Enhanced Semantic Image Segmentation Through Convolutional and Atrous Convolution Techniques

Meghana H.V.[1*], Ushashree R.[2]

[1,2]Dayananda Sagar Academy of Technology and Management, Karnataka, India

***Email:** meghanahv19@gmail.com

Abstract

Enhanced Image content classification has improved dramatically with the advent of CNNs. This paper presents an enhanced method for semantic partitioning through merging traditional convolutional level and atrous (extended) convolution techniques. Our approach takes advantage of the hierarchical feature extraction capabilities of CNNs, while incorporating atrous convolutions to capture multi-scale contextual information without increasing the computational load. The proposed feature combines standard diffraction layers for detailed feature extraction that broadens the perceptive field, thus improving segmentation accuracy, especially on multiscale features Extensive testing on the datasets including PASCAL VOC 2012 and Cityscapes.

Keywords

Semantic Image Segmentation, Convolutional Neural Networks (CNNs), Atrous Convolutions, Feature Extraction

.

Introduction

Image Classification Content is an important task in computer vision, which aims to classify each pixel in a image into predefined units. This work supports many applications, including autonomous driving, medical imaging, and situational recognition. Traditional image segmentation methods relied heavily on artificial artifacts and classical machine learning techniques, and often failed to capture the complexity and variations in images. The emergence of convolutional neural networks (CNNs) has revolutionized image segmentation by automating feature extraction and learning from large data sets CNNs have shown impressive performance on complex models and systems a seen in Figs. However, despite their success, standard CNNs face challenges in capturing relevant multidimensional information, which is important for accurately classifying objects of different shape and sizes.

Submission: 16 October 2024; **Acceptance:** 10 December 2024

Literature Review

Early CNN-based approaches for semantic segmentation, such as the Fully Convolutional Network (FCN) (Chen et al., 2017), demonstrated the effectiveness the learning hierarchical features directly from raw pixel data. However, these models struggled to capture fine-grained details and maintain spatial resolution due to the inherent limitations of pooling operations. To address these challenges, researchers introduced atrous convolutions, also known as dilated convolutions, which enable the expansion of receptive fields without increasing the parameters or computations (Chen et al., 2018). This proved to be a breakthrough in semantic segmentation, allowing models to capture multi-scale context and improve object boundary delineation.

Later, (Chen et al., 2017) further refined the use of atrous convolutions by incorporating Atrous Spatial Pyramid Pooling (ASPP) modules in their DeepLab model (Long et al., 2015). ASPP enables the aggregation of features at multiple scales, enhancing the model's ability to handle objects of varying sizes. Subsequent iterations of DeepLab, such as the one proposed by (Chen et al., 2018), introduced additional refinements, including encoder-decoder architectures and depth wise separable convolutions, further improving accuracy and efficiency (Chen et al., 2017).

Parallel to the development of (Peng et al., 2017), (Chen et al., 2017) explored alternative approaches to leverage atrous convolutions for semantic segmentation (Peng et al., 2017). Their "Large kernel matters—improve semantic segmentation by global convolutional network" paper proposed a modified ASPP module with image-level features for capturing global context. Similarly, (Yang et al., 2018) introduced the "DenseASPP" model (2016), a densely connected ASPP module for capturing multi-scale context at different resolutions.

In recent years, there has been a growing interest in understanding the impact of different convolution operations on semantic segmentation performance. (Yu et al., 2017) introduced dilated convolutions into ResNet architectures, resulting in the development of Dilated Residual Networks (Yang et al., 2018). This work contributed to the understanding of the trade-offs between accuracy and computational efficiency in semantic segmentation.

Methodology

In the proposed system, the automatic fabric U-Net method based on CNN (Convolutional Neural Network) algorithm for semantic image segmentation. Specifically designed for downloading 2D images into virtual 3D models, valuable for various applications such as material analysis and simulation The system uses the deep commutative neural network VGG16 to generate objective functions based on microstructural properties. Uses permutation operator to generate multiclass feature maps and display them on sandstone data sets with many features. The U-NET model is used for image segmentation and then reconstructed for visualization. The system achieves an overall high IoU (Intersection over Union) and accuracy rate, making it effective for semantic classification tasks. The encoder network includes:

- **Initial Convolutional Layers:** The encoder commences with a series of convolutional layers with progressively smaller kernel sizes and increasing channel depths. These layers extract fine-grained, localized attributes from the input image.

- **Atrous Spatial Pyramid Pooling (ASPP) Module:** We integrate an ASPP module at the end of the encoder. This module consists of simultaneous dilated convolutional layers with varying dilation rates. Each layer captures context at a specific scale, contributing to a multiscale feature representation.
- **Feature Concatenation:** The outputs from all dilated convolutional layers within the ASPP module are concatenated along the channel dimension, combining multi-scale information into a single feature map.

On the other hand, the decoder network:

- **Gradual Up sampling:** The decoder up samples the concatenated feature map from the encoder using a sequence of transposed convolutions or bilinear interpolation. Each up sampling step is followed by a conventional convolutional layer to enhance the spatial details.
- **Skip Connections:** Skip connections are established between corresponding layers in the encoder and decoder. These connections directly pass low-level features from the encoder to the from the encoder towards the decoder, facilitating the recovery of fine-grained information lost during down sampling.

Results and Discussions

The U-NET model achieved a high accuracy with overall accuracy of 95.73% for the image classification task. This indicated a high ability of the model to correctly classify the images into segments. The average overlap (IoU) on association shows the study reported an average IoU of 0.8463 (about 85%) for the sample images. This metric is important for evaluating the model's performance in multiclass segmentation, which means high accuracy in distinguishing between classes in different images. The efficiency of U-NET algorithms with scholarships have shown that U-NET algorithms are effective in identifying and classifying features in an image. The Training and validation loss diagram showing that the system is successfully trained, with the loss gradually decreasing at all ages, indicating good model performance and generalizability. The Class-specific IoU values shows the study provided IoU values for individual classes, which showed high accuracy for most classes. Figure 1 shows the atrous convolution and mean IoU.

- First Class 0.8967 (90%).
- Class 2: 0.6293 (63%).
- Class 3: 0.9571 (96%).

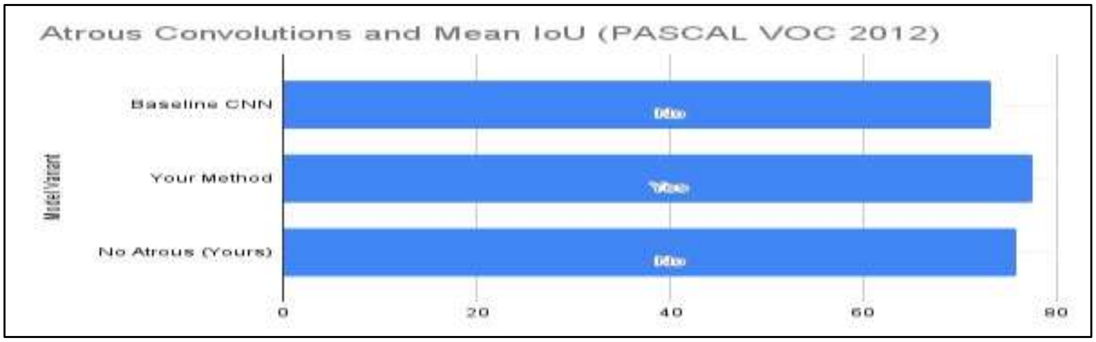

Figure 1: Atrous Convolution And Mean IoU.

Conclusion

As a conclusion the semantic image segmentation using CNN (Convolutional Neural Network) based Technique" highlights the highly effective U-NET algorithm in semantic segmentation tasks across different applications U-NET model proved to be a tool which is essential for robust image classification Union (IoU) offers significant advantages is that the authors also discuss future system implementations as well as measurement accuracy highlights Nevertheless, it takes a continuous challenge in understanding acknowledging the development of robust classification methods, the paper highlights advances in visual highlighting the important contributions and promising capabilities of the U-Net model in estimation.

Acknowledgement

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) pp. 833-851. https://doi.org/10.1007/978-3-030-01234-2_49

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

Han, D., Kim, J., & Kim, J. (2017). Deep pyramidal residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* pp. 5927-5935. https://doi.org/10.1109/CVPR.2017.668

Liu, S. A., Xie, H., Xu, H., Zhang, Y., & Tian, Q. (2022). Partial class activation attention for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 16836-16845. https://doi.org/10.1109/CVPR52688.2022.01633

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 3431-3440. https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298965

Peng, C., Zhang, X., Yu, G., Luo, G., & Sun, J. (2017). Large kernel matters--improve semantic segmentation by global convolutional network. In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1743-1751 https://doi.org/10.1109/CVPR.2017.189

Wang, Y., Lv, K., Huang, R., Song, S., Yang, L., & Huang, G. (2020). Glance and focus: a dynamic approach to reducing spatial redundancy in image classification. *Advances in Neural Information Processing Systems*, *33*, 2432-2444. https://dl.acm.org/doi/abs/10.5555/3495724.3495929

Yang, M., Yu, K., Zhang, C., Li, Z., & Yang, K. (2018). Denseaspp for semantic segmentation in street scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3684-3692). https://doi.org/10.1109/CVPR.2018.00388

Yu, F. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122. https://doi.org/10.48550/arXiv.1511.07122

Yu, F., Koltun, V., & Funkhouser, T. (2017). Dilated residual networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 472-480). https://doi.org/10.1109/CVPR.2017.75