BMC Medical Imaging

RESEARCH

Open Access

Medical image analysis using improved SAM-Med2D: segmentation and classification perspectives

Jiakang Sun[1,2], Ke Chen[1,2], Zhiyi He[1,2], Siyuan Ren[1,2], Xinyang He[1,2], Xu Liu[1,2] and Cheng Peng[1,2*]

Abstract

Recently emerged SAM-Med2D represents a state-of-the-art advancement in medical image segmentation. Through fine-tuning the Large Visual Model, Segment Anything Model (SAM), on extensive medical datasets, it has achieved impressive results in cross-modal medical image segmentation. However, its reliance on interactive prompts may restrict its applicability under specific conditions. To address this limitation, we introduce SAM-AutoMed, which achieves automatic segmentation of medical images by replacing the original prompt encoder with an improved MobileNet v3 backbone. The performance on multiple datasets surpasses both SAM and SAM-Med2D. Current enhancements on the Large Visual Model SAM lack applications in the field of medical image classification. Therefore, we introduce SAM-MedCls, which combines the encoder of SAM-Med2D with our designed attention modules to construct an end-to-end medical image classification model. It performs well on datasets of various modalities, even achieving state-of-the-art results, indicating its potential to become a universal model for medical image classification.

Keywords SAM-Med2D, Segmentation, Classification and grading, Attention mechanism, Multimodal

Introduction

Segmentation and classification are two foundational tasks in medical image analysis. Segmentation involves extracting regions of interest such as organs, tissues, or lesions from medical images, while classification entails determining disease types or conducting disease grading based on image features such as color, texture, and shape. The accuracy of segmentation and classification is crucial for clinical applications, including disease diagnosis, treatment progress monitoring, and rehabilitation assessment. Traditionally, medical image segmentation or classification relied on manual recognition and annotation by knowledgeable medical professionals, which is time-consuming and labor-intensive. In recent years, machine learning, particularly deep learning, has played a significant role in medical image analysis. Its end-to-end model structure enables the automatic learning of complex image features.However, this approach has limitations, namely its task-specific nature. As illustrated in Fig. 1, unlike natural images, medical images come in multiple modalities, including computed tomography (CT) scans, endoscopic imaging, magnetic resonance imaging (MRI), ultrasound imaging, microscopic imaging and others. The pixel intensities, textures, and color layer characteristics of images from different modalities vary significantly, leading to notable performance discrepancies of the same deep learning model across different modal data. Hence, it becomes imperative to design

*Correspondence:
Cheng Peng
pengcheng@casit.com.cn
[1] Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610213, Sichuan, China
[2] School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 101499, China

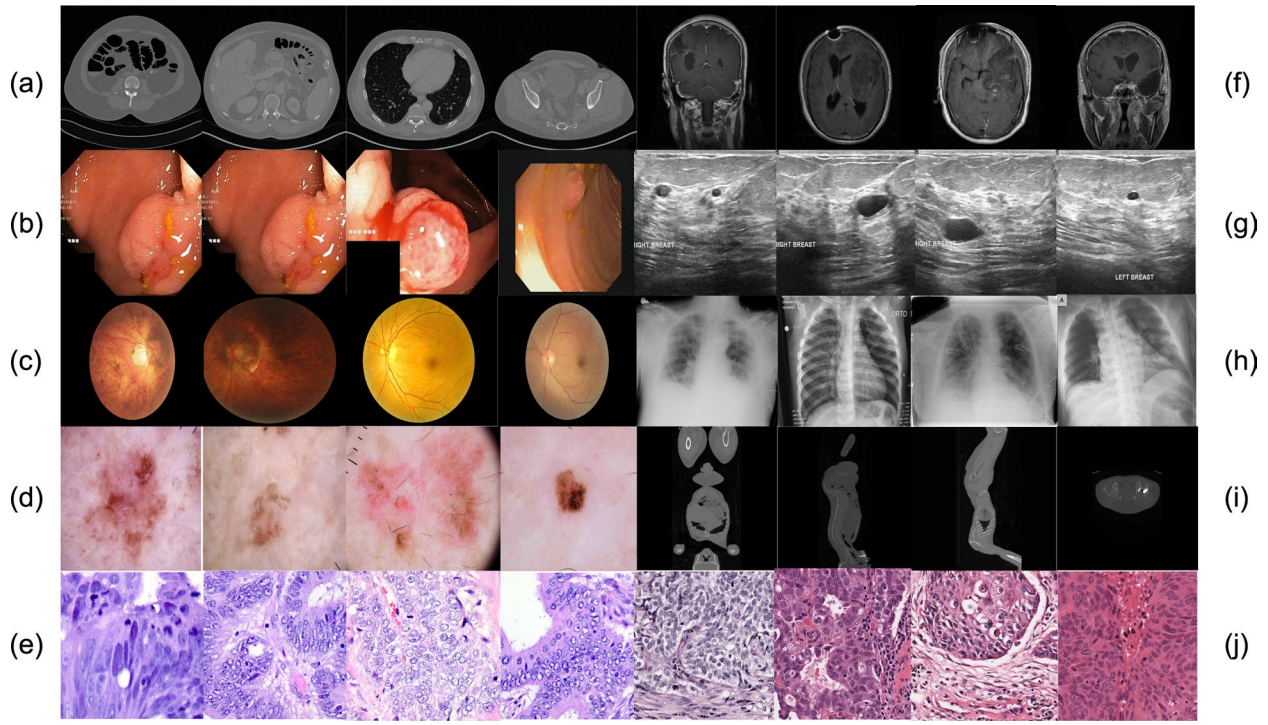

Fig. 1 **a** CT. **b** Endoscopy. **c** Fundus. **d** Dermoscopy. **e** Microscopy. **f** MRI. **g** Ultrasound. **h** X-ray. **i** PET. **j** Histopathology

specific model structures and preprocessing pipelines tailored to individual modalities.

The recent emergence of the Large Visual Model SAM [1] has transformed this landscape. Its zero-shot and few-shot generalization capabilities, honed on extensive image segmentation datasets, have garnered widespread attention. Particularly, the SAM model, fine-tuned on large-scale medical segmentation datasets like MedSAM [2] and SAM-Med2D [3], has established itself as the foundational universal model in medical image segmentation. This advancement enables a single model to adeptly handle images from various modalities. However, SAM and its derivatives necessitate users to provide prompts for segmentation masks through interactive clicks, bounding boxes, etc., which may pose limitations on certain devices and may not deliver optimal performance on specific medical image datasets. To address this, we have enhanced it into an end-to-end automatic segmentation model. Concurrently, we utilize the encoder of SAM-Med2D to establish a medical classification model, thereby expanding SAM's utility in downstream tasks. In summary, this paper contributes the following:

1. We propose SAM-AutoMed, an entirely automated medical image segmentation model based on SAM-Med2D. We have devised a MobileNet v3 architecture, enhancing the SE block, to substitute the interactive prompt encoder. This substitution enables automatic segmentation of medical images, demonstrating outstanding performance across various medical image segmentation datasets.
2. By leveraging the encoder from SAM-Med2D, we introduce an end-to-end model, SAM-MedCls, which incorporates prior knowledge for medical image classification. Figure 2 delineates the architecture of SAM-MedCls. By amalgamating the encoder from SAM-AutoMed with straightforward attention modules, SAM-MedCls augments expression and feature extraction capabilities, surpassing state-of-the-art methods across various medical image classification datasets.

Related work

Segment anything model

Inspired by large language models like ChatGPT, researchers have developed large visual models capable of rapidly adapting to target tasks and exhibiting excellent zero-shot and few-shot generalization capabilities [1, 4–7]. One such model is the Segment Anything Model (SAM), proposed by Meta AI. As illustrated in Fig. 2, SAM consists of three sub-networks: an image encoder based on Vision Transformer (ViT) [8], a prompt encoder, and a transformer-based [9] mask decoder. The image

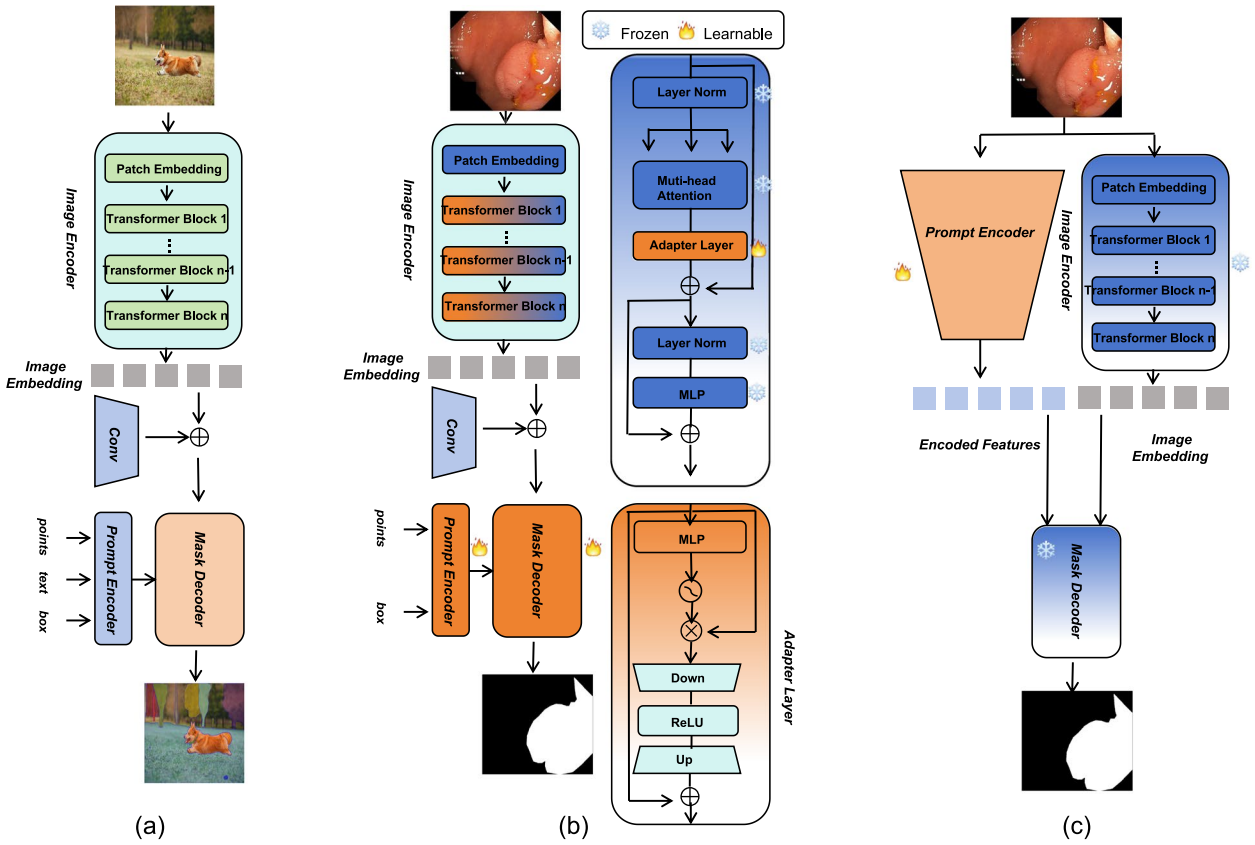

Fig. 2 **a** The model structure of SAM. **b** The model structure of SAM-Med2D. **c** The model structure of AutoSAM

encoder encodes the image into vectors, which are then combined with prompt vectors from the prompt encoder and decoded by the mask decoder to generate the final mask prediction. Prompts can take various forms, including point prompts, prompt boxes, arbitrary masks, and text with positional information. SAM is trained on the SA-1B dataset, which comprises 110 million high-resolution images and over 1 billion high-quality segmentation masks. This training data empowers SAM with robust segmentation capabilities, demonstrating strong versatility in downstream tasks such as instance segmentation [10], object hiding segmentation [11], and video tracking [12]. However, SA-1B lacks images from specialized domains such as medical imaging and remote sensing, leading to suboptimal segmentation performance in corresponding professional fields. Consequently, many researchers have fine-tuned the SAM base model with domain-specific knowledge, achieving outstanding results [2, 3, 10, 13].

Segment medical images based SAM

SAM provides an excellent general framework for interactive image segmentation, but there exists a significant gap between natural images and medical images. This leads to a notable decrease in performance when applying SAM to medical image segmentation. Therefore, current research focuses on fine-tuning SAM with specialized medical image datasets. Ma et al. [2] collected 11 different modality medical datasets with over 1 million masks to fine-tune SAM's mask decoder while retaining the original box prompts. Zu et al. [13] fine-tuned SAM's encoder and mask decoder using adapter layer. As illustrated in Fig. 2, Cheng et al. [3, 14] proposed the SAM-Med2D model, which constructs a large-scale medical image segmentation dataset, SA-Med2D-20M, containing ten modalities (CT, Endoscopy, PET, Fundus, Microscopy, MR, Dermoscopy, X-ray, Ultrasound, Histopathology), over 100 categories, and 4.6 million images and 19.7 million masks ,and comprehensively fine tune the encoder and decoder by adding adapter layer to the encoder.As shown in Fig. 2, Shaharabany et al. [15] introduced AutoSAM, which replaces the prompt encoder. It inputs the image into the encoder simultaneously as a prompt, freezing the encoder and mask decoder during training, thereby achieving automatic segmentation of medical images. Although AutoSAM has achieved automatic segmentation based on SAM, its encoder has not been fine-tuned with medical data, which may reduce

segmentation performance when transferring to the medical domain. We believe the SAM encoder, trained on the SA-1B dataset and fine-tuned on large-scale medical datasets, already possesses visual potential and generalization capabilities. However, there is currently a lack of research on applying SAM to medical classification. Based on these existing issues, we propose a solution in this paper.

Methods

This section presents the methodology of using SAM-AutoMed for medical image segmentation, including details of improved Squeeze and Excitation block [16] and the loss function used for training. Then we introduce the methodology of using SAM-MedCls for medical image classification and propose the attention module.

SAM-AutoMed

SAM-Med2D represents an enhanced iteration of SAM, specifically tailored for medical image segmentation. This enhancement is accomplished through fine-tuning on the extensive medical image segmentation dataset, SA-Med2D-20M. As depicted in Fig. 2, SAM-Med2D is designed with three components: an encoder with adapter layers added for fine-tuning, a prompt encoder supporting sparse prompts such as points and boxes, as well as dense mask prompts, and a mask decoder.Initially, the encoder computes the image embedding E_I of the input image I. Subsequently, the prompt encoder generates the corresponding prompt vector P_I based on the selected prompt interaction type. Finally, the mask decoder predicts the segmentation mask M_I based on the image embedding E_I and the prompt vector P_I. Similar to SAM, SAM-Med2D still requires manual prompts to generate predicted masks. To address this, we propose an end-to-end approach to automate mask generation. As illustrated in Fig. 5, we introduce a MobileNet v3 [17] with an improved Squeeze and Excitation block as the image prompt encoder, replacing the original prompt encoder. The image prompt encoder takes the image I as input and generates prompt encodings as query vectors Q_I for the transformer module of the mask decoder. As shown in Figs. 2 and 3, unlike AutoSAM, during fine-tuning, we only freeze the encoder, while the image encoder and mask decoder are trained. It is worth noting that SAM-AutoMed employs a lightweight convolutional neural network. This is because SAM-AutoMed is fine-tuned for a specific dataset, which can lead to overfitting when the dataset has too few samples. This issue is particularly pronounced since the mask decoder has a certain number of parameters and also needs to be fine-tuned, exacerbating the problem. Unlike AutoSAM, which faces similar issues, we did not use simple convolutional block stacking but opted for MobileNet v3. This choice allows us to fully leverage the pre-trained advantages of MobileNet v3 while maintaining a smaller parameter size, resulting in better performance compared to AutoSAM. Similarly, SAM-MedCls also uses MobileNet v3 as the image prompt encoder to address the overfitting problem when the dataset has too few samples.

Improved squeeze and excitation block

Incorporating the Squeeze and Excitation block into the MobileNet v3 block, as depicted in Fig. 4, involves two main steps: Squeeze and Excitation. In the Squeeze step, a global average pooling operation is performed on the original feature map, yielding a compressed feature vector called the squeezed feature map. In the Excitation step, the squeezed feature map is processed through two 1×1 convolutional layers to first reduce and then increase the dimensionality, resulting in channel-wise weights for each channel of the squeezed feature map. These weights are then multiplied with the original feature map to restore its original size. The motivation behind

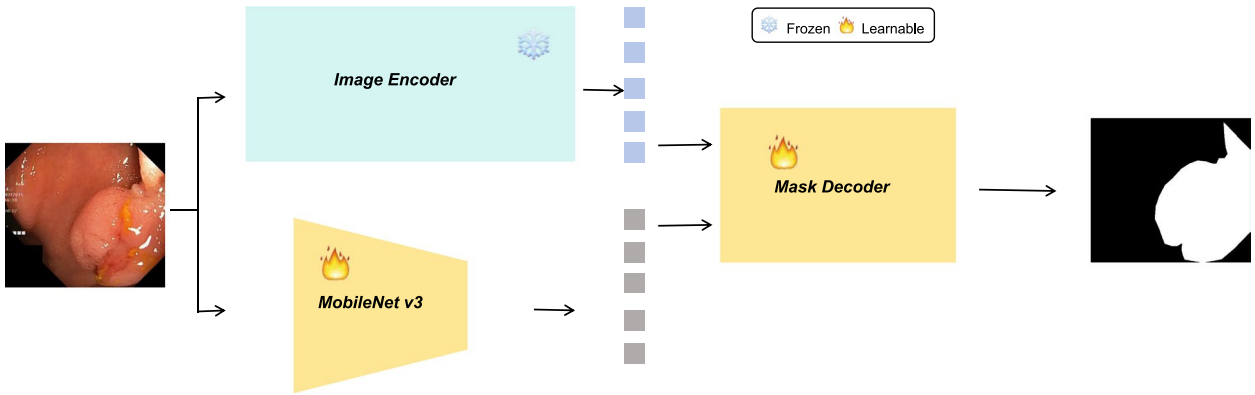

Fig. 3 SAM-AutoMed consists of MobileNet v3 with improved SE block, SAM-Med2D encoder, and SAM-Med2D mask decoder

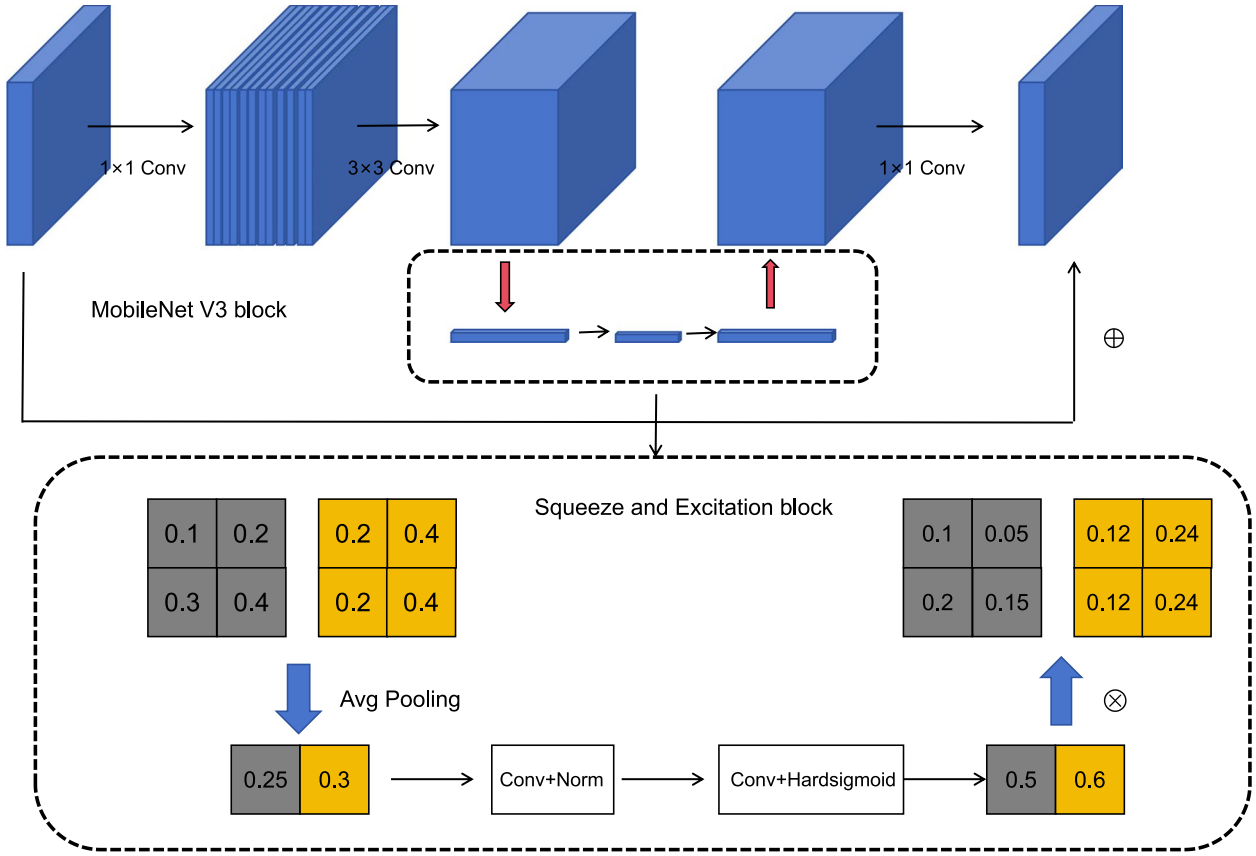

Fig. 4 The structure diagram of MobileNet v3 block. The feature map is first dimensionalized through convolutional layers, then SE blocks, then dimensionality is reduced through 1x1 convolutional layers, and finally added to the original feature map. SE block is essentially a channel attention mechanism

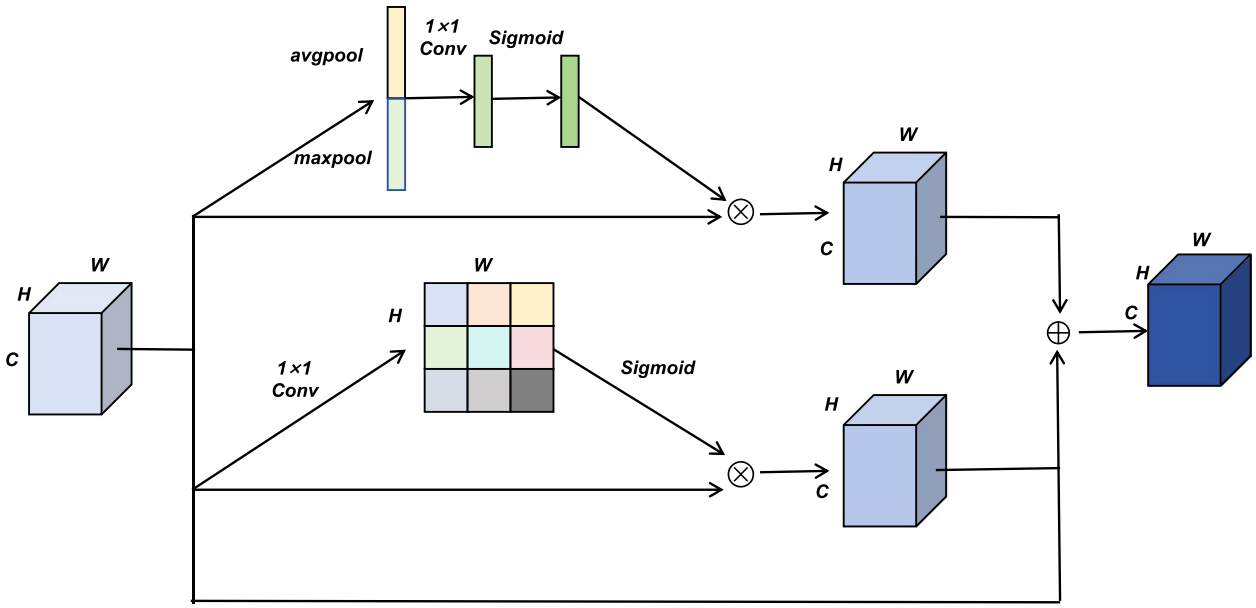

Fig. 5 The improved SE block. adds a parallel spatial attention mechanism and uses residual design to add it to the original feature map before inputting the next module

introducing MobileNet v3, a lightweight convolutional neural network, is to quickly obtain feature vectors containing spatial information from images for input into the mask decoder of SAM-AutoMed and the channel attention module of SAM-MedCls. However, the Squeeze and Excitation block inherently lacks spatial information as it primarily focuses on channel attention mechanisms. To address this limitation, we improve the Squeeze and Excitation block as shown in Fig. 5. We introduce two parallel branches to compute channel attention and spatial attention separately. In the channel attention branch, the feature $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ is processed through global average pooling and global maximum pooling operations to obtain $\mathbf{C_{avg}} \in \mathbb{R}^{C\times 1\times 1}$ and $\mathbf{C_{max}} \in \mathbb{R}^{C\times 1\times 1}$, respectively. These two results are concatenated to obtain $\mathbf{C} \in \mathbb{R}^{2C\times 1\times 1}$, which is then processed through a 1×1 convolutional layer and a Sigmoid activation function to obtain channel weights $\mathbf{C} \in \mathbb{R}^{C\times 1\times 1}$. These weights are then multiplied with the original feature map to restore its original size, resulting in $\mathbf{X_C} \in \mathbb{R}^{C\times H\times W}$. In the spatial attention branch, the feature $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ is processed through a 1×1 convolutional layer and a Sigmoid activation function to obtain spatial weights $\mathbf{S} \in \mathbb{R}^{H\times W}$, which are then multiplied with the original feature map to restore its original size, resulting in $\mathbf{X_S} \in \mathbb{R}^{C\times H\times W}$. Before being passed to the next layer, we add $\mathbf{X}$, $\mathbf{X_C}$, and $\mathbf{X_S}$. Throughout this process, we avoid using fully connected layers, resulting in minimal increase in computational complexity.

Loss function

To address the issue of class imbalance in medical image segmentation, when training SAM-AutoSAM, we utilize two loss functions to compute the loss: Focal Loss and Dice Loss.

$$L_{Seg} = L_{Focal} + L_{Dice} \tag{1}$$

$$L_{Focal} = \begin{cases} \alpha\left(1 - S_{pread}\right)^{y} \log\left(S_{pread}\right), \ y = 1 \\ -(1-a)S_{pread}^{y} \log\left(1 - S_{pread}\right), y = 0 \end{cases} \tag{2}$$

$$L_{Dice} = 1 - \frac{2TP(S_{pred}, S_{gt}) + 1}{2TP(S_{pred}, S_{gt}) + FN(S_{pred}, S_{gt}) + FP(S_{pred}, S_{gt}) + 1} \tag{3}$$

where α and γ are 0.25 and 1, respectively. TP, FN, and FP respectively represent true positives, false negatives, and false positives between the ground truth mask S_{gt} and the output mask S_{pred}.

SAM-MedCls

The structure of SAM-MedCls is illustrated in Fig. 6, consisting of the encoder SAM-Med2D, MobileNet v3, attention modules, and the final MLP layer. Initially, medical images are fed through SAM-Med2D encoder and enhanced MobileNet v3 to obtain features enriched with prior knowledge and fundamental image characteristics (color features, texture features, and crucial spatial information). Subsequently, the features enriched with prior knowledge are passed through the spatial attention module [18] to enhance the representation of important region information. Following this, they are combined with the fundamental features obtained from MobileNet v3 and fed into the channel attention module to enhance semantic features and recognition capability. Finally, classification is performed through the MLP layer. During training, we freeze the encoder SAM-Med2D. Next, we introduce the spatial and channel attention modules.

Spatial attention module

Traditional spatial attention mechanisms require calculating similarity between pixels, which not only is inefficient but also tends to bias towards larger objects due to their higher pixel count. Therefore, we devised

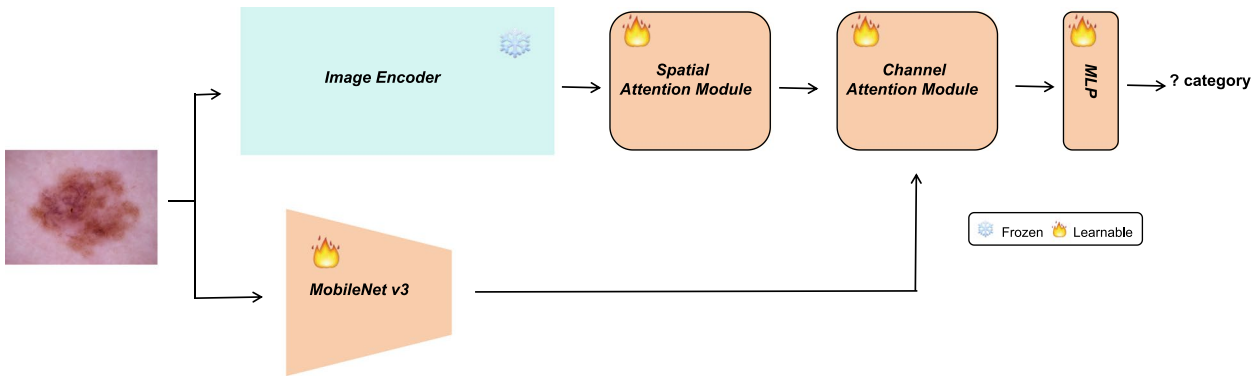

Fig. 6 SAM-MedCls consists of MobileNet v3 with improved SE block, SAM-Med2D encoder, attention module, and finally MLP layer

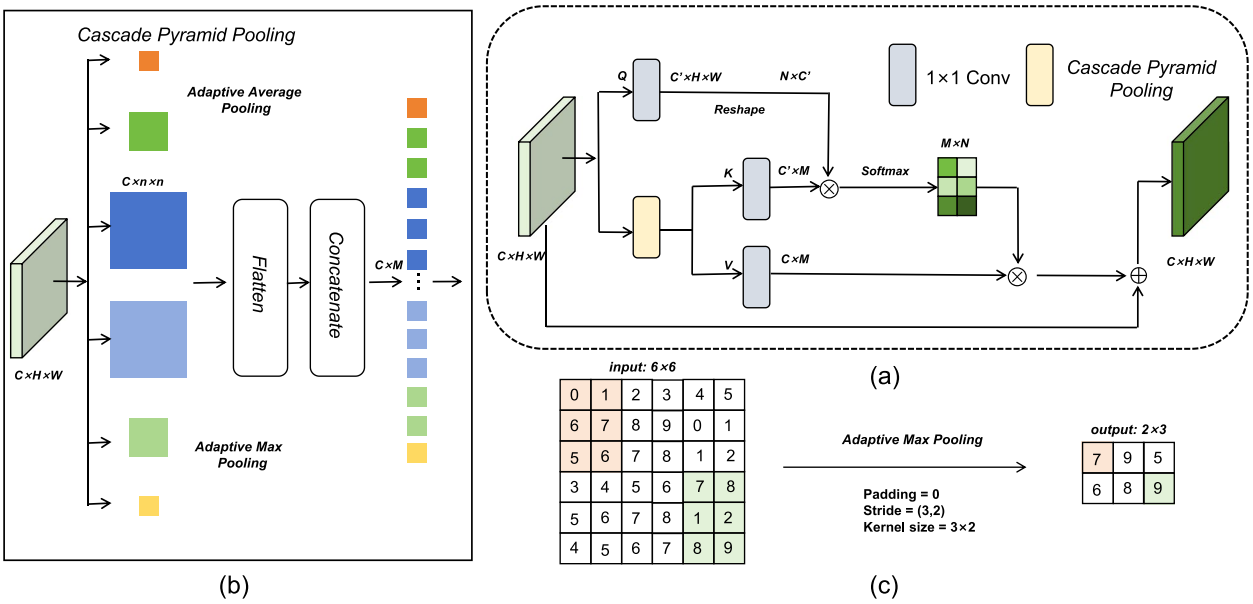

Fig. 7 **a** Spatial attention module. **b** Cascaded pyramid pooling layer. **c** The process of adaptive max pooling:adjusting kernel-size,stripe, and padding based on the size of the output feature map

a cascaded pyramid pooling layer to acquire highly abstracted multiscale context. Then, instead of computing pixel-to-pixel similarity, we calculate similarity between pixels and multiscale context to obtain the spatial attention map. As depicted in Fig. 7, the cascaded pyramid pooling layer takes the given feature $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ and produces feature maps of different sizes through adaptive average and adaptive maximum 2D pooling. These feature maps are unfolded and concatenated to form multiscale context $\mathbf{Z} \in \mathbb{R}^{C\times M}$, where M is the number of contexts. In particular, the output length of the adaptive 2D pooling layer is $n \in [1, 2, 3, 6]$ (the length and width of the output are equal, and the number of channels is equal to the input), that is, each feature map is $C \times n \times n$. Expand and concatenate all feature maps to obtain multi-scale context $\mathbf{Z} \in \mathbb{R}^{C\times M}$, $M = \sum_{n\in[1,2,3,6]} n^2 = 50$. Due to the simultaneous use of adaptive average pooling and adaptive 2D pooling, $M = 100$. Initially, we reduce the dimensionality of the feature X with a 1×1 convolutional layer (original feature dimension: 256, reduced to 128) to obtain $\mathbf{Q} \in \mathbb{R}^{C'\times H\times W}$. Then, Q is reshaped to $N \times C'$, where N is the total number of pixels. Subsequently, the feature X is passed through the cascaded pyramid pooling layer to obtain multiscale context Z, which is further processed by 1×1 convolutional layers to derive $\mathbf{K} \in \mathbb{R}^{C'\times M}$ and $\mathbf{V} \in \mathbb{R}^{C\times M}$. Next, we perform matrix multiplication between Q and K, followed by a softmax layer to compute the similarity $\theta \in \mathbb{R}^{M\times N}$ between pixels and context. Finally, matrix multiplication between θ and V yields semantic features X_S calibrated by spatial attention. The spatial attention module adopts a residual design, where the output is added to the original feature map before being sent out. The overall formula can be expressed as (4):

$$\mathbf{X}_{Si} = \sum_{j=1}^{M} f\left(\mathbf{x}_i, \mathbf{z}_j\right) \cdot \mathbf{z}_j + \mathbf{x}_i \tag{4}$$

where f represents the impact of multi-scale context on pixels, i ranges in [1, ⋯, H × W] represents the number of pixels, and M represents the number of multi-scale contexts.

Channel attention module

Each channel map can be viewed as a representation of some abstract feature, and the semantic information of different channel features is interrelated. Therefore, we constructed a channel attention module to model the interdependencies between channels and enhance the representation of semantic features. We concatenate the features enriched with prior knowledge processed by the SAM-Med2D encoder and the representation of fundamental image characteristics processed by the MobileNet v3 encoder, then input them into the channel attention module. The structure of the channel attention module, as depicted in Fig. 8, reshapes the original feature $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ to $\mathbb{R}^{C\times N}$. Then, matrix multiplication is performed between X and its transpose, followed by a softmax layer to obtain the channel attention map $\theta \in \mathbb{R}^{C\times C}$. The formula is as (5), where θ_{ij} represents the

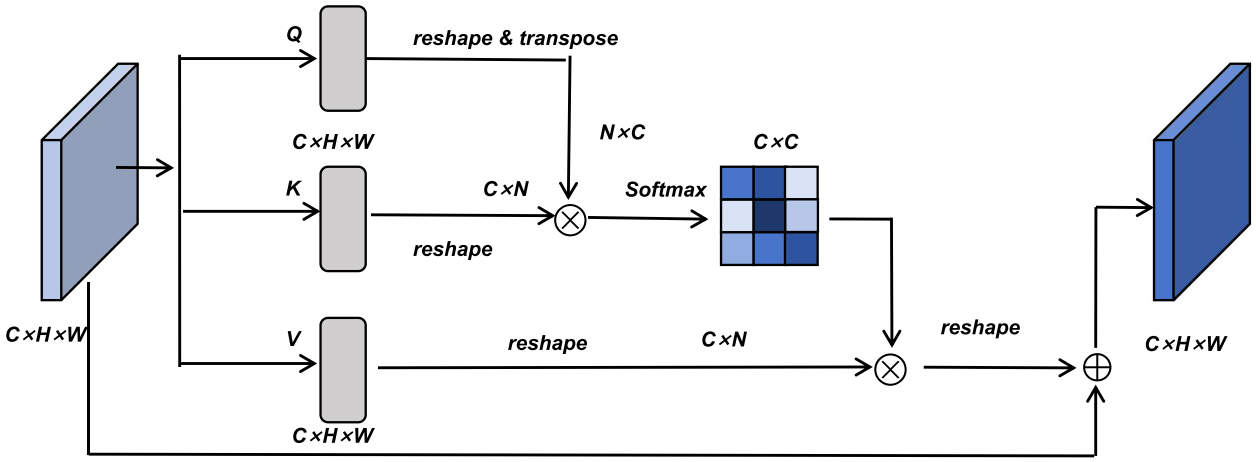

Fig. 8 Channel attention module

influence of channel j on channel i. Subsequently, we perform matrix multiplication between θ and the reshaped X to obtain the features calibrated by channel attention. The channel attention module also adopts a residual design, where the original feature map is added before the final output is produced. The overall formula can be expressed as (6):

$$\theta_{ij} = \frac{exp(X_i \cdot X_j)}{\sum_{j=1}^{C} exp(X_i \cdot X_j)} \tag{5}$$

$$X_{Ci} = \sum_{j=1}^{C} (\theta_{ij} A_j) + X_i \tag{6}$$

Experimental results

In this section, we present the experimental details and results of the proposed SAM-AutoMed and SAM-MedCls, and compare them with the latest research findings.

Datasets

As shown in the Figs. 9 and 10, we conducted experiments on six different medical image datasets, including three for medical image segmentation and three for medical image classification, selecting datasets from various modalities whenever possible. Three tasks were employed to validate the performance of SAM-AutoMed, including optic disc segmentation in retinal images, colon polyp segmentation in endoscopic images, and segmentation of pigment skin lesions in dermoscopy images. Three tasks were utilized to assess the performance of SAM-MedCls, including brain tumor classification on brain MRI images, classification of X-ray images for pneumonia detection, and classification of lung and colon cancer tissues in microscopic images.For optic disc segmentation, experiments were conducted on the REFUGE2 [19] dataset. For colon polyp segmentation, we use the clinic [20] dataset. For segmentation of pigment skin lesions, experiments were performed on the HAM10000 [21] dataset, a subset of ISIC. For brain tumor classification, experiments were conducted on the Brain Tumor MRI dataset obtained from the Kaggle repository, which comprises

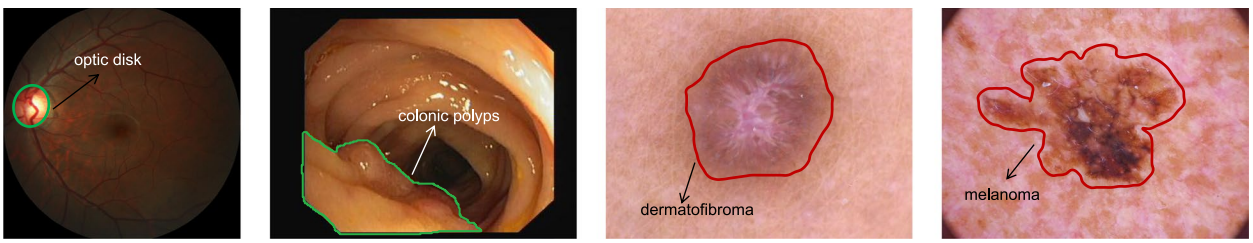

Fig. 9 **a** Sample image from REFUGE2. **b** Sample image from clinic. **c**, **d** Sample images from HAM10000

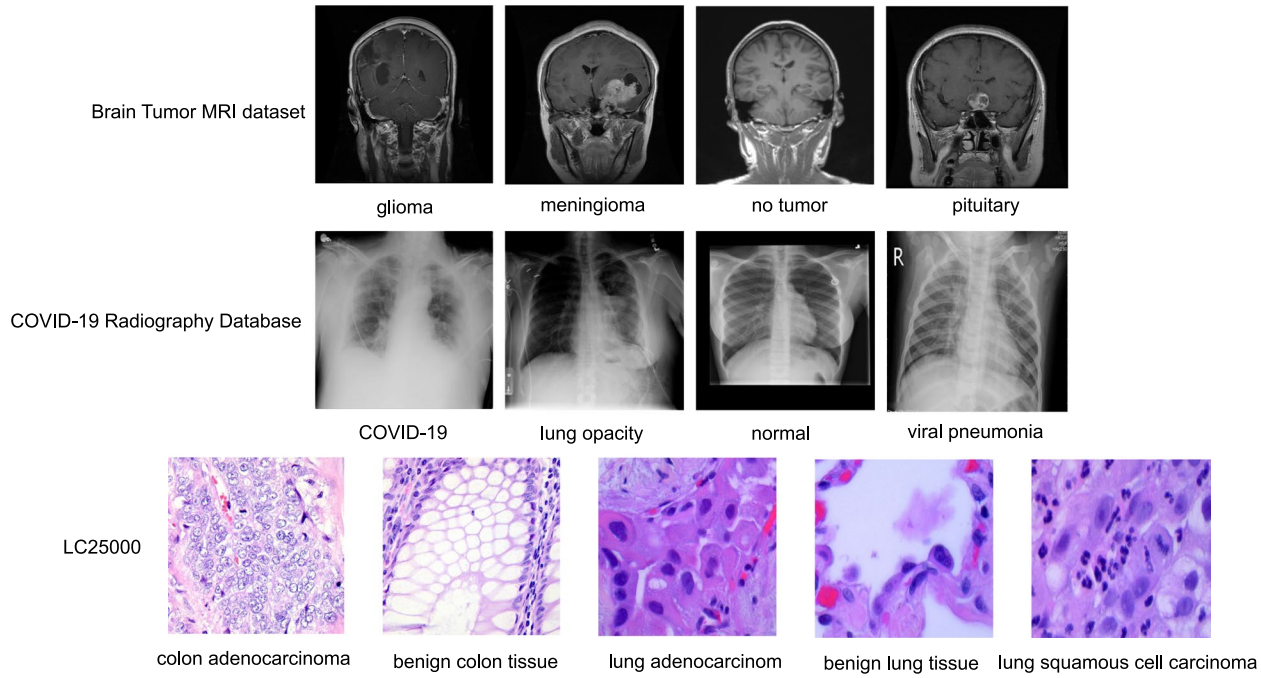

Fig. 10 Overview of brain tumor MRI dataset, COVID-19 radiography database and LC25000

a combination of datasets from Figshare [22], SARTAJ [23], and Br35H [23], containing a total of 7023 images categorized into glioma, meningioma, no tumor, and pituitary tumor classes. For pneumonia image classification, experiments were carried out on the COVID-19 Radiography Database [24, 25], consisting of 21165 images categorized into normal, viral pneumonia, lung opacity, and COVID-19 classes. Finally, for lung and colon cancer tissue classification, experiments were conducted on the LC25000 [26] dataset, which includes 25000 images categorized into colon adenocarcinoma, benign colon tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue classes.

Experiment details

In the preprocessing stage, we standardized the images and resized them to 256×256 dimensions. If both the height and width of an image were less than zero, we padded the edges of the image with zeros. Otherwise, we adjusted the image size using bilinear interpolation. For SAM-MedCls, we employed the same data augmentation techniques, including random horizontal flipping, random vertical flipping, and random rotation (with a scale range of (-40,40)). During the training phase, we conducted training on an NVIDIA2080 GPU. For SAM-AutoMed, we utilized the Adam optimizer with an initial learning rate of 1e-4 and trained for 30 epochs with a batch size of 2. For SAM-MedCls, we employed the Adam optimizer with an initial learning rate of 1e-3, a weight decay regularization parameter set to 0.4, and an epoch size of 50 with a batch size of 10.

Evaluation metrics

To evaluate the performance of our SAM-AutoMed model in image segmentation tasks, we used Mean Intersection over Union (mIoU) and Mean Dice Similarity Coefficient Score (mDSC). The mathematical expressions for IoU and DSC are as (7) and (8). Where S is the predicted segmentation mask, and G is the original ground truth mask of the image.

$$IoU(S,G) = \frac{|S \cap G|}{|S \cup G|} \tag{7}$$

$$DSC(S,G) = \frac{2 \times |S \cap G|}{|S| + |G|} \tag{8}$$

For evaluating the performance of our SAM-MedCls model in image classification tasks, we mainly used four evaluation metrics: Accuracy, Precision, Recall, and F1-Score. The formulas for these evaluation metrics are provided below, where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

$$\text{Precision(pre)} = \frac{TP}{TP + FP} \tag{9}$$

$$\text{Recall(rec)} = \frac{TP}{TP + FN} \tag{10}$$

$$\text{F1-score(F1)} = \frac{2 \times pre \times rec}{pre + rec} \tag{11}$$

$$\text{Accuracy(acc)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

Evaluation results on SAM-AutoMed

We validated the effectiveness of our SAM-AutoMed on medical image segmentation tasks using three different datasets, as shown in Tables 1, 2, and 3. Our model achieved mIoU scores of 88.1%, 86.5%, and 89.9% on the REFUGE2, clinic, and HAM10000 datasets, respectively, with corresponding mDSC scores of 93.6%, 92.2%, and 94.2%. In all three tables, we observed that on the REFUGE2 dataset, our model outperformed the previous state-of-the-art (Swin-UNetr), by 0.2% in mIoU. On the HAM10000 dataset, our model surpassed the SOTA by 6.2% in mIoU and 3.9% in mDSC. Although our model performed 2.5% lower in mIoU and 1.6% lower in mDSC compared to SOTA on the clinic dataset, it is worth noting that these methods often excel in their respective modalities or specific datasets. When applied to other domains, their performance may degrade. In contrast, our approach constructs an end-to-end model with simple and generalizable image preprocessing methods, which we believe can fully unleash its potential in the future. It is worth noting that, as shown in Table 4, SAM-AutoMed outperformed AutoSAM by 2.5%, 11.2%, and 1.6% in mIoU on the three datasets, respectively (under the same conditions for comparison). We attribute this improvement to two factors: first, SAM-AutoMed is an improvement over SAM-Med2D, with its encoder and mask decoder fine-tuned on large-scale medical datasets and thus having learned medical knowledge. As demonstrated in Table 4, the segmentation performance of MedSAM-2D was significantly improved over SAM after fine-tuning on all three medical datasets. Second, we designed an image encoder with parallel channel and spatial attention mechanisms SEblock, which provides stronger spatial perception compared to AutoSAM's encoder consisting of only simple convolutional layers.We visualized the segmentation results, as shown in Fig. 11, demonstrating the segmentation performance of SAM, SAM-Med2D, AutoSAM, and SAM-MedSAM on three datasets. From (c), it can be seen that SAM's segmentation performance is suboptimal across all three datasets, especially in fitting the segmentation masks to

Table 1 The comparison of SAM-AutoMed with state-of-the-art segmentation methods on REFUGE2

Method	mDSC(%)	mIoU(%)
ResUNet [27]	92.9	85.8
BEAL [28]	93.7	86.1
TransBTS [29]	94.1	87.2
EnsemDiff [30]	94.3	87.8
UltraUNet [31]	91.5	82.8
FAT-Net [32]	91.8	84.8
SegDiff [33]	92.6	85.2
nnUNet [34]	94.7	87.3
TransUNet [35]	95.0	87.7
UNetr [36]	94.9	87.5
Swin-UNetr [37]	**95.3**	87.9
SAM-AutoMed	93.6	**88.1**

Table 2 The comparison of SAM-AutoMed with state-of-the-art segmentation methods on clinic

Method	mDSC(%)	mIoU(%)
U-Net [38]	82.3	75.5
U-Net++ [39]	79.4	72.9
SFA [40]	70.0	60.7
MSEG [41]	90.9	86.4
DCRNet [42]	89.6	84.4
ACSNet [43]	88.2	82.6
PraNet [44]	89.9	84.9
EU-Net [45]	90.2	84.6
SANet [46]	91.6	85.9
Polyp-PVT [47]	93.7	88.9
FCN-Hardnet85 [48]	92.0	86.9
3P-SEG [49]	**93.8**	**89.0**
SAM-AutoMed	92.2	86.5

Table 3 The comparison of SAM-AutoMed with state-of-the-art segmentation methods on HAM10000

Method	mDSC(%)	mIoU(%)
Double U-Net [50]	84.3	81.2
U-Net [38]	78.1	77.4
SegNet [51]	81.6	82.1
Saha et al. [52]	89.1	81.9
Abraham et al. [53]	85.6	-
Shahin et al. [54]	90.3	83.7
Bissoto et al. [55]	87.3	79.2
Ibtehaz et al. [56]	-	80.3
SAM-AutoMed	**94.2**	**89.9**

Table 4 Comparison of SAM-AutoMed with SAM, SAM-Med2D, and AutoSAM on REFUGE2, clinic, and HAM10000

	REFUGE2		clinic		HAM10000	
Method	**mDSC(%)**	**mIoU(%)**	**mDSC(%)**	**mIoU(%)**	**mDSC(%)**	**mIoU(%)**
SAM	71.1	57.1	85.3	75.8	87.1	77.5
SAM-Med2D	91.9	85.1	89.6	82.4	93.1	87.2
AutoSAM	92.0	85.6	83.5	75.3	93.3	88.3
SAM-AutoMed	**93.6**	**88.1**	**92.2**	**86.5**	**94.2**	**89.9**

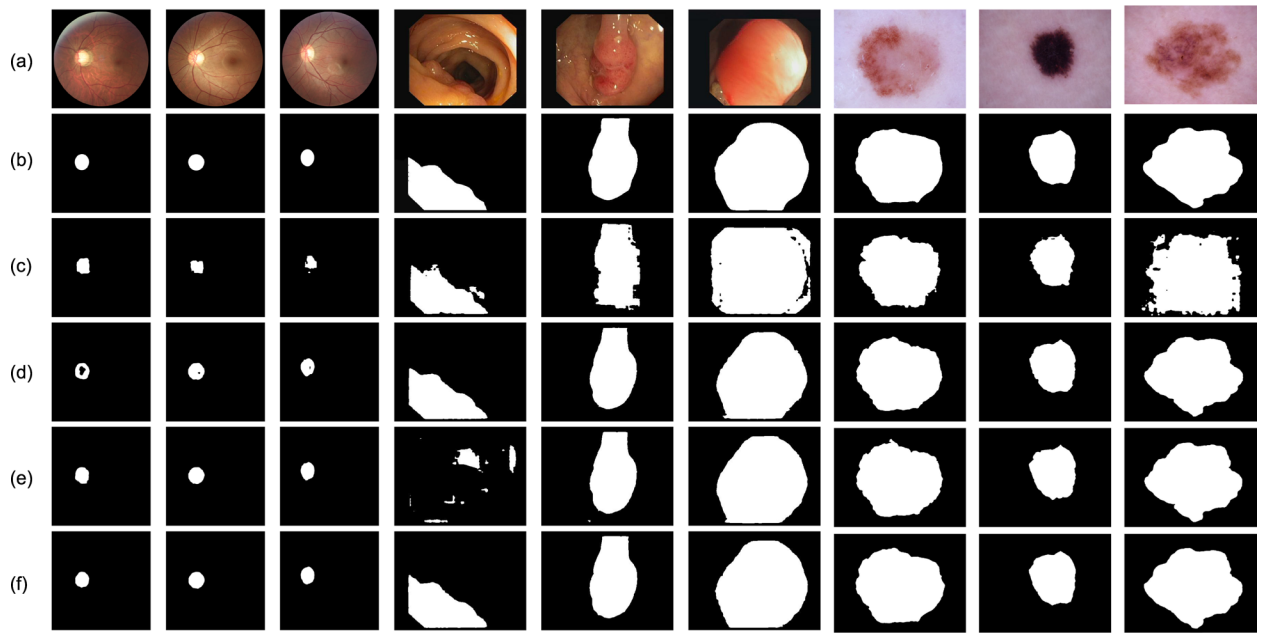

Fig. 11 Visualization of segmentation effects on three datasets(REFUGE2,clinic,HAM10000). **a** Images. **b** Ground truth. **c** SAM. **d** SAM-Med2D. **e** AutoSAM. **f** SAM-AutoMed

ground truth for polyp segmentation and pigment skin lesion segmentation tasks, attributed to the distinctive textures and boundaries in medical images compared to natural images. From (d) and (e), it is evident that SAM-Med2D and AutoSAM possess the ability for cross-modal medical image segmentation. However, SAM-Med2D sometimes excludes the optic cup during optic disc segmentation, showcasing its sensitivity to color features, which, though demonstrating its capability, is not medically trustworthy, emphasizing the necessity of task-specific fine-tuning. While AutoSAM undergoes task-specific fine-tuning, its lack of spatial awareness and training the decoder in a frozen manner may result in completely erroneous segmentations for certain samples, albeit with a low probability, which is intolerable. (f) presents the segmentation results of SAM-AutoMed, which essentially addresses the aforementioned issues.

Evaluation results on SAM-MedCls

The experimental results of SAM-MedCls on the three datasets, Brain Tumor MRI dataset, COVID-19 Radiography Database, and LC25000, are illustrated in Fig. 12 and presented in Table 5. The training loss curve and accuracy curve on three datasets are shown in Fig. 13. Based on the confusion matrix in Fig. 12, we calculated the accuracy, recall, F1-score, and overall accuracy for each class in the three datasets as listed in Table 5. From the results, except for slightly lower recall rates in the pituitary class for the Brain Tumor MRI dataset, COVID-19 class for the COVID-19 Radiography Database, and lung adenocarcinoma class for LC25000, all other classification metrics are high, indicating effective classification performance.To provide more convincing experimental results, we compared SAM-MedCls with state-of-the-art methods, as shown in Tables 6, 7, and 8. Our method achieved accuracies of 98.2%, 97.5%, and 97.7% on the three datasets, respectively. Under the

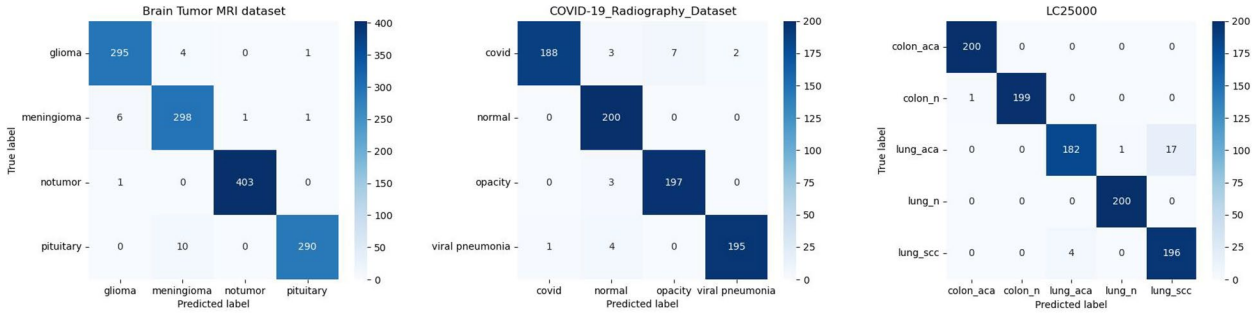

Fig. 12 **a** Confusion matrix of Brain Tumor MRI dataset. **b** Confusion matrix of COVID-19 Radiography Database. **c** Confusion matrix of LC25000

Table 5 Classification performance evaluation metrics (Precision, Recall, F1-Score and Accuracy) of SAM-MedCls on Brain Tumor MRI Dataset, COVID-19 Radiography Dataset and LC25000

Dataset	Class	Precision(%)	Recall(%)	F1-Score(%)
Brain Tumor MRI Dataset	glioma	97.7	98.3	98.0
	meningioma	95.5	97.4	96.4
	no tumor	99.8	99.8	99.8
	pituitary	99.3	96.7	98.0
Accuracy(%)				98.2
COVID-19 Radiography Dataset	COVID-19	99.5	94.0	96.7
	normal	95.2	100.0	97.6
	lung opacity	96.6	98.5	97.5
	viral	99.0	97.5	98.2
Accuracy(%)				97.5
LC25000	colon adenocarcinoma	99.5	100.0	99.8
	benign colon tissue	100.0	99.5	99.7
	lung adenocarcinom	97.8	91.0	94.3
	benign lung tissue	99.5	100.0	99.8
	lung squamous cell carcinoma	92.0	98.0	94.9
Accuracy(%)				97.7

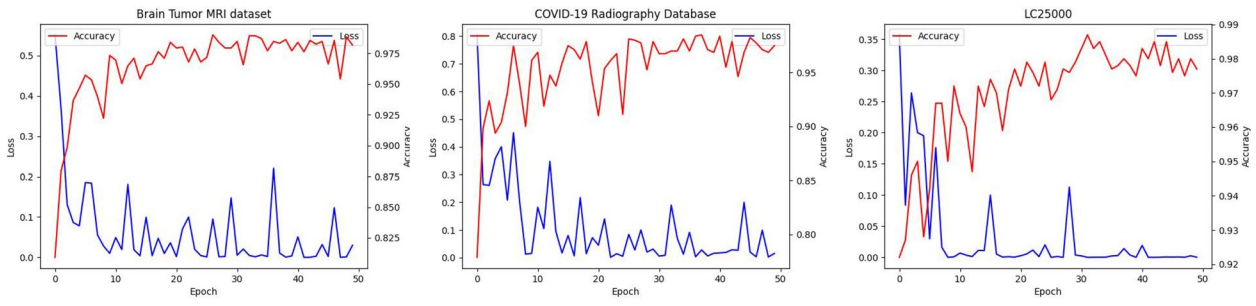

Fig. 13 **a** The training loss curve and accuracy curve of Brain Tumor MRI dataset. **b** The training loss curve and accuracy curve of COVID-19 Radiography Database. **c** The training loss curve and accuracy curve of LC25000

Table 6 The comparison of SAM-MedCls with state-of-the-art classification methods on Brain Tumor MRI Dataset

Reference	Model	Dataset	Classes	Best Model	Accuracy (%)
ref. [57]	CNN Multi Scale	Nanfang Hospital	3	-	97.3
ref. [58]	CNN	REMBRANDT	3	-	96.1
ref. [59]	TL	SARTAJ	3	InceptionResNetV2	98.9
ref. [60]	CNN and SVM	Figshare	3	-	95.8
ref. [61]	Dense Efficient-Net	Figshare	3	Dense EfficientNet	99.9
ref. [62]	LeNet Inspired Model	Figshare,Brainweb, Radiopedia	3	-	88.0
ref. [63]	TL and DeepTumorNet	Figshare	3	DeepTumorNet	99.7
ref. [64]	MobileNetV2 and TL	Figshare and BraTS 2018	3	Hybrid MobileNetV2	98.9
ref. [65]	GoogLeNet and TL	BR35H	2	GoogLeNet	99.1
ref. [66]	TumorResNet and TL	BTD-MRI dataset	2	TumorResNet	99.3
ref. [67]	Generic CNN and six TL models	Brain Tumor MRI dataset	4	InceptionV3	97.1
this work	SAM-MedCls	Brain Tumor MRI dataset	4	-	98.2

Table 7 The comparison of SAM-MedCls with state-of-the-art classification methods on COVID-19 Radiography Database

Reference	Dataset CXR Images	Classes	Best Model	Accuracy (%)
ref. [24]	3487	3(COVID-19, Viral Pneumonia, Normal)	DenseNet201	97.9
ref. [68]	7406	2(COVID-19, Viral Pneumonia)	ResNet101	99.5
ref. [69]	6432	3(COVID-19, Pneumonia, Normal)	Xception	98.0
ref. [25]	18479	3(COVID-19, Lung opacity, Normal)	ChexNet	96.3
ref. [70]	458	3(COVID-19, Pneumonia, Normal)	SqueezeNet	99.3
ref. [71]	1125	3(COVID-19, Pneumonia, No Findings)	DarkCovidNet	98.1
ref. [72]	13975	3(COVID-19, Pneumonia, Normal)	COVID-Net	93.3
ref. [73]	1251	4((COVID-19, Viral Pneumonia, Bacterial Pneumonia, Normal)	CoroNet	89.6
ref. [74]	9000	3(COVID-19, Pneumonia, No Findings)	Multiscale deep CNN	97.2
ref. [75]	450	3(COVID-19, Pneumonia, Normal)	QuNet	92.9
ref. [76]	21165	4(COVID-19, Viral Pneumonia, Normal, Lung Opacity)	Modified MobileNetV2	95.8
SAM-MedCls	21165	4(COVID-19, Viral Pneumonia, Normal, Lung Opacity)	-	97.5

Table 8 The comparison of SAM-MedCls with state-of-the-art classification methods on LC25000

Reference	Method	Accuracy(%)
ref. [77]	RESNET50	93.9
	RESNET18	93.0
	RESNET34	93.0
ref. [78]	Ensemble	91.0
ref. [79]	CNN-D	94.6
ref. [80]	CNN	97.2
ref. [81]	Shallow-CNN	97.9
ref. [82]	DL-based CNN	96.3
ref. [83]	UKSSL	98.9
this work	SAM-MedCls	97.7

same dataset conditions for brain tumor classification (MRI), our method outperformed the state-of-the-art by 1.1% in accuracy. Although SAM-MedCls did not achieve the highest accuracy among the methods compared on the respective datasets, it is worth noting that the datasets we used are larger in terms of both image and class quantities, which, to some extent, also reflects the advancement of SAM-MedCls. Similarly, on the COVID-19 Radiography Database, SAM-MedCls outperformed the state-of-the-art by 1.7% in accuracy. On the LC25000 dataset, our method achieved a slightly lower accuracy by 1.2% compared to the state-of-the-art. It is worth mentioning that our method achieved good results on datasets from three different modalities

Table 9 Results of ablation experiments using SAM-AutoMed on HAM10000

Encoder for SAM-Med2D	fine-tune mask decoder	MobileNet v3 with improved SE block	mDSC(%)	mIoU(%)
×	✓	✓	88.3	80.6
✓	×	✓	92.3	86.9
✓	✓	×	93.4	88.5
✓	✓	✓	94.2	89.9

Table 10 Results of ablation experiments using SAM-MedCls on Brain Tumor MRI dataset

Encoder for SAM-Med2D	MobileNet v3 without improved SE block	MobileNet v3 with improved SE block	attention module	Accuracy(%)
✓	×	×	×	67.9
×	✓	×	×	86.2
×	×	✓	×	89.5
✓	×	✓	×	93.1
✓	×	✓	✓	98.2

(MRI, X-ray, and microscopic imaging) without specific image processing tailored to a particular modality or dataset. This demonstrates the advancement of our method and its potential as a universal medical classification model.

Ablation experiment

We conducted ablation experiments to validate the effectiveness of our proposed SAM-AutoMed and SAM-MedCls design methods, improved SE block, and designed attention modules. Ablation experiments on ham10000 were performed on SAM-AutoMed, as shown in Table 9. When we replaced the encoder of SAM-Med2D with SAM's encoder, the mIoU decreased by 9.3% and mDSC decreased by 5.9% on ham10000, demonstrating that using SAM-Med2D's encoder significantly improves segmentation performance. When we trained SAM-AutoMed using the frozen decoder training method of AutoSAM, the performance decreased by 3% and 1.9% in mIoU and mDSC, respectively, indicating the effectiveness of our training method. When we did not use the improved SE block, the performance decreased by 1.4% and 0.8% in mIoU and mDSC, respectively, demonstrating that our improved SE block effectively extracts spatial information from the images.

We conducted ablation experiments on SAM-MedCls using the Brain Tumor MRI dataset, as shown in Table 10. When only the encoder of SAM-Med2D was used, followed by pooling layers and two MLP layers, the accuracy on the Brain Tumor MRI dataset was only 67.9%. Using MobileNet v3 for classification without the improved SE block yielded an accuracy of only 86.2%. However, incorporating the improved SE block into MobileNet v3 increased the classification accuracy by 3.3%. When concatenating the output features of SAM-Med2D's encoder with those of MobileNet v3 equipped with the improved SE block and feeding them into fully connected layers without passing through our designed attention module, the classification accuracy was 93.1%. In contrast, the classification accuracy of our SAM-MedCls was 98.2%, demonstrating the effectiveness of our proposed attention module.

Conclusions

In this paper, we propose two models based on SAM-Med2D for medical image analysis: SAM-AutoMed for automatic segmentation and SAM-MedCls for general medical image classification. For SAM-AutoMed, we replace the original encoder with a MobileNet v3 equipped with an improved SE block and design multiple loss functions to address class imbalance issues. This model achieves automatic segmentation of medical images, outperforming SAM, SAM-Med2D, and the previous method AutoSAM on multiple datasets. As for SAM-MedCls, we introduce a novel model structure that combines features with prior knowledge from SAM-Med2D and spatial information from MobileNet. These features pass through our designed spatial attention module and channel attention module before classification. This model achieves state-of-the-art performance on various datasets with different modalities, demonstrating its potential to become a universal model for medical image classification.

Acknowledgements

Not applicable.

Authors' contributions

J.S was responsible for conceptualizing the paper, designing research methods, programming, designing experiments and writing the initial draft.K.C and Z.H are responsible for conducting experiments and writing initial drafts.S.R, X.H, and X.L are responsible for verifying the experiment and visualizing the results.C.P is responsible for supervising and guiding the research process.

Funding

This research received no external funding.

Availability of data and materials

Datasets can be obtained from the Kaggle official website.
REFUGE2: https://www.kaggle.com/datasets/victorlemosml/refuge2.
clinic: https://www.kaggle.com/datasets/balraj98/cvcclinicdb.
HAM10000: https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification.

Brain Tumor MRI dataset: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.
COVID-19 Radiography Database: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.
LC25000: https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images.

Code availability
Code are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The dataset used in the study was obtained from a public repository (Kaggle), therefore ethical approval from the ethical committee is not applicable here. Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 March 2024 Accepted: 14 August 2024
Published online: 16 September 2024

References

1. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, et al. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision.2023. p. 4015–4026.
2. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nat Commun. 2024;15(1):654.
3. Cheng J, Ye J, Deng Z, Chen J, Li T, Wang H, et al. Sam-med2d. 2023. arXiv preprint arXiv:230816184.
4. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. In: International conference on machine learning. New York: PMLR; 2021. p. 8748–63.
5. Yuan L, Chen D, Chen YL, Codella N, Dai X, Gao J, et al. Florence: A new foundation model for computer vision. 2021. arXiv preprint arXiv:211111432.
6. Liu S, Zeng Z, Ren T, Li F, Zhang H, Yang J, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. 2023. arXiv preprint arXiv:230305499.
7. Li J, Li D, Xiong C, Hoi S. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In: International Conference on Machine Learning. New York: PMLR; 2022. p. 12888–900.
8. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020. arXiv preprint arXiv:201011929.
9. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:5998–6008.
10. Chen K, Liu C, Chen H, Zhang H, Li W, Zou Z, Shi Z. Rsprompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model. IEEE Transactions on Geoscience and Remote Sensing. 2024;62:1–17. https://doi.org/10.1109/TGRS.2024.3356074.
11. He C, Li K, Zhang Y, Xu G, Tang L, Zhang Y, Guo Z, Li X. Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. Adv Neural Inf Process Syst. 2024;36:30726–37.
12. Yang J, Gao M, Li Z, Gao S, Wang F, Zheng F. Track anything: Segment anything meets videos. 2023. arXiv preprint arXiv:230411968.
13. Wu J, Fu R, Fang H, Liu Y, Wang Z, Xu Y, et al. Medical sam adapter: Adapting segment anything model for medical image segmentation. 2023. arXiv preprint arXiv:230412620.
14. Ye J, Cheng J, Chen J, Deng Z, Li T, Wang H, et al. Sa-med2d-20m dataset: Segment anything in 2d medical imaging with 20 million masks. 2023. arXiv preprint arXiv:231111969.
15. Shaharabany T, Dahan A, Giryes R, Wolf L. AutoSAM: Adapting SAM to Medical Images by Overloading the Prompt Encoder. 2023. arXiv preprint arXiv:230606370.
16. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE; 2018. p. 7132–41.
17. Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, et al. Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. Piscataway: IEEE; 2019. p. 1314–24.
18. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et al. Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE; 2019. p. 3146–54.
19. Fang H, Li F, Fu H, Sun X, Cao X, Son J, et al. Refuge2 challenge: Treasure for multi-domain learning in glaucoma assessment. 2022. arXiv preprint arXiv:220208994.
20. Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput Med Imaging Graph. 2015;43:99–111.
21. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018;5:180161. Search in. 2018;2.
22. Cheng. Brain tumor dataset.figshare. Dataset. 2017. https://doi.org/10.6084/m9.figshare.1512427.v5.
23. Msoud Nickparvar. Brain Tumor MRI Dataset [Data set]. Kaggle. 2021. https://doi.org/10.34740/KAGGLE/DSV/2645886.
24. Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access. 2020;8:132665–76.
25. Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Kashem SBA, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput Biol Med. 2021;132:104319.
26. Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mastorides SM. Lung and colon cancer histopathological image dataset (lc25000). 2019. arXiv preprint arXiv:191212142.
27. Diakogiannis FI, Waldner F, Caccetta P, Wu C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens. 2020;162:94–114.
28. Wang S, Yu L, Li K, Yang X, Fu CW, Heng PA. Boundary and entropy-driven adversarial learning for fundus image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Cham: Springer; 2019. p. 102–10.
29. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. Transbts: Multimodal brain tumor segmentation using transformer. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Cham: Springer; 2021. p. 109–19.
30. Wolleb J, Sandkühler R, Bieder F, Valmaggia P, Cattin PC. Diffusion models for implicit image segmentation ensembles. In: International Conference on Medical Imaging with Deep Learning. New York: PMLR; 2022. p. 1336–48.
31. Chu C, Zheng J, Zhou Y. Ultrasonic thyroid nodule detection method based on U-Net network. Comput Methods Prog Biomed. 2021;199:105906.
32. Wu H, Chen S, Chen G, Wang W, Lei B, Wen Z. FAT-Net: Feature adaptive transformers for automated skin lesion segmentation. Med Image Anal. 2022;76:102327.
33. Amit T, Shaharbany T, Nachmani E, Wolf L. Segdiff: Image segmentation with diffusion probabilistic models. 2021. arXiv preprint arXiv:211200390.
34. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11.

35. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. 2021. arXiv preprint arXiv:210204306.
36. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. Piscataway: IEEE; 2022. p. 574–84.
37. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. Cham: Springer; 2021. p. 272–84.
38. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Cham: Springer; 2015. p. 234–41.
39. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Cham: Springer; 2018. p. 3–11.
40. Fang Y, Chen C, Yuan Y, Tong Ky. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Cham: Springer; 2019. p. 302–10.
41. Huang CH, Wu HY, Lin YLHM. A Simple Encoder-Decoder Polyp Segmentation Neural Network that Achieves over 0.9 Mean Dice and 86 FPS. 2021. arXiv preprint arXiv:210107172.
42. Yin Z, Liang K, Ma Z, Guo J. Duplex contextual relation network for polyp segmentation. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). Piscataway: IEEE; 2022. p. 1–5.
43. Zhang R, Li G, Li Z, Cui S, Qian D, Yu Y. Adaptive context selection for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. Cham: Springer; 2020. p. 253–62.
44. Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J, et al. Pranet: Parallel reverse attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention. Cham: Springer; 2020. p. 263–73.
45. Patel K, Bur AM, Wang G. Enhanced u-net: A feature enhancement network for polyp segmentation. In: 2021 18th Conference on Robots and Vision (CRV). Piscataway: IEEE; 2021. p. 181–8.
46. Wei J, Hu Y, Zhang R, Li Z, Zhou SK, Cui S. Shallow attention network for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Cham: Springer; 2021. p. 699–708.
47. Dong B, Wang W, Fan DP, Li J, Fu H, Shao L. Polyp-pvt: Polyp segmentation with pyramid vision transformers. 2021. arXiv preprint arXiv:210806932.
48. Chao P, Kao CY, Ruan YS, Huang CH, Lin YL. Hardnet: A low memory traffic network. In: Proceedings of the IEEE/CVF international conference on computer vision. Piscataway: IEEE; 2019. p. 3552–61.
49. Shaharabany T, Wolf L. End-to-End Segmentation of Medical Images via Patch-Wise Polygons Prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer; 2022. p. 308–18.
50. Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD. Doubleu-net: A deep convolutional neural network for medical image segmentation. In: 2020 IEEE 33rd International symposium on computer-based medical systems (CBMS). Piscataway: IEEE; 2020. p. 558–64.
51. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(12):2481–95.
52. Saha A, Prasad P, Thabit A. Leveraging adaptive color augmentation in convolutional neural networks for deep skin lesion segmentation. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Piscataway: IEEE; 2020. p. 2014–17.
53. Abraham N, Khan NM. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). Piscataway: IEEE; 2019. p. 683–7.
54. Shahin AH, Amer K, Elattar MA. Deep convolutional encoder-decoders with aggregated multi-resolution skip connections for skin lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Piscataway: IEEE; 2019. p. 451–4.
55. Bissoto A, Perez F, Ribeiro V, Fornaciali M, Avila S, Valle E. Deep-learning ensembles for skin-lesion segmentation, analysis, classification: RECOD titans at ISIC challenge 2018. 2018. arXiv preprint arXiv:180808480.
56. Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. 2020;121:74–87.
57. Díaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D. A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network. In: Healthcare. vol. 9. Basel: MDPI; 2021. p. 153.
58. Sultan HH, Salem NM, Al-Atabany W. Multi-classification of brain tumor images using deep neural network. IEEE Access. 2019;7:69215–25.
59. Ullah N, Khan JA, Khan MS, Khan W, Hassan I, Obayya M, et al. An effective approach to detect and identify brain tumors using transfer learning. Appl Sci. MDPI, Basel, Switzerland. 2022;12(11):5645.
60. Deepak S, Ameer P. Automated categorization of brain tumor from mri using cnn features and svm. J Ambient Intell Humanized Comput. 2021;12:8357–69.
61. Nayak DR, Padhy N, Mallick PK, Zymbler M, Kumar S. Brain tumor classification using dense efficient-net. Axioms. 2022;11(1):34.
62. Wahlang I, Maji AK, Saha G, Chakrabarti P, Jasinski M, Leonowicz Z, et al. Brain magnetic resonance imaging classification using deep learning architectures with gender and age. Sensors. 2022;22(5):1766.
63. Raza A, Ayub H, Khan JA, Ahmad I, S Salama A, Daradkeh YI, et al. A hybrid deep learning-based approach for brain tumor classification. Electronics. 2022;11(7):1146.
64. Maqsood S, Damaševičius R, Maskeliūnas R. Multi-modal brain tumor detection using deep neural network and multiclass SVM. Medicina. 2022;58(8):1090.
65. Amran GA, Alsharam MS, Blajam AOA, Hasan AA, Alfaifi MY, Amran MH, et al. Brain Tumor Classification and Detection Using Hybrid Deep Tumor Network. Electronics. 2022;11(21):3457.
66. Ullah N, Khan MS, Khan JA, Choi A, Anwar MS. A robust end-to-end deep learning-based approach for effective and reliable BTD using MR images. Sensors. 2022;22(19):7575.
67. Gómez-Guzmán MA, Jiménez-Beristaín L, García-Guerrero EE, López-Bonilla OR, Tamayo-Perez UJ, Esqueda-Elizondo JJ, et al. Classifying Brain Tumors on Magnetic Resonance Imaging by Using Convolutional Neural Networks. Electronics. 2023;12(4):955.
68. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Anal Applic. 2021;24:1207–20.
69. Jain R, Gupta M, Taneja S, Hemanth DJ. Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl Intell. 2021;51:1690–700.
70. Toğaçar M, Ergen B, Cömert Z. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. Comput Biol Med. 2020;121:103805.
71. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020;121:103792.
72. Wang L, Lin ZQ, Wong A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Sci Rep. 2020;10(1):19549.
73. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Prog Biomed. 2020;196:105581.
74. Muralidharan N, Gupta S, Prusty MR, Tripathy RK. Detection of COVID19 from X-ray images using multiscale Deep Convolutional Neural Network. Appl Soft Comput. 2022;119:108610.

75. Asghar U, Arif M, Ejaz K, Vicoveanu D, Izdrui D, Geman O. [retracted] an improved covid-19 detection using gan-based data augmentation and novel qunet-based classification. BioMed Res Int. 2022;2022(1):8925930.
76. Sanida T, Sideris A, Tsiktsiris D, Dasygenis M. Lightweight neural network for COVID-19 detection from chest X-ray images implemented on an embedded system. Technologies. 2022;10(2):37.
77. Bukhari SUK, Syed A, Bokhari SKA, Hussain SS, Armaghan SU, Shah SSH. The histological diagnosis of colonic adenocarcinoma by applying partial self supervised learning. MedRxiv. 2020;2020–08.
78. Phankokkruad M. Ensemble transfer learning for lung cancer detection. In: 2021 4th international conference on data science and information technology. New York: ACM; 2021. p. 438–42.
79. Hlavcheva D, Yaloveha V, Podorozhniak A, Kuchuk H. Comparison of CNNs for lung biopsy images classification. In: 2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE; 2021. pp. 1–5.
80. Hatuwal BK, Thapa HC. Lung cancer detection using convolutional neural network on histopathological images. Int J Comput Trends Technol. 2020;68(10):21–4.
81. Mangal S, Chaurasia A, Khajanchi A. Convolution neural networks for diagnosing colon and lung cancer histopathological images. 2020. arXiv preprint arXiv:200903878.
82. Masud M, Sikder N, Nahid AA, Bairagi AK, AlZain MA. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors. 2021;21(3):748.
83. Ren Z, Kong X, Zhang Y, Wang S. Ukssl: Underlying knowledge based semi-supervised learning for medical image classification. IEEE Open J Eng Med Biol. 2024;5:459–66.

Publisher's Note