remains constant in both manifolds the generator does not learn anything useful by backpropagating through it. This results in slow convergence and poor practical performance. To put more succinctly, randomly initialized parameters of generator may not necessarily generate samples that fall within the support of real distribution. However, the proposed expert regularization will update the generator's parameters in such a way that the two supports are connected in $\mathbb{R}^{M \times N}$, i.e., $\mathbb{X} \cap \mathbb{Y} \neq \emptyset$ and $\bar{\mathbb{X}} \cap \mathbb{Y} \neq \emptyset$. This ensures that the generator receives useful gradients from the discriminator.

D. Supervised and Adversarial Objective

Supervised learning based on fixed objective function, such as RMSE performs reasonably well in many computer vision applications. However, this RMSE metric faces a common issue as suggested by statistical theory. As per this theory, for a given $z \in \mathbb{P}_S$ there exist many $\hat{x}_i \in \mathbb{X}, i = 1, 2, 3, \dots, N$. The predicted sample, $\hat{x}$ therefore is obtained by implicit optimization, $\arg_{\hat{x}} \min \sum_{i=1}^N \|\hat{x}_i - y\|$. Thus, the predicted R band will be expected value of all possible R bands corresponding to given input bands: G, NIR, SWIR. This is mirrored by poor perceptual quality though PSNR remains quite high. On the contrary, the supervised generator with discriminator as an adversary selects one high quality sample from the pool of possible targets [17].

Further, since $G(\cdot)$ is a hierarchical composition of continuous constituent functions mapping from $\mathbb{Z} \rightarrow \mathbb{Y}$, it is continuous in $\mathbb{Z} \rightarrow \mathbb{Y}$. Therefore, $z \in N_{\delta}(z_o)$ implies $d_Y(G_{\theta}(z), y_o) < \epsilon$ for $\delta > 0$ and $\epsilon > 0$. Here, z_o and y_o are the paired source and target bands, respectively. Hence, $\nabla_{\theta} d_Y(G_{\theta}(z), y_o) < \hat{\epsilon}$ for arbitrarily small $\hat{\epsilon}$. This exacerbates the learning progress. On the contrary, the adversarial objective

under this circumstance provides useful gradients that help improve the image quality further by pushing it towards natural manifold of realistic data. The major contribution for non-negligible gradients in adversarial objective comes from $\nabla_{\theta_G} [\mathbb{E}_{x \sim \mathbb{P}_R} [D(x)] - \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{R}}} [D(\hat{x})]]$, which will only be zero under constant discrimination, or upon convergence. This proves that the augmentation of adversarial learning into a supervised framework improves the perceptual quality of generated images and attain relatively minimal empirical risk. Within a tiny landscape of optimal empirical risk, the non-diminishing gradient in ALERT also provides an explanation for faster convergence compared to purely supervised framework.

In [19], the authors claim that one of the most compelling benefits of WGANs is that it does not require a careful design of the network architecture. The authors claim is proven to be effective in many computer vision tasks, however, we observe that it does require a thoughtful design especially for satellite image processing. We discuss these architectural details in the following section.

E. Architecture

Recent advancement in deep neural networks, such as Densenet [22] and ResNet [23] are very much successful in Imagenet Large Scale Visual Recognition Challenge (ILSVRC) [24]. It seems natural to wonder whether these architectural innovations would be beneficial when incorporated in an adversarial framework. For this reason, we design a new architecture by borrowing key ingredients from [22] and [23]. We carefully combine the dense connection of DenseNet and skip connection of ResNet to develop a new Residual Dense Block (RDB) that enables faster convergence and better stability due to efficient gradient flow. We develop a custom Residual Block (RB) that performs reasonably well on satellite image processing. Though batch normalization reduces the covariate shift and improves performance on a wide variety of computer vision tasks, it is proven to achieve sub-optimal results for image reconstruction, such as super resolution [25]. We observed similar degradation in performance while reconstructing missing band from deep priors and hence, we advocate not to use batch normalization in satellite image reconstruction. The generator network consists of 2 convolutional blocks to extract shallow features. The shallow features are then passed through 3 RDBs and finally concatenated with global features extracted by a single convolution layer of kernel size (1x1). This helps the generator preserve spatial characteristics of source distribution while pushing the generated samples towards the realistic manifold of R band. The discriminator is equipped with similar configuration as that of the generator so as to untangle the generated samples properly. In addition to that it has a Multi-Layer Perceptron (MLP) with 3 linear layers to map the discriminated features into required format. The overall configuration and intrinsic architectural details are illustrated in Fig. 2.

TABLE I
TRAINING AND INFERENCE

Metrics	Parameters	Training epochs	Training RMSE	Inference Time
AeroGAN	2 M	100	41.57	64 s
DeepAWiFS	4.7 M	1000	18.2	495 s
DSen2	4.7 M	1000	21.2	456 s
ALERT	2.8 M	50	10.3	70 s
% gain	40.42	95.00	51.41	84.64

TABLE II
QUANTITATIVE ANALYSIS

Metrics	RMSE	SAM (deg)	SSIM	PSNR (dB)	SRE (dB)
AeroGAN	42.77	8.91	0.74	38.08	52.90
DeepAWiFS	27.76	6.63	0.86	31.60	56.6
DSen2	20.01	5.72	0.95	35.01	59.84
ALERT	15.85	4.93	0.96	36.63	61.60
% gain	20.72	13.81	1.05	4.63	2.94

IV. EXPERIMENT

A. Dataset and Study Area

We use multi-spectral satellite imagery of AWiFS on board Resourcesat-2A payload of Indian Space Research Organisation (ISRO). AWiFS operates in VNIR and SWIR at 56 m spatial resolution with a large swath of 730 km. In this paper, we focus on two study areas in India. As per Worldwide Reference System (WRS), we use path & row combinations of 99 & 53 for training and 99 & 55 for testing purpose. We process large satellite imagery by using small crops of size ($M = 64, N = 64$) each. Thus, we create a total of 21375 paired concurrent samples randomly split into training (18168 samples) and validation (3207 samples). We test the performance of the proposed model on physically isolated 21250 test samples. In order to reconstruct the actual full resolution image, we predict adjacent patches with a fixed stride of (16,16) and employ Gaussian feathering for image stitching [5].

B. Implementation Details

Here, we provide the necessary implementation details along with exact hyper-parameter settings for faithfully reproducing and evaluating the reported results. We use a local machine with 8 GB gpu memory (Quadro P4000), 256 GB ram, and i7 processor for conducting all experimentation. We use Adam optimizer with fixed learning rate of 0.0001. We pretrain generator with expert regularization alone for 2 iterations before updating discriminator 5 times and generator once, as given by equation (4) and equation (5), respectively. Note that the pretraining of generator, updates of discriminator, and generator take place sequentially in every epoch. We set $\lambda = 10$ and $\alpha = 100$. We use fixed convolution kernels of size (3x3) throughout except for global feature extraction where we use (1x1) kernels. The entire framework has been developed using Pytorch.

V. RESULTS AND DISCUSSION

In this study, we evaluate the model's performance both qualitatively and quantitatively. We analyze the training and

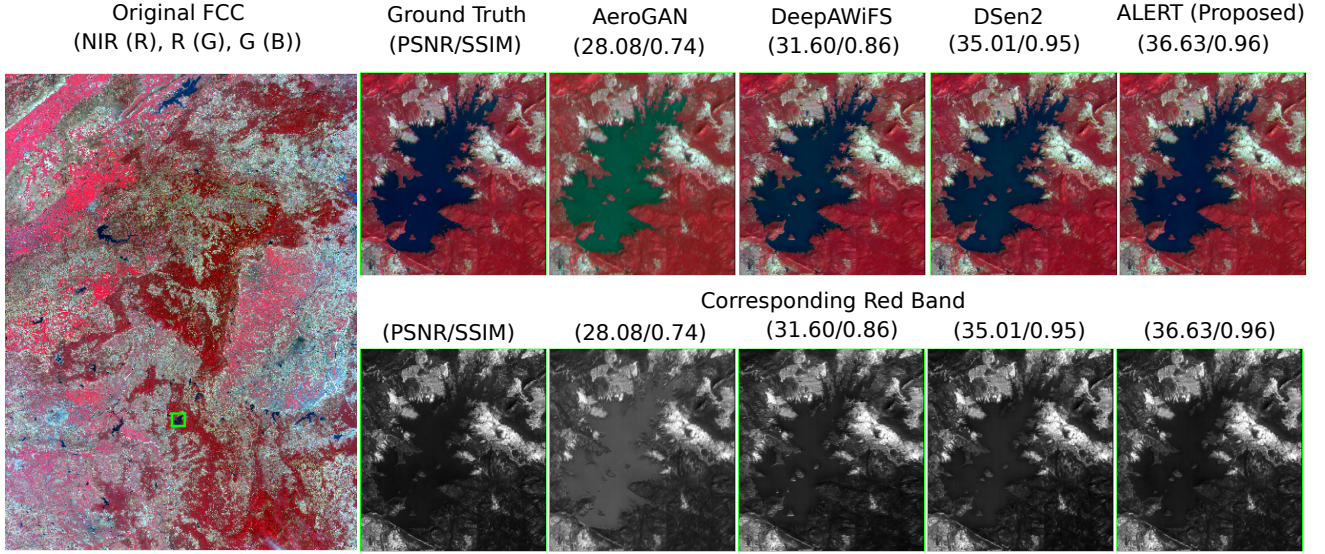


Fig. 3. Qualitative analysis of reconstructed band in simulation. ALERT performs favourably against supervised learning. The region within bounding box is enlarged for the sole purpose of better visualization.

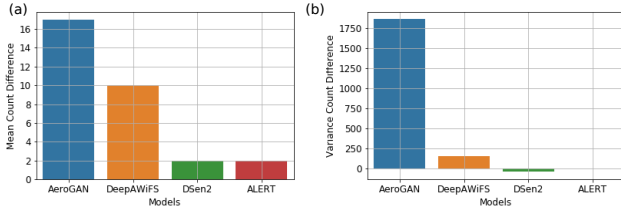


Fig. 4. (a) First and (b) Second order moment difference between original and reconstructed band.

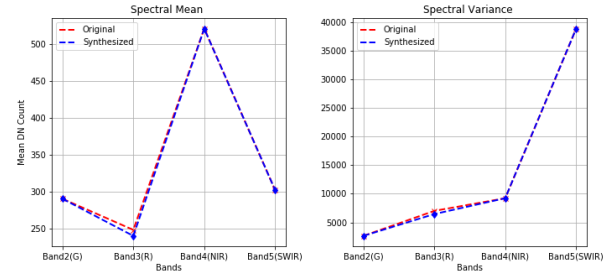


Fig. 6. First and Second order moments. The reconstructed band shows spectral consistency with existing bands.

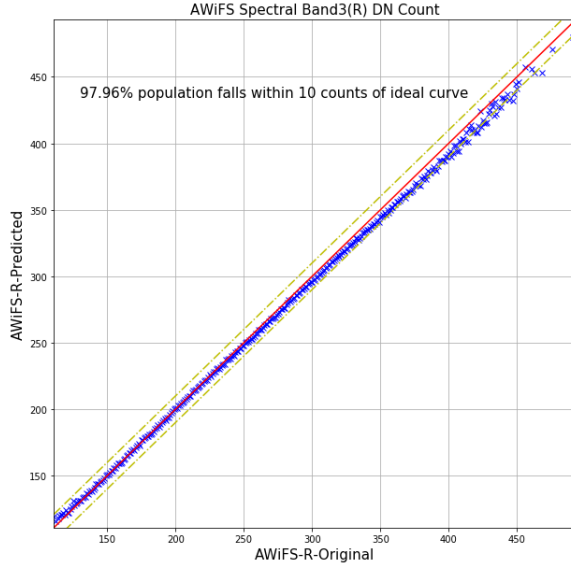


Fig. 5. Per-pixel DN count cross correlation of ALERT. The DN values of reconstructed band closely matches the original.

inference of state-of-the-art methods and proposed ALERT. As training can be performed beforehand, we focus particularly on inference time. As given in Table I, top performing method DSen2 [26] with complete supervision achieves sub-optimal

results compared to adversarial learning based ALERT. It is worth noticing that ALERT offers significant gain in trainable parameters (40.42%), training epochs (95.00%), RMSE (51.41%), and more importantly¹ inference time (84.64%). The percentage gain is computed relative to the best performing model DSen2 in every category.

Further, we use state-of-the-art image reconstruction assessment metrics to critically analyze the performance quantitatively. As given in Table II, ALERT achieves considerable improvement over its counterparts. Of particular interest, it achieves 20.72%, 13.81%, 1.05%, 15.91%, and 2.94% improvement in RMSE, SAM, SSIM, PSNR, and SRE respectively over state-of-the-art model, DSen2. It is worth mentioning that the adversarial learning with JSD as proposed in AeroGAN [13] fails to capture higher order moments and thereby, produces sub-optimal results. However, the adversarial learning with Wasserstein distance and expert regularization produces considerably better results. To qualitatively assess the performance, we compare ground truth with these approaches in Fig. 3 in a simulated environment where we have access to corresponding red band. The supervised learning based

¹important for near real time generation of bundled data product

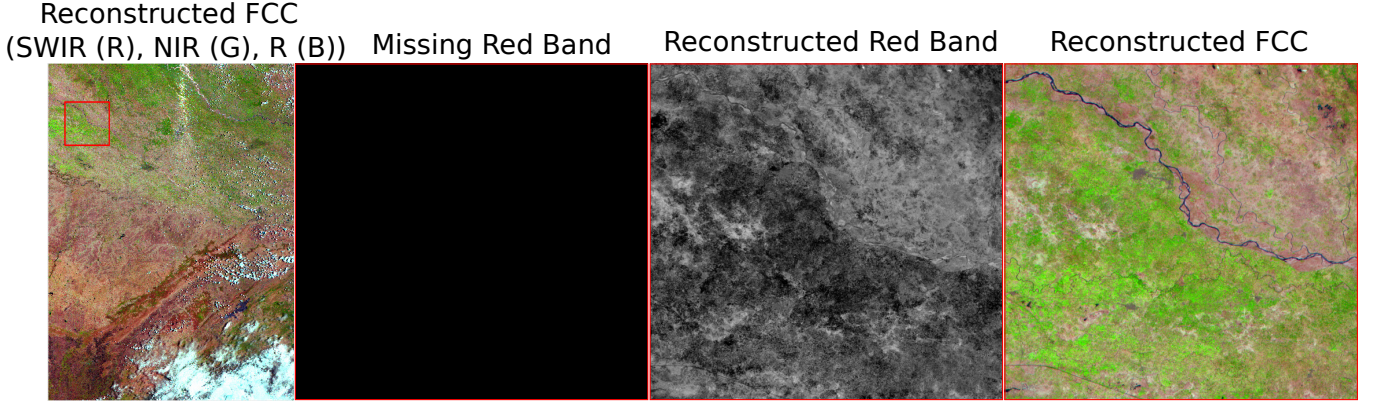


Fig. 7. Qualitative analysis of reconstructed band in real. ALERT reconstructs photo-realistic missing red band. The reconstructed band is auto-registered and coherent with the existing bands: SWIR and NIR.

reconstructed image has relatively inferior image quality as compared to adversarial learning which is consistent with our aforementioned claim of realistic manifold. In addition, we compare various models in terms of difference between original and reconstructed statistical properties in Fig. 4. Among the compared models, the proposed approach captures both mean and variance statistics reasonably well.

Justifying the efficiency of ALERT, we extend our analysis towards DN values. We cross correlate the reconstructed Red band with the original band pixel-wise. We observe that the reconstructed DN values closely follow the original DN values, as shown in Fig. 5. Further, we analyze the consistency of reconstructed band with existing bands in terms of mean spectral characteristics in Fig. 6. We use all the 21250 test samples to calculate the first and second order moments of reconstructed and original bands while comparing the mean spectral characteristics. The reconstructed DN values follow near identical spectral signature as that of the original band.

Intent on checking generalization, we employ ALERT in a real environment where we do not have access to corresponding red band. As a matter of fact, we report that it generalizes well to real environment and successfully reconstructs missing band entirely from deep priors, as shown in Fig. 7. Also, the reconstruction process does not inherently inject any sort of geometric distortion and resolution degradation as a part of multiple convolutions. We report that the proposed architecture contributes significantly to achieve this feat.

Here, we provide more experimental results on adversarial band reconstruction aiming at various spectral signatures. As shown in Fig. 8A, ALERT successfully generates missing red band in real environment where the synthetic band is consistent with existing bands in terms of spectral and spatial resolution. In Fig. 8A (a), (b), and (c), we observe that ALERT captures the essence of several individual targets including barren land, vegetation, lake water, and river water. Further, we analyse the reconstructed band in NIR, R, and G false colour composites in Fig. 8B. We report that the reconstructed band is coherent across multiple spectral bands of real sensors. In Fig. 8B (a), (b), and (c), we illustrate some of the aforementioned key signatures in different colour composites to analyse the spectral consistency and visual perceptual quality. We report

that the reconstructed missing band based on spectral priors, as generated by ALERT, consists of essential spatial and spectral components that make it visually indistinguishable from real band.

A. Spectral Signature Analysis

To this end, we observed the efficacy of ALERT in reconstructing photo-realistic missing band. However, it would be interesting to study whether the reconstructed band follows desired spectral patterns. In Fig. 9, we show the mean of spectral minimum and maximum of 21250 test samples. We observe that the DN values of reconstructed red band follows the exact trend as that of the actual red band. Though the mean of spectral maximum of reconstructed DN values incurs a slight offset in Fig. 9 (b), it is reasonably minimal as compared to the dynamic range. Despite that it preserves similar pattern as that of the actual red band.

Further, we analyse the spectral signature of particular targets, such as cloud and snow to make the spectral consistency more concrete. In this experiment, we randomly select a small patch of pixels from a particular target of actual red band and correlate its mean with the corresponding mean of synthesized red band. As shown in Fig. 10 (a) and (b), we observe that the DN values of reconstructed band closely follows that of the original band in the simulated environment, suggesting high confidence on spectral similarity.

B. Results on WorldView-2

Here, we discuss the efficacy of ALERT on very high resolution multi-spectral imagery. To realize this, we have used 12208 multi-spectral images of WorldView-2 in testing. These images are acquired over Washington DC with 1.6 m product level GSD. Since WorldView-2 does not provide SWIR band, we retrained the model with B, G, and NIR(1) as source distribution and R as the target distribution. We used the same learning algorithm in both the mid (56 m) and high resolution (1.6 m) datasets to study its generalization. As shown in Fig. 11, the reconstructed band has 63.25 dB PSNR and 0.9986 SSIM which is a favourable indicator of image perceptual quality. Note that the training data consists of images which

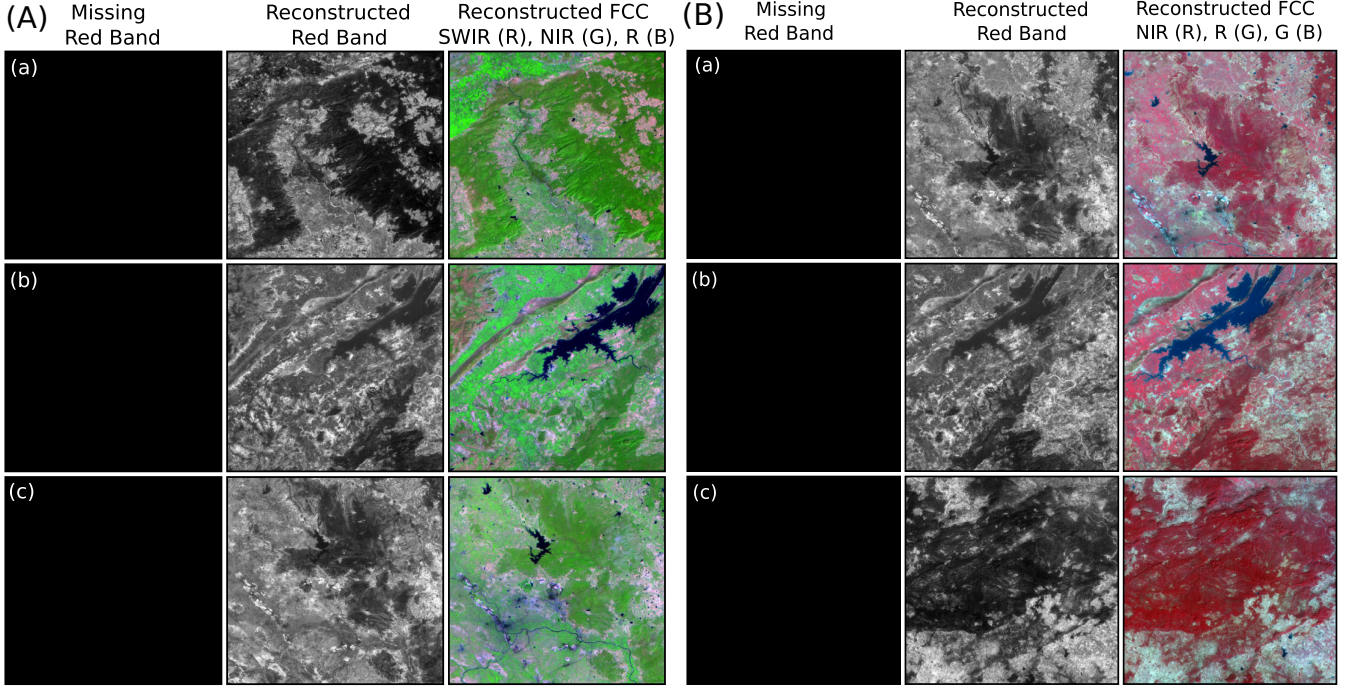


Fig. 8. Qualitative analysis of reconstructed band in real environment. ALERT reconstructs photo-realistic missing red band. The reconstructed R band is auto-registered and coherent with the existing bands: G, NIR and SWIR.

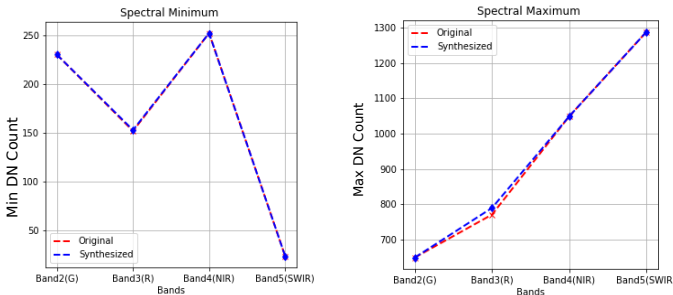


Fig. 9. Spectral minimum and maximum analysis. The synthesized missing red band closely follows the original band.

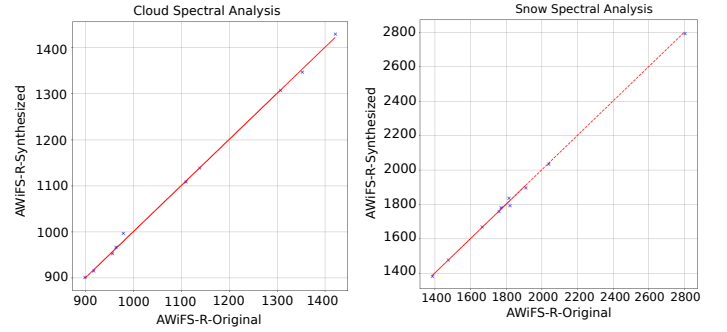


Fig. 10. Spectral characteristics analysis of cloud and snow targets. The reconstructed red band has almost identical DN values when compared with original red band.

are acquired only over Indian territory and the test data is from a completely different physical location. However, the underlying features of test images loosely follow Independent and Identically Distributed (I.I.D.) statistics which is mainly the reason for such favourable outcome.

VI. CONCLUDING REMARK

In this manuscript, we studied the efficacy of supervised and adversarial learning to address the complex task of missing band reconstruction with the sole supervision of existing spectral and spatial prior distribution. We analysed the difficulty in reconstructing a missing band in comparison with missing partial information. Thereafter, we proposed an adversarial learning with expert regularization framework for addressing this issue with sufficient degree of realism necessary to meet scientific fidelity. We showed compelling

benefits of using proposed ALERT framework in remote sensing applications as compared to supervised learning. Further, we conducted extensive experimentation both qualitatively and quantitatively using state-of-the-art evaluation metrics to support our aforementioned claims. Also, we proposed an architectural innovation, namely RDB to assist adversarial learning especially in satellite image processing. Based on our extensive experimentation, we showed considerable gain in image quality, training, and inference as compared to completely supervised learning. Though we have demonstrated the adversarial learning with expert regularization for missing band reconstruction, we believe it can be further extended to other complex computer vision tasks without having to lose any performance benefits.

ACKNOWLEDGEMENT

The author would like to thank all the members of Optical Data Processing Division (ODPD), Signal and Image Processing Group (SIPG), Space Applications Centre (SAC) for their continuous support throughout this research.

REFERENCES

- [1] W. Xie and Y. Li, "Hyperspectral imagery denoising by deep learning with trainable nonlinearity function," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1963–1967, 2017.
- [2] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, 2018.
- [3] P. Xiao, Y. Guo, and P. Zhuang, "Removing stripe noise from infrared cloud images via deep convolutional networks," *IEEE Photonics Journal*, vol. 10, no. 4, pp. 1–14, 2018.
- [4] C. Tuna, G. Unal, and E. Sertel, "Single-frame super resolution of remote-sensing images by convolutional neural networks," *International journal of remote sensing*, vol. 39, no. 8, pp. 2463–2479, 2018.
- [5] L. Rout, Y. Bhateja, A. Garg, I. Mishra, S. M. Moorthi, and D. Dhar, "Deepswir: A deep learning based approach for the synthesis of short-wave infrared band using multi-sensor concurrent datasets," *arXiv preprint arXiv:1905.02749*, 2019.
- [6] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [7] N. Komodakis, "Image completion using global optimization," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 442–452, IEEE, 2006.
- [8] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE transactions on image processing*, vol. 19, no. 5, pp. 1153–1165, 2010.
- [9] H. Shen, C. Zeng, and L. Zhang, "Recovering reflectance of aqua modis band 6 based on within-class local fitting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 1, pp. 185–192, 2010.
- [10] X. Li, H. Shen, L. Zhang, H. Zhang, and Q. Yuan, "Dead pixel completion of aqua modis band 6 using a robust m-estimator multiregression," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 768–772, 2013.
- [11] C. Zeng, H. Shen, and L. Zhang, "Recovering missing pixels for landsat etm+ slc-off imagery using multi-temporal regression analysis and a regularization method," *Remote Sensing of Environment*, vol. 131, pp. 182–194, 2013.
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [13] A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, and M. Hoffman, "Aerial spectral super-resolution using conditional adversarial networks," *arXiv preprint arXiv:1712.08690*, 2017.
- [14] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision*, pp. 19–34, Springer, 2016.
- [15] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 518–525, 2017.
- [16] B. Arad, O. Ben-Shahar, and R. Timofte, "Ntire 2018 challenge on spectral reconstruction from rgb images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 929–938, 2018.
- [17] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, 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*, pp. 4681–4690, 2017.
- [18] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.
- [19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, pp. 214–223, 2017.
- [20] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, pp. 5767–5777, 2017.
- [21] M. Henaff, A. Canziani, and Y. LeCun, "Model-predictive policy learning with uncertainty regularization for driving in dense traffic," *arXiv preprint arXiv:1901.02705*, 2019.
- [22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [24] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
- [25] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 864–873, 2018.
- [26] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, pp. 305–319, 2018.

WorldView-2 Multi-Spectral FCC (NIR (R), R (G), G (B))

Original
(PSNR/SSIM)

Reconstructed
(63.25/0.9986)

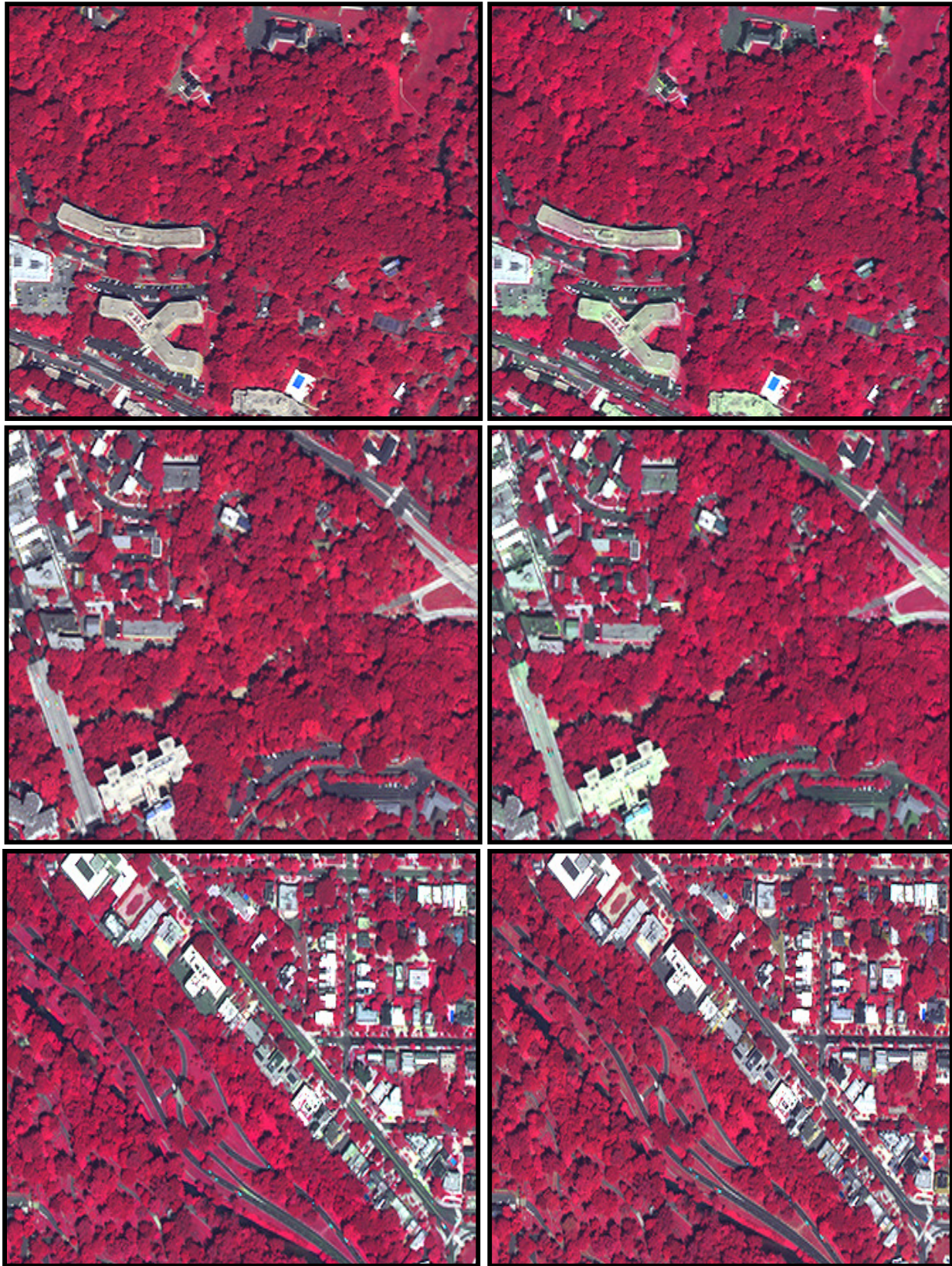


Fig. 11. Qualitative analysis of reconstructed band on WorldView-2. ALERT reconstructs very high resolution missing red band with spectral and spatial consistency.