International Journal of Information Technologies and Systems Approach
Volume 17 • Issue 1

Integrated Design of Building Environment Based on Image Segmentation and Retrieval Technology

Zhou Li, Hubei University of Technology, China

Hanan Aljuaid, Princess Nourah bint Abdulrahman University, Saudi Arabia*

https://orcid.org/0000-0001-6042-0283

ABSTRACT

Existing models still exhibit a deficiency in capturing more detailed contextual information when processing architectural images. This paper introduces a model for architectural image segmentation and retrieval based on an image segmentation network. Primarily, spatial attention is incorporated into the U-Net segmentation network to enhance the extraction of image features. Subsequently, a dual-path attention mechanism is integrated into the U-Net backbone network, facilitating the seamless integration of information across different spaces and scales. Experimental results showcase the superior performance of the proposed model on the test set, with average dice coefficient, accuracy, and recall reaching 94.67%, 95.61%, and 97.88%, respectively, outperforming comparative models. The proposed model can enhance the U-Net network's capability to identify targets within feature maps. The amalgamation of image segmentation networks and attention mechanisms in artificial intelligence technology enables precise segmentation and retrieval of architectural images.

KEYWORDS

Architectural Environment Integration Design, Artificial Intelligence, Attention Mechanism, Image Segmentation, U-Net

Image-segmentation techniques are extensively employed in architectural design due to the rapid advancement of artificial-intelligence technology. Researchers commonly segment image regions through manual or machine-learning methods. The threshold-based segmentation method (Du et al., 2023; Wang et al., 2023) delineates the image's grayscale histogram by selecting various grayscale thresholds. Pixels within the same grayscale range are considered part of the same class, sharing inherent similarities. The edge-based segmentation method (Li et al., 2010; Khan et al., 2023; Maican et al., 2023) necessitates the identification of the edge starting point, followed by a search for and connection of surrounding edge points from the starting point based on a similarity criterion. The region-based segmentation method (Liu et al., 2020; Xu et al., 2022; Li et al., 2023) relies on the spatial information of the image, constructing the segmentation region based on the similarity features of pixels. The graph partitioning approach in the graph theory-based segmentation method (Pei et al., 2020; Mamatha et al., 2022) completes the segmentation process by determining the optimal

DOI: 10.4018/IJITSA.340774

*Corresponding Author

solution to the goal function. Thus, the application of image-segmentation techniques in architectural design facilitates a more intuitive understanding of architectural construction and design methods for relevant practitioners.

The evolution of cities is contingent upon the interplay between the architecture people inhabit and their environment. As population density escalates, the development and layout of cities become increasingly intricate. Modern cities resemble networks, encompassing elements such as population, transportation, architecture, nature, industry, land, and water. Consequently, in urban planning and architectural design, designers must not only contemplate architectural design but also address and enhance the natural environment upon which we rely. The integration of the concept of environmental protection with architectural design is imperative. Architectural images harbor comprehensive information about architecture and environments, underscoring the necessity of employing image segmentation technology to precisely delineate architectural styles and environmental elements. This ensures the provision of more-nuanced image information and design concepts for architectural designers (Yin et al., 2022; Wang et al., 2022; Lüddecke & Ecker, 2022).

This paper introduces a model for architectural image segmentation and retrieval based on an image-segmentation network, specifically tailored for multi-perspective scenarios. The primary contributions include:

(1) Integration of spatial attention and dual-path contextual attention into the U-Net segmentation network to extract additional contextual architectural image features, thereby enhancing the network's ability to identify targets within feature maps
(2) Training the improved U-Net segmentation network on a self-constructed dataset of architectural images in this study, leading to the optimization of network parameters

The proposed approach proves effective in achieving accurate segmentation and retrieval of architectural images. This capability holds promise in assisting architectural designers in crafting superior urban architectural environment integration design solutions.

RELATED WORK

Prior to the integration of image-segmentation and retrieval technology into the realms of architectural design and urban development, architects seeking a comprehensive understanding of the overall design schemes and aesthetic styles of integrated architectural environments typically engaged in discussions with peers and consulted relevant literature. However, these conventional methods proved inadequate in meeting the sensory requirements of urban design concerning architectural style and green environments. With the introduction and application of image-segmentation and -retrieval technology, a novel solution has emerged for this challenging issue. For architects, the segmentation and retrieval of architectural images offer a superior means of acquiring relevant knowledge about integrated architectural environment design, thereby propelling the development of green cities. Consequently, the key focus shifts to the construction of an intelligent and efficient architectural image-segmentation and -retrieval model.

In recent years, the application of deep learning–based semantic image segmentation has become widespread across various domains. This approach is employed primarily to address issues such as fuzzy boundaries, low precision, and low resolution in images. When image-segmentation techniques are applied to architectural images, the model is expected not only to accurately delineate specific architectural features and refine architectural categories but also to assist designers in obtaining more-precise design solutions.

Deep learning–based semantic segmentation of images (Ulku and Akagündüz, 2022; Hemamalini et al., 2022) has witnessed widespread adoption across various domains, effectively addressing issues such as fuzziness and low resolution in images. Erdi et al. (1997) introduced an end-to-end neural network for semantic image segmentation. Li et al. (2019) proposed a U-Net network structure based on fully convolutional networks (FCNs), better suited for fine image processing. Unlike the summation mechanism of FCN, U-Net utilizes multiple upsampling and downsampling operations to gradually acquire high-level semantic information. It also incorporates jump connections (stitching dimensions of the same channels together), thereby enhancing feature fusion and significantly improving segmentation performance. While U-Net has demonstrated success in image segmentation, its limitations in extracting detailed contextual information have led to the proposal of new structures with U-Net as a variant. For instance, Duan et al. (2018) designed a lightweight SegNet model, introducing a novel upsampling method for efficient image segmentation.

The UNet++ network, an extension of U-Net, represents a notable breakthrough in image-segmentation technology. This network efficiently addresses the adaptive selection of sampling depth among different samples, accelerating the extraction of feature information at various levels. However, it comes with a drawback of an abrupt increase in the number of model parameters, leading to heightened computational costs and a significant demand for GPU resources (Zhou et al., 2018). As network models deepen, Tan et al. (2021) proposed an AcuNet network, utilizing depth-separable convolution to reduce model parameters. Trebing et al. (2021) introduced an At-UNet segmentation network, incorporating an attention mechanism based on U-Net and employing depth-wise convolution instead of traditional convolution. Cao and Zhang (2020) proposed an updated Res-UNet model for high-resolution image segmentation. He et al. (2020) presented a hybrid attention approach for effective architecture segmentation. Zhao et al. (2022) introduced an Inception v3–based image-segmentation method to enhance the segmentation accuracy of small target images effectively. Zhao et al. (2017) proposed a pyramid-shaped scene-parsing network, integrating contextual data and fully exploiting global features for semantic segmentation of diverse scenes. He et al. (2017) introduced mask R-CNN for image segmentation, achieving high-quality semantic segmentation while performing target detection.

However, the aforementioned study on establishing intelligent and efficient architectural image-segmentation and -retrieval models also exhibits some crucial shortcomings and research gaps. First, despite the incorporation of various advanced neural network structures and their variants, the model still lacks in-depth extraction of more-detailed contextual information when processing architectural images. There is a need for a more comprehensive consideration of effective global-information extraction. Second, some structures introducing additional parameters result in an increase in the computational cost of the model, potentially requiring substantial GPU resources, thereby limiting the model's applicability in certain environments. Additionally, the segmentation accuracy for small target images is not adequately emphasized, warranting further in-depth research. Furthermore, while the review mentions the model's requirements in terms of segmenting architectural features and refining architectural categories, it does not delve into how the model practically supports architects in obtaining more-precise design solutions. This represents a research gap that merits attention in terms of the practical support for architectural design.

METHODOLOGY

First, the improved U-Net segmentation network is trained on a self-constructed dataset of architectural images to optimize network parameters. Subsequently, semantic segmentation is applied to architectural images to acquire corresponding image features. On this foundation, an architectural image feature library is established. Finally, based on weighted Euclidean distance, a similarity measure is employed on architectural images to determine corresponding retrieval results. This process aims to provide architects with enhanced design insights for architectural environment integration.

U-Net

The encoder segment of the U-Net network is referred to as the compression path. In this compressed path, the feature map undergoes a total of four downsamplings, employing a 2×2 max-pooling method. Before each downsampling, the feature map goes through two convolution operations, reducing its edge length by 4 in each step. After four downsamplings and two convolution operations, the size of the feature map is diminished to 32×32. The decoder segment of the network is denoted as the extended path, taking a 32×32 feature map as input. The extended path involves four upsamplings, each followed by two convolution operations. To reduce feature dimensionality, the upsampled feature maps are concatenated with the feature maps from the compressed path. The resulting feature map then undergoes two convolution operations in the final module, generating the ultimate prediction map. The structure of the U-Net is illustrated in Fig. 1.

ResNet Backbone Network

The feature-extraction network employed in the enhanced U-Net image-segmentation framework in this paper is ResNet34, incorporating a skip connection mechanism between every two convolutional layers to provide the neural network with backward capability. The network structure is depicted in Fig. 2. A softmax classifier has been added to the final layer,

Figure 1. U-Net Structure

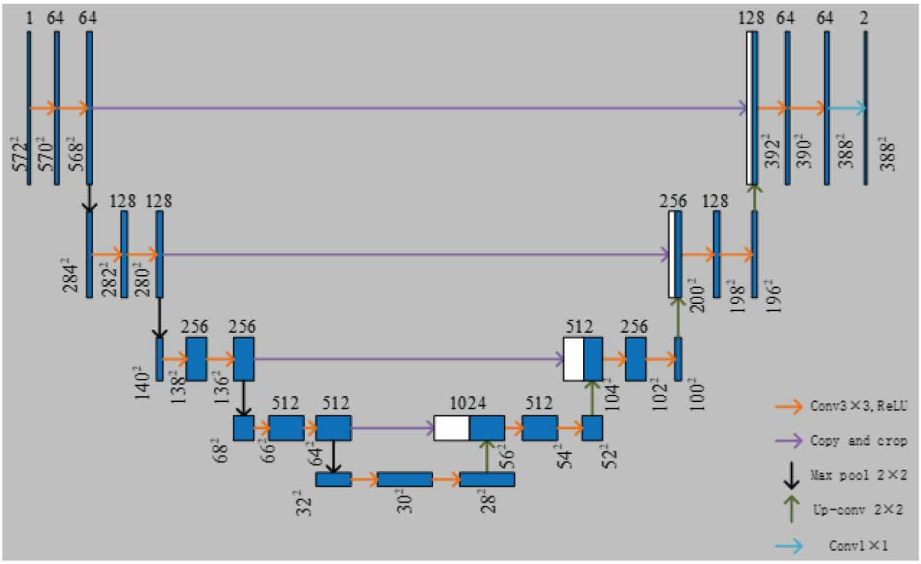
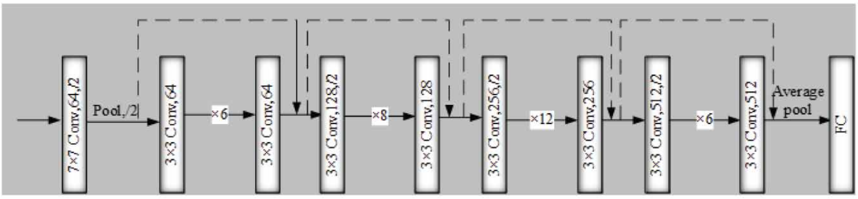


Figure 2. ResNet34 Network Structure



representing a fully connected layer. While the dropout layer is not explicitly displayed, it is integrated into the convolutional layer.

During operation, the pooling layer samples the input feature maps, effectively reducing the number of connected units in adjacent convolutional layers to facilitate subsequent computational analysis. Notably, in consideration of the unique characteristics of architectural images, the maximum pooling method is selected when designing the network. This choice aims to mitigate the impact caused by parameter errors in the convolutional layer and ensure the adequate retention of texture information. To help architecture designers perceive intricate details within architectural images, the output of the fully connected layers is thoughtfully preserved.

Spatial Attention

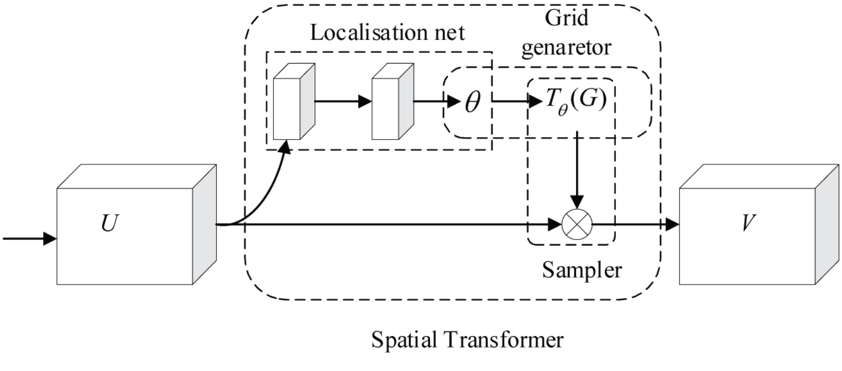
During the segmentation of architectural images, the network anticipates effective recognition of the target even when it undergoes rotation and distortion. Spatial transformer networks (Lin & Lucey, 2017) represent an early and impactful contribution to spatial attention. Prior to input of the feature map into the convolutional neural network, the option exists to rotate and scale the target. This augmentation enhances the convolutional neural network's proficiency in recognizing the target within the feature map. The spatial attention structure is delineated in Fig. 3.

Following the input of the feature map into spatial attention, the target within the image undergoes a sequence of operations, cropping, rotation, scaling, and translation, facilitated by the localization network. To effectuate these transformations, the coordinate values of the target are treated as a two-row, one-column matrix, multiplied by a two-row, two-column matrix, and subsequently added to a two-row, one-column matrix. Once the localization network completes the aforementioned transformations on the target, a grid generator converts the target image's coordinates back to the original image's coordinates. Subsequently, the gradient descent problem is addressed by a sampler. The collective interplay of the localization network, grid generator, and sampler constitutes a comprehensive spatial-attention mechanism.

Dual-Path Attention

Achieving accurate image segmentation requires consideration of the size and location of the target itself. However, the convolution operation in previous segmentation networks is limited to a local field of perception, hindering effective collection of contextual information and resulting in suboptimal network segmentation. To address this limitation, we propose the incorporation of dual-path attention to aggregate contextual information, replacing the bridge module in U-Net. The inclusion of dual-path attention in the model framework enhances the aggregation of context information by allowing the model to simultaneously focus on local content details and broader contextual information.

Figure 3. Spatial Attention



The dual-path attention comprises two parallel residual channel attention modules, emphasizing the contextual information of feature mapping, as illustrated in Fig. 4. The structure of the residual channel attention is detailed in Fig. 5.

Given an input feature mapping, dual-path attention scrutinizes the location of the specified target within the feature mapping. It selectively disregards other less relevant target features while learning essential information about the specified target. The application of dual-path attention enhances the overall model architecture, providing a more advantageous approach to improving model performance compared to merely increasing the depth of the model.

The residual channel attention module zeroes in on valuable information within the feature matrix, extracting more-discriminative features. In conventional convolution operations, constrained by the perceptual field's limited range, the convolution kernel can focus on only one local region of the feature mapping at a time. Consequently, it fails to capture contextual information beyond that localized region. To more effectively explore the semantics of the entire space of feature mappings, the residual channel attention module is employed.

This paper first uses a global averaging pooling operation to fuse the entire spatial information of the input feature mapping $M \in \mathbb{R}^{H \times W \times C}$ into a single channel identifier $I_c \in \mathbb{R}^{1 \times 1 \times C}$. From an implementation point of view, a channel identifier $I_c \in \mathbb{R}^{1 \times 1 \times C}$ is obtained by reducing a feature map from a two-dimensional space to a one-dimensional space.

To extract meaningful information from channel identifiers, this paper captures the nonlinear relationship between each channel within the identifier using two convolution processes. The first convolution layer is designed as a dimensionality-reduction layer, aiming to decrease the channel dimension of the channel identifier and regulate the complexity of the subsequent two convolution operations. Subsequently, the input feature mapping undergoes dimensional expansion through the second convolution layer. Following two convolution operations utilizing the sigmoid function, a

Figure 4. Dual-Path Attention

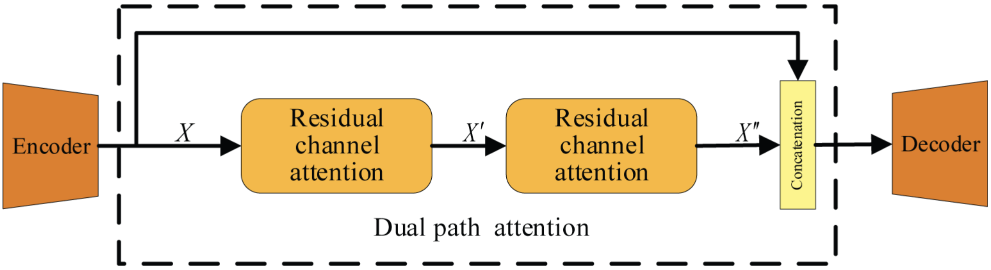
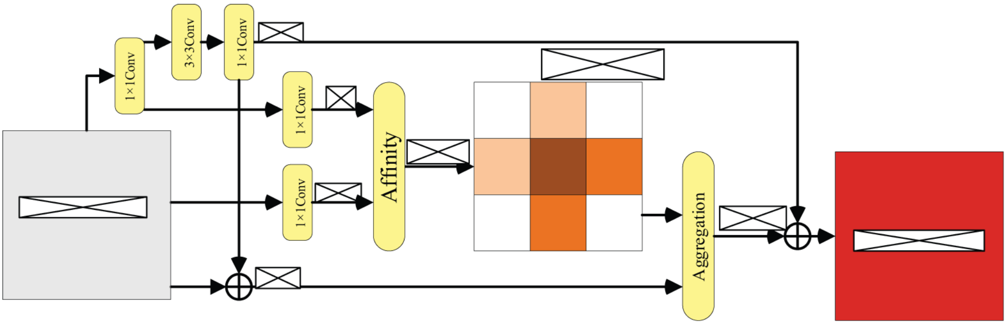


Figure 5. Residual Channel Attention



straightforward gating control is implemented on the channel. This process results in the creation of a logicalized channel information structure U'_c. The input feature mapping and the logicalized spatial structure $U_c \in \mathbb{R}^{H\times W\times C}$ are then used in a weighted sum operation to create the channel attention feature matrix:

$$U'_c = \sigma\left(W_2 L\left(W_1 I\right)\right) \tag{1}$$

$$U_c = U'_c \times M \tag{2}$$

$L\left(.\right)$ denotes the leaky ReLU function, W_1 and W_2 denote the weights of the first and second convolutions, respectively, and $\sigma\left(.\right)$ denotes the sigmoid function.

Ultimately, this paper integrates the information from the acquired channel attention feature mapping with the input feature mapping through a residual structure. This process aims to obtain the final channel attention feature mapping, mitigating the loss of information from the input feature mapping $\tilde{U}_c \in \mathbb{R}^{H\times W\times C}$.

$$\tilde{U}_c = U_c + M \tag{3}$$

Weighted Euclidean Distance

Upon the completion of architectural-image segmentation, this paper conducts similarity measures on the feature vectors between two architectural images utilizing weighted Euclidean distance. Subsequently, it discerns the specific features of architectural images along with their corresponding retrieval results. The traditional expression for Euclidean distance is as follows:

$$dist\left(X,Y\right) = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2} \tag{4}$$

The traditional Euclidean distance is observed to overlook the impact of differences between corresponding individual elements when representing the cumulative difference between two spatial vectors. Utilizing the Euclidean distance directly for similarity measures between feature vectors results in significant errors in metric accuracy, as it neglects the influence of element similarity within the feature vector. Since each element signifies a specific attribute of the target with a distinct meaning, it becomes essential to consider both the cumulative inaccuracy of the feature vector and the influence of its elements when measuring feature similarity.

To address this, the paper proposes a weighted Euclidean distance, offering enhanced distinguishability and greater alignment with human sensory perception compared to the traditional Euclidean distance-calculation method. The expression for the weighted Euclidean distance is as follows:

$$dist\left(X,Y\right) = \sqrt{\sum_{i=0}^{n} \frac{w_i}{W} (x_i - y_i)^2} \tag{5}$$

W is the normalization parameter, σ is the adjustment parameter, $\sigma = 1$ in the experiment, and w_i represents the weight, which is calculated as follows:

$$w_i = e^{\frac{\left|x_i - y_i\right|}{\sigma}} \tag{6}$$

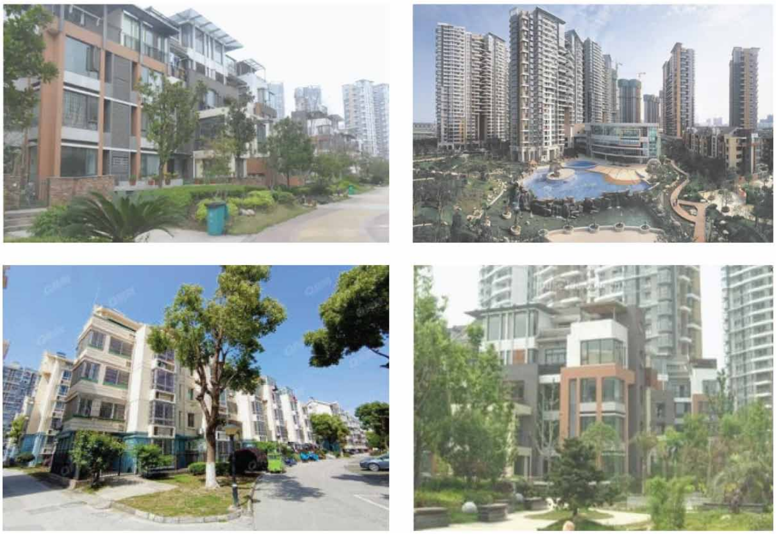
EXPERIMENTS AND RESULTS

The experimental environment for this paper is Windows 10, with an Intel i7-10875H processor, 16GB of memory, and an Nvidia GeForce RTX 2070 graphics card. TensorFlow 2.3 and CUDA 10.2 are utilized. The evaluation metrics commonly employed in image segmentation and retrieval—dice coefficient, accuracy, retrieval rate, and recall—are adopted as evaluation criteria. The proposed model's segmentation and retrieval performance is assessed by crawling 2,000 architectural images of urban areas from Baidu and Weibo. The images encompass various perspectives of the main body, each containing distinct green elements. An illustrative sample image is presented in Fig. 6. In addition, the ICCV 2017 detecting symmetry in the wild dataset is used for model performance comparison.

Data preprocessing plays a crucial role in computer vision and deep-learning research, aiming to enhance model performance and ensure the quality of input data. Our preprocessing steps include:

- Image cleaning and denoising to eliminate potential noise or irrelevant information
- Image scaling and cropping to standardize image dimensions and reduce unnecessary background
- Brightness and contrast adjustment to normalize image attributes
- Color standardization to ensure consistent color distribution
- Data augmentation through techniques such as rotation, flipping, and scaling to augment the dataset

Figure 6. Sample Dataset Diagram



Model-Segmentation Performance Analysis

The model's segmentation capability, as presented in this research, was evaluated on the dataset created for this paper. The dataset comprises 2,000 architectural images evenly distributed across five groups. Throughout the training process, one group was randomly selected as the test set, and this experiment was repeated five times. The results of these five experiments are detailed in Table 1.

The results presented in Table 1 highlight the impressive performance of the proposed model. Specifically, the average dice coefficient, accuracy, and recall metrics were observed to be 94.67%, 95.61%, and 97.88%, respectively. These metrics serve as robust indicators of the model's efficacy in image-segmentation tasks, particularly in the context of architectural images. The high values achieved across these key evaluation measures underscore the model's superior capability in accurately delineating and segmenting elements within architectural imagery. Such precision is indicative of the model's advanced ability to discern intricate details and boundaries, establishing it as a promising solution for tasks demanding precise image segmentation in the domain of architectural analysis and recognition.

Model-Retrieval Performance Analysis

To assess the effectiveness of the weighted Euclidean distance for retrieving architectural images in this paper's model, a test set of 100 images was randomly selected from the dataset. The retrieval results were utilized to calculate the area under the receiver operating characteristic curve (AUC). The proposed weighted Euclidean distance was compared to the Euclidean distance, semantic similarity metric, and visual similarity metric to evaluate its retrieval performance. Table 2 provides a comprehensive comparison of classification performance, showcasing the AUC with corresponding mean and standard deviation values for various metrics. Notably, the proposed model in this paper, labeled as "ours," outperforms other methods.

The AUC for the weighted Euclidean distance in the presented model is notably higher at 0.964 with a standard deviation of 0.008. In contrast, the AUC values for alternative methods such as Euclidean distance, semantic similarity metric, and visual similarity metric are 0.923, 0.931, and

Table 1. Model-Segmentation Performance

Evaluation Metrics	Cross-Validation					Average Value
	1	2	3	4	5	
Dice (%)	95.79	94.52	94.91	93.85	94.12	94.67
Accuracy (%)	96.66	95.48	95.85	94.89	94.97	95.61
Recall (%)	98.42	97.76	98.43	97.34	97.51	97.88

Table 2. Classification Performance Comparison

Name	AUC (Mean ± SD)
Euclidean distance	0.923 ± 0.008
Semantic similarity metric	0.931 ± 0.011
Visual similarity metric	0.930 ± 0.010
Ours	0.964 ± 0.008

0.930, respectively. This substantiates that the model introduced in this paper excels in architectural-image retrieval, surpassing the individual performance of semantic and visual-similarity measures. The higher AUC underscores the model's superior ability to retrieve relevant architectural images effectively.

Performance Comparison of Different Models

To evaluate the segmentation and retrieval capabilities of the model proposed in this paper, the following networks were selected for comparison: U-Net, ResU-Net (Sabir et al., 2022), AU-Net (Liu et al., 2022), UNet++ (Zhou et al., 2019), and CE-Net (Chen et al., 2021). Cross-comparison experiments were conducted between the aforementioned networks and the model in this paper. The mean values were taken as the comparison results, and these results are presented in Table 3.

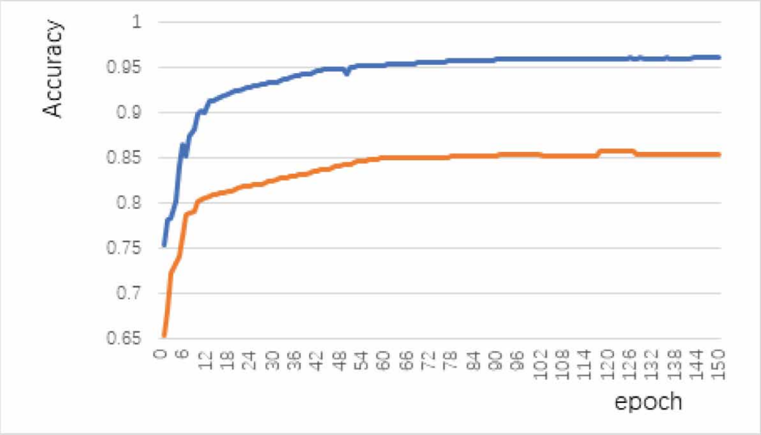
This paper demonstrates optimal performance across all evaluation metrics. Compared to the original U-Net, the model in this paper exhibits a 9.68% improvement in dice coefficient, a 12.36% enhancement in accuracy, and a 14.09% increase in recall. The introduced spatial attention and dual-path attention in this paper contribute to the U-Net segmentation network's ability to extract more precise characteristics from architectural images.

To visually highlight the segmentation performance superiority of the model in this paper, a comparative analysis is conducted through 150 training iterations between the model in this paper and the original U-Net. Accuracy curves are plotted for comparison, with the blue curve representing the model in this paper and the yellow curve representing the original U-Net. The accuracy comparison results are depicted in Fig. 7.

Table 3. Comparison of Different Models

Model	DICE	Accuracy	Recall
U-Net	86.31	85.09	85.79
ResU-Net	92.39	93.41	93.58
UNet++	93.16	93.78	94.02
AU-Net	93.17	94.59	94.74
CE-Net	93.49	94.71	94.89
Ours	94.67	95.61	97.88

Figure 7. Accuracy Curve



The convergence analysis reveals that the model in this paper exhibits a noteworthy trend. Accuracy stabilizes and approaches a consistent state starting from the 24th iteration, reaching a stable condition by the 60th iteration. This observation suggests that the model not only achieves accurate architectural-image segmentation and retrieval but also demonstrates robustness and generalization capabilities. The stability in accuracy after a relatively small number of iterations implies that the model has effectively learned the underlying patterns in the data, showcasing its ability to generalize well to new instances. This stability in performance signifies the reliability of the proposed model, instilling confidence in its consistent and accurate performance in architectural image–related tasks.

Model Segmentation and Retrieval Example Analysis

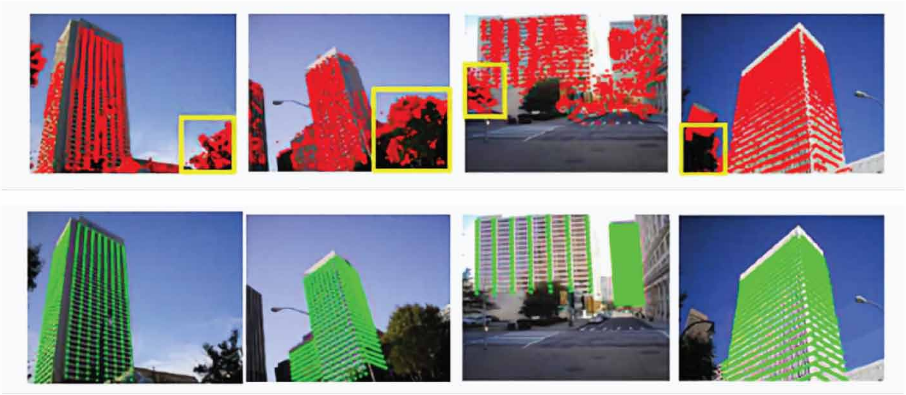
To underscore the outstanding capability of this model in segmenting and retrieving architectural images and surrounding environmental elements, the architectural-image data from the ICCV 2017 detecting symmetry in the wild dataset was employed to validate the feasibility of the model. Fig. 8 presents a comparison of the segmentation and retrieval results between this model and the original U-Net on this dataset. The retrieved architecture is labeled with green dots in U-Net, while this paper's model utilizes red dots for labeling. Additionally, the environment elements are designated with yellow borders.

The comparative analysis between the model proposed in this paper and the original U-Net underscores the enhanced capabilities of the former. Specifically, the presented model excels in identifying both environmental and architectural elements within images. In contrast, the original U-Net exhibits limitations in retrieving environmental elements surrounding architecture and displays partial omissions in architectural identification. This improvement in retrieval and segmentation, achieved by the model in this paper, holds significant implications for urban construction and design. The heightened accuracy in distinguishing architectural and environmental elements contributes to a more comprehensive understanding of the image content. Such advancements are crucial for informing decisions in architectural environment integration, enhancing the model's utility in urban planning and construction scenarios.

CONCLUSION

This paper introduces a U-Net-based model for architectural-image segmentation and retrieval. The deep fusion of spatial attention and dual-path contextual attention with the U-Net segmentation network enables the collection of contextual information from different scale feature maps, aiding the model in extracting higher-quality architectural image features. The model achieves an average dice

Figure 8. Comparison of Segmentation and Retrieval Results



coefficient, accuracy, and recall of 94.67%, 95.61%, and 97.88%, respectively, on a self-constructed dataset of architectural images, demonstrating accurate segmentation and retrieval capabilities.

The advantage of this deep fusion lies in its ability to suppress noise and redundant information, enhancing the network's robustness and generalization by reducing sensitivity to noncritical information. The introduction of multilevel contextual information empowers the network with enhanced semantic structure resolution, contributing to improved performance in handling fuzzy object boundaries and complex scenes. Moreover, this fusion strategy makes the network more adaptable to objects of different scales, simultaneously increasing segmentation efficiency while maintaining high accuracy. From the application level, the proposed method can help architectural designers to create more-excellent architectural environment integration design schemes and further optimize urban design concepts and paths.

However, for particularly complex scenarios such as dense occlusion and highly variable lighting conditions, the deep fusion network may still face challenges. This could result in insufficient robustness of the model in extreme situations, requiring further improvements to adapt to more-complex scenes. Additionally, the fused attention mechanisms may lead to increased computational and storage overhead when dealing with large-scale data, limiting their application in real-time scenarios and on edge devices. Addressing this issue may necessitate more-efficient algorithm design and model optimization.

AUTHOR NOTE

The authors thank the anonymous reviewers whose comments and suggestions helped improve this manuscript.

The authors declare that there are no conflicts of interest.

This work received no funding.

REFERENCES

Cao, K., & Zhang, X. (2020). An improved Res-UNet model for tree species classification using airborne high-resolution images. *Remote Sensing (Basel)*, *12*(7), 1128–1135. doi:10.3390/rs12071128

Chen, S., Niu, J., Deng, C., Zhang, Y., Chen, F., & Xu, F. (2021). CE-Net: A coordinate embedding network for mismatching removal. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 147634–147648. doi:10.1109/ACCESS.2021.3123942

Du, Y., Yuan, H., Jia, K., & Li, F. (2023). Research on threshold segmentation method of two-dimensional Otsu image based on improved sparrow search algorithm. *IEEE Access : Practical Innovations, Open Solutions*, *11*, 70459–70469. doi:10.1109/ACCESS.2023.3293191

Duan, L., Xiong, X., Liu, Q., Yang, W., & Huang, C. (2018). Field rice panicle segmentation based on deep full convolutional neural network. *Nongye Gongcheng Xuebao (Beijing)*, *34*(12), 202–209. doi:10.11975/j.issn.1002-6819.2018.12.024

Erdi, Y. E., Mawlawi, O., Larson, S. M., Imbriaco, M., Yeung, H., Finn, R., & Humm, J. L. (1997). Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer*, *80*(S12), 2505–2509. doi:10.1002/(SICI)1097-0142(19971215)80:12+<2505::AID-CNCR24>3.0.CO;2-F PMID:9406703

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2980–2988). IEEE Press. doi:10.1109/ICCV.2017.322

He, N., Fang, L., & Plaza, A. (2020). Hybrid first and second order attention Unet for building segmentation in remote sensing images. *Science China. Information Sciences*, *63*(4), 140305. doi:10.1007/s11432-019-2791-7

Hemamalini, V., Rajarajeswari, S., Nachiyappan, S., Sambath, M., Devi, T., Singh, B. K., & Raghuvanshi, A. (2022). Food quality inspection and grading using efficient image segmentation and machine learning-based system. *Journal of Food Quality*, *5262294*, 1–6. Advance online publication. doi:10.1155/2022/5262294

Khan, A., Garner, R., La Rocca, M., Salehi, S., & Duncan, D. (2023). A novel threshold-based segmentation method for quantification of COVID-19 lung abnormalities. *Signal, Image and Video Processing*, *17*(4), 907–914. doi:10.1007/s11760-022-02183-6 PMID:35371333

Li, C., Xu, C., Gui, C., & Fox, M. D. (2010). Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, *19*(12), 3243–3254. doi:10.1109/TIP.2010.2069690 PMID:20801742

Li, J., Chen, S., Zhang, Q., Song, Y., Gador, C., Sun, J., Xu, D., Zhao, X., Yuan, C., & Li, R. (2019). The study on the segmentation method of carotid vessel wall in multicontrast MR images based on U-Net neural network. *Chinese Journal of Radiology*, *12*, 1091–1095.

Li, N., Guan, C., Huang, X., Zhen, Q., Wang, A., Dai, X., & Zhang, Y. (2023). Factors influencing the willingness to accept health behavior and psychological monitoring systems in the milieu of information management technology. *Journal of Organizational and End User Computing*, *35*(1), 1–19. doi:10.4018/JOEUC.330020

Lin, C. H., & Lucey, S. (2017). Inverse compositional spatial transformer networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2568–2576). doi:10.1109/CVPR.2017.242

LiuC.GuoX.JiangJ. (2022). AU-Net: A deep learning network for precise water body extraction in the middle and lower reaches of the yellow river. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-3/W1-2022*, 107–113. 10.5194/isprs-annals-X-3-W1-2022-107-2022

Liu, X., Zhang, Y., Jing, H., Wang, L., & Zhao, S. (2020). Ore image segmentation method using U-Net and Res_Unet convolutional networks. *RSC Advances*, *10*(16), 9396–9406. doi:10.1039/C9RA05877J PMID:35497237

Lüddecke, T., & Ecker, A. (2022). Image segmentation using text and image prompts. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7086–7096). doi:10.1109/CVPR52688.2022.00695

Maican, C. I., Sumedrea, S., Tecau, A., Nichifor, E., Chitu, I. B., Lixandroiu, R., & Bratucu, G. (2023). Factors influencing the behavioural intention to use AI-generated images in business: A UTAUT2 perspective with moderators. *Journal of Organizational and End User Computing*, *35*(1), 1–32. doi:10.4018/JOEUC.330019

Mamatha, S. K., Krishnappa, H. K., & Shalini, N. (2022). Graph theory based segmentation of magnetic resonance images for brain tumor detection. *Pattern Recognition and Image Analysis*, *32*(1), 153–161. doi:10.1134/S1054661821040167

Pei, Y., Yang, W., Wei, S., Cai, R., Li, J., Guo, S., Li, Q., Wang, J., & Li, X. (2020). Automated measurement of hip–knee–ankle angle on the unilateral lower limb X-rays using deep learning. *Physical and Engineering Sciences in Medicine*, *44*(1), 53–62. doi:10.1007/s13246-020-00951-7 PMID:33252719

Sabir, M. W., Khan, Z., Saad, N. M., Khan, D. M., Al-Khasawneh, M. A., Perveen, K., Qayyam, A., & Azhar Ali, S. S. (2022). Segmentation of liver tumor in CT scan using ResU-Net. *Applied Sciences (Basel, Switzerland)*, *12*(17), 8650. doi:10.3390/app12178650

Tan, L., Ma, W., Xia, J., & Sarker, S. (2021). Multimodal magnetic resonance image brain tumor segmentation based on ACU-Net network. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 14608–14618. doi:10.1109/ACCESS.2021.3052514

Trebing, K., Staṅczyk, T., & Mehrkanoon, S. (2021). SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, *145*, 178–186. doi:10.1016/j.patrec.2021.01.036

Ulku, I., & Akagündüz, E. (2022). A survey on deep learning-based architectures for semantic segmentation on 2D images. *Applied Artificial Intelligence*, *36*(1), 2032924. doi:10.1080/08839514.2022.2032924

Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., & Nandi, A. K. (2022). Medical image segmentation using deep learning: A survey. *IET Image Processing*, *16*(5), 1243–1267. doi:10.1049/ipr2.12419

Wang, X., Wang, S., Guo, Y., Hu, K., & Wang, W. (2023). Coal gangue image segmentation method based on edge detection theory of star algorithm. *International Journal of Coal Preparation and Utilization*, *43*(1), 119–134. doi:10.1080/19392699.2021.2024173

Xu, Y., He, X., Xu, G., Qi, G., Yu, K., Yin, L., Yang, P., Yin, Y., & Chen, H. (2022). A medical image segmentation method based on multi-dimensional statistical features. *Frontiers in Neuroscience*, *16*, 1009581. doi:10.3389/fnins.2022.1009581 PMID:36188458

Yin, X. X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). U-Net-based medical image segmentation. *Journal of Healthcare Engineering*, *4189781*, 1–16. Advance online publication. doi:10.1155/2022/4189781 PMID:35463660

Zhao, H. S., Shi, J. P., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2881–2890). IEEE. doi:10.1109/CVPR.2017.660

Zhao, X., Li, X., Ye, S., Li, X., Feng, W., & You, X. (2022). Multi-scale tomato disease segmentation algorithm based on improved U-Net network. *Computer Engineering and Applications*, *58*(10), 216–223. doi:10.3778/j.issn.1002-8331.2105-0201

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, & A. Madabhushi (Eds.), Lecture notes in computer science: Vol. 11045. *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer. doi:10.1007/978-3-030-00889-5_1

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2019). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, *39*(6), 1856–1867. doi:10.1109/TMI.2019.2959609 PMID:31841402