RESEARCH ARTICLE OPEN ACCESS

Segmentation and Extraction of Parcels from Satellite Images Using a U-Net CNN Model

K. Arthi [1], K. Brintha [2]
[1](Department of CSE, Arunachala College of Engineering for Women, and Manavilai
Email: arthikk2000@gmail.com)
[2] (Department of CSE, Arunachala College of Engineering for Women, and Manavilai
Email: brinthak13@gmail.com)

Abstract:

The identification and division of parcels within satellite images form a crucial aspect of image interpretation. Successful segmentation of these parcels holds the potential to offer valuable insights for environmental preservation, agricultural advancements, and urban development. In this study, we developed a U-Net convolutional neural network model using the Tensor flow framework specifically tailored for satellite image parcel segmentation. To bolster the model's adaptability, a specialized data augmentation strategy was devised during the training phase. Evaluation metrics such as Intersection Ratio, Recall, and the Kappa coefficient were employed, resulting in a remarkable Kappa coefficient of 0.934 for the final model, surpassing the performance of commonly utilized methods like random forest and convolutional neural networks in satellite image segmentation. However, some segmented image areas exhibit incompleteness, necessitating improvements in image connectivity. The suggested approach demonstrates the capability to finely divide high-resolution satellite pictures, serving as a valuable benchmark for future investigations into segmenting satellite images.

Keywords — **Segmentation, Extraction, deep learning, U-Net CNN.**

I. INTRODUCTION

Satellite image analysis holds significant importance in image processing, particularly within military, and environmental sciences [1]. The advancements in the technology of space have granted access to numerous high-resolution satellite images, presenting fresh challenges for their effective interpretation. Prior to the emergence of deep learning, traditional methods for satellite image segmentation relied on digital image processing, topology, and mathematics [2].

These methods mainly comprised Segmentation methods that rely on setting thresholds and methods that detect edges within an image to delineate different regions. The former, while straightforward and effective, solely considers pixel gray value features and often overlooks spatial features, rendering it susceptible to noise and lacking robustness [3]. On the other hand, edge detection-based segmentation offers precise edge localization and rapid processing. However, it struggles with maintaining edge continuity and closure, resulting in numerous fragmented edges in high-detail regions. Consequently, it faces difficulties in forming cohesive larger regions and isn't suitable for segmenting high-detail areas into smaller, more coherent fragments.

Lately, there has been rapid progress in deep learning, leading to the emergence of various highly effective algorithms for semantic segmentation in images. Common among these are FCN,U-Net [4]. The FCN algorithm, while widely used, tends to yield less precise results as it disregards relationships of pixel-to-pixel and lacks consistency [5-6]. Seg-Net and U-Net employ different strategies for up sampling feature maps in image segmentation. Seg-Net uses de-pooling techniques to maintain detailed information, while U-Net employs a direct connection between the encoder and decoder feature maps, forming a trapezoidal structure. The innovative aspect of U-Net lies in its

jump-connected architecture, which facilitates the recovery of lost correlations between encoding and pooling in the decoder stage. This mechanism significantly enhances boundary segmentation accuracy and expedites training speed [8].

This study delves into evaluating the performance of the U-Net CNN model for satellite image segmentation. We put this model to the test using experimental data obtained from the "Satellite Image Parcel Segmentation" dataset released for the 2020 CCF BDCI competition. Our experiments reveal that the final model demonstrates substantial enhancements in the test set. The resulting intersection ratio (IoU), recall rate (Recall), and kappa coefficient significantly outperform those achieved by commonly employed methods such as the random forest algorithm and conventional convolutional neural networks in satellite image segmentation.

II. METHODOLOGY

A. Model for U-Net Neural Network

CNN are extensively utilized in segmentation of image and edge detection. U-Net, an advancement over FCN, boasts a well-defined network architecture characterized by a "contract-expand" structure [9]. Initially, it employs convolution for down sampling, extracting layer-specific features, and subsequently up samples to derive pixel-specific image types. The "contraction" phase primarily diminishes feature dimensions and gradually reduces parameters via pooling.

In U-Net, the fusion technique involves stitching together features along the channel dimension, creating denser, combined features. On the other hand, FCN utilizes a fusion method that involves the addition of corresponding points without forming thick features [10].

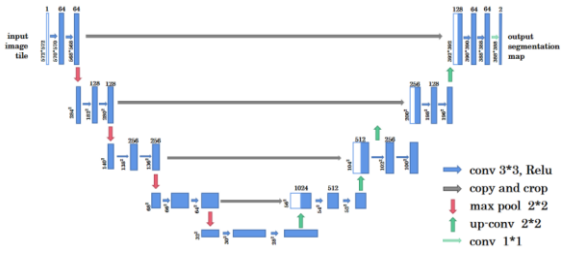

Fig .1. Structure Diagram of U-Net Structure

B. Structure of the model

TheU-Net architecture used comprises 23 convolution layers 4 down sampling layers, 4 up sampling layers, and integrates 4 skip connections connecting the feature maps from deep and shallow networks, demonstrating a structure named "contract-expand". Each down sampling step in U-Net involves two convolution layers, followed by a pooling layer. The activation function utilized is ReLU , employing zero-padding (padding='same') to address any missing data, and the weight initialization mode is set to 'he_normal'

Within the U-shaped architecture, the fusion of feature information from both deep and shallow networks is executed through the Concatenate function, ensuring the preservation of image dimensions. In this study encompasses two key points:

1. Effective Utilization of Labeled Samples: U-Net maximizes the utilization of corresponding labeled samples, allowing for highly accurate segmentation even with a limited number of training images. Leveraging its fully convolutional neural network design, U-Net excels in achieving precise segmentation.

2. Binary Classification Proficiency: The U-Net model demonstrates robust performance in binary classification tasks, further supporting its efficacy for the intended objectives of this research.

During training, the final energy function involves the utilization of both the soft-max function and the cross-entropy function. The Soft-max function is defined as.

$$p_k(x) = \frac{e^{(a_k(x))}}{\sum_{k1}^{k} . e^{(a_k(x))}} \qquad (1)$$

This paper employs the Adam optimizer, setting the learning rate at '1e-4' based on empirical findings and results.

The Adam Optimizer calculates the update step by considering both the mean and uncentered variance of the gradient. Its primary advantages lie in well-interpreted hyperparameters that often require minimal adjustment. It proves effective for scenarios with sparse or noisy gradients and is often regarded as a default optimizer with superior performance. In the context of image boundaries,

Adam optimizer addresses boundary blurring issues and enhances model accuracy. Additionally, controlling the learning rate helps prevent overfitting, making Adam optimizer well-suited for the specific problem addressed in this paper, considering its practical implications.

III. DESCRIPTION OF EXPERIMENT AND METHODS FOR DATA IMPROVEMENT

C. Description and Statistical Overview of the Dataset

The information utilized "Satellite Image Parcel Segmentation." This dataset comprises 130,000 satellite images, each with a resolution of 2.5 meters per pixel as depicted in Fig 2, formatted as JPG. Correspondingly, the dataset includes label files equal in number to the image files.

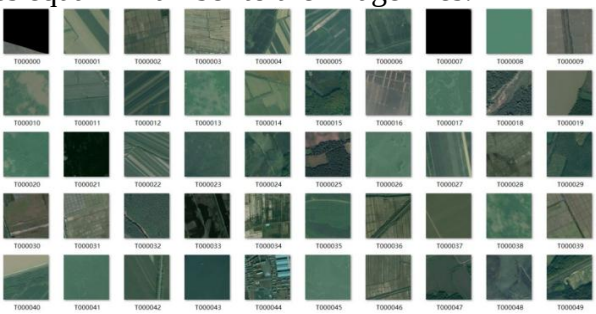

Fig .2. Satellite images

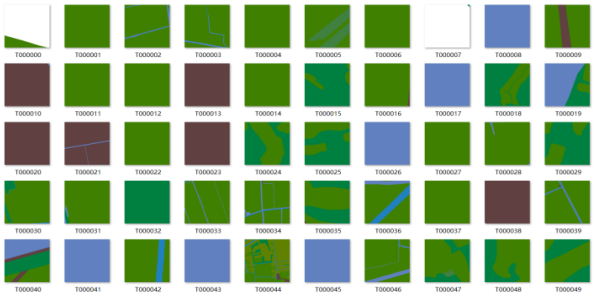

Fig .3. Labels assigned to specific data points within satellite image plots

The dataset is categorized into seven categories: forest, farmland, water, road, grass, building and others, with their distribution depicted in Figure 4.

Prior to training the model, it's crucial to comprehend the distribution of these categories within the dataset. All labelled files were analysed by counting the pixel points within the labelled areas, and the findings are outlined in Table I. Upon reviewing Table I, an imbalance in the sample data is evident, with cultivated land representing the highest percentage at 50.87%, while roads account for merely 0.35% of the pixel points.

Addressing this issue of imbalanced distribution among the dataset's sample categories stands out as a significant focus of the experiment. Table I also presents the association between categories and pixel values in the tag file, as well as the proportion or percentage of each category., along with the percentage representation of each type of tag file.

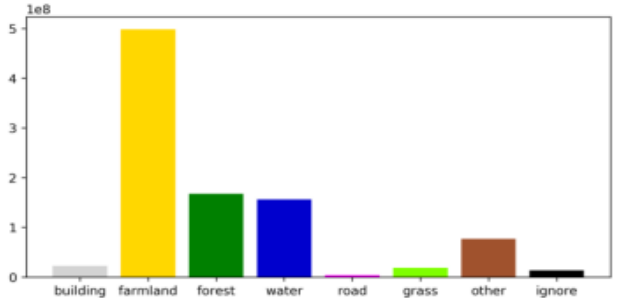

Fig .3. Information regarding the statistical distribution or characteristics of pixel labels within a dataset.

TABLE I
THE ASSOCIATION BETWEEN CATEGORIES AND PIXEL VALUES IN THE TAG FILE, AS WELL AS THE PROPORTION OR PERCENTAGE OF EACH CATEGORY

Category	Label	Percentage %
Building	0	2.79
Farmland	1	50.87
Forest	2	17.87
Water	3	17.74
Road	4	0.35
Grass	5	1.96
Other	6	7.38
Ignore	255	1.03

For enhanced annotation visualization, three images that are used in training are presented in this study as follows: gray for building, yellow for farmland, forest is represented as green, blue for water, pink for road, cyan for grass, brown for other categories, and black for ignored areas.

D. Methods to Enhance Data

Diverse strategies for augmenting data are utilized to bolster the model's ability to generalize and endure varied conditions, especially when the dataset size is constrained. These methods include:

1. Up-and-down flipping: Provides rotation invariance to the model.
2. Cropping and stretching: Introduces size invariance to the model.
3. Adding noise: Prevents the model from learning irrelevant high-frequency features and minimizes overfitting.

Given that the dataset already consists of 130,000 images, further expansion could significantly increase training time. Hence, a strategy of random data augmentation during data loading is adopted for experimentation. Specific strategies implemented are:

1. Random rotation: Rotation within an angle span of [0, 45°].
2. Horizontal and vertical position shifting: Translation within a distance range of [0, 0.1].
3. Staggered tangent transformation: Maintains one coordinate while adjusting the other in proportion to the vertical distance from the point to the axis.
4. Zooming: Equal expansion or contraction in both directions of the image.
5. Random horizontal and vertical mirroring.
6. Filling empty image spaces: Using a reflection mode to maintain image consistency.

When images undergo flipping or scaling transformations, the corresponding labels are transformed in a similar manner to maintain alignment.

IV. INVESTIGATION AND ANALYSIS THROUGH PRACTICAL STUDY

The setup in experiment involves a server running a 64-bit Windows 10 operating system, equipped with an NVIDIA graphics card boasting video memory of 16GB. Specifically, the hardware used is the GPU GeForce GTX1080Ti, while the software consists of Python 3.6 and Tensorflow 2.4.3.

For assessing accuracy, this study employs several key metrics: Intersection over Union (IoU), recall, and the Kappa coefficient. IoU, commonly used in semantic segmentation, gauges the proportion of overlap between predicted and actual values in relation to their combined area.

Recall computes the ratio of correctly classified pixels to the total number of pixels in a category. TP denotes instances where the network correctly predicts a positive sample, FP represents instances where a positive prediction is incorrect, and FN signifies instances where a negative prediction is inaccurate.

The evaluation methodology includes assessing the model's accuracy that are acquired using the coefficient of Kappa— which is used as a tool for evaluation of accuracy. This matrix, being a 2x2 representation in this study due to the dichotomous classification problem, shows correct and misclassified pixels for each class. The Kappa coefficient, ranging between -1 and 1, indicates the level of agreement between predicted and actual data, with values exceeding 0.8 suggesting strong classification. The formula for the Kappa coefficient utilizes the confusion matrix size and the total number of pixels.

In the preprocessing stage, the enhanced images aren't directly fed into network. Instead, the images that are given as input and their respective labels undergo normalization and mapping and gets converted to standardized format.

This paper outlines a training data generator tasked with producing training and validation sets while the model trains, upholding a 4:1 ratio. Furthermore, a separate test data generator manages the preprocessing of images that are being used in testing.

Several parameters are established for the model, including dropout set at 0.20, a learning rate of 1e-4, a batch size of 8, 50 epochs for training. After training, the validation set samples are used for prediction, generating various accuracy metrics: IoU , Recall rate, and Kappa coefficient, which are subsequently used for providing the accuracy.

The images that are being predicted produced by the final model, trained, exhibit an intersection ratio reaching 91.3%. Fig. 4 displays the prediction effect for some images. Observing the figure, it's evident the model being trained showcases robust segmentation capabilities. The delineation between different types displays notable accuracy, effectively addressing the issue of uneven sample sizes to a considerable amount.

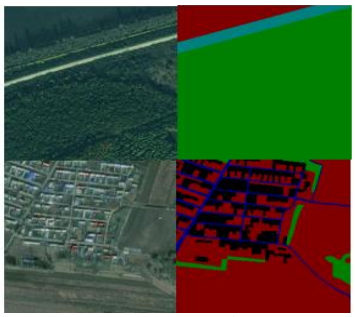

Fig .4. Satellite Images Classification by U-Net

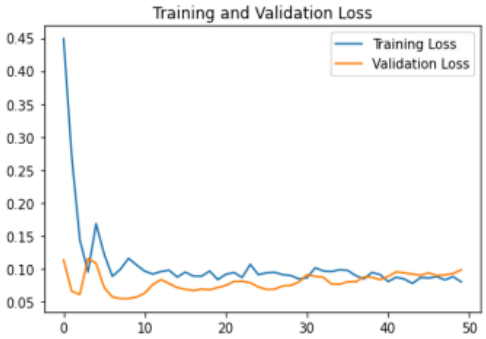

Fig .5. Training Loss and Validation Loss

The model described in this paper was trained for 50 iterations, reaching a concluding accuracy of 94.02%. The depicted figures in Fig. 9 showcasing the loss of both the training and validation sets per training round signify a well-iterated training process.

To thoroughly assess the methods outlined in this paper, two supplementary comparative experiments were carried out:

1. Random Forest Model: The random forest algorithm is extensively used in satellite image segmentation, relying on classification recognition. It generates multiple decision trees from randomly chosen data subsets. By using a voting mechanism, this algorithm categorizes samples based on the most frequently chosen class. Its effectiveness stems from its ability to diminish bias and variance, resulting in more accurate feature classification outcomes compared to individual decision trees. Additionally, it demonstrates resilience against missing data and noise, albeit with increased complexity.

2. Unaltered Convolutional Neural Network (CNN) Model: Trained without employing data enhancement techniques. The backbone network's pre-training dataset remains consistent across all three sets of experiments.

This observation suggests that to a certain degree, the issue of unbalanced sample distribution can be mitigated during training with the approach presented in this paper. Consequently, the model derived from this method is notably more robust, thereby demonstrating finer segmentation abilities.

V. CONCLUSIONS

In this study, the segmentation method namely satellite image segmentation, based on the U-Net CNN, was constructed within a deep learning framework. Following the model's construction, it underwent training in the "Satellite Image Parcel Segmentation" dataset, integrating a random data augmentation technique during the training stage.

The conclusive model displayed notable enhancements when evaluated on the test set. Metrics such as Intersection over Union, Recall rate, and Kappa coefficient demonstrated substantial enhancement compared to commonly used methods like the random forest and conventional CNN for satellite image segmentation. Notably, it achieved the KAPPA accuracy reached 0.934, underscoring of the robust segmentation capabilities of the model developed in this paper.

However, certain limitations were observed in this method. Some regions displayed insufficient continuity in the predicted images, resulting in incomplete segmentation outcomes.

Future efforts will concentrate on improving the consistency of segmentation results and tackling the issue of model overfitting as the primary areas of focus. Improving these aspects stands as a crucial step toward enhancing prediction accuracy.

REFERENCES

[1] Van der Meer, Freek. "Remote-sensing image analysis and geostatistics." International Journal of Satellite 33.18 (2012): 5644- 5676.

[2] Sheykhmousa, Mohammadreza, et al. "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review." IEEE Journal of Selected Topics in Applied Earth

Observations and Remote Sensing 13 (2020): 6308-6325.
[3] Ma, Lei, et al. "Deep learning in Remote Sensing applications: A meta-analysis and review." ISPRS journal of photogrammetry and remote sensing 152 (2019): 166-177.
[4] Bera, Somenath, and Vimal K. Shrivastava. "Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. " International Journal of Remote Sensing 41.7 (2020): 2664-2683.
[5] Chen, Guanzhou, et al. "SDFCNv2: An Improved FCN Framework for Remote Sensing Images Semantic Segmentation." Remote Sensing 13.23 (2021): 4902.
[6] Schuegraf, Philipp, and Ksenia Bittner. "Automatic building footprint extraction from multi-resolution remote sensing images using a hybrid FCN." ISPRS International Journal of Geo-Information 8.4 (2019): 191.
[7] Abdollahi, Abolfazl, Biswajeet Pradhan, and Abdullah M. Alamri. "An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images." Geocarto International 37.12 (2022): 3355-3370.
[8] Feng, Wenqing, et al. "Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a super pixel based conditional random field model." IEEE Geoscience and Remote Sensing Letters 16.4 (2018): 618-622.
[9] Abdollahi, Abolfazl, et al. "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review." Remote Sensing 12.9 (2020): 1444.
[10] Abdollahi, Abolfazl, Biswajeet Pradhan, and Abdullah M. Alamri. "An ensemble architecture of deep convolutional Seg-net and U-net networks for building semantic segmentation from high-resolution aerial images." Geocarto International 37.12 (2022): 3355-3370.
[11] Gao, Lin, et al. "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network." Remote Sensing 11.5 (2019): 552S.