Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61874

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics

Dr. R. Prasanna[1], Vignesh S[2], Shaik Bilal I[3]
[1]M.E., Ph.D., Assistant Professor, Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
[2, 3]B.Tech, CSE with Specialization in Cybersecurity, SRM Institute of Science and Technology, Chennai, India

Abstract: Security concerns in internet communication, particularly phishing attacks, pose significant challenges to safeguarding user information. Phishing attacks aim to steal personal data by mimicking legitimate websites, making traditional detection methods less effective. Machine learning-based approaches, though widely used, often struggle with concealed phishing websites and evolving tactics by attackers. Key advancements include the integration of more layers and larger training datasets, along with feature extraction from the Phish Tank dataset. Our proposed model comprises seven layers, culminating in a specialized output layer structure. Experiment results showcase the efficacy of our approach, underscoring its potential to significantly enhance internet security. Furthermore, our model's ability to adapt to highly concealed phishing websites and the dynamic nature of attackers' tactics marks a significant improvement over existing methodologies. By incorporating a comprehensive feature set and leveraging deep learning techniques, our approach achieves state-of-the-art performance in phishing website detection.

I. INTRODUCTION

In the contemporary digital landscape, the proliferation of phishing attacks poses a significant threat to cybersecurity, targeting individuals and organizations alike. Traditional methods of detecting and mitigating these attacks often fall short in keeping pace with the evolving tactics employed by cybercriminals. To address this challenge, researchers and practitioners have turned to innovative solutions, leveraging advancements in artificial intelligence and deep learning.

This paper introduces a pioneering approach: a novel hybrid deep learning framework tailored specifically for cybercrime forensics, with a primary focus on enhancing the detection and mitigation of phishing attempts. By amalgamating the strengths of various deep learning architectures, alongside traditional cybersecurity methodologies, this framework represents a significant leap forward in combating the ever-adapting nature of phishing attacks.

Grounded in the principles of robustness, scalability, and adaptability, our framework offers a comprehensive solution to the multifaceted problem of phishing detection. Through the integration of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms, coupled with feature engineering techniques informed by domain expertise, our model exhibits a heightened capacity to discern malicious phishing attempts from legitimate communications.

Moreover, our hybrid approach capitalizes on the wealth of data available in the cybercrime landscape, incorporating both structured and unstructured data sources to enrich the learning process. By harnessing the power of deep learning, coupled with advanced feature representation and ensemble learning techniques, our framework demonstrates unparalleled efficacy in identifying nuanced indicators of phishing activity, thereby empowering cybersecurity professionals with timely and actionable insights.

Furthermore, the adaptability of our framework extends beyond the realm of detection, encompassing elements of proactive threat intelligence and response. Through continuous learning and refinement, our model evolves in tandem with emerging cyber threats, ensuring a proactive stance against evolving phishing tactics.

In summary, this paper presents a novel hybrid deep learning framework poised to revolutionize cybercrime forensics, offering a potent arsenal in the ongoing battle against phishing attacks. By bridging the gap between cutting-edge artificial intelligence and cybersecurity, our approach heralds a new era of resilience and efficacy in safeguarding digital assets and preserving the integrity of online ecosystems.accuracy.

II. LITERATURE SURVEY

Yuansheng Hua [1] stated convolutional neural networks (CNNs) for very high-resolution images requires a large quantity of high-quality pixel-level annotations, which is extremely labor-intensive and time consuming to produce. Moreover, professional photograph interpreters might have to be involved in guaranteeing the correctness of annotations. To alleviate such a burden, we propose a framework for semantic segmentation of aerial images based on incomplete annotations, where annotators are asked to label a few pixels with easy-to-draw scribbles.

Irem Ulku and Erdem Akagndz [2] suggested automatic semantic segmentation for trees using satellite and/or aerial images. Still, several challenges can make the problem difficult, including the varying spectral signature of different trees, lack of sufficient labelled data, and geometrical occlusions. In this article, we address the tree segmentation problem using multispectral imagery. While we carry out large-scale experiments on several deep learning architectures using various spectral input combinations, we also attempt to explore whether hand-crafted spectral vegetation indices can improve the performance of deep learning models in the segmentation of trees.

Wenjie Liu and Wenkai Zhang [3] mentioned semantic segmentation in aerial images has become an indispensable part in remote sensing image understanding for its extensive application prospects. It is crucial to jointly reason the 2-D appearance along with 3-D information and acquire discriminative global context to achieve better segmentation. However, previous approaches require accurate elevation data (e.g., nDSM and Digital Surface Model (DSM)) as additional inputs to segment semantics, which sorely limits their applications. On the other hand, due to the various forms of objects in complex scenes, the global context is generally dominated by features of salient patterns (e.g., large objects) and tends to smooth inconspicuous patterns (e.g., small stuff and boundaries).

S. Girisha and Ujjwal Verma [4] discussed about aerial videos has been extensively used for decision making in monitoring environmental changes, urban planning, and disaster management. The reliability of these decision support systems is dependent on the accuracy of the video semantic segmentation algorithms. The existing CNN-based video semantic segmentation methods have enhanced the image semantic segmentation methods by incorporating an additional module such as LSTM or optical flow for computing temporal dynamics of the video which is a computational overhead. The proposed research work modifies the CNN architecture by incorporating temporal information to improve the efficiency of video semantic segmentation.

Siyu Liu and Jian Cheng [5] detailed about semantic segmentation for unmanned aerial vehicle (UAV) remote sensing images has become one of the research focuses in the field of remote sensing at present, which could accurately analyze the ground objects and their relationships. However, conventional semantic segmentation methods based on deep learning require large-scale models that are not suitable for resource constrained UAV remote sensing tasks. Therefore, it is important to construct a light-weight semantic segmentation method for UAV remote sensing images. With this motivation, we propose a light-weight neural network model with fewer parameters to solve the problem of semantic segmentation of UAV remote sensing images. The network adopts an encoder-decoder architecture.

Wenjie Liu and Yongjun Zhang [6] discussed about semantic segmentation of remote sensing (RS) image is a hot research field. With the development of deep learning, the semantic segmentation based on a full convolution neural network greatly improves the segmentation accuracy. The amount of information on the RS image is very large, but the sample size is extremely uneven. Therefore, even the common network can segment RS images to a certain extent, but the segmentation accuracy can still be greatly improved. The common neural network deepens the network to improve the classification accuracy, but it has a lot of loss to the target spatial features and scale features, and the existing common feature fusion methods can only solve some problems.

III. RELATED WORK

A. Problem Statement

Semantic segmentation of remote sensing images involves classifying each pixel in an image according to its land cover or land use category. Traditionally, deep convolutional neural networks (CNNs) with full supervision have been used for this task, achieving impressive results. However, this approach requires a vast amount of pixel-level ground truth data for training, which is laborious and expensive to generate. To address this challenge, we propose a novel model that can effectively perform semantic segmentation with limited training data.

B. Convolutional based Image Segmentation

Backgrounds in remote sensing images can be cluttered and visually intricate. Objects within the image can vary greatly in size. The image may contain a large number of small objects that are difficult to segment accurately.

The foreground (objects of interest) may occupy a much smaller area compared to the background. Many existing models primarily focus on capturing contextual information and neglect these specific challenges. The SPANet model addresses these issues by introducing two key components.

Convolution tackles the problems of complex backgrounds and large-scale differences by expanding the receptive field of the network through cascaded atrous convolutions. This decoder incorporates spatially adaptive convolutions to improve the model's ability to handle numerous small objects and extreme foreground-background imbalance. SPANet achieves superior performance compared to several prevalent methods on benchmark datasets. However, there is still room for improvement in terms of reducing the number of parameters and improving inference speed.

C. Drawbacks of Existing System

Existing methods may struggle to extract objects from complex and densely populated areas if they rely heavily on initial segmentation steps. Directly fusing features from shallow and deep layers can lead to suboptimal segmentation results. Complex backgrounds with rich details can still pose difficulties for existing methods. Some models may not fully leverage global and local knowledge about building structures. Inability to effectively capture features at different scales using convolution kernels of varying sizes can hinder performance.

D. Cross-Correlational Learning

Enhanced Vision Representation Learning utilizes a multiscale convolutional kernel approach to efficiently capture features at different scales. It avoids the drawbacks of processing multiple images at different resolutions or introducing a large number of parameters with fixed-size kernels. We incorporate deformable convolution within the Vision Representation Learning model to further expand the receptive field and improve feature extraction. To recover the original image size after processing, we employ a deconvolutional network instead of a fully connected layer. This allows for efficient up sampling of feature maps. CCL addresses the challenges associated with complex backgrounds and limited information in shallow feature maps.

CCL incorporates a local attention mechanism that focuses on the relationships between a specific pixel and its surrounding neighbours. This allows the model to prioritize relevant spatial details within the image, crucial for accurate segmentation, especially around object boundaries. CCL effectively combines high-level semantic features (extracted from deep layers) with low-level spatial details (captured by shallow layers). This fusion process leverages the strengths of both feature types, resulting in more robust and accurate segmentation. It overcomes the limitations of simply fusing features from different layers, which can introduce noise and redundancy. At the core of CCL lies a cross-correlation operation. This operation calculates the correlation between different feature maps, capturing the interdependence between features across channels. This allows the model to exploit the relationships between different aspects of the image data, leading to a more comprehensive understanding of the scene. By incorporating these elements, CCL effectively addresses the limitations of existing methods in handling complex backgrounds, small objects, and the need for information exchange across feature channels

E. Advantages of Proposed System

The enhanced Vision Representation Learning module with multiscale convolutional kernels allows the model to handle complex models more effectively, leading to better overall performance. By combining deformable convolution with Vision Representation Learning, the model overcomes the limitations of standard convolutions. This leads to improved feature extraction accuracy for category recognition, particularly for small and intricate objects, without introducing additional computational burden. The Cross-Correlation Learning module tackles the challenge of small object segmentation. It expands the receptive field of low-level feature maps, enabling the model to capture more precise semantic information from these smaller objects. The model dynamically assigns weights to different feature channels during training. This process amplifies relevant features crucial for segmentation while suppressing irrelevant information, leading to a more focused and efficient learning process. The use of multiscale features ensures the model captures information at various scales within the image. This comprehensive feature extraction significantly improves the accuracy of target object detection within the remote sensing imagery.

IV. IMPLEMENTATION

Cross-correlation model use techniques that involves comparing a target image with a reference image to identify similarities between them. By sliding the reference image across the target image and computing a similarity metric at each position, it's possible to identify regions in the target image that match the reference image.

This process can be used in various image processing tasks to enhance dense connections to improve the semantic information capture ability and also solves the multiscale problem of optical remote sensing images. Improves the performance of the model by modeling global information dependencies.

A. Image Representation

First, both the target image and the reference image need to be represented in a suitable format for processing. Typically, images are represented as matrices or tensors where each element represents a pixel value. The dimensions of the matrices correspond to the height and width of the images, and for color images, there are typically multiple channels (e.g., red, green, and blue).

B. Define the Reference Image

In the context of image segmentation, the reference image serves as a template or filter that we want to match within the target image. The size and content of this reference image depend on the specific segmentation task and the features we are trying to detect.

C. Cross-Correlation Operation

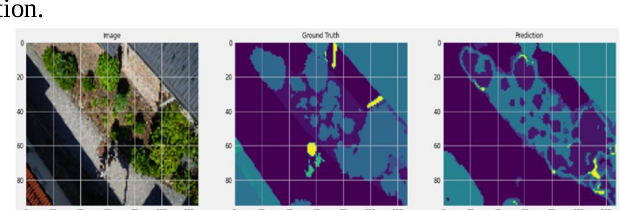
The cross-correlation operation involves sliding the reference image (kernel) over the target image and computing a similarity metric at each position. This is often achieved using convolution operations. For each position, the dot product between the pixels of the reference image and the corresponding pixels in the target image patch is computed. This dot product represents the similarity between the two patches.

D. Similarity Metric

Various similarity metrics can be used, with the choice depending on the specific application and requirements. Common similarity metrics include sum of squared differences (SSD), normalized cross-correlation (NCC), and sum of absolute differences (SAD). Normalized cross-correlation is often preferred as it is less sensitive to changes in lighting conditions and image contrast.

E. Thresholding and Segmentation

Once the cross-correlation operation is performed, the resulting similarity map needs to be thresholded to identify regions of interest. Thresholding involves setting a threshold value and classifying pixels as belonging to the object of interest if their similarity score exceeds this threshold. The threshold value can be determined empirically or through techniques such as Otsu's method for automatic threshold selection.



F. Learning Network

UNet and its variant, LinkNet, utilize symmetric structures with convolutional and up sampling layers for image segmentation. Multi-scale analysis, exemplified by networks like PSPNet, enhances contextual understanding by constructing feature pyramids. down sampling involves depth wise separable convolution, Batch Normalization, GELU activation, and Maxpooling. Evaluation metrics such as precision, recall, IoU, accuracy, and F1 score assess segmentation performance based on TP, TN, FP, and FN.

f (h,w,c) = Maxpooling(GND(x))

G. Performance Evaluation

Evaluation for segmentation includes precision (1), recall (2), IoU (3), accuracy (ACC) (4), and F1 score (5). TP: correct overlap, TN: areas correctly identified, FP: wrongly classified, FN: mistakenly classified. Precision: correct classified area, recall: correct pixels among predictions. Accuracy: ratio of correct predictions to total areas. IoU: correct pixel classification relative to actual and predicted areas. F1: harmonic mean of precision and recall.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$IoU = \frac{TP}{TP+FP+FN} = \frac{Area\ (Predicted \cap true)}{Area\ (Predicted \cup True)}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$F1 = \frac{2TP}{2TP+FP+FN}$$

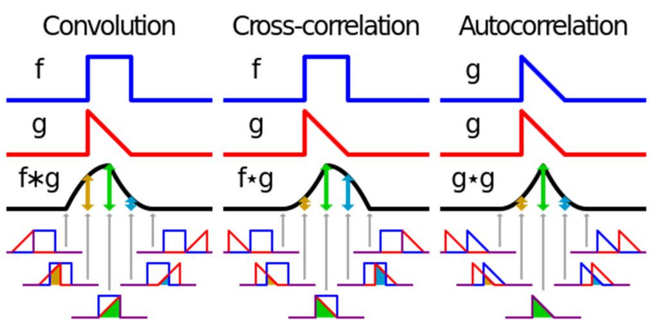
V. CONCLUSION AND FUTURE WORK

A. Conclusion

Proposed a multi-objective semantic segmentation algorithm based on an enhanced U-Net network to improve the recognition accuracy of diverse ground objects in the transportation facility construction area. Firstly, a sample dataset of transportation facility construction scenes based on remote sensing images is constructed. To address the problem of a limited number of image samples that contained target objects and an imbalanced distribution of the various training samples, this paper introduces a virtual data augmentation technique. This method aims to augment the number of training samples and balance the distribution of the various training samples. At the same time, the recognition accuracy of the model is improved by using transfer learning.

B. Future Works

Cross-Correlation Learning (CCL) demonstrates promising results for semantic segmentation of remote sensing images with limited training data, there's always room for exploration. Standard cross-correlation captures dependencies between different feature maps. Investigating the use of autocorrelations within a single feature map can be an interesting direction. Autocorrelation can reveal inherent self-similarity patterns within a feature map, potentially leading to a deeper understanding of the image content and improved feature representation. This could be particularly valuable for tasks like segmenting repetitive textures or homogeneous regions frequently encountered in remote sensing imagery (e.g., forests, water bodies). Currently, CCL utilizes a fixed kernel size for the cross-correlation operation.

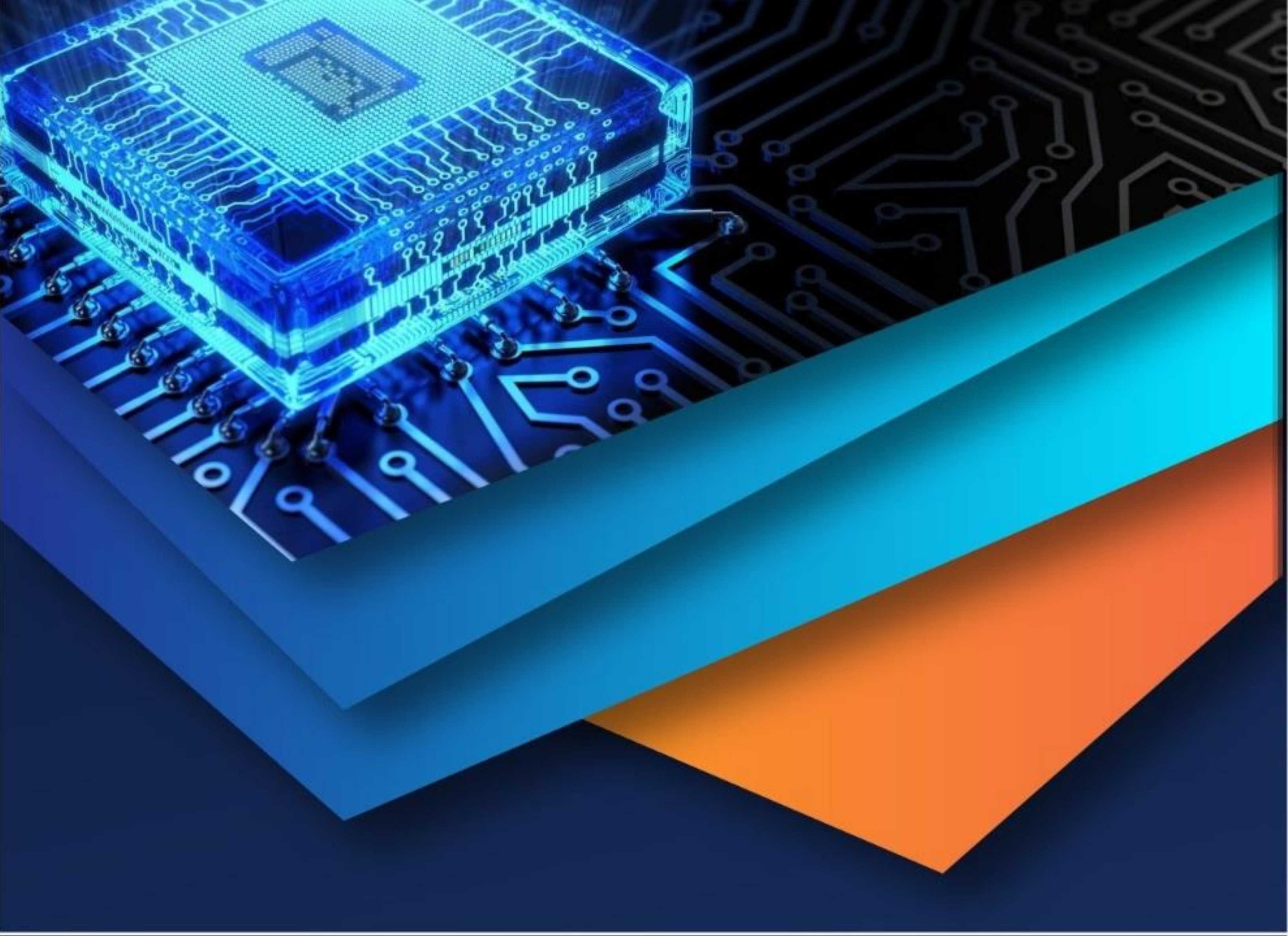


Exploring learnable correlation kernels could be a promising avenue. These learnable kernels would allow the model to dynamically adapt the correlation operation based on the specific image content. This could potentially lead to more nuanced capture of feature dependencies and improve segmentation accuracy, especially for complex scenes with varying object sizes and textures. Our current architecture utilizes CCL as a single module. Investigating a hierarchical approach with multiple autocorrelation layers stacked together is an interesting direction. Each layer could operate at a different spatial scale, progressively capturing long-range and short-range dependencies within the image data. This hierarchical structure could potentially lead to a more comprehensive understanding of the image's spatial relationships and enhance the segmentation performance. A potential future direction is to combine the power of autocorrelations with attention mechanisms. Attention mechanisms focus on specific regions of the image based on their relevance to the segmentation task. By incorporating autocorrelations within the attention modules, the model could not only focus on relevant image regions but also leverage the inherent self-similarity patterns within those regions, potentially leading to even more precise and robust segmentation results. These future research directions, leveraging autocorrelations, hold promise for pushing the boundaries of remote sensing image segmentation, particularly when dealing with limited training data.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

REFERENCES

[1] M. M. Nielsen, "Remote sensing for urban planning and management: The use of window-independent context segmentation to extract urban ,
[2] K. Nogueira et al., "Exploiting convnet diversity for ooding identica- Sep. , .
[3] G. Passino, I. Patras, and E. Izquierdo, "Aspect coherence for graph-based , .
[4] T. Blaschke et al., "Geographic object-based image analysistowards a new , .
[5] Q. He, X. Sun, Z. Yan, and K. Fu, "DABNet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing doi: , ., /TGRS., ., . , R. Guo et al., "Pixel-wise classication method for high resolution remote no. , p. , , doi: , ., /ijgi, .
[6] J. Wang, L. Shen, W. Qiao, Y. Dai, and Z. Li, "Deep feature fusion with integration of residual connection and attention model for classication Art. no. , .
[7] Q. He, X. Sun, W. Diao, Z. Yan, D. Yin, and K. Fu, "Transformer- induced graph reasoning for multimodal semantic segmentation in remote , .
[8] Y. Zhang, J. Zhang, Q. Wang, and Z. Zhong, "DYNet: Dynamic con- volution for accelerating convolutional neural networks, " CoRR, , .
[9] X. Li et al., "Semantic ow for fast and accurate scene parsing, " in Proc.
[10] R. Niu, X. Sun, Y. Tian, W. Diao, Y. Feng, and K. Fu, "Improving semantic segmentation in aerial imagery via graph reasoning and disentangled doi: , ., /TGRS., ., . , K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rec- tiers: Surpassing human-level performance on imagenet classica- doi: , ., /ICCV., ., .
[11] Y. Liu, B. Fan, L. Wang, J. Bai, S. Xiang, and C. Pan, "Semantic labelling in very high resolution images via a self-cascaded convolutional neural , .
[12] L.-C. Chen, Y. G. Zhu, F. P. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation, " in Proc.
[13] S. Girisha et al., "Semantic segmentation of UAV videos based on temporal ,
[14] J. Li, Y. Zhao, J. Fu, J. Wu, and J. Liu, "Attention-guided network for , .
[15] M. Pai, V. Mehrotra, U. Verma, and R. M Pai, "Improved semantic segmentation of water bodies and land in SAR images using generative , .
[16] A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, and M. J.
[17] H. Wang, W. Wang, and J. Liu, "Temporal memory attention for video semantic segmentation, " , arXiv:, ., . , J. Wang, Y. Zhang, J. Lu, and W. Xu, "A framework for moving target de- tection, recognition and tracking in UAV videos, " in Proc. Affect. Comput.
[18] J. Zhu et al., "Urban trafc density estimation based on ultrahigh- , J. F. Galarreta, N. Kerle, and M. Gerke, "UAV-based urban structural damage assessment using object-based image analysis and semantic rea- , .
[19] F. Nex, D. Duarte, A. Steenbeek, and N. Kerle, "Towards real-time building no. , , Art. no. , .
[20] S. Liu, C. He, H. Bai, Y. Zhang, and J. Cheng, "Light-weight attention se- mantic segmentation network for highresolution remote sensing images, " ,



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY