A Modification Method Based on U-Net for The Distorted Pseudo Edge of Aerial Initial Orthophoto

JIA XIONG¹, JIANXIN HU², JIANCHENG Hu³

¹Chongqing Normal University Affiliated Middle School, Chongqing, China
²CCTEG Chongqing Engineering Co., Ltd., Chongqing, China
³Tianshui Huatian Technology Co., Ltd., Gansu, China

Abstract. The images captured by UAV camera have serious non-perspective distortion, and the overlap rate of heading and side direction is high. Only about 30% area of the image is available. The characteristics of small image frame and small ratio of base to height also lead to of model connection. In addition, the accuracy of image segmentation based on feature extraction is far from enough. Although the classical convolutional neural network can achieve effective image segmentation and edge calculation, but the resolution is declining in the process of forward propagation, which makes it difficult to achieve accurate segmentation of building edge when only using the features of the last layer. The above problems are the main reasons for the poor accuracy and serious distortion of the final synthetic aerial orthophoto image. To solve this problem, this paper proposes a U-Net based method to calculate and correct the distorted pseudo edge of aerial orthophoto. The object of study is the initial Orthophoto Image which is not synthesized by aerial photography. Firstly, based on the idea of U-Net, a neural network model with excellent performance in the field of image segmentation, the symmetrical network structure is used to fuse the high-dimensional and low-dimensional features of the depth network to restore the high-fidelity real boundary. Secondly, before the true value output, for the distorted features, brown method is used to find the superposition constraint positions of ideal feature points and corrected feature points, calculate the pseudo edge between distorted and undistorted, extract and prune, and retain the undistorted true value region. Finally, nested overlay and constraint detection are performed by combining the DEM of satellite images with the combined aerial orthophoto results. In the research and test, the detection accuracy statistics of internal industry encryption points and field control points with different scale accuracy are adopted, and the total coverage area is 0.5km², more than 4000 building target data sets. The results show that the DOM detection error of the new aerial composite image and satellite image is less than 3m and 9m, which shows that the edge calculation and correction of aerial orthophoto composite image based on U-Net is efficient and feasible.

Key words: U-Net, Aerial orthophoto, Distorted pseudo edge

¹ Corresponding author e-mail: olivejiajia@gmail.com
² Ordinary author e-mail: hu760303672@gmail.com
³ Corresponding author e-mail: 1370468574@qq.com
1. Introduction

In the past few "Five-Year Plan", China has basically established the national, provincial and District Street basic geographic information database, and on this basis, it has been improved to a certain extent. However, because the data update cycle of the system library is once a year, the map information is not updated in time, so the data acquisition of the construction site usually uses UAV to take the original film, and synthesize a complete orthophoto map in the later stage. However, due to the UAV image, the middle is small, the edge is large, the non perspective distortion is serious, and the offset can reach 20-40 pixels. In photogrammetry, the overlap rate of heading and side direction is high, and the ratio of unavailable image area is large. In addition, another feature of UAV image is small image frame and small base to height ratio, which will lead to a larger amount of seam work of orthophoto Image and more relative models which increase the workload of model switching and model edge joining. At the same time, the small base to height ratio also leads to the lack of stability of the aerial triangulation, which further aggravates the degree of distortion. In order to solve the distortion problem of the original image, we try to prune the peripheral distortion area of the distorted pseudo edge by combining Brownian algorithm and U-Net network structure, extract the available true value boundary and retain the distortion area. Because the distorted region is no longer involved in image synthesis, it is hopeful to improve the efficiency of orthophoto synthesis.

The full convolutional neural network (FCN) [9] improves the boundary accuracy to a great extent. In this model [2], the full connected layer of classical CNN is removed, and the features of the last layer are deconvoluted to generate the output layer consistent with the input image. At present, FCN model has been widely used in buildings. Combined with the decreasing resolution of CNN in the process of forward propagation, the segmentation results only using the features of the last layer tend to have low edge accuracy. Many subsequent image segmentation oriented models further extend the end-to-end idea of FCN, including ResNet [10], SegNet [11], DeconvNet [12] and U-Net. Taking the U-Net as an example, the model not only achieves the resolution consistency of the output layer, but also integrates the low dimension and high dimension features of CNN network through the symmetrical result design, so as to achieve higher precision image edge feature segmentation effect in orthophoto. The resolution of the initial image captured by UAV is basically consistent with that of the input image, and the high latitude and low dimensional features of CNN are fused through the symmetrical structure design to achieve high-precision segmentation effect.

In this paper, we use DJI wizard 4 pro, a 4K high-definition UAV device with a pixel ratio of
1920 * 1080p, to shoot the construction site of Evergreen International Cultural City, Chongqing, China. Through the analysis of Orthophoto Image synthesized by traditional algorithm, it is found that the more concentrated the feature points are, the worse the mosaic quality is. In this paper, taking the original spiral roof building image taken by UAV as an example, we try to use U-Net high-precision to extract the building features of the original aerial Orthophoto Image, restore the high fidelity real boundary, calculate the distorted and undistorted pseudo boundary according to the distortion features, extract and prune, and retain the undistorted true value area, and finally combine with the satellite image digital elevation model to carry out superimposed constraint test.

![FIG 2. Orthophoto Image of the construction site of Evergreen International Cultural City by DJI wizard 4 Pro](image)

2. Introduce Relevant Theories

CNN has been widely used in feature extraction, but it cannot participate in distortion correction on the premise of high fidelity feature extraction. This type of research is still in exploring stage. This section will compare FCN model[17] and U-type convolution network in image segmentation, and finally introduce the satellite image digital elevation model (DEM).

2.1. Fully Convolutional Network (FCN) and U-Shaped Convolutional Network (U-Net)

The purpose of full convolution network is to improve the performance of classical CNN model in semantic segmentation. Before that, CNN model has achieved the most advanced level of classification accuracy in the field of target classification, but this classical model structure is generally only applicable to image classification and regression, often connecting multiple fully connected layers after several convolution layers and pooling layers, so as to map the features generated in the convolution layer into an n-dimensional vector[6]. However, for the orthophoto feature detection and semantic segmentation of distorted pseudo edge class, we need to obtain the classification results of each pixel in the image.

In view of this, FCN model makes a deep improvement on the classic CNN. As shown in Figure 2, compared with the network structure of the classical model, the biggest difference of FCN is that after the convolution layer at the end, it does not use the full connection layer to generate a fixed length feature vector, but uses the deconvolution layer to up sample the high-dimensional feature image generated by the convolution layer at the front end, so as to restore it to the same resolution as the input image, which effectively retains the spatial information of the input image, and saves the need for data processing due to the complex window by window calculation process, the processing efficiency of CNN has been greatly improved[4].
Although one deconvolution in FCN network structure can generate the output layer of the original image size, but the segmentation is often too smooth to restore many details. This is mainly because the input image has been pooled many times before the final feature image is generated. The solution proposed by FCN model is to fuse the lower level features in the feature pyramid[4] with the deconvoluted up sampling results and apply them. This method has been verified to improve the segmentation accuracy. The U-Net model further extends the idea of FCN, which is the fusion of high-dimensional and low dimensional features. As shown in Figure 3, in the network structure of U-Net, the input image first passes through several convolution layers and pooling layers to obtain a high-dimensional feature map with low resolution. After a series of anti convolution layers, the feature map corresponding to the original feature pyramid is generated step by step, and the final output and input image resolution are consistent. In the process of upsampling the high-dimensional feature graph, the dimension of the feature graph will be halved with each deconvolution operation. Before the next upsampling, these reduced features will be fused with the corresponding level feature graph in the feature pyramid by matrix cascading. The fused features not only contain the abstract data at the top of the pyramid, but also inject the low-level data Detail information extracted from each level. If the pyramid is inverted, this symmetrical network structure is close to a “U” shape in shape, so it is called U-Net[4], as shown in Figure 4.

\[ y = \frac{x-k}{\text{stride}} + 1(1) \]

Where, the size of the output image is obtained by the formula.

The mean square error function of the output layer is

\[ L(x, y) = \frac{(x - y)^2}{2} \quad (2) \]

The loss function formula of U-Net model for multiple training samples is

\[ J(w, b) = -\frac{1}{m} \sum_{i=1}^{m} [x_i \log y_i + (1 - x_i) \log (1 - y_i)] \quad (3) \]
2.2. Distorted Pseudo Edge

When wide angle camera is used, lens distortion should be considered to keep the linear imaging process of projection. Adial, decentering and thin prism distortions are the main three types of distortions which are essentially considered, as shown in Figure 7.

Considering the distortion characteristics of UAV lens, we hope to deal with the image distortion from the source. For an aerial original image, the real boundary is the boundary of the whole image. The distorted boundary is relative to the boundary of the whole image. It distinguishes the distorted region from the undistorted region, which is called distorted pseudo edge. Where is the distorted pseudo edge and how to define it? Our idea is to use Brownian method to find the superposition constraint positions of ideal feature points and corrected feature points, and use U-Net to find the distorted pseudo edge in a single graph. The biggest significance of this study is that the raw image is deformed, and the distorted image areas are also marked by features, which participate in the image synthesis problem. The experimental results not only improve the accuracy of the composite image, but also improve the synthesis efficiency. Finally, because the aerial image acquisition is highly overlapped, we do not need to consider the problem of missing image after cutting off the deformed edge image.

2.2.1. Lens Distortion Model Analysis
1. Prepare the coordinates of corresponding points on the plane pattern and image buffer plane.
2. Calculate the initial image \( h \) between the plane pattern and the image buffer plane. In the two-dimensional mapping space, the distorted image point \( X \) and the corresponding feature point \( X_P \) are approximately related to \( H[8] \).

\[ X = H X_P \]

\( H \) is a 3 * 3 matrix, defined as a scale factor. It can be calculated by least square (LMS) method.[8]

Calculate the initial cost function

\[ f_0 = \frac{1}{N} \sum_{i=1}^{N} d(x_i, H x_{ip})^2 \]

4. Compute distortion parameter \( ms \) by \( x \) and \( Hx_p \). From Eqs. (11) and (12), we have the equations as.

\[ k_1 x_d \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} \right) + p_1 \left( 3 \frac{x^2}{a^2} + y^2 \right) + 2 p_2 x y = x_d - \bar{x}_d \]

\[ k_1 y_d \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} \right) + 2 p_1 x y + p_2 \left( \frac{x^2}{a^2} + 3 y^2 \right) = y_d - \bar{y}_d \]

2.2.2. Acquisition and Pruning of Distorted Pseudo Edge

As shown in Figure 8, the red region is the undistorted region, and the red outer region is the distorted region. Taking the distorted pseudo edge as the pruning boundary, the distorted region outside the pseudo edge is pruned to retain the undistorted true value region.

![FIG 8. distortion pseudo edge detection](image)

2.3. Digital Elevation Model (Dem) of Satellite Image

DOM is a digital orthophoto data set generated by digital elevation model (DEM) through pixel by pixel projection correction and mosaic in line with the national basic map scale. Orthophoto accuracy detection uses high-precision field control results as the basic data of sampling detection, and randomly selects digital orthophoto for accuracy detection according to different factors such as terrain types. The detection point should be selected in the boundary area as far as possible. Among them, the accuracy statistics is calculated according to the average error formula of high precision detection in GB / t24356 <= provisions for inspection and acceptance of surveying and mapping results > [3].

DEM production:

1. Aerial triangulation
   GXL is used in aerial triangulation. After automatic aerial triangulation, the beam is used to adjust the whole area, so as to obtain the regional encryption results and the image parameters after regional network adjustment[7].

2. DEM automatic extraction
   Using GXL to automatically extract DSM data with grid spacing of 10m, and filter DSM to DEM[7].

3. Feature point line acquisition
   The data collection of stereo mapping is carried out where the DEM obtained by automatic extraction is not accurate enough. Coordinate all kinds of feature points and lines to make their relationship reasonable.
4. DEM detection

The tin is constructed by using the feature points, lines, contour lines, elevation points and static pictures.

3 Experiment and Analysis

Tens of millions of original images with the boundary of effective truth value are extracted. After synthesis, the accuracy of two new digital orthophotos is detected and compared by using the digital elevation model of satellite image and its nested overlay constraint.

3.1. Exact Boundary Analysis of Orthophoto Composite Image

3.1.1. Analysis of Orthodontic Examination

Taking 1: 50,000 as the basic unit, comparing the corrected aerial orthophoto results with the latest satellite remote sensing images, the measurement error is reduced. The results are shown in Table 1 and figure 1. It can be seen from table 1 that the method based on U-Net network structure and pruning the deformed area in advance effectively improves the synthesis accuracy of Orthophoto projection, and has a small error compared with the latest satellite remote sensing images.

Table 1: Comparison of accuracy of plane position boundary between aerial Orthophoto Image and latest satellite remote sensing image

| Serial number | Terrain category | Check points | Error in the frame |
|---------------|-----------------|--------------|--------------------|
|               | Aerial images   | Satellite imagery | Aerial images DOM | Satellite imagery DOM |
| P12           | The building    | 38488        | 1848              | 1.860               | 6.219               |
| P13           | The building    | 18414        | 1678              | 2.748               | 5.244               |
| P14           | The building    | 19736        | 562               | 2.076               | 6.616               |
| P21           | The building    | 19621        | 2532              | 2.088               | 4.835               |
| P33           | The grass       | 18995        | 2461              | 2.136               | 5.074               |
| P34           | The grass       | 19720        | 2064              | 1.608               | 4.280               |
| P41           | The grass       | 25948        | 1080              | 1.932               | 6.638               |
| P42           | The hill        | 31942        | 5888              | 1.884               | 4.631               |
| P43           | mountain        | 8898         | 6308              | 2.496               | 4.983               |
| P44           | mountain        | 23960        | 3880              | 2.556               | 5.425               |
3.1.2. Boundary Accuracy Detection and Analysis

The accuracy test is carried out in zone 1:10000 encryption results and 1:2000 the control points, the accuracy test have passed the two-level acceptance inspection level, meeting the accuracy requirements of corresponding scale topographic map in aerial survey. The statistical results of precision detection are shown in Table 2 and Table 3. The results show that the detection accuracy of the new DOM is less than 3m for aerial orthophoto images, and less than 9m for the latest satellite remote sensing images. The DOM accuracy meets the requirements of the new generation technology. It can be seen that our scheme is efficient and feasible.

| Serial number | Terrain category | Error in specification | Check points | Error in the frame |
|---------------|------------------|------------------------|--------------|--------------------|
| P31           | The ground       | 27.200                 | 42.600       | 2.124              |
| P32           | The hills        | 27.200                 | 52.500       | 2.796              |
| P23           | The hills        | 27.200                 | 52.500       | 2.700              |
| P24           | Mountain         | 40.950                 | 48.100       | 9.684              |

| Serial number | Terrain category | Error in specification | Check points | Error in the frame |
|---------------|------------------|------------------------|--------------|--------------------|
| E31           | The ground       | 24.23                  | 33           | 2.79               |

3.2. Efficiency Comparison

It can be seen that the U-Net based correction method for correcting pseudo edges of distortion in aerial orthophotos not only improves the accuracy of the results, but also greatly improves the efficiency, and the boundary accuracy has largely approached the boundary accuracy of the latest satellite remote sensing images.

| Serial number | Job content                                    | Time of Use (Working day) | Improved performance |
|---------------|-----------------------------------------------|---------------------------|----------------------|
|               | Aerial image control point measurement         | Field control             | 100%                 |
| 1             | DOM production of aerial image(no IMU/DGPS)   | 191                       | 75%                  |
| 2             | Satellite image control point measurement     | Field control             | 100%                 |
|               | DOM production of satellite image(with RPC     | 23                        | 62%                  |
|               | parameters)                                    | 10                        |                      |
Compared with the Orthophoto Image synthesized by the traditional classical algorithm, the existing algorithm can avoid the problem that the distorted regions participate in feature merging. Pruning the invalid distorted regions in advance, also greatly improve the edge accuracy and quality of the synthetic image. As shown in Figure 10.

FIG 10. comparison of composite diagram

4. Conclusion

From the above research, we can see that after learning from the excellent neural network model U-Net modeling idea, we first obtain the high fidelity real boundary. Secondly, before the output of the true value, according to the distortion characteristics, the brown method is used to get the distorted pseudo edge, and the image of the distorted area is pruned. Finally, DOM detection is carried out simultaneously for the DEM of satellite image and the combined aerial Orthophoto Image. The results show that the DOM detection error of the new aerial composite image and satellite image is less than 3m and 9m, which is significantly improved compared with the previous composite image. Therefore, the edge calculation and correction method of aerial orthophoto composite map based on U-Net is efficient and feasible. In the future, priority should be given to further improve the accuracy and fast mapping performance of 3D orthophoto synthetic image map. At the same time, the U-Net network algorithm should be applied to the target detection of moving objects. Therefore, the probability threshold range and tracking query technology of moving objects are also the next research direction.

[References]

[1] Shao Z F, Tang P H, Wang Zhong Y, et al. BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images[J]. Remote Sens, 2020. 12(6), 1050.

[2] Han Z M, Dian Y Y, Xia H, et al. Comparing Fully Deep Convolutional Neural Networks for Land Cover Classification with High-Spatial-Resolution Gaofen-2 Images[J]. ISPRS Int. J. Geo-Inf, 2020, 9, 478.

[3] Hou B W, Yan D M, Hao W, Huang Q Q, Su X Q and Li Q W. Urban built-up area extraction using high-resolution remote sensing images with an improved convolutional neural network. Journal of Image and Graphics, 2020.25(12):267-2689.

[4] Wu G M, Chen Q, Ryosuke SHIBASAKI, et al. High Precision Building Detection from Aerial
Imagery Using a U-NetLike Convolutional Architecture[J].Acta Geodaetica et Cartographica Sinica,2018.47(6):864-872.

[5] Liu C,Zhang J,Lin J P.Single-pixel edge extraction of image based on improved fully convolutional neural network[J].Computer Engineering,2020.46(1):262-270.

[6] Zhong H,Li H,Li Z J,et al.Image manipulation detection based on convolutional neural networks[J].Journal of Jilin University(Engineering and Technology Edition),2020.50(4):1428-1434.

[7] Cuan L,Fan S,Chen S p,et al.Precision Analysis of Digital Elevation Model Based on ZY -3 Satellite Imagery[J].GEOMATICS & SPATIAL INFORMATION TECHNOLOGY,2017.40(8):143-146.

[8] Gao D P,Yin F L .Computing a complete camera lens distortion model by planar homography[J].Optics & Laser Technology,49(2013):95-107.

[9] LONGJ,SHELHAMERE,DARRELLT.Fully Convolutional Networks for Semantic Segmentation [C]//Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition . Boston,MA:IEEE,2015:3431-3440.

[10] BADRINARAYANANAN V,KENDALLA,CIPOLLAR.Segnet:A Deep Convolutional Encoder-decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(12):2481-2495.

[11] NOH H,HONG S,HAN B . Learning Deconvolution Network for Semantic Segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision . Santiago,Chile:IEEE,2015:1520-1528.

[12] RONNEBERGER O,FISCHER P,BROX T.U-net: Convolutional Networks for Biomedical Image Segmentation [M]//NAV AB N,HORNEGGER J, WELLS W,et al.Medical Image Computing and Computer-Assisted Intervention MICCAI2015.Cham:Springer,2015:234-241.

[13] MARMANIS D,WEGNER J D,GALLIANI S,et al.Semantic Segmentation of Aerial Images with an Ensemble of CNNs[J].ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,2016, III-3:473-480.