



ASD-Net: a novel U-Net based asymmetric spatial-channel convolution network for precise kidney and kidney tumor image segmentation

Zhanlin Ji¹ · Juncheng Mu¹ · Jianuo Liu¹ · Haiyang Zhang² · Chenxu Dai¹ · Xueji Zhang³ · Ivan Ganchev^{4,5,6}

Received: 10 October 2023 / Accepted: 15 January 2024 / Published online: 8 February 2024
© The Author(s) 2024

Abstract

Early intervention in tumors can greatly improve human survival rates. With the development of deep learning technology, automatic image segmentation has taken a prominent role in the field of medical image analysis. Manually segmenting kidneys on CT images is a tedious task, and due to the diversity of these images and varying technical skills of professionals, segmentation results can be inconsistent. To address this problem, a novel ASD-Net network is proposed in this paper for kidney and kidney tumor segmentation tasks. First, the proposed network employs newly designed Adaptive Spatial-channel Convolution Optimization (ASCO) blocks to capture anisotropic information in the images. Then, other newly designed blocks, i.e., Dense Dilated Enhancement Convolution (DDEC) blocks, are utilized to enhance feature propagation and reuse it across the network, thereby improving its segmentation accuracy. To allow the network to segment complex and small kidney tumors more effectively, the Atrous Spatial Pyramid Pooling (ASPP) module is incorporated in its middle layer. With its generalized pyramid feature, this module enables the network to better capture and understand context information at various scales within the images. In addition to this, the concurrent spatial and channel squeeze & excitation (scSE) attention mechanism is adopted to better comprehend and manage context information in the images. Additional encoding layers are also added to the base (U-Net) and connected to the original encoding layer through skip connections. The resultant enhanced U-Net structure allows for better extraction and merging of high-level and low-level features, further boosting the network's ability to restore segmentation details. In addition, the combined Binary Cross Entropy (BCE)-Dice loss is utilized as the network's loss function. Experiments, conducted on the KiTS19 dataset, demonstrate that the proposed ASD-Net network outperforms the existing segmentation networks according to all evaluation metrics used, except for recall in the case of kidney tumor segmentation, where it takes the second place after Attention-UNet.

Keywords Kidney tumor · Image segmentation · Medical image analysis · Neural network · U-Net

1 Introduction

Recently, the cases of kidney tumors have increased worldwide [1, 2]. As one of the malignant tumors affecting the human urinary reproductive system, radical nephrectomy is the most commonly used method in clinical treatment.

However, this procedure can result in post-operative issues, causing daily life inconveniences. Hence, kidney-preserving surgeries are commonly used in tumor treatments [3, 4], necessitating the accurate segmentation of the kidney and kidney tumor before surgery [5, 6]. However, challenges such as low contrast in the tissue details and the irregular

✉ Xueji Zhang
zhangxueji@szu.edu.cn

✉ Ivan Ganchev
ivan.ganchev@ul.ie

¹ Department of Artificial Intelligence, North China University of Science and Technology, Tangshan 063009, People's Republic of China

² Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, People's Republic of China

³ School of Biomedical Engineering, Shenzhen University Health Science Center, Shenzhen, Guangdong 518060, People's Republic of China

⁴ Telecommunications Research Centre (TRC), University of Limerick, Limerick V94 T9PX, Ireland

⁵ Department of Computer Systems, University of Plovdiv "Paisii Hilendarski", Plovdiv 4000, Bulgaria

⁶ Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia 1040, Bulgaria

shape of the kidney and tumor in abdominal scanning images can make precise segmentation difficult [7–10]. Artificial intelligence (AI) has shown potential in assisting clinical care as an effective way to detect and diagnose abnormal areas [11–13]. The most advanced AI algorithms can effectively distinguish between benign and malignant kidney masses on computed tomography (CT) scans [14–16]. However, the performance of AI algorithms still needs to be improved when distinguishing different subtypes of kidney cancer or evaluating tumors [17, 18].

With the significant progress made by deep learning (DL) in the fields of computer vision [19], natural language processing (NLP), and other areas, many problems in the field of medical image segmentation are being addressed using convolutional neural networks (CNNs) [20–22]. CNNs are essentially based on feature extraction algorithms that can learn the high-level semantic characteristics of various tissue/organ images from a large number of annotated CT images [23], realizing semantic segmentation of CT images [24, 25]. It is particularly noteworthy that the emergence of U-Net [26], a type of encoder-decoder structured neural network, has laid an important foundation for the development of subsequent biomedical image semantic segmentation networks. The introduction of U-Net marks that, in complex medical image analysis tasks, researchers are no more limited to traditional segmentation strategies, but can use the power of DL to extract and utilize richer and more complex feature information, thereby improving the accuracy and robustness of segmentation [27].

U-Net, proposed by Ronneberger et al. in [26], is a DL-based image segmentation network that is able to achieve precise image segmentation by utilizing an encoder-decoder architecture and skip connections. U-Net has shown excellent performance in many medical image segmentation tasks, including kidney tumor segmentation. Zhou et al. [28] proposed U-Net++, which consists of a series of U-Nets with different depths and decoders. These decoders are densely connected at the same resolution through redesigned skip connections. Although with better segmentation performance, U-Net++ is a very complex network, requiring additional learnable parameters, and some of its components are redundant for specific tasks [29]. Oktay et al. [30] incorporated skip connections with focusing gates into a U-shaped structure for medical image segmentation, whereby an attention gate (AG) implicitly generates soft region suggestions, emphasizing relevant task features. Diakogiannis et al. [31] proposed a deep residual U-Net based network, named ResUNet-a, which uses a series of stacked residual units to replace the ordinary neural units as basic blocks, effectively deepening the network training layers. However, as the network depth increases, the training time also significantly increases. Researchers also

consider introducing a self-attention mechanism in CNNs to improve network performance [29]. Çiçek et al. [32, 33] extended U-Net to three dimensions, proposing 3D U-Net. Chen et al. [34] proposed the DeepLab network, a DL network used for image segmentation, which utilizes dilated convolution and Atrous Spatial Pyramid Pooling (ASPP) structure to capture multi-scale contextual information, and can better handle complex image segmentation tasks such as kidney tumor segmentation. However, this method cannot allow the network to efficiently use the target features extracted. The abdominal CT images contain a large amount of complex background. If the features of the kidney and tumor images cannot be accurately extracted and efficiently used, irrelevant background information will greatly affect the final segmentation result.

This paper proposes a novel ASD-Net network, based on U-Net, which shows an enhanced segmentation performance. The main contributions of the paper are reflected in the following aspects:

- 1) An innovative combination of the asymmetric convolution and the concurrent spatial and channel squeeze & excitation (scSE) attention gate is proposed, forming a novel Adaptive Spatial-channel Convolution Optimization (ASCO) block for incorporation into U-Net. Without adding much to the computational complexity, this block enhances the network's ability in complex pattern recognition and global context information understanding, thereby improving its image segmentation performance.
- 2) A novel Dense Dilated Enhancement Convolution (DDEC) block is proposed for incorporation into U-Net, in which the last 3×3 convolution of the dense convolution block is replaced with a dilated convolution, followed by a spatial and channel dual attention. This allows the network to effectively enlarge the receptive field, enabling it to capture a broader range of contextual information while keeping the number of parameters and computational complexity unchanged. In addition, it can also recalibrate the feature maps, enhance the important features, suppress the irrelevant features, and enhance the network's segmentation performance.
- 3) To reduce the impact of noise, it is proposed to get rid of the skip connections at the top layer of the network in order to pay more attention to the learning of deep features. This way the lower-level features, which may contain a lot of noise and irrelevant information, are directly passed to the upper layers, thus positively affecting the segmentation performance of the network.
- 4) Inspired by N-Net [35], a dual-encoder/single-decoder backbone structure is utilized for the proposed ASD-Net network.

2 Related work

2.1 Image preprocessing

Data preprocessing and data augmentation are two key strategies for improving the training efficiency of DL networks. These strategies mainly improve the quality of the training sample set and make the networks more effectively adapted to the distribution of training data characteristics. Through data augmentation, one can generate more complex training data, which improves the network generalization ability and helps prevent overfitting [36]. Since the KiTS19 dataset, used in the conducted experiments, consists of three-dimensional (3D) images, but the elaborated segmentation network is two-dimensional (2D), the data were first converted from 3D to

2D. During the slicing conversion process, aimed at enhancing the details of the target tissue in the images, the window width and position were initially adjusted as follows: any CT value greater than 500 HU (Hounsfield units) was set to 500 HU, and any CT value less than -200 HU was set to -200 HU. After completing the slicing, pictures with kidneys or tumors were selected based on the marked images, invalid images were deleted, and random rotation, horizontal or vertical flipping on the training set’s images were performed before the network training, as shown in Fig. 1.

An illustrative example of adjusting the window width and window level is depicted in Fig. 2.

A sample random rotation of an input image is shown in Fig. 3 and a sample random horizontal and vertical flipping (with a probability of 0.5) are illustrated in Fig. 4.

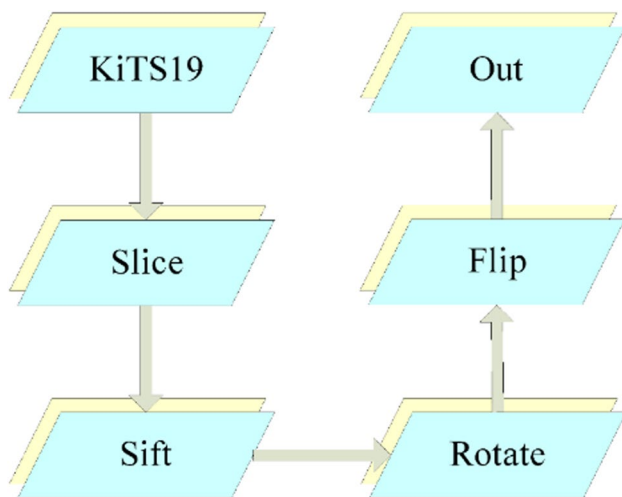


Fig. 1 The utilized image preprocessing sequence

2.2 U-Net

The classic architecture of U-Net is composed of two parts: (i) feature extraction, designed with inspiration from the VGG network [37, 38], where each pooling layer contains a specific scale, with a total of five different scales based on the original image; and (ii) upsampling, where each stage performs an upsampling operation and fuses information with the corresponding channels from the feature extraction part through skip connecting [39]. The encoder-decoder structure of U-Net allows it to capture and merge contextual information at different abstraction levels. This contextual information plays a key role in enhancing semantic understanding and accuracy of image segmentation.

2.3 scSE attention mechanism

The attention mechanism, initially used in machine translation, was quickly applied to the field of computer vision due

Fig. 2 An illustrative result of adjusting the window width and window level: **a** before the adjustment; **b** after the adjustment

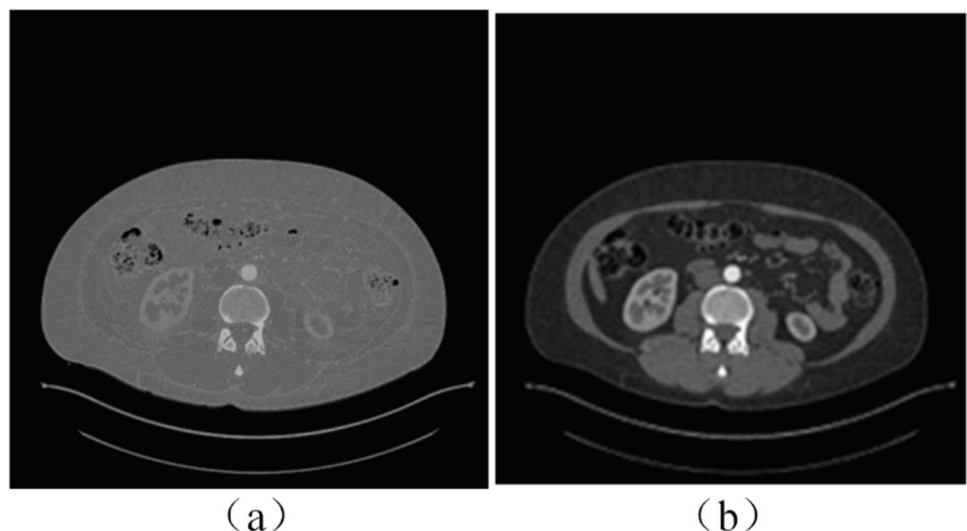


Fig. 3 Sample image rotation: **a** original image; **b** clock-wise rotation of 90°

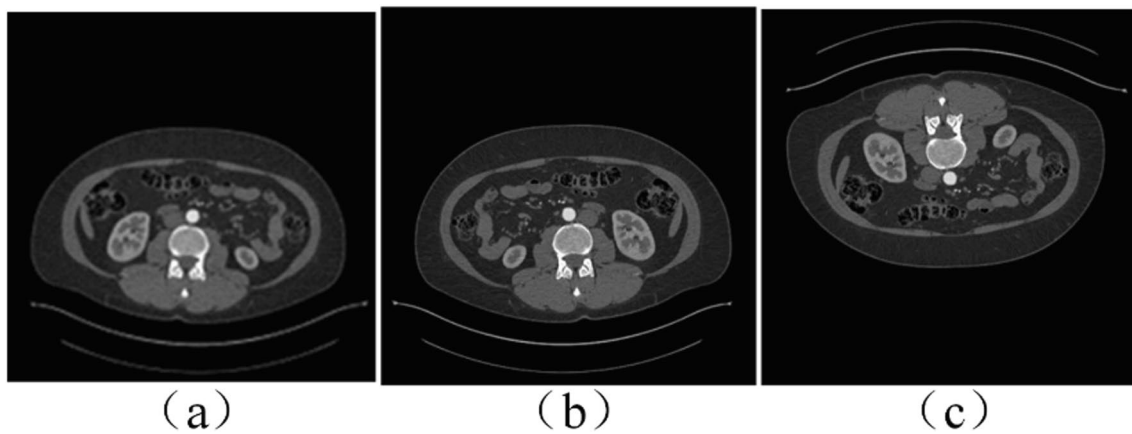
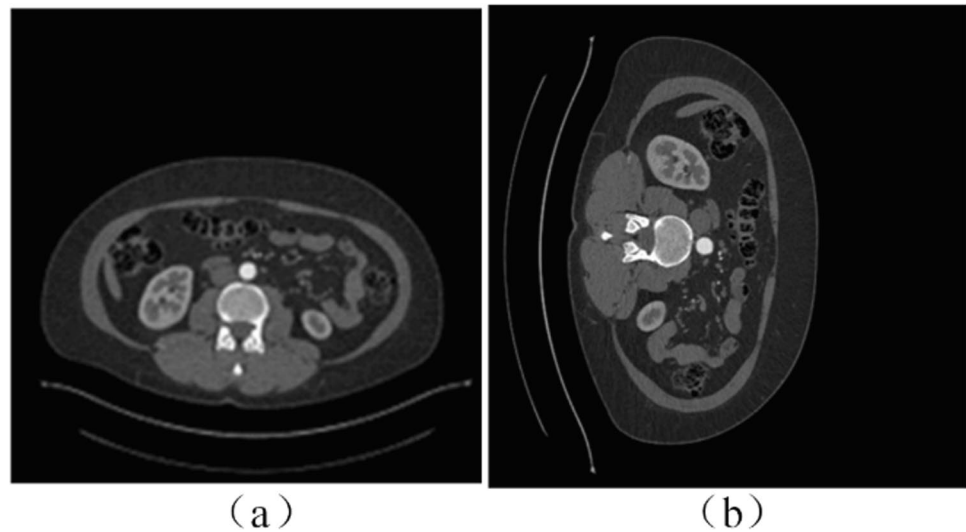


Fig. 4 Sample image horizontal and vertical flipping: **a** original image; **b** horizontal flipping; **c** vertical flipping

to its outstanding performance. Today, the attention mechanism has become a common means for enhancing the neural networks [40]. By combining channel and spatial attention, the scSE attention mechanism [41], depicted in Fig. 5, significantly enhances the network learning ability. Its design strategy adjusts the importance of features at different levels, allowing a network to learn more valuable high-level features, while paying less attention to features that have less impact on the target task. This strategy allows to obtain richer spatial and channel information at the pixel level. The relative importance of attention in both dimensions is simultaneously adjusted, thus further improving the accuracy of downstream registration tasks [42].

2.4 Loss functions

In the conducted experiments, described further in this paper, a combined Binary Cross Entropy (BCE)-Dice

loss [43] is used as a loss function. This design of the loss function aims to take into account both the pixel-level classification loss and the region-level overlap loss in order to assess the segmentation performance more comprehensively.

Specifically, BCE is a commonly used loss function for binary classification problems, measuring the gap between the predicted output p_i and the actual label y_i , as follows [44]:

$$L_{BCE} = - \sum_i (y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)) \quad (1)$$

However, when dealing with imbalanced data—for instance, when most areas in the image are background with only a small part being tumors—it could cause a network to predict the majority categories, potentially “ignoring” the minority categories. Thus, when dealing with image segmentation tasks with uneven category

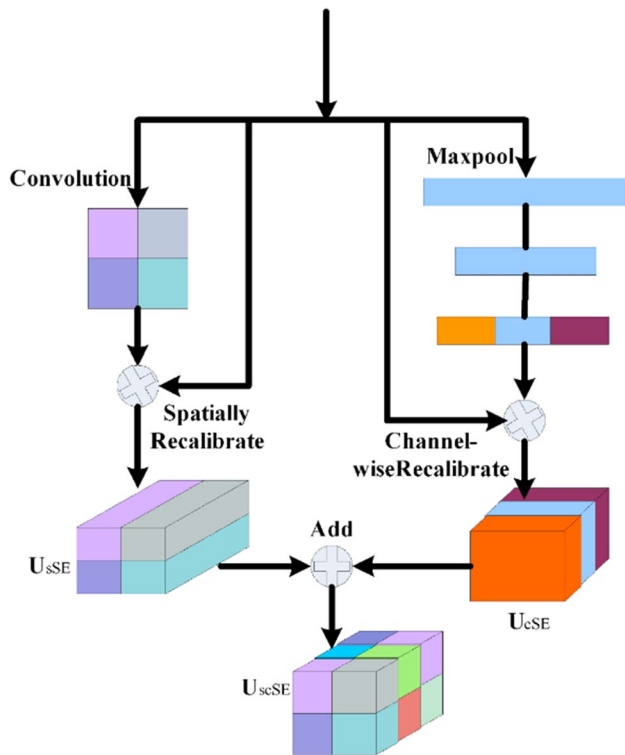


Fig. 5 The scSE structure

distribution or large blank background areas, using BCE alone may result in poor predictive performance for minority categories.

To solve this problem, the Dice loss could be used in addition to the BCE loss. It is based on the Dice similarity coefficient (DSC), which is a commonly used metric to measure the similarity between two samples, especially suitable for dealing with category imbalance problems in medical image segmentation tasks. The Dice loss is defined in [45] as follows:

$$L_{Dice} = 1 - 2 \frac{\sum_{i=1}^N y_i p_i}{\sum_{i=1}^N y_i^2 + \sum_{i=1}^N p_i^2} \tag{2}$$

To attain precise segmentation of the kidney and tumors, and to overcome challenges such as slow network convergence, gradient disappearance during backpropagation, and class imbalance, a combined loss function made of these two loss functions is used for optimizing the network training, as follows:

$$L = 1/2L_{BCE} + L_{Dice} \tag{3}$$

The benefit of using this combined loss function is that when the prediction results are closer to the actual labels, or the overlap between the predicted area and the actual area is higher, the value of the BCE-Dice loss is smaller than that of each individual loss.

3 Proposed network

3.1 Overall structure

The ASD-Net network, proposed in this paper, is based on U-Net, with the following improvements, as shown in Fig. 6:

- 1) The incorporation of newly designed ASCO blocks into the U-Net structure allows to effectively capture the anisotropic properties of input images, thereby adapting the network to their inherent asymmetry and scale differences. This lays a solid foundation for subsequent feature extraction and fusion.
- 2) The addition of newly constructed DDEC blocks to U-Net not only enables drawing on the advantages of the dense connection structures in feature propagation and feature reusing but also effectively expanding the receptive field and deeply integrating and extracting different levels of features without increasing the computational complexity related to the use of dilated convolution, thereby improving the segmentation performance.
- 3) The incorporation of a scSE attention mechanism into U-Net endows it with an ability to analyze and process contextual information in input images more deeply. The scSE mechanism significantly enhances the network’s ability to capture key spatial and channel information, thus further improving the segmentation results.
- 4) The addition of an extra encoding layer in the U-Net left side and the removal of the top skip connection between the encoding and decoding layers allows emphasizing on the network’s extraction and utilization of advanced features by fusing features at different levels, thus further enhancing the network’s segmentation performance and its ability to recover details.
- 5) The incorporation of an ASPP pooling pyramid module [34] between the encoding and decoding layers, for performing multi-scale spatial sampling on the input in parallel, enables extracting rich global contextual information that positively affecting the network’s segmentation ability at multiple scales.

These U-Net improvements are described in the following subsections in greater detail.

3.2 ASCO block

The design of the ASCO block aims to more effectively extract and use feature spatial and channel information in order to improve the segmentation performance. The ASCO core idea is to use scSE with asymmetric convolution (Fig. 7).

The scSE component acts on the spatial and channel axes of the input feature maps, generating two activation

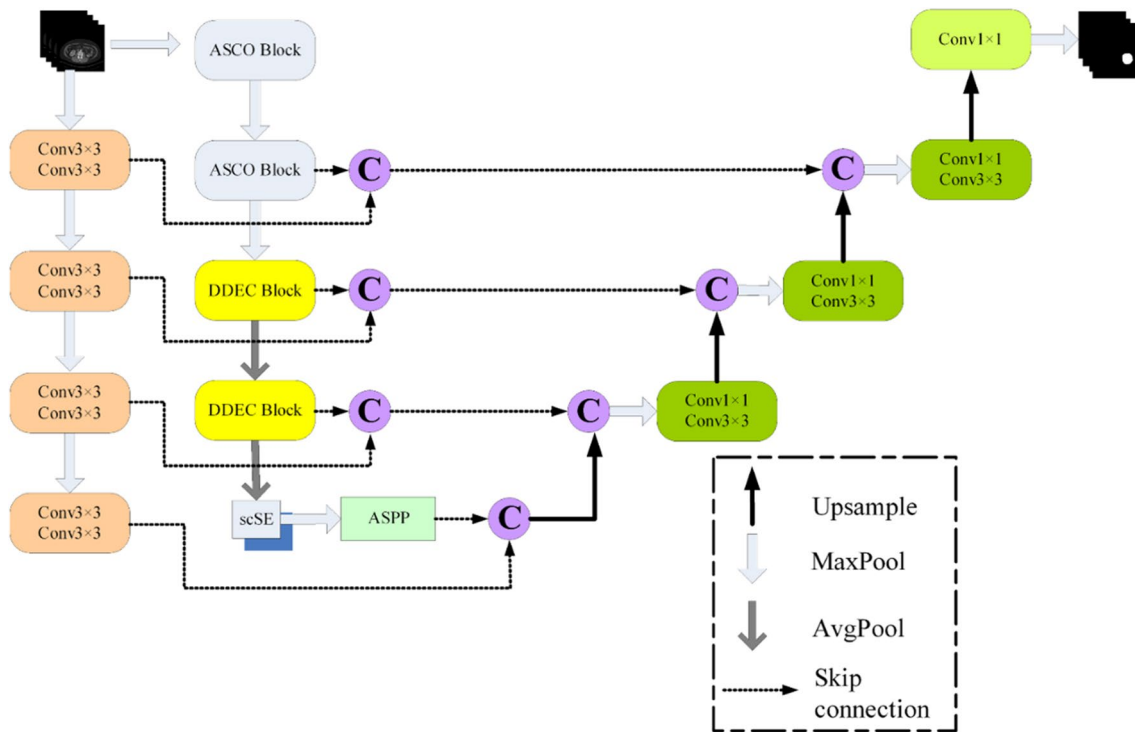
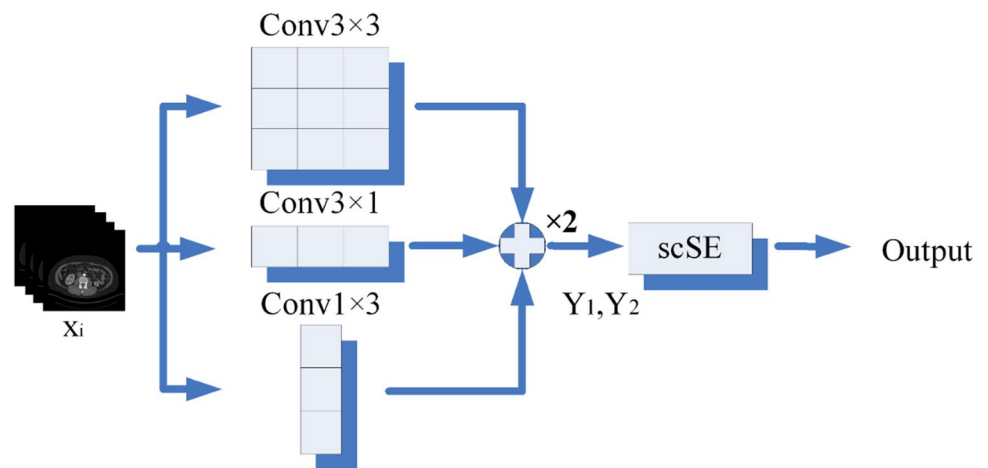


Fig. 6 The proposed ASD-Net network

Fig. 7 The proposed ASCO block



maps, and then multiplies these with the input feature map to recalibrate the features. This adaptive adjustment allows the proposed ASD-Net network to obtain richer spatial and channel information at the pixel level.

In addition, the ASCO block uses asymmetric convolution (vertical and horizontal) to better integrate spatial and channel information and deeply mine the correlation between features. For this, ASCO integrates complex asymmetric convolution operations and attention gates into a standard convolution operation. This design helps reduce computational complexity in practical applications and improve network runtime efficiency.

The design of the ASCO block is well-grounded theoretically. Existing research has shown that for parallel processing of space and channels, both spatial and channel information play significant roles in image processing tasks [46]. Moreover, asymmetric convolution can better extract and integrate spatial and channel information, which is also supported by related research [47].

The effectiveness of the ASCO block is validated through ablation study experiments, as described further in Subsection 4.4.4.

For image x_i , the calculation formulae used by ASCO at the first and second layers of asymmetric convolution are the following:

$$Y_1 = BN(x_i * W_{3 \times 3} + b_{3 \times 3} + x_i * W_{3 \times 1} + b_{3 \times 1} + x_i * W_{1 \times 3} + b_{1 \times 3}) \tag{4}$$

$$Y_2 = BN(Y_1 * W_{3 \times 3} + b_{3 \times 3} + Y_1 * W_{3 \times 1} + b_{3 \times 1} + Y_1 * W_{1 \times 3} + b_{1 \times 3}) \tag{5}$$

where BN denotes batch normalization operation, $W_{i \times j}$ denotes convolution operation of $i \times j$, and $b_{i \times j}$ denotes the bias of $i \times j$.

Then, following the scSE operation, the final output Y is obtained as follows:

$$Y_{cSE} = Y_2 \times Sigmoid[W_1 * (W_2 * Avg(Y_2))] \tag{6}$$

$$Y_{sSE} = Y_2 \times Sigmoid(Conv_{1 \times 1}(Y_2)) \tag{7}$$

$$Y = Y_{cSE} + Y_{sSE} \tag{8}$$

where Y_{cSE} represents the output obtained through the Spatial Squeeze and Channel Excitation (cSE) part of the concurrent spatial and channel squeeze & excitation (scSE) mechanism. Y_{sSE} represents the output obtained through the Channel Squeeze and Channel Excitation (sSE) part of the concurrent scSE mechanism, and W_1 and W_2 being the weights of two fully connected layers.

3.3 DDEC block

The proposed DDEC block enhances traditional dense connections by replacing the original 3×3 standard convolution

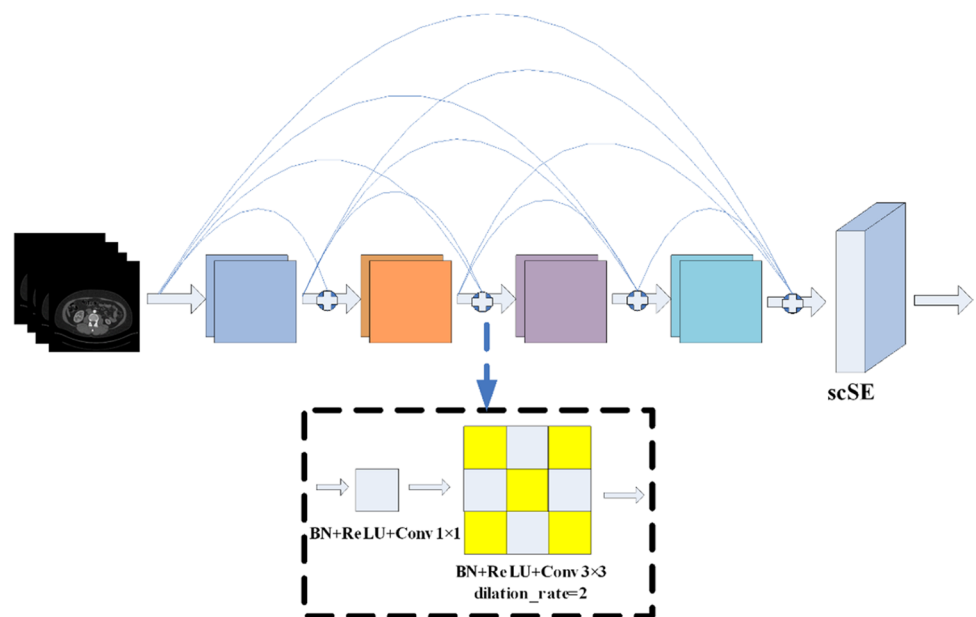
with dilated convolution and incorporating an scSE attention gate (Fig. 8). This seemingly simple replacement led to an unexpected result, having a significant impact on the network segmentation performance.

Firstly, the DDEC block can effectively expand the receptive field of the network by using dilated convolution, without a need of additional parameters or extra computational complexity. This is crucial for capturing and understanding substantial contextual information in images, particularly for the target task of kidney and kidney tumor segmentation. A large receptive field can help the network acquire a broader context, thus achieving a better balance between details and global information. At the same time, the addition of the scSE attention gate allows the network to perform “squeeze-and-excitation” operations in both the spatial and channel dimensions. This allows the network to emphasize not only on the important channels but also to highlight key areas in the input feature map.

Secondly, compared to traditional dense connections, the proposed DDEC block retains an advantage of enhancing feature propagation and reuse. This allows each convolution layer to receive feature maps from all previous layers, conducive to acquiring richer feature information and ensuring better gradient backpropagation, thus improving the network training stability.

Finally, the incorporation of the DDEC block makes the network more effective in acquiring multi-scale features. Dilated convolution can obtain features of multiple scales while keeping the number of parameters unchanged, thus allowing the network to handle kidneys and kidney tumors with large-scale changes more effectively.

Fig. 8 The proposed DDEC block



3.4 Dual-encoder/single-decoder backbone

The inspiration about this improvement comes from the N-Net network [38], where two parallel paths of the dual encoder are interconnected layer-by-layer, through standard skip connections. As some information might be lost during the encoding and decoding, N-Net introduces a dual encoder network in order to reduce such losses. However, while applying this strategy, we introduced another innovation related to the removal of the top skip connection between the original encoding and decoding layers. Such dual encoder network not only deepens the network depth but also integrates more comprehensive information [38].

In the traditional U-Net architecture, features at all levels are treated equivalently and passed and fused through skip connections. However, we observed that the top-level feature information mainly contains global, low-level semantic information, which contributes little to the details of the segmentation task. Therefore, we chose to remove the top skip connection between the encoding and decoding layers, thereby putting more attention and resources on higher-level, more distinctive, and detailed features.

In the proposed network, the first encoder utilizes maximum average pooling, two 3×3 convolution layers, and the Rectified Linear Unit (ReLU) activation function to extract features. In contrast, the second encoder, composed of an ASCO block, a DDEC block, and an scSE attention mechanism [45], is responsible for extracting and reconstructing more complex features. These two encoders are interconnected through skip connections, facilitating the full utilization and integration of features at different levels and scales. This fusion method helps improve the network's expressiveness and prediction accuracy and also makes the network more robust when dealing with complex tasks.

This meticulous design enables the proposed ASD-Net network to capture more fine-grained details and more intense semantic information, thus achieving a significant performance improvement in image segmentation tasks.

4 Experiments and results

4.1 Datasets

To evaluate segmentation performance of the proposed ASD-Net network in comparison to other existing networks, experiments were conducted on a dataset, sourced from the 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19), which is a diverse dataset in terms of the voxel dimensions, contrast timing, table signature, and scanner field of view [48]. As stated in its original paper [49], this dataset was reviewed and approved by the Institutional Review Board at the University of Minnesota as Study 1611M00821. Additionally, the KiTS19

dataset is made available under the CC BY-NC-SA (Creative Commons Attribution-NonCommercial-ShareAlike) license, as of its publication date. We diligently adhered to the terms of this license throughout our research process to ensure compliance.

The KiTS19 dataset includes both original abdominal CT images and label images manually annotated by doctors. The dataset incorporates a range of different cases from 210 patients, thereby increasing its complexity and diversity, which in turn provides a more challenging environment for network training. In the experiments, the images of the first 170 patients were selected to form the training set, the images of the next 20 patients formed the validation set, and the images of the last 20 patients comprised the test set. All utilized CT scans contained kidney or tumor images, with a resolution of 512×512 pixels.

Due to variations in the number of 2D images extracted per patient, we selectively chose for analysis the images containing either kidneys or tumors. Consequently, in the tumor segmentation experiments, we utilized 4857 images for network training, 305 images for validation, and 294 images for testing. In the case of kidney segmentation experiments, we employed 13840 images for network training, 766 images for validation, and 842 images for testing. This selection ensured utilization of only those images, which are relevant to the specific tasks under consideration, while accounting for the uneven distribution of data contributed by each patient. Such a strategy helped maintain consistency in network training and evaluation, enhancing the interpretability and comparability of experimental results.

Although the KiTS19 dataset was the only one used in the experiments, its richness and diversity of patient cases ensured that the elaborated network has a good generalization ability. Firstly, this dataset covers a wide variety of cases with different disease characteristics and stages, thus training the network to handle various situations. Secondly, a strict training/validation/testing division strategy was adopted to avoid overfitting and conduct evaluations on unseen test sets, ensuring an accurate estimate of the network's generalization ability. Finally, the superior segmentation performance demonstrated by the proposed ASD-Net network further substantiates its strong generalization ability when handling unseen, real medical image data.

4.2 Evaluation metrics

In the experiments, four evaluation metrics, including Intersection over Union (IoU), DSC, recall, and precision, were used to quantitatively evaluate the performance of the proposed ASD-Net network, compared to other networks.

IoU is a widely used evaluation metric for image segmentation that measures the degree of overlap between the detected segmentation mask and the ground truth mask, calculated as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{9}$$

where TP (true positives) represents the number of correctly identified pixels as being part of an object (i.e., a kidney/tumor in our case), FN (false negatives) represents the number of incorrectly identified pixels as being not part of an object, and FP (false positives) represents the number of incorrectly identified pixels as being part of an object.

DSC is another widely used evaluation metric for image segmentation, which describes the degree of similarity between the detected segmentation and its corresponding ground truth, calculated as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{10}$$

Recall refers to the proportion of the boundary pixels in the ground truth that are successfully detected by a network, calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

Precision refers to the proportion of the boundary pixels in the segmentation corresponding to the boundary pixels in the ground truth, calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

4.3 Experimental setup

The hardware configuration used in the experiments included an InterCore i5-12490 processor with a main frequency of 3.0 GHz and a single NVIDIA RTX3060 graphics card with 12 GB of memory. To ensure normal training, the following settings were used: a batch size set to 4, the number of epochs set to 100, validation was performed on every epoch, the adaptive moment estimation (Adam) optimizer was used to train the network, the initial learning rate was set to 1×10^{-4} , the decay coefficient was set to 1×10^{-4} to prevent overfitting, momentum was set to 0.9, the minimum learning rate was 1×10^{-5} , and the network structure was implemented by Pytorch.

4.4 Results and analysis

4.4.1 Kidney segmentation comparison with classic networks

First, the kidney segmentation performance of the proposed network was compared to that of classic segmentation networks, such as U-Net, Attention-UNet, U-Net++,

ResNet18, TransUNet, and scSEU-Net. The obtained results are presented in Table 1 (the best result on each metric is shown in bold). Based on these, it is clear that the proposed ASD-Net network outperformed all other networks according to all evaluation metrics. More specifically, the second-best performing network (scSEU-Net) was surpassed by 1.44 percentage points on IoU, 0.84 percentage points on DSC, and 0.75 percentage points on recall, and by 0.14 percentage points on precision, where U-Net took the second place.

While the quantitative comparison highlights the performance improvement achieved by the proposed network, it may not fully convey its advantages. Thus, in Fig. 9, a visual comparison is provided of segmentation results achieved by different networks in segmenting kidneys on the KiTS19 dataset.

4.4.2 Kidney tumor segmentation comparison with classic networks

Then, the proposed network was compared in terms of the kidney tumor segmentation with the classic segmentation networks, participating in the previous experiment. The obtained results are shown in Table 2 (the best result on each metric is shown in bold). Again, the proposed ASD-Net network outperformed all other networks according to all evaluation metrics, except for recall, where it took the second place after Attention-UNet. More specifically, the second-best performing network (scSEU-Net) was surpassed by 4.98 percentage points on IoU, 4.20 percentage points on DSC, and 3.65 percentage points on precision, which demonstrates the higher superiority of ASD-Net in terms of the kidney tumor segmentation. While Attention-UNet outperforms the proposed network in terms of recall, ASD-Net excels it by far on commonly used segmentation metrics such as IoU, DSC, and precision, highlighting its superiority in specific tasks. Higher values of IoU and DSC achieved by ASD-Net indicate that it is more accurate at the pixel level, while high precision demonstrates its ability in predicting positive instances. This balanced performance

Table 1 Kidney segmentation performance comparison of ASD-Net with classic networks

Networks	IoU (%)	DSC (%)	Recall (%)	Precision (%)
U-Net	90.08	94.58	92.74	96.90
Attention-UNet	91.45	95.35	94.24	96.85
U-Net++	90.75	94.96	93.87	96.46
ResNet18	91.02	95.03	94.25	96.35
TransUNet	91.04	95.11	94.32	96.32
scSEU-Net	91.91	95.60	95.32	96.20
ASD-Net (<i>proposed</i>)	93.35	96.44	96.07	97.04

The best results are shown in bold

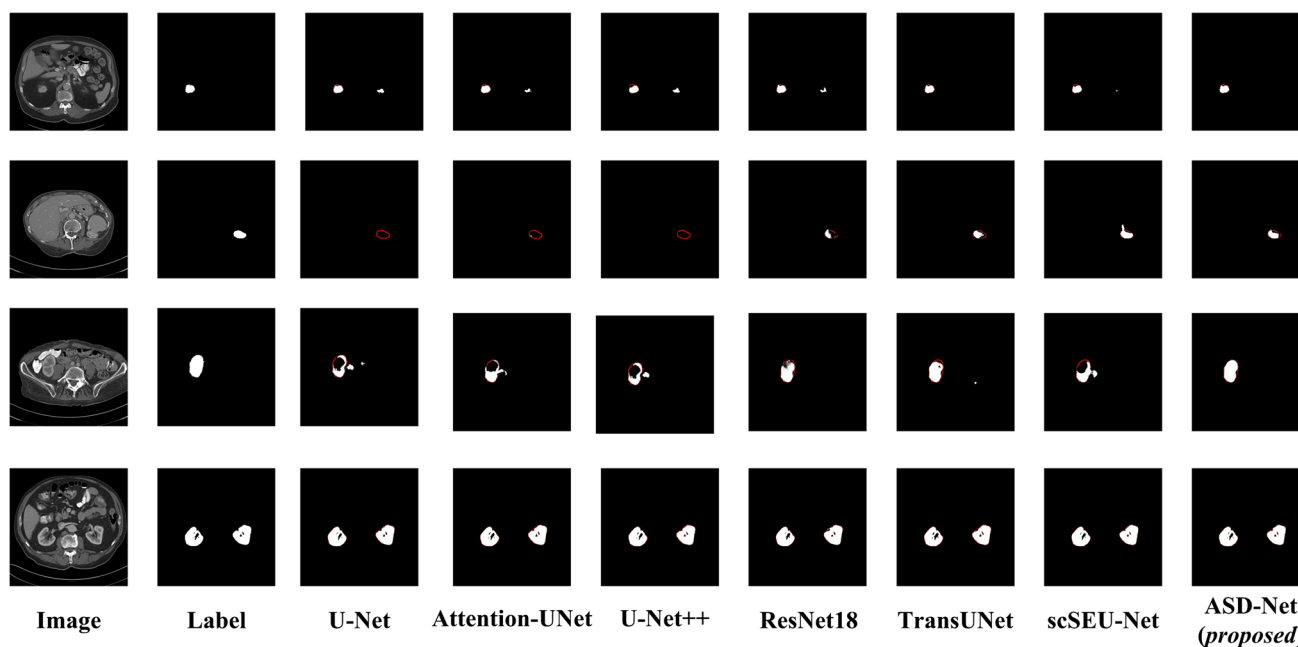


Fig. 9 Sample kidney segmentations performed by the compared networks (the red contour lines highlight the regions that should be successfully segmented)

makes ASD-Net more competitive in overall segmentation tasks. Attention-UNet may be more suitable for scenarios emphasizing high recall, such as initial screening or rapid detection in medical imaging, where a preliminary rough segmentation can be efficiently accomplished by a high-recall network. However, in the final detailed segmentation, the proposed ASD-Net network remains more competitive. Therefore, it is capable of solving (to some extent) the problem of separating small tumor areas from kidneys and can compensate for the problems of missed detection due to incorrect segmentation of the lesion location performed by other networks, thus playing an auxiliary guiding role in the clinical diagnosis of kidney tumors.

Figure 10 shows a visual comparison of results achieved by different networks in segmenting kidney tumors on the

KiTS19 dataset (the white area represents the tumor segmentation result). As can be seen from Fig. 10, traditional classic networks such as U-Net and U-Net++ have rough segmentation of the edges of complex-shaped, small-volume tumor targets, and there are cases of erroneous segmentation. In contrast, the proposed ASD-Net network performs well in segmenting small and irregular lesion areas.

4.4.3 Kidney and kidney tumor segmentation comparison with state-of-the-art networks

Next, the proposed network was compared in terms of both kidney segmentation and kidney tumor segmentation with state-of-the-art networks, based on their DSC results reported in the corresponding literature sources, as shown in Table 3 (the best DSC result achieved in each task is shown in bold). The proposed ASD-Net network outperforms all state-of-the-art networks by scoring 2.82 and 0.05 percentage points more than the second-best performing network in kidney segmentation and kidney tumor segmentation, respectively, according to DSC.

4.4.4 Ablation study

To verify the effectiveness of each module added to the baseline (U-Net) when designing the proposed ASD-Net network, ablation study experiments were conducted separately for kidney segmentation and kidney tumor segmentation. In these experiments, we incrementally added first the

Table 2 Kidney tumor segmentation performance comparison of ASD-Net with classic networks

Networks	IoU (%)	DSC (%)	Recall (%)	Precision (%)
U-Net	64.49	76.93	81.80	76.59
Attention-UNet	69.20	80.29	91.85	74.21
U-Net++	65.94	77.95	85.26	74.83
ResNet18	67.45	79.36	84.15	78.10
TransUNet	64.86	77.01	82.44	76.53
scSEU-Net	70.48	81.02	85.57	80.54
ASD-Net (<i>proposed</i>)	75.46	85.22	88.55	84.19

The best results are shown in bold

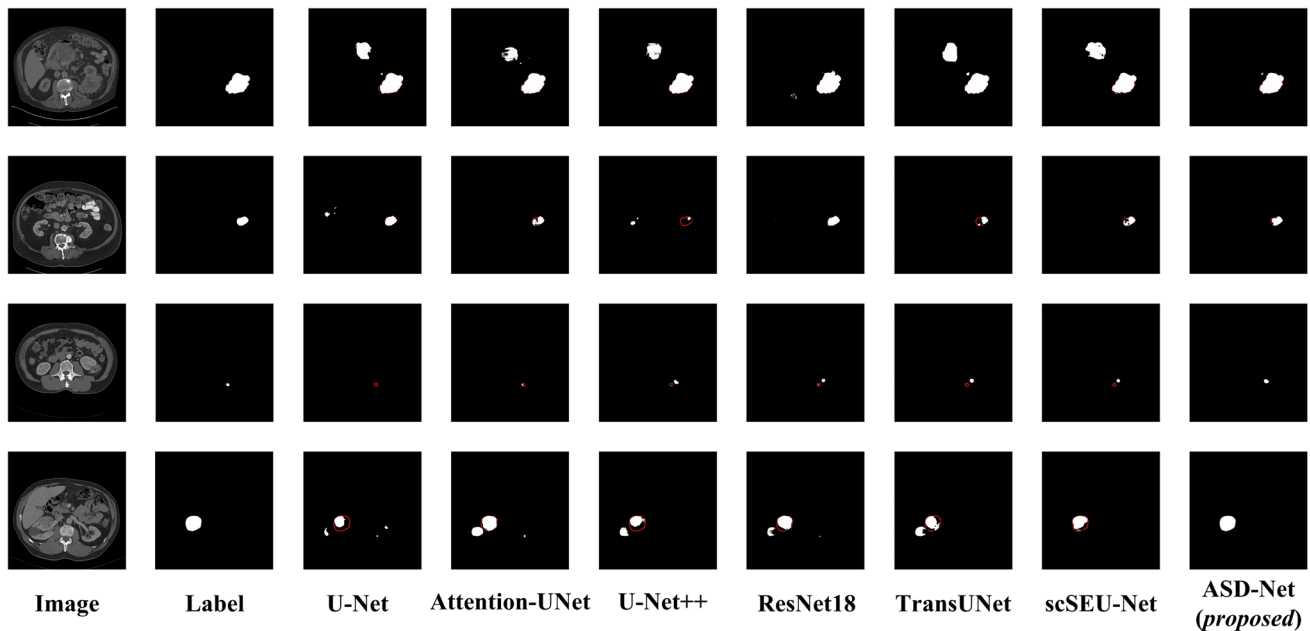


Fig. 10 Sample kidney tumor segmentations performed by the compared networks (the red contour lines highlight the regions that should be successfully segmented)

Table 3 Kidney and kidney tumor segmentation performance comparison of ASD-Net with state-of-the-art networks

Networks	DSC (%) for kidney tumor segmentation	DSC (%) for kidney segmentation
RAU-Net [50]	76.75	96.19
LC-Unet [51]	78.91	96.39
MSVRL+nnU-Net [52]	74.50	95.98
O2M-UDA [53]	75.40	92.30
SLM-SA [54]	82.40	-
MedNeXt [55]	-	91.02
ASD-Net (<i>proposed</i>)	85.22	96.44

The best results are shown in bold

ASCO block and then the DDEC block, then removed the original top-level skip connection, added an ASPP module, and finally added a dual encoder to form the final network, all under the same experimental environment. The results obtained in the kidney segmentation are shown in Table 4, whereas Table 5 presents the obtained results in the kidney tumor segmentation (the best result on each metric is shown in bold). As can be seen from these tables, in the first step, the addition of ASCO block to U-Net led to improving the values of all evaluation metrics, except for precision. The additional incorporation of the DDEC block in the second step further improved the values of all evaluation metrics, except for recall in case of kidney tumor segmentation. The removal of the original top-level skip connection

Table 4 Ablation study results for kidney segmentation

Networks	IoU (%)	DSC (%)	Recall (%)	Precision (%)
U-Net	90.08	94.58	92.74	96.90
+ ASCO	91.18	95.20	94.28	96.50
+ DDEC	91.71	95.50	94.60	96.80
-cat	91.87	95.57	94.95	96.58
+ ASPP	92.61	96.00	95.20	97.11
ASD-Net (<i>proposed</i>)	93.55	96.44	96.07	97.04

The best results are shown in bold

Table 5 Ablation study results for kidney tumor segmentation

Networks	IoU (%)	DSC (%)	Recall (%)	Precision (%)
U-Net	64.49	76.93	81.80	76.59
+ ASCO	69.29	80.35	88.19	76.52
+ DDEC	72.70	82.26	88.15	79.54
-cat	73.99	83.49	87.66	82.12
+ ASPP	74.80	84.37	86.11	85.26
ASD-Net (<i>proposed</i>)	75.46	85.22	88.55	84.19

The best results are shown in bold

was beneficial for all metrics, except for precision in kidney segmentation and recall in kidney tumor segmentation. The addition of the ASPP module improved all evaluation metrics, except for recall in kidney tumor segmentation. The last

step of adding the dual encoder, by which the final network was formed, allowed to improve all metrics, except for precision (according to this metric, the result of the previous step was the best one among all others). By using majority voting (3 out of 4 evaluation metrics reported the best results), it was decided to promote the result of the last step as a novel ASD-Net network, which is the subject of this paper.

5 Conclusions and future directions

To address the issue of inaccurate kidney and kidney tumor segmentation, this paper has proposed improvements to the U-Net network structure, leading to a more advanced network, named ASD-Net. Firstly, an encoder of context-aware features was designed, by combining efficient attention channels and the newly designed ASCO and DDEC blocks, and depth and spatial information, to extract multi-scale feature images. Then, based on the U-Net structure, additional encoding layers were added and connected by skip connections in order to optimize feature extraction and fusion, thus further enhancing detail restoration capabilities of the proposed network. Subsequently, a combined BCE-Dice loss was utilized to mitigate the issue of unbalanced positive and negative samples and to enhance the accuracy of boundary segmentation. The experimental results demonstrated that even the hard-to-segment areas of kidneys and kidney tumors can be completely delineated, exhibiting smooth boundary contours.

While ASD-Net has made significant strides in addressing kidney and kidney tumor segmentation, it is worth also noting its limitations, which provide valuable directions for future improvements.

Firstly, ASD-Net may face challenges in handling small regions, particularly when these are covered by overlapping structures, leading to less accurate segmentation results. Additionally, despite the network's outstanding performance in segmentation, its processing speed may not be ideal, limiting its applicability in real-time applications or large-scale dataset processing. In addition to the mentioned limitations, another drawback of ASD-Net is its inability to directly process 3D graphics. The current workflow requires the conversion of 3D to 2D graphics before inputting them into the network for segmentation, potentially causing information loss and introducing complexity and inaccuracies when dealing with 3D medical images.

Future research directions will focus on addressing these limitations while further enhancing the performance of ASD-Net. Firstly, efforts will be directed towards optimizing segmentation time to meet the demands of real-time applications and large-scale dataset processing. Secondly, there will be a concentrated effort to improve ASD-Net's ability to handle segmentation of small areas, potentially through the introduction of more intelligent context information capture or specifically designed structures. Simultaneously, future

plans include also enhancing the robustness of the network, reducing dependency on parameter selection, and increasing its universality. Future developments may also involve expanding ASD-Net's support for multi-modal images and exploring the integration of other advanced technologies, such as transfer learning, reinforcement learning, or self-supervised learning, to further improve the network's performance and applicability.

Funding Open Access funding provided by the IReL Consortium. This paper has emanated from research conducted with the financial support of the National Key Research and Development Program of China under the Grant No. 2017YFE0135700, the Bulgarian National Science Fund (BNSF) under the Grant No. КП-06-ИП-КИТАЙ/1 (КР-06-ИР-CHINA/1), and the Telecommunications Research Centre (TRC) of University of Limerick, Ireland.

Declarations

Ethics approval This paper does not contain any experiments with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Checcucci E, De Cillis S, Granato S, Chang P, Afyouni AS, Okhunov ZJCOIU (2020) Applications of neural networks in urology: a systematic review. *Curr Opin Urol* 30:788–807
2. Sun J, Zhang H, Yan Y, Xu S, Fan X (2021) Optimal regulation strategy for nonzero-sum games of the immune system using adaptive dynamic programming. *IEEE Trans Cybern* 53(3):1475–1484. <https://doi.org/10.1109/TCYB.2021.3103820>
3. Farjana A, Liza FT, Pandit PP, Das MC, Hasan M, Tabassum F, Hossen MH (2023) Predicting chronic kidney disease using machine learning algorithms. In: 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, pp 1267–1271. <https://doi.org/10.1109/CCWC57344.2023.10099221>
4. Campbell S, Uzzo Robert G, Allaf Mohamad E, Bass Eric B, Cadeddu Jeffrey A, Chang A, Clark Peter E, Davis Brian J, Derweesh Ithaar H, Giambaresi L et al (2017) Renal mass and localized renal cancer: AUA Guideline. *J Urol* 198:520–529. <https://doi.org/10.1016/j.juro.2017.04.100>

5. Gillies RJ, Kinahan PE, Hricak H (2015) Radiomics: images are more than pictures they are data. *Radiology* 278:563–577. <https://doi.org/10.1148/radiol.2015151169>
6. Kerdvibulvech C, Chen L (2020) The power of augmented reality and artificial intelligence during the Covid-19 outbreak. In: Stephanidis C, Kurosu M, Degen H, Reinerman-Jones L (eds) *HCI International 2020 - Late breaking papers: multimodality and intelligence*. HCII 2020. Lecture Notes in Computer Science(), vol 12424. Springer, Cham. https://doi.org/10.1007/978-3-030-60117-1_34
7. Jin S, Zhang X, Li X, Cheng M, Cui, X, Liu J (2023) Development and application of teaching model for medical humanities education using artificial intelligence and digital humans technologies. In: 2023 IEEE 6th Eurasian Conference on Educational Innovation (ECEI), Singapore, pp 119–122. <https://doi.org/10.1109/ECEI57668.2023.10105419>
8. Wang Y, Zhou Y, Shen W, Park S, Fishman EK, Yuille AL (2019) Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *Med Image Anal* 55:88–102. <https://doi.org/10.1016/j.media.2019.04.005>
9. Pandey M, Gupta A (2021) A systematic review of the automatic kidney segmentation methods in abdominal images. *Biocybern Biomed Eng* 41:1601–1628. <https://doi.org/10.1016/j.bbe.2021.10.006>
10. Ashok M, Gupta A, Pandey M (2023) HCIU: Hybrid clustered inception-based UNET for the automatic segmentation of organs at risk in thoracic computed tomography images. *Int J Imaging Syst Technol* 33:2203–2217
11. Liu J, Cao L, Akin O, Tian Y (2019) 3DFPN-HS²: 3D feature pyramid network based high sensitivity and specificity pulmonary nodule detection. In: Shen D et al (eds) *Medical image computing and computer assisted intervention – MICCAI 2019*. MICCAI 2019. Lecture Notes in Computer Science(), vol 11769. Springer, Cham. https://doi.org/10.1007/978-3-030-32226-7_57
12. Rasmussen R, Sanford T, Parwani AV, Pedrosa I (2022) Artificial intelligence in kidney cancer. *Am Soc Clin Oncol Educ Book* 42(42). https://doi.org/10.1200/EDBK_350862
13. Soomro TA, Zheng L, Afifi AJ, Ali A, Yin M, Gao J (2022) Artificial intelligence (AI) for medical imaging to combat coronavirus disease (COVID-19): a detailed review with direction for future research. *Artif Intell Rev* 55:1409–1439. <https://doi.org/10.1007/s10462-021-09985-z>
14. Said D, Hectors SJ, Wilck E, Rosen A, Stocker D, Bane O, Beksaç AT, Lewis S, Badani K, Taouli B (2020) Characterization of solid renal neoplasms using MRI-based quantitative radiomics features. *Abdominal Radiol* 45:2840–2850. <https://doi.org/10.1007/s00261-020-02540-4>
15. Xi IL, Zhao Y, Wang R, Chang M, Purkayastha S, Chang K, Huang RY, Silva AC, Vallières M, Habibollahi P et al (2020) Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. *Clin Cancer Res* 26:1944–1952. <https://doi.org/10.1158/1078-0432.CCR-19-0374>
16. Nassiri N, Maas M, Cacciamani G, Varghese B, Hwang D, Lei X, Aron M, Desai M, Oberai AA, Cen SY et al (2022) A radiomic-based machine learning algorithm to reliably differentiate benign renal masses from renal cell carcinoma. *Eur Urol Focus* 8:988–994. <https://doi.org/10.1016/j.euf.2021.09.004>
17. Liu J, Yildirim O, Akin O, Tian Y (2023) AI-driven robust kidney and renal mass segmentation and classification on 3D CT images. *Bioengineering* 10. <https://doi.org/10.3390/bioengineering10010116>
18. Conze PH, Andrade-Miranda G, Singh VK, Jaouen V, Visviki D (2023) Current and emerging trends in medical image segmentation with deep learning. *IEEE Trans Radiat Plasma Med Sci* 7:545–569. <https://doi.org/10.1109/TRPMS.2023.3265863>
19. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y et al (2023) A survey on vision transformer. *IEEE Trans Pattern Anal Mach Intell* 45:87–110. <https://doi.org/10.1109/TPAMI.2022.3152247>
20. Liu Y, Sun Y, Xue B, Zhang M, Yen GG, Tan KC (2023) A survey on evolutionary neural architecture search. *IEEE Trans Neural Networks Learn Syst* 34:550–570. <https://doi.org/10.1109/TNNLS.2021.3100554>
21. Huang X, Chen J, Chen M, Chen L, Wan Y (2022) TDD-UNet: transformer with double decoder UNet for COVID-19 lesions segmentation. *Comput Biol Med* 151:106306
22. Kerdvibulvech C, Dong ZY (2021) Roles of artificial intelligence and extended reality development in the post-COVID-19 era. In: Stephanidis C et al (eds) *HCI International 2021 - Late breaking papers: multimodality, extended reality, and artificial intelligence*. HCII 2021. Lecture Notes in Computer Science(), vol 13095. Springer, Cham. https://doi.org/10.1007/978-3-030-90963-5_34
23. Huang Z, Wang X, Huang L, Huang C, Wei Y, Liu W (2019) Ccnet: criss-cross attention for semantic segmentation. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, pp 603–612. <https://doi.org/10.1109/iccv.2019.00069>
24. Zhuang Y, Jiang N, Xu Y (2022) Progressive distributed and parallel similarity retrieval of large CT Image sequences in mobile telemedicine networks. *Wirel Commun Mob Comput* 2022:6458350. <https://doi.org/10.1155/2022/6458350>
25. Zhuang Y, Chen S, Jiang N, Hu H (2022) An effective WSENet-based similarity retrieval method of large lung CT image databases. *KSII Trans Internet Inf Syst* 16(7):2359–2376. <https://doi.org/10.3837/tiis.2022.07.013>
26. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) *Medical image computing and computer-assisted intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
27. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp 3431–3440. <https://doi.org/10.1109/cvpr.2015.7298965>
28. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov D et al (eds) *Deep learning in medical image analysis and multimodal learning for clinical decision support*. DLMIA ML-CDS 2018. Lecture Notes in Computer Science(), vol 11045. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1
29. Yuan L, Song J, Fan Y (2023) FM-Unet: Biomedical image segmentation based on feedback mechanism Unet. *Math Biosci Eng* 20:12039–12055
30. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B (2018) Attention U-Net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*
31. Diakogiannis FI, Waldner F, Caccetta P, Wu C (2020) ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J Photogramm Remote Sens* 162:94–114
32. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds) *Medical image computing and computer-assisted intervention – MICCAI 2016*. MICCAI 2016. Lecture Notes in Computer Science(), vol 9901. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49
33. Pandey M, Gupta A (2023) Tumorous kidney segmentation in abdominal CT images using active contour and 3D-UNet. *Irish*

- J Med Sci (1971 -) 192:1401–1409. <https://doi.org/10.1007/s11845-022-03113-8>
34. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 40:834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>
 35. Liang B, Tang C, Zhang W, Xu M, Wu T (2023) N-Net: an UNet architecture with dual encoder for medical image segmentation. *SIViP* 17:3073–3081. <https://doi.org/10.1007/s11760-023-02528-9>
 36. Li J, Liu K, Hu Y, Zhang H, Heidari AA, Chen H, Zhang W, Algarni AD, Elmannai H (2023) Eres-UNet++: Liver CT image segmentation based on high-efficiency channel attention and Res-UNet++. *Comput Biol Med* 158:106501. <https://doi.org/10.1016/j.compbiomed.2022.106501>
 37. Sengupta A, Ye Y, Wang R, Liu C, Roy K (2019) Going deeper in spiking neural networks: VGG and residual architectures. *Front Neurosci* 13. <https://doi.org/10.3389/fnins.2019.00095>
 38. Wu X, Hong D, Chanussot J (2023) UIU-Net: U-Net in U-Net for infrared small object detection. *IEEE Trans Image Process* 32:364–376. <https://doi.org/10.1109/TIP.2022.3228497>
 39. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39:2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
 40. Dong L, Liu H (2021) Segmentation of pulmonary nodules based on improved UNet++. In: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, pp 1–5. <https://doi.org/10.1109/CISP-BMEI53629.2021.9624438>
 41. Roy AG, Navab N, Wachinger C (2018) Concurrent spatial and channel ‘Squeeze & Excitation’ in fully convolutional networks. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G (eds) Medical image computing and computer assisted intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science(), vol 11070. Springer, Cham. https://doi.org/10.1007/978-3-030-00928-1_48
 42. Xiao X, Dong S, Yu Y, Li Y, Yang G, Qiu Z (2023) MAE-TransRNet: an improved transformer-ConvNet architecture with masked autoencoder for cardiac MRI registration. *Front Med* 10. <https://doi.org/10.3389/fmed.2023.1114571>
 43. Montazerolghaem M, Sun Y, Sasso G, Haworth A (2023) U-Net architecture for prostate segmentation: the impact of loss function on system performance. *Bioengineering* 10. <https://doi.org/10.3390/bioengineering10040412>
 44. Ruby U, Yendapalli V (2020) Binary cross entropy with deep learning technique for image classification. *Int J Adv Trends Comput Sci Eng* 9. <https://doi.org/10.30534/ijatcse/2020/175942020>
 45. Milletari F, Navab N, Ahmadi SA (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, pp 565–571. <https://doi.org/10.1109/3DV.2016.79>
 46. Han X, Zhang T (2022) Spatial steganalysis based on non-local block and multi-channel convolutional networks. *IEEE Access* 10:87241–87253. <https://doi.org/10.1109/ACCESS.2022.3199351>
 47. Haris M, Hou J, Wang X (2023) Lane line detection and departure estimation in a complex environment by using an asymmetric kernel convolution algorithm. *Vis Comput* 39:519–538. <https://doi.org/10.1007/s00371-021-02353-6>
 48. Heller N, Isensee F, Maier-Hein KH, Hou X, Xie C, Li F, Nan Y, Mu G, Lin Z, Han M et al (2021) The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge. *Med Image Anal* 67:101821. <https://doi.org/10.1016/j.media.2020.101821>
 49. Heller N, Sathianathen N, Kalapara A, Walczak E, Moore K, Kaluzniak H, Rosenberg J, Blake P, Rengel Z, Oestreich, M (2019) The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445. <https://doi.org/10.48550/arXiv.1904.00445>
 50. Guo J, Zeng W, Yu S, Xiao J (2021) RAU-Net: U-Net model based on residual and attention for kidney and kidney tumor segmentation. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, pp 353–356. <https://doi.org/10.1109/ICCECE51280.2021.9342530>
 51. Kang L, Zhou Z, Huang J, Han W (2022) Renal tumors segmentation in abdomen CT Images using 3D-CNN and ConvLSTM. *Biomed Signal Process Control* 72:103334. <https://doi.org/10.1016/j.bspc.2021.103334>
 52. Zheng R, Zhong Y, Yan S, Sun H, Shen H, Huang K (2023) MsVRL: self-supervised multiscale visual representation learning via cross-level consistency for medical image segmentation. *IEEE Trans Med Imaging* 42:91–102. <https://doi.org/10.1109/TMI.2022.3204551>
 53. Jiang Z, He Y, Ye S, Shao P, Zhu X, Xu Y, Chen Y, Coatrieux J-L, Li S, Yang G (2023) O2M-UDA: Unsupervised dynamic domain adaptation for one-to-multiple medical image segmentation. *Knowl-Based Syst* 265:110378. <https://doi.org/10.1016/j.knosys.2023.110378>
 54. Wen M, Zhou Q, Tao B, Shcherbakov P, Xu Y, Zhang X (2023) Short-term and long-term memory self-attention network for segmentation of tumours in 3D medical images. *CAAI Transact Intell Technol* n/a. <https://doi.org/10.1049/cit2.12179>
 55. Roy S, Koehler G, Ulrich C, Baumgartner M, Petersen J, Isensee F, Jaeger PF, Maier-Hein K (2023) MedNeXt: Transformer-driven scaling of ConvNets for medical image segmentation. In: Greenspan H et al (eds) Medical image computing and computer assisted intervention – MICCAI 2023. MICCAI 2023. Lecture Notes in Computer Science, vol 14223. Springer, Cham. https://doi.org/10.1007/978-3-031-43901-8_39

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Zhanlin Ji (Member, IEEE) received his MEng and PhD degrees from Dublin City University and University of Limerick in 2006 and 2010, respectively. He is a Professor at the North China University of Science and Technology (China) and an Associated Researcher at the Telecommunications Research Centre (TRC), University of Limerick (Ireland). Prof. Ji's research interests include Ubiquitous Consumer Wireless World (UCWW), Internet of Things (IoT), cloud computing,

big data management, and data mining. He has authored/co-authored 100+ research papers in refereed journals and conferences.



Juncheng Mu was born in 2000, he received B.S degree from the North China University of Science and Technology in 2022. He is currently pursuing the master's degree at the North China University of Science and Technology. His research interests include machine vision and graphic image processing.



Jianuo Liu was born in 2000. She received his B.S. degree from North China University of Science and Technology in 2022. Currently, she is pursuing a master's degree at North China University of Science and Technology. Her research interests include machine vision and graphic image processing.



Haiyang Zhang is currently a lecturer at Xi'an Jiaotong-Liverpool University, China. She received her PhD degree in the Department of Electronic & Computer Engineering, University of Limerick, Ireland, in 2018 and her B.S. degree from the school of Software Engineering, JiLin University, China, in 2013. Dr. Zhang's current research interests include recommender systems, data mining, collaborative filtering, and natural language processing.



Chenxu Dai was born in 2000, she received B.S degree from the North China University of Science and Technology, in 2022. She is currently pursuing the master's degree at the North China University of Science and Technology. Her research interests include machine vision and graphic image processing.



Xueji Zhang is vice president of Shenzhen University and Professor at the School of Biomedical Engineering, China. Professor Zhang's research interests span the disciplines of chemistry, biology, materials, and medicine, with an emphasis on studies of biosensing, biomedicine, and biomaterials. He serves as the co-Editor in Chief of Sensors & Diagnostics and has been an editorial member of 24 international journals. He has received numerous national and international awards and honors including

Member of the Russian Academy of Engineering; Fellow of the American Institute for Medical & Bioengineering; Fellow of the Royal Chemical Society; a National Innovation Award in China; Scientist of the Year in China; and Simon Fellow of ICSC-World Lab.



Ivan Ganchev is an IEEE Senior Member, an ITU-T Invited Expert, and an IET Invited Lecturer, currently affiliated with University of Limerick, Ireland, University of Plovdiv "Paisii Hilendarski" and Institute of Mathematics and Informatics—Bulgarian Academy of Sciences, Bulgaria. He received his doctoral and engineering (*summa cum laude*) degrees from Saint-Petersburg University of Telecommunications in 1995 and 1989, respectively. Prof. Ganchev was involved in more

than 40 international and national research projects. He has served on the TPC of more than 400 prestigious international conferences/symposia/workshops and has authored/co-authored 1 monographic book, 3 textbooks, 4 edited books, and more than 300 research papers in refereed international journals, books, and conference proceedings. Prof. Ganchev is on the editorial board of and has served as a Guest Editor for multiple prestigious international journals.