Biomedical Signal Processing and Control 110 (2025) 108251

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Robust semantic learning for precise medical image segmentation

Snehashis Chakraborty [a], Komal Kumar [a], Ankan Deria [a], Dwarikanath Mahapatra [b], Behzad Bozorgtabar [c], Sudipta Roy [a,d,*]

[a] Artificial Intelligence & Data Science, Jio Institute, Sector 4, Ulwe, Navi Mumbai, 410206, Maharashtra, India
[b] Department of Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates
[c] Swiss Federal Institute of Technology Lausanne (EPFL) University Hospital Center (CHUV), Lausanne, Switzerland
[d] Department of Computer Science Engineering, Mahindra University, Hyderabad, India

ARTICLE INFO

Keywords:
Semantic feature enhancement
Precise segmentation
Computational efficient
Multimodal images

ABSTRACT

Precisely localizing anomalies in medical images remains a significant challenge due to their heterogeneous nature across modalities and organs. While initial efforts excelled in identifying prominent anomalies, detecting minute target lesions posed significant limitations. These minute anomalies are particularly elusive and demand advanced detection techniques. Additionally, many existing models demand high computational resources, limiting their practicality in real-world clinical settings. In this study, we present REUnet, a novel Unet based architecture designed to address these obstacles by providing precise segmentation while also exhibiting strong generalization across diverse modalities and organs. The core advantage of REUnet resides in its resilient encoding pathway, constructed upon a module called dynamic mobile inverted bottleneck convolution. This module introduces a gating signal that significantly enhances semantic information, enabling the model to focus on specific regions of interest. The encoding pathway of REUnet is also linked strategically with the decoder to ensure efficient processing of these robust features which facilitates better communication between the two. Furthermore, the use of depth-wise separable convolution and dropout layers further makes REUnet computationally efficient for clinical use. Extensive experiments conducted on five publicly available datasets, including DUKE, BRATS2020, KiTS2023, INBreast, and FracAtlas, demonstrate REUnet's strong generalization capabilities and superior performance, establishing a new state-of-the-art in medical image segmentation. The source code is available at GitHub link: https://github.com/labsroy007/RobustSemanticLearning.

1. Introduction

Over the past few decades, there has been an exponential increase in the utilization of artificial intelligence, particularly in the field of medicine, driven by the growing volume of available data [1–4]. Numerous attempts have been conducted in this domain to address a wide array of challenges, including classification and segmentation of various anomalies and target lesions. But the reason why medical image segmentation is still challenging is due to the unique characteristics of diverse multi modal images, including variations in data distribution, resolution, acquisition protocols, and more, which hinder the generalization of deep learning models across them. Another significant limitation of these models is the inability in understanding the heterogeneous appearance of anomalies or target lesions such as variations in shape, size, texture, contrast, and other visual characteristics in various anatomical regions. Even expert radiologists often face challenges in diagnosing these anomalies, potentially impacting patient outcomes. Fig. 1 shows some examples of visually challenging anomalous regions pointed by arrow across different multi-modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Mammography and X-ray spanning different body parts like kidney, breast, brain and bone.

The development of U-Net [5] marked a significant milestone in the field of medical image segmentation, demonstrating exceptional proficiency in segmenting cells and organs, and setting new benchmarks for precision. However, U-Net faced two major challenges: overfitting on small datasets due to its large parameter count, and a loss of spatial information caused by excessive downsampling. These limitations encouraged the creation of numerous U-Net variants [6–11] aimed at addressing these issues. Enhancements included improvements to skip connections through multi-layer feature fusion, the use of dilated convolutions to expand the receptive field, the addition of residual or dense blocks, and the incorporation of transformer architectures.

* Corresponding author.
E-mail addresses: Snehashis1.C@jioinstitute.edu.in (S. Chakraborty), suryavansi8650@gmail.com (K. Kumar), ankanderia01@gmail.com (A. Deria), dmahapatra@gmail.com (D. Mahapatra), behzad.bozorgtabar@epfl.ch (B. Bozorgtabar), sudiptaroy01@yahoo.com (S. Roy).

https://doi.org/10.1016/j.bspc.2025.108251
Received 20 December 2024; Received in revised form 17 May 2025; Accepted 23 June 2025
Available online 10 July 2025

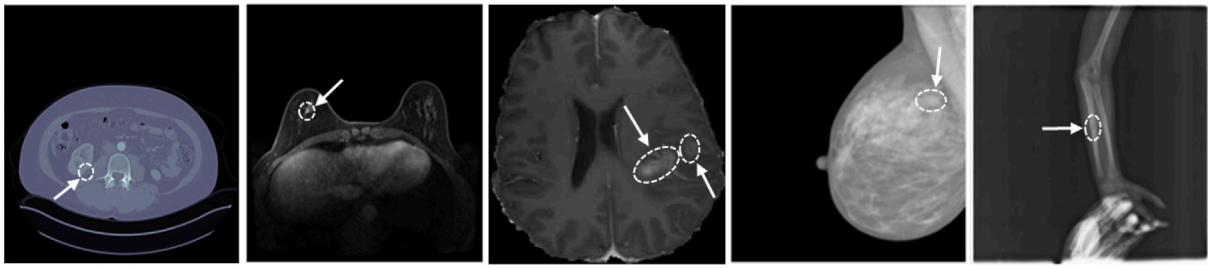

Fig. 1. The heterogeneous nature of abnormalities across different modalities as well as body parts poses significant challenges and often lead to misdiagnosis, even by expert clinicians.

These modifications significantly boosted accuracy and redefined the standard U-Net architecture, leading to state-of-the-art performance. However, high false alarm rates, incapability to comprehend inter-class relationships in the context of anomalies, and the inability to pinpoint minute target lesions, even in cases where medical expertise is required are their major drawbacks. Additionally, the computationally intensive nature of these models make them unsuitable for clinical practice. Therefore, the development of a robust, light weighted and generalized model is very much required that can not only segment target lesions regardless of any size or shape, but is also ready for clinical deployment.

In this study, we introduce a novel architecture named REUnet, that stands out for its robust encoding pathway complemented by an effective bottleneck mechanism. REUnet features strong encoding blocks that replace the conventional encoding layers of U-net, with an additional encoding block inserted as an Intermediator in between the encoder and decoder paths. The role of this intermediate block becomes crucial as it not only provides the highest level of feature abstraction but also strongly binds with the encoding and decoding blocks present at the lower level. Each of these encoding blocks in REUnet is built upon a series of module called dynamic mobile inverted bottleneck convolution (DMBC), a modification over mobile inverted bottleneck convolution from the EfficientNet-b7 architecture [12]. DMBC incorporates Gating Signal along with squeeze-and-excite attention mechanism [13], enhancing the model's ability to focus on intrinsic features and understand inter-class relationships. The integration of these robust encoding blocks within the encoding pathway not only excel in extracting salient semantic features but also significantly boosts the processing capabilities of these features into the decoder, showcasing the strength and effectiveness of our model in segmenting even complex regions of interest across various modalities and body parts after training. Moreover, we introduce attention gates via skip connections between each encoding and decoding block, helping the model to focus on specific regions of interest. Additionally, to optimize REUnet for clinical practices, we use depth-wise separable convolutions within DMBC to reduce computational complexity.

In summary, this paper presents the following key contributions:

- We propose REUnet, a robust architecture that provides precise segmentation of target lesions because of its seamless procession and integration of robust semantic features throughout the architecture.
- We introduce Gating Signal in DMBC that not only enhances the semantic flow of information but also improves the model's grasp of inter-class relationships, leading to finer separation of target lesions from its surroundings.
- Execution of experiments on diverse multi modal datasets, encompassing CT, MRI, Mamo, and XRay imaging, to assess the model's efficiency in accurately segmenting and localizing anomalies within multiple organs.

2. Related work

Precise segmentation of medical image anomalies is vital for guiding subsequent medical interventions. Traditional techniques like Graph-Based Methods [14], Region Growing [15], Markov Random Fields [16] and many more, while foundational, relied on handcrafted features, lacked context awareness, and lacked end-to-end learning capabilities. This led to the rise of deep neural networks, particularly convolutional neural networks (CNNs) that addressed these limitations by automating feature extraction and delivering cutting-edge performance in semantic segmentation, marking a transformative shift in medical image analysis.

One of the ground-breaking inventions of CNN in the field of medical image segmentation was U-net which was an encoder–decoder based fully convolutional network (FCN). U-net addressed the problem of data availability which is very common in medical domain. However, several drawbacks like, inability to effectively address small lesion segmentation and limitation in identifying regions of interest with irregular or non-standard shapes led to the development of several modified versions of U-net which includes integration of various CNN based architectures like VGG16 [17], ResNet [18], Xception [19] and so on as feature extractor. Though these architectures improved the performance of U-net by a good margin, they are computationally expensive and tend to overfit due to presence of large number of parameters. To prevent overfitting while improving both global and local feature extraction, several works [20,21] incorporated more advanced skip connections by aggregating features from multiple levels. Although these models provide better performance as compared to the previously mentioned architecture, they fail to generalize well across other data distributions as they are unable to explore a sufficient amount of information from full scales. This issue was resolved by [22] by introducing dilated convolution which helped the architecture in preserving both local details and global context, which is useful for tasks with complex backgrounds. But the method oversmooth the finer details in the segmentation results, especially for small objects that leads to high false alarms. Another modification includes the use of attention gates in medical image segmentation tasks [23] that was previously limited to natural language processing. Utilizing attention gates recursively, served to augment the receptive fields of convolutional filters while fostering a global perspective on tissue relationships. Yet, the challenge of curbing false positive predictions for smaller objects remained a barrier to the successful generation of segmentation masks for these specific objects of interest. Further improvements include the integration of transformers [9,10] which fasten model convergence by focusing on regions overlooked by CNN layers due to their large kernel size. These models demonstrated the ability to precisely segment target lesions by reducing the number of falsely predicted pixels. Recent advancements in foundational models, such as MedSAM [24], which utilizes Vision Transformer [25] as its core architecture have shown impressive performance across a wide range of medical images. However, these models require additional inputs like bounding boxes, masks, or points indicating suspected areas to achieve precision. Such inputs may not always be available or could be inaccurate, as they depend on domain experts. This introduces potential errors, especially when dealing with minute or hard-to-detect anomalous regions and tumors. Moreover, their reliance on significant computational resources and high-performance computing infrastructure for efficient training or fine-tuning constitutes another significant drawback.

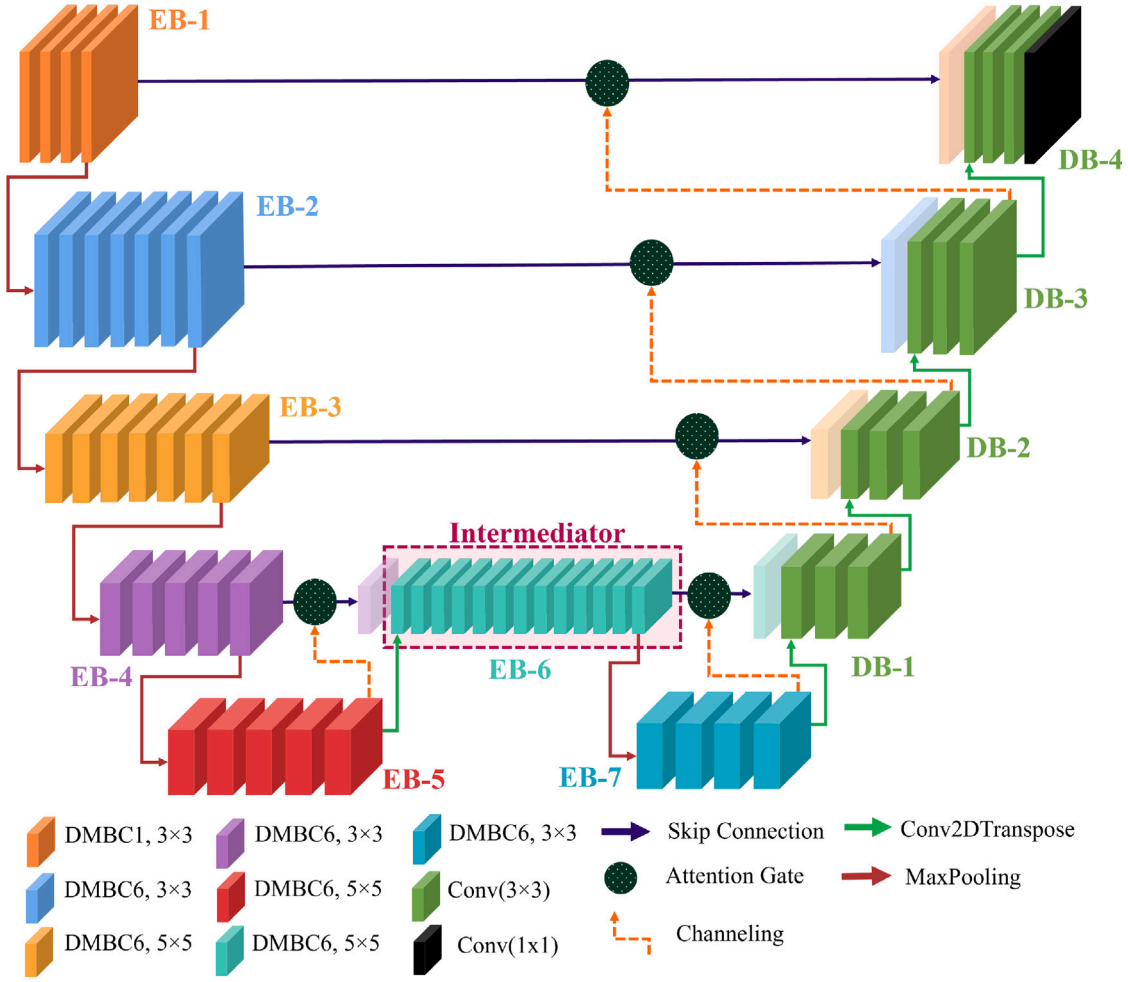

Fig. 2. The architecture of REUnet with encoding and decoding route. Within the encoding route, each block (EB) is composed of a sequence of DMBC modules. Each DMBC module (DMBCi) is accompanied by its respective filter size, and it employs either the standard ReLU activation function (**i** = 1) or the ReLU6 activation function (**i** = 6). The decoding route on the other side comprises of multiple decoding blocks (DB) for obtaining the segmentation mask.

All the works mentioned above addressed various challenges with respect to semantic segmentation of target lesions and achieved SOTA performance. However, challenges like tackling heterogeneous appearance of target lesions or anomalies across different modalities and different body parts, improving fine edges segmentation by proper understanding of inter-class relationship when dealing with multiple regions of interest and complexity optimization of the model are still open problems in medical image segmentation task.

3. Methodology

We present an overview of REUnet in Fig. 2. The encoding pathway of REUnet comprises of seven robust encoding blocks (EBs) while the decoding pathway is similar to that of U-net. Each of these EBs is further made up of a series of module named DMBC. The integration of DMBC within each EBs plays a pivotal role in constructing a resilient encoder that concentrates solely on relevant features resulting in robust feature refinement. To boost the processing of these features into the decoder, we strategically position EB-6 (containing thirteen DMBC modules) in such a way that it acts as an Intermediator between both the pathways. The role of EB-6 becomes pivotal as it not only captures the high-level semantic features but also strongly binds with EB-4, EB-5, EB-7 and DB-1, strengthening the bottleneck region. The inclusion of dropout layers after each EB further prevents the model from over fitting. The utilization of depth-wise separable convolution within DMBC also contributes to a reduced number of parameters, making the model well-suited for clinical applications. More details about the architecture are provided in the following sections.

3.1. Encoding route

The encoding route comprises of seven encoding blocks. These blocks collectively form the structure of EfficientNet-b7 architecture based on filter size, striding and number of channels. The fundamental unit of each of these encoding blocks is DMBC which is a modified version of mobile inverted bottleneck convolution. The architecture of DMBC is shown in Fig. 3. DMBC is formulated as follows, for an input tensor $x \in \mathbb{R}^{H_i \times W_i \times C_i}$, where H_i is the height, W_i is the width and C_i is the number of channels of the input respectively. We start by defining a Conv(1 × 1) followed by BatchNorm and activation using a function $x' = \{f_1(x),\ f_1 : \mathbb{R}^{H_i \times W_i \times C_i} \rightarrow \mathbb{R}^{H_i \times W_i \times d}\}$ and calculated d as follows:

$$\mu = C_i \times \eta \times w \tag{1}$$

$$\upsilon = \max(\text{divisor}, \left\lfloor \mu + \frac{\text{divisor}}{2} \right\rfloor / \text{divisor}^2) \tag{2}$$

$$d = \begin{cases} \upsilon & \text{if } \upsilon \geq 0.9\mu \\ \upsilon + \text{divisor} & \text{otherwise} \end{cases} \tag{3}$$

where η is expansion factor which is different for all block, w is the width coefficient and 'divisor' is a constant value which is set to 8 for all

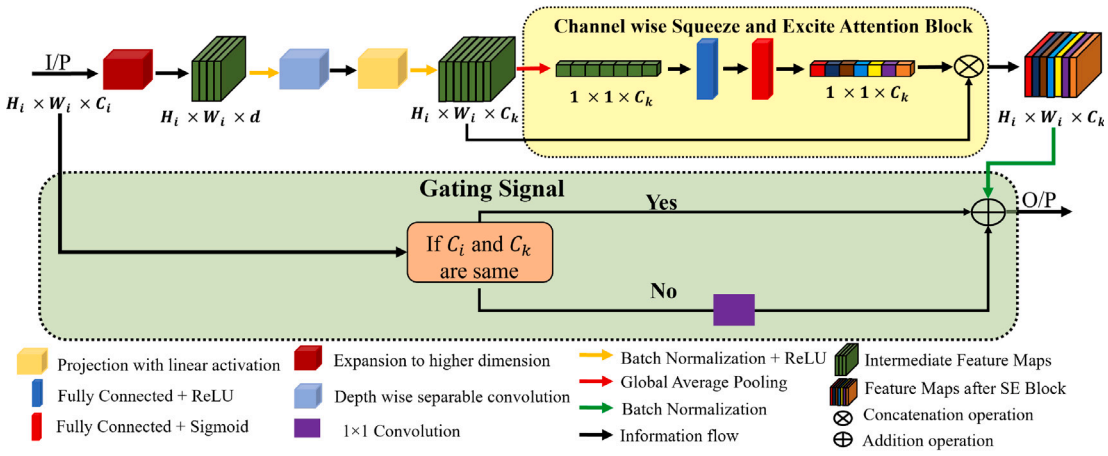

Fig. 3. The architectural design of Dynamical Mobile Inverted Bottleneck Convolution (DMBC) which implements Gating Signal along with depth wise separable convolution and squeeze and excite attention block. DMBC takes an input tensor of shape $(H_i \times W_i \times C_i)$ and outputs a tensor of shape $(H_i \times W_i \times C_k)$.

the blocks, unless specified explicitly [12]. The above process enables the capture of intricate and higher-level features which is denoted by x'.

Furthermore, $x'' = \{f_2(x'),\ f_2 : \mathbb{R}^{H_i \times W_i \times d} \rightarrow \mathbb{R}^{H_i \times W_i \times c_k}\}$ consists of depth wise separable convolution followed by a linearly activated Conv(1 × 1) for projecting x' back to low-dimensional space x''. This enables us to effectively reduce the parameter count which makes the model computationally efficient. We also employ squeeze-excite-scale (SES) for channel wise feature response [13] on the output of f_2:

$$\begin{aligned} &\text{S:} \quad Z_S = \text{GlobalAvgPool}(x'') \\ &\text{E:} \quad S_E = \sigma(W_2 \delta(W_1 \times Z_S)) \\ &\text{S:} \quad Z'_S = S_E \odot x'' \end{aligned} \tag{4}$$

Where W_1 and W_2 are weights of fully connected layers which is activated by ReLu (δ), and Sigmoid (σ) respectively. Z'_S enhances the model's performance by prioritizing the most relevant information (channel wise) based on global context. All the above operations (f_1, f_2 and SES) till now in the block can be formulated by supposing a composite function, $\mathcal{F} = \text{SES} \circ f_2 \circ f_1$.

Sometimes training deep architectures often lead to vanishing gradient problem where the gradient flow in the earlier layers becomes negligible which leads to information loss. To address this issue, we employ a Gating Signal (highlighted in green) in Fig. 3 which adds the output tensor from the function $\mathcal{F}$ (after batch normalization) with the initial input tensor x. The Gating Signal first checks the shape of both the tensors and performs addition if their shapes are same or performs a pointwise convolution operation on x, followed by addition with $\mathcal{F}(x)$. Finally the overall operation after incorporation of gating signal in DMBC can be formulated as:

$$\text{DMBC(x)} = \begin{cases} \mathcal{F}(x) + x & \text{try} \\ \mathcal{F}(x) + \text{Conv}(1 \times 1)(x) & \text{otherwise} \end{cases} \tag{5}$$

This improvement (Eq. (5)) not only mitigates the vanishing gradient problem, but also improves the learning of interclass relationship which results in fine separation of multiple target lesions from each other. We also added dropout layer with a rate of 0.2 after each EB to further prevent the model from over fitting.

3.2. Decoding route

The decoding route starts after EB-7 and has 4 decoding blocks (DBs), as seen in Fig. 2. The feature maps derived from EB-7 are up-sampled and concatenated with the attention guided feature maps and are channeled into the 1st decoder block (DB-1). Within this initial decoding block, the feature maps undergo a sequence of two convolution layers which are consecutively followed by ReLU activations. The resulting feature maps are again up-sampled and concatenated with its corresponding attention guided feature maps and are passed to the next decoding block. This process continues till the final decoding block (DB-4) which has three convolution layers. The feature maps after the 2nd convolution layer are passed to the last convolution layer which has a 1 × 1 filter to transform the feature vector into the desired number of classes.

3.3. Attention gate

REUnet makes use of attention guided feature maps to enhance the localization of target lesions. The generation of attention guided feature maps is demonstrated in Fig. 4. The attention gate starts by taking two feature vectors as inputs, which are the encoding signal E_{signal} and the decoding signal D_{signal}. E_{signal} are the feature maps that come from the corresponding encoding block through skip connection ($E_{signal} \in \mathbb{R}^{H_e \times W_e \times C_e}$). These feature maps already contain a substantial amount of semantic information regarding the input, as it goes through a series of DMBC within the same block. D_{signal} on the other side is the convoluted and transformed feature maps channeled from the decoder block positioned immediately below the corresponding encoding block ($D_{signal} \in \mathbb{R}^{H_d \times W_d \times C_d}$). Inside the attention gate, D_{signal} passes through an upsampling operation U while E_{signal} passes through a Conv (3 × 3) f. Then both signals are added and passed through a ReLU (δ), followed by linear transformation using a Conv(1 × 1). The above process is formulated as:

$$z = \text{Conv}(1 \times 1)[\delta\{U(D_{\text{signal}}) + f(E_{\text{signal}})\}] \tag{6}$$

The resultant activation map $z \in \mathbb{R}^{H_d \times W_d \times 1}$ is then passed through a Sigmoid (σ) function followed by an up-sampling operation to generate attention coefficients $A_{coef} \in \mathbb{R}^{H_e \times W_e \times 1}$. Finally, attention-guided feature maps A_{maps} are produced by repeating A_{coef} channel-wise and then multiplying it point-wise $\otimes$ with E_{signal}. The formula to generate A_{maps} is shown as:

$$A_{maps} = A_{coef} \otimes E_{signal} \tag{7}$$

4. Experiments

4.1. Datasets

To assess REUnet's performance, we utilized five diverse datasets covering various modalities across different anatomical regions. We

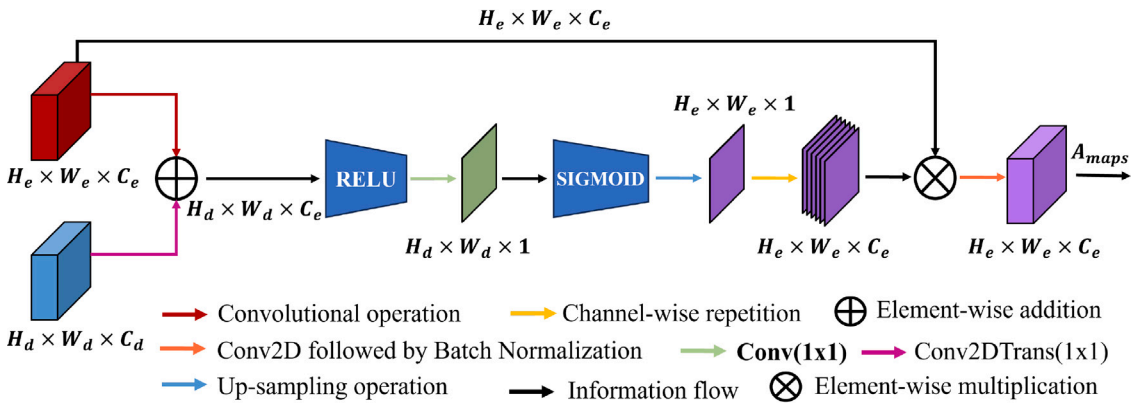

Fig. 4. A pictorial illustration of the generation of attention guided feature maps (A_{maps}) through attention gate in REUnet.

Table 1
Summarized description of the datasets used in this work. The fourth column represents the number of images used from the corresponding dataset.

Dataset	Modality	Target location	$\mathbb{I}_N$
DUKE	MRI (DCE)	Breast	700
BraTS2020	MRI (T1w, T2, FLAIR)	Brain	2400
KiTS2023	CT	Kidney	600
INBreast	Mammography	Breast	250
FracAtlas	X-ray	Bone	500

provide the summarized table of dataset details in Table 1. For MRI analysis, we used the Duke Breast Cancer [26] and BraTS2020 [27] datasets. The DUKE dataset includes DCE (Dynamic Contrast Enhanced) and T1 MRI breast images of 529 patients. Manual tumor annotations for 30 randomly selected patients were conducted using the MicroDICOM viewer and exported in DICOM format. BraTS2020 provides annotations for native (T1), post-contrast T1-weighted, T2-weighted, and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) sequences, covering brain anomalies like enhancing tumor (ET), peritumoral edema (ED), and the necrotic and non-enhancing tumor core (NCR/NET). Scans are available in NifTI format. For CT evaluation, we utilized the KiTS2023 [28] dataset, containing annotations of kidney, renal tumors, and renal cysts for 589 patients, also available as NifTI files. For Mammogram, we used the INBreast dataset [29], which comprises 410 breast images in CC (craniocaudal) and MLO (mediolateral oblique) views, stored in DICOM format, with tumor annotations provided in XML files. Finally, for X-ray analysis, the FracAtlas dataset [30] was utilized, containing 717 annotated images of bone fractures across various body parts, stored in JPG format, with annotations provided in JSON format.

4.2. Implementation details

All the experiments are conducted on a NVIDIA RTX A4000 GPU with 16 GB RAM and for calculating the processing speed of the models in low computational environment, an Intel chip with i9 processor and 128 GB RAM was used as CPU. All the models are built with the help of Tensorflow 2.0. We train REUnet till 150 epochs with 4 as batch size and an early stopping with patience of 5 is used to prevent further over fitting. As an optimizer, Adam is used with a default learning rate of 0.001, and crossentropy is used as loss function. For multi sequential datasets like BraTS2020, we concatenated and stacked the images from three sequences (T1-weighted, T2-weighted and T2-FLAIR) from the same position. This was done to utilize different information from these sequences in model training. Prior to training, all the images of the above datasets underwent similar pre-processing steps including normalization and resizing to dimensions of 256 × 256x3. And the training, validation and testing sets for each dataset is obtained by splitting them randomly with a ratio of 60:20:20.

4.3. Comparison with SOTA methods

Quantitative Analysis. The performance of the proposed REUnet was compared with other SOTA architectures. As quantitative measures, all the methods were evaluated under mean Intersection over Union (IoU) and mean Dice Similarity Coefficient (Dice) for all the datasets. Table 2 provides a comprehensive view of the quantitative analysis results, where the performance of REUnet is at the bottom.

REUnet outperforms all the methods mentioned in Table 2 by a good margin which showcases its efficiency in the field of segmentation. The reason behind this improvement is the robust encoding blocks incorporated into the model which increases the semantic information of the input thus empowering the decoder to pinpoint the area of interest with high precision. The evidence for this advancement is evident in the table where it improves the mean Dice score by a substantial margin of no less than 3% for breast (MRI), 5% for brain (MRI), 3% for kidney (CT), 2% for bone (X-ray), and 4% for breast (Mammography) datasets. Apart from using mean IoU and mean Dice, we further compared the performance of REUnet in terms of mean precision and mean recall for all the datasets. Fig. 5, highlights the performance of the models in the form of stacked plots where it can be observed that REUnet achieves highest mean precision and mean recall scores for DUKE, BraTS2020 and FracAtLas datasets respectively. It is also noteworthy that, although REUnet's precision matches that of Unet3+ on the KiTS2023 dataset, it surpasses Unet3+ by a remarkable 4% margin overall. A similar pattern is seen with the INBreast dataset, where REUnet's precision trails Swin-Unet by 1%, but it outperforms Swin-Unet in recall by an impressive 12%. This highlights REUnet's strength in producing accurate segmentation masks by reducing misclassified pixels

We also calculated the receiver operating characteristic curve (ROC curve), Fig. 6, which compares REUnet, Transunet, Unet 3+, Unet++, DR-Unet104 and Swin-unet for three different datasets. As seen in Fig. 6, REUnet performs better than all the other models, covering the highest area under the curve (AUC) score of 0.997, 0.995 and 0.966 for DUKE, INBreast and FracAtLas datasets respectively.

Qualitative Analysis. In terms of qualitative analysis, Fig. 7 demonstrates the superiority of REUnet in generating fine predicted masks. Along with segmenting target lesions of large and prominent shapes (shown in Fig. 8), REUnet outperforms other models in segmenting minute target lesions as well which is shown in white box. For example, in the 1st and the last rows of Fig. 7, its predicted masks not only locate the tumor in the breast precisely but also differentiates it from its surrounding tissue which has a similar intensity but is not tumor. Similar observation is seen in the 4th row where REUnet segments the hair line fracture present in the leg instead of getting bias towards the artifacts (plates) which was observed in case of other models. The incorporation of Gating Signal in DMBC further leads to a notable enhancement in the model's capacity to grasp interclass relationships. This observation is supported by the results shown in the 2nd and

Table 2
Comparison with SOTAs on breast MRI, brain MRI, Kidney CT, Bone X-ray and Breast Mammogram (Mamo) datasets. The best performance is highlighted in **bold**, while the second best is underscored. Standard deviations (multiplied by 10) are also reported alongside the mean scores across 5 runs, with deviations indicated in green.[31–35].

Method	MRI				CT		X-ray		Mamo	
	Breast		Brain		Kidney		Bone		Breast	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Attention Unet [8]	0.66 ± 0.12	0.70 ± 0.15	0.68 ± 0.1	0.72 ± 0.09	0.69 ± 0.08	0.73 ± 0.09	0.68 ± 0.14	0.72 ± 0.13	0.65 ± 0.19	0.70 ± 0.2
Unet++ [20]	0.80 ± 0.07	0.82 ± 0.06	0.75 ± 0.08	0.79 ± 0.08	0.73 ± 0.06	0.77 ± 0.05	0.67 ± 0.1	0.71 ± 0.08	0.64 ± 0.11	0.68 ± 0.1
DeepLabv3+ [31]	0.78 ± 0.1	0.81 ± 0.08	0.75 ± 0.11	0.80 ± 0.1	0.70 ± 0.17	0.74 ± 0.14	0.68 ± 0.17	0.71 ± 0.13	0.66 ± 0.2	0.71 ± 0.18
Unet 3+ [21]	0.83 ± 0.08	0.88 ± 0.06	0.80 ± 0.05	0.85 ± 0.06	0.76 ± 0.09	0.82 ± 0.07	0.77 ± 0.11	0.83 ± 0.09	0.75 ± 0.16	0.81 ± 0.13
DR-Unet104 [22]	0.79 ± 0.08	0.83 ± 0.06	0.80 ± 0.05	0.84 ± 0.06	0.74 ± 0.08	0.78 ± 0.07	0.72 ± 0.18	0.77 ± 0.16	0.70 ± 0.2	0.74 ± 0.18
Pspnet [32]	0.80 ± 0.07	0.85 ± 0.05	0.72 ± 0.08	0.78 ± 0.06	0.67 ± 0.14	0.72 ± 0.13	0.58 ± 0.23	0.62 ± 0.21	0.64 ± 0.21	0.69 ± 0.22
Transunet [9]	0.85 ± 0.04	0.90 ± 0.03	0.81 ± 0.03	0.86 ± 0.03	0.75 ± 0.06	0.80 ± 0.05	0.76 ± 0.1	0.81 ± 0.08	0.74 ± 0.08	0.79 ± 0.07
Swin-unet [10]	0.76 ± 0.12	0.82 ± 0.1	0.63 ± 0.24	0.69 ± 0.21	0.75 ± 0.06	0.81 ± 0.05	0.71 ± 0.11	0.76 ± 0.09	0.71 ± 0.09	0.75 ± 0.07
Mobile-Unet [33]	0.74 ± 0.13	0.77 ± 0.11	0.62 ± 0.19	0.65 ± 0.17	0.73 ± 0.06	0.77 ± 0.06	0.70 ± 0.09	0.74 ± 0.08	0.69 ± 0.07	0.73 ± 0.07
Mamba-UNet [34]	0.84 ± 0.03	0.88 ± 0.03	0.82 ± 0.03	0.86 ± 0.02	0.74 ± 0.07	0.79 ± 0.06	0.72 ± 0.07	0.75 ± 0.06	0.67 ± 0.08	0.71 ± 0.06
VM-UNet [35]	0.82 ± 0.03	0.85 ± 0.04	0.81 ± 0.04	0.85 ± 0.04	0.71 ± 0.08	0.74 ± 0.06	0.68 ± 0.1	0.71 ± 0.09	0.65 ± 0.08	0.68 ± 0.08
REUnet	**0.88** ± 0.03	**0.93** ± 0.02	**0.85** ± 0.02	**0.91** ± 0.02	**0.79** ± 0.04	**0.85** ± 0.03	**0.80** ± 0.04	**0.85** ± 0.05	**0.78** ± 0.09	**0.85** ± 0.07

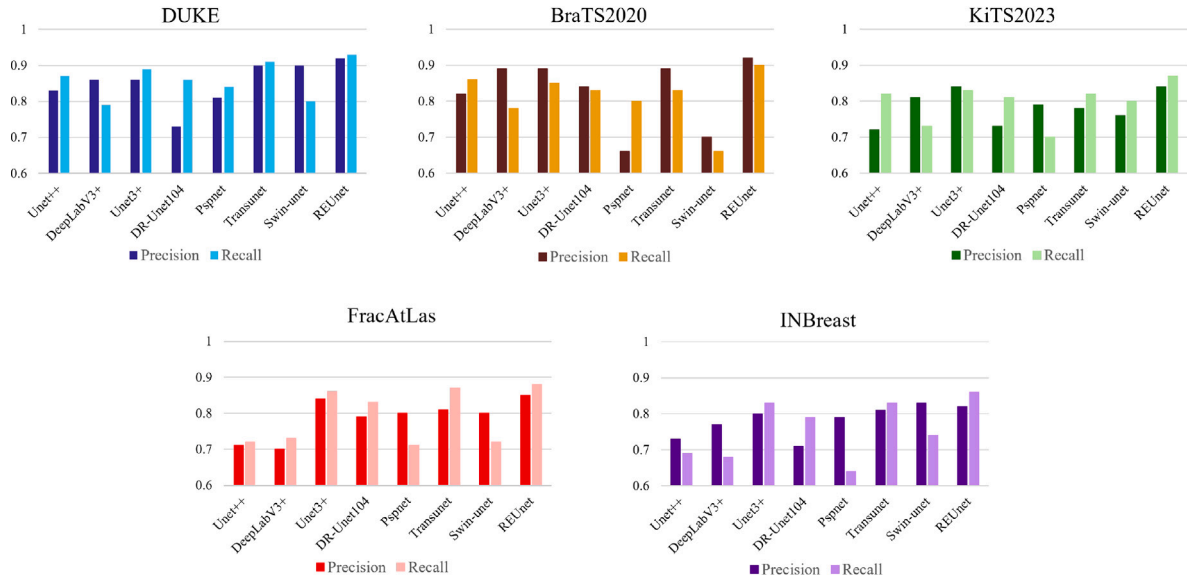

Fig. 5. Stacked plots showing the performance comparison of REUnet with other SOTAs in terms of mean precision and mean recall for DUKE, BraTS2020, KiTS2023, FracAtLas, and INBreast datasets respectively.

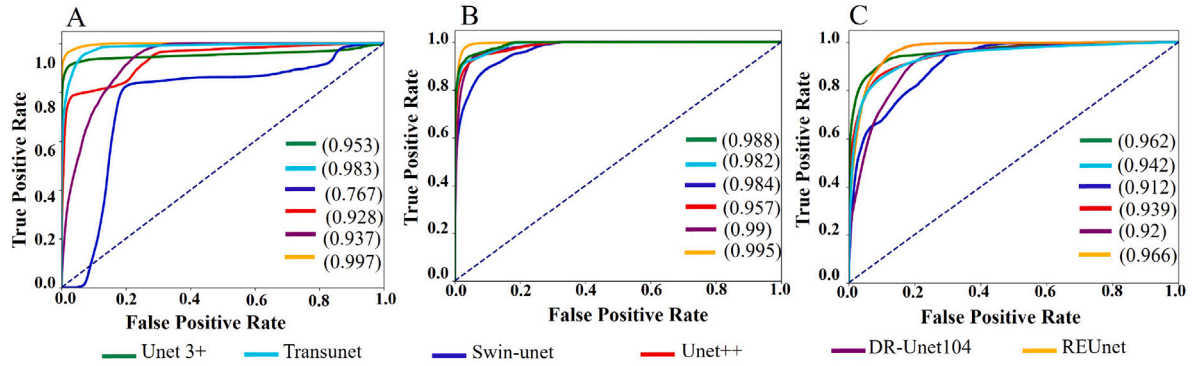

Fig. 6. ROC curve comparison of different segmentation models for (A) DUKE, (B) INBreast and, (C) FracAtLas. Values at the right side of each sub-figures represent the comparisons of AUC scores of REUnet with other models in that particular dataset.

3rd rows of Fig. 7, where REUnet is capable to separate multiple classes by their boundaries which leads to fine separation of multiple target lesions. However, this enhanced performance was not replicated by other methods, where the separation was irregular due to pixel misclassification.

Complexity Analysis. To highlight the potential of REUnet in clinical application, we compared its complexity with the above-mentioned models in terms of total number of parameters (in million), memory required for loading (in MB) and number of floating-point operations per second (FLOPS). We present the result in the form of bubble plot in Fig. 9(A) where each model is denoted by a unique circle of a distinct color. The circles were plot with respect to FLOPS (in x axis) and number of parameters (in y axis) whereas the memory size was denoted by the size of the bubble. From Fig. 9(A), it is evident that REUnet (orange bubble) exhibits significantly lower computational complexity

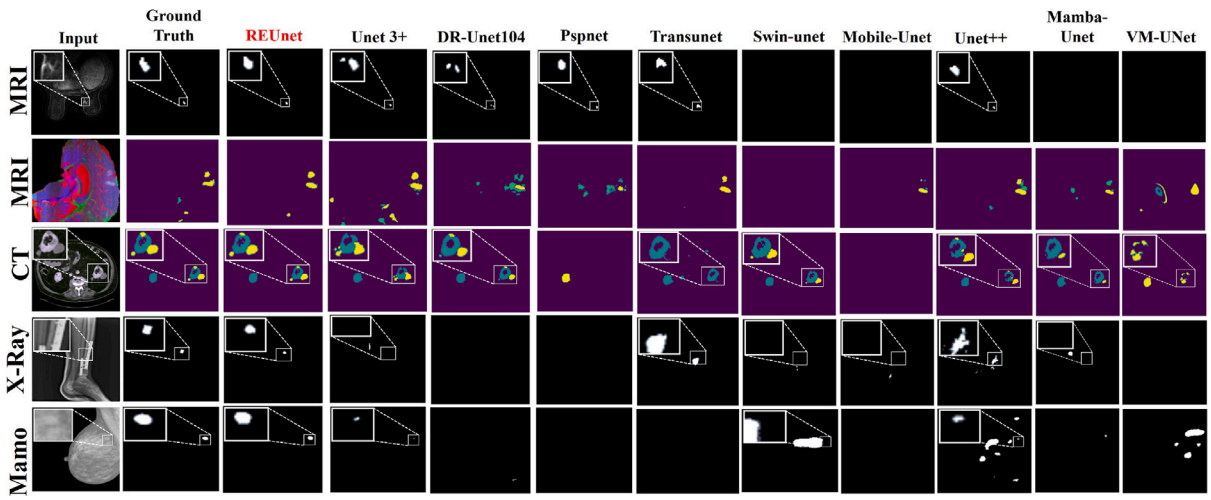

Fig. 7. Comparison of predicted masks generated by REUnet to those of other SOTAs. The comparison covers the following modalities, listed from top to bottom: MRI (DCE), MRI (stack of T1CE, T2 and FLAIR), CT, X-ray and Mammography. The white box displays a magnified view of a complex visual region.

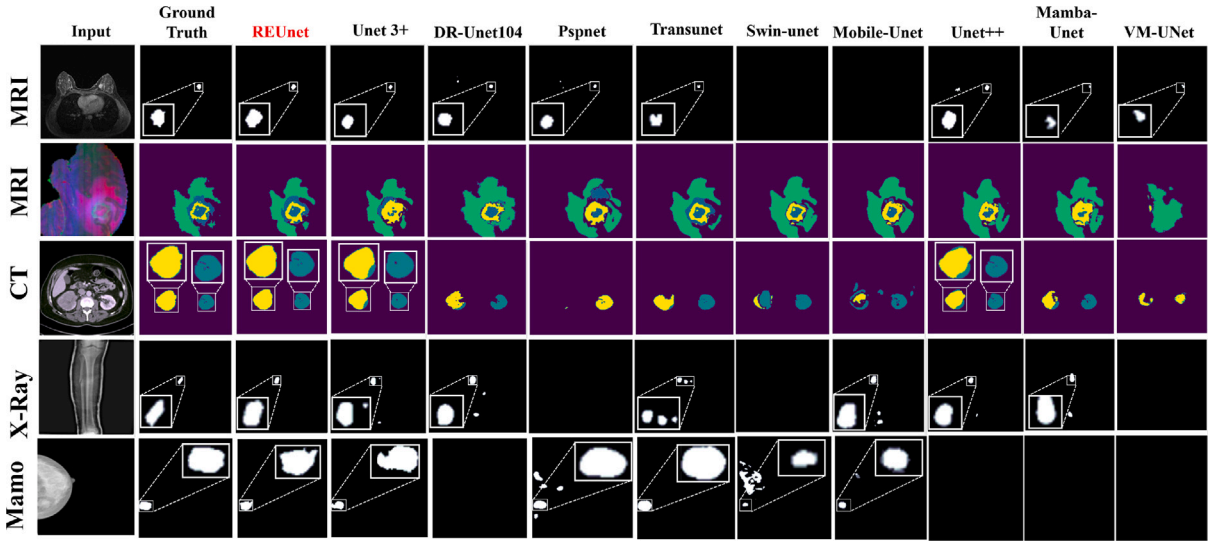

Fig. 8. Comparison of predicted masks generated by REUnet to those of other SOTAs for prominent sized target lesions. The comparison covers the following modalities, listed from top to bottom: MRI (DCE), MRI (stack of T1CE, T2 and FLAIR), CT, X-ray and Mammography. The white box displays a magnified view of a complex visual region.

compared to most of the SOTA methods. However, it should be noted that it falls slightly behind Mobile-Unet (light green bubble) in terms of memory size and number of parameters.

To further highlight the computational efficiency of REUnet over other models for deploying in different clinical scenarios, we analyze the processing speed of REUnet with other models in terms of frames per second (Fig. 9(B)) and processing time taken per image by the model (Fig. 9(C)). Both of the experiments were conducted separately on CPU and GPU (the specification of systems are provided in the implementation details). We conducted the first experiment on KiTS2023 dataset whereas for the second one, INBreast dataset was used. From both the sub-figures, it is clearly evident that REUnet not only outperforms all the mentioned models in terms of processing speed, but also exhibit quick inferencing in terms of processing time. Moreover, REUnet achieves the highest number of frames processed per second (6) and lowest time taken to process an image (167 ms) in CPU which highlights REUnet's ability to perform efficiently even in the absence of high performance computing systems.

Data Growth Analysis. We conducted a study on dataset magnitude's impact on model performance, illustrated in Fig. 10. REUnet's performance was compared with Transunet, Swin-unet, Unet 3+ and DR-Unet104 across varied data scales. Initially trained on 20% of the data and tested on the remaining 80%, we iteratively increased training data by 10% while decreasing testing data by 10% until reaching a 90% training and 10% testing configuration.

Fig. 10 demonstrates REUnet's superior scalability compared to other models as training data increases, showcasing its robust generalization even with limited training data and its ability to mitigate overfitting. These findings highlight REUnet's advantages, particularly in scenarios where less annotated data is available.

5. Ablation study

In this section, we analyze various components of REUnet through a series of ablation studies to evaluate their contribution in the full architecture. Despite of reporting the results on only one dataset, we used all the datasets (selected randomly) as REUnet is well generalized across them.

Effect of EBs. To analyze the impact of the encoding pathway, we replace each EB of REUnet with basic encoding layers of U-net. The visual illustration of this process is shown in Fig. 11, where the encoding block from EB-7 to EB-4 were successively replaced one by

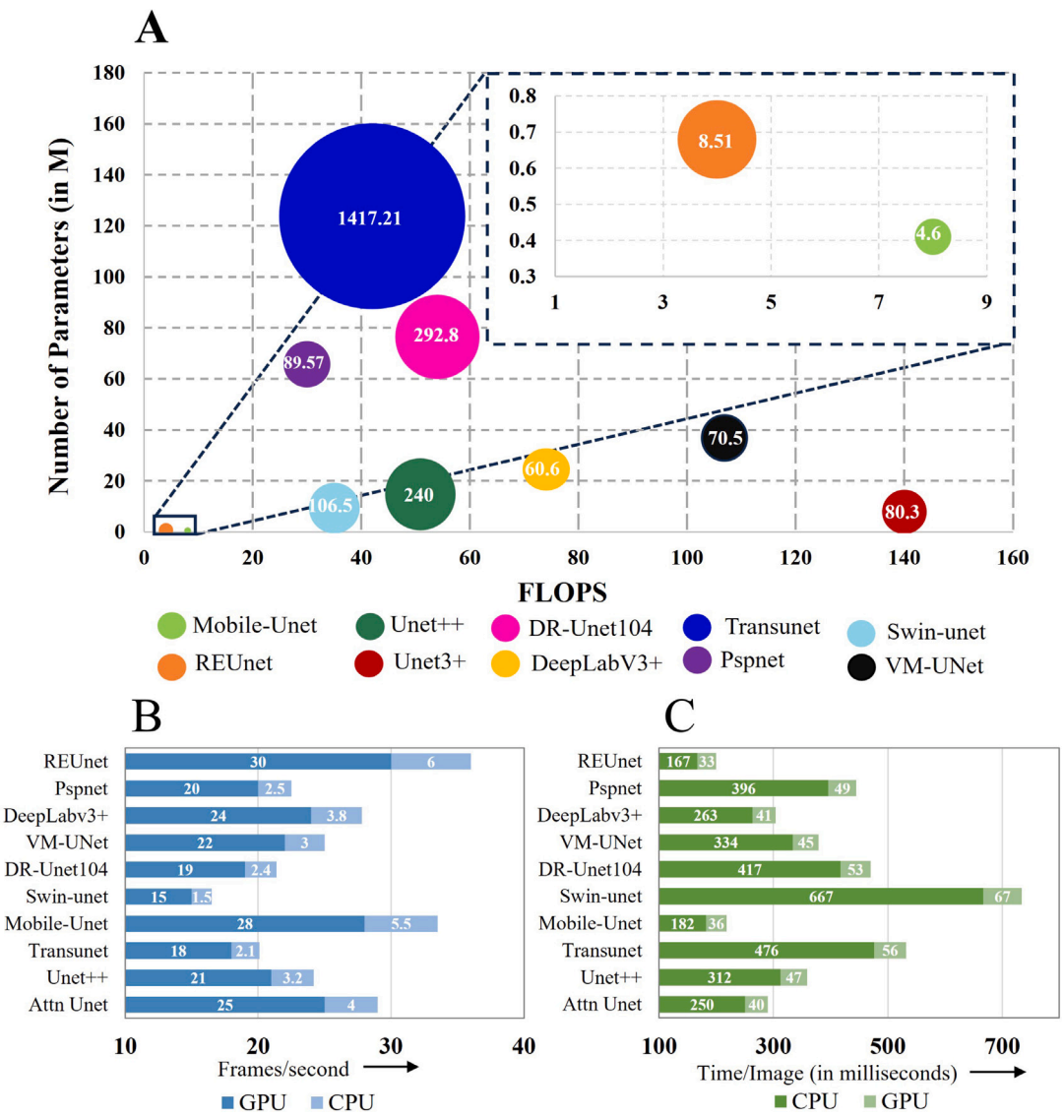

Fig. 9. Complexity analysis of REUnet compared to SOTA models. (A) Model comparison in terms of number of parameters (y axis), FLOPS (x axis) and, memory required for loading (size of the bubble). (B) Inference speed analysis showing frames per second (FPS) on both GPU (light blue) and CPU (dark blue). (C) Per-image inference time comparison on GPU (dark green) and CPU (light green).

one with basic encoding blocks, highlighted as yellow boxes. We limited replacements to EB-4 to assess the influence of deeper-level blocks responsible for capturing complex spatial information. We denoted these replaced variants as A1, A2, A3 and A4 respectively. Table 3 presents the performance of each variant across BraTS2020 classes. The table shows a decrease in performance from A1 to A4 as more EBs are replaced. This decline underscores the importance of EBs in extracting robust salient features, with their significance diminishing from A1 to A4. Hence, the role of these EBs emerges as pivotal in constructing a resilient encoding pathway.

Size of Intermediator. In Table 4, we analyze the influence of Intermediator size on REUnet's predictions for the DUKE dataset. The Intermediator plays a vital role in reinforcing the model's bottleneck region and provides higher level feature abstraction. This fact is validated in the mentioned table where the performance of REUnet gets boosted by a huge margin of 20% in Precision and Recall and by 23% and 25% in IoU and Dice, when the Intermediator (of size 13) is introduced in the architecture. Experimenting with Intermediator sizes ranging from 5 to 15 DMBC units, we find that the 13-unit configuration yields the best results. Decreasing Intermediator size adversely affects encoding, while increasing it results in model overfitting.

Table 3
The performance (mean IoU and mean Dice) scores of REUnet variants (A1, A2, A3, and A4) on each class of the BraTS2020 dataset.

Variants	NCR/NET		ET		ED	
	IoU	Dice	IoU	Dice	IoU	Dice
A1	0.78	0.84	0.89	0.94	0.82	0.90
A2	0.76	0.82	0.88	0.93	0.83	0.91
A3	0.75	0.81	0.85	0.90	0.79	0.87
A4	0.73	0.80	0.82	0.88	0.80	0.88

Influence of Attention Gate. We conduct an ablation study to assess the impact of incorporating attention gate in our model architecture. Table 5 shows the performance of the model with and without attention gates for FracAtLas dataset. As shown in Table 5, the mean Dice score increases by 3% when the attention gate is incorporated. This enhancement highlights the significance of the attention coefficients, which combine encoding and decoding feature maps to focus on specific regions of interest, thus improving model performance.

Impact of Gating Signal. The significance of Gating Signal in DMBC is showcased in Table 6 where the performance of REUnet was

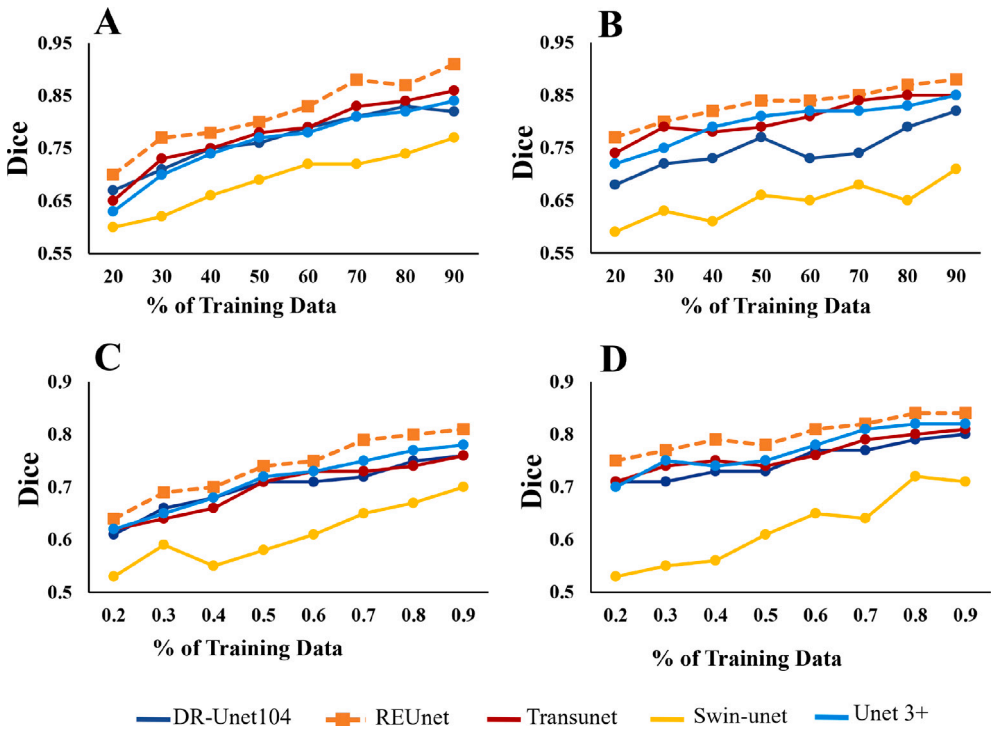

Fig. 10. Illustration of the performance of REUnet in the form of line plots under varying segments of training data for (A) DUKE, (B) BraTS2020, (C) INBreast and, D) FracAtLas. Here 0.2 means 20% of the data is used for training and so on.

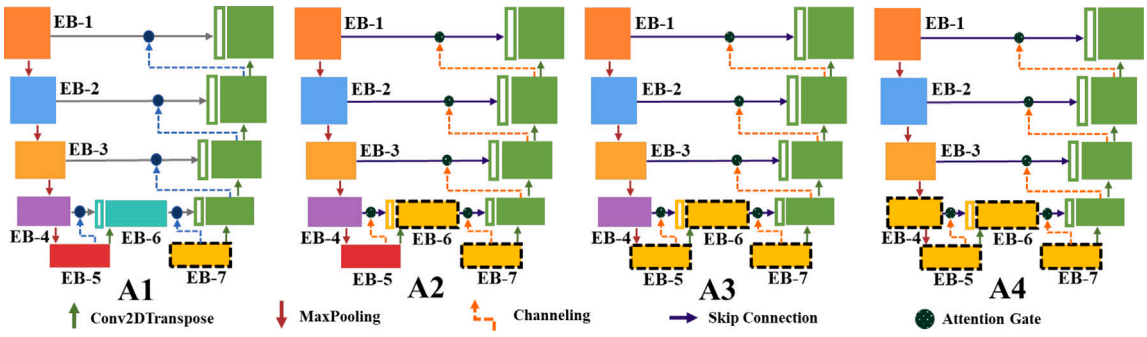

Fig. 11. Generation of REUnet Variants (**A1**, **A2**, **A3**, and **A4**) by sequential replacement of its encoding blocks (**EB7** to **EB4**) with basic encoding blocks.

Table 4
The performance of REUnet on DUKE dataset under different Intermediator sizes. Here size of 0 means REUnet without the Intermediate block.

No of DMBC	Mean precision	Mean recall	Mean IoU	Mean dice
0	0.73	0.68	0.59	0.64
5	0.89	0.78	0.67	0.78
7	0.90	0.88	0.80	0.87
9	0.91	0.82	0.76	0.84
11	0.89	0.89	0.80	0.88
13 (ours)	0.93	0.92	0.88	0.93
15	0.90	0.88	0.82	0.89

compared in its presence and absence for KiTS2023 dataset. From the table it is worth noting that the presence of Gating Signal in the DMBCs improves the performance of the model by a significant margin of 1%, 4%, and 7% in Dice score for kidney, tumor and cyst respectively.

This showcases its impact in segmenting target lesions of multiple types where it efficiently separates the boundaries of different target lesions from each other by reducing pixel misclassification. The reason behind this is the robust feature refinement that this signal offer when present in DMBC which helps the decoder to precisely segment the target lesions.

Fig. 12 provides a visual interpretation of this claim where we compare the outputs as well as the feature maps (after EB-7) in presence and absence of Gating Signal for two random test images from BraTS2020 and KiTS2023 datasets respectively. In both the sub-figures,

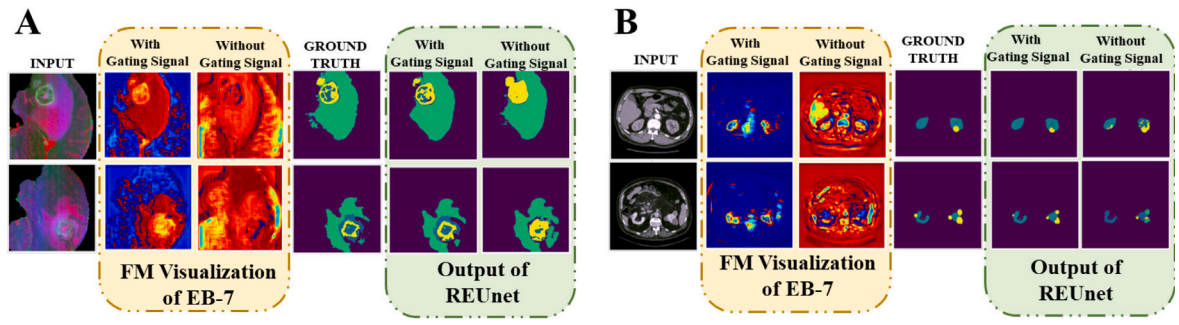

Fig. 12. Visual comparison of encoding feature map as well as the output of REUnet with and without Gating Signal for (A) BraTS2020 and, (B) KiTS2023 dataset respectively.

Table 5
The performance of REUnet with and without attention gates on FracAtLas dataset.

Attention gate	Mean precision	Mean recall	Mean IoU	Mean dice
With	0.91	0.84	0.80	0.85
Without	0.91	0.81	0.78	0.82

Table 6
The performance of REUnet for each class of KiTS2023 dataset in the presence and absence of gating signal.

Gating signal	Kidney		Tumor		Cyst	
	IoU	Dice	IoU	Dice	IoU	Dice
Present	0.89	0.93	0.80	0.84	0.71	0.76
Absent	0.88	0.92	0.75	0.80	0.63	0.69

the encoding feature map in presence of Gating Signal (2nd columns) extracts the most informative features that concentrates strongly on the regions of interest compared to the encoding feature map in absence of Gating Signal (3rd columns). This feature refinement enables the decoder to attain precision in segmentation which can be observed in its output (2nd last columns).

6. Conclusion

The paper introduces REUnet, a novel architecture designed for precise segmentation of complex target lesions across various medical imaging modalities. REUnet's strength lies in its robust encoding pathway, featuring the DMBC module. The inclusion of Gating Signals and a channel squeeze-and-excite block enhances and refines semantic information, facilitating accurate lesion localization by the decoder. The model is also computationally efficient due to the use of depthwise separable convolutions and dropout, making it well-suited for real-world applications. Extensive experiments on five benchmark datasets validate REUnet's superiority over other state-of-the-art models. Multiple ablation studies further highlight the effectiveness of various components in the overall architecture.

However, we acknowledge certain limitations of the proposed model. For instance, REUnet may not perform optimally when applied to imaging modalities that were not included during training. This could limit its direct applicability to unseen data distributions. A potential solution to this challenge is to fine-tune REUnet on the target modality, which can help it adapt to the specific characteristics of the new data.

Looking ahead, several future research directions can be explored to enhance REUnet's generalization and robustness. While REUnet demonstrates strong generalization ability, its performance on diverse and unseen distributions could be further improved by integrating self-supervised learning techniques followed by fine-tuning or by incorporating online-training strategies. Additionally, the incorporation of metaheuristic learning strategies or noise-aware loss functions during training may improve the model's resilience to noisy or imperfect data, which is often encountered in clinical scenarios.

It is our belief that REUnet signifies a significant step towards the goal of making medical image segmentation more precise, robust, and universally applicable.

CRediT authorship contribution statement

Snehashis Chakraborty: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Komal Kumar:** Writing – original draft, Visualization, Validation, Formal analysis. **Ankan Deria:** Writing – review & editing, Visualization, Validation, Formal analysis. **Dwarikanath Mahapatra:** Writing – review & editing, Supervision. **Behzad Bozorgtabar:** Writing – review & editing, Supervision. **Sudipta Roy:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

Author have no conflict of interest to declare.

Data availability

All the datasets that are used in this study are open sourced and are mentioned in Section 4.1.

References

[1] S. Chakraborty, K. Kumar, B.P. Reddy, T. Meena, S. Roy, An explainable AI based clinical assistance model for identifying patients with the onset of sepsis, in: 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science, IRI, IEEE, 2023, pp. 297–302.
[2] K. Kumar, B. Pailla, K. Tadepalli, S. Roy, Robust MSFM learning network for classification and weakly supervised localization, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2442–2451.
[3] S. Roy, T. Meena, S. Lim, Demystifying supervised learning in healthcare 4.0: a new reality of transforming diagnostic medicine. Diagnostics 12 (10): 2549, 2022.
[4] A. Rahman, J.M.J. Valanarasu, I. Hacihaliloglu, V.M. Patel, Ambiguous medical image segmentation using diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11536–11546.
[5] O. Ronneberger, P. Fischer, T. Brox, 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, Springer, 2015, pp. 234–241.
[6] Z. Zhou, M. Siddiquee, N. Tajbakhsh, J.U. Liang, A nested U-net architecture for medical image segmentation, 2018, arXiv preprint arXiv:1807.10165.
[7] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, 2018, arXiv preprint arXiv:1802.06955.
[8] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.
[9] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.
[10] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: European Conference on Computer Vision, Springer, 2022, pp. 205–218.

[11] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[12] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[13] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[14] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vis. 59 (2004) 167–181.
[15] E.S. Biratu, F. Schwenker, T.G. Debelee, S.R. Kebede, W.G. Negera, H.T. Molla, Enhanced region growing for brain tumor MR image segmentation, J. Imaging 7 (2) (2021) 22.
[16] Z. Kato, T.C. Pong, A Markov random field image segmentation model for color textured images, Image Vis. Comput. 24 (10) (2006) 1103–1114.
[17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[18] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[19] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
[20] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, 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, Springer, 2018, pp. 3–11.
[21] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.W. Chen, J. Wu, Unet 3+: A full-scale connected unet for medical image segmentation, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 1055–1059.
[22] J. Colman, L. Zhang, W. Duan, X. Ye, DR-unet104 for multimodal MRI brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 410–419.
[23] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, S. Escalera, Bi-directional ConvLSTM U-net with densley connected convolutions, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
[24] J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, Nat. Commun. 15 (1) (2024) 654.
[25] A. Dosovitskiy, An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
[26] A. Saha, M.R. Harowicz, L.J. Grimm, C.E. Kim, S.V. Ghate, R. Walsh, M.A. Mazurowski, A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features, Br. J. Cancer 119 (4) (2018) 508–516.
[27] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, C. Davatzikos, Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection, Cancer Imaging Arch. 286 (2017).
[28] N. Heller, F. Isensee, D. Trofimova, The KiTS21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase CT, 2023, arXiv:2307.01984.
[29] I.C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M.J. Cardoso, J.S. Cardoso, Inbreast: toward a full-field digital mammographic database, Academic Radiol. 19 (2) (2012) 236–248.
[30] I. Abedeen, M.A. Rahman, F.Z. Prottyasha, T. Ahmed, T.M. Chowdhury, S. Shatabda, FracAtlas: A dataset for fracture classification, localization and segmentation of musculoskeletal radiographs, Sci. Data 10 (1) (2023) 521.
[31] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.
[32] L. Yan, D. Liu, Q. Xiang, Y. Luo, T. Wang, D. Wu, H. Chen, Y. Zhang, Q. Li, PSP net-based automatic segmentation network model for prostate magnetic resonance imaging, Comput. Methods Programs Biomed. 207 (2021) 106211.
[33] J. Jing, Z. Wang, M. Rätsch, H. Zhang, Mobile-Unet: An efficient convolutional neural network for fabric defect detection, Text. Res. J. 92 (1–2) (2022) 30–42.
[34] Z. Wang, J.Q. Zheng, Y. Zhang, G. Cui, L. Li, Mamba-unet: Unet-like pure visual mamba for medical image segmentation, 2024, arXiv preprint arXiv:2402.05079.
[35] J. Ruan, J. Li, S. Xiang, Vm-unet: Vision mamba unet for medical image segmentation, 2024, arXiv preprint arXiv:2402.02491.