Sparse Pixel Training of Convolutional Neural Networks for Land Cover Classification

NOURELDIN LABAN$^{1}$, (Student Member, IEEE), BASSAM ABDELLATIF$^{1}$, HALA M. EBEID$^{2}$, HOWIDA A. SHEDEED$^{2}$, AND MOHAMED F. TOLBA$^{2}$, (Senior Member, IEEE)

$^{1}$Data Reception and Analysis Division, National Authority for Remote Sensing & Space Sciences, Cairo 11843, Egypt
$^{2}$Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

Corresponding author: Noureldin Laban (nourlaban@narss.sci.eg)

This work was supported in part by the Global Environment Facility (GEF)/World Bank Project “Regional Co-ordination for Improved Water Resources Management and Capacity Building” alongside the National Authority for Remote, Sensing & Space Science, Egypt.

ABSTRACT

Convolutional Neural Networks (CNN) have become the core of modern machine learning approaches. In addition to its inspiring interior design idea, the success of CNN depends mainly on two factors, the first is the availability of training data and the second is the computing power of the used devices. In the field of remote sensing, data availability is difficult and expensive. Furthermore, processing large remote sensing data to accommodate different models is a laborious process. At the same time, training data is often collected in the form of points distributed over crop fields rather than regions, which results in the scarcity of training data. To specifically address the scarcity of training points, in this paper we present a sparse pixel-based training of U-Net convolutional neural networks for land cover classification. Training images are reconstructed from the points’ collection in a random manner, they are used as an input for the convolution networks. Based on this proposed method, the amount of training data is reproduced from the different spectral signals for each land cover. We conducted extensive experiments on eight classes, using ground truth data collected from several locations in Fayoum Governorate, Egypt. The obtained results showed the superiority of the proposed method over other methods.

INDEX TERMS

Satellite image, land cover classification, U-Net convolutional neural network, sparse training, pixel-based classification.

I. INTRODUCTION

The use of satellite imagery has become one of the basic tools for sustainable development. That is why we see it in many agricultural, industrial, urban, geological fields, etc. One of these vital areas is the classification of land cover, which all depends on semantic segmentation [1]. Regarding satellite imagery, the use of classification is an essential part of understanding these images and their uses in different applications. In this manner, the decision makers receive the required information in a shorter time, lower cost and higher accuracy equal to or more than the manual method [2], [3].

The way a machine learning model is trained varies from model to model. Training methodology can be classified into two main types [4]: the first is a pixel-based training model [5] and the second is an object-based training model [6]. One is fed by training labeled pixels and the second is fed by pre-classified images. This all depends primarily on our ability to provide training data. Regarding the field of remote sensing images, the first type is considered easier and less expensive, while the second type is more difficult and more expensive when considering the large geographical areas covered by satellite images.

Numerous approaches for land cover classification in remote sensing imagery have been used. Among these methods found, commonly, in satellite image processing software and supported by much research: k-Nearest Neighbors (kNNs) [7], Random Forests (RFs) [8], MultiLayers Perceptrons (MLPs) [9], and Support Vector Machines (SVMs) [10].

In addition to these traditional methods, Convolutional Neural Networks (CNNs) have played a prominent role in recent years. Recently, the use of different CNNs architectures has become the state-of-the-art techniques used in machine learning. It aims to build parameterized multi-layer CNNs to automatically learn the best possible weights that
express all feature spaces [11]–[13]. It is, in fact, a numerical solution to large sets of simple and direct equations [14], [15]. One of the main advantages of neural networks is that they provide us with an end-to-end framework to train the classification model. Much less manual effort is required in the feature engineering process [16]. In return, we had to supply it with a large amount of training data to complete that sacred task of self-learning. From here comes the greatest challenge in this area. The problem of scarcity of labeled data.

Several models based on deep learning have been used in many fields. Among the most popular and widely used CNN models, we find Fully Convolutional Networks (FCNs) [17], SegNet [18], EfficientNet [19], U-Net [20], and extensions of these frameworks. These models rely on an encoder-decoder approach to result in an end-to-end classification model. These models have played a pivotal role in the progress we are witnessing in the modern classification. U-Net and its successors are the most successful of these methods in many areas, especially in the field of remote sensing.

The U-Net convolutional neural network is an architecture of a symmetric encoding-decoding U shaped framework, to create a segmentation mask capable of precisely locating each class [21]. This ability helps us determine the exact location of each class in the image. Through the encoding (downsampling) path to obtain the class and the decoding (upsampling) path to obtain the location [22]. Therefore, U-Net is not a mere pixel-based method, but it takes into consideration the relationship within each area of the image through the consistency between neighbors in the training process. At the same time, noise and interruptions in the data are resisted as the resulting classification is finally regulated and smoothed [23]. Although U-Net training depends mainly on a small number of labeled images. Each image has a bunch of training pixels that belongs to each class. This is a fundamental difference from the way satellite imagery is practically trained.

The remainder of this paper is divided as follows. Section II deals with the work related to the study of the land cover in satellite images, followed by section III on explaining the proposed method for training U-Net and its various details, and then section IV on the practical experiments of the proposed method, its results, and its comparison with other methods, then section V containing the conclusion of this paper.

II. RELATED WORK

Great progress has been made over the past years in the field of deep learning, especially in the subfield of image classification and interpretation. In parallel, these techniques have been extensively transferred for use in the field of satellite image processing. Remote sensing has benefited from this development through two complementary perspectives [24]. The first is to modify the techniques used to adapt unique characteristics of satellite images, either based on Convolutional Neural Networks CNNs or Recurrent Neural Networks RNNs. The second perspective is dealing with data where large datasets are built to be used in the training process of different deep learning architectures. DeepGlobe [25] and BigEarthNet [26] are powerful recent examples of these datasets. Then, combined with the great efforts made in international competitions, which provides a suitable environment for the development and exchange of ideas, and led to great development in this context.

Despite the great successes achieved by different CNN models, there are still many challenges, whether in the field of general image understanding and interpretation or specifically in the field of remote sensing imagery. Generally, the biggest and most prominent challenge in dealing with the CNNs models is the vast number of parameters required to be learned. Which in turn requires an enormous amount of labeled data. If we move to the effect of this on working with remote sensing data, specifically a land cover classification, it means a great deal of effort and money.

Wu et al. restructured U-Net, adding some new parts to create what is called Attention Dilation-LinkNet (AD-LinkNet) to be able to capture different sizes of the same semantic [1]. Cue et al. have used transfer learning relying on the pre-trained model and modified U-Net to overcome the training problem [2]. Giang et al. have studied the best U-Net Optimizer that can be used in the classification of some places in Vietnam [22]. Qiu et al. have increased the data used in training by using different dates for these images in the same location, which means that we have an increasing number of spectral signatures for the same location due to the abundance of data available from the navigating satellites [27], They realized their idea using residual convolutional neural networks (ResNet). Rußwurm et al. are based on the same multi-temporal satellite data idea in Land Cover Classification, but this time based on recurrent networks models (RNNs) [28].

We still have the idea of maximizing the already existing data and how to deal with the scattered remote sensing data of the model training. These data are often in the form of scattered points that have been collected in field trips, mostly using mobile applications. For example, Kemker et al. have reconstructed a set of satellite images and manipulated them in different ways to create synthesized images and applied the same manipulation to the labeled images in which the CNN model is trained. A fine tuning process is then applied using real, unprocessed dataset [29]. Another method is by increasing training data, through an additional step, is conducting an initial classification to the data using a method other than deep Learning. Thereafter, the CNN model is fed with the labeled data. So, it is better trained [30]. Authors do not only make an initial classification, but also used more than one classifier and vote among scores to choose the best label using a decision fusion technique.

Other methods have been also used in the classification of land cover, including SVMs, RFs, kNNs, and new CNN models. They are still the basic methods used in many satellites image displays and analysis software such as ERDAS®, ENVI®, and ESRI ArcMap®. They are actively receiving a number of studies to develop and integrate them with
CNNs frameworks [31]–[35]. Note that all these methods are pixel-based methods and do not use spatial context in their predictions. At the same time, one does not need a massive amount of training data as in CNNs models.

Authors in [24] used U-Net in their classification by utilizing training points based on data collected from the area of study and issued by the US Department of Agriculture from the mission of automatic classification of crops in the United States. They classified the study area into only two classes, whether cropland or non-cropland. They have modified the U-Net model to match these separate points by hiding all pixels in the image except those that are being trained on, or in other words not to calculate the loss except at training points. Here, an important note in this work is that the study area was very large, and therefore the number of points that were taken for training was relatively large, which improved the training process and remedied to some extent the problem of scarcity of labeled data.

It is noticeable from the aforementioned methods, whether in the method of processing training data or the machine learning process that depends on fully labeled images, and not points collected from a field, and there is no doubt that building a fully labeled image dataset is much more expensive than relying on points only. So even those who relied on synthesis images did not rely on points, which is the practical strength of our proposed method.

Meanwhile, we adapted U-Net to train synthesized images as it becomes the baseline method for semantic segmentation in many areas of machine learning. U-Net has the advantage of being a simple and straightforward Convolutional Network in addition to that it does not require large amounts of training data and thus it is more suitable for solving the problem of scarcity of training data in remote sensing field.

III. PROPOSED METHOD
In this section, we will review the most important features of the proposed method, including the method of collecting and distributing training images, then a modified architecture of U-Net, followed by a method for testing and training the subject images. FIGURE 3 illustrate the overall architecture of sparse pixel training including U-Net training to produce the final classified labeled image.

A. COLLECT, MANIPULATE GROUND TRUTH DATASET
It is known that the most common and accurate method for obtaining a ground truth dataset for satellite imagery is the collection of in-situ data through a field campaign. Unfortunately, human work is expensive and in sometimes is inconsistent. Therefore, the amount of data obtained in this way is considerably less than fulfilling the requirements of voracious methods such as deep learning methods. Thus, the first stage of collecting ground truth data, whereby expert people move on field trips and determine points on different crop fields using some useful mobile applications. These applications automatically and accurately determine the geographical locations of these points. Note that it is possible to place more than one point in the same crop field.

The manual collection process will generate a reasonable amount of land cover points that will be considered the ground truth dataset on which the training process is based. By placing the collected points at a certain date on the satellite images taken on the same date, we can get the correct spectral signature for each point. For each point collected, it will be represented as a pixel in the satellite image. It is noted that for the same crop here has many spectral signatures, as there are few differences between the different sites of the same.
crop, which enriches the training process and makes it more realistic.

B. CREATE A SYNTHESIZED IMAGE FROM SCATTERED PIXELS
The idea behind using individual points to classify land cover in satellite images is that they represent the actual fieldwork. At the same time, it considers as pixel-based image classification considering the surrounding pixels as it depends on U-Net. To produce synthesized images from individual points, we have many easy approaches. We have relied here on three basic methods to produce synthesized images as they are easy and straightforward at the same time as they simulate real images.

In the phase of collecting and building the ground truth dataset, we now have a set of arrays containing spectral signature for each class with spatial reference for each pixel. To prepare these arrays for a training process that requires the existence of two basic elements; the first is ground truth images, and the second is the classification mask that corresponds to these images. Many synthesized images are constructed using different spectral signature for each class. We have relied on several strategies to build new synthesized dataset. These strategies differ according to the method of distributing the collected points on the new image. We have adopted three strategies for the distribution of points as follows:

1) SINGLE SIZE CHESSBOARD IMAGE
As shown in FIGURE 4, First the size of the newly chessboard image is determined and at the same time the size of each square in this board, which will represent the footprint for a given class signature. The signature is chosen from the signatures bowl in a random manner, each time a new square is added. Each time a new image is constructed, it is not identical to the other composite images, because of randomly selecting squares each time. In the end we have a lot of pictures to train our U-net Architecture. Of course, while building chessboard images of signatures, chessboard images of corresponding labels are built in parallel.
2) VARIETY SIZE CHESSBOARD IMAGE
Different sizes of the chessboard widely rely upon one size as stated above. These sizes range from 1 to 16 pixels. This is done randomly so that every time a new manufactured image is produced completely different from other images in the distribution of signature pixels, as shown in FIGURE 5.

3) TEMPLATE-LIKE IMAGE
In this strategy, a similar model is adopted for the actual distribution of the main classes in the image. When a distribution similar to a template distribution is obtained using any classification method, this distribution is taken and filled with the spectral signature of the grouped points, and a corresponding mask is created using the same template. Of course, this template is large enough to generate smaller images that are used in the training process as shown in FIGURE 5.

C. MODIFIED MULTI-CLASS U-NET ARCHITECTURE
U-Net architecture was first introduced in 2015 by Ronneberger et al. [21] for medical image segmentation. The proposed model offers promising results for the image segmentation and classification process. The idea is simply a coding-decoding model, using fully convolutional neural networks that Long et al. [17] have proposed. More clearly, the original U-Net consisted of four encoding (down-sampling) blocks followed by the same number of decoding (upsampling) blocks, and in between the two U shape branches, there is a bottleneck block [36]. Note that the original U-Net was originally designed to classify grayscale images into two classes.

Several modifications have been added to the original U-Net architecture. Among the most important of these modifications, especially in the field of satellite image processing, is the multi-class image classification on the one hand, and on the other hand, using different encoding branch architectures such as various version VGG Net [37], [38] and ResNet [39]. All these modifications and additions to the encoder-decoder model made it more robust and capable of classifying extremely complex images full of semantic details such as satellite images.

1) ENCODING BRANCH
The coding branch includes two basic processes, convolution, and max-pooling, as it does in other neural networks. At the end of this process, a map with the basic features of the images is obtained automatically during training. To reach deeper layers, the residual skip connections are added without diverging the model. To make our model deeper and can distinguish between classes, we relied on a model that has more layers. Therefore, the encoder branch is based on twelve convolution operations two for each layer. At the same time, five max-pooling was added to move from one layer to another. Additionally, residual skip connections have been made to prevent any weight vanishing effect. This turns U-Net into Deep U-Net. Full details are as shown in FIGURE 8.
2) DECODING BRANCH
To reproduce a new image from the features map generated by the encoding branch, the decoding process must be performed. The most important characteristic of this process is that it is completely opposite to the encoding process, making it symmetrically mirrored. This new image is not the original image, but it is a semantic abstraction of the original image, or in other words, just general pointers for the different parts of the original image, which we call classes. What we will actually get is a mask image of the original image. To this end, we should not only reverse the encoding process, specifically upsampling operation, but also use the original image to help guide the classification process, and thus we need a concatenation process. We have three main operations during decoding, namely: upsampling, concatenation, and convolution.

3) BASIC OPERATIONS
- **CONVOLUTION** is used mainly to re-represent image features, depending on the required classification task. This process relies on different sets of filters in each layer. The proposed method uses filters with convolution matrix of $3 \times 3$.
- **MAX-POOLING** is used to downsample the input image and the successive encoding branch layers. This is done using $2 \times 2$ pooling matrix which lead to $2 \times 2 \times 1024$ features matrix cube. This cube is used to predict spatial classification map later.
- **UP-CONVOLUTION** is used to upsample the feature cube generated by encoding branch in an opposite order. The is done recursively until obtain the required classification map.
- **COPY AND CONCATINATE** is used to transmit data from the encoding branch to the corresponding position in the decoding branch. The transmit process led to achieve an accurate prediction of spatial distribution of the required classification map.
- **BATCH NORMALIZATION** is used to normalize the output of each convolutional layer. It reduces the distribution shifts of the activation values during the training process.
- **DROPOUT** is used to reduce the overfit training. Some neurons in the inner layers are dropout randomly during the training process. The use dropout rate is 0.25.
4) U-NET KEY FUNCTIONS

- **ACTIVATION FUNCTION** There are many activation functions that are used in deep learning architectures, and each one of them has its own features, depending on the type of problem you are facing. Among these functions is the Rectified Linear Unit (RELU). The most important and preferred feature of RELU is that it is computationally efficient as it quickly converges, which facilitates finding the best solutions. SO, we used it in our own structure.

- **LOSS FUNCTION** The loss function is one of the main pillars of deep learning techniques. Its good selection is an essential factor in the speed and success of CNN architectures training. Pixel-wise cross-entropy loss is the most widely used function for training U-Net architecture. This function can calculate class prediction for each pixel individually.

- **OPTIMIZER FUNCTION**

  Adadelta optimizer is one of the best optimizers used to train U-Net[22]. It is an extension to Adagrad algorithm as it reduces the extensive calculations by using a small and fixed window of past gradients. It needs a small number of epochs to reach a local minimum. We have used Adadelta optimizer to train our U-Net Model.

**TABLE 1** explains the detailed mathematical structure of successive layers of our U-Net architecture. The architecture is formed of 55 layers. It consists of 23 convolutional layers and 22 non-convolutional layers.

**D. EVALUATION METRICS**

Since U-Net creates a mask for the classification results, we should compare each classified pixel with a ground truth one. Therefore, our comparison unit here, is the pixel, which means that we have counted four sets of pixels. The first is the number of pixels whose class is classified correctly (True Positive), the second is the number of pixels whose class is not classified correctly (False Positive) and the third is the number of pixels whose class is classified (True Negative), and finally, the number of points that have also been incorrectly classified (False Negative).

We used three metrics to evaluate the proposed method and compare it to the other existing methods. These three matrices are recall, precision, and F-score. The following equations [1]–[3] explain the method adopted for each metric:

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}
\]

\[
\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}
\]

\[
F1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

**IV. EXPERIMENTAL RESULTS**

**A. STUDY AREA AND MATERIALS**

We have used data for the same area in our previous work [32], which represents sentinel 2 images of Fayoum Governorate, Egypt as shown in **FIGURE 9**, in addition to field trips data for the period from January to March 2016. Four bands of Sentinel satellite with 10 meters spatial resolution and six bands with 20 meters which were resampled
to 10 meters, so we have 10 bands as a spectral signature for each pixel. The study area is about 2,000 square kilometers and includes about 3.8 million pixels. The field trips data includes eight classes that include four agricultural crops specifically clover, sugar beet, wheat, and fruit trees, in addition to three land cover: urban, bar soil and water, and finally the background. Error! Reference source not found. shows the number of points for each crop, whether used in training or testing.

As shown in TABLE 2, the amount of data that the research team was able to collect is 508 ground truth points, which were divided in two divisions: training and testing, with a total of 254 each. This small amount of data, collected over a period of three months at a high cost, is required to be used in the process of classifying the entire area. This is what further illustrates the problem of model training in the presence of scarcity of data.

### B. IMPLEMENTATION DETAILS

We have implemented our U-Net framework using Keras based on Tensorflow [40] as a backend which is pre-installed on Linux operating system with NVIDIA Tesla K80 GPU and 12GB RAM. We used a learning rate equal to 0.0001 using 300 epochs with patch size equal to 64. Meanwhile, we used ADAM optimizer to update weights of the U-Net model to reduce backpropagation losses. To deal efficiently with remote sensing imagery.

### C. CREATING TRAINING DATASETS USING DIFFERENT STRATEGIES

As we now have several strategies for training CNN using sparse pixel training, we had to test the accuracy of these strategies to demonstrate their power in the classification process. For the first strategy, which is to use a fixed-sized tile, we tested several sizes ranging from 2 × 2 to 16 × 16. As for the second strategy, which is to use the variant tile size chessboard, we also used different sizes but on the same board. And third, regarding the template image, we used the template from SVM, and we extracted it from its values and matched it with our data in the same way in the two previous strategies.

### D. ACCURACIES OF DIFFERENT SPARSE TRAINING STRATEGIES

TABLE 3 illustrates the weighted average accuracies for recall, precision, and F-score as illustrated in equation 4. It appears clearly here that the more we use a board made of tiles of different sizes and in an overlapped manner, the more useful it will be in learning our model in different classes. So, we will find variant size better than fixed-size board, at the same time using a similar template with real spatial distribution is better than variant size.

$$\text{Weighted average} = \frac{\sum_{i=1}^{n} (x_i \times w_i)}{\sum_{i=1}^{n} w_i} \tag{4}$$

$n$ is the number of classes.

$x$ single class accuracy for each land cover.

$w$ weight of class relative to other classes.

### E. COMPARISON WITH OTHER CLASSIFICATION METHODS

The use of sparse pixel training of U-Net has shown remarkable progress in the results we obtained compared to the very small number of training points, which demonstrates the
TABLE 4. Precision, recall, F1-score of five methods for land cover classification.

| Method      | Clover | Soil | Bare | Grass | Trees | Urban | Water | Wheat |
|-------------|--------|------|------|-------|-------|-------|-------|-------|
| kNN         | 0.87   | 0.95 | 1    | 0.89  | 0.95  | 1     | 0.92  |
| RFs         | 0.89   | 0.91 | 1    | 0.86  | 0.9   | 1     | 0.91  |
| MLPs        | 0.94   | 0.88 | 0.85 | 0.9   | 0.89  | 1     | 0.93  |
| SVMs        | 0.96   | 0.74 | 0.87 | 0.89  | 1     | 0.97  |
| 1D-CNN      | 1      | 0.94 | 0.94 | 0.89  | 1     | 0.81  | 0.96  |
| SegNet      | 0.73   | 0.83 | 0.71 | 0.42  | 0.62  | 1     | 0.69  |
| U-Net       | 0.97   | 0.92 | 0.93 | 0.91  | 1     | 1     | 0.96  |

TABLE 5. Results of total accuracy and kappa for different methods.

| Method      | Total Accuracy | Kappa |
|-------------|----------------|-------|
| kNN         | 0.91           | 0.91  |
| RFs         | 0.91           | 0.91  |
| MLPs        | 0.93           | 0.93  |
| SVMs        | 0.96           | 0.96  |
| 1D-CNN      | 0.95           | 0.95  |
| SegNet      | 0.96           | 0.96  |
| U-Net       | 0.96           | 0.96  |

strength and suitability of using a sparse pixel in increasing U-Net’s ability to learn. With the diversity of land cover, whether urban or agricultural, and the convergence of spectral signature between classes such as clover and wheat, or between urban and bare soil, the differential ability was high compared to other methods.

We have compared the proposed method with other existing methods used mainly in various image analysis packages. From an experimental comparison, U-Net superiority emerged using sparse pixel training, showing the superiority of the proposed method. Specifically, the methods compared are the k-nearest neighborhoods (kNNs) - Random Forests (RFs) - Multi-Layer Perceptrons (MLPs) and Support Vector Machines (SVMs). Classification pixels have been used to train these classifiers. Meaning that the Pixels in the training set were used to train them and then their accuracies were tested using the testing set of pixels. In addition to visual verification of the result in the classified (mask) image.

The detailed configuration of comparison methods used is illustrated in our previous work [31] which investigates the best parameters for each method. The results for each one is shown as follows: for k-Nearest Neighbor (k-NN), the best number of neighbors (k) is 3. for Random Forests (RFs), the best number of trees is 50, and the best number of parallel processes is 6. For Multilayers Perceptrons (MLPs), the best number of hidden layers is 4 with 144 neurons in each one. for Support Vectors Machines (SVMs), the best gamma parameter is $10^{-0.8}$ and the best C parameter is 4000.

We also compared our suggested method with other deep learning methods, whether based on pixel-based training or image-based training. Where we used one-dimension convolutional neural network (1DCNN) and SegNet architecture. Regarding 1DCNN we have used 2 convolutional layers and 2 fully connected layers and finally softmax layer for results.1DCNN was train using 10,000 iterations. According to SegNet architecture, we have used the traditional one with our synthesized image dataset used with U-Net. SegNet was train using 700 epochs.

Through TABLE 4, which shows us the results of the different accuracy metrics represented by recall, Precision
and F1-Score for each class of land cover except for the background. The practical results for recall show that the proposed method is superior to or equal to the rest of the methods in four classes of seven. This is the same percentage in the case of precision, noting that in other cases it is close to the highest results. Whereas if we look at the F1-Score, we find that the proposed method exceeds almost all other methods in a noticeable way, as the F1-Score metric depends on Recall and precision at the same time.

**TABLE 5** shows the overall results for each method, including total accuracy and kappa statistics. This table demonstrates the superiority of our proposed approach to training with U-Net while the results of SegNet were considerably low. On the other hand, none of the approved pixel-based training methods outperformed our proposed method, including 1DCNN, as our experiments demonstrated. **TABLE 5** shows the total accuracy of each method, the proposed method has outperformed the best other methods by about 2%, with the total accuracy reaching 96%. Also, kappa statistics of our proposed method is exceeding the other with also 2%. In addition, **TABLE 6** shows the confusion matrix results for our proposed methods. This explains which land cover class was weaker in recognition and with which classes the confusion occurred. The results showed overlap between wheat and clover, also between wheat and trees.

Regarding the visual evaluation of the results, the results are very close, but some subtle differences exist, especially between closely related classes such as wheat and clover or urban and bare soil. The most important thing about U-Net is that with its ability to classify by pixels, it considers the spatial relationships between adjacent image areas in the image. This ability causes the results to be more correlated and to reduce the appearance of scattering found in the other methods, which is clearly shown in **FIGURE 10**.

**V. CONCLUSION**

In this paper, we proposed a method for training U-Net to overcome the scarcity of training data for classifying satellite images by using redistribution of training points collected from field trips. We set the collected points in more than one form, whether a simple chessboard image, multi-size chessboard or using a template-like image, to be used later in the U-Net training process. Although these distribution methods varied among themselves, they show a remarkable result. The proposed template-like image method showed a remarkable superiority in the classification of the Sentinel−2 satellite image for Fayoum Governorate in Egypt, surpassing other methods in total accuracy by about 2 %. And at the same time improved the classification of related categories such as clover and wheat or urban and bare soil.
Generally, experiments showed that the use of synthesized images consisting of collected points efficiently solved the problem of the lack of training data. In the future, other CNNs structures could be used to train on such images.

REFERENCES

[1] M. Wu, C. Zhang, J. Liu, L. Zhou, and X. Li, “Towards accurate high resolution satellite image semantic segmentation,” *IEEE Access*, vol. 7, pp. 55609–55619, 2019.

[2] B. Cui, X. Chen, and Y. Yu, “Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection,” *IEEE Access*, vol. 8, pp. 116744–116755, 2020.

[3] S. Chakraborty, J. Phukan, M. Roy, and B. B. Chaudhuri, “Handling the class imbalance in land-cover classification using bagging-based semisupervised neural approach,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1493–1497, Sep. 2020.

[4] R. Khattami, G. Mountarakis, and S. V. Stethman, “A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research,” *Remote Sens. Environ.*, vol. 177, pp. 89–100, May 2016.

[5] R. Goldblatt, M. F. Stihihmacher, B. Tellman, N. Clinton, G. Hanson, M. K. Foggescu, C. Wang, F. Serrano-Candela, A. K. Khandelwal, W.-H. Cheng, and R. C. Balling, “Using landsat and nighttime lights for supervised pixel-based image classification of urban land cover,” *Remote Sens. Environ.*, vol. 205, pp. 253–275, Feb. 2018.

[6] C. Zhang, P. Yue, D. Tapete, B. Shangguan, M. Wang, and Z. Wu, “A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 88, Jun. 2020, Art. no. 102086.

[7] N. S. Kothari and S. K. Meher, “Semisupervised classification of remote sensing images using efficient neighborhood learning method,” *Eng. Appl. Artif. Intell.*, vol. 90, Apr. 2020, Art. no. 103520.

[8] B. Li, H. Xie, X. Tong, Z. Zhang, and S. Liu, “Extracting satellite laser altimetry footprints with the required accuracy by random forest,” *IEEE Geosci. Remote Sens. Lett.*, early access, Jun. 24, 2020, doi: 10.1109/LGRS.2020.2999201.

[9] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, “A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 133–144, Jun. 2018.

[10] C. Geiß, P. A. Pelizari, L. Blickensdörfer, and H. Taubenböck, “Virtual support vector machines with self-learning strategy for classification of multispectral remote sensing imagery,” *ISPRS J. Photogramm. Remote Sens.*, vol. 151, pp. 42–58, May 2019.

[11] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsadai, “A survey of deep neural network architectures and their applications,” *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[12] N. Khattab, S. Rashwan, H. M. Ebied, and W. Sheta, “Shedeed, H. Shedeed, and W.-M. Pham, “U-Net convolutional networks for mining land cover classification based on high-resolution UAV imagery,” *IEEE Access*, vol. 8, pp. 186257–186273, 2020.

[13] X. Yao, H. Yang, Y. Wu, P. Wu, B. Wang, X. Zhou, and S. Wang, “Land use classification of the deep convolutional neural network method reducing the loss of spatial features,” *Sensors*, vol. 19, no. 12, p. 2792, Jun. 2019.

[14] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, “Weakly supervised deep learning for segmentation of remote sensing imagery,” *Remote Sens.*, vol. 12, no. 2, pp. 1–25, 2020.

[15] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, “DeepGlobe 2018: A challenge to parse the earth through satellite images,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 18, pp. 172–177, Oct. 2018.

[16] G. Sunbul, M. Charfueian, B. Demir, and V. Markl, “BigEarthNet: A large-scale benchmark archive for remote sensing image understanding,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5901–5904.

[17] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, “Fusing multiseasonal Sentinel-2 imagery for urban land cover classification with multibranch residual convolutional neural networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1787–1791, Oct. 2020.

[18] M. Rußwurm and M. Körner, “Multi-temporal land cover classification with sequential recurrent encoders,” *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 4, p. 129, Mar. 2018.

[19] R. Kemker, C. Salvaggio, and C. Kanan, “Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning,” *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, Nov. 2018.

[20] Y. Wang, B. Zhou, and P. Dai, “Semi-supervised deep learning for hyperspectral image classification,” *Remote Sens. Lett.*, vol. 10, pp. 353–362, Apr. 2019.

[21] N. Laban, B. Abdellatif, H. M. Ebeid, H. A. Shedeed, and M. F. Tolba, “Improving land-cover and crop-types classification of Sentinel-2 satellite images,” in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.* vol. 723, 2018, pp. 449–458.

[22] N. Laban, B. Abdellatif, H. M. Ebeid, H. A. Shedeed, and M. F. Tolba, “Seasonal multi-temporal pixel based crop types and land cover classification for satellite images using convolutional neural networks,” in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 21–26.

[23] A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier, “Land cover classification using extremely randomized trees: A kernel perspective,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1702–1706, Oct. 2020.

[24] Y. Ishii, H. Bagan, K. Iwao, and T. Kinoshita, “A new land cover classification method using gray-adled rough sets,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 8–12, Jan. 2021.

[25] G. Lantzanakis, Z. Mitraka, and N. Chrysoulakis, “X-SVM: An extension of C-SVM algorithm for classification of high-resolution satellite imagery,” *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 28, 2020, doi: 10.1109/TGRS.2020.3017937.

[26] L. Liebel, K. Bittner, and M. Körner, “A generalized multi-task learning approach to stereo DSM filtering in urban areas,” *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 213–227, Aug. 2020.

[27] L. El Mendili, A. Puissant, M. Chougard, and I. Sebari, “Towards a multi-temporal deep learning approach for mapping urban fabric using sentinel 2 images,” *Remote Sens.*, vol. 12, no. 3, p. 423, Jan. 2020.

[28] A. Rahklin, A. Davydow, and S. Nikolenko, “Land cover classification from satellite imagery with U-Net and Lovász-softmax loss,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 257–261.
image processing, pure pixel extractions, spectral unmixing and information sciences.

His research interests include the use of machine learning in the field of remote sensing, by clouds and shadows, machine learning and deep learning for remote sensing images to extract valuable information, restoration of contaminated images and data mining.

He is currently working as a Researcher with the Department of Scientific Computing and Information Sciences, Ain Shams University, Cairo, Egypt. He is also interested in preprocessing and processing of satellite images and their use in sustainable development in Egypt. In addition to his work at the Geographic Information Systems Center in the Municipality of Jeddah, Saudi Arabia, as a specialist in analyzing satellite images from 2002 to 2009, respectively. From 2006 to 2009, he was a Ph.D. Student with the Heinz Nixdorf Institute, University of Paderborn, Germany. Since December 2016, she has been working as the Director of the Quality Unit, Faculty of Computer and Information Sciences (FCIS), ASU. Since January 2018, she has also been working as the Head of the Scientific Computing Department, FCIS, ASU. Since August 2018, she has also been working as the Vice-Dean for Education & Students Affairs with FCIS, ASU. She has also been a Professor with FCIS, ASU, since 2019. She has coauthored about 70 research articles in refereed journals and international conferences. Her research interests include neural networks and deep learning, computer vision, machine learning, and robotics.

[39] Y. Yi, Z. Zhang, W. Zhang, C. Zhang, W. Li, and T. Zhao, “Semantic segmentation of urban buildings from VHR remote sensing imagery using a deep convolutional neural network,” Remote Sens., vol. 11, no. 15, p. 1774, 2019.

[40] M. Abadi et al., “TensorFlow: A system for large-scale machine learning,” in Proc. 12th USENIX Symp. Operating Syst. Design Implement., 2016, pp. 265–283.