Multiple-Source Remote Sensing Image Classification Using Generative Adversarial Network

N R Nantha Priya[1], D Lathisha[2]
[1, 2] Dept of ECE
[1, 2] Vins Christian Women's College Of Engineering

Abstract*- Deep learning has made spectacular breakthroughs in many computer vision tasks, such as semantic segmentation and object detection which are driven by the progress of deep convolutional neural networks (CNNs) and large quantities of available training data sets. Unfortunately, CNNs rarely generalize learned knowledge to new data sets, especially when there is a wide domain gap between the source and target images. The accuracy of remote sensing image segmentation and classification is known to dramatically decrease when the source and target images are from different sources; while deep learning-based models have boosted performance, they are only effective when trained with a large number of labeled source images that are similar to the target images. In this article, we propose a generative adversarial network (GAN) based domain adaptation for land cover classification using new target remote sensing images that are enormously different from the labeled source images. In GANs, the source and target images are fully aligned in the image space, feature space, and output space domains in two stages via adversarial learning. The source images are translated to the style of the target images, which are then used to train a fully convolutional network (FCN) for semantic segmentation to classify the land cover types of the target images. The domain adaptation and segmentation are integrated to form an end-to-end framework.*

Keywords- deep convolutional neural network, generative adversarial network

I. INTRODUCTION

Deep learning has made spectacular breakthroughs in many computer vision tasks, such as semantic segmentation and object detection, which are driven by the progress of deep convolutional neural networks (CNNs) and large quantities of available training data sets. Unfortunately, CNNs rarely generalize learned knowledge to new data sets, especially when there is a wide domain gap between the source and target images. For example, there are usually significant appearance differences between the remote sensing images acquired from different satellites. It is never expected that a pretrained model can acquire satisfactory performance when applied directly to another quite different remote sensing data set, at least in the current stage. Hence, labeled samples in each upcoming satellite image are required to train a new deep learning-based model. Although a large number of historical GIS maps are available, these vector maps are outdated soon. As a result, an extensive amount of manual work is applied to update vector maps, which can then be provided as up-todate training samples. The other problem is the access to updated high-quality vector maps is not easy. To address these problems and make the best of the available vector maps, unsupervised domain adaptation (UDA) has been proposed to reduce the domain gap between the source and target data to enable a pretrained deep learning model to achieve competitive results on unlabeled target data.

Recently introduced methods have focused on cross domain semantic segmentation from high-quality close-range or medical images. A generative adversarial network (GAN)-based feature space domain adaptation using adversarial training to align the CNN features from the source and target domains to improve the segmentation performance of target images, where only the source images had labeled samples. Following this approach, further studies addressed the domain transfer in semantic segmentation using different GAN-based adversarial training techniques in different space dimensions, such as the combination of image translation techniques (image space domain adaptation) and feature alignment (feature space domain adaptation) for lane image segmentation, image synergizing and feature adaptation for medical image segmentation, and synthetic-to-real adaptation at the output space domain.

Very recently, several GAN-based UDA methods have been shown to have some effects on processing remote sensing images obtained from different data sources. However, this area of research is still in its initial stages, and more sophisticated methods are expected to be developed for handling remote sensing data that are more complex than close-range data. For example, the work was only an application of CycleGAN in remote sensing image transfer at the image space. Another method named ColorMapGAN attempted to learn the color transformation from source images to target images. Both approaches stay in the image space domain transfer without touching the more advanced feature space and output space domain adaptation.

Remote sensing simply means obtaining information about an object without touching the object itself. It has two facets; acquiring data by a device at a distance from the object, and analyzing data of the object to interpret its physical properties. These two aspects are closely connected to each other. The basic fact in remote sensing is that different wavelength ranges of the electromagnetic spectrum is reflected or emitted from an object at certain intensity, which is dependent upon the physical and compositional attributes of the object.

Remote sensing today plays an important role in geological analysis of large areas which utilizes electromagnetic spectrum not only within the visible range but also beyond the visible range that human eye can‟t perceive. The unique spectral signatures of rocks, minerals, and other geological elements are used to map these geological elements in large areas in a short time using remote sensing data. Earth observation systems generally include infrared region of the electromagnetic spectrum, which include the visible-near infrared (VNIR) and shortwave infrared (SWIR). Further some imaging systems such as LANDSAT and ASTER cover thermal infrared (TIR) region, which is a mid-long wave infrared region in the spectrum.

TIR radiance values of objects can also be used for mapping similar to VNIR and SWIR. As useful as it may be, remote sensing like any tool requires continuously increasing improvement. Similarly advances in the technology necessitate the improvement of the methods accordingly, both in terms of accuracy and precision. Image fusion is one of the techniques that are employed to increase spatial and/or spectral resolution of remotely sensed data by fusing a high spatial but low spectral resolution image with a low spatial high spectral resolution image.

The purpose of this study is to fuse TIR and SWIR bands of ASTER with VNIR bands, while evaluating the multispectral infrared data with increased resolution for lithological discrimination and mapping using the basic image fusion techniques. Evaluation techniques will reveal which methods provide the best information and show how they compare to the original non-fused data. To accomplish these objectives, a graphical user interface (GUI) was prepared using the commercial software MATLAB and its Image Processing toolbox which contains commands and utilities that are commonly used in image processing applications.

II. LITERATURE SURVEY

This The relationship between impervious land cover and tree development is an important component to understanding urban ecological systems. While impervious surfaces are associated with degraded soil conditions, rerouted hydrological networks and urban microclimates, the overall impact of these effects on tree development is highly variable landscape. Using a fusion of airborne hyperspectral imagery and light detection and ranging (LiDAR) data, a 1.0 m spatial resolution classified land cover map (accuracy of 88.6%) was produced for the city of Surrey, British Columbia, Canada, from which landscape imperviousness was then derived. The stem heights of 1914 trees were estimated from the LiDAR data, to which species-specific height models were fit using planting dates recorded by city authorities. Having accounted for the age of the trees, the residuals from these models (i.e.: the difference between modelled and measured height) were then used as indicators of tree development. When aggregated to 0.5 km2 spatial units, negative relationships (r2 between 0.292 and 0.753) were found between height model residuals and the degree of land cover imperviousness. These relationships did not persist when examined at the individual tree level, for which imperviousness was measured within the direct vicinity of each tree using the same imperviousness map. We conclude that, imperviousness does not appear to be a significant driver of tree height variation, with broad-scale relationships likely due to correlations with other environmental variables associated with the urban-rural gradient. Some limitations, the integration of hyperspectral and LiDAR data proved to be a powerful tool for mapping imperviousness, with LiDAR metrics being particularly important for distinguishing between types of urban land cover. The drawback of this method is imperviousness in urban tree development within the city of Surrey is low.

Airborne laser scanning (ALS), often combined with passive multispectral information from aerial images, has shown its high feasibility for automated mapping processes. The main benefits have been achieved in the mapping of elevated objects such as buildings and trees. Recently, the first multispectral airborne laser scanners have been launched, and active multispectral information is for the first time available for 3D ALS point clouds from a single sensor. This article discusses the potential of this new technology in map updating, especially in automated object-based land cover classification and change detection in a suburban area. Results from an object-based random forests analysis suggest that the multispectral ALS data are very useful for land cover classification, considering both elevated classes and ground-level classes. The overall accuracy of the land cover classification results with six classes was 96% compared with validation points. Compared to classification of single-channel data, the main improvements were achieved for ground-level classes. According to feature importance analyses, multispectral intensity features based on several channels were more useful than those based on one channel. Automatic change detection for buildings and roads was also demonstrated by utilizing the new multispectral ALS data in combination with old map vectors. In change detection of buildings, an old digital surface model (DSM) based on single-channel ALS data was also used. Overall, our analyses suggest that the new data have high potential for further increasing the automation level in mapping. Unlike passive aerial imaging commonly used in mapping, the multispectral ALS technology is independent of external illumination conditions, and there are no shadows on intensity images produced from the data. These are significant advantages in developing automated classification and change detection procedures. The drawback of this method is the role of laser intensity has been relatively small in the development.

In this paper, we propose a multiscale deep feature learning method for high-resolution satellite image scene classification. Specifically, we first warp the original satellite image into multiple different scales. The images in each scale are employed to train a deep convolutional neural network (DCNN). However, simultaneously training multiple DCNNs is time-consuming. To address this issue, we explore DCNN with spatial pyramid pooling (SPP-net). Since different SPP-nets have the same number of parameters, which share the identical initial values, and only fine-tuning the parameters in fully connected layers ensures the effectiveness of each network, thereby greatly accelerating the training process. Then, the multiscale satellite images are fed into their corresponding SPP-nets, respectively, to extract multiscale deep features. Finally, a multiple kernel learning method is developed to automatically learn the optimal combination of such features. Experiments on two difficult data sets show that the proposed method achieves favorable performance compared with other state-of-the-art methods. The drawback of this method is spatial pyramid pooling (SPP-net) inevitably poses the overfitting problem.

The deep convolutional neural network (CNN) can provide excellent performance in hyperspectral image classification when the number of training samples is sufficiently large. In this paper, a novel pixel-pair method is proposed to significantly increase such a number, ensuring that the advantage of CNN can be actually offered. For a testing pixel, pixel-pairs, constructed by combining the center pixel and each of the surrounding pixels, are classified by the trained CNN, and the final label is then determined by a voting strategy. The proposed method utilizing deep CNN to learn pixel-pair features is expected to have more discriminative power. Several hyperspectral image data sets demonstrate that the proposed method can achieve better classification performance than the conventional deep learning-based method. Deep-learning models are usually heavily parameterized and enormous amounts of training data are required to ensure the performance; however, through reorganizing the available training samples, the proposed pixel-pair strategy is able to overcome this problem. A deep CNN architecture is designed with multiple layers, and then employed to learn deep PPFs, which tend to be more discriminative and reliable. The proposed testing procedure is implemented by a voting strategy based on the fact that neighboring pixels belong to the same class with high probability; the voting fashion that determines the final label makes classification performance more robust, particularly in heterogeneous regions. The drawback of this method is the execution time of CNN is much less than that of CNN-PPF for both the training and testing procedures CNN-Training(0.5), Testing(0.21), CNN-PPF-Training(6.0), Testing(4.76).

In this paper, we propose a method using a three dimensional convolutional neural network (3-D-CNN) to fuse together multispectral (MS) and hyperspectral (HS) images to obtain a high resolution hyperspectral image. Dimensionality reduction of the hyperspectral image is performed prior to fusion in order to significantly reduce the computational time and make the method more robust to noise. Experiments are performed on a data set simulated using a real hyperspectral image. The proposed approach is very promising when compared to conventional methods. This is especially true when the hyperspectral image is corrupted by additive noise. An important component of the method is dimensionality reduction via PCA prior to the fusion. This decreases the computational cost significantly while having no impact on the quality of the fused image. In the presence of noise, the dimensionality reduction can improve the result. The proposed method is compared to two methods based on MAP estimation. Experiments using a simulated dataset demonstrated that the proposed method gives good results and is also tolerant to noise in the HS image. The drawback of this method is MAP1method, which does not use PCA prior to the fusion, performs significantly worse than the other methods in the presence of noise.

A novel spatiotemporal fusion method based on deep convolutional neural networks (CNNs) under the application background of massive remote sensing data. In the training stage, we build two five-layer CNNs to deal with the problems of complicated correspondence and large spatial resolution gaps between MODIS and Landsat images. Specifically, we first learn a nonlinear mapping CNN between MODIS and low-spatial-resolution (LSR) Landsat images and then learn a super-resolution CNN between LSR Landsat and original Landsat images. In the prediction stage, instead of directly taking the outputs of CNNs as the fusion result, we design a fusion consisting of high-pass modulation and a weighting strategy to make full use of the information in prior images. Specifically, we first map the input MODIS images to transitional images via the learned nonlinear mapping CNN and further improve the transitional images to LSR Landsat images via the fusion model; then, via the learned SR CNN, the LSR Landsat images are supersolved to transitional images, which are further improved to Landsat images via the fusion model. Compared with the previous learning-based fusion methods, mainly referring to the sparse-representation-based methods, our CNNs-based spatiotemporal method has the following advantages:1) automatically extracting effective image features; 2) learning an end-to-end mapping between MODIS and LSR Landsat images; and 3) generating more favorable fusion results. The proposed fusion method, we conduct experiments on two representative Landsat–MODIS datasets by comparing with the sparse-representation-based spatiotemporal fusion model.

III. PROPOSED SYSTEM

Propose a novel end-to-end GAN-based full-space domain adaptation learning framework, which, to the best of knowledge, is the first full-space domain transfer method. It is not only applicable in land cover classification from multisource remote sensing images but also can be generic to many UDA tasks. Second, authors propose an image style transfer method that generates stylized images similar to the target domain images for improving the performance of deep learning tasks. Third, in its implementation in two different large-scale data sets, one consists of bitemporal satellite images in Wuhan city, and the other consists of open-source aerial image sets, method demonstrated superior performance for segmentation adaptation from multiple-source remote sensing images and markedly improved accuracy over current state-of-the-art methods.

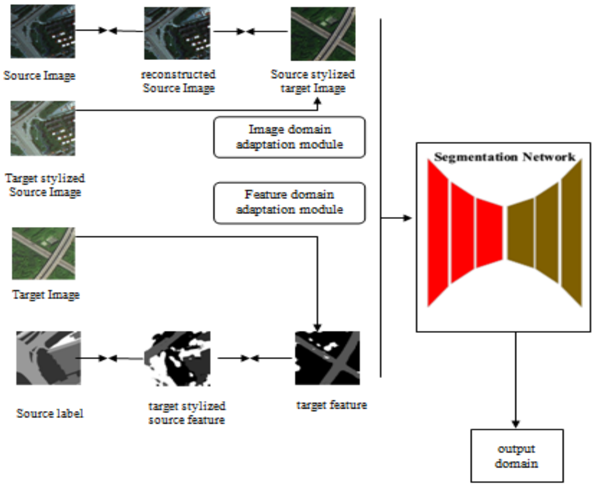

Source Image
reconstructed Source Image
Source stylized target Image
Target stylized Source Image
Image domain adaptation module
Feature domain adaptation module
Target Image
Segmentation Network
Source label
target stylized source feature
target feature
output domain

Fig 1 Block Diagram

Propose a novel end-to-end GAN-based full-space domain adaptation framework for land cover classification of the unlabeled target remote sensing images with labeled source data, i.e., existing remote sensing images with corresponding GIS data. The core idea of this approach is using the adversarial learning in all of the image space, feature space, and output space to transfer the source domain to the target domain and to realize the segmentation of the target images simultaneously. The proposed method consists of two stages.

Stage I: The image style transfer model aligns the distribution of the source images and target images both in the image space and the feature space.

Stage II: The output space domain transfer model further reduces their distribution gaps between the output spaces, and the segmentation network predicts the output results.

The main contributions of this article are as follows. First, authors propose a novel end-to-end GAN-based full-space domain adaptation learning framework, which, to the best of knowledge, is the first full-space domain transfer method. It is not only applicable in land cover classification from multisource remote sensing images but also can be generic to many UDA tasks. Second, authors propose an image style transfer method that generates stylized images similar to the target domain images for improving the performance of deep learning tasks. Third, in its implementation in two different large-scale data sets, one consists of bitemporal satellite images in Wuhan city, and the other consists of open-source aerial image sets in Potsdam and Vaihingen; method demonstrated superior performance for segmentation adaptation from multiple-source remote sensing images and markedly improved accuracy over current state-of-the-art methods.

SYSTEM MODULES

Dataset
Image domain adaptation module
Feature domain adaptation module
Segmentation network
Output domain adaptation module

MODULES DESCRIPTION

Dataset

Land cover classification from multisource remote sensing images but also can be generic to many UDA tasks. An image style transfer method that generates stylized images similar to the target domain images for improving the performance of deep learning tasks. Implementation in two different large-scale data sets, one consists of bitemporal satellite images and the other consists of open-source aerial image sets.

Image domain adaptation module

The image space domain adaptation module focuses on the transformation between the source and target images at the image-level. To translate a source image set XS with known labels YS , to the domain of a target image set XT ,which is lacking labels and needs to be classified. By doing this, a task model trained on the translated source data can be well adapted to the target data, and an unsupervised classification problem can be adequately supervised. In image space domain adaptation, follow the classic GAN structure proposed . In the output of discriminator DT is a binary value of 0 or 1, indicating that the discriminator judges the input images that are from the source domain or target domain.

Feature domain adaptation module

Feature space domain adaptation then further enhances the domain-invariance of the extracted CNN features. An adversarial learning to achieve the feature space domain adaptation. The adaptation in the feature space is not the optimal choice for semantic segmentation.

The image space domain adaptation module focuses on the transformation between the source and target images at the image-level. However, it is not an easy task for a GAN-based module to learn complete transferable knowledge and achieve the desired image style transfer, especially when the domain shift is complicated. Inspired by SIFA, authors developed a collaborative training framework for image style transfer that transforms the appearance of the images across the domains both in image space and feature space, where the feature space domain adaptation then further enhances the domain-invariance of the extracted CNN features.

As multilayer high-dimension features are difficult to align, in SIFA, low-dimensional features are extracted by compressing the high-dimensional features, i.e., the output of encoder E. Authors borrowed the idea of the shared encoder from SIFA; however, in that work, the compression is implemented by a simple rough decoder C, which is composed of a single convolutional layer and a direct 16× upsampling. Inspired by Deeplabv3+, authors added three upsampling blocks that gradually recover the spatial information to capture finer and sharper object boundaries.

3.4.4 Segmentation network

The segmentation network predicts the output results. In the test procedure, the input is only the target images to be fed into the segmentation network. A stylized image and an original image may exhibit some difference in visual appearance, regardless of whether the images were stylized or not, they should have the same semantic information outputted from the segmentation networks, which would lead to perceptual loss. The output maps that are produced by the same segmentation network and adjust the output maps of target images to be the final segmentation maps.

3.4.5 Output domain adaptation module

Output space domain adaptation consists of a segmentation network, which is supervised with labels in the source domain and predicts the segmentation outputs of both the source and target images, and an adversarial learning module, which aligns the two outputs. The output space domain alignment feature can be useful supervision to guide finer training. Here align the target-stylized source images and target images in the output space; in other words, here align the output maps that are produced by the same segmentation network and adjust the output maps of target images to be the final segmentation maps.

Here only tested the output space alignment between the SYNTHIA data set and the Cityscapes data set, both of which have high-quality close-range images and relatively simple segmentation maps. Alignment only at the output space may not be suitable for remote sensing images due to the relatively rougher image quality, complex land covers, and nonlinear radiometric changes between two data sets. In method, authors integrate output space domain adaptation into the image and feature space domain adaptation in stage I.

The basic idea for output space domain adaptation consists of a segmentation network, which is supervised with labels in the source domain and predicts the segmentation outputs of both the source and target images, and an adversarial learning module, which aligns the two outputs. In Fig. 1, the target-stylized source images GS→T (xs) and reconstructed source images U(E(GS→T (xs))) generated from stage I, as well as the source images xs and target images xt , are separately sent to a multiscale segmentation network, which is represented as S. The first task is to train the segmentation network with target-stylized source images GS→T (xs) and labels ys. The segmentation loss is formulated.

FEATURE SPACE DOMAIN ADAPTATION MODULE

The image space domain adaptation module focuses on the transformation between the source and target images at the image-level. However, it is not an easy task for a GAN-based module to learn complete transferable knowledge and achieve the desired image style transfer, especially when the domain shift is complicated. Inspired by SIFA, developed a collaborative training framework for image style transfer that transforms the appearance of the images across the domains both in image space and feature space, where the feature space domain adaptation then further enhances the domain-invariance of the extracted CNN features.

As multilayer high-dimension features are difficult to align, in SIFA, low-dimensional features are extracted by compressing the high-dimensional features, i.e., the output of encoder E. Authors borrowed the idea of the shared encoder from SIFA, however, in that work, the compression is implemented by a simple rough decoder C, which is composed of a single convolutional layer and a direct 16× upsampling. Inspired by Deeplabv3+, added three Upsampling blocks that gradually recover the spatial information to capture finer and sharper object boundaries.

OUTPUT SPACE DOMAIN TRANSFER

A segmentation map of the target images created, the map is rough and cannot be treated as the final results. On the other hand, although authors obtained target-stylized source images from stage I, a domain shift may still exist, as and reported, because the adaptation in the feature space is not the optimal choice for semantic segmentation. The output space domain alignment feature can be useful supervision to guide finer training. In this section, authors align the target-stylized source images and target images in the output space; in other words, authors align the output maps that are produced by the same segmentation network and adjust the output maps of target images to be the final segmentation maps.

Output Space Domain Adaptation Module

Output space domain adaptation, which was first mentioned, is based on the observation that the segmentation results of the source and target images usually share a significant amount of similarities in the spatial layout and local context. However, only tested the output space alignment between the SYNTHIA data set and the Cityscapes data set, both of which have high-quality close-range images and relatively simple segmentation maps. Alignment only at the output space may not be suitable for remote sensing images due to the relatively rougher image quality, complex land covers, and nonlinear radiometric changes between two data sets.

The basic idea for output space domain adaptation consists of a segmentation network, which is supervised with labels in the source domain and predicts the segmentation outputs of both the source and target images, and an adversarial learning module, which aligns the two outputs.

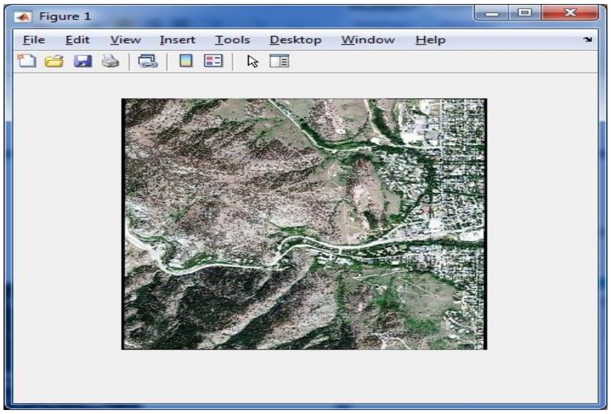
Up to this point, the data flow is forward-propagated from stage I to stage II. Authors needed gradient propagation from updated stage II back to stage I to refine the latter. Therefore, authors introduced a new loss function called the perceptual loss L. The original perceptual loss was used to measure the distance of the features. Inspired by this idea, authors proposed perceptual loss to maintain the semantic consistency between the original images and stylized images. Although a stylized image and an original image may exhibit some difference in visual appearance, regardless of whether the images were stylized or not, they should have the same semantic information outputted from the segmentation networks, which would lead to perceptual loss to tie stage I and stage II together. Especially, perceptual loss consists of two parts: 1) the gradient from updated stage II modifies the parameters of generator G to generate more target-like source images and 2) the gradient also forces generator {E, U} to output the reconstructed source images approaching to the source images.

NETWORK ARCHITECTURE

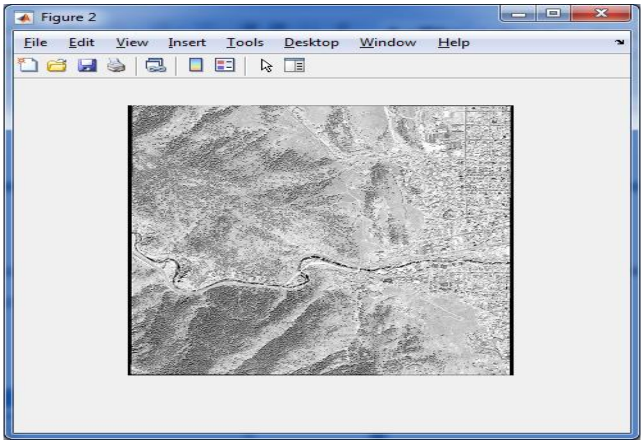
In stage I, the structure of gener follows CycleGAN, which consists of three convolutional layers, nine residual blocks, two deconvolutional layers, and one convolutional output layer to obtain the target-stylized source images. The structures of Eand Ufollow SIFA. Encoder Eis comprised of the residual connections and dilated convolutions (dilation rate = 2) to enlarge the model’s field-of-view while preserving the spatial information for dense predictions. Decoder Uconsists of one convolutional layer, four residual blocks, and three deconvolutional layers, followed by one convolutional output layer to obtain the source-stylized target images. The decoder C that we designed is composed of three upsampling blocks, each of which consists of one upsampling layer by a factor of 2; two convolutional layers each with a 3 × 3 kernel size, a batch normalization (BN) layer, and an ReLU activation; and one convolutional output layer with a 1 × 1 kernel size and without activation. The size of the output feature is the same as the original image size. The pair {E, C} constitute the feature extractor for feature space domain transfer.

In stage II, the segmentation network is modified from a multiscale segmentation network MA-FCN. The backbone and the encoder are the same, while the difference in the decoder is that we do not upsample the output feature maps of each scale and concatenate them.

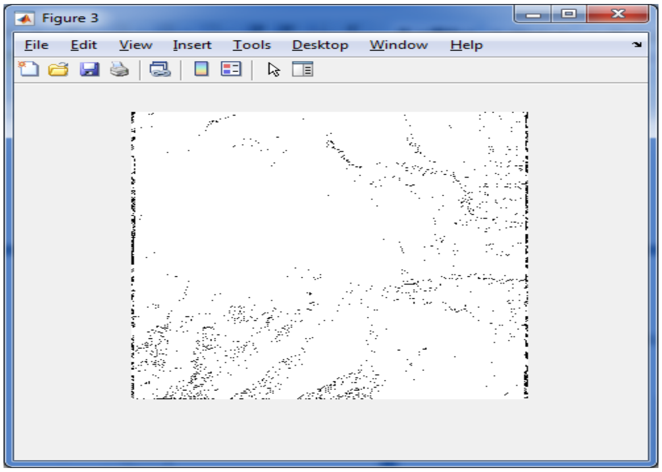
IV. SCREEN SHOTS



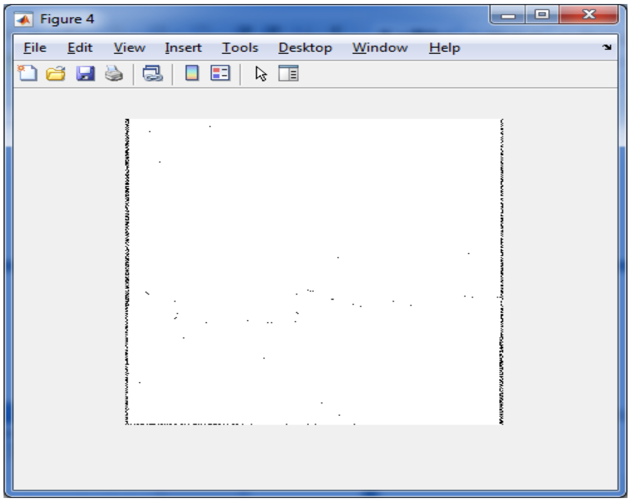
Source Image



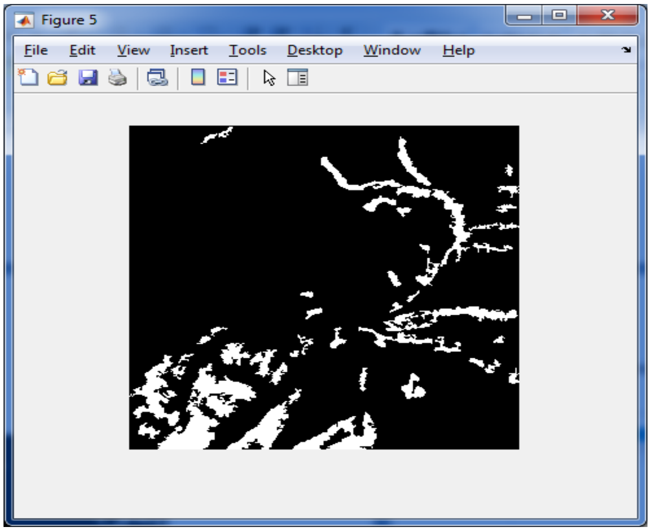
Target Image



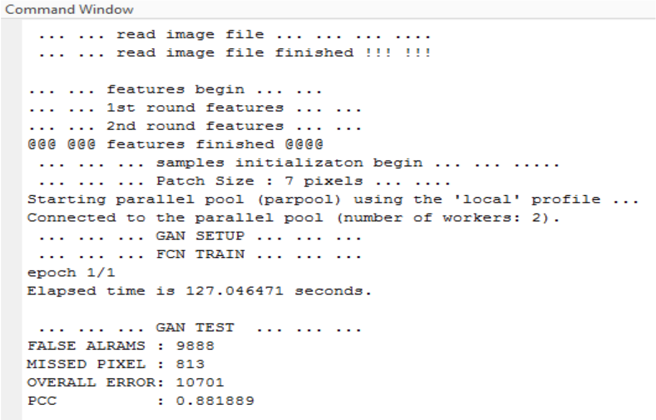
Encoded Features



Decoded Features



Output Image



```
 ... ... read image file ... ... ... ....
 ... ... read image file finished !!! !!!

... ... features begin ... ...
... ... 1st round features ... ...
... ... 2nd round features ... ...
@@@ @@@ features finished @@@@
 ... ... ... samples initializaton begin ... ... .....
 ... ... ... Patch Size : 7 pixels ... ....
Starting parallel pool (parpool) using the 'local' profile ...
Connected to the parallel pool (number of workers: 2).
 ... ... ... GAN SETUP ... ... ...
 ... ... ... FCN TRAIN ... ... ...
epoch 1/1
Elapsed time is 127.046471 seconds.

 ... ... ... GAN TEST  ... ... ...
FALSE ALRAMS : 9888
MISSED PIXEL : 813
OVERALL ERROR: 10701
PCC          : 0.881889
```

Performance Window

V. CONCLUSION

Thus, generative adversarial network (GAN) based domain adaptation for land cover classification using new target remote sensing images that are enormously different from the labeled source images. In GANs, the source and target images are fully aligned in the image space, feature space, and output space domains in two stages via adversarial learning. The source images are translated to the style of the target images, which are then used to train a fully convolutional network (FCN) for semantic segmentation to classify the land cover types of the target images. The domain adaptation and segmentation are integrated to form an end-to-end framework. Authors also proved that GAN is a generic framework that can be implemented for other domain transfer methods to boost their performance.

REFERENCES

[1] Benjdira B., Bazi Y., Koubaa A and Ouni K., (2019), “Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images,” Remote Sens., vol. 11, no. 11, p. 1369, Jun.

[2] Chen C., Dou Q., Chen H., Qin J and Heng P.A., (2019), “Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation,” in Proc. AAAI Conf. Artif. Intell, pp. 865–872.

[3] Chen L., Zhu Y., Papandreou G., Schroff F and Adam H., (2018), “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), pp. 833–851.

[4] Chen L.C., Papandreou G., Kokkinos I., Murphy K and Yuille A. L., (2018), “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr.

[5] Cordts M. et al., (2016), “The cityscapes dataset for semantic urban scene understanding,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun, pp. 3213–3223.

[6] Girshick R., (2015), “Fast R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec, pp. 1440–1448.

[7] Hoffman J. et al., (2018), “CyCADA: Cycle-consistent adversarial domain adaptation,” in Proc. Int. Conf. Mach. Learn. (ICML), pp. 1989–1998.

[8] Hoffman J., Wang D., Yu F and Darrell T., (2016) “FCNs in the wild: Pixel-level adversarial and constraint-based adaptation”, arXiv:1612.02649

[9] Huang X., Liu M., Belongie S and Kautz J., (2018), “Multimodal unsupervised image-to-image translation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), pp. 172–189.

[10] Ioffe S and Szegedy C., (2015), “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in Proc. Int. Conf. Mach. Learn. (ICML), pp. 448–456.

[11] Isola P., Zhu J.Y, Zhou T and Efros A. A., (2017), “Image-to-image translation with conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul, pp. 5967–5976.

[12] Long J., Shelhamer E and Darrell T., (2015), “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun, pp. 3431–3440.

[13] Luo Y., Liu P., Guan T., Yu J and Yang Y., (2019), “Significance-aware information bottleneck for domain adaptive semantic segmentation,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct, pp. 6778–6787.

[14] Luo Y., Zheng L., Guan T., Yu J and Yang Y., (2019), “Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun, pp. 2507–2516.

[15] Maas A. L., Hannun A. Y and Ng A. Y., (2013), “Rectifier nonlinearities improve neural network acoustic models,” in Proc. Int. Conf. Mach. Learn. (ICML), pp. 1–3.

[16] Ronneberger O., Fischer P and Brox T., (2015) “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., Cham, Switzerland, pp. 234–241.

[17] Ros G., Sellart L., Materzynska J., Vazquez D and Lopez A. M., (2016), “The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun, pp. 3234–3243.

[18] SaitoK., WatanabeK., Ushiku Y and HaradaT., (2108), “Maximum classifier discrepancy for unsupervised

domain adaptation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun, pp. 3723–3732.

[19] Tasar O., Happy S. L., Tarabalka Y and Alliez P., (2019), "ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks", arXiv:1907.12859.

[20] Tran V.H and Huang C.C., (2019), "Domain adaptation meets disentangled representation learning and style transfer," in Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC), Oct, pp. 2998–3005.