e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

A DUAL-ATTENTION NETWORK WITH GATED FUSION FOR ROAD SEGMENTATION IN SATELLITE IMAGES

Goutham Kondagurle*1, Dr. S. Shanthi*2

*1Student Of B. Tech Computer Science And Engineering, Department Of Computer Science And Engineering, Malla Reddy College Of Engineering & Technology, Hyderabad, Telangana, India.

*2Professor, Department Of Computer Science And Engineering, Malla Reddy College Of Engineering & Technology, Hyderabad, Telangana, India.

ABSTRACT

The automatic recognition and extraction of roads from high-resolution satellite images are vital tasks in remote sensing and computer vision. As remote sensing technology progresses, more detailed ground object information becomes available in images, making road extraction more challenging due to increased interference. This paper proposes a novel approach called the Gated Fusion and Dual Attention Network (GFDANet) to tackle this issue. Our method employs an encoder-decoder structure with a full-stage gated fusion module in the skip connection. This module selectively fuses feature maps of different scales using the Gated Fusion (GFF) unit, thereby expanding the network's receptive field and enhancing the accuracy of road extraction. Additionally, the context extraction and dual attention module are introduced to incorporate rich global information and to appropriately weight features from both spatial and channel dimensions. This addresses the limitations of existing road extraction models, which often focus solely on local information, thereby improving semantic segmentation. Experimental results on two public datasets demonstrate the effectiveness of GFDANet in accurately extracting roads even in complex scenes.

Keywords: Road Extraction, Gated Fusion, Very High Resolution (VHR) Satellite Imagery.

I. INTRODUCTION

Realizing automatic, intelligent, reliable, and accurate road extraction methods from very high resolution (VHR) satellite images is an important research field of remote sensing image intelligence interpretation. Automatic road extraction from VHR satellite images can be used for disaster emergency management, smart cities, autonomous driving, high-precision map acquisition, and other missions.

Road extraction from remote sensing images is an important task with many potential applications. In recent years, deep learning models have been widely used to achieve highquality results. Some of the most popular models include the Full Convolutional Neural Network (FCN), U-Net, DeepLabV3, and D-LinkNet. FCN uses multiple convolutional and pooling layers to gradually expand the perceptual field of the network, making it easier to extract road information. U-Net connects different levels of feature maps to effectively extract road information. DeepLabV3 uses dilated convolution to capture the distance between pixels, preserving spatial structure information. It also proposes a spatial pyramid pooling layer to aggregate remotely sensed images. d-LinkNet uses an encoder-decoder architecture to obtain a coarse localization of road extraction through highlevel features and refine boundaries through low-level features containing spatial structure details.

Simultaneously, the attention mechanism has been fully exploited in computer vision tasks. Hu et al proposed a new structural unit, squeeze-excitation (SE), which adaptively recalibrates channel feature responses by specifying the interdependencies between channels. Sanghyun et al proposed a simple but effective approach: giving an intermediate feature map with mutually independent channel attention and spatial attention, the weights obtained from the attention mechanism are then multiplied by the input feature map for adaptive feature refinement. Hou et al developed a channel attention mechanism that combines spatial information and enables more accurate localization and detection.

Nevertheless, the high-resolution remote sensing images captured by satellites have feature information of varying scales and complex structures. This complexity, combined with the presence of other features such as pedestrians, vehicles on the road, shading from trees, and tall buildings, make road extraction challenging. Most existing methods for extracting multiscale features of roads use dilated convolution to expand the perceptual field and stitch multiscale information directly into the module, without considering the semantic differences

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

between features at different convolution kernel scales, which affects the accuracy of road extraction from remote sensing images. Moreover, the existing attention mechanism does not account for the slender nature of the road itself and its highly complex topology.

This paper presents a novel approach called GFDANet, which uses a gated fusion and Dual Attention Network to accurately extract roads from very high resolution (VHR) satellite images. Inspired by GFFNet and RCFSNet, this paper employs a gated mechanism to aggregate multi-scale information of convolutional kernel with different expansion rates and controls the propagation of information between different feature mappings by selecting channels to obtain better information aggregation results. The main contributions of this paper are as follows:

- Multiple full-stage gated fusion units (FSGF) are constructed in the skip connection to fuse different stage features and provide accurate road structure information. Unlike direct feature stitching or summation, FSGF uses selective channels to control information propagation between different feature mappings and to fuse fine-grained features for improved accuracy.
- A context extraction and dual attention module (CEDA) are proposed to capture long-distance dependencies, which can effectively extract long-distance roads and enhance the channel and spatial characteristics of roads.
- Experiments are conducted on two public datasets, and the results demonstrate that GFDANet can significantly improve road extraction accuracy in complex scenarios.

II. LITERATURE REVIEW

This paper presents a novel approach called GFDANet, which uses a gated fusion and Dual Attention Network to accurately extract roads from very high resolution (VHR) satellite images. Inspired by GFFNet and RCFSNet, this paper employs a gated mechanism to aggregate multi-scale information of convolutional kernel with different expansion rates and controls the propagation of information between different feature mappings by selecting channels to obtain better information aggregation results. The main contributions of this paper are as follows:

- Multiple full-stage gated fusion units (FSGF) are constructed in the skip connection to fuse different stage features and provide accurate road structure information. Unlike direct feature stitching or summation, FSGF uses selective channels to control information propagation between different feature mappings and to fuse fine-grained features for improved accuracy.
- A context extraction and dual attention module (CEDA) are proposed to capture long-distance dependencies, which can effectively extract long-distance roads and enhance the channel and spatial characteristics of roads.

The segmentation-based remote sensing image road extraction method generally regards the extraction of road maps as a pixel-by-pixel binary classification problem, i.e., to identify whether each pixel in the image belongs to the road region. After obtaining the segmentation result, the road graph is obtained through postprocessing algorithms such as skeletonization.

Deep neural networks are widely used in segmentation based road extraction methods. Early deep-learning-based segmentation method divides the aerial image into small patches and then uses a deep classification network to classify each patch to determine whether it belongs to the road areas. With more fully convolutional networks architecture being proposed, such as FCN, U-Net, DeepLab, and SegNet, the subsequent segmentation-based roadextraction methods, also adopt similar network architectures. Cheng et al. utilized SegNet as the network backbone and proposed a cascaded road centerline extraction method. Diakogiannis et al. designed a U-Net network based on residual modules.

Later, some methods incorporating the characteristics of the road were proposed. Shi et al. applied the generative adversarial network to the road extraction task, and used the discriminator network to make the segmentation result similar to the structure of the real road map. Gao et al. considered the roads of different scales and shapes and proposed a multiscale pyramid network that consists of convolutions on different scales and shapes. Zao and Shi proposed a detailed feature fusion module and a road topological loss function and obtained a more refined road segmentation result. Batra et al. developed a stacked multibranch convolutional module to effectively utilize the mutual information between orientation learning and segmentation tasks. To

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

better integrate context information, some methods, adopt network architecture with a larger receptive field. Zhou et al. proposed D-LinkNet that introduces a codec structure and an intermediate field of the view expansion module. Wang et al. used a nonlocal correlation module to enable the network to obtain a larger receptive field. Zao et al. proposed a network with differentiable Hough transform modules that enhance global line features. Although the above segmentation-based methods are widely studied, these methods are still limited to improving pixel-bypixel classification results, and the topology of roads is rarely considered. Different from these methods, our method directly treats road graph extraction as a vertex and edge extraction problem, so road topology is effectively incorporated into the process.

III. METHODOLOGY

A. GFDANet

The GFDANet architecture is illustrated in Figure 1, comprising an encoder-decoder, FSGF module, and CEDA module. To begin with, we utilize the ResNet34 [18] architecture as the encoder to derive the initial feature E1 via convolution. Subsequently, the feature maps E2, E3, E4, and E5 are obtained using Maxpooling and multiple residual blocks. After the CEDA module, a feature map with richer global information is obtained, which captures more information about long-distance roads. The skip connection FSGF module incorporates a gating mechanism to suppress irrelevant information and aggregates usage information from the entire stage while combining the output of the preceding stage decoder to serve as input for the subsequent stage decoder. During prediction, the feature maps D5, D4, D3, and D2 are adjusted to the same resolution and number of channels as D1 via upsampling and convolution operations while conducting feature fusion. Subsequently, the fused results undergo 3×3 convolution and upsampling operations to achieve a semantic segmentation output that corresponds to the input image's size.

B. Full-Stage Gated Fusion

Module With the increasing availability of remotely sensed road image data and improving image resolution, extracting road from these complex scenes has become more challenging. Since encoders at different stages capture varying levels of information, the feature map derived from the shallow network is abundant in spatial information, allowing for precise road positioning. On the other hand, the high-level feature map obtained from the deep network contains accurate road semantic information, enabling better distinction of roads from the background. Therefore, in this paper, we propose the FSGF module that aggregates the feature maps of the entire stage, empowering the decoder to extract sufficient road features, resulting in a clearer road boundary prediction. First, coarse-grained feature information (E1, E2) and fine-grained feature information (E4, E5) are adjusted to the same size as E3 through convolution, upsampling, and pooling operations and the number of channels is adjusted to 64. Then the feature maps are concatenated, and the spliced feature maps are convolved with different expansion rates to obtain information on the feature maps of different scales. After the GFF unit emphasizes useful information, inhibition of useless information, and strengthens the said road characteristics. Finally, the obtained feature maps were concatenated, and the accurate road structure feature maps were generated by convolution, batch Normalization (BN), and activation function (ReLU).

C. Gated Fully Fusion (GFF)

Unit To improve the accuracy of road extraction, it is essential to expand the receptive field and acquire multi-scale information. Many existing road extraction methods use multiple parallel expansion rates to construct convolution kernels with different sensitivity fields to obtain multi-scale information and splice multi-scale information of different branches directly. However, this method is often ineffective because it does not consider the semantic differences between different scale features of the convolution kernel. Inspired by GFFNet, we used GFF units in the FSGF module to selectively fuse feature maps containing context information of different scales. In the GFF unit, each branch feature is enhanced by the remaining features to control information propagation through a gating mechanism. The gating mechanism effectively restrains useless information and only allows useful information to be sent to the correct position. This mechanism avoids redundant information and ensures that only features that are necessary for the current position receive information.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

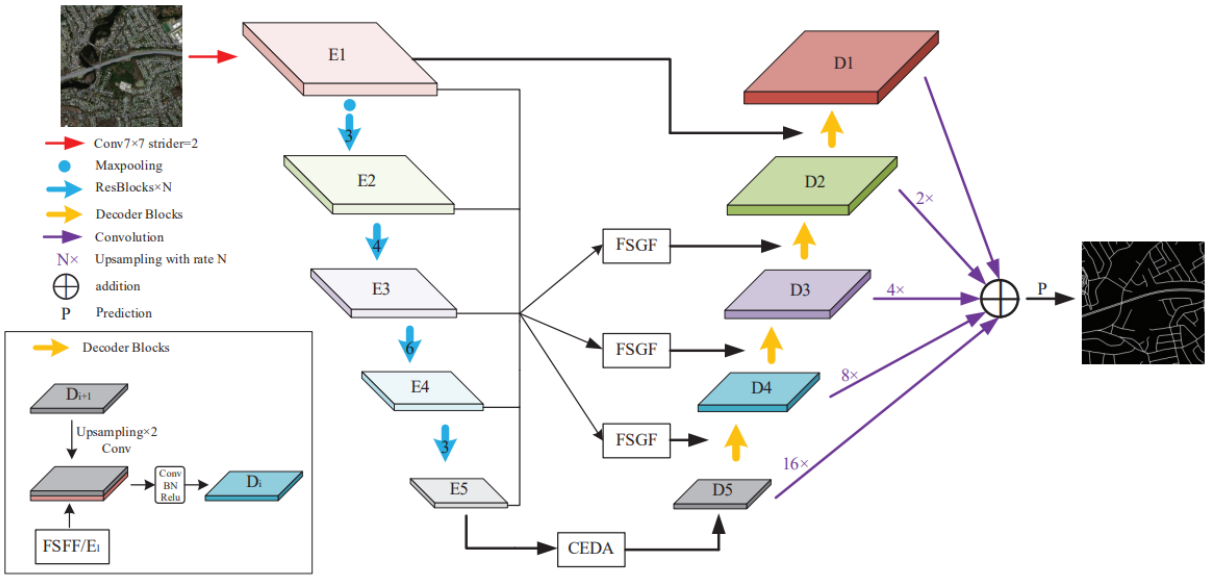

Figure 1: GFDANet Architecture

D. Context Extraction and Dual Attention Module

Due to the variety and complexity of feature information in high resolution remotely sensed imagery, the road structure is more complex. Some of the roads are obscured by obstacles such as trees and buildings, resulting in incomplete extraction of roads over long distances. To solve this problem, a context extraction and dual attention module (CEDA) is proposed in this paper.

The part in the red dotted box in Figure 4, road features are extracted to introduce a rich road environment by using 3×3 convolution branches with an expansion rate of {1,2,4}for E5, and global information in horizontal and vertical directions are preserved by two pooling branches. Then, the features of different branches are concatenated and convolved before being combined with E5 to obtain a feature map with rich context information, where p is a learnable parameter.

The part in the blue dotted box carries out feature weighting on the above feature map from two dimensions of space and channel. Make the network more focused on road-useful parts and suppress irrelevant semantic information. It improves the semantic segmentation of the common road extraction model which only focuses on local information and improves the efficiency and accuracy of road extraction. Finally, the final road structure feature map D5 is obtained after concatenation, convolution, BN, and ReLU of the obtained results.

IV. RESULTS AND ANALYSIS

To evaluate the proposed FSGF module and CEDA module, we conducted ablation experiments on the Massachusetts data set, and the experimental results are shown in Table IV. When two modules are removed at the same time, Recall and IOU of the model decrease significantly, which indicates that our work is effective for road extraction. When only the CEDA module is added, Recall of the model increases by 2.83%, IOU by 2.98%, and F1-Score by 2.31%, which indicates that rich road background and global information are very effective for complete road extraction. When only increased FSGF module, the model of a Recall of 2.55%, IOU rose by 3.27%, and F1 - Score increases by 2.48%, indicating that all stages of road information in improving road label accuracy are very important. The best experimental results are obtained by using two modules simultaneously, which indicates that our model can extract complete roads with high accuracy and clear boundaries from satellite images.

V. CONCLUSION

This paper proposes a GFDANet network for road extraction of satellite images. Through the full-stage gated fusion of feature maps and context extraction with dual attention, our network is very effective for long-distance road extraction, and the extracted road boundary is clearer. A large number of experiments on the SpaceNet dataset and the Massachusetts dataset fully demonstrate the effectiveness of our network. In the

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

subsequent work, considering the limitations of labeled data, we plan to use labels that are easier to obtain than pixellevel labels combined with the idea of weak supervision to further improve the accuracy of the network.

VI. REFERENCES

[1] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," J. Traffic Transp. Eng., vol. 3, no. 3, pp. 271–282, 2016.

[2] Y.-Y. Chiang and C. A. Knoblock, "Extracting road vector data from raster maps," in Proc. Int. Workshop Graph. Recognit. Cham, Switzerland: Springer, 2009, pp. 93–105.

[3] Y. Wang, J. Seo, and T. Jeon, "NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations," IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, 2022.

[4] C. Heipke, H. Mayer, C. Wiedemann, and O. Jamet, "Evaluation of automatic road extraction," Int. Arch. Photogramm. Remote Sens., vol. 32, no. 3, pp. 151–160, 1997.

[5] J. B. Mena, "State of the art on automatic road extraction for GIS update: A novel classification," Pattern Recognit. Lett., vol. 24, no. 16, pp. 3037–3058, Dec. 2003.

[6] H. Mayer, S. Hinz, U. Bacher, and E. Baltsavias, "A test of automatic road extraction approaches," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol. 36, no. 3, pp. 209–214, 2006.

[7] S. Das, T. T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from highresolution multispectral satellite images," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 10, pp. 3906–3931, Oct. 2011.

[8] J. Shen, X. Lin, Y. Shi, and C. Wong, "Knowledge-based road extraction from high resolution remotely sensed imagery," in Proc. Congr. Image Signal Process., vol. 4, May 2008, pp. 608–612.

[9] D. Guo, A. Weeks, and H. Klee, "Robust approach for suburban road segmentation in high-resolution aerial images," Int. J. Remote Sens., vol. 28, no. 2, pp. 307–318, Jan. 2007.

[10] R. Marikhu, M. Dailey, S. Makhanov, and H. Kiyoshi, "A family of quadratic snakes for road extraction," in Proc. Asian Conf. Comput. Vis., vol. 4843, Nov. 2007, pp. 85–94.

[11] M. Song and D. Civco, "Road extraction using SVM and image segmentation," Photogramm. Eng. Remote Sens., vol. 70, no. 12, pp. 1365–1371, Dec. 2004.

[12] M. Amo, F. Martinez, and M. Torre, "Road extraction from aerial images using a region competition algorithm," IEEE Trans. Image Process., vol. 15, no. 5, pp. 1192–1201, May 2006.

[13] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 1, pp. 1–14, Jan. 1996.

[14] P. Gamba, F. Dell'Acqua, and G. Lisini, "Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts," IEEE Geosci. Remote Sens. Lett., vol. 3, no. 3, pp. 387–391, Jul. 2006.

[15] M. Barzohar and D. B. Cooper, "Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation," IEEE Trans.