Cloud Imputation for Multi-sensor Remote Sensing Imagery with Style Transfer

Yifan Zhao, Xian Yang, and Ranga Raju Vatsavai(✉)

Computer Science Department, North Carolina State University, Raleigh, USA
{yzhao48,xyang45,rrvatsav}@ncsu.edu

Abstract. Widely used optical remote sensing images are often contaminated by clouds. The missing or cloud-contaminated data leads to incorrect predictions by the downstream machine learning tasks. However, the availability of multi-sensor remote sensing imagery has great potential for improving imputation under clouds. Existing cloud imputation methods could generally preserve the spatial structure in the imputed regions, however, the spectral distribution does not match the target image due to differences in sensor characteristics and temporal differences. In this paper, we present a novel deep learning-based multi-sensor imputation technique inspired by the computer vision-based style transfer. The proposed deep learning framework consists of two modules: (i) cluster-based attentional instance normalization (CAIN), and (ii) adaptive instance normalization (AdaIN). The combined module, CAINA, exploits the style information from cloud-free regions. These regions (land cover) were obtained through clustering to reduce the style differences between the target and predicted image patches. We have conducted extensive experiments and made comparisons against the state-of-the-art methods using a benchmark dataset with images from Landsat-8 and Sentinel-2 satellites. Our experiments show that the proposed CAINA is at least 24.49% better on MSE and 18.38% better on cloud MSE as compared to state-of-the-art methods.

Keywords: Cloud imputation · Multi-sensor · Deep learning · Style transfer

1 Introduction

Remote sensing imagery has been widely used as an important research material and information source in many applications ranging from crop monitoring, disaster mapping, nuclear proliferation, and urban planning since 1950's. However, since more than 50% of Earth's surface is covered by clouds [10] at any time, the performance of various downstream tasks such as segmentation, recognition, and classification on remote sensing images could be seriously affected because of the cloud-contaminated pixels. Fortunately, the advancing remote sensing technology and increasing number of satellite collections have significantly increased the spatial and temporal density of multi-sensor images.

G. De Francisci Morales et al. (Eds.): ECML PKDD 2023, LNAI 14175, pp. 37–53, 2023.
https://doi.org/10.1007/978-3-031-43430-3_3

Given the limitations of cloud-contaminated remote sensing images in the downstream applications (e.g., classification, change detection), a large number of techniques have been developed for imputation under clouds by exploiting multi-sensor imagery collections [1,2,4,14,34]. Multi-sensor imagery is preferred for cloud imputation as the large revisit cycle of a single satellite (more than 15 days) makes it hard to find temporally close-by images. In contrast, the chance of finding temporally close-by (less than a week) cloud-free images significantly increases if images from several satellites (that is, multi-sensor) are used [14]. Multi-sensor cloud imputation problem is often formulated as an image restoration task with a triplet consisting of the target, and before and after images [2,14,34]. Additionally, it is assumed that necessary cloud masks are often given beforehand, as these masks help focus the imputation to cloudy regions [17,34]. Given the computation and memory limitations, deep learning approaches often work with small images (e.g., 384×384). However, the typical size of a remote sensing image is more than 7000×7000 for Landsat and 10000×10000 for Sentinel satellites. In order to use these large images in training deep learning models, we often split them into smaller patches. For convenience, we call these input patches as images. Any subpart of this image is called as a patch (for example, the cloudy portion of an image is called a patch, and any small portion of the background (non-cloudy image) is also referred to as a patch). Usually, an image with a cloudy patch is treated as the target, and two temporally nearby geo-registered cloud-free images as input. These nearby images may come from the same sensor as the target image or a different sensor.

Though recent advances in deep learning-based multi-sensor cloud imputation methods have improved imputation performance significantly against single-sensor cloud imputation methods, they still have limitations. In particular, these methods can preserve the spatial structure of the imputed patches close to the input images. However, to the best of our knowledge, the existing multi-sensor cloud imputation models can't preserve the pixel-level spectral properties of the target image. As a result, the imputed patches are not close to the target images in terms of color style (spectral values). To address this issue, we propose a novel deep learning framework that harmonizes the imputed cloudy patches to the target image. The multi-sensor component of the network preserves the structure of the imputed patch and the harmonization component learns to transfer the color style by utilizing the cloud-free background and the land cover information of the target image.

From computer vision literature, earlier methods of style transfer between images can be attributed to the work of [6]. They used VGG-based deep learning architecture with a goal of style transfer to synthesize a texture from a source image, called "style," while preserving the semantic content of a target image called "content." Later works by [3,8,24] found that the feature statistics such as mean and standard deviation are highly effective in controlling the "style" of the output images. In particular, the adaptive instance normalization (AdaIN) method proposed by [8] can accommodate arbitrary style images without pre-training using adaptive affine transformations learned from the style

inputs. This AdaIN approach gave us the idea of adopting it to the multi-sensor cloud imputation problem by transferring the style of cloud-free background to imputed patches. In the multi-sensor cloud imputation problem, the cloud-free background of the target image will be the style and the input images from which the imputed patches are derived will be the content.

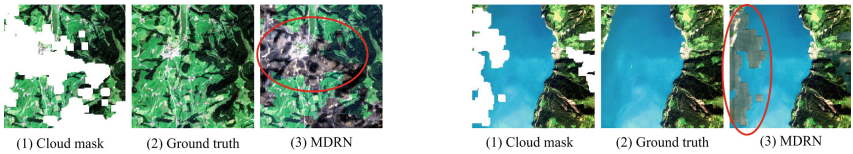

Fig. 1. Figure shows the spectral (color style) differences (red circles) in the imputed regions. From left to right: (1) the cloud-masked image, (2) the ground truth, (3) the imputed image by a state-of-the-art method called MDRN [34] (Color figure online).

However, the existing AdaIN only takes the mean and standard deviation of the whole style feature as the transferring style information. Remote sensing images often contain multiple types of land covers (e.g., forests, crops, buildings) and thus multiple and complicated styles in a single image. Therefore, AdaIN could only provide limited improvement for the multi-sensor cloud imputation task. To address this limitation, we propose a novel extension to the AdaIN that exploits the land cover information of the target image and transfers style information from targeted patches called cluster-based attentional instance normalization (CAIN). Without requiring extra land cover data, the land cover information can be extracted with an unsupervised clustering method such as K-means [30]. Recall that a patch is a small portion of the image, these smaller patches can effectively capture the individual land cover types. CAIN first splits both the cloud-free style and imputed content portions of the image into smaller patches and matches each of the style and content patches according to their land cover clustering results. For each imputed content patch, the cloud-free background patches with the same land cover cluster are selected for transferring the feature statistics, that is, the mean and standard deviation of the cloud-free patches. This way, each patch of the imputed feature will be transferred to the style of the patches with the same land cover cluster and, thus, to the style closer to the target image. However, CAIN could be prone to the noise and bias contained by a single land cover cluster. Therefore to overcome the limitations of both AdaIN and CAIN, we combine them using a weighted combination scheme called CAINA (CAIN + AdaIN). Thorough experimentation showed that both the bias (via MSE) and variance (via box-plots) have significantly reduced as the CAINA captures both general global and particular land cover style information.

Another advantage of the style transfer modules described above is that they can be easily plugged into various deep learning architectures. In this paper, we incorporated CAIN and CAINA modules in the deep learning networks inspired by MDRN [34] and MSOP$_{unet}$ [2] and named the resulting architecture

as MDRN$_{unet}$. While MDRN$_{unet}$ has the same multi-stream-fusion structure and composite upsampling structure as of MDRN, it also has U-Net [18] components inspired by MSOP$_{unet}$ [2].

Overall, the contributions of this paper are two-fold. First, a novel style transfer module, CAINA, is designed to exploit the remote sensing feature statistics for harmonizing the imputed cloudy patches using the cloud-free background. Second, a new deep learning network architecture was proposed by combining the merits of two state-of-the-art multi-sensor cloud imputation models (MDRN and MSOP$_{unet}$) for testing our proposed style transfer module. Finally, extensive experiments are conducted on a multi-sensor cloud imputation benchmark dataset for evaluating the performance of our proposed style transfer module. Experimental results showed that our proposed CAINA outperformed the state-of-art methods by at least 18.38% and 24.49% using mean squared error (MSE) in cloudy regions and the entire images, respectively.

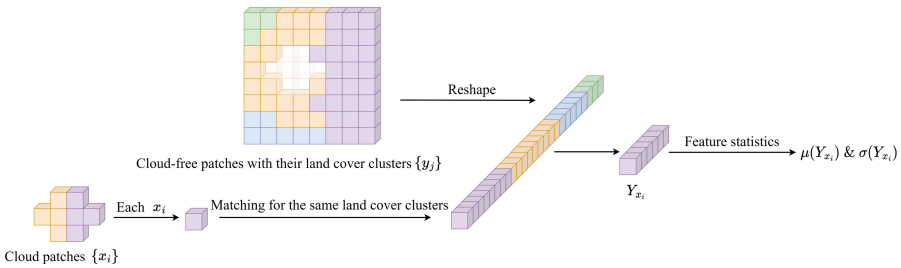

Fig. 2. The illustration for cluster-based attentional instance normalization (CAIN). We split the content feature X and the style feature Y into two sets of patches, $\{x_i\}$ and $\{y_j\}$, respectively. Then a lightweight K-means clustering method is employed to extract each patch's land cover type (Illustrated with different colors in $\{x_i\}$ and $\{y_j\}$). Then for each x_i, all the patches with the same land cover type in $\{y_j\}$ are selected and denoted as Y_{x_i}. Then Y_{x_i} can be aggregated for transferring the mean $\mu(Y_{x_i})$ and standard deviation $\sigma(Y_{x_i})$ to x_i.

2 Related Work

2.1 Multi-sensor Cloud Imputation

The remote sensing cloud imputation problem has been primarily considered in single-sensor or single-image settings previously in [9,21,22,29,31,33]. Although these works made significant improvements on the cloud imputation task, single-sensor settings can only be adopted to limited practical situations as it has more restrictions for input compared to multi-sensor settings. In contrast, cloud imputation with multi-sensor data was considered in [1,4,14,19]. [2,4,14] used optical and SAR channels for cloud imputation tasks. However, they did not explicitly address the multi-resolution issue between SAR and optical images.

Instead, they artificially down-sampled the SAR images to the same lower resolution as optical images and thus caused a loss of spectral and spatial information. The multi-resolution settings in remote sensing imagery were tackled while other problems such as land cover classification and segmentation were addressed in [13,16,20,23,25–27,36]. Recently, the multi-resolution issue in the cloud imputation problem was tackled by a multi-stream deep residual network (MDRN) [34]. MDRN used a multi-stream-fusion structure to process inputs with different resolutions separately and achieved state-of-the-art performance. However, MDRN could not effectively harmonize the imputed patches to the same color style as the target image. Therefore, in this paper, we attempt to improve the harmonization of imputed patches with our proposed style transfer modules, CAIN and CAINA while keeping the effective components of MDRN in our deep learning network, MDRN_{unet}.

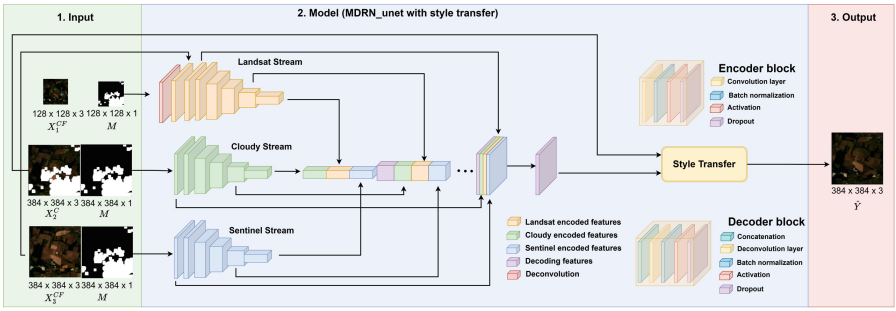

Fig. 3. The architecture and data flow of our testing deep learning network. The detailed structures of the encoder and decoder blocks are shown in the right-hand side. X_1^{CF} (CF stands for cloud-free) is the cloud-free Landsat-8 input, X_2^C is the cloudy Sentinel-2 input, and X_3^{CF} is the cloud-free Sentinel-2 input. $\hat{Y}$ is the predicted target cloud-free image.

2.2 Style Transfer

Style transfer between images was tackled with deep learning networks first in [6]. The goal of style transfer is to synthesize a texture from a source image, called "style," while preserving the semantic content of a target image called "content." Later works done by [3,8,24] discovered that the feature statistics such as mean and standard deviation in a deep learning network can be effective in controlling the style of the output images. In particular, adaptive instance normalization (AdaIN) was proposed by [8] for arbitrary style transfer. AdaIN has no learnable affine parameters. Each content could be paired with a style in every data instance. AdaIN's adaptiveness enabled the possibility of improving multi-sensor cloud imputation with style transfer ideas, as the cloud-free background could be the style and the imputed patches could be the content. However, AdaIN computed the statistics over the entire style feature and could contain tangent

information in the feature statistics. The tangent information could compromise the performance as it is not expected in the cloud patch. In contrast, our proposed CAINA extracts feature statistics more accurately from the semantically similar cloud-free regions with the same land cover cluster as the cloud patch so that the cloud patch could be transferred with reduced tangent style information.

In addition to the instance normalization methods that directly inspired our work, style transfer has also been tackled with other works focusing on the innovation of deep learning architectures [12,15]. Multi-level interactive Siamese filtering (MISF) [12] aims at the high-fidelity transformation of background in image inpainting by exploiting the semantic information with a kernel prediction branch and filling details with an image filtering branch. Whereas contrastive unpaired translation (CUT) [15] proposed a patchwise contrastive loss based on the famous Cycle-consistent GAN [37] to overcome the restriction of bijective assumption with more accurate contrastive translation in the style transfer task. However, while applying to the remote sensing imagery, these methods didn't exploit the valuable information in land cover clusters and, thus, cannot achieve optimal cloud imputation performance.

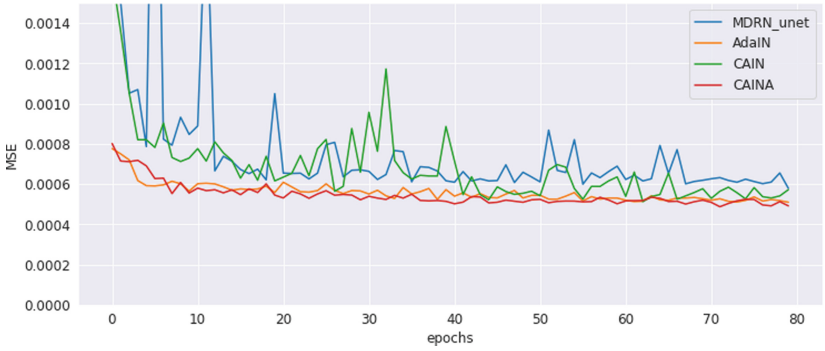

Fig. 4. The validation MSE loss curves of the same deep learning architecture without style transfer, with AdaIN, with CAIN, with CAINA.

The idea and methods of style transfer were considered to be helpful for cloud imputation in remote sensing imagery only starting from recent years [32,35]. AdaIN was adopted and applied to a cloud imputation model in [35] for controlling the global information of the images. Two parameters generated by a pre-trained MLP network were used to replace the feature statistics used in [8]. Another example of employing AdaIN for cloud imputation is presented by [32]. AdaIN enabled incorporating physical attributes such as cloud reflection, ground atmospheric estimation, and cloud light transmission to the deep learning networks in [32]. However, the usages of AdaIN in [32,35] relied on the reinforcement of pre-trained models and external physical information. Additionally, they still applied identical style information for all cloud patches. In contrast, our proposed

CAINA applied the style of the corresponding land cover type by clustering techniques to each cloud patch and do not rely on any pre-trained models or external physical information.

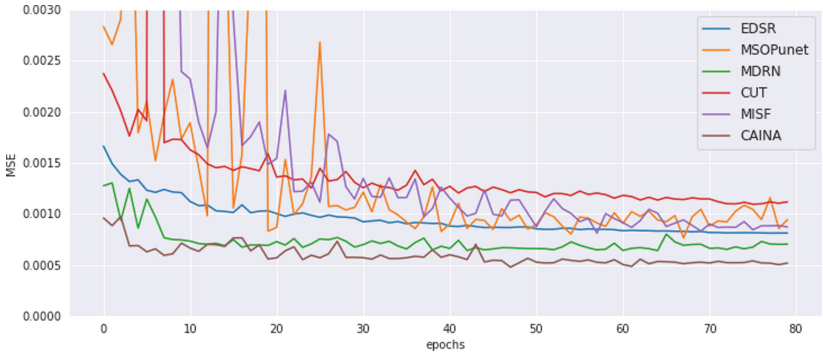

Fig. 5. The validation MSE loss curves of the state-of-the-art deep learning cloud imputation model, EDSR, MSOP$_{unet}$, MDRN, CUT, MISF comparing with CAINA.

3 Methodology

Existing multi-sensor cloud imputation methods could generally detect the missing values and derive the spectral content from the temporally-nearby cloud-free images reasonably well. Though the spatial structure under the cloud patches is close to the target image, the existing models do not effectively preserve the pixel-level spectral properties of the target image due to spectral and temporal differences. Figure 1 shows some examples of the cloud patches imputed by an existing cloud imputation method (MDRN). As can be seen from the images, the imputed patches do not match the spectral distribution (color style) of the target image.

As the pixel-level spectral properties of remote sensing images tend to depend on time and the sensor collection, the patches imputed by existing deep learning networks often do not match the surrounding regions in the target image. Thus, to make the imputed patches consistent with the target image, we need to transfer the style of the cloud-free background to the imputed patches. Therefore, the style transfer techniques in the computer vision (CV) area were considered and evaluated. In this section, we demonstrate our attempts to bridge the style transfer area to the multi-sensor cloud imputation problem and propose new style transfer modules that serve the multi-sensor cloud imputation problem better.

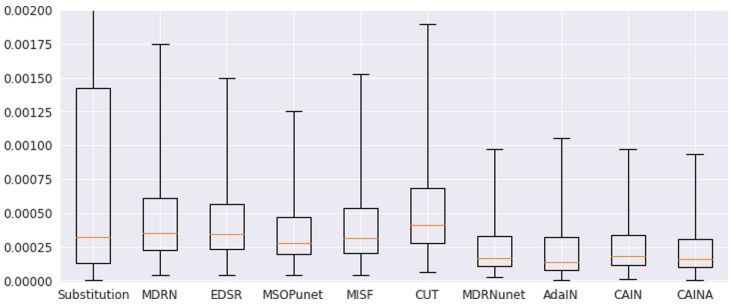

Fig. 6. The validation cloud MSE boxplots of Substitution, MDRN, EDSR, MSOP$_{unet}$, MISF, CUT, MDRN$_{unet}$, AdaIN, CAIN, and CAINA. The whiskers extend from the box by 3x the inter-quartile range (IQR). Outliers (around 10% of the total validation set size) that pass the end of the whiskers are omitted. It is shown that the variance of CAINA is lower than all other methods, which is why CAINA outperformed the state-of-the-art methods on averaged cloud imputation performance.

3.1 Adaptive Instance Normalization (AdaIN)

AdaIN [8] is an arbitrary style transfer technique that could take an arbitrary style image as input without pre-training. The goal of style transfer is to synthesize a texture from a source image, called "style," while preserving the semantic content of a target image called "content." The intuition of AdaIN is to make the content image aligned with the mean and standard deviation of the "style" image. More formally, suppose X and Y are content and style features, respectively, then AdaIN aligns the feature-wise mean (μ) and standard deviation (σ) of X to those of Y.

$$\mathbf{AdaIN}(X, Y) = \sigma(Y)\left(\frac{X - \mu(X)}{\sigma(X)}\right) + \mu(Y) \tag{1}$$

In the case of cloud imputation, we are dealing with the following image triplets similar to [31,34], X_1^{CF}, $Y^C = X_2^C$, and X_3^{CF} (CF stands for cloud-free and C stands for cloudy), where Y^C is the target image containing the cloud patches, and X_1^{CF} and X_3^{CF} are nearby cloud-free images which could be from a different sensor than the target image Y^C. From the perspective of style transfer notation, the content feature X comes from $X_{\{1,3\}}^{CF}$, and the style feature Y comes from the cloud-free region of the target image Y^C.

3.2 Cluster-Based Attentional Instance Normalization (CAIN)

Although experiments show that transferring the global mean and standard deviation of the cloud-free background to cloud patches could improve the cloud imputation performance, the improvement is still limited since remote sensing images often contain multiple types of land covers (e.g., forests, crops, buildings). Thus more focused and accurate style information for the cloudy patches

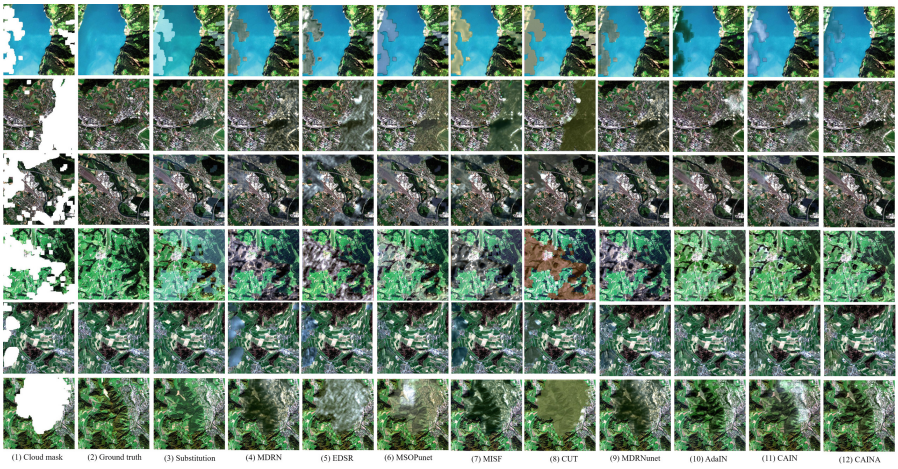

Fig. 7. Few examples of cloud imputed images showing the comparison across the state-of-the-art deep learning cloud imputation models and our testing methods, MDRN$_{unet}$, AdaIN, CAIN, and CAINA. From the left to the right: (1) the cloud-masked images; (2) the ground truths; the restored images by: (3) Substitution, (4) MDRN, (5) EDSR, (6) MSOP$_{unet}$, (7) MISF, (8) CUT, (9) MDRN$_{unet}$, (10) AdaIN, (11) CAIN, (12) CAINA.

is expected to further reduce the style inconsistency between predicted images and the target images.

Therefore, we propose a new module called cluster-based attentional instance normalization (CAIN). Instead of simply normalizing all pixels in the content feature X with the global mean and standard deviation of the style feature Y, we only transfer the feature statistics of the pixels with the same land cover type as the cloudy pixels. Specifically, we first employ a lightweight clustering model such as K-means [30] on a temporally close cloud-free image, X_3^{CF}, for obtaining the land cover information. Then we split X and Y into two sets of patches $\{x_i\}$ and $\{y_j\}$ as shown in Fig. 2.

Then for each x_i, we extract all the cloud-free patches in the same land cover cluster, Y_{x_i},

$$Y_{x_i} = \tau_{\{y_j\}}\Big(C(y_j) = C(x_i)\Big) \tag{2}$$

where $\tau(\cdot)$ is the choice function. $C(y_j)$ and $C(x_i)$ are y_j and x_i's land cover clusters, respectively. Then x_i is aligned to the mean and standard deviation, Y_{x_i}, as given by CAIN(,):

$$\mathbf{CAIN}(x_i, Y) = \sigma(Y_{x_i})\left(\frac{x_i - \mu(x_i)}{\sigma(x_i)}\right) + \mu(Y_{x_i}) \tag{3}$$

This way, the content patches $\{x_i\}$ are transferred to the mean and standard deviation of the style patches in the same land cover cluster, $\{Y_{x_i}\}$. And the

content feature X could be more consistent with the cloud-free background of the target image Y.

3.3 Composite Style Transfer Module, CAIN + AdaIN (CAINA)

Though experiments show that CAIN could provide more accurate style information, the mean and standard deviation of the same land cover type are aggregated from a subset of cloud-free patches and could be prone to the employed clustering method's limitations. Therefore, a weighted combination of CAIN and AdaIN (CAINA) is proposed to overcome their disadvantages and utilize their advantages simultaneously,

$$\mathbf{CAINA}(X, Y) = \mathbf{Convolution}(\mathbf{CAIN}(X, Y) \oplus \mathbf{AdaIN}(X, Y)) \quad (4)$$

We employ a convolution layer to perform an automatic weighted combination of the concatenated features returned by CAIN and AdaIN. In this setup, the style information from the same land cover type is focused, while the style information from the entire image could also contribute to the predictions. Experiments show that the variance of the predictions is reduced with CAINA and the average error of cloud imputation is further reduced as well.

3.4 The Deep Learning Network Architecture

Since the style transfer modules demonstrated above are independent from any particular deep learning networks, they can be easily plugged into various deep learning architectures. In this paper, we incorporate AdaIN, CAIN, and CAINA in a deep learning architecture (MDRN_{unet}), inspired by MDRN [34] and MSOP_{unet} [2]. As the multi-sensor, multi-resolution cloud imputation problem is considered in this paper, the architecture has the same multi-stream-fusion structure and composite upsampling structure as MDRN. On the other hand, inspired by MSOP_{unet} [2], more U-Net [18] components were employed. Figure 3 shows the architecture and dataflow of our deep learning network.

4 Experiments

4.1 Dataset and Environmental Configuration

We test and compare all the methods on the benchmark dataset introduced in [34] with remote sensing images from Landsat-8[1] and Sentinel-2[2]. This collection includes the temporally closest image triplets, thus ideal for testing our proposed methods. Another recent cloud imputation benchmark dataset, Sen12MS-CR-TS [5], also includes temporally close-by multi-sensor image collections. However, Sen12MS-CR-TS uses SAR channels, two airborne microwave channels that

[1] https://landsat.gsfc.nasa.gov/satellites/landsat-8/.
[2] https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

can penetrate clouds but contain non-negligible noises. Thus, Sen12MS-CR-TS is not an ideal benchmark dataset for evaluating the proposed methods here.

The most widely used RGB channels are used for training our model. However, the proposed architecture does not depend on any specific channel combination and it could be readily trained on any number of channels and their combinations as long as the system permits (memory and compute power). The dataset is split into independent training (consisting of 4,003 instances) and validation (1,000 instances). All the models are trained with the following parameters: batch-size = 16, epochs = 80, mean squared error (MSE) loss, ADAM optimizer, and a step learning rate scheduler starting from 0.01 and every 10 epochs decreases at the rate of 0.75. The source code is implemented with PyTorch[3] and has been deployed to our sponsor's system.

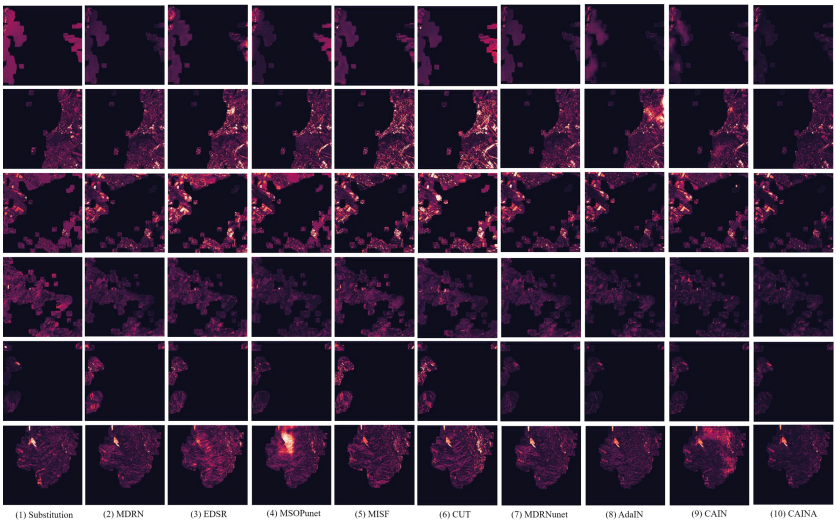

Fig. 8. The cloud residual maps of the same examples in Fig. 7 showing the comparison across the state-of-the-art deep learning cloud imputation models and our proposed CAINA. From the left to the right: the cloud residual maps by: (1) Substitution, (2) MDRN, (3) EDSR, (4) MSOP$_{unet}$, (5) MISF, (6) CUT, (7) MDRN$_{unet}$, (8) AdaIN, (9) CAIN, (10) CAINA. The darker residual maps implies better cloud imputation results.

4.2 Experiment Settings

We perform two sets of experiments. In the first set of experiments, we compare the cloud imputation performance of the baseline method MDRN$_{unet}$ and the style transfer extensions: MDRN$_{unet}$ + AdaIN, MDRN$_{unet}$ + CAIN, and MDRN$_{unet}$ + CAINA. In the second set of experiments, we compare our best-performing cloud imputation model with the style transfer module, MDRN$_{unet}$

[3] https://github.com/YifanZhao0822/CAINA.

+ CAINA, with other state-of-the-art deep learning cloud imputation methods, namely, MDRN [34], MSOP$_{unet}$ [2], EDSR [14], MISF [12], and CUT [15]. Besides, we also compare with the imputation results by simply substituting the cloud region of the target image with the corresponding area from the temporally closest Sentinel-2 image and call this baseline method as "Substitution."

We report the quantitative comparison results using two types of error metrics: pixel-wise metrics and structural metrics. For pixel-wise metrics, we are using three well-known measures. These include Mean Square Error (MSE) for the entire image and the cloud area separately, whereas peak-signal-to-noise ratio (PSNR) [7] and spectral angle mapper (SAM) [11] on full images. The MSE shows how close the predicted pixels are with respect to the ground truth; the peak-signal-to-noise ratio (PSNR) approximates the human perception of the restored image; the spectral angle mapper (SAM) is used for evaluating the spectral difference over RGB channels. For structural metric, we used the structural similarity index (SSIM) [28] for measuring the image restoration quality from a visual perception standpoint. For the hyperparameters in CAIN and CAINA, we tuned the patch size= 3×3 and the K-means # clusters $k = 4$.

4.3 Quantitative Results of the First Set of Experiments

In this section, we present the experimental results of various extension and their relative performance over the baseline method: MDRN$_{unet}$, MDRN$_{unet}$ + AdaIN, MDRN$_{unet}$ + CAIN, and MDRN$_{unet}$ + CAINA. For simplicity, we are omitting the prefix of MDRN$_{unet}$, for example MDRN$_{unet}$ + CAINA is simply referred to as CAINA. Figure 4 shows the validation MSE loss curve of each model at the end of each epoch. We observe that both CAIN and AdaIN outperform MDRN$_{unet}$ in a significant way. CAINA further improves the performance on the basis of CAIN and AdaIN. Table 1 shows the comparison using MSE, cloud MSE, PSNR, SSIM, and SAM. As can be seen, CAINA outperforms all other methods on the pixel-wise metrics and it outperforms all other methods on structural metrics except for AdaIN on SAM measure.

Table 1. The comparison on MSE, cloud MSE, PSNR, and SSIM for MDRN$_{unet}$, AdaIN, CAIN, and CAINA. The best result of each metric is bolded.

Methods	MSE (10^{-4})	Cloud MSE (10^{-4})	PSNR	SSIM	SAM (10^{-2})
MDRN$_{unet}$	5.7871	16.2712	42.0875	0.9876	4.6549
AdaIN	5.0939	15.1207	42.6586	0.9878	**3.9988**
CAIN	5.1214	15.0891	42.2273	0.9871	4.3729
CAINA	**4.8222**	**14.3214**	**42.9390**	**0.9881**	4.1365

4.4 Quantitative Results of the Second Set of Experiments

In this section, we further compare our best-performing CAINA with the baseline method, Substitution, and the state-of-the-art deep learning cloud imputation models, namely MDRN [34], MSOP$_{unet}$ [2], EDSR [14], MISF [12], and CUT [15]. We used the same quantitative metrics as in Sect. 4.3, namely MSE, cloud MSE, PSNR, SSIM, and SAM for comparing the performance of each model. Figure 5 shows the validation MSE loss curves of each model except Substitution at the end of each training epoch. EDSR has the most stable convergence among the state-of-the-art models. However, its best MSE is still suboptimal. We observe that MSOP$_{unet}$ outperforms EDSR significantly, however, its validation loss is not stable. MDRN outperforms the other state-of-the-art methods on MSE but its validation MSE loss is still higher than CAINA. CAINA outperforms all methods significantly and has also reached a stable convergence after 40 epochs. Table 2 shows the comparison using MSE, cloud MSE, PSNR, SSIM, and SAM. As can be seen, CAINA outperforms all other methods on both the pixel-wise and structural metrics. Compared to the state-of-the-art cloud imputation models in Table 2, using the primary measure of MSE, CAINA shows at least 18.38% and 24.49% improvement in cloudy regions and the entire image, respectively. Additionally, MISF achieved the second-best cloud MSE although its performance on other metrics is limited. In our understanding, it could be the significant contribution of the novel kernel prediction module in MISF, which could be an inspiring point that leads to future innovations.

Table 2. The comparison on MSE, cloud MSE, PSNR, SSIM, and SAM for Substitution, MDRN, EDSR, MSOP$_{unet}$, CAINA. The best result of each metric is bolded.

Methods	MSE (10^{-4})	Cloud MSE (10^{-4})	PSNR	SSIM	SAM (10^{-2})
Substitution	25.6594	83.2660	38.6478	0.9704	4.6629
MDRN	6.3895	19.0507	39.5097	0.9810	5.1440
EDSR	8.1018	22.8269	39.3500	0.9805	4.9847
MSOP$_{unet}$	7.6326	21.4471	39.1222	0.9803	5.2841
CUT	10.9325	28.5818	36.1848	0.9701	8.6621
MISF	8.1022	17.5454	35.3311	0.9573	9.6810
CAINA	**4.8222**	**14.3214**	**42.9390**	**0.9881**	**4.1365**

4.5 Analysis on Variances Among the Compared Methods

We further analyze the results using boxplots to understand the performance gains of the CAINA better. Figure 6 shows the boxplot for Substitution, MDRN, EDSR, MSOP$_{unet}$, MISF, CUT, MDRN$_{unet}$, AdaIN, CAIN, and CAINA on the cloud MSE. Both CAIN and CAINA have the lowest third quartile (the upper bound of the boxes). In addition, CAINA has the lowest median among the boxplots. Therefore, the variance of CAINA is lower than all other methods,

which is why CAINA outperformed the state-of-the-art methods on averaged cloud imputation performance.

4.6 Qualitative Results and Residual Maps

Figure 7 and 8 shows a few examples of the restored RGB images and cloud residual maps for comparing across the state-of-the-art deep learning cloud imputation models and the testing methods in the same order as in Sect. 4.5. Images in Fig. 8 are residual maps generated by subtracting the predicted image from the ground truth in Fig. 7. The darker residual maps implies better cloud imputation results. Our proposed CAINA outperformed the state-of-the-art models consistently by achieving the darkest residual maps.

5 Conclusions

In this paper, we presented an effective cloud imputation model with a novel style transfer function (CAINA) that harmonizes imputed patches by exploiting image style information from the cloud-free region of the image to reduce the style differences between the target and predicted image patches. We have experimentally shown that our method not only brings improvements as an add-on module to the MDRN$_{unet}$, but also provides an improved cloud imputation performance in comparison to the several state-of-the-art deep learning models on a benchmark dataset. In particular, CAINA is at least 24.49% better on MSE as compared to the state-of-the-art models, and 18.38% better on cloud MSE. However, the current proposed CAINA relies on the results of K-means clustering and cloud-free regions of the target image. In the future, we will work on introducing land cover segmentation maps to replace K-means clustering for improving the reliability of the cloud imputation method. Additionally, our future work will also try to reduce the dependence on cloud-free regions of the target image by possibly exploiting sensor-level metadata.

Acknowledgments. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Contract #2021-21040700001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. We would like to thank Benjamin Raskob at ARA for useful feedback on this project.

Ethical Statement

Our proposed method improves cloud imputation performance. Remote sensing imagery has been widely used in applications ranging from land-use land-cover mapping to national security. By improving the imputation performance, we are

directly improving the downstream applications such as assessing damages due to natural disasters, forest fires, and climate impacts. Our work does not have direct ethical implications or adverse impacts on humans.

References

1. Cresson, R., Ienco, D., Gaetano, R., Ose, K., Minh, D.H.T.: Optical image gap filling using deep convolutional autoencoder from optical and radar images. In: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 218–221. IEEE (2019)
2. Cresson, R., et al.: Comparison of convolutional neural networks for cloudy optical images reconstruction from single or multitemporal joint SAR and optical images. arXiv preprint arXiv:2204.00424 (2022)
3. Dumoulin, V., Shlens, J., Kudlur, M.: A learned representation for artistic style. arXiv preprint arXiv:1610.07629 (2016)
4. Ebel, P., Meraner, A., Schmitt, M., Zhu, X.X.: Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. IEEE Trans. Geosci. Remote Sens. **59**(7), 5866–5878 (2020)
5. Ebel, P., Xu, Y., Schmitt, M., Zhu, X.X.: SEN12MS-CR-TS: a remote-sensing data set for multimodal multitemporal cloud removal. IEEE Trans. Geosci. Remote Sens. **60**, 1–14 (2022)
6. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423 (2016)
7. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369. IEEE (2010)
8. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510 (2017)
9. Kang, S.H., Choi, Y., Choi, J.Y.: Restoration of missing patterns on satellite infrared sea surface temperature images due to cloud coverage using deep generative inpainting network. J. Mar. Sci. Eng. **9**(3), 310 (2021)
10. King, M.D., Platnick, S., Menzel, W.P., Ackerman, S.A., Hubanks, P.A.: Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites. IEEE Trans. Geosci. Remote Sens. **51**(7), 3826–3852 (2013)
11. Kruse, F.A., et al.: The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data. Remote Sens. Environ. **44**(2–3), 145–163 (1993)
12. Li, X., Guo, Q., Lin, D., Li, P., Feng, W., Wang, S.: MISF: multi-level interactive siamese filtering for high-fidelity image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1869–1878 (2022)
13. Ma, W., et al.: A novel adaptive hybrid fusion network for multiresolution remote sensing images classification. IEEE Trans. Geosci. Remote Sens. **60**, 1–17 (2021)
14. Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M.: Cloud removal in sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. ISPRS J. Photogramm. Remote. Sens. **166**, 333–346 (2020)

15. Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 319–345. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_19
16. Qu, J., Shi, Y., Xie, W., Li, Y., Wu, X., Du, Q.: MSSL: hyperspectral and panchromatic images fusion via multiresolution spatial-spectral feature learning networks. IEEE Trans. Geosci. Remote Sens. **60**, 1–13 (2021)
17. Requena-Mesa, C., Benson, V., Reichstein, M., Runge, J., Denzler, J.: Earthnet 2021: a large-scale dataset and challenge for earth surface forecasting as a guided video prediction task. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1132–1142 (2021)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
19. Roy, D.P., et al.: Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of landsat data. Remote Sens. Environ. **112**(6), 3112–3130 (2008)
20. Rudner, T.G., et al.: Multi3Net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 702–709 (2019)
21. Singh, P., Komodakis, N.: Cloud-Gan: cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 1772–1775. IEEE (2018)
22. Stock, A., et al.: Comparison of cloud-filling algorithms for marine satellite data. Remote Sens. **12**(20), 3313 (2020)
23. Sun, Z., Zhou, W., Ding, C., Xia, M.: Multi-resolution transformer network for building and road segmentation of remote sensing image. ISPRS Int. J. Geo Inf. **11**(3), 165 (2022)
24. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6924–6932 (2017)
25. Varshney, D., Persello, C., Gupta, P.K., Nikam, B.R.: Multiresolution fully convolutional networks to detect clouds and snow through optical satellite images. arXiv preprint arXiv:2201.02350 (2022)
26. Wang, L., Weng, L., Xia, M., Liu, J., Lin, H.: Multi-resolution supervision network with an adaptive weighted loss for desert segmentation. Remote Sens. **13**(11), 2054 (2021)
27. Wang, L., Zhang, C., Li, R., Duan, C., Meng, X., Atkinson, P.M.: Scale-aware neural network for semantic segmentation of multi-resolution remote sensing images. Remote Sens. **13**(24), 5015 (2021)
28. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
29. Weiss, D.J., Atkinson, P.M., Bhatt, S., Mappin, B., Hay, S.I., Gething, P.W.: An effective approach for gap-filling continental scale remotely sensed time-series. ISPRS J. Photogramm. Remote. Sens. **98**, 106–118 (2014)
30. Yadav, J., Sharma, M.: A review of k-mean algorithm. Int. J. Eng. Trends Technol. **4**(7), 2972–2976 (2013)

31. Yang, X., Zhao, Y., Vatsavai, R.R.: Deep residual network with multi-image attention for imputing under clouds in satellite imagery. In: 2022 27th International Conference on Pattern Recognition (ICPR). IEEE (2022)
32. Yu, W., Zhang, X., Pun, M.O., Liu, M.: A hybrid model-based and data-driven approach for cloud removal in satellite imagery using multi-scale distortion-aware networks. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 7160–7163. IEEE (2021)
33. Zhang, Q., Yuan, Q., Zeng, C., Li, X., Wei, Y.: Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. IEEE Trans. Geosci. Remote Sens. **56**(8), 4274–4288 (2018)
34. Zhao, Y., Yang, X., Vatsavai, R.R.: Multi-stream deep residual network for cloud imputation using multi-resolution remote sensing imagery. In: 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 97–104. IEEE (2022)
35. Zhao, Y., Shen, S., Hu, J., Li, Y., Pan, J.: Cloud removal using multimodal GAN with adversarial consistency loss. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2021)
36. Zhu, H., Ma, W., Li, L., Jiao, L., Yang, S., Hou, B.: A dual-branch attention fusion deep network for multiresolution remote-sensing image classification. Inf. Fusion **58**, 116–131 (2020)
37. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)