

Image translation between SAR and optical imagery for Arctic wildfire analysis

Xikun Hu

*Department of Urban Planning & Environment
KTH Royal Institute of Technology
Teknikringen 10A, 100 44 Stockholm, Sweden
xikun@kth.se*

Maryam Rahnemoonfar

*College of Engineering and Information Technology
University of Maryland, Baltimore County
1000 Hilltop Cir, Baltimore, MD 21250
maryam@umbc.edu*

Tian Jin

*College of Electronic Science and Technology
National University of Defense Technology
No.109, Deya Road, Changsha 410003, China
tianjin@nudt.edu.cn*

Abstract—Climate change by anthropogenic warming leads to the Arctic regions being vulnerable. Though deep learning has been widely applied to remote sensing and climate science, most applications are centered on the usage of optical satellite imagery rather than radar sensors that can penetrate through clouds. The performance of burned area detection depends on the SAR sensor characteristics (e.g. relative orbit number), dominated land cover (e.g. unburned or burned), and the environmental conditions (e.g. soil moisture) during SAR data acquisition. As the large-scale spatial-temporal variation of SAR images results in the vague synthesis from SAR to optical, we present a novel model based on CyclePix2Pix architecture with style matching to improve the quality of image translation and avoid mode collapse with one-to-many optical-SAR images in pairs. Unlike the existing CycleGAN model, the CyclePix2Pix combines the advantages of Pix2Pix and segregates the content features and style features. The enhancement of style features in the homogeneous transformation between two generators prevents the corruption of image semantic content (i.e. information of burned disturbance). The relatively high SSIM (0.48) and PSNR (19.65 dB) on average demonstrates that the proposed CyclePix2Pix is a promising alternative that facilitates disaster management related to wildfire. After semantic segmentation, the Kappa and mIoU achieved 84.59% and 86.18%, respectively with style matching involved.

Index Terms—Synthetic Aperture Radar, image translation, generative adversarial network, burned areas

I. INTRODUCTION

Wildfires generate considerable research interest due to its high frequency of occurrence around the globe recently, even in Arctic Circle [1]. Northern peatlands cool the climate but their disturbances caused by the fire would trigger losses of vegetation and an increase of greenhouse gases [2]. Optical satellite images usually failed in these northern regions to observe peatland wildfire because of high cloud coverage (see Fig. 9 in the Appendix). Therefore, Synthetic Aperture Radar (SAR) images become an alternative solution with more SAR satellites launched [3]. In contrast to optical imaging sensors depending on the Sun’s illumination, radar (e.g. SAR) is an active imaging system. Although there is no color associated

with raw radar imagery (unlike optical imagery illuminated by various colors from the visible light.), it provides at least two significant advantages: the ability to see through clouds, and the ability to image at night.

Burned area mapping algorithms from radar images are mostly based on temporal differences between pre- and post-fire backscatter values. These bi-temporal change detection approaches have several limitations in terms of spatio-temporal variations. Due to a reduction in grass biomass and grass/soil moisture, a systematic backscatter decrease from pre-fire to post-fire conditions might be observed, e.g. in southern African deciduous savannahs [4]. However, previous related studies have mostly demonstrated that enhanced returns from burned areas can be observed concerning unburned areas at C-band in the case of Mediterranean ecosystems, due to the increased soil moisture after rainfall [5]. On the other hand, temporal decorrelation is of the utmost importance for areas detected as burned immediately in Mediterranean ecosystems, which also constrains the usage of bi-temporal SAR images [3]. Therefore, to avoid the temporal variance between SAR images, uni-temporal detection base on a single post-event SAR image might be promising as burned area detection using a single optical image in [6], if we could represent the characteristics of SAR images with optical spectral information.

Comprehensive interpretation of the SAR image on wild-land fire is still challenging because of the spatio-temporal variations in terms of different temporal information (e.g. various soil moisture and relative orbit numbers in different acquisition dates) and spatial information (e.g. different landscapes with/without burned area pixels), which hinders the wide application of classical approaches used in optical data. Deep learning has been developed rapidly to solve fire-related tasks in image processing for either SAR or multi-spectral satellite images. An implicit Radar Convolutional Burn Index under a convolutional network-based classification framework was proposed based on multitemporal Sentinel-1 SAR data for mapping the burnt areas [7]. A Lightweight U-

Net was implemented in [8] for active fire spots segmentation based on B7, B6, and B2 bands of Landsat OLI image. Recently, the performance of deep convolutional autoencoders (U-Net and ResU-net) was compared using a bi-temporal image pair of the Landsat scenes and recommended the sampling window size of 256 by 256 pixels [9]. Generative Adversarial Networks (GAN)-based optical synthesis from Sentinel-1 data was employed to classify regions with the burned area in land cover classes but it additional required multi-temporal data as conditioning variables [10].

As a breakthrough development in deep learning, GAN resolves the image-to-image translation tasks as typical probability generation models [11]. GANs explores the internal data distribution via learning abundant data. Conditional Generative Adversarial Networks (cGANs) is proposed to control the data generation process, which is applied to many image-to-image GAN architectures such as Pix2Pix [12], Cycle-Consistent Generative Adversarial Networks (CycleGAN) [13]. Regarding the earth observation, GANs have been widely used to translate the SAR images to optical images in [14]–[17], produce samples for SAR-to-optical matching [18], [19], implement SAR image colorization [20] and SAR image classification [21], [22]. Most SAR-to-optical translation methods are based on an independent pair of images in the same regions without considering the temporal connection within paired images, and few studies focus on the GAN generation from uncertain source SAR images with a temporal variation. In reality, taking wildfire detection as an example, SAR images might be acquired under various conditions such as different climates, or land cover change before or after the event. It leads to limited applications in climate science practically.

To overcome this limitation, in this work, GAN is conducted to generate the pseudo optical images from the SAR images over fire-disturbed areas. The main contributions of this study include: (1) analyze the impacts of SAR characteristics and environmental conditions on SAR to optical translation based on one-to-many strategy; (2) propose CyclePix2Pix model to translate between SAR and optical images with style matching to fuse diversity (i.e. avoid a collapse into one mode).

II. DATASET

In this work, three fire-sensitive bands of Sentinel-2 MSI Level 1-C are selected (i.e. B12, B11, and B8A) with a spatial resolution of 20 meters. More details about Sentinel-2 data can be found at [23]. On the other hand, Sentinel-1 SAR Ground Range Detected (GRD) C-band data are also acquired from ESA Copernicus [24]. The region of interest (ROI) is located in the extreme north of Yakutia, Russia above the Arctic Circle with a latitude of 67°23'05" North and longitude 154°05'07" East, where a large wildfire near to Kolyma River ignited on 17th of June, 2020 and extinguished around 6th of July after heavy rainfall. Calibrated Global Precipitation Measurement (GPM) [25] provides next-generation observations of rain and snow worldwide every three hours as Fig. 1 given.

We chose one cloud-free Sentinel-2 optical image (acquired on 21st of July) as the target and 10 post-event Sentinel-1

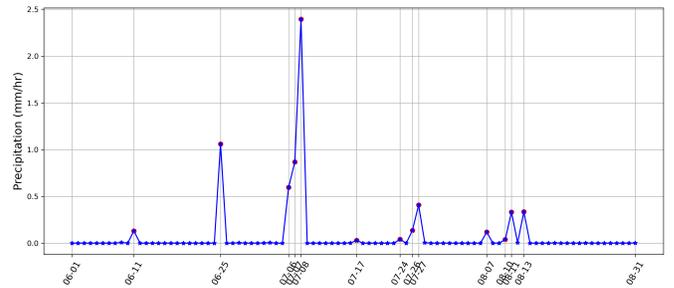


Fig. 1. Calibrated Precipitation (mm/hr) in ROI 2020 based on GPM v6

radar images between the 9th of July and 28th of August are downloaded with Vertical-horizontal (VH) and Vertical-vertical (VV) backscattering coefficient (i.e. σ_{VV}^0 and σ_{VH}^0) in descending pass orbit. The third channel of is cross-polarization ratio ($R_{cross} = \sigma_{VV}^0 / \sigma_{VH}^0$). A false-color composite scheme (R-G-B, B12-B11-B8A for optical image and VH-VV- R_{cross} for SAR image, respectively) has been adopted and shown in Fig. 2.

According to sensor metadata, precipitation dataset, as well as a visual inspection, we characterize the SAR image spatio-temporal properties, namely different temporal information (e.g. various soil moisture and relative orbit numbers in different acquisition dates) and spatial information (e.g. patches with burned area pixels and patches without/with few burned areas pixels). Therefore, we divided test image patches into 4 sub-categories (see also Fig. 2). The detail of the acquisition date of images belonging to each sub-category is summarised in Table.I

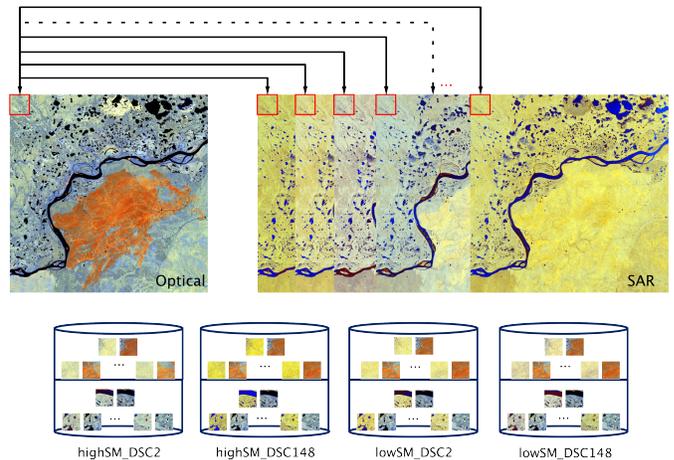


Fig. 2. Taking one representative patch as an example, here is one-to-many strategy (i.e. one cloud-free optical image is paired with 10 SAR images before or after the acquisition date of optical one) to group the sub-categories. These paired images with $3840 \times 3840 \times 3$ are randomly cropped into $256 \times 256 \times 3$ patches, respectively. The whole number of dataset patches is 2250. Properties in abbreviation in test stage: highSM - high soil moisture; lowSM: low soil moisture; DSC2/148 - Orbit 2/148 in descending pass.

TABLE I
SUB-CATEGORY DATASETS WITH SAR IMAGE DESCRIPTION

Sub-categories	Date	Image ID
lowSM_DSC2	07-11	S1B_IW_GRDH_1SDV_20200711T200916_20200711T200941_022428_02A915_0BB7
	07-23	S1B_IW_GRDH_1SDV_20200723T200916_20200723T200941_022603_02AE65_E486
	08-04	S1B_IW_GRDH_1SDV_20200804T200917_20200804T200942_022778_02B3B1_5890
	08-28	S1B_IW_GRDH_1SDV_20200828T200919_20200828T200944_023128_02BEA1_9D1C
lowSM_DSC148	07-21	S1B_IW_GRDH_1SDV_20200721T202526_20200721T202551_022574_02AD88_9813
		S1B_IW_GRDH_1SDV_20200721T202551_20200721T202619_022574_02AD88_AAC6
		S1B_IW_GRDH_1SDV_20200802T202527_20200802T202552_022749_02B2D2_E591
	08-02	S1B_IW_GRDH_1SDV_20200802T202552_20200802T202620_022749_02B2D2_E6DE
	08-26	S1B_IW_GRDH_1SDV_20200826T202528_20200826T202553_023099_02BDC0_22A8
	S1B_IW_GRDH_1SDV_20200826T202553_20200826T202621_023099_02BDC0_D53F	
highSM_DSC2	08-16	S1B_IW_GRDH_1SDV_20200816T200918_20200816T200943_022953_02B91F_C001
highSM_DSC148	07-09	S1B_IW_GRDH_1SDV_20200709T202550_20200709T202618_022399_02A835_B923
		S1B_IW_GRDH_1SDV_20200709T202550_20200709T202618_022399_02A835_B923
	08-14	S1B_IW_GRDH_1SDV_20200814T202527_20200814T202552_022924_02B83C_91EC
	S1B_IW_GRDH_1SDV_20200814T202552_20200814T202621_022924_02B83C_A691	

III. CYCLEPIX2PIX ARCHITECTURE

Our image translation network is based on two state-of-art image-to-image translation networks (i.e. Pix2Pix [12] and CycleGAN [13]). As shown in Fig. 3, the major structure of our architecture follows the cycle-consistent design of the CycleGAN, to achieve the translation between SAR and optical image (OPT). The Patch-GAN and the U-Net [26] in the Pix2Pix are employed as discriminator and generator respectively (see Fig. 4). The input and output of the two generators are the Real SAR/OPT and the Translated OPT/SAR, respectively with a size of $256 \times 256 \times 3$. The input of the discriminator is the Real SAR/OPT and the Translated SAR/OPT, and the output of the discriminator is a probability map with the shape of $32 \times 32 \times 1$, respectively. Both the generator and discriminator are assembled by a set of units of convolution filter, deconvolution filter, activation function, and concatenation. The arrows indicate the flow of the data, and the symbol ‘+’ denotes concatenation. ‘Leaky ReLU (LReLU)’, ‘Tanh’, and ‘Sigmoid’ are three activation functions. ‘Conv’ and ‘DeConv’ represent the convolution filter and deconvolution filter, respectively. ‘IN’ and ‘Dropout’ mean the operations of Instance Normalization and Dropout, respectively. The alphanumeric characters combined by ‘ngf’, ‘ndf’, ‘k’, ‘s’ are used to denote the number of generator filters and discriminator filters, the kernel size, and the stride, respectively.

The network consists of two branches highlighted in red and blue, respectively (see also Fig. 3). In the red branch, the generator G_1 aims to translate the Real SAR into the Translated OPT, and the discriminator D_1 is trained to distinguish the Real OPT and the Translated OPT. The Translated OPT will be again converted into the Reconstructed SAR by the generator G_2 . The distance between the Reconstructed SAR and the Real SAR is employed as the cyclic loss 1. A conventional binary classification loss is employed to train the discriminator D_1 , while the combination of cGAN loss 1 and cyclic loss 1 is applied to train the generator G_1 . Conversely, the blue branch is trained to translate the Real OPT into Translated SAR and then generate Reconstructed OPT. The distance between the

Reconstructed OPT and the real OPT is employed as the cyclic loss 2. The combination of cGAN loss 2 and cyclic loss 2 is utilized to train the generator G_2 . The generators of G_1 and G_2 are trained jointly, while the discriminators of D_1 and D_2 are trained separately. The style matching between generators in two branches is to fuse the spatio-temporal variation of multiple SAR images based on their Gram matrix of features maps of selected layers to avoid the potential model collapse into one specific condition.

IV. LOSS FUNCTION WITH STYLE MATCHING

The training of the generators of G_1 and G_2 and the discriminators of D_1 and D_2 are adversarial. The loss of the

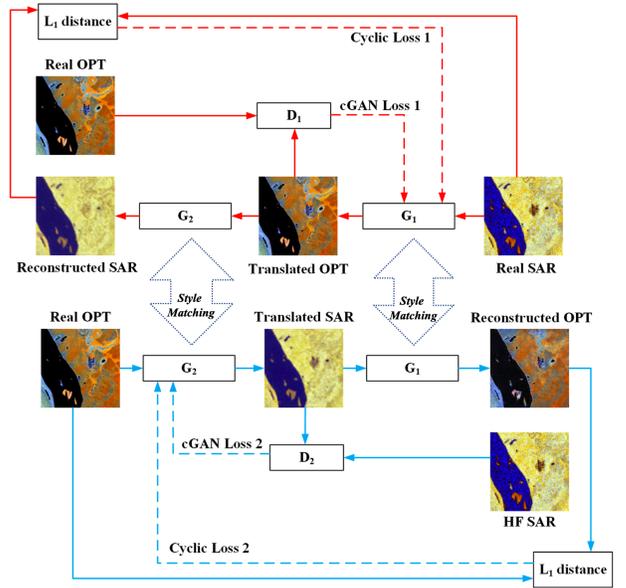


Fig. 3. Proposed CyclePix2Pix architecture.

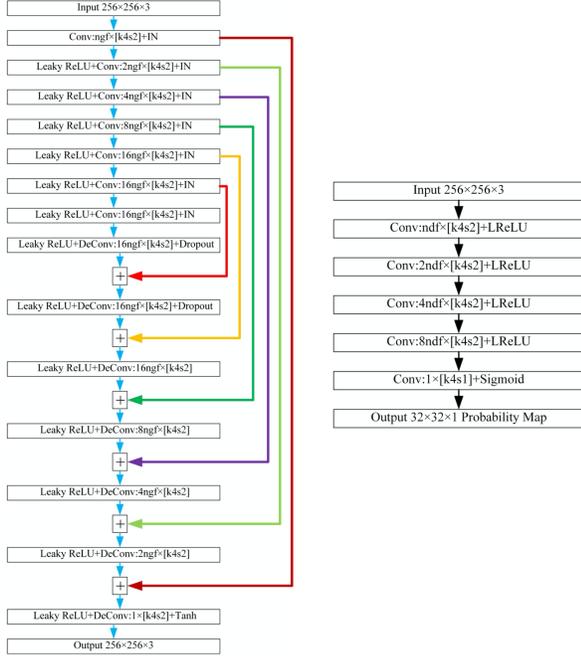


Fig. 4. Left: Generator (U-Net); right: Discriminator (Patch-GAN).

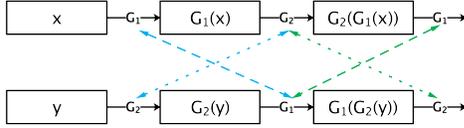


Fig. 5. Proposed style matching (SM) loss. The blue dashed line linking the G_1 denotes the forward style matching $\mathcal{L}_{\text{style}}(G_1)$, while the other blue dotted line connecting the G_2 denotes backward style matching $\mathcal{L}_{\text{style}}(G_2)$. Enhanced style matching is represented as green connection which further strengthens the cyclic translation with iterative operation.

generators can be defined as:

$$\begin{aligned} \mathcal{L}(G_1, G_2) = & \mathbb{E}_{x \in p_{\text{data}}} [\log(D_1(G_1(x)))] \\ & + \mathbb{E}_{y \in q_{\text{data}}} [\log(D_2(G_2(y)))] \\ & + \beta \cdot [\|x - G_2(y)\|_1 + \|y - G_1(x)\|_1] \\ & + \lambda \cdot \mathcal{L}_{\text{cyclic}}(G_1, G_2) + \mathcal{L}_{\text{style}}(G_1) \end{aligned} \quad (1)$$

where x and y represent the input paired SAR and OPT, and p_{data} and q_{data} are the underlying data distributions of x and y . $G_1(x)$ and $G_2(y)$ represent the Generated OPT and Generated SAR, respectively. The first term denotes the parts of cGAN loss of G_1 , and the second one represents the parts of cGAN loss of G_2 . β is a regularization weight that controls the weight of the L1 distance loss. λ is a regularization weight that controls the weight of the cyclic loss, i.e. $\mathcal{L}_{\text{cyclic}}(G_1, G_2)$ is the cyclic consistent loss in the whole network:

$$\begin{aligned} \mathcal{L}_{\text{cyclic}}(G_1, G_2) = & \mathcal{L}_{\text{cyclic}}(G_1) + \mathcal{L}_{\text{cyclic}}(G_2) \\ = & \mathbb{E}_{x \in p_{\text{data}}, y \in q_{\text{data}}} [\|x - G_2(G_1(x))\|_1] \\ & + \mathbb{E}_{x \in p_{\text{data}}, y \in q_{\text{data}}} [\|y - G_1(G_2(y))\|_1] \end{aligned} \quad (2)$$

The style matching loss (see Fig.5) adapts from texture synthesis with x and Translated x , replacing the source and the synthesized output image in [27], [28]:

$$\mathcal{L}_{\text{style}}(G_1) = \sum_{l=1}^L \frac{\alpha_l}{|x_l|} \|M(x_l) - M(G_2(y)_l)\|_F^2 \quad (3)$$

Here α_l are user parameters that weight terms in the style matching loss; $|\cdot|$ is the number of elements in a tensor, and $\|\cdot\|_F$ is the Frobenius norm. Gram matrix $M(x_l)$ is defined over any feature map of G_1 with x as input (i.e. Real SAR), while $M(G_2(y)_l)$ determines the Gram matrix of corresponding feature maps of G_1 with $G_2(y)$ as input (i.e. Translated SAR). This results in feature maps for the activations of the first L convolutional layers, which we denote as $x_1 \dots x_L$ with $G_2(y)_1 \dots G_2(y)_L$. Then we minimize a loss $\mathcal{L}_{\text{style}}(G_1)$ over the layers during the two translation processes based on G_1 . Conversely, G_2 also shows the inverse process in terms of y and $G_1(x)$ to match the style of two-generation process as (4). The style matching on G_2 focuses on the style from OPT to SAR, thus, it is recommended to use style matching for G_2 if the study case centers on translation from SAR to OPT rather than OPT to SAR, thus the $\mathcal{L}_{\text{style}}(G_2)$ is skipped here. Further, enhanced style matching is proposed based on the iterative operation (see also Fig. 5 in green dotted connection), even though it is not involved in the translation process of generators. The enhanced matching can be involved, depending on the focus of image reconstruction (style or context) during generators training in specific application.

$$\mathcal{L}_{\text{style}}(G_2) = \sum_{l=1}^L \frac{\alpha_s}{|y_l|} \|M(y_l) - M(G_1(x)_l)\|_F^2 \quad (4)$$

The losses of the discriminators D_1 and D_2 are defined as

$$\begin{aligned} \mathcal{L}(D_1) = & \mathbb{E}_{y \in q_{\text{data}}} [\log(D_1(y))] \\ & + \mathbb{E}_{x \in p_{\text{data}}} [\log(1 - D_1(G_1(x)))] \\ & + \beta_2 \cdot \mathcal{L}_{\text{SSIM}}(y, G_1(x)) \end{aligned} \quad (5)$$

$$\begin{aligned} \mathcal{L}(D_2) = & \mathbb{E}_{x \in p_{\text{data}}} [\log(D_2(x))] \\ & + \mathbb{E}_{y \in q_{\text{data}}} [\log(1 - D_2(G_2(y)))] \\ & + \beta_2 \cdot \mathcal{L}_{\text{SSIM}}(x, G_2(y)) \end{aligned} \quad (6)$$

where the first two terms of $\mathcal{L}(D_1)$ and $\mathcal{L}(D_2)$ are from discriminator parts of cGAN loss 1 and cGAN loss 2, respectively. They are augmented by the structural similarity (SSIM) loss [29]. $\|\cdot\|_1$ denotes L1 distance. β_2 is the regularization weight of $\mathcal{L}_{\text{SSIM}}$ between the paired P_1 and P_2 :

$$\mathcal{L}_{\text{SSIM}}(P_1, P_2) = 1 - \frac{2\mu_m\mu_n + C_1}{\mu_m^2 + \mu_n^2 + C_1} \cdot \frac{2\sigma_{mn} + C_2}{\sigma_m^2 + \sigma_n^2 + C_2} \quad (7)$$

where μ_m , μ_n , σ_n , σ_m and σ_{mn} denote the mean, standard, deviation and covariance of image P_1 and P_2 and respectively. The image pairs $P_1 \leftrightarrow P_2$ can be regarded as $y \leftrightarrow G_1(x)$ or $x \leftrightarrow G_2(y)$. C_1 and C_2 are set as constants of 6.50 and 58.52. Using SSIM as loss function strengthens the connection

between the generated image and the target image. To this end, the training objective is to solve:

$$G_1^*, G_2^* = \arg \min_{G_1, G_2} \min_{D_1, D_2} [\mathcal{L}(G_1, G_2), \mathcal{L}(D_1), \mathcal{L}(D_2)] \quad (8)$$

V. EXPERIMENT SET-UP

Using the same evaluation datasets and metrics, we compare our method against the baseline both qualitatively and quantitatively. The tasks include qualitative evaluation of translated images between SAR \leftrightarrow OPT, quantitative assessment on different sub-category test patches for SAR \rightarrow OPT, and semantic segmentation on translated OPT images in different set-up. The followed evaluation metrics are applied to compare the performance.

PSNR (Peak Signal to Noise Ratio) is a traditional image quality assessment (IQA) index based on (9):

$$\begin{cases} \text{PSNR}(P, Q) = 10 \log_{10} \left(\frac{(2^n - 1)^2}{\text{MSE}} \right) \\ \text{MSE} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (P(i, j) - Q(i, j))^2 \end{cases} \quad (9)$$

where **MSE** is to evaluate the squared error between images P and Q with height of H and width of W .

SSIM [29], as coefficient embodies the structural similarity between the two images and reflects the visual effect of the image. Similar with (7), the SSIM between two images P and Q can be defined as:

$$\text{SSIM}(P, Q) = \frac{2\mu_P\mu_Q + C_1}{\mu_P^2 + \mu_Q^2 + C_1} \cdot \frac{2\sigma_{PQ} + C_2}{\sigma_P^2 + \sigma_Q^2 + C_2} \quad (10)$$

“FCN-score” Although perceptual studies may be the gold standard for assessing graphical realism, we also seek an automatic quantitative measure that does not require human experiments. To this end, we adopt the popular HRNet [30] architecture for semantic segmentation on BA (burned area) and uBA (unburned area) segments. We then score synthesized OPT by the segmentation accuracy against the thresholded dNBR ground truth [31]. This segmented map can then be evaluated using standard semantic segmentation matrices.

Semantic segmentation metrics To evaluate the performance of SAR \rightarrow OPT, we use the standard metrics from the Cityscapes benchmark [4], including overall accuracy, Kappa, and mean class Intersection-Over-Union (Class IOU). In detail, the HRNet was trained based on the Sentinel-2 MSI data acquired over two sites, where two large wildfire events occurred in Elephant Hill, British Columbia, Canada in 2017 and Enskogen, Sweden in 2018, respectively.

VI. EXPERIMENT RESULTS AND DISCUSSION

A. Image translation between SAR and optical

With the enhancement of style matching, we train the CyclePix2Pix model based on a one-to-many strategy with 200 patches (including uBA and BA) remained to test the model. In Fig.6, despite spatio-temporal variations within these patches from different sub-categories, the CyclePix2Pix model could reconstruct the optical image reliably with rich contextual information such as the boundaries of lakes and irregular fire scar. Each subgraph among (a), (b), (c), and

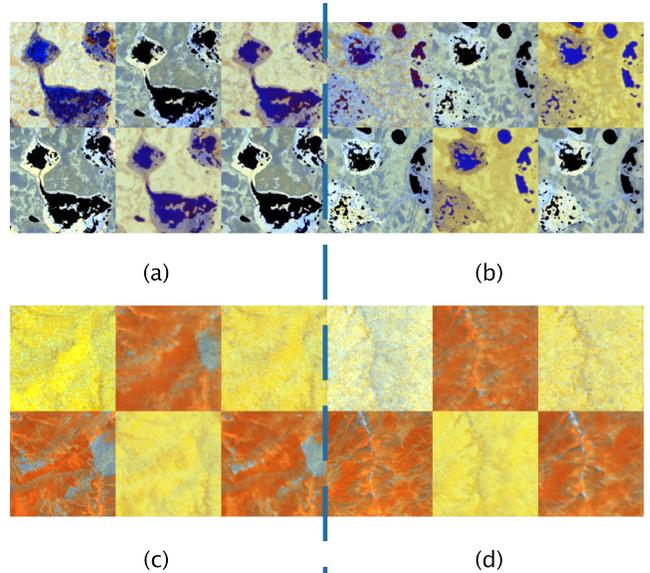


Fig. 6. Image translation results. (a) and (b) are samples without burned areas while the others contain burned area pixels.

(d) presents the Real SAR, Translated OPT, Reconstructed SAR in the first row, and Real OPT, Translated SAR, and Reconstructed OPT in the second row from left to right. No obvious variation is observed from three OPTs because of only one target for generation, while the Real SAR differs from the other two pseudo SAR because of multiple sources of SAR. Style matching enhancement for SAR \rightarrow OPT might result in the average effects in style features, which blurs the details like narrow branches in Fig.6 (b).

The average SSIM values between Real OPT and Translated OPT are 0.4767 for all 200 testing patches, and PSNR values are 19.65 dB, respectively. The performance evaluation for each sub-categories are shown in Fig.7. After applying the style matching, patches in each category can be translated into an optical image with relatively high similarity. And the categories with high soil moisture reach higher quantitative performance than the low one. Interestingly, it was observed that relative orbit 148 showed higher accuracy than orbit 2 in terms of the high soil moisture but conversely in terms of the low soil moisture condition.

B. Burned area segmentation

Fig. 8 demonstrated the burned area segmentation results based on the optical images including the original optical image and two translated images from original SAR using CyclePix2Pix with or w/o SM. It can be observed that CyclePix2Pix with style matching can improve the concentration of image translation with fewer commission errors. The translated optical image based on CyclePix2Pix can keep style features of burned areas which facilitate the potential burned area mapping.

Table. II illustrated the FCN-score based on different models including CycleGAN, CyclePix2Pix w/o SM, and Cy-

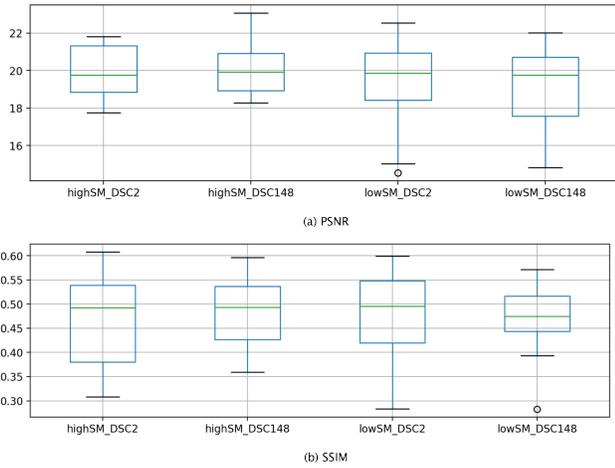


Fig. 7. Quantitative evaluation based on SSIM and PSNR on pairs of translated optical image and real one under different sub-categories.

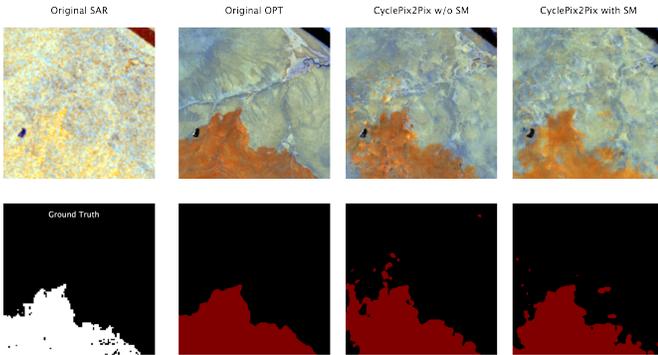


Fig. 8. Semantic segmentation based on the HRNet.

clePix2Pix with SM. As expected, semantic segmentation based on original optical images can reach the highest accuracy at the upper level due to the same source images. With the style matching process, mIoU of CyclePix2Pix increases by 0.54% than CyclePix2Pix w/o SM, reaching 86.18%. Despite the minor quantitative improvement, the visual inspection from the translated images in Fig. 8 can prove the better reconstruction performance of CyclePix2Pix with SM. On the other hand, CycleGAN as the lower baseline shows no potential to reconstruct the optical images in terms of fire disturbance changes.

TABLE II
PERFORMANCE (%) OF FCN-SCORE USING DIFFERENT MODELS.
BA-BURNED AREA CLASS; UBA-UNBURNED AREA CLASS

	Kappa	OA	mIoU	IoU (BA)	IoU (uBA)
Original	95.59	98.30	95.70	93.67	97.73
CycleGAN	13.31	87.59	48.21	8.98	87.44
CyclePix2Pix w/o SM	83.93	96.14	85.64	75.67	95.62
CyclePix2Pix with SM	84.59	96.32	86.18	76.54	95.82

VII. CONCLUSION AND FUTURE WORK

SAR becomes an alternative approach to monitor the fire-disturbed regions under dense clouds. As SAR images are different from the natural optical ones, large-scale spatial-temporal variation results in the vague synthesis in image translation from SAR to optical under complex applicable conditions. This work aims to fill the gap between SAR and optical using GAN-based methods. It helps us monitor the fire disturbance or progression using generated optical images under the cloud to avoid the potential damage we can not see. We recognize that these results are preliminary due to the limited training dataset in terms of study areas, but they indicate the potential success of GAN-based methods in fire-related analysis.

We propose a CyclePix2Pix architecture with style matching to reduce the impacts of various source images due to sensor characteristics, soil moisture, local vegetation. High soil moisture results in better translation quality than low soil moisture in SSIM and PSNR. Based on style matching, spatio-temporal variation causing these unstable generations can be minimized with acceptable performance despite the potential blurring effects existing. Accounting for the semantic segmentation in burned areas from translated optical images, the proposed approach can map most burned areas with relatively low commission and omission errors, which contributes to disaster management in future applications.

ACKNOWLEDGMENT

Thank CCAI mentorship program for the opportunity to start this work. We acknowledge the use of data from Sentinel-2 and Sentinel-1 operated by the Copernicus Programme.

APPENDIX

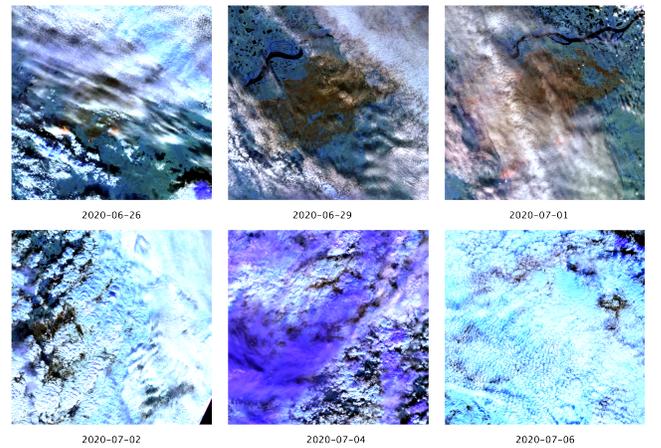


Fig. 9. Sentinel-2 optical images acquired during fire events (false-color with B12-B11-B8A in RGB composite).

REFERENCES

- [1] D. M. Bowman, G. J. Williamson, J. T. Abatzoglou, C. A. Kolden, M. A. Cochrane, and A. M. Smith, "Human exposure and sensitivity to globally extreme wildfire events," *Nature ecology & evolution*, vol. 1, no. 3, p. 0058, 2017.

- [2] G. Hugelius, J. Loisel, S. Chadburn, R. B. Jackson, M. Jones, G. MacDonald, M. Marushchak, D. Olefeldt, M. Packalen, M. B. Siewert, C. Treat, M. Turetsky, C. Voigt, and Z. Yu, "Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 34, pp. 20438–20446, 2020.
- [3] M. A. Belenguer-Plomer, E. Chuvieco, and M. A. Tanase, "Temporal decorrelation of c-band backscatter coefficient in mediterranean burned areas," *Remote Sensing*, vol. 11, no. 22, p. 2661, nov 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/22/2661>
- [4] R. Mathieu, R. Main, D. Roy, L. Naidoo, and H. Yang, "Detection of burned areas in southern African savannahs using time series of C-band sentinel-1 data," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2018-July, 2018, pp. 5337–5339. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8517838/>
- [5] M. Kurum, "C-band SAR backscatter evaluation of 2008 gallipoli forest fire," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1091–1095, 2015.
- [6] C. Quintano, A. Fernández-Manso, O. Fernández-Manso, and Y. E. Shimabukuro, "Mapping burned areas in Mediterranean countries using spectral mixture analysis from a uni-temporal perspective," *International Journal of Remote Sensing*, vol. 27, no. 4, pp. 645–662, 2006.
- [7] P. Zhang, A. Nascetti, Y. Ban, and M. Gong, "An implicit radar convolutional burn index for burnt area mapping with Sentinel-1 C-band SAR data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 50–62, 2019.
- [8] G. H. D. A. Pereira, A. M. Fusioka, B. T. Nassu, and R. Minetto, "Active fire detection in satellite imagery: a large-scale dataset and a cnn-based approach," *IEEE Geoscience and Remote Sensing Letters*, vol. (under review), pp. 1–5, 2020.
- [9] P. P. de Bem, O. A. de Carvalho Júnior, O. L. F. de Carvalho, R. A. T. Gomes, and R. F. Guimarães, "Performance analysis of deep convolutional autoencoders with different patch sizes for change detection from burnt areas," *Remote Sensing*, vol. 12, no. 16, 2020.
- [10] 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.
- [11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 3, no. January, 2014, pp. 2672–2680.
- [12] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks Phillip," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-October, 2017, pp. 2242–2251.
- [13] 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*, vol. 2017-October, 2017, pp. 2242–2251.
- [14] 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," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797, 2018.
- [15] 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.
- [16] 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. 60338–60343, 2020.
- [17] 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.
- [18] L. H. Hughes, M. Schmitt, and X. X. Zhu, "Mining hard negative samples for SAR-optical image matching using generative adversarial networks," *Remote Sensing*, vol. 10, no. 10, 2018.
- [19] 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.
- [20] G. Ji, Z. Wang, L. Zhou, Y. Xia, S. Zhong, and S. Gong, "SAR Image Colorization Using Multidomain Cycle-Consistency Generative Adversarial Network," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.
- [21] L. Li, C. Wang, H. Zhang, and B. Zhang, "SAR Image Ship Object Generation and Classification With Improved Residual Conditional Generative Adversarial Network," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.
- [22] Z. Ren, S. Member, B. Hou, and Q. Wu, "Generative Adversarial Network for SAR Image Classification," vol. 58, no. 6, pp. 3864–3880, 2020.
- [23] *Sentinel-2 MSI Introduction*, Accessed October 6, 2020, <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi>.
- [24] *Sentinel-1 SAR User Guide Introduction*, Accessed October 6, 2020, <https://earth.esa.int/web/sentinel/user-guides/sentinel-1-sar>.
- [25] G. Huffman, E. Stocker, D. Bolvin, E. Nelkin, and T. Jackson, "GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC)," *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, 2019. [Online]. Available: [10.5067/GPM/IMERG/3B-HH/06](https://doi.org/10.5067/GPM/IMERG/3B-HH/06)
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351. Springer Verlag, may 2015, pp. 234–241.
- [27] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 2015-January, pp. 262–270, 2015.
- [28] E. Risser, P. Wilmot, and C. Barnes, "Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses," 2017. [Online]. Available: <http://arxiv.org/abs/1701.08893>
- [29] D. Marmanis, W. Yao, F. Adam, M. Datcu, P. Reinartz, K. Schindler, J. D. Wegner, and U. Stilla, "Artificial Generation of Big Data for Improving Image Classification: A Generative Adversarial Network Approach on SAR Data," 2017. [Online]. Available: <http://arxiv.org/abs/1711.02010>
- [30] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *CVPR*, 2019.
- [31] C. H. Key and N. C. Benson, "Landscape assessment: ground measure of severity; the Composite Burn Index, and remote sensing of severity, the Normalized Burn Index," in *FIREMON: Fire effects monitoring and inventory system*, 2006.