A Review on Remote Sensing Data Fusion with Generative Adversarial Networks (GAN)

Peng Liu, Jun Li *IEEE Fellow*, Lizhe Wang *IEEE Fellow*, Guojin He,

Abstract—In the past decades, remote sensing (RS) data fusion has always been an active research community. A large number of algorithms and models have been developed. Generative Adversarial Networks (GAN), as an important branch of deep learning, show promising performances in variety of RS image fusions. This review provides an introduction to GAN for remote sensing data fusion. We briefly review the frequently-used architecture and characteristics of GAN in data fusion and comprehensively discuss how to use GAN to realize fusion for homogeneous RS data, heterogeneous RS data, and RS and ground observation data. We also analyzed some typical applications with GAN-based RS image fusion. This review takes insight into how to make GAN adapt to different types of fusion tasks and summarizes the advantages and disadvantages of GAN-based RS data fusion. Finally, we discuss the promising future research directions and make a prediction on its trends.

Index Terms—Generative Adversarial Networks (GAN), Remote Sensing Data Fusion.

I. INTRODUCTION

REMOTE sensing data fusion has always been a hot topic that is extensively and deeply studied, since our way of acquiring RS data is becoming increasingly diverse. In many situations, Earth-observation data can be acquired by both airborne and space-borne missions, by frame cameras or line scanners, by synthetic aperture radar (SAR) sensors, by video satellite, or by spectrometers providing infrared or multi or hyperspectral imagery [1]. In addition, accompanying with them, there might be information about atmospheric parameters, environmental parameters, or optical imagery of smart phones, UAV images, point clouds or depth imagery, or geographic information system (GIS) data, or global positioning system (GPS) coordinates. All these data show great diversity in their attributes such as spatial, spectral, and temporal resolution, scale, measurement accuracy, and way of observation (extensive observations or point-wise measurements). These diversities lead to higher demands on the algorithm and model of data fusion.

The technology of RS data fusion has been in progress with the development of machine learning, computer vision, or signal processing, etc. Many advanced theories or methods such as Bayesian theory, variation method, sparse representation, compressed sensing, non-local means, low rank, etc. gave a huge boost to the development of remote sensing data fusion. Deep learning is a new milestone for the study of data fusion. Automatic feature learning and end-to-end architectures show promising performances and are widely spreading among

every branch of data fusion. However, many difficulties in RS data fusion study are remaining, such as the mutual restraint between spatial, temporal or spectral resolution, the inconsistency of heterogeneous data, etc. For multi-source data, there are often both opposites and complementary, which is similar to the unity of opposites in contradiction. However, in a common deep architecture such as CNN, there is no coordination mechanism for the unity of opposites relationship of a fusion process. We need a new architecture to model the relationship of multi-source data.

As an important sub-branch of deep learning, Generative Adversarial Networks (GAN) is opening of a new horizon of the study on RS data fusion. GAN is a system of two neural networks contesting with each other in a zero-sum game framework. They were introduced by Ian Goodfellow et al. in 2014 [2]. The main idea of GAN can be explained by Nash equilibrium in game theory [3]. GAN has been successfully applied to image style translation in computer vision, and even data fusion. As mentioned above, the unity of opposite relationships widely exist in most RS fusions. GAN is born with the mechanism to model this unity of opposite relationships. As result, GAN is attracting more and more attentions of the RS data fusion community. This survey will summarize how we use GAN to realize RS data fusion for different types of data and discuss advantages or disadvantages in GAN-based RS fusion.

We believe that this review will not only contribute to a summarization of the current situation of remote sensing data fusion algorithms with GAN, but will also guide the reader to thoroughly think about the nature of fusion and point out development directions in the future. We also hope that our survey will be beneficial to the different applications based on remote sensing data fusion.

The rest of this review is organized as follows. In Section II, we made a brief description of the definition, architecture, and loss function of GAN. In Sections III, we discussed the degraded model of remote sensing data. In Section IV, we discussed the taxonomy of remote sensing data fusion. Sections V to VII reviewed the algorithms of GAN for homogeneous RS data, heterogeneous RS data, and RS and ground observation data. Section VIII, some of applications with GAN for image fusion were introduced. In Section X we made some discussions about GAN for RS data fusion. In Section XI we concluded the review.

II. GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Networks (GAN) [2] are types of neural network architectures consists of generator and discriminator, which is capable of generating new data that conforms

Peng Liu is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

This work was supported by the National Natural Science Foundation of China (Nos. 41971397 and 61731022). Corresponding author Lizhe Wang (lizhe.wang@gmail.com)

to learned patterns by both generative and adversarial process. The architecture of GAN is simply illustrated as Fig. 3.

Fig. 1: The architecture of GAN.

Let P_x denote the distribution of random variable x , and P_z denote the the distribution of random variable z . GAN learns to find a balance relationship through the competition of the generator G and the discriminator D . Specifically, G maps the random variable z to data space of x , trying to produce $G(z)$ that is subject to the real x distribution, while D distinguishes whether its input is the real data x or the generated data $G(z)$. D is trained to maximize the probability of assigning value 1 to real data and 0 to fake data from G , while G is trained to produce data who make D think them as real. D and G play the two-player minmax game that can be presented by:

$$L_{GAN}(\theta_D, \theta_G) = E_{x \sim P_x}[\log D(x)] - E_{z \sim P_z}[\log(1 - D(G(z)))] \quad (1)$$

where θ_D and θ_G are the parameters of discriminator D and generator G . GAN have received wide attentions in the machine learning field for their potential to learn high-dimensional, complex real data distribution [4]. The reader can also refer to some surveys on GAN as [5] [6] [7] [8] [9] [10][11] and [12]. Specifically, they do not rely on any assumptions about the distribution and can generate real-like samples from latent space in a simple manner. This special property makes GAN be applied to various applications such as image synthesis, image attribute editing, image translation, domain adaptation, and other academic fields. In this article, we explain how GAN is reformed and applied to RS data fusion. We mainly focus on how homogeneous RS data, heterogeneous RS data, or ground observation data are integrated with GAN.

A. Commonly Used Architectures of GAN in RS Fusion

Variety of GAN architectures have been developed in recent years. For examples, BEGAN [20], EBGAN [33] and MAGAN [34] are auto-encoder architecture, StackedGAN [35], GoGAN [36], Progressive GAN [27] are hierarchy architecture, etc. We summarized some of the milestones in the development of GAN as Fig. 2. They show different advantages and disadvantages in generating data. For the problem of remote sensing data fusion, most of the studies did not directly use existing GAN architectures but modify their architectures and make them more suitable for the process of fusing multi-source data. Among them, conational GAN (CGAN) and cycle-GAN are two mostly used GAN architectures in many fusion studies.

Fig. 3: The architecture of CGAN.

CGAN: GAN can be extended to a conditional model if both the generator and discriminator are conditioned on some extra history data c . In this review, variable c is defined as an auxiliary image (or data) with the same scene of z from the different satellites or sensors. When performing the conditioning by feeding c into both the discriminator and generator as additional inputs, the conditional generative adversarial nets (CGAN) [15] is defined as

$$L_{CGAN}(\theta_D, \theta_G) = E_{x \sim P_x}[\log D(x/c)] - E_{z \sim P_z}[\log(1 - D(G(z/c)))] \quad (2)$$

where P_x is the probability distribution of the real data x , P_z is the probability distribution of the input data z , and $E[\cdot]$ is the function of expectation. The condition variable c is introduced into both the generator and discriminator. In the discriminator, x and c are combined to be presented as inputs of the discriminator. The CGAN framework allows for considerable flexibility in how this hidden representation is composed. It means that the output of the generator will be not far from the feature space of the reference image c .

CycleGAN: CycleGAN [26] is an extension of the GAN architecture that simultaneously trains two generator models and two discriminator models. The CycleGAN learns a mapping between input and output images using an unpaired dataset.

Fig. 4: Architecture of CycleGAN.

CycleGAN is composed of two generator models: one generator (Generator-A) for generating images for the first domain (Domain-A) and the second generator (Generator-B) for generating images for the second domain (Domain-B). Each generator has a corresponding discriminator model (Discriminator-A and Discriminator-B). Discriminator models are then used to determine how plausible the generated images are and update the generator models accordingly.

The CycleGAN introduced cycle consistency into the architecture. This is the idea that an image output of Generator-A can be used as input to Generator-B, and then the output of Generator-B should match the input of Generator-A (original

Fig. 2: The development of GAN. LAPGAN[13], LSGAN[14], CGAN [15], AAE [16], ACGAN [17], info-GAN [18], StarGAN [19], ImprGAN, BEGAN[20], WGAN [21], WGAN-GP [22], DCGAN[23], BiGAN [24], SGAN [25], CycleGAN [26], PGGAN[27], StyleGAN[28], SAGAN [29], BigGAN [30], SinGAN [31], YLG [32]

image). The reverse is also true. Cycle consistency means that Generator-A and Generator-B are mutual inverse mapping.

CycleGAN is to learn mapping functions between two domains X and Y given training samples $\{\mathbf{x}_i\}_1^N$ and $\{\mathbf{y}_i\}_1^N$ where $\{\mathbf{x}_i\}_1^N \subset X$ and $\{\mathbf{y}_i\}_1^N \subset Y$. As illustrated in Fig. 3, CycleGAN model includes two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. At the same time, there are also two adversarial discriminators D_X and D_Y . With these definitions, CycleGAN is denoted as

$$L_{GAN}(G, F, D_X, D_Y) = L_\alpha(G, D_Y, X, Y) + L_\beta(F, D_X, Y, X) + L_{cyc}(G, F) \quad (3)$$

where

$$L_\alpha(G, D_Y, X, Y) = E_{\mathbf{y} \sim p_Y} [\log(D_Y(\mathbf{y}))] + E_{\mathbf{x} \sim p_X} [\log(1 - D_Y(G(\mathbf{x})))] \quad (4)$$

and

$$L_\beta(F, D_X, Y, X) = E_{\mathbf{y} \sim p_Y} [\log(D_X(\mathbf{y}))] + E_{\mathbf{x} \sim p_X} [\log(1 - D_X(F(\mathbf{x})))] \quad (5)$$

and

$$L_{cyc}(G, F) = E_{\mathbf{y} \sim p_Y} [\|\mathbf{F}(G(\mathbf{y})) - \mathbf{y}\|_1] + E_{\mathbf{x} \sim p_X} [\|\mathbf{G}(F(\mathbf{x})) - \mathbf{x}\|_1] \quad (6)$$

For CGAN, there are more than one inputs. In many fusion scenarios, the multi-source data can be taken as conditional data which are matching the characteristic of CGAN. For Cycle-GAN, it does not need to be one-by-one mapping between the domain X and Y . Its weak supervision is very convenient to many fusion problems. CGAN and Cycle-GAN are two mostly used GAN in many RS data fusion studies.

B. Loss Function of GAN in RS Fusion

In common GAN, the loss function l_G^{ADV} of generator is a statistical expectation which is an entirety but not individual. Obviously, it is not enough for the image fusion task because we want to obtain an individual fusion image from multi-source images. Loss function provides the supervised information for GAN. Except for statistical expectation, content loss, perception loss, spectral loss, etc. all can be introduced

into the loss function of the generator of data fusion. Let the total loss be denoted as

$$l_G = l_G^{ADV} + \lambda_1 l_G^{PXL} + \lambda_2 l_G^{FTR} + \lambda_3 l_G^{SAM} + \dots, \quad (7)$$

where l_G^{ADV} is the adversarial loss, l_G^{PXL} is the pixel content loss, l_G^{FTR} is the feature perception loss, and l_G^{SAM} is the spectral loss. We define them as

$$l_G^{ADV} = -E[\log(1 - D(G(\mathbf{x}_1, \mathbf{x}_2)))] \quad (8)$$

$$l_G^{SAM} = SAM(\mathbf{x}_F, G(\mathbf{x}_1, \mathbf{x}_2)) \quad (9)$$

$$l_G^{PXL} = \|\mathbf{x}_F - G(\mathbf{x}_1, \mathbf{x}_2)\|_p \quad (10)$$

$$l_G^{FTR} = \|F(\mathbf{y}_1, \mathbf{y}_2) - F(\mathbf{x}_1, \mathbf{x}_2)\|_p \quad (11)$$

Pixel content loss is a L_p norm. We denote it as l_G^{PXL1} if it is L_1 norm or l_G^{PXL2} if it is L_2 norm. Structural similarity (MS-SSIM) [37] is also a common loss in GAN for RS image fusion. Most metrics for evaluating fusion results all can be introduced into GAN to be loss function.

We notice that there are many methods [38][39][40] that use the perceptual loss calculated by the feature of a self-designed pre-training network to ensure the quality of image generation. Similar to the style transfer and super-resolution [41][42][43] for natural images, the features of the pre-trained VGG network are often used to calculate the perceptual loss to obtain better image results.

We find that GAN can be understood as a special way to establish the relationship between two distributions. How to use GAN depends on how to understand the nature of the fusion problem. Since most fusions have more than one input and one output, very few of study directly use vanilla GAN in a traditional way. The architecture, hidden layer, active function, loss function, and optimization process are all reformed to adapt to RS data fusion in most of the studies. However, different from traditional fusion methods, they all belong to implicit model and data-driven fusion. In the next section, we simply review the two different main currents for understanding RS data fusion.

III. EXPLICIT AND IMPLICIT MODEL IN RS IMAGE FUSION

A. Explicit Model and Model-driven

For RS data, the degradation process can be explicitly denoted as

$$\mathbf{y} = \mathcal{H} * \mathbf{x} + \mathbf{n} = \mathbf{H}_1 \mathbf{H}_2 \cdots \mathbf{H}_k \mathbf{x} + \mathbf{n} \quad (12)$$

where $\mathbf{y}$ is the degraded and observed image, $\mathbf{x}$ is the original image, $\mathbf{H}$ is the degradation matrix, and $\mathbf{n}$ is the additive noise. The operator $*$ is the convolution. The noise $\mathbf{n}$ is often assumed as a zero-mean white Gaussian process with variance σ^2 . The total degradation $\mathcal{H} = \mathbf{H}_1 \mathbf{H}_2 \cdots \mathbf{H}_k$, where $\mathbf{H}_k$ is the degradation by a single factor, for examples, $\mathbf{H}_1$, $\mathbf{H}_2$ and $\mathbf{H}_3$ can be the degradation in temporal domain, spatial domain, spectral domain, etc. Image $\mathbf{y}$ and $\mathbf{x}$ can be denoted by vectors, that $\mathbf{y} = [y_1, \dots, y_N]^T$ and $\mathbf{x} = [x_1, \dots, x_N]^T$. $\mathbf{H}_k \mathbf{x}$ is the matrix-vector form, where $\mathbf{H}_k$ is convolution matrix which is approximated by a block-circulant matrix.

For many fusion problems, we need first obtain different $\mathbf{y}$ and then reconstruct original data $\mathbf{x}$. This explicit model is suitable for homogeneous RS fusion, because the multi-source data often come from different spatial, temporal or spectral sampling. For example, in reference [44], the author proposed an integrated framework for the spatio-temporal-spectral fusion, where the relationship between $\mathbf{y}$ and $\mathbf{x}$ can be comprehensively analyzed to construct the explicit spatio-temporal-spectral relationship model.

Different type of degradation usually presents different problems and lead to different algorithms. The explicit models have very clear physical meaning, so that they have almost all the advantages of model-driven methods. However, for some heterogeneous RS data fusion or RS and ground observation data fusion, they can not be denoted as an explicit model. Therefore, the effect of the explicit model will be very limited in some fusion scenarios.

B. Implicit Model and Data-driven

For the early studies of RS data fusion, explicit model and model-driven play more important roles. Either observation model or data model is the fundamental consideration in most algorithms or processes of fusion. However, with the development of deep learning and automatic feature learning, it is possible to weaken the observation model $\mathcal{H}$. The process of fusion can be easily realized in the hidden layers of CNN or other deep architectures, where the physical meaning of the observation or data distribution model is dramatically weakened. We call this kind of trend of fusion as implicit model and data-driven, which is denoted as

$$\mathbf{x} = \mathcal{F}_1(\cdots \mathcal{F}_k(\cdots \mathcal{F}_K(\mathbf{y}))) \quad (13)$$

where $\mathcal{F}_k(\cdots)$ is the k -th hidden layer of an implicit model such as CNN. GAN for RS data fusion is a typical method of implicit model and data-driven. In most cases, we do not need to consider the observation equation $\mathbf{y} = \mathcal{H} * \mathbf{x} + \mathbf{n}$. One of the advantages is that the fusion process can avoid some problems resulted from the complexity of $\mathcal{H}$ especially for heterogeneous RS data etc. At the same time, the explainability

or interpretability of the fusion becomes so weak that we can not judge how and why the fusion produces better performances in some cases. Another problem is that data-driven fusion usually needs more training data and more computation sources. Now, data-driven fusion is becoming more and more popular, but there are also some studies who start to rethink the relationship between data-driven and model-driven. There are already studies that try to combine the two schemes in fusion.

IV. TAXONOMY OF REMOTE SENSING IMAGE FUSION

There are different taxonomies for remote sensing image fusion if we stand in a different perspective. A very common taxonomy of fusion is: observation level, feature level, and decision level [1]. This taxonomy depends on: which stage does the fusion occurs at. This kind of taxonomy is not limited to remote sensing data. Medical image, industry computer vision, social media data, etc are all can be included in this taxonomy.

Fig. 5: It is homogeneous RS data fusion if the two types of data are with a overlaid spectral range.

Another taxonomy of remote sensing data fusion is: Homogeneous RS data, Heterogeneous RS data, RS and Ground Observation data, and RS Data Assimilation [45]. This taxonomy is based on the different characteristics of the data.

Homogeneous RS data mainly refers to PAN image, multi-spectral image and hyperspectral image, etc, who are basically optical images but with different band range or band number. **Heterogeneous RS data** mainly refers to SAR-Optical image fusion, LiDAR-optical image fusion, etc, whose imaging mechanisms are so different that it is impossible to produce homogeneous images with a clear physical meaning. **RS and Ground Observation data fusion** refers to the fusion of remote sensing observation and ground station observation. For a specific land surface or environmental parameter, station observation is accurate but sparse, while remote sensing observation is dense but not so accurate. **RS Data Assimilation** refers to the integration of satellite observations into the model of weather prediction, hydrological forecast, climate impact studies, ocean dynamics, carbon cycle monitoring, etc. to provide them more initial or bound conditions, so that the uncertainty and the unsteadiness of the model will be reduced by the integration. In this survey, we adopt this

Fig. 6: A taxonomy of remote sensing image fusion with GAN.

kind of taxonomy for remote sensing data fusion (without assimilation) since it is convenient to address the problem of how to use GAN in fusion.

V. HOMOGENEOUS RS DATA FUSION

One of the most important characteristics for homogeneous RS data is that they have intersection areas in imaging spectral, spatial, or temporal domain. For example, there are overlap regions of the spectral band between PAN data and MS data, so that they can be used as spectral-spatial fusion. If we want to perform temporal-spatial image fusion, the two sequences of RS data usually need to have similar spectral bands and overlap temporal ranges such as Landsat and MODIS. It means that the foundation of homogeneous RS data fusion is their common physical meaning and similar imaging mechanism. The goal of most homogeneous RS data fusion is to enhance the resolution, which may be spatial resolution, spectral resolution, and temporal resolution, or even angle resolution. At the same time, most homogeneous RS data fusion is pixel-level or observation level fusion, because in many cases the fusion can be realized by the combination solution of several different observation equations as Eq. (14).

$$\begin{cases} \mathbf{y}_1 = \mathcal{H}_1 * \mathbf{x} + \mathbf{n}_1 \\ \mathbf{y}_2 = \mathcal{H}_2 * \mathbf{x} + \mathbf{n}_2 \\ \dots \\ \mathbf{y}_k = \mathcal{H}_k * \mathbf{x} + \mathbf{n}_k \end{cases} \quad (14)$$

where $\mathcal{H}_n$ is the k th degradation process, $\mathbf{y}_k$ is the k th observation. Most heterogeneous RS data can not be represented as combination solution of Eq. (14), because heterogeneous RS data have no common $\mathbf{x}$. Maybe, invisible-infrared RS

data fusion is an exceptional case. They have a similar imaging mechanism but without overlap region of spectral band, which makes it belong to heterogeneous RS data. Different from heterogeneous RS data, only homogeneous data can be integrated with a clear physical meaning. Therefore, pixel-level resolution fusion (spatial, spectral or temporal, etc.) are the most common type in homogeneous RS data fusion.

A. Spatial-spectral image fusion

The spatial resolution and spectral resolution are often mutually restricted for the sensors of most satellites. For example, the PAN image can achieve higher spatial resolution, while MS or HS images can achieve higher spectral resolution. It was natural to think that PAN and Low Resolution Multi-Spectral (LRMR) images would be complementary. Spatio-spectral fusion may be the most common approach in remote sensing image fusion. Its goal is to obtain a fused image with both high spatial and spectral resolutions, as well as High Resolution Multi-Spectral (HRMS) image. The classical spatio-spectral fusion methods include PAN-MS fusion, PAN-HS fusion, and MS-HS fusion, etc. In this survey, we category Infrared-visible light image fusion into heterogeneous RS fusion although their complementarity is related to spatial resolution and spectrum.

1) *Pan-sharpening image fusion*: Pan-sharpening image fusion has a very long history in the remote sensing community. Deep learning such as CNN was already successfully introduced into Pan-sharpening. Most of them [46] [47] focus on generating the high-quality fusion image with both accurate spectral distributions and reasonable spatial structures.

Fig. 7: The x-axis is the wavelength. The y-axis is the relative response. The spectral spectral response function (SRF) describes the sensitivity of the photosensor to optical radiation of different wavelengths. (a) The SRF of IKONOS. (b) The SRF of Quick-Bird.

Fig. 8: (a) PAN image. (b) MS image. (c) Fusion image.

Usually, the spectrum of PAN images should cover the range of the combination spectrum of MS images. In most model-driven PAN-sharpening studies, there is an approximation assumption:

$$P \approx \omega_1 \mathbf{x}_1 + \omega_2 \mathbf{x}_2 + \dots + \omega_n \mathbf{x}_n = \Omega X_h \quad (15)$$

where P is the PAN image, X_h are the band images of high resolution MS images. Ω is the weight matrix for bands of MS image X_h , where $X_h = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and $\Omega = \{\omega_1, \dots, \omega_n\}$. This expression is already not necessary in current deep learning-based PAN-sharpening fusion because it is hidden within the data-driven process. However, its physical meaning illustrated the source of difficulty in PAN-sharpening, that is the spectral response function of PAN usually can not perfectly cover the range of spectral response function of MS. The corresponding phenomenons are that there are spectral distortions in the fusion images when the spatial resolution was improved.

When CNN becomes the major stream in PAN-sharpening, the observation equation is no longer represented explicitly. As far as we know, a common GAN only has one input and one output. It means that we can not directly use common GAN to perform PAN-sharpening in most cases. At the same time, GAN does really have some advantages in establishing the relationship between multi-source data since it is a good model-free statistical estimation method. To make GAN suitable for image PAN-sharpening, many studies mainly improve GAN in its architectures and loss functions.

The structure of PSGAN [48] and HPGAN [49] as in Fig. 14 (a) and (b) are very straightforward and concise.

PSGAN [48] uses two subnetworks to extract features. HPGAN [49] induces the Spectral Response Function (SRF) into generator network and use skip connection between generator and discriminator. Considering the oppositional relationship between spatial-resolution and spectral-resolution, Fig. 14 (c) Pan-GAN [50] proposed a new unsupervised framework for pan-sharpening, where the generator separately establishes the adversarial games with the spectral discriminator and the spatial discriminator. Fig. 14 (d) PanColorGAN [51], the pan-sharpening process was guided by colorization of PAN Images via GAN, so that the structure is simplified and similar to PSGAN [48]. In Fig. 14 (e), PercepPan [52], a constructor was combined with generator to produce PAN and HRMS images who enter discriminator together. In Fig. 14 (f), PAN images were utilized as condition in CGAN to form frame of RED-CGAN [53]. The architecture of generator can also be designed by exploring the posteriori distribution, which assists selecting the appropriate generator parameters [54] based on Bayesian GAN. We can notice that, Fig. 14 (c) and (e) do not need the HRMS images as supervision information which are different from Fig. 14 (a), (b), (d) and (f). Most of the studies re-designed the structure of generator of GAN, but the structures of discriminator were relative simple. The image fusion process mainly occurred in the generator but not discriminator.

In common GAN, the loss function l_G^{ADV} of generator is a statistical expectation which is an entirety but not individual. Obviously, it is not enough for the image fusion task because we want to obtain an individual HRMS image from the PAN and LRMS images. The loss function provides the supervised information for generators. Except for statistical expectation, content loss, perception loss, spectral loss, etc. all can be introduced into the loss function of the generator. The total loss is denoted as

$$l_G = l_G^{ADV} + \lambda_1 l_G^{PXL} + \lambda_2 l_G^{FTR} + \lambda_3 l_G^{SAM} + \dots, \quad (16)$$

where l_G^{ADV} is the adversarial loss, l_G^{PXL} is the pixel content loss, l_G^{FTR} is the feature perception loss, and l_G^{SAM} is the spectral loss.

TABLE I: Loss Functions for Different Methods.

Dataset	l_G^{ADV}	l_G^{PXL1}	l_G^{PXL2}	l_G^{FTR}	l_G^{SAM}
PSGAN [48]	✓	✓			
HPGAN [49]	✓		✓		✓
Pan-GAN [50]	✓		✓		✓
PanColorGAN [51]	✓	✓			
PercepPan [52]	✓	✓		✓	
RED-CGAN [53]	✓	✓			

In the training process, most PAN-sharpening methods rely on the ground-truth image, which usually is unavailable for neural networks training. Unlike the remote sensing image classification or segmentation problem, it is impossible to provide ground-truth HRMS images for Pan-sharpening by manual annotation. Therefore, the author in [52] pointed out that: neural networks-based methods usually follow Walds protocol [55] to take the original LRMS images as labels by

Fig. 9: (a) PSGAN [48], (b) HPGAN [49], (c) Pan-GAN [50], (d) PanColorGAN [51], (e) PercepPan [52], (f) RED-CGAN [53].

degrading the original LRMS and PAN images into a lower resolution space as input. In this supervising manner, the pan-sharpening network G is trained in the lower resolution space. Strictly speaking, this operation may not make sense. The features of LRMS in lower resolution space are fewer than that of HRMS in high resolution space, so it may not enough to supervise GAN to generate HRMS with more details.

Some studies believe that it is not necessary to use the degradation step in training a Pan-sharpening network. In [52], the authors proposed an unsupervised Pan-sharpening framework. As in Fig. 14 (e), the degradation step for training was eliminated, but it leverages an auxiliary reconstructor network instead. Degradation images and LRMS images are used as input of discriminator to avoid degradation steps before training. In [50], The authors propose an unsupervised framework for Pan-sharpening based on GAN, which does not rely on ground truth during network training. In this method, the generator separately establishes the adversarial games with the spectral discriminator and the spatial discriminator so as to generate the rich spectral information and the spatial information for HRMS. The problem of degradation steps in training is worthy to notice. However, moving the degradation step before the discriminator as [52] and [50] may lead to a weak discriminator in GAN.

Except for GAN, many other different types of deep archi-

tures all can be applied to Pan-sharpening fusion of remote sensing. We can take GAN as one of the special instances. Please refer to [56] in which the authors make a comprehensive review of Pan-sharpening satellite imagery.

2) *Multispectral-Hyperspectral image fusion*: Hyperspectral (HS) image usually has a low spatial resolution although it is with a large number of bands. On the contrary, MS sensors can obtain an image with a higher spatial resolution but with fewer spectral bands. This is similar to the relationship between PAN and MS images. There are already some surveys on HS-MS fusion such as [57][58]. But they did not focus on GAN methods for fusion. In reference [58], HS-MS fusion methods are categorized as four classes, including methods extended by Pan-sharpening, MF-based methods, tensor representation (TR) based methods, and deep CNN based methods. In reference [57], HS-MS fusion methods are categorized into four classes such as component substitution (CS), multiresolution analysis (MRA), spectral unmixing, and Bayesian probability. In reference [59], spectral and spatial information fusion are reviewed and the current trends and challenges for hyperspectral image classification were discussed. Among them, deep CNN-based methods are developing very fast and have received more and more attention in the community of HS-MS fusion. We would take Pan-sharpening as a particular instance of the HS-MS fusion, and then many Pan-sharpening

approaches can be extended to the fusing of HS and MS images. GAN-based HS-MS fusion should be taken as a sub-branch of deep learning-based fusion methods.

However, there is still not so much published work with GAN for standard HS-MS fusion. We can only find some similar research such as Hyperspectral Pansharpening with GAN [60][61], HS-MS fusion with generative model (not GAN) [62] or MS-MS fusion with GAN [63][64][65], etc. In reference [66], the authors proposed a new adversarial HS-MS fusion method, which took both the spectral and spatial correlations into account and designed a Spectral Spatial Quality (SSQ) index as the guidance for subsequent adversarial selection processes. It is not a GAN-based HS-MS fusion method, but it used the adversarial between spectral and spatial feature and we believe that it can be easily extended to GAN-based HS-MS fusion method.

There are some different characteristics between HS-MS fusion and Pan-sharpening, which would also be as challenges for fusion algorithms. **First**, HS-MS fusion often faces with multi-temporal problems, since the HS and MS images of the same scenario are often acquired at different times. The difference from multi-temporal images may be a very challenging problem, because it will lead to obvious errors in fusion, but there is still no mechanism to deal with it in most current fusion methods. Especially, when the ground objects are changed, How to use GAN or more auxiliary data to elevate errors in fusion is still an open problem. **Second**, HS-MS fusion often faces with larger spatial resolution differences than Pan-sharpening. The spatial downsampling factor between PAN images and MS images is often 4. However, for HS and MS images, the factor is often much higher than 4. Too large spatial differences will make the ill-posed problem more serious so that it will lead to severe spatial distortions in the fusion results. We believe that more efforts need to be made to solve this problem. The effectiveness of the traditional spatial degradation model may be very limited in reducing spatial distortions. GAN-based methods provide us a new way to solve the problem of large spatial differences in HS-MS image fusion.

B. Spatiotemporal image fusion

Owing to technical and budget limitations, spatial and temporal resolution capabilities are mutually restricted. In many cases, satellites with high-spatial-resolution sensors require long revisit cycles, which provide low temporal resolutions, while satellites employing short-time revisit cycles often provide only low- or middle-spatial resolution images. Therefore, it is often difficult to acquire high-spatiotemporal resolution (HSTR) data from a single satellite.

Spatiotemporal data fusion is a feasible solution for the above-mentioned problem. This technique provides high-spatiotemporal resolution data by fusing high-frequency low-spatial resolution images with low-frequency high-spatial-resolution images. The land-surface reflectance between Landsat Enhanced Thematic Mapper Plus (ETM+) and Moderate Resolution Imaging Spectroradiometer (MODIS) basically

Fig. 10: Two kinds of spatiotemporal data fusion manners. (a) Two-pairs fusion manner. (b) One-pair fusion manner.

satisfied with the complementarity of spatial and temporal resolution. Many studies took Landsat and MODIS data as examples of spatial-temporal image fusion.

Over the past decades, a variety of spatiotemporal data-fusion methods have been developed. The existing spatiotemporal data fusion methods can be categorized into different groups: linear blending methods such as [67] [68] [69], unmixing methods [70–72], Bayesian methods [73] [74], sparse approximation methods [75] [76], and deep-learning methods [77] etc.

Most of these mentioned methods have made some progress in Spatiotemporal fusion. However, it is still an open problem since it is hard to accurately establish the complex relationship between high- and low-resolution images. Current methods often suffer from one of the important limitations [78]: All of them have to make an assumption that some key variables keep on unchanging in the Spatiotemporal images. For example, linear mixture methods assume that land cover type does not change during the data observation period[67, 68, 79], unmixing methods assume that the abundances of the endmember keep unchanging[70–72], sparse approximation methods assume that the dictionary keeps unchanging [75] [76] etc. These unchanging hypotheses make the solving of fusion more feasible but it also introduced some errors or distortion into the results.

Meanwhile, the challenging mainly comes from three important factors [80] [81]: 1) In the temporal dimension, the dramatic uncertainty of land feature change (such as flood, phenology, etc., and land types may also change) is hard to predict by images of adjacent time; 2) In spatial dimension, the huge difference between high-resolution images and low-resolution images, brings great difficulties to reconstruct the detailed textures; 3) For different sensors, there are inevitably systematic errors in imaging process, such as imaging conditions (atmosphere, or solar zenith angle), device differences (spectral response, or modulation transfer function) and so on. Deep learning methods have promoted the development of spatiotemporal RS image fusion. The data-driven nature of deep learning can provide more accurate predictions for spatiotemporal sequence images. However, the above three challenging factors are still existing.

The GAN-based method is also a subbranch of deep learning methods. The contradiction between spatial and temporal resolution can be modeled by the generative adversarial mechanism in GAN. It enhances the ability to find the rela-

tionship between the two different Spatiotemporal sequence images. More and more researchers have made attempts to use GAN in Spatiotemporal fusion. For the fusion of Landsat and MODIS, there are three studies such as CycleGAN-STF [82], STFGAN [83], and GAN-STFM [84]. In CycleGAN-STF [82], the fusion was model as a process of data augmentation and data selection. The authors use Cycle-GAN to generate sufficient simulated images between two times, and then some selection metrics such as entropy are adopted to select the suitable images among the generated image to perform further image fusion. In STFGAN [83], considering the huge spatial resolution gap between Landsat and MODIS imagery, it adopts a two-stage framework to develop an end-to-end image fusion. The two-stage framework by GAN makes image fusion more steady, but the down-sampling operator between the first and second stage is not very necessary and may lead to some conflicts with the up-sampling in the first stage. In GAN-STFM [84], it took the reference images as condition inputs in GAN, which introduces the CGAN and switchable normalization technique into the spatiotemporal fusion problem. It claimed that it can break the time restriction on reference image selection, especially when it is not easy to collect adequate data pairs for image fusion because of the time inconsistency or the bad weather condition.

Similar to spectral-spatial image fusion or other types of fusion, conditional GAN was already introduced into Spatiotemporal image fusion. Furthermore, pixel content loss, feature perception loss, and spectral loss are often combined with the adversarial loss in GAN-based spatial-temporal image fusion, which has almost become a very common technic in the fusion community. We believe that there is still room for improvement for GAN-based methods since most of them did not clearly represent the contradiction of spatiotemporal resolution with the generative and adversarial relationship, which is the nature and advantage of GAN.

In this section, we mainly reviewed some GAN based homogeneous RS data fusions including spatial-spectral and spatiotemporal image fusion. The goal of these type of fusion is mainly in pixel level, as well as resolution enhancement. Therefore, the function of GAN in this family fusions is to reconstruct more detail such as edges and textures. In the view point of inverse problem, GAN play the role of both anti-gradation and regularization, which are also a contradiction with the unity of opposites.

VI. HETEROGENEOUS RS DATA FUSION

Heterogeneous RS Data mainly refers to the data with different imaging mechanism and physical meaning. The goal of heterogeneous RS data fusion is often not to make up for insufficiency of the spatial, temporal or spectral resolutions but object recognition, change detection, classification etc. Most of heterogeneous RS data do not belong to pixel level fusion but for feature level or decision level fusion. However, in this review, heterogeneous RS data are still remote sensing data but not as the relationship between RS and ground observation data. If some of data are not remote sensing data but ground observation data or others, they are categorized into another

type of fusion but not heterogeneous RS data fusion. The most common type heterogeneous RS data are SAR and optical images, point clouds and optical images, and infrared-visible light images. We will discuss the three type of fusion with GAN in the following subsections.

A. SAR and Optical Image Fusion

SAR as a microwave sensor can penetrate through cloud cover, dust, and many bad climatic conditions. The shortcoming is that it does not provide spectral information which is important to many remote sensing applications and its interpretability is influenced by speckle noise. Optical images can provide very abundant spectral information and do not tend to be corrupted by noise. However, optical images are seriously affected by cloud cover, dust, and many bad climatic conditions. It is pointed that: two identically structured objects may appear different in optical imagery due to their spectral responses, which can be identical in SAR imagery [85]. Therefore, SAR and optical imagery can offer complementary information to each other and the fusion of these images can generate an image with both rich spatial structure and spectral informat

Similar to literature survey [85], many traditional SAR-optical image fusion in pixel level can also be classified into four categories namely component substitution methods (CS), multiscale decomposition methods (MSD), hybrid methods, and learning-based methods. Component substitution methods and multiscale decomposition methods almost can also be applied to any fusion scenario such as Pan-sharpening, spatial-temporal or infrared-visible light image fusion. These two types of methods are totally unsupervised and very easy to implement. Learning-based methods can be taken as the pioneer of data-driven methods before the emerging of deep learning. GAN was introduced into SRA-optical image fusion mainly including SAR-to-optical Image Translation, Cloud Removal with SAR-optical fusion, and SAR-optical image registration, etc.

(a) SAR image (b) MS image (c) Fusion image

Fig. 11: The images are from reference [85].

SAR-to-optical image translation is a very promising method to improve the interpretation of the SAR images since the human experts are generally trained by comparing SAR images side-by-side with the corresponding optical images. In the computer vision community, image-to-image translation is very popular, which involves generating a new synthetic version of a given image with a specific modification, such as collection style transfer, object transfiguration, season transfer,

TABLE II: GAN Architectures and Loss Functions for Different Landsat-MODIS fusion Methods.

Method	Structure	l_G^{ADV}	l_G^{PXL1}	l_G^{PXL2}	l_G^{FTR1}	l_G^{CYC}
CycleGAN-STF [82]	Cycle-GAN	✓	✓		✓	✓
STFGAN [83]	Two-Stage	✓		✓		
GAN-STFM [84]	CGAN	✓		✓	✓	

and photo enhancement. Most of these studies are very suitable for SAR-to-optical image translation. Traditional manual feature such as wavelet was already applied to SAR-to-Optical image translation [86]. Except that, CNN architecture can be directly applied to SAR-to-optical image translation, such as in reference [87] where VGG-19 network is used as the feature extractor in the task of cross-modal image translation. GAN-based method is attracting more attention because it usually can generate realistic features of visual perception. The most often used architecture may be CGAN such as [88][89][90][91][92], since both the input and output are images but not noise vectors. Furthermore, similar to pix2pix [93] method, the content loss is necessary for image translation, so that the L_1 norm is always combined with CGAN in this kind of problem. Similarly to other fusion study, image translation also combine CGAN with multi-loss-functions such as L_1 and SSIM [94]. Both CNN and CGAN need to be supervised image pairs in their training, which could be uneasy to acquire in many cases of heterogeneous RS data. In reference [95], the authors investigate the utilization of an adapted version of the CycleGAN architecture for the SAR-to-optical image translation task and use the value of domain knowledge for improving results. In [96], an extension towards unsupervised learning is tested with the CycleGAN loop, where a modified image translation GAN architecture with multiscale cascaded residual connections is proposed. The advantages of CycleGAN and CGAN can also be combined. For example, in [97], a supervised Cycle-Consistent Adversarial Network (S-CycleGAN) was proposed to generate large optical images from the SAR images, where the generated optical images can be alternative data that aid in land cover visual recognition for untrained people [97]. Another very interesting application for SAR-to-optical image translation is multi-temporal data generation or prediction. As in [98], the authors demonstrate that the CGAN with CNN can successfully simulate the multi-temporal optical image with the aid of SAR data.

Cloud Removal with SAR-optical Fusion. Both thin and thick clouds can be removed with SAR-optical fusion method. For thick cloud removal, it is similar to SAR-to-optical image translation, which SAR image will provide a prediction for the missing data. These studies will be reviewed in Section VIII-A: Missing Data Reconstruction. For thin cloud removal, the SAR image is usually used to help to estimate the thickness of the thin clouds, which will be reviewed in Section VIII-B: Thin Cloud Removal.

SAR-optical image registration. Some studies on SAR-optical image registration are also related to the translation of SAR-to-optical images such as deep learning framework for matching of SAR and optical imagery [99] and distribution and structure match GAN for SAR image classification [100]. However, they are indirectly related to GAN for image fusion.

We do not review them in this survey.

Overall, SAR-optical image fusion is often related to feature enhancement, cloud removal, temporal-spatial fusion, image translation, or even multi-modal image registration. They all need to establish the multi-model relationship between SAR and optical images. GAN is playing a very important role in the translation of either pixel distribution or feature correlation for the multi-model data.

B. Point Clouds and Optical Image Fusion

Point clouds and optical images are heterogeneous RS data but they share complementary characteristics, which makes the models with their fusion to be more effective and popular in satellite remote sensing and autonomous driving etc. To be more specific, an optical imaging system usually cannot provide reliable 3D geometry, which is essential for many applications such as autonomous driving. Although stereo cameras can provide 3D geometry, its computational cost is high and was seriously affected by the lighting condition of imaging and the richness of texture features. Point clouds can provide high-precision 3D geometry and is not limited by lighting condition and textureless. However, some point clouds such as LiDARs are limited by low-resolution, low-refresh rates, and severe weather conditions. For many applications, it is worthy to combine these two complementary sensors and it demonstrated significant advantages comparing with a single modal approach. In reference [101], the author made a substantial review on deep learning for image and point cloud fusion in autonomous driving. In [102], it made a review on advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. **Usually, point clouds and optical image fusion includes LiDARs and optical image fusion, stereo vision clouds and optical image fusion, depth clouds and optical image fusion, and InSAR clouds and optical image fusion, etc.** In this survey, we mainly review the GAN-based fusion of point clouds and optical images.

The fusion of point cloud and optical image has been applied to many fields such as depth complete, object detection, semantic segmentation, land use, and land cover classification, etc. This kind of fusion is different from spatial-spectral fusion (homogeneous) or SAR-optical fusion (heterogeneous), whose images are composed of pixels of raster form. On one hand, point clouds are vector images but not raster images, which provide geometry information but not intensity information. On the other hand, it is not necessary a pixel (or voxel) fusion, because its output would be optical image, saliency map, object bound box, segmentation map, or 3D model. Fusion of point cloud and optical image belongs to the fusion of heterogeneous data.

Fig. 12: Four different fusion manners for the point cloud and optical image [103].

There are sufficient studies on many applications of fusion of point cloud and optical image, but for a GAN model, it was only applied to some of its branches up to now. This may be attributed to the characteristics of the generative model or we have not found how to use GAN in some special applications. In [104], GAN was used to translate the input of RGB image into a synthetic representation of the missing one (synthetic depth), and make an improvement to semantic segmentation of building footprints with missing modalities. In reference [103], to compensate for the poor quality of the point clouds, in generator, the proposed method adds the cross-modal guidance from the side-output features of the RGB stream to the decoder network of depth stream. In addition, the discriminator network adaptively fuses the features of double streams using a gated fusion module. They are all good tries for the fusion of point cloud and optical image.

In the early works, the classification of multi-modal fusion strategies has various taxonomic methods. In this survey, as in Fig. 12, we think they are suitable to be categorized into four classes: early fusion, late fusion, intermediate fusion, and hybrid fusion. At the same time, we find that GAN was already introduced into depth completion, saliency map, and semantic segmentation [105], contour extraction [106], image translation [107], [108]. Most of them are belong to low and middle-level vision problems. Many deep learning methods of point cloud and optical image fusion are proposed for high-level vision problems such as object detection, classification, 3D reconstruction [109] and change detection etc., but most of them are not with GAN. This is partly because most detection or classification tasks need feature level fusion but we still do not find a very good way to generate accuracy and steady feature by GAN.

C. Infrared-visible Image Fusion

Fig. 13: These images are from reference [110].

Both infrared images and visible light images are raster images, but they have no spectral intersection. Some studies category them into spectral fusion. In this paper, we category

infrared-visible light image fusion as heterogenous RS fusion because they have no spectral intersection and are with different physical meanings.

Infrared images, which are captured by sensors to record the thermal radiations, are minimally affected by illumination variations and environment disguises. Infrared images can be easily acquired at both daytime and nighttime but they usually lack texture, which is seldom related to the heat emissions. By contrast, visible images contain rich texture information, while visible imaging sensors are susceptible to the environment. The perceptual scene in visible images is similar to the descriptions from human eyes. The visible images may be easily influenced by the external environment, such as disguises, nighttime conditions, smoke, haze, etc [111]. The purpose of infrared-visible light image fusion is to obtain a single complementary fused image that has both rich texture information (from the visible image) and saliency target areas (from the infrared image) [111]. The fused infrared-visible image has many applications such as depth prediction [112], object tracking [113], etc. **There is no spectral intersection between infrared image and visible image. This is the most obvious point different from Pan-sharpening fusion. It also makes infrared-visible image fusion to be different in the structure, loss function, and training or supervising process.**

There are at least three kinds of structures for infrared-visible image fusion with GAN. **The first** type as in Fig. 14 (a), has one generator and one discriminator such as [114] [111] [115]. Their inputs are the concatenated infrared and visible images. Since the ground truth of fusion is un-available, the discriminator has no reference to make a decision on the fusion image. In [111] the discriminator directly compares fused and visible images; In [114], the discriminator compares fused and visible images in feature space; In [115] the discriminator compares the fused image with visible image and infrared image altogether. **The second** type as Fig. 14 (b) has one generator and two discriminators such as [116] [117] [118]. Their inputs are the concatenated infrared and visible images, while one discriminator compares fusion image with visible image and the other compares fusion image with infrared image. **The third** type as Fig. 14 (c) has two generators and two discriminators such as [119] [120]. For this type, the coupled GAN [119] shared some features, while PCSGAN [120] is a cyclic model.

Similar to other types of image fusion with GAN, most of the infrared-visible image fusion methods employed multiple loss functions but not a single loss function. For example, the PCSGAN [120] uses the combination of the perceptual (i.e.,

Fig. 14: Different GAN structures for infrared-visible image fusion. (a) is the architecture with one generator and one discriminator as [114][111][115]. (b) is the architecture with one generator and two discriminators as [116][117][118]. (c) is the coupled GAN architecture as [119][120].

feature-based) losses along with the pixel-wise and adversarial losses. Loss functions for different infrared-visible image fusion methods are summarized in Table III. l_G^{ADV} is the adversarial loss. l_G^{FTR1} is the implicit feature such as hidden layer in CNN. l_G^{FTR2} is the explicit feature such as gradient or local variation. l_G^{PXL1} is the L_1 norm contents loss for pixels. l_G^{PXL2} is the L_2 norm contents loss for pixels. It is worthy to notice that there is no spectral loss function l_G^{SAM} for infrared-visible image fusion. In fact, the meanings of l_G^{PXL1} , l_G^{PXL2} , l_G^{FTR1} and l_G^{FTR2} are also different from that of Pan-sharpening fusion because infrared-visible image fusion aims to resolve the conflict between salience and texture details but not the conflict between spatial resolution and spectral resolution.

TABLE III: Loss Functions for Different Infrared-visible Image Fusion Methods.

Dataset	l_G^{ADV}	l_G^{PXL1}	l_G^{PXL2}	l_G^{FTR1}	l_G^{FTR2}
PCSGAN [120]	✓	✓		✓	
RCGAN [119]	✓		✓		✓
DPAL [111]	✓		✓	✓	✓
FusionGAN [115]	✓		✓		✓
D2WGAN [117]	✓		✓		✓
AttentionFGAN[116]	✓		✓		✓
MgAN-Fuse [118]	✓		✓	✓	✓

About the training process, different from Pan-sharpening fusion, infrared-visible image fusion not only have no ground-truth but also can not be simulated in training. Furthermore, Walds protocol [55] can not be applied to infrared-visible image fusion, so that there is no resolution degradation step for all infrared-visible image fusion. In many cases, both original visible and infrared images are used as supervising information such as [110]. How to provide supervising information sometimes depends on the application scenarios of the fusion. This leads to a very flexible and diversified evaluation for fusion results. The training process often depends on the goal of infrared-visible image fusion such as object detection [113], depth prediction [112], whose training will be different.

In this section, we reviewed three types of heterogenous RS

data fusion: SAR and optical image fusion, point clouds and optical image fusion, and infrared-visible light image fusion. Different from homogeneous RS data, the heterogenous RS data can not be unified into an observation model because they have totally different physical meanings. These characteristics make the heterogenous RS data fusion guided by variety of applications but not resolution enhancement. As result, many heterogenous RS data fusions are very close to feature-level fusion but not pixel-level fusion. Therefore, the function of GAN in this type of fusion often is to produce features which have all of the advantages of different heterogenous RS data.

VII. REMOTE SENSING (RS) AND GROUND OBSERVATION (GO) DATA FUSION

In most situations, Remote Sensing (RS) data and Ground Observation (GO) data both have their own advantages and disadvantages. Remote sensed data from satellites such as multi-spectral, hyperspectral, or radar images, or metrics calculated from airborne laser scanning, are densely sampled data, as well as they are available for all the pixels of the investigated area. RS data are competent to provide observation of large-scale and dense sampling. However, the accuracy of RS data may not very high due to their complicated imaging process. RS data also are limited by their spatial, temporal, and spectral resolution because of the recycling of satellite or sensor manufacturing. In contrast with RS data, the Ground Observation (GO) data are often sparsely and irregularly sampled data, which are with the value only for the sampled portion of the area. GO data are with both very high accuracy and very high temporal frequency, but they are sparse and very difficult to set up densely sampling by GO sensors. There are obvious complementary characteristics between RS data and GO data.

RS and GO data fusion is different from both homogeneous RS data fusion and heterogeneous RS data fusion. The dimension of GO data is often not consistent with RS data. Furthermore, the structures between GO data and RS data are often not matching, which makes them more difficult to fuse. The most common GO data are Physical Geography (PG) and Human Geography (HG) data. We will review the state-of-the-art of RS and GO data fusion and analysis the roles of GAN for the fusion in the following subsections.

A. Remote Sensing (RS) and Physical Geography (PG) Data Fusion

Physical Geography (PG) data include soil, atmosphere, hydrology, temperature, etc. data. The relationship between RS and PG data will determine their mode of fusion. In reference [121], the authors summarized three type relationships: Complementary, Redundant, and Cooperative. 1. Complementary, when data and/or information obtained from distinct sources represent different parts or aspects of a particular issue, so contributing to obtain more exhaustive global information. 2. Redundant, when two or more input sources provide information about the same target (e.g. study area), that can be fused to reduce the uncertainty. 3. Cooperative, when the data and/or information provided by different sources are combined to form exhaustive information about the study target (e.g. derived variables, processes, features, etc.). We believe that most RS and PG data fusion belongs to complementary and redundant relationships and a few of them are cooperative relationships.

For example, direct PM_{2.5} data from ground observation are redundant and complementary for the PM_{2.5} estimating with remote sensing data, while the ground observation data such as surface reflectance, temperature, wind speed, and relative humidity are cooperative in estimating PM_{2.5}. In [122], to deal with the problem of lacking enough ground PM_{2.5} measurements, the authors developed a national-scale geographically weighted regression (GWR) model to estimate daily PM_{2.5} concentrations in China with fused satellite AOD and PG data of the national monitoring network. In [123], to obtain spatially continuous ground-level PM_{2.5} concentrations on national scale, several models established by the point-surface fusion of station measurements and satellite observations have been developed, where a pixel-based merging strategy is proposed and comprehensively evaluated by different experiments. In [124], the authors make attempt to employ deep learning for air temperature mapping mainly based on space remote sensing and ground station observations, in which a 5-layers deep belief network (DBN) is used to search the complicated and non-linear relationships between air temperature and different predictor variables. In [125], the study developed a point-surface collaborative inversion method for the estimation of regional surface soil moisture (SSM) using a generalized regression neural network (GRNN) which is trained on sparse ground-based measurements. We can found that there are very large challenges in the traditional model-driven fusion of RS and GP data, because the atmosphere, hydrology, temperature, etc. data all have their own characteristics and the RS data are often need to inversion in fusion. Deep learning as a data-driven method is attracting much attention in this community and shows promising performances in many studies, which can be a better way to bridge the gap between RS and GP data.

Up to now, we still have not found any published paper that used GAN to perform RS and PG data fusion. The reason may be two points: 1. The structure and physical property of RS and GP data often dramatically differ, so that it is hard to take them as inputs of GAN at the same time; 2. For RS

and GP data, we still do not find a suitable way with GAN to model their relationship such as complementary, redundant, cooperative. However, in this survey, we believe that no matter complementary, redundant, or cooperative, they all can be transformed into the generative and adversarial relationship of the two parts of GAN, which will promote the fusion of RS and PG data in future research.

B. Remote Sensing (RS) and Human Geography Data Fusion

Human Geography data includes communication, social media, location, disease transmission, public sentiment, population distribution, economic, etc. Human Geography (HG) data are totally different from traditional data collection methods such as remote sensing. Their relationships are so complicated that they can not be represented in a uniform formula or structure. HG data, especially social sensing data, are developing very fast in recent years. They enable all citizens to become part of a large sensor network, which is low-cost, more comprehensive, and always broadcasting situational awareness information.

However, data collected with social sensing is often massive, heterogeneous, noisy, unreliable from some aspects, as well as comes in continuous streams, and often lacks geospatial reference information. Together, these issues represent a grand challenge toward fully leveraging social sensing for RS and HG data fusion. Meanwhile, deep learning such as LSTM, CNN, GAN, etc architecture become critical components of fusing social sensing and remote sensing data to understand the human geography problem in a timely fashion.

There is a very wide range of studies for RS and HG data fusion. In [126], the authors summarized some applications of RS-HG data fusion in ecological monitoring, air monitoring, and disaster monitoring, etc. It also pointed out that the fusion of heterogeneous remote sensing and social media data exhibits huge potential for different applications, especially for problems in which (near) real-time response is needed. Very popular and typical applications include population distribution, public sentiment, flood inundation, etc. In [127], the close relationship between the number of Twitter users and brightness of nighttime lights (NTL) over the contiguous United States is calculated and geotagged tweets are then used to upsample a stable light image for 2013. In [128], the authors present and evaluate a method for rapidly estimating flood inundation extent based on a model that fuses remote sensing, social media, and topographic data sources. In [129], using this extensive flooding (2013) of the City of Calgary as a case study, this work illustrates how to fuse authoritative RS and volunteered geographical data to estimate flood extent and identify affected roads during a flood disaster. In [130], it proposes a multi-model fusion neural network for estimating fine-resolution population estimates from multi-source data, which takes into account the local spatial information and global information of each geographic unit by fusion of remote sensing and social sensing data.

Similar to RS and PG fusion, up to now, we have still not found any published paper that used GAN to perform RS and HG data fusion. The reason may be similar to it:

because the structure and physical property for HG data are more complicated and it is also hard to model the relationship between HG and RS data by GAN. However, we still believe GAN architectures are promising in the future studies of HG and RS fusion.

Some of the papers also take RS and GO data as heterogeneous data, while in this survey we use the conception of "heterogeneous RS" data (such as SAR and optical image) but not heterogeneous data. We do not think RS and GO data are homogeneous, but this is a review of RS image fusion. Actually, we think the difference between RS and GO data is more obvious than that between SAR and optical images. Therefore, the representation of RS or GO data becomes the first keyword in this type of fusion. The models or schemes of the fusion of RS and GO data are another very challenging problem. The information between RS and GO data need to be extracted or integrated in a suitable way no matter they are complementary, redundant, or cooperative. Lastly, the way of result validation will also need to be noticed since they are not easy to carry simulating experiments like homogeneous RS data fusion.

VIII. SOME OF APPLICATIONS OF DATA FUSION WITH GAN

There are many applications of data fusion with GAN. For the page limitation, we only address two typical applications that are directly related to image fusion with GAN: missing data reconstruction and thin cloud removal. Usually, there are some overlaps between missing data reconstruction, spatial-temporal image fusion, or cloud removal etc. To be more clear, we give out their relationships as in Fig. 15. We can observe that for some applications they can be realized by both homogeneous and heterogeneous RS image fusion.

A. Missing Data Reconstruction

There are often missing data in remote sensing images because of defective sensors, shadows, thick clouds, etc., so that the value or application of the acquired remote sensing data is greatly reduced. Image fusion is one of the most popular methods to reconstruct the missing data. As in reference [131], the methods of the missing data reconstruction in remote sensing images can be divided into four categories: spatial-based methods, spectral-based methods, temporal-based methods, and hybrid methods. Both homogeneous RS data and heterogeneous data can be fused to reconstruct missing data.

Before introducing GAN into missing data reconstruction, there are already some studies which are based on deep learning such as CNN. For example, in reference [132], a unified spatial-temporal-spectral framework based on a deep convolutional neural network (STS-CNN) was proposed for missing information reconstruction, which can be applied to deadlines in Aqua MODIS, Landsat ETM+ Scan Line Corrector (SLC)-off problem, and thick cloud removal. Missing data reconstruction is not necessary to use image fusion. Without GAN or fusion, it belongs to a simple and direct problem of data reconstruction, such as [133] which translates incomplete data into their corresponding complete data. There are some

overlaps between missing data reconstruction, spatial-temporal fusion, and SAR-optical fusion which are illustrated in Fig. 15.

In remote sensing, the key problem for missing image reconstruction is how to integrate the auxiliary information (spatial, spectral, temporal, etc.) or priors into the reconstruction by current generators. However, different from spatial-spectral fusion or Spatiotemporal fusion, missing data are often large or irregular areas but not common down-sampling. Please refer to the observation Eq.(12), the rank of matrix $\mathcal{H}$ for missing data is often lower than that in other fusion problems. It means that the ill-posedness of the equation is more serious from the viewpoint of solving an inverse problem. Without more auxiliary information from fusion, we usually can only reconstruct small areas.

Implicit and data-driven methods such as CNN or U-net can improve the performance on the large missing area but it still plays a limited role. GAN opens up new prospects for missing data reconstruction. Parts of its ability come from its schema of data generating. In reference [134], the author summary a common model of GAN-based missing data imputation: the generator's goal is to accurately impute missing data, and the discriminator's goal is to distinguish between observed and imputed components. The discriminator is trained to minimize the classification loss, but the generator is trained to maximize the discriminators misclassification rate.

There are substantial studies on missing data reconstruction in the field of computer vision which are known as image inpainting. Inpainting is more like a spatial-based method. In remote sensing applications, more reference information or priors can be utilized in missing data reconstruction. Therefore, there are often two or more inputs in reconstruction, so that it can be taken as an image fusion problem. For simplicity, we category them into small and medium scale, and large-scale missing data reconstruction.

Small and medium scale data missing mainly refers to the data missing phenomena caused by sensor defective [132], orbit gaps [135] [136] and temporal gaps [137]. In most cases, the thick clouds are not small and medium scale data missing but large area data missing. The missing data of voids in mountainous areas by Shuttle Radar Topography Mission (SRTM) are small and medium-scale data missing. The authors in [138] incorporated the shadow geometric constraints into the CGAN. The shadow boundary loss function, shadow ceiling loss function, and shadow entrance curvature loss function are combined with an adversarial loss to guide the generator to well predict the value in void areas. The missing data in comprehensive traffic flow data [139] is small-scale data missing. In [139], the authors proposed a Spatiotemporal Learnable Bidirectional Attention Generative Adversarial Networks (ST-LBAGAN) to implement data fusion for missing traffic data imputation. In this study, the masked reconstruction loss, perceptual loss, discriminative loss, and adversarial loss are combined as a new objective function and optimized to improve the data imputation ability. **In some applications**, the data collection process is likely to generate data with missing, incomplete, or corrupted modalities. In [140], the authors focus on semantic segmentation of building footprints with missing modalities, in which GAN are effectively used to synthesis the

Fig. 15: Missing data reconstruction and thin cloud removal.

Fig. 16: Several examples of missing information in remote sensing data. (a) Shuttle Radar Topography Mission (SRTM) tends to have an abundance of voids in mountainous areas where the elevation measurements are missing [138]. (b) Very high resolution (VHR) satellite and aerial images often suffer from scene occlusion caused by redundant objects [141]. (c) DN value of Landsat ETM+ with the SLC-off problem. (d) Ozone of Aura OMI with the row anomaly problem [131]. (e) Restoration of band 130 of an EO-1 Hyperion dataset [142]. (f) Missing areas caused by thick clouds[143].

missing or incomplete data in the depth map.

Large scale data missing mainly refers to the data missing phenomena caused by clouds [144], shadow and large area occlusion [141]. In reference [141], to deal with VHR images with large-scale missing regions, this paper proposed that the reconstruction process is divided into two connected parts: structure prediction and texture generation. In the first part, one generator predicts the edges of objects in missing regions; then, in the second part, the other generator predicts the textures based on edge structural information from the first part. There are two cascaded generators and one discriminator in this method and it is a spatial-based method without any auxiliary spectral or temporal data. Thick cloud removal is one of the important subbranches of large-scale missing data reconstruction. In the last decades, many approaches have been proposed for the specific task of cloud removal in optical imagery [145]. In parallel to traditional model-driven approaches, data-driven methods based on deep learning have become popular recently. Different from stripe missing or deadline, it is hard to recover without using image fusion because most of the cloud areas are large. Similar as [131], we can also category thick cloud removal into spatial, temporal, spectral, etc. methods. However, if we focus on fusion-based

methods it depends on what type of auxiliary data such as infrared, SAR, or history images.

SAR images are unaffected by clouds. However, the texture and physical meaning in SAR images are different from that of optical images. It is very popular to use deep learning and SAR-optical image fusion to solve cloud removal [145]. In [146], [144] and [147], they all used conditional generative adversarial network which SAR images are conditions. The generators in [146] and [147] are U-net architectures. To avoid the limitation of requiring pixel-wise correspondences between cloudy and noncloudy optical training data, [148] used the architecture of cycle-consistent GAN for cloud removal in all-season Sentinel-2 imagery, which is a global data set. Cycle-consistent GAN can also be used to remove thin clouds [149].

Except for SAR images, Near Infra-Red (NIR) spectrum images can also help to realize cloud removal. In [150], Filmy cloud is removed on satellite imagery with NIR image and conditional generative adversarial nets, where the generator also adapts U-net architecture. Multi-temporal data can also be used in thick cloud removal [151]. In reference [152], the gated convolutional networks were used for cloud removal from bi-temporal remote sensing images, where U-net is used but not in GAN architecture. Multi-temporal fusion can not

generate on-time cloud-free images but mainly visual quality improvements. We can find that CGAN and U-net are the most commonly used architectures in thick cloud removal. CGAN is suitable for multi-inputs in many image fusion problems.

Missing data problem occurs not only in remote sensing but also in many other types of signal or data. In reference [142], the authors make an analysis on theory problems (such as noise and unstable process) about the missing value imputation in multivariate time series. It is not for remote sensing images but can provide some theory supports for Spatiotemporal methods in remote sensing missing data reconstruction.

In missing data reconstruction, we are hard to summarize their GAN structure as Fig. 14 for Pan-sharpening because the inputs and output data may be totally different in gap-filling, shadow removal, cloud removal, etc. Diverse auxiliary reference data (SAR, Optical, NIR) make their GAN architecture with no obvious comparability. However, what we can make sure is that the fusion stage usually happens in the generator and the auxiliary information is often used as a condition of GAN.

B. Thin Cloud Removal

GAN can also be applied to thin cloud removal by image fusion. According to the thickness, clouds can be divided into two classes: thick cloud and thin cloud. A thick cloud almost blocks all information of ground objects, while a thin cloud often appears to be semitransparent. Thin clouds result in the visual effect of a loss of contrast in the subject, as well as the haze effect. Usually, we can reduce the haze effect by data correction for the distorted pixels.

Fig. 17: Hazing image and clear image.

Many studies have been carried out for thin cloud removal. Since both the thickness and the clear image are unknown, it is seriously ill-posed and hard to directly solve. Therefore, more priors are required if we want to find a reasonable solution for thin cloud removal problems. The correlation between the different bands of multispectral images is one of the very important priors. The haze optimized transformation (HOT) [153] method employed the linear relationship between red band and blue band. HOT method was successfully applied to Landsat data and MODIS data. However, for other sensors, there may not be a very clear linear relationship between

bands. Thin clouds are mainly generated by the atmospheric scattering of large particles and the haze effect is usually assumed to locate in the low frequency of a cloudy image. Therefore, it is possible to remove thin clouds by a high pass filter, such as homomorphic filter [154]. Dark channel [155] is also a very effective prior. It shows that at least one channel has some very low-intensity pixels in most non-haze patches. With dark channel prior, the thickness of haze can be estimated and the equation for the atmospheric scattering model will be solved. Dark Channel Prior has been successfully applied to the images of consumer electronic cameras. However, for remote sensing images, it has many limitations.

There are some differences between natural image dehazing (NID) and Remote Sensing Image Dehazing (RSID). The most important difference comes from their way of imaging. First, the natural image is usually captured with a strong three-dimensional perspective as illustrated in Fig. 18 (a), where haze depth mainly depends on its scene depth. The changing of scene depth is usually obvious from targets to backgrounds. However, remote sensing images are usually a top view captured by satellite far from the earth as illustrated in Fig. 18 (b), where its haze depth mainly depends on the thickness of the thin cloud. Second, the contents or targets in their view of the field are very different. There are usually haze-opaque regions such as the sky in natural images but there is usually no sky region in remote sensing images. There are usually haze-free regions in remote sensing images but there are often no haze-free regions in natural images. Their haze distribution characteristics are different, so that it leads to different priors in NID and RSID. Although many studies for NID have been applied to RSID, they often cannot directly and efficiently remove the hazes of remote sensing images because of the aforementioned problems.

More recently, deep learning based dehazing has been developed extensively such as DehazeNet [156], MS-CNN [157], AOD-Net [158], Ranking-CNN [159], and coding of contours and colors [160]. At the same time, many new attempts also have been made in dehazing studies such as attention [161] [162], weakly supervised learning [163], self-filtering [164], image decomposition [165], and airlight component etc [166]. Most of them are data-driven methods for natural images, which do not rely too much on the physical model of atmosphere transmission. They directly learn transmission maps or haze-free images based on given databases that consist of image pairs (clean image and hazy image). However, for both natural images and remote sensing images, ground-truth images are hard to obtain. For natural images, the difficulties mainly come from accurate scene depth measurement. For remote sensing images, we can usually simulate the thin cloud scenes with the re-visit images from satellites. Although some models of natural image dehazing methods have been successfully applied to remote sensing images, the specific distribution characteristic of the thin clouds in remote sensing images is always reducing the performances of these methods. The NID may be well applied to some UAV-based railway images [167] that are similar to natural images. It is clearly pointed out that, in reference [168], the difference between NID and RSID should be well addressed if we want to extend

Fig. 18: (a) In a nature image, the extent of hazes is related to the depth in perspective imaging. (b) In a remote sensing image, the thickness of thin clouds is mainly caused by atmosphere distortion.

the NID methods to RSID for the best performance.

At the same time, more and more attention has been paid to dehazing with GAN. For natural images, cGAN-Dehaze [169] utilized CGAN and DCPDN constructs a generator by combining a densely connected pyramid network with U-net structure. To avoid the requirement of a paired input for hazy image and clear image, some methods such as [170] [171] and [172] proposed to use cycle-GAN to train an unpaired fashion with clear and hazy images altogether. They improve the flexibility in the training of dehazing but they are also subject to distortion of detail texture because of the weakly supervised model in cycle-GAN

For remote sensing images, GAN is also becoming important in the study of thin cloud removal. The cycle-GAN is also applied to the removal for sentinel-2 imagery [150], which not only rejects the necessity of any paired (cloud/cloud-free) training dataset but also avoids the need for any additional (expensive) spectral source. However, similar to [172], the shortcoming of cycle-GAN dehazing still existed. In practice, there are more priors that can be used in thin cloud removal of remote sensing images. Similar to missing data reconstruction, at least there are often more temporal and spectral images such as in [149], or even atmosphere physical model [173] that can provide aid for thin cloud removal.

Different from spatial-spectral or spatial-temporal fusion, thin cloud removal does not enhance the spatial or temporal resolution of images; different from missing data reconstruction, thin cloud removal does not recover the missing data. We believe that cloud removal may be easier than resolution enhancement and missing data reconstruction. Similar to missing data reconstruction, fusion is a very important way to thin cloud removal but it is not the only way. Most dehazing methods for natural images are not fusion-based methods. Many GAN architectures in spatial-temporal-spectral fusion or missing data reconstruction can be directly transferred to thin cloud removal of remote sensing images. The main difficulty for thin cloud removal is from the uncertainty distribution of the thickness of thin cloud, which is both similar to and different from the scene depth in natural images. GAN method may be a new effective way to estimate the thickness of the thin clouds in remote sensing images.

IX. RELATED RESEARCH

A. Super-Resolution

Some studies also take super-resolution as a special case of image fusion, who is especially very similar to homogeneous RS data fusion. However, super-resolution mainly focus on blurring, down-sampling, or noising problem. Both single frame or multi-frame super-resolution are almost all based on the images from the same kind of sensors. Therefore, there are still some differences between super-resolution and RS image fusion. In computer vision community, research on super-resolution has achieved much progress in recent years. Both CNN and GAN architectures have been successfully applied to super-resolution. Many of them can be introduced into RS image fusion. For the page of limitation, we do not address super-resolution in detail. The review for super-resolution can be referred to [174] [175]. The review for GAN based super-resolution can be referred to [176] [177].

B. RS Data Assimilation

Data assimilation is also related research of RS data fusion. Data assimilation is a mathematical discipline that seeks to optimally combined information with different observations. Satellite remote sensing is based on a wide range of platforms and provide various type of observations on large scales and now covering more than decades of measurements. RS data assimilation approaches open possibilities to further exploit satellite observations to serve for modeling the earth system (land, urban surfaces, ocean, or sea-ice).

Both RS data fusion and RS data assimilation need to combine different observation data. However, most of homogeneous and heterogeneous RS data fusion never directly serve for earth system. Even RS and Ground Observation (GO) data fusion usually only pursue the goal of enhancing observation data or providing more features by machine learning, image analysis, and statistical methods. Many RS data assimilations are to minimize modeling uncertainties but not generate more information from different sources of data.

Deep learning has been applied to data assimilation, but we did not find any studies that use GAN in RS data assimilation up to now. It may need more time for us to find how to take advantage of GAN in the problem of data assimilation.

X. DISCUSSION

A. Model-driven, Data-driven and Knowledge-driven

Most GAN-based RS data fusion belongs to data-driven methods, which are different from model-driven methods in nature. For example, in the early studies, many homogeneous data fusions are pixel-level fusion, which can be summarized as solving an observation equation. They are explicit models, as well as model-driven methods. If the fusion process can be modeled as an observation equation, it can be understood as an inverse problem. The characteristic of most inverse problems is that they are ill-posed and their solving processes mainly center around three fundamental problems: observation equation, regularization scheme, and regularization parameter. However, in the family of data-driven methods, there are no obvious boundaries between the three fundamental problems. In CNN-based fusion methods, observation equation, regularization scheme, and regularization parameter have no explicit forms but all by hidden layers of implicit form. Apparently, GAN-based RS image fusion belongs to data-driven methods.

In data-driven fusion methods, the ill-posed problem is solved by feature learning with large data sets. The major advantage is that they do not rely too much on the experience of humans to construct features. The disadvantage is that they need large data sets and power computation abilities. At the same time, model-driven has very clear psychical meaning, which may be very important to some fusion problems. There is already some research integrating model- and data-driven for fusion such as [178]. Furthermore, data-driven often can not be directly applied to some problem. In this case, we may need to convert data into knowledge. Knowledge-driven fusion will be a very important direction following model-driven and data-driven.

B. How to convert the multi-source relationship into generative and adversarial relationship

In reference [121], the authors summarized three type relationships: complementary, redundant, and cooperative. For example: infrared and visible light image fusion belongs to complementary type relationship; Most spatiotemporal image fusion also belongs to complementary type relationship; The fusion of MS and PAN belongs to redundant type. Most point clouds and optical image fusion belong to the cooperative type. All of them face the problem of what should be generative relation and what should be adversarial relation. It is easy to understand the function of generative networks that are responsible to establish the relationship between multi-source inputs and the target outputs. However, it is still hard to clearly address the function of the adversarial scheme in GAN for fusion, because sometimes there are no obvious adversarial relationships in complementary, redundant, and cooperative parts. There do have some unity of opposites in RS image fusion. For example, when we take a common type of pixel-level fusion as an ill-posed problem, their anti-degradation and regularization are a pair of contradictions. But they are explicit contradictions, which need to be transformed if we want to model them by implicit GAN. Therefore, in RS image

fusion, the function of the adversarial architecture needs to be researched further in future research.

C. What role does the discriminator play in the GAN based image fusion?

Most of the GAN for RS image fusion has a very complex generator but a relative sample discriminator. It is a strange phenomenon since generator and discriminator should be similar to the two sides of a coin. Why do we need a very strong generator and a very weak discriminator? One explanation is that generate a fusion image just be more difficult than discriminate against a fusion image. Another explanation is that we still do not know the nature about how to discriminate a real fusion image from a fake fusion image. It is worthy to re-thinking that what role the discriminator plays in the GAN-based image fusion. Maybe we need to open a new way of using GAN for RS image fusion in the future, which is based on totally different understanding of the principle and mechanism of GAN.

XI. CONCLUSION

In this review, we have performed a comprehensive analysis on how to utilize GAN for RS data fusion. We mainly analyzed different characteristics of homogeneous RS data, heterogeneous RS data, and RS and ground observation data, and systematically review how to use GAN in these types of fusions. We point out that in many methods the architecture, loss function, and training scheme of GAN are all reformed to adaptive to RS data fusion. We also discussed some open problems in GAN for fusion such as data-driven or model-driven, the relationship of generative and adversarial parts, and the special role of a discriminator, etc. We believe that the classification and systematization of these methods of RS data fusion with GAN can help the reader to better understand and use them.

XII. ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (Nos. 41971397 and 61731022)

REFERENCES

- [1] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 4, pp. 6–23, 2016.
- [2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.
- [3] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.
- [4] Y. Hong, U. Hwang, J. Yoo, and S. Yoon, "How generative adversarial networks and their variants work:

- An overview,” *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–41, 2019.
- [5] Z. Wang, Q. She, and T. E. Ward, “Generative adversarial networks in computer vision: A survey and taxonomy,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.
 - [6] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, “A review on generative adversarial networks: Algorithms, theory, and applications,” *arXiv preprint arXiv:2001.06937*, 2020.
 - [7] S. W. Park, J. S. Ko, J. H. Huh, and J. C. Kim, “Review on generative adversarial networks: Focusing on computer vision and its applications,” *Electronics (Switzerland)*, vol. 10, no. 10, pp. 1–40, 2021.
 - [8] Y.-J. Cao, L.-L. Jia, Y.-X. Chen, N. Lin, C. Yang, B. Zhang, Z. Liu, X.-X. Li, and H.-H. Dai, “Recent advances of generative adversarial networks in computer vision,” *IEEE Access*, vol. 7, pp. 14 985–15 006, 2019.
 - [9] —, “Recent advances of generative adversarial networks in computer vision,” *IEEE Access*, vol. 7, pp. 14 985–15 006, 2018.
 - [10] V. Sorin, Y. Barash, E. Konen, and E. Klang, “Creating artificial images for radiology applications using generative adversarial networks (gans)—a systematic review,” *Academic radiology*, vol. 27, no. 8, pp. 1175–1185, 2020.
 - [11] A. Yadav and D. K. Vishwakarma, “Recent developments in generative adversarial networks: A review (workshop paper),” in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2020, pp. 404–413.
 - [12] L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, “A state-of-the-art review on image synthesis with generative adversarial networks,” *IEEE Access*, vol. 8, pp. 63 514–63 537, 2020.
 - [13] E. Denton, S. Chintala, A. Szlam, and R. Fergus, “Deep generative image models using a laplacian pyramid of adversarial networks,” *arXiv preprint arXiv:1506.05751*, 2015.
 - [14] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, “Least squares generative adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.
 - [15] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.
 - [16] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, “Adversarial autoencoders,” *arXiv preprint arXiv:1511.05644*, 2015.
 - [17] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier gans,” in *International conference on machine learning*. PMLR, 2017, pp. 2642–2651.
 - [18] X. Chen, Y. Duan, R. Houthoofd, J. Schulman, I. Sutskever, and P. Abbeel, “Infogan: Interpretable representation learning by information maximizing generative adversarial nets,” in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 2180–2188.
 - [19] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.
 - [20] D. Berthelot, T. Schumm, and L. Metz, “Began: Boundary equilibrium generative adversarial networks,” 2017.
 - [21] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*. PMLR, 2017, pp. 214–223.
 - [22] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein gans,” *arXiv preprint arXiv:1704.00028*, 2017.
 - [23] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434*, 2015.
 - [24] J. Donahue, P. Krähenbühl, and T. Darrell, “Adversarial feature learning,” *arXiv preprint arXiv:1605.09782*, 2016.
 - [25] A. Odena, “Semi-supervised learning with generative adversarial networks,” *arXiv preprint arXiv:1606.01583*, 2016.
 - [26] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
 - [27] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. [Online]. Available: <https://openreview.net/forum?id=Hk99zCeAb>
 - [28] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
 - [29] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-attention generative adversarial networks,” in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.
 - [30] A. Brock, J. Donahue, and K. Simonyan, “Large scale gan training for high fidelity natural image synthesis,” *arXiv preprint arXiv:1809.11096*, 2018.
 - [31] T. R. Shaham, T. Dekel, and T. Michaeli, “Singan: Learning a generative model from a single natural image,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4570–4580.
 - [32] G. Daras, A. Odena, H. Zhang, and A. G. Dimakis, “Your local gan: Designing two dimensional local attention mechanisms for generative models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 531–14 539.
 - [33] J. J. Zhao, M. Mathieu, and Y. LeCun, “Energy-based generative adversarial networks,” in *5th International*

- Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. [Online]. Available: <https://openreview.net/forum?id=ryh9pmcee>
- [34] R. Wang, A. Cully, H. J. Chang, and Y. Demir, "MAGAN: margin adaptation for generative adversarial networks," *CoRR*, vol. abs/1704.03817, 2017. [Online]. Available: <http://arxiv.org/abs/1704.03817>
- [35] X. Huang, Y. Li, O. Poursaeed, J. E. Hopcroft, and S. J. Belongie, "Stacked generative adversarial networks," *CoRR*, vol. abs/1612.04357, 2016. [Online]. Available: <http://arxiv.org/abs/1612.04357>
- [36] F. Juefei-Xu, V. N. Boddeti, and M. Savvides, "Gang of gans: Generative adversarial networks with maximum margin ranking," *CoRR*, vol. abs/1704.04865, 2017. [Online]. Available: <http://arxiv.org/abs/1704.04865>
- [37] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.
- [38] Z. Tan, L. Di, M. Zhang, L. Guo, and M. Gao, "An enhanced deep convolutional model for spatiotemporal image fusion," *Remote Sensing*, vol. 11, no. 24, p. 2898, 2019.
- [39] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [40] Z. Tan, M. Gao, X. Li, and L. Jiang, "A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [41] 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*, 2017, pp. 4681–4690.
- [42] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.
- [43] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018.
- [44] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatio-temporal-spectral fusion of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7135–7148, 2016.
- [45] L. Zhang and H. Shen, "," *Yaogan Xuebao/Journal of Remote Sensing*, vol. 20, no. 5, pp. 1050–1061, 2016.
- [46] H. Zhang and J. Ma, "Gtp-pnet: A residual learning network based on gradient transformation prior for pansharpening," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 172, pp. 223 – 239, 2021.
- [47] M. Jiang, H. Shen, J. Li, Q. Yuan, and L. Zhang, "A differential information residual convolutional neural network for pansharpening," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 257 – 271, 2020.
- [48] Q. Liu, H. Zhou, Q. Xu, X. Liu, and Y. Wang, "Psgan: A generative adversarial network for remote sensing image pan-sharpening," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2020.
- [49] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "Hpgan: Hyperspectral pansharpening using 3-d generative adversarial networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 463–477, 2021.
- [50] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-gan: An unsupervised pan-sharpening method for remote sensing image fusion," *Information Fusion*, vol. 62, pp. 110 – 120, 2020.
- [51] F. Ozcelik, U. Alganci, E. Sertel, and G. Unal, "Rethinking cnn-based pansharpening: Guided colorization of panchromatic images via gans," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2020.
- [52] C. Zhou, J. Zhang, J. Liu, C. Zhang, R. Fei, and S. Xu, "Perceppean: Towards unsupervised pansharpening based on perceptual loss," *Remote Sensing*, vol. 12, no. 14, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/14/2318>
- [53] Z. Shao, Z. Lu, M. Ran, L. Fang, J. Zhou, and Y. Zhang, "Residual encoderdecoder conditional generative adversarial network for pansharpening," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1573–1577, 2020.
- [54] J. Chen, Y. Pan, and Y. Chen, "Remote sensing image fusion based on Bayesian GAN," 2020. [Online]. Available: <http://arxiv.org/abs/2009.09465>
- [55] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolution: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sensing*, pp. 691–699, 1997.
- [56] F. Dadrass Javan, F. Samadzadegan, S. Mehravar, A. Toosi, R. Khatami, and A. Stein, "A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 101 – 117, 2021.
- [57] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.
- [58] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 69, pp. 40–51, 2021. [Online]. Available: <https://doi.org/10.1016/j.inffus.2020.11.001>
- [59] M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," *Information Fusion*, vol. 59, pp. 59 – 83, 2020.

- [60] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral Pansharpening Using 3-D Generative Adversarial Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 463–477, 2021.
- [61] W. Dong, Y. Yang, and J. Qu, "Fusion of Hyperspectral and Panchromatic Images Using Generative Adversarial Network," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2021.
- [62] Q. Wang, L. Wang, Z. Li, X. Tong, and P. M. Atkinson, "Spatial-Spectral Radial Basis Function-Based Interpolation for Landsat ETM+ SLC-Off Image Gap Filling," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–17, 2020.
- [63] K. G. Lore, K. K. Reddy, M. Giering, and E. A. Bernal, "Generative adversarial networks for spectral super-resolution and bidirectional rgb-to-multispectral mapping," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019-June, pp. 926–933, 2019.
- [64] X. Luo, X. Tong, and Z. Hu, "Improving Satellite Image Fusion via Generative Adversarial Training," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2020.
- [65] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Single Sensor Image Fusion Using A Deep Convolutional Generative Adversarial Network," *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, vol. 2018-Septe, 2018.
- [66] X. Luo, J. Yin, X. Luo, and X. Jia, "A Novel Adversarial Based Hyperspectral and Multispectral Image Fusion," *Remote Sensing*, vol. 11, no. 5, p. 492, Feb. 2019.
- [67] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2207–2218, aug 2006.
- [68] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sensing of Environment*, vol. 114, no. 11, pp. 2610–2623, 2010.
- [69] Q. Wang, Y. Zhang, A. O. Onojeghuo, X. Zhu, and P. M. Atkinson, "Enhancing Spatio-Temporal Fusion of MODIS and Landsat Data by Incorporating 250 m MODIS Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4116–4123, sep 2017.
- [70] Z. Niu, "Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model," *Journal of Applied Remote Sensing*, vol. 6, no. 1, p. 063507, mar 2012. [Online]. Available: <http://remotesensing.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JRS.6.063507>
- [71] R. Zurita-Milla, J. G. Clevers, and M. E. Schaepman, "Unmixing-based landsat TM and MERIS FR data fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 453–457, jul 2008.
- [72] J. Amorós-López, L. Gómez-Chova, L. Alonso, L. Guanter, R. Zurita-Milla, J. Moreno, and G. Camps-Valls, "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, no. 1, pp. 132–141, 2013.
- [73] H. Shen, X. Meng, and L. Zhang, "An Integrated Framework for the Spatio-Temporal-Spectral Fusion of Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7135–7148, dec 2016.
- [74] J. Xue, Y. Leung, and T. Fung, "A bayesian data fusion approach to spatio-temporal fusion of remotely sensed images," *Remote Sensing*, vol. 9, no. 12, p. 1310, dec 2017.
- [75] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10 PART1, pp. 3707–3716, 2012.
- [76] J. Wei, L. Wang, P. Liu, X. Chen, W. Li, and A. Y. Zomaya, "Spatiotemporal Fusion of MODIS and Landsat-7 Reflectance Images via Compressed Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7126–7139, dec 2017.
- [77] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal Satellite Image Fusion Using Deep Convolutional Neural Networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 821–829, 2018.
- [78] X. Li, F. Ling, G. M. Foody, Y. Ge, Y. Zhang, and Y. Du, "Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps," *Remote Sensing of Environment*, vol. 196, pp. 293–311, jul 2017.
- [79] T. Hilker, M. A. Wulder, N. C. Coops, J. Linke, G. McDermaid, J. G. Masek, F. Gao, and J. C. White, "A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sensing of Environment*, vol. 113, no. 8, pp. 1613–1627, aug 2009.
- [80] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, "Spatio-temporal fusion for remote sensing data: an overview and new benchmark," *Science China Information Sciences*, vol. 63, no. 4, p. 140301, apr 2020. [Online]. Available: <http://link.springer.com/10.1007/s11432-019-2785-y>
- [81] J. L. L. W. L. Z. J. L. Bingze Song, Peng Liu, "Mlff-gan: A multi-level feature fusion with gan for spatiotemporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [82] J. Chen, L. Wang, R. Feng, P. Liu, W. Han, and X. Chen, "CycleGAN-STF: Spatiotemporal fusion via cycleGAN-based image generation," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.
- [83] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative

- adversarial network,” *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2020.
- [84] Z. Tan, M. Gao, X. Li, and L. Jiang, “A Flexible Reference-Insensitive Spatiotemporal Fusion Model for Remote Sensing Images Using Conditional Generative Adversarial Network,” no. 1, pp. 1–13, 2021.
- [85] S. C. Kulkarni and P. P. Rege, “Pixel level fusion techniques for SAR and optical images : A review,” *Information Fusion*, vol. 59, no. July 2019, pp. 13–29, 2020.
- [86] W. Zhang and M. Xu, “Translate SAR Data into Optical Image Using IHS and Wavelet Transform Integrated Fusion,” *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 1, pp. 125–137, 2019. [Online]. Available: <https://doi.org/10.1007/s12524-018-0879-7>
- [87] J. Zhang, J. Zhou, and X. Lu, “Feature-guided SAR-to-optical image translation,” *IEEE Access*, vol. 8, pp. 70 925–70 937, 2020.
- [88] K. Enomoto, K. Sakurada, W. Wang, N. Kawaguchi, M. Matsuoka, and R. Nakamura, “Image translation between SAR and optical imagery with generative adversarial nets,” *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2018-July, pp. 1752–1755, 2018.
- [89] N. Merkle, S. Auer, R. Muller, and P. Reinartz, “Exploring the potential of conditional adversarial networks for optical and SAR Image Matching,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 6, pp. 1811–1820, 2018.
- [90] M. F. Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, “SAR-to-optical image translation based on conditional generative adversarial networks-optimization, opportunities and limits,” *Remote Sensing*, vol. 11, no. 17, pp. 1–19, 2019.
- [91] J. N. Turnes, J. D. B. Castro, D. L. Torres, P. J. S. Vega, R. Q. Feitosa, and P. N. Happ, “Atrous cGAN for SAR to Optical Image Translation,” *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.
- [92] M. Schmitt, L. H. Hughes, and X. X. Zhu, “The sen1-2 dataset for deep learning in sar-optical data fusion,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-1, pp. 141–146, 2018.
- [93] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [94] Y. Li, R. Fu, X. Meng, W. Jin, and F. Shao, “A SAR-to-Optical Image Translation Method Based on Conditional Generation Adversarial Network (cGAN),” *IEEE Access*, vol. 8, pp. 60 338–60 343, 2020.
- [95] M. F. Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt, “SAR-to-optical image translation based on conditional generative adversarial networks-optimization, opportunities and limits,” *Remote Sensing*, vol. 11, no. 17, pp. 1–19, 2019.
- [96] S. Fu, F. Xu, and Y. Q. Jin, “Reciprocal translation between sar and optical remote sensing images with cascaded-residual adversarial networks,” *arXiv*, 2019.
- [97] L. Wang, X. Xu, Y. Yu, R. Yang, R. Gui, Z. Xu, and F. Pu, “SAR-to-optical image translation using supervised cycle-consistent adversarial networks,” *IEEE Access*, vol. 7, pp. 129 136–129 149, 2019.
- [98] W. He and N. Yokoya, “Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation,” pp. 1–11, 2018.
- [99] L. Haydn, D. Marcos, S. Lobry, D. Tuia, and M. Schmitt, “ISPRS Journal of Photogrammetry and Remote Sensing A deep learning framework for matching of SAR and optical imagery,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, no. September, pp. 166–179, 2020. [Online]. Available: <https://doi.org/10.1016/j.isprsjprs.2020.09.012>
- [100] Z. Ren, B. Hou, Q. Wu, Z. Wen, and L. Jiao, “A distribution and structure match generative adversarial network for sar image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3864–3880, 2020.
- [101] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, and D. Cao, “Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review,” *arXiv*, pp. 1–19, 2020.
- [102] J. Zhang and X. Lin, “Advances in fusion of optical imagery and lidar point cloud applied to photogrammetry and remote sensing,” *International Journal of Image and Data Fusion*, vol. 8, no. 1, pp. 1–31, 2017.
- [103] Z. Liu, W. Zhang, and P. Zhao, “A cross-modal adaptive gated fusion generative adversarial network for rgb-d salient object detection,” *Neurocomputing*, vol. 387, pp. 210 – 220, 2020.
- [104] B. Bischke, P. Helber, F. Koenig, D. Borth, and A. Dengel, “Overcoming missing and incomplete modalities with generative adversarial networks for building footprint segmentation,” no. 2, 2018.
- [105] R. Zhang, G. Li, M. Li, and L. Wang, “Fusion of images and point clouds for the semantic segmentation of large-scale 3d scenes based on deep learning,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 143, pp. 85 – 96, 2018, iSPRS Journal of Photogrammetry and Remote Sensing Theme Issue Point Cloud Processing.
- [106] W. Zhang, L. Chen, Z. Xiong, Y. Zang, J. Li, and L. Zhao, “Large-scale point cloud contour extraction via 3d guided multi-conditional generative adversarial network,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 97 – 105, 2020.
- [107] S. Milz, M. Simon, K. Fischer, and M. Ppperl, “Points2pix: 3d point-cloud to image translation using conditional generative adversarial networks,” 2019.
- [108] V. Lekic and Z. Babic, “Automotive radar and camera fusion using Generative Adversarial Networks,” *Computer Vision and Image Understanding*, vol. 184, pp. 1–8, 2019. [Online]. Available: <https://doi.org/10.1016/j.cviu.2019.04.002>
- [109] N. Qin, X. Hu, and H. Dai, “Deep fusion of multi-view and multimodal representation of als point cloud

- for 3d terrain scene recognition,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 143, pp. 205 – 212, 2018, iSPRS Journal of Photogrammetry and Remote Sensing Theme Issue Point Cloud Processing.
- [110] J. Ma, H. Zhang, Z. Shao, P. Liang, and H. Xu, “GANMcC: A Generative Adversarial Network with Multiclassification Constraints for Infrared and Visible Image Fusion,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021.
- [111] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, J. Wu, and J. Jiang, “Infrared and visible image fusion via detail preserving adversarial learning,” *Information Fusion*, vol. 54, pp. 85 – 98, 2020.
- [112] Y. Li, H. Zhao, Z. Hu, Q. Wang, and Y. Chen, “Iv-fusenet: Fusion of infrared and visible light images for depth prediction,” *Information Fusion*, vol. 58, pp. 1 – 12, 2020.
- [113] X. Zhang, P. Ye, H. Leung, K. Gong, and G. Xiao, “Object fusion tracking based on visible and infrared images: A comprehensive review,” *Information Fusion*, vol. 63, pp. 166 – 187, 2020.
- [114] B. Liao, Y. Du, and X. Yin, “Fusion of infrared-visible images in ue-iot for fault point detection based on gan,” *IEEE Access*, vol. 8, pp. 79 754–79 763, 2020.
- [115] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, “Fusion-gan: A generative adversarial network for infrared and visible image fusion,” *Information Fusion*, vol. 48, pp. 11 – 26, 2019.
- [116] J. Li, H. Huo, C. Li, R. Wang, and Q. Feng, “AttentionFGAN: Infrared and Visible Image Fusion using Attention-based Generative Adversarial Networks,” *IEEE Transactions on Multimedia*, vol. 14, no. 8, pp. 1–1, 2020.
- [117] J. Li, H. Huo, K. Liu, and C. Li, “Infrared and visible image fusion using dual discriminators generative adversarial networks with wasserstein distance,” *Information Sciences*, vol. 529, pp. 28 – 41, 2020.
- [118] J. Li, H. Huo, C. Li, R. Wang, C. Sui, and Z. Liu, “Multigrained Attention Network for Infrared and Visible Image Fusion,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021.
- [119] Q. Li, L. Lu, Z. Li, W. Wu, Z. Liu, G. Jeon, and X. Yang, “Coupled gan with relativistic discriminators for infrared and visible images fusion,” *IEEE Sensors Journal*, pp. 1–1, 2019.
- [120] K. K. Babu and S. R. Dubey, “Pcsgan: Perceptual cyclic-synthesized generative adversarial networks for thermal and nir to visible image transformation,” *Neurocomputing*, vol. 413, pp. 41 – 50, 2020.
- [121] D. D. Curzio, S. Fountas, M. Romić, and R. A. V. Rossel, “Multi-source data fusion of big spatial-temporal data in soil , geo-engineering and environmental studies,” *Science of the Total Environment*, vol. 788, 2021.
- [122] X. Hu, L. Huang, J. Bi, and Y. Liu, “Estimating Ground-Level PM 2.5 in China Using Satellite Remote Sensing,” *ENVIRONMENTAL SCIENCE and TECHNOLOGY*, vol. 48, pp. 7436–7444, 2014.
- [123] T. Li, H. Shen, C. Zeng, Q. Yuan, and L. Zhang, “Point-surface fusion of station measurements and satellite observations for mapping PM 2.5 distribution in China : Methods and assessment,” *Atmospheric Environment*, vol. 152, pp. 477–489, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.atmosenv.2017.01.004>
- [124] H. Shen, Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang, “Remote Sensing of Environment Deep learning-based air temperature mapping by fusing remote sensing , station , simulation and socioeconomic data,” *Remote Sensing of Environment*, vol. 240, no. November 2019, p. 111692, 2020. [Online]. Available: <https://doi.org/10.1016/j.rse.2020.111692>
- [125] Q. Yuan, H. Xu, T. Li, H. Shen, and L. Zhang, “Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental U . S,” *Journal of Hydrology*, vol. 580, no. November 2019, p. 124351, 2020. [Online]. Available: <https://doi.org/10.1016/j.jhydrol.2019.124351>
- [126] J. Li, Z. Liu, X. Lei, and L. Wang, “Distributed Fusion of Heterogeneous Remote Sensing and Social Media Data: A Review and New Developments,” *Proceedings of the IEEE*, pp. 1–14, 2021.
- [127] N. Zhao, W. Zhang, Y. Liu, E. L. Samson, Y. Chen, and G. Cao, “Improving Nighttime Light Imagery With Location-Based Social Media Data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2161–2172, 2019.
- [128] J. F. Rosser, D. G. Leibovici, and M. J. Jackson, “Rapid flood inundation mapping using social media, remote sensing and topographic data,” *Natural Hazards*, vol. 87, no. 1, pp. 103–120, 2017.
- [129] E. Schnebele, G. Cervone, S. Kumar, and N. Waters, “Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data,” *water*, pp. 381–398, 2014.
- [130] L. Cheng, L. Wang, R. Feng, and J. Yan, “Remote sensing and social sensing data fusion for fine-resolution population mapping with a multi-model neural network,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 8, pp. 1–1, 2021.
- [131] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, “Missing Information Reconstruction of Remote Sensing Data: A Technical Review,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, 2015.
- [132] 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.
- [133] X. Chai, G. Tang, S. Wang, R. Peng, W. Chen, and J. Li, “Deep Learning for Regularly Missing Data Reconstruction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4406–4423, 2020.

- [134] J. Yoon, J. Jordon, and M. Van Der Schaar, "Supplementary materials GAIN: Missing data imputation using generative adversarial nets," *35th International Conference on Machine Learning, ICML 2018*, vol. 13, pp. 9052–9059, 2018.
- [135] X. Liu and M. Wang, "Gap filling of missing data for VIIRS global ocean color products using the DINEOF Method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4464–4476, 2018.
- [136] R. Yao, L. Wang, X. Huang, S. Member, L. Sun, R. Chen, and X. Wu, "A Robust Method for Filling the Gaps in MODIS and VIIRS Land Surface Temperature Data," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2021.
- [137] A. Pepe, A. B. Ortiz, P. R. Lundgren, P. A. Rosen, and R. Lanari, "The stripmap-scansar sbas approach to fill gaps in stripmap deformation time series with ScanSAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12 PART 1, pp. 4788–4804, 2011.
- [138] G. Dong, W. Huang, W. A. Smith, and P. Ren, "A shadow constrained conditional generative adversarial net for SRTM data restoration," *Remote Sensing of Environment*, vol. 237, no. December 2019, 2020.
- [139] B. Yang, Y. Kang, Y. Yuan, X. Huang, and H. Li, "ST-LBAGAN: Spatio-temporal learnable bidirectional attention generative adversarial networks for missing traffic data imputation," *Knowledge-Based Systems*, vol. 215, p. 106705, 2021.
- [140] B. Bischke, P. Helber, F. Koenig, D. Borth, and A. Dengel, "Overcoming missing and incomplete modalities with generative adversarial networks for building footprint segmentation," *arXiv*, no. 2, 2018.
- [141] H. Xu, X. Tang, B. Ai, X. Gao, F. Yang, and Z. Wen, "Missing data reconstruction in vhr images based on progressive structure prediction and texture generation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 266 – 277, 2021.
- [142] Y. Zhang, B. Zhou, X. Cai, W. Guo, X. Ding, and X. Yuan, "Missing value imputation in multivariate time series with end-to-end generative adversarial networks," *Information Sciences*, vol. 551, pp. 67–82, 2021.
- [143] "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, no. February, pp. 148–160, 2020.
- [144] J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su, "Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks," *Remote Sensing*, vol. 12, no. 1, 2020.
- [145] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333 – 346, 2020.
- [146] J. D. Bermudez, P. N. Happ, R. Q. Feitosa, and D. A. Oliveira, "Synthesis of Multispectral Optical Images from SAR/Optical Multitemporal Data Using Conditional Generative Adversarial Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1220–1224, 2019.
- [147] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2018-July, pp. 1726–1729, 2018.
- [148] P. Ebel, A. Meraner, M. Schmitt, and X. X. Zhu, "Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2020.
- [149] P. Singh and N. Komodakis, "Cloud-GAN: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2018-July, pp. 1772–1775, 2018.
- [150] K. Enomoto, K. Sakurada, W. Wang, H. Fukui, M. Matsuoka, R. Nakamura, and N. Kawaguchi, "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets," *CVPR 2017*, pp. 48–56, 2017.
- [151] S. Chen, X. Wang, H. Guo, P. Xie, and A. M. Sirelkhatim, "Spatial and temporal adaptive gap-filling method producing daily cloud-free ndsi time series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2251–2263, 2020.
- [152] P. Dai, S. Ji, and Y. Zhang, "Gated convolutional networks for cloud removal from bi-temporal remote sensing images," *Remote Sensing*, vol. 12, no. 20, pp. 1–18, 2020.
- [153] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images," *Remote Sensing of Environment*, vol. 82, no. 2-3, pp. 173–187, 2002.
- [154] H. Shen, H. Li, Y. Qian, L. Zhang, and Q. Yuan, "An effective thin cloud removal procedure for visible remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 96, pp. 224–235, 2014.
- [155] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
- [156] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [157] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 154–169.
- [158] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017,

- pp. 4780–4788.
- [159] Y. Song, J. Li, X. Wang, and X. Chen, “Single image dehazing using ranking convolutional neural network,” *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1548–1560, 2018.
 - [160] M. Tan, T. Fang, Y. Fan, W. Wu, Q. She, and H. Gan, “Image-dehazing method based on the fusion coding of contours and colors,” *IEEE Access*, vol. 7, pp. 147 857–147 871, 2019.
 - [161] S. Zhang, F. He, and W. Ren, “Nlcn: Non-local dehazing network for dense haze removal,” *Neurocomputing*, vol. 410, pp. 363 – 373, 2020.
 - [162] S. Yin, Y. Wang, and Y.-H. Yang, “A novel image-dehazing network with a parallel attention block,” *Pattern Recognition*, vol. 102, p. 107255, 2020.
 - [163] C. Wang, W. Fan, Y. Wu, and Z. Su, “Weakly supervised single image dehazing,” *Journal of Visual Communication and Image Representation*, vol. 72, p. 102897, 2020.
 - [164] P. Huang, L. Zhao, R. Jiang, T. Wang, and X. Zhang, “Self-filtering image dehazing with self-supporting module,” *Neurocomputing*, vol. 432, pp. 57 – 69, 2021.
 - [165] Y. Liu, A. Wang, H. Zhou, and P. Jia, “Single nighttime image dehazing based on image decomposition,” *Signal Processing*, p. 107986, 2021.
 - [166] K. Borkar and S. Mukherjee, “Single image dehazing by approximating and eliminating the additional airlight component,” *Neurocomputing*, vol. 400, pp. 294 – 308, 2020.
 - [167] Y. Wu, Y. Qin, Z. Wang, X. Ma, and Z. Cao, “Densely pyramidal residual network for uav-based railway images dehazing,” *Neurocomputing*, vol. 371, pp. 124 – 136, 2020.
 - [168] L. Xu, D. Zhao, Y. Yan, S. Kwong, J. Chen, and L.-Y. Duan, “Iders: Iterative dehazing method for single remote sensing image,” *Information Sciences*, vol. 489, pp. 50 – 62, 2019.
 - [169] R. Li, J. Pan, Z. Li, and J. Tang, “Single image dehazing via conditional generative adversarial network,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8202–8211.
 - [170] J. Park, D. K. Han, and H. Ko, “Fusion of heterogeneous adversarial networks for single image dehazing,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4721–4732, 2020.
 - [171] Y. Z. Su, Z. G. Cui, C. He, A. H. Li, T. Wang, and K. Cheng, “Prior guided conditional generative adversarial network for single image dehazing,” *Neurocomputing*, vol. 423, pp. 620 – 638, 2021.
 - [172] B. Chaitanya and S. Mukherjee, “Single image dehazing using improved cyclegan,” *Journal of Visual Communication and Image Representation*, vol. 74, p. 103014, 2021.
 - [173] J. Li, Z. Wu, Z. Hu, J. Zhang, M. Li, L. Mo, and M. Molinier, “Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 373 – 389, 2020.
 - [174] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, 2020.
 - [175] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, “Deep learning for single image super-resolution: A brief review,” *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.
 - [176] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, “The 2018 pirm challenge on perceptual image super-resolution,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.
 - [177] K. Zhang, S. Gu, and R. Timofte, “Ntire 2020 challenge on perceptual extreme super-resolution: Methods and results,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 492–493.
 - [178] S. Lin, Z. Han, D. Li, J. Zeng, X. Yang, X. Liu, and F. Liu, “Integrating model- and data-driven methods for synchronous adaptive multi-band image fusion,” *Information Fusion*, vol. 54, pp. 145 – 160, 2020.