Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-026-01931-1

Deep Learning for Medical Ultrasound Image Segmentation: A Systematic Review of the Current Research

Oona Rainio[1] · Ehsan Roshan[1,2] · Seyed Mohammedreza Hosseini[1] · Rida Rehman[1,3] · Joanna Okenwa[1,3] · Riku Klén[1]

Received: 26 November 2025 / Revised: 18 February 2026 / Accepted: 11 March 2026

Abstract

Deep learning (DL) has enabled automated segmentation of ultrasound images, and due to the rapid development of DL models, we want to offer a comprehensive overview of the current state of research. Following PRISMA 2020 guidelines, we systematically selected and analyzed 296 recent scientific articles on DL-based ultrasound segmentation from the PubMed database. According to our results, the most common targets of DL-based ultrasound segmentation are breast tumors, organs, and cardiovascular structures. Other major application categories include orthopedics, thyroid nodules, obstetrics-gynecology, and oncology in general. Convolutional neural networks (CNNs) and especially U-shaped architectures have preserved their popularity, even though vision transformers (ViTs), CNN/ViT hybrids, and segment anything models have also become well-established within a few years of their release. The newer models are given significantly more data, but no association between the method type and the reported values of the evaluation metrics can be detected across several studies. Most common limitations of the current research include a lack of information on computational requirements and issues related to model performance evaluation. DL-based ultrasound segmentation is a quickly developing field, supported by increased use of ultrasound imaging, new public datasets, and methodological advancements.

Keywords Deep learning · Segmentation · Sonography · Ultrasound

✉ Oona Rainio
ormrai@utu.fi

Ehsan Roshan
seyedehsan.roshan@utu.fi

Seyed Mohammedreza Hosseini
seyed.m.hosseini@utu.fi

Rida Rehman
Rida.Rehman@abo.fi

Joanna Okenwa
Joanna.okenwa@abo.fi

Riku Klén
riku.klen@utu.fi

1 Turku PET Centre, University of Turku and Turku University Hospital, Turku, Finland

2 Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

3 Turku PET Centre, Åbo Akademi, Turku, Finland

Introduction

Ultrasound imaging, also known as ultrasonography or sonography, is an increasingly popular medical imaging technique [1]. With the use of high-frequency sound waves, it allows the visualization of the body's internal structures in a way that is cost-effective, non-invasive, and does not involve ionizing radiation [2]. Due to its real-time nature, ultrasound is often the first-line imaging option for the purposes of screening, diagnostics, treatment planning, surgical guidance, and follow-up monitoring, especially in the fields of gynecology, obstetrics, and emergency medicine [3].

One of the routine tasks of medical image analysis is image segmentation [1]. It means partitioning a digital image into different segments on the pixel level or, in case of a single region of interest, separating it from the background. In the context of ultrasound imaging, segmentation is used for potentially malignant lesions, internal organs, bones, joints, blood vessels, fetal structures, and other such targets, in order to collect information on their locations, allow automatic area or volume calculations, and enable computer-assisted diagnosis. While relatively simple in theory, ultrasound segmentation has its unique challenges due to the characteristic speckle texture of ultrasound images, anechoic shadows, possible imaging artifacts, and limited resolution.

To facilitate medical image processing, numerous deep learning (DL) approaches have been developed to perform fully or semi-automatic image segmentation. The automated segmentation of medical images via convolutional neural networks (CNNs) was studied already in the early 1990s [4] and, enabled by the increased computational power, access to digitized image datasets, and methodological innovations, DL-based segmentation has become a popular subject of study in the time since then. Notably advancements in the field include introduction of the lightweight segmentation architecture U-Net in 2015 [5], new explainable artificial intelligence (AI) solutions in segmentation via Grad-CAM in 2017 [6], refinement of earlier DeepLab architecture into more effective DeepLabV3 and DeepLabV3+ CNNs in 2017 [7] and 2018 [8], introduction of U-Net-based nnU-Net CNN in 2018 [9], adaptation of transformers and generative AI models into image segmentation tasks in early-2020s [10, 11], and more recent hybrid approaches between CNNs and vision transformers (ViTs) [12].

Due to the rapid progress on DL approaches, further research on ultrasound-based segmentation requires timely information on which of the new DL methods have already become established in the research on ultrasound images, how they compare with each other, and which methods could be improved. To answer these questions, we systematically review nearly 300 recent scientific articles about ultrasound image segmentation and analyze them in terms of application category, DL methodology, ultrasound data, and model evaluation. Our aim is to offer the reader a comprehensive overview of the current state of research on ultrasound image segmentation, identify research gaps, and assess potential future directions.

Article Selection and Methods of Analysis

To select articles for the review, we performed a PubMed search on October 18th, 2025. We only used two search words, ultrasound and segmentation, in order to obtain a wide selection of ultrasound segmentation studies. In particular, we did not include any DL-related keywords to avoid the results being biased in favor of a specific method. To limit our focus on the current research, we included only the results within the 18-month time period between 1.4.2024 and 30.9.2025, which yielded 670 initial results. The titles, abstracts, and full-text versions were screened to select relevant results according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) 2020 guidelines [13]. We excluded results without DL-based segmentation of medical, non-simulated ultrasound data on humans, leaving us with 296 articles. Our workflow is summarized in Fig. 1.

The included articles were categorized based on the application and the DL method. We also collected details on DL architectures, datasets, and evaluation metrics. Chi-squared tests, Mann–Whitney *U* tests, and correlation tests were performed with the standard level of significance of 5% to detect statistically significant association between variables.

Applications of Ultrasound Image Segmentation

Based on our review of 296 articles, we identified ten major application categories shown in Fig. 2a.

Breast Tumors

Breast tumor ultrasound is commonly used to evaluate potential tumors and guide biopsies, and DL methods can be applied to perform segmentation for both benign and malignant breast tumors. According to our analysis, the breast tumor segmentation was the most common application in the current research of DL-based ultrasound image segmentation: It was studied in 74 (25.0%) of the 296 reviewed articles. Out of these 74 articles, 49 articles (66.2%) were solely focused on breast tumor segmentation. Conversely, a DL method was applied both for breast tumor segmentation and other segmentation tasks in 25 articles, which was 33.8% of the 74 articles with breast tumor segmentation and 75.8% of the 33 articles with several segmentation tasks of multiple targets in different datasets.

Organ Segmentation

The task of organ segmentation is used to assist ultrasound image analysis in many different subfields of medicine. Thirty-two (10.8%) of the reviewed articles studied organ segmentation, and out of them, six (18.8%) were about multi-organ segmentation, six (18.8%) about prostate segmentation, five (15.6%) about kidney segmentation, and another five (15.6%) about liver segmentation.

Cardiology

In cardiology, ultrasound, or specifically echocardiography, is used to detect heart defects, fluid buildup, blood clots, and functional issues of the heart. Twenty-nine (9.8%) of the reviewed articles were related to the segmentation of cardiac structures. Fifteen (51.7%) of them focused on segmentation

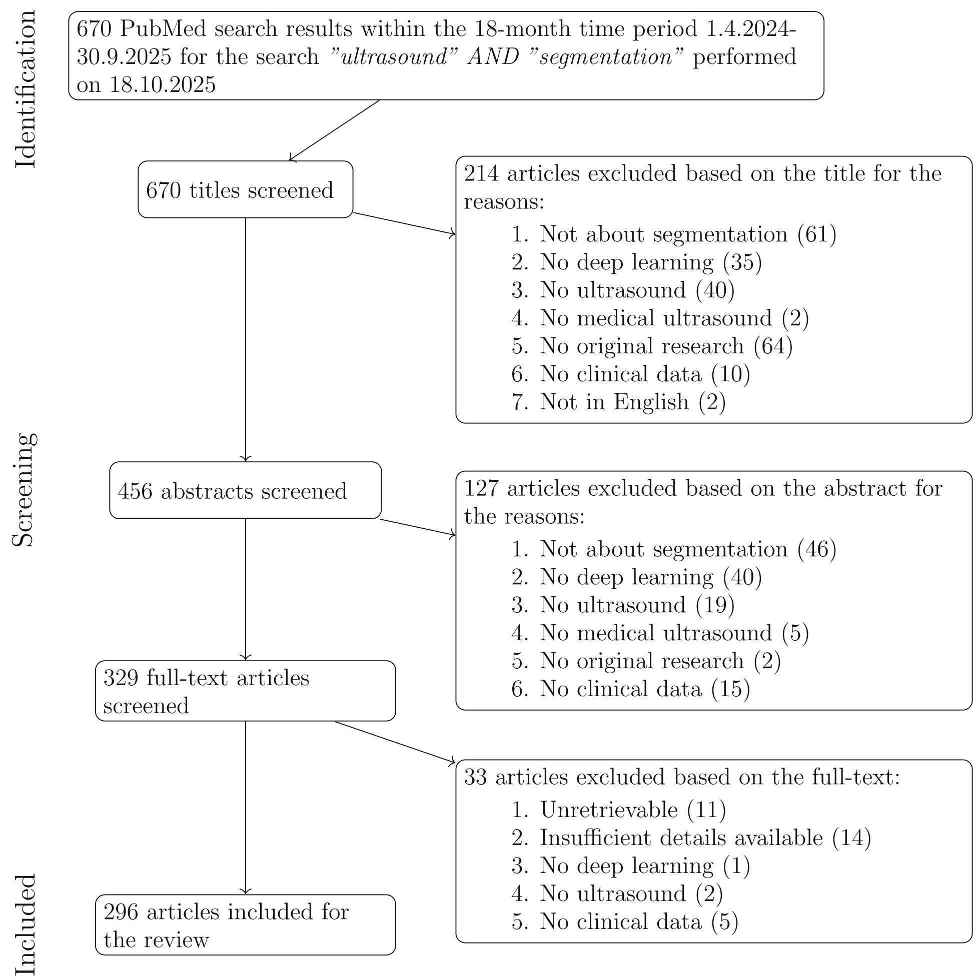

Fig. 1 The article selection process visualized as a flow diagram according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) 2020 guidelines

of the left ventricle, whereas the rest studied segmentation of all four chambers, left atrium, myocardium, or mitral regurgitation flow.

Angiology

Vascular ultrasound is used to study blood flow, identify possible clots, and gain structural information on arteries and veins. Twenty-nine (9.8%) of the reviewed articles were about the segmentation of vascular structures. Out of them, 12 (41.4%) articles were related to segmentation of blood vessels or their walls, nine (31.0%) to segmentation of atherosclerotic plaque, and four (13.8%) to segmentation of the lumen. Among the 21 studies on blood vessel or plaque segmentation, 14 (66.7%) were related to the segmentation of carotid arteries or carotid plaque.

Obstetrics

Ultrasound is used in obstetrics to confirm a pregnancy, check the position of the placenta, estimate the gestational age, monitor the development of the fetus, and check the baby's position before birth. Twenty-eight (9.5%) of the reviewed articles were related to obstetrics. A majority of them, 25 articles, were related to segmentation of fetal structures, such as head and brain, nuchal translucency, and heart and vessels. Additionally, there were three articles related to placenta segmentation.

Orthopedics

In orthopedics, ultrasound can be used to study different bones, joints, muscles, and tendons in order to detect fractures, bone erosion, tears, fluid buildup, and sprains. In 26 (8.8%) of the reviewed articles, the target of segmentation was either bone, joint, muscle, or musculoskeletal structure. Out of these articles, nine (34.6%) were related to the segmentation of bones, eight (30.8%) to joints, six (23.1%) to muscles, and the remaining three (10.3%) to musculoskeletal structures.

Thyroid Nodules

A thyroid nodule is a small lesion that forms in the thyroid gland and is typically benign and only rarely caused by pap-

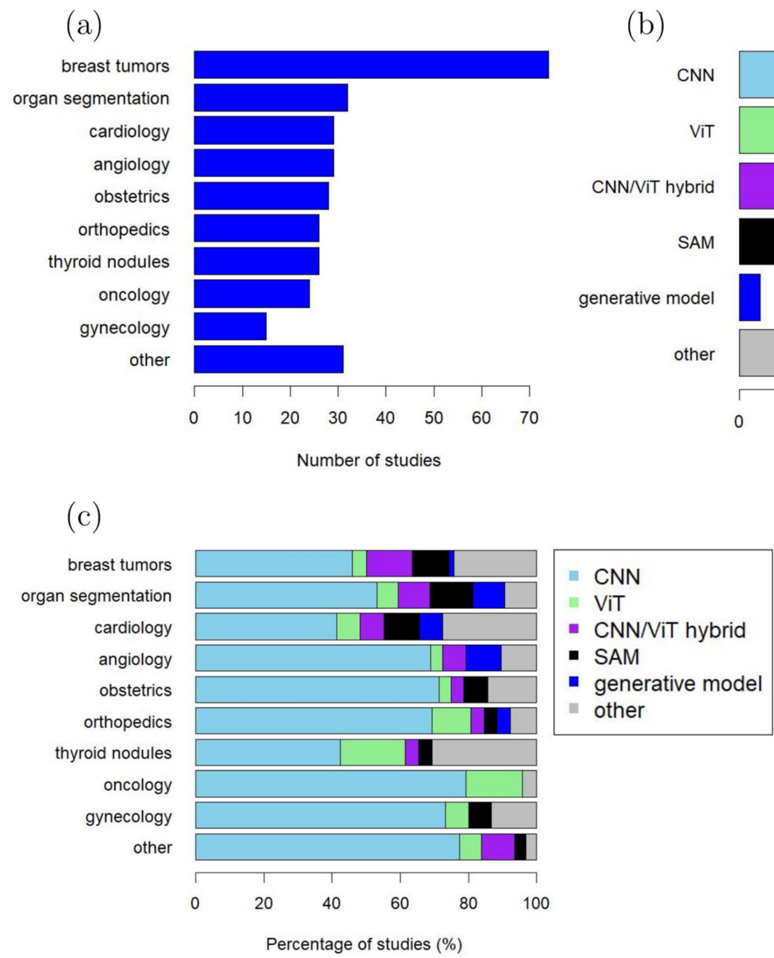

Fig. 2 The number of studies within different categories based on **(a)** the application and **(b)** the deep learning method, and **(c)** the distribution of studies between different deep learning methods for each application category. Note that there is overlap between certain application categories due to the presence of studies on gynecological oncology or on both breast tumor and thyroid nodule segmentation

illary cancer. An ultrasound is used to characterize thyroid nodules, and thyroid nodule segmentation via DL can be used to assist in the related image analysis. Thyroid nodule segmentation was studied in 26 (8.8%) of the reviewed articles, and out of them, 14 (53.8%) were solely focused on thyroid nodules. Among the 33 articles with multiple segmentation tasks, 12 (36.4%) included thyroid nodule segmentation as one task.

Oncology

Ultrasound is used to study different cancers. If not counting studies on breast tumors and thyroid nodules, most of which are benign, there were 21 different studies on segmentation of cancer or potentially cancerous tumors or lesions, making up 7.1% of the reviewed articles. Out of these 21 articles, six (28.6%) were about adnexal, ovarian, or uterine tumors, four (19.0%) about tumors in abdominal organs (pancreas, liver, gastrointestinal tract), four (19.0%) about prostate or colorectal tumors, and three (14.3%) about brain tumors.

Gynecology

In gynecology, ultrasound is a standard diagnostic tool for possible diseases affecting the cervix, uterus, ovaries, and fallopian tubes. The segmentation tasks in 15 (5.1%) of the reviewed articles were related to gynecology. A majority of them (nine articles, 60.0%) were about segmentation of potentially cancerous lesions, overlapping our previous category. The remaining six articles were about the segmentation of ovarian cysts, uterine layers, or pelvic floor structures.

Other

Additionally, there were 31 studies outside the aforementioned nine categories, making up 10.5% of the reviewed articles. They included segmentation of nerves (in five articles), brain structures (in four articles), lymph nodes (in four articles), and surgical needles (in one article), for instance.

Deep Learning Methods

The DL methods in the reviewed articles could be divided into the following six categories, also shown in Fig. 2b.

Convolutional Neural Networks

A CNN is an artificial neural network specifically designed to process image data in a way that accounts for the spatial relationships between adjacent image points. An image is first given to a CNN as a matrix whose elements correspond to the pixels of the image, and by utilizing matrix convolutions with different filters, the CNN is able to detect small features such as edges or certain textures. With suitable training data, the parameters within the CNN are optimized to focus on relevant features for the task of interest, and based on the extracted features, they produce desired outputs by minimizing the value of a loss function. In case of medical image segmentation, these outputs are typically binary matrices showing the regions of interest.

In 2015, Long et al. [14] introduced a fully convolutional network (FCN) that, unlike earlier CNNs, only used convolutional, pooling, and upsampling layers instead of fully connected layers, allowing the model to produce pixel-wise predictions for accurate segmentation. The idea of FCN was refined in the U-Net architecture by Ronneberger et al. [5], which included a symmetric encoder-decoder structure with skip connections. When an image is given to the U-Net, the encoder first performs the feature extraction requiring downsampling, after which the decoder performs segmentation while also gradually returning the data matrix to its original size, as shown in Fig. 3. The skip connections of the U-Net allow the direct transfer of data between encoder and decoder layers to prevent loss of information in the downsampling process. Containing only convolutions, maximum pooling operations, and transpose convolutions, the U-Net is a very lightweight design.

In the decade since the U-Net's introduction in 2015, it has become very popular, and several architectures have been developed based on it. The well-known modifications and versions of U-Net include Attention U-Net by Oktay et al. [15], UNet++ by Zhou et al. [16], residual U-Net (ResU-Net) by Zhang et al. [17], and nnU-Net by Isensee et al. [18]. Other popular CNN architectures besides U-Net and its modifications include the DeepLab architectures, in particular DeepLabV3 [7] and DeepLabV3+ [8].

A total of 175 (59.1%) of the 296 reviewed articles studied a CNN for ultrasound image segmentation. Among 166 articles where more details could be retrieved from the CNN architecture, 108 articles (65.1%) introduced a novel method or a new modification of an established method while the other 58 articles (34.9%) proposed the use of an existing method. In the 108 articles proposing a novel method, a great majority (93 articles, 86.1%) introduced a new CNN based on the U-Net architecture. In the 58 articles using well-established CNNs, the most popular designs were U-Net (in 23 articles), nnU-Net (in 12 articles), ResU-Net (in 4 arti-

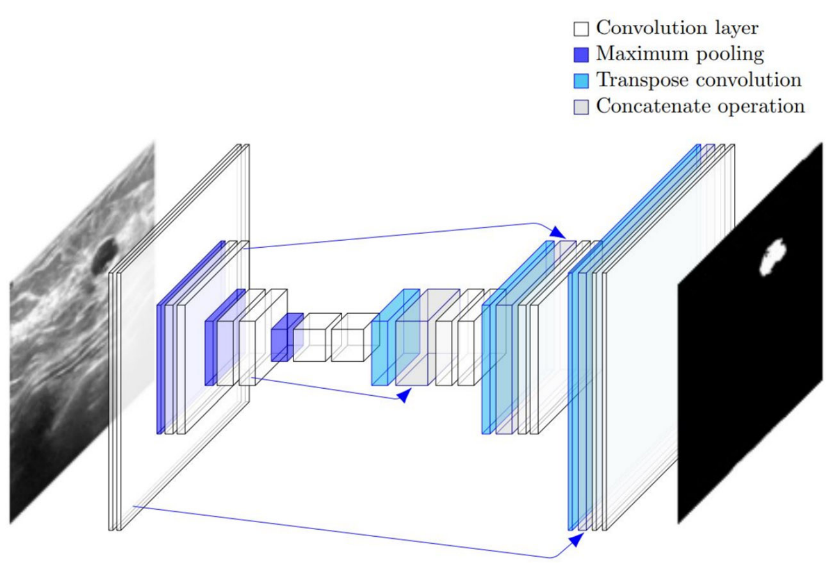

Fig. 3 The architecture of the U-Net convolutional neural network illustrated. The process starts on the ultrasound image (on the left) as an input and ends with the binary segmentation mask as an output. The skip connections are depicted by arrows. The ultrasound image is a breast ultrasound image from the public Breast Ultrasound Images (BUSI) dataset

cles), Attention U-Net (in 3 articles), FCN (in 3 articles), and DeepLabV3+ (in 3 articles). In total, at least 143, equivalent to 48.3% of the 296 reviewed articles and 81.7% of the 175 articles studying a CNN, proposed some U-shaped architecture, meaning either U-Net, its well-established modification such as nnU-Net or U-Net++, or a novel U-Net modification.

By comparing the distribution of the studies utilizing a CNN in the nine application categories, it was observed that the use of a CNN was most popular in oncology, gynecology, and obstetrics, where around 75% of the studies used a CNN for segmentation. Conversely, only 41.4% of the cardiac studies used a CNN. See Fig. 2c. According to a chi-squared test, the differences in the use of a CNN between the different application categories were statistically significant (p-value, 0.0019). For the chi-squared test, the overlap in application categories caused by studies on both breast tumor and thyroid nodule segmentation or topics related to gynecological oncology was solved by dividing the two overlapping categories into three new categories.

Vision Transformers

A transformer is a DL model characterized by its multi-head attention mechanism that allows the model to focus on the relevant parts of the input and process them in parallel with each other. In 2017, Vaswani et al. [19] introduced the modern version of the transformer with an encoder-decoder structure, leading to great success in the field of natural language processing. The first ViT was created three years later when Dosovitskiy et al. [20] adapted a transformer to perform computer vision tasks. A ViT decomposes an image into a grid of patches and utilizes a self-attention mechanism to capture both global and local features. ViTs offer an alternative to CNNs in computer vision tasks, but their use often requires higher computational resources and larger datasets [21].

In 2021, Liu et al. [22] introduced Swin Transformer, a new hierarchical ViT architecture utilizing shifted windows in its self-attention mechanism. Swin Transformer could perform a wide range of computer vision tasks, including image segmentation. Later in the same year, Xie et al. [23] introduced SegFormer, a new lightweight transformer with multilayer perception decoders specifically designed for image segmentation. In 2022, Cao et al. [24] introduced Swin-Unet, which was a U-shaped encoder-decoder model built with a pure Swin Transformer architecture.

Twenty-two (7.4%) of the reviewed articles performed ultrasound image segmentation via a ViT. The most popular ViT designs were Swin-Unet (in six articles), SegFormer (in five articles), and Swin Transformer (in five articles). Nineteen (86.4%) of the 22 articles proposing a ViT introduced a new version of an existing ViT architecture whereas the remaining three utilized the existing models with no further modifications. The studies with ViTs were quite evenly distributed between different application categories with thyroid nodule segmentation as the most popular application.

CNN/ViT Hybrids

A CNN/ViT hybrid is a DL model formed as a combination of a CNN and a ViT. The aim is to benefit from the CNN's ability to find local patterns in limited datasets and the ViT's higher capacity to utilize global context and generalize findings across diverse datasets. The hybrid models can be created as two-step processes or by integrating the parts of one model into the other. For instance, TransUNet by Chen et al. [25] is a ViT receiving its image patches from U-Net-produced feature maps.

There were 22 different studies proposing the use of a CNN/ViT hybrid for the segmentation task, making up 7.4% of the reviewed articles. Out of these 22 studies, six (27.2%) performed segmentation via a modified version of TransUNet. Sixteen (72.7%) of the 22 studies with a CNN/ViT hybrid used U-Net as the CNN part of model, and nine (40.9%) were related to breast tumor segmentation.

Segment Anything Model

Segment anything model (SAM) is a vision foundation model introduced in 2023 by Kirillov et al. [26]. Foundation models are first pre-trained on broad datasets to perform generalizable tasks and later assigned a specific task by the user via prompt engineering. They had become established in the field of natural language processing, and SAM was an adaptation of their idea into segmentation. Trained with over a billion masks for 11 million images, SAM can perform segmentation of diverse targets, and the user can specify the desired region of interest with a prompt that can be either a bounding box, a robust segmentation mask, or a few points clicked on the desired target [26]. A new, more accurate and faster version of SAM called SAM 2 was released in 2024 by Ravi et al. [27].

Twenty (6.8%) of the reviewed articles used SAM or SAM 2 to perform segmentation. Out of these 20 articles, 13 (65.0%) used a new, modified version of SAM. SAM 2 was used in only two articles, though it was released only 14 months prior to our PubMed search to select the reviewed articles. Eight (40.0%) of the articles with SAM studied breast tumor segmentation.

Generative Models

A generative model is a DL model that learns the underlying patterns of an existing dataset and produces new instances similar to the original data. One of the major types of generative models is a generative adversarial network (GAN), originally introduced in 2014 by Goodfellow et al. [28]. It consists of two neural networks performing opposite tasks: A generator aims to create synthetic images that cannot be differentiated from the real images, and a discriminator tries to separate the real and the synthetic images. During the training process, the performance of both networks is optimized with two competing loss functions, giving the user highly realistic synthetic images. As an alternative to GANs, Sohl-Dickstein et al. [29] introduced a diffusion model, which produces new images from pure noise after it has been trained to gradually increase random noise and perform denoising for a set of original images.

While the original aim of both GANs and diffusion models was simply to create new instances resembling an existing dataset as a whole, both model types were later used to build conditional models that could generate a new instance based on a specific request, such as a class label or a textual prompt. This idea was further developed into image-to-image generation [30], enabling a new way to perform image segmentation. When given a specific input image, a GAN or a diffusion model can generate a segmentation mask as a synthetic image that should be indistinguishable from the ground-truth segmentation mask of this given image.

Only 10 (3.4%) of the reviewed articles proposed either a generative model or a hybrid between a generative model and another DL model. A diffusion model was used in four studies and a GAN in three studies, and the remaining three studies introduced a diffusion model/state space model (SSM) hybrid [31], a GAN/ViT hybrid [32], and a GAN/CNN hybrid [33]. The application categories of the generative models were varied, including segmentation of breast tumors, carotid plaque, and bone structures.

Other

Forty-seven (15.9%) used a DL method outside the previous categories. These articles formed a relatively heterogeneous collection of various DL methods, including neural cellular automata [34], a novel version of a Kolmogorov-Arnorld network [35], and hybrid approaches between a CNN and an extended long short-term memory model [36], neural ordinary differential equations [37], or a SSM [38].

Data

Types of Ultrasound Imaging

Modern ultrasound imaging is based on sending ultrasound pulses into the body and recording their echo [2]. The ultrasound is transmitted and recorded via an ultrasonic transducer inside an ultrasound probe. Due to the differences in the reflection properties between tissues, the data collected by the probe can be used to create an image of the internal structures of the body. Ultrasound imaging can be categorized by the mode used, the most common of which is the B-mode. In B-mode, an array of transducers simultaneously scans a plane of the body, resulting in a two-dimensional (2D) image [2]. The B-mode ultrasound studies were also the most common mode in the reviewed articles, though some studies also specified using Doppler data [39, 40].

Ultrasound can also be used to create videos or three-dimensional (3D) images. There were 30 articles performing segmentation for 3D ultrasound images and 18 performing segmentation for ultrasound videos, making up together 16.2% of the 296 reviewed articles. A majority of the studies involving either 3D images or videos utilized a CNN as the method for segmentation.

The ultrasound probe is typically placed outside the body against the skin, but internal ultrasound studies can also be performed via an endocavitary probe. The use of an endocavitary ultrasound data was specified in 12 (4.1%) of the reviewed articles. This included five studies with transrectal ultrasound, three studies with transvaginal ultrasound, three with endoscopic ultrasound, and one with intraoral ultrasound, most related to either organ segmentation or oncology.

Public Datasets

Out of the 296 reviewed articles, 158 articles (53.4%) used a private dataset, 114 articles (38.5%) used an existing public dataset, 22 articles (7.4%) used both public and private datasets, and two articles [32, 51] published their newly collected dataset alongside the results of the DL-based segmentation. In total, we identified over 30 different ultrasound image datasets, 10 of which are the most popular, and are summarized in Table 1. Example ultrasound images from public datasets can be viewed from Fig. 4. As shown in Fig. 5a, the use of public data was most common within breast tumor segmentation studies, where there were over three times more studies with public data than there were with private data. Conversely, less than a fourth of the segmentation studies related to orthopedics and angiology utilized public data. The difference in the use of public data between the different application categories was statistically significant according to a chi-squared test (p-value, 1.5e−15).

Number of Data Instances

The number of data instances (either ultrasound images or videos) ranged from 26 in a study by Strohm et al. [52] to 282,000 in a study by Meyer et al. [53] with a mean and standard deviation (SD) of 4932±18,352 and a median of 1098. See Fig. 5b. According to a Mann–Whitney U test, the studies utilizing public datasets had significantly more data instances than the studies with private data (p-value, 8.8e−5), and the studies with ViTs and CNN/ViT hybrids

Table 1 A reference, name, and description for the most common public datasets in the reviewed articles on ultrasound image segmentation. See examples from Fig. 4

Ref	Name	Description
Al-Dhabyani et al. [41]	BUSI	780 breast ultrasound images of 600 female patients (487 with benign tumors, 210 with malignant tumors, 133 normal)
Yap et al. [42]	UDIAT/Dataset B	163 ultrasound breast images with lesions (53 malignant, 110 benign)
Piotrzkowska-Wróblewska et al. [43]	OASBUD	Breast ultrasound images from 78 patients with 100 clinically confirmed lesions
Pedraza et al. [44]	DDTI	347 thyroid ultrasound images of 299 patients with thyroid disorders
Gong et al. [45]	TN3K	3493 annotated thyroid ultrasound images with thyroid nodules
Leclerc et al. [46]	CAMUS	2D two- and four-chamber view echocardiography sequences from 500 patients
Bernard et al. [47]	CETUS	45 3D ultrasound images of one cardiac cycle
Ouyang et al. [48]	EchoNet-Dynamic	10,030 annotated echocardiogram videos
Singla et al. [49]	Open Kidney US Dataset	514 2D B-mode abdominal ultrasound images from 514 patients
Baum et al. [50]	μ-RegPro	108 3D transrectal ultrasound volumes of prostate

(p-value, 0.031) included significantly more data instances than the studies with CNNs. Notably, we did not account for the potential confounding factors in this comparison, so the identified significant differences might be caused by differences in the application categories or the exact tasks of segmentation.

Pre-processing and Augmentation

The 2D ultrasound images were often resized into 128×128, 256×256, or 512×512 pixels, especially if a CNN was used [32, 54, 55]. Individual frames were extracted from the video data and processed as static 2D images [56], whereas

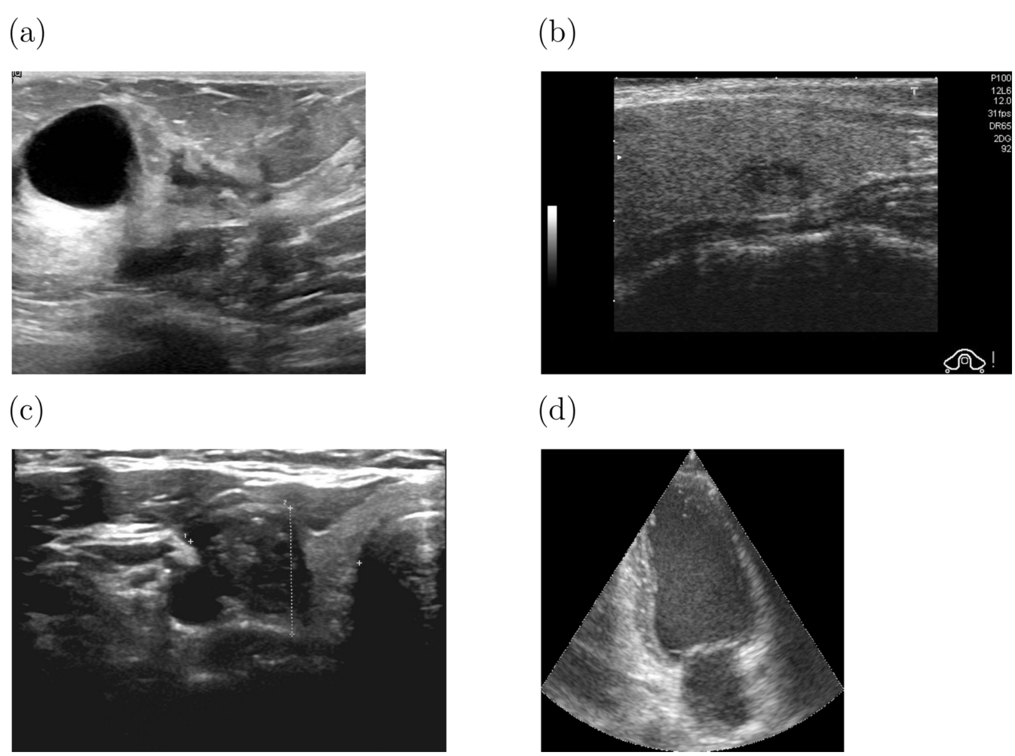

Fig. 4 Example ultrasound images from four public datasets of Table 1, including **(a)** a breast image with a benign tumor from Breast Ultrasound Images (BUSI), **(b)** a thyroid nodule image from Digital Database Thyroid Image (DDTI), **(c)** a thyroid nodule image from thyroid nodule segmentation dataset (TN3K), and **(d)** an echocardiography image from Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS)

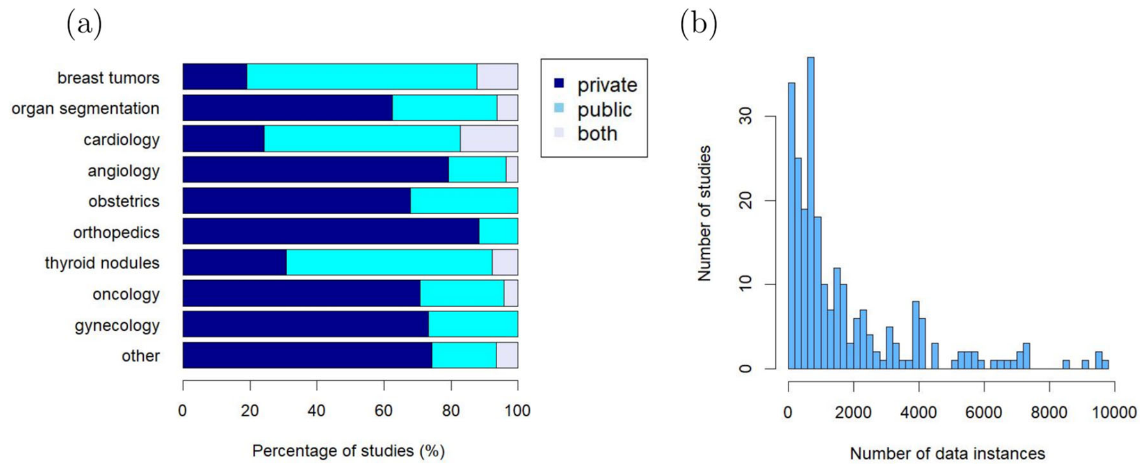

Fig. 5 **(a)** The distribution of studies using private data, public data, or both for each application category and **(b)** the histogram showing the number of studies using a certain number of data instances. There were 30 studies with more than 10,000 data instances not visible in this histogram

the 3D volumes were either divided into transaxial slices [57] or processed with 3D DL methods [58]. Either no data augmentation was used or only simple augmentation transformations, such as rotations less than 10 or 15° [59, 60], vertical and horizontal flips [60–62], zooms [59, 60], translations [60, 62], intensity scaling [62], and Gaussian blurring [59, 61], were utilized.

Evaluation

Quantitative estimation of the predicted segmentation masks requires evaluation metrics. Two common metrics, the Dice score and the Intersection over Union (IoU), both measure a relative overlap between segments annotated by a human expert and predicted by the DL method. These metrics can be defined as

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|}; \quad \text{IoU} = \frac{|X \cap Y|}{|X \cup Y|}, \tag{6.1}$$

where X denotes the set of image points within the ground-truth segment, Y is the set of points within the predicted segment, $|S|$ is the number of points within a set S, $\cap$ is the set intersection, and $\cup$ is the set union. Notably, each image point is either positive or negative based on whether it belongs to the ground-truth segment or not, and true or false based on whether it was predicted correctly by the method of interest. Consequently, we can also write the formula above as

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}}; \quad \text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}, \tag{6.2}$$

where TP is the number of true positive points, FN the number of false negative points, and FP the number of false positive points. The values of both Dice and IoU range between 0 and 1, so that 0 means no overlap and 1 is the perfect overlap [63].

A third possible evaluation metric commonly used in medical image segmentation is the Hausdorff distance. Formally defined as [63]

$$\text{HD} = \max\left\{\max_{x \in X} d(x, Y), \max_{y \in Y} d(y, X)\right\}, \tag{6.3}$$

it means the longest possible distance from a point of one point set (either the ground-truth or the predicted segment) to the closest point of the other set. The Hausdorff distance can be expressed in image points or converted into millimeters.

The most common evaluation metric in the reviewed studies was the Dice score, reported in 252 articles (85.1%). Other used evaluation metrics included IoU, Hausdorff distance, average symmetric surface distance, pixel-level accuracy, and mean average precision. The reported mean Dice scores ranged from 0.426 in an organ segmentation study by Hamedi et al. [64] to 0.991 in a fetal head segmentation study by Degala et al. [65]. The reported mean Dice scores had a mean and SD 0.862±0.097 and a median of 0.887. Figure 6 shows the distribution of Dice scores, and Table 2 summarizes seven example studies. According to Mann–Whitney U tests, there were no significant differences in Dice scores between studies with public and private data (p-value, 0.959), studies with CNNs and either ViTs or CNN/ViT hybrids (p-value, 0.112), nor studies with CNNs and SAMs (p-value, 0.198). There was also no statistically significant association between the number of data instances and the reported mean Dice scores (Spearman's correlation coefficient of -0.00991 with a p-value of 0.879).

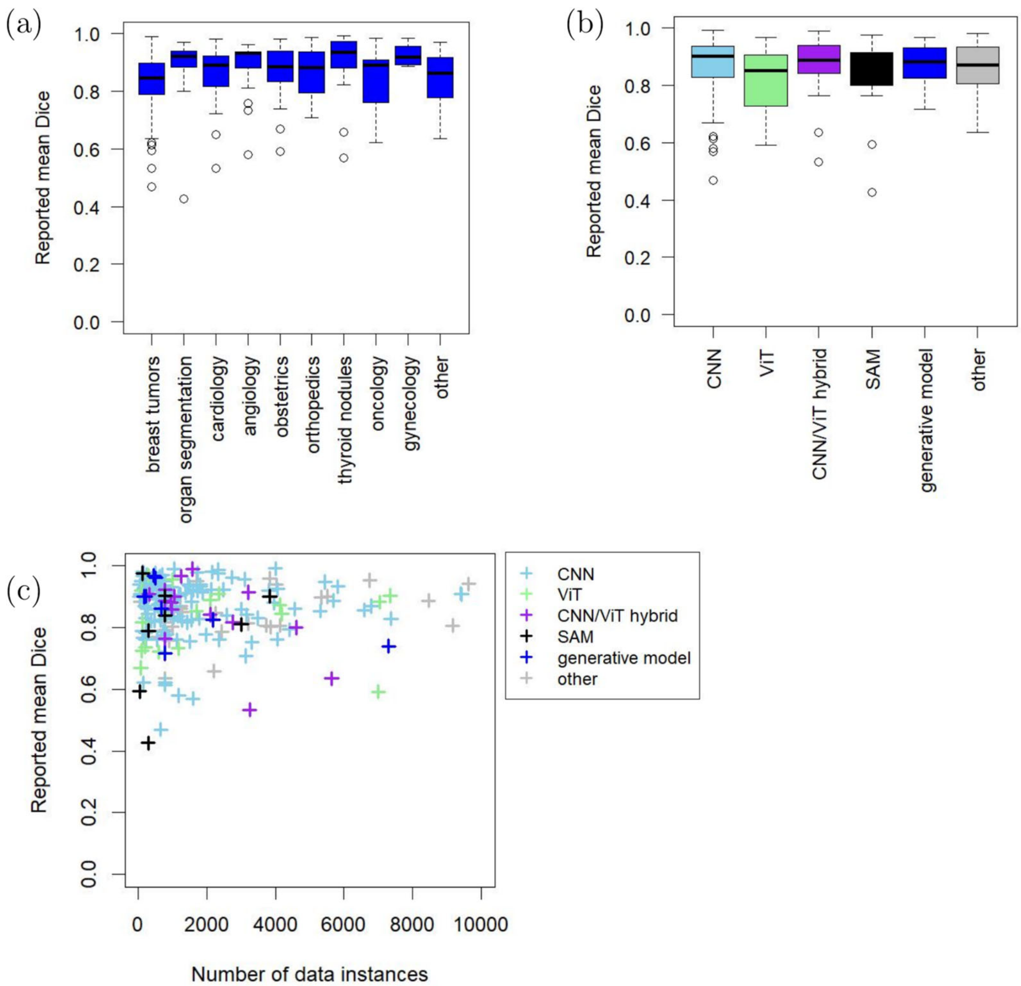

Fig. 6 The boxplots of reported mean Dice scores within **(a)** application and **(b)** deep learning method categories, and **(c)** a scatter plot showing the reported mean Dice scores against the number of data instances. There were 30 studies with more than 10,000 data instances not visible in this scatter plot

Table 2 Examples of studies summarized in terms of the target of segmentation, the deep learning method, the used data, and the reported evaluation metrics. These examples were chosen to present different application categories among studies that had good segmentation performance based on the reported Dice scores

Study	Target	Method	Data	Evaluation
Madhu et al. [55]	Breast tumors	Modified U-Net (UCapsNet)	780 images from BUSI	Dice: 95.14%, IoU: 94.22%
Karimi et al. [32]	Spleen	GAN/ViT hybrid with Pix2Pix and SegFormerB0	450 images from the newly assembled Spleenex dataset	Dice: 96.82%, IoU: 94.17%
Hu et al. [66]	Thyroid nodules	CNN/SSM hybrid with ResNet-34 and Mamba	5,511 images from TN3K, TNUI-2021, and DDTI	Dice: 90.02%, IoU: 81.85% on TN3K
Shi et al. [67]	Left ventricle	CNN/ViT hybrid with ResNet-50	20,060 video frames from EchoNet-Dynamic	Dice: 92.4%
Chen et al. [68]	blood vessels	ResU-Net	570 ultrasound images from 152 patients	Dice: c. 98%
Degala et al. [65]	Fetal head	Modified U-Net (ADU-GD)	999 ultrasound images from HC18	Dice: 99.1±0.8%
Hers et al. [69]	Femoral head and pelvis	Modified SegFormer (SegFormer3D)	98 3D ultrasound images from 34 patients	Dice: 90.8±3.8%, IoU: 83.5±6.0%
Wei et al. [70]	Ovarian tumor	Modified UNet++ (Res-ECA-UNet++)	350 ultrasound images	Dice: 95.63%, IoU: 91.84%

Discussion

Prevalence of Ultrasound Applications

Our results on the prevalence of different application categories can be used to assess which applications are treated with research interest proportional to their clinical use. Naturally, there are many points of comparison here, including the applications of medical ultrasound imaging in general, the tasks that would benefit from automated segmentation, and the topics that have enough scientific interest to be published. Systematic analysis in many respects is limited by the amount of existing statistics.

Common applications of medical ultrasound in research and clinical use have been categorized in earlier reviews: According to a systematic literature review by Stewart et al. [71] on medical ultrasound use in low- and middle-income countries (LMIC), the primary categories of medical ultrasound studies are cardiology and obstetrics, followed by pediatrics, gastroenterology, and internal medicine. In a more recent survey conducted by Ginsburg et al. [72], over a third of 177 respondents from LMIC reported using ultrasound for intravenous line (IV) insertion, studies on lung, cardiology, trauma assessment, and thoracentesis. Compared to our results, this implies that segmentation of ultrasound images of the lung might be an understudied topic. On the other hand, it is understandable that a task as simple as placing an IV does not require manual or automated segmentation, and trauma assessment might contain multiple subtasks such as organ segmentation already included in our categories.

Conversely, the number of DL studies for breast tumor segmentation is very high. According to our results, one-fourth of the DL-based ultrasound segmentation studies were about breast tumors. In particular, new methods developed for multiple segmentation tasks or also for modalities besides ultrasound were primarily evaluated in breast tumor segmentation to assess whether they work for ultrasound in general. While breast cancer is a significant health issue, the evaluation of multi-purpose segmentation methods on additional targets besides breast tumors would help to assess their performance in a more reliable way. Breast tumor segmentation was also identified as a major category of DL methods in an earlier review by Komatsu et al. [73].

Deep Learning Methodology

Based on our results, the ViTs, ViT/CNNs, and SAMs have become relatively established in the research on ultrasound image segmentation despite their recent release. However, while ViT-type models and SAMs are given significantly more data, the reported results are similar to those from CNNs, at least when comparing the evaluation metrics over several studies. The CNNs, in particular U-Net variants, have still preserved their popularity, especially in oncology and obstetrics. Notably, several articles introduce a novel, slightly modified version of the 10-year-old U-Net rather than test any of the already existing improvements of this architecture. This is likely explained by the pressure of scientific novelty present in any field of science: If the specific topic has already been studied previously, methodological development might be required for the study to be novel enough to become published.

Limitations of the Current Review

There are certain limitations in this work. Firstly, the article selection was based on only a single database, PubMed, which is focused on medical literature rather than technical DL publications. This might have introduced bias to the article selection process and, subsequently, to the results. Additionally, there was also a risk of subjective bias during the screening process. The search results were divided between three different authors so that each of them screened a certain number of studies, followed by the first author checking that the included articles fit the scope of the review. After this, the articles were divided between five authors to summarize. While several authors performing the screening independently and subsequently comparing their results would have reduced the risk of subjective judgments and errors, it was not feasible here due to the high number of search results and included articles.

Limitations of the Reviewed Studies

During the review, it was noted that most of the articles reported a very limited amount on technical constraints. Considering both ViTs and generative models are known to have very high requirements in terms of computational power and memory, the researchers in this field would benefit from practical information on what is needed to train and run the models. In addition to details on the software and hardware, the inference times reported in [60] would be very valuable. Furthermore, the use of explainable AI models would give the readers a more thorough understanding of the DL methods.

Additionally, there were limitations in the model evaluation. Some studies did not report any estimates of Dice score, IoU, nor Hausdorff distance, relying entirely on metrics such as pixel-level accuracy that are highly dependent on the amount of negative background within the images. Certain studies also only reported the Hausdorff distances in image points despite the fact that their values in millimeters would be more informative to the reader, especially in cases where the pixel resolution was unclear. Additionally, it is difficult to say how Dice or IoU values compare between different application categories, and what score is high enough for a certain application to be of use in clinical practice.

There was also a limited amount of information about certain details. For instance, with resized model inputs, it was sometimes unclear if the predicted masks were resized to the original size of the images prior to computing the evaluation metrics. Furthermore, the use of augmentation transformations was rarely accompanied by any analysis of whether it produced significant improvement in the model performance.

Future Directions

The use of ultrasound has been increasing steadily and, due to cheaper and portable devices, ultrasound imaging is also becoming accessible to a larger amount of the global population [1, 72]. This allows more research on health conditions that are only prevalent in certain populations. The current interest in DL and computer vision is also very high [73], so automated segmentation is a very fast-evolving topic of research. Additionally, new publicly available datasets would allow more methodological development.

Conclusion

The major applications of DL-based ultrasound segmentation include breast tumors, organ segmentation, and cardiovascular structures, followed by categories of obstetrics, orthopedics, thyroid nodules, oncology, and gynecology. Nearly half of the current studies use CNNs, and in particular, the U-Net architecture is still the single most common design despite being released already a decade ago. While less popular, ViTs, ViT/CNN hybrids, and SAMs have become well-established within the last few years. The limitations of the current research include a lack of details on technical requirements and inconsistencies in the model evaluation. Overall, the research on ultrasound image segmentation is a fast-developing field of research supported by increased use of ultrasound, new datasets, and advancements in DL.

Acknowledgements We wish to thank the reviewers for their constructive comments.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital). O.R. received funding from Otto A. Malm Foundation.

Data and Code Availability Supplementary data about the reviewed articles and the code used to analyze it are available at https://github.com/rklen/ultrasound_review. No new patient data was collected. The ultrasound image included in Fig. 3 is from the public BUSI dataset [41], and the images in Fig. 4 are from the public datasets BUSI [41], DDTI [44], TN3K [45], and CAMUS [46].

Declarations

Ethical Approval Not applicable. No new patient data was collected.

Informed Consent Not applicable. No new patient data was collected.

Conflict of Interest The authors declare no competing interests.

References

1. Danilo Avola et al. "Ultrasound medical imaging techniques: a survey". In: *ACM Computing Surveys (CSUR)* 54.3 (2021), pp. 1–38.
2. Aladin Carovac, Fahrudin Smajlovic, and Dzelaludin Junuzovic. "Application of ultrasound in medicine". In: *Acta Informatica Medica* 19.3 (2011), p. 168.
3. Ruud JG Van Sloun, Regev Cohen, and Yonina C Eldar. "Deep learning in ultrasound imaging". In: *Proceedings of the IEEE* 108.1 (2019), pp. 11–29.
4. Wei Zhang et al. "Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network". In: *Medical physics* 21.4 (1994), pp. 517–524.
5. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
6. Ramprasaath R Selvaraju et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.
7. Liang-Chieh Chen et al. "Rethinking atrous convolution for semantic image segmentation". In: *arXiv preprint* arXiv:1706.05587 (2017).
8. Liang-Chieh Chen et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.
9. Fabian Isensee et al. "nnu-net: Self-adapting framework for u-net-based medical image segmentation". In: *arXiv preprint* arXiv:1809.10486 (2018).
10. Hanguang Xiao et al. "Transformers in medical image segmentation: A review". In: *Biomedical Signal Processing and Control* 84 (2023), p. 104791.
11. Jiwoong J Jeong et al. "Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation". In: *Journal of Digital Imaging* 35.2 (2022), pp. 137–152.
12. Xiayu Guo et al. "UCTNet: Uncertainty-guided CNN-Transformer hybrid networks for medical image segmentation". In: *Pattern Recognition* 152 (2024), p. 110491.
13. Matthew J Page et al. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews". In: *bmj* 372 (2021).
14. Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

15. Ozan Oktay et al. "Attention u-net: Learning where to look for the pancreas". In: *arXiv preprint* arXiv:1804.03999 (2018).
16. Zongwei Zhou et al. "Unet++: A nested u-net architecture for medical image segmentation". In: *International workshop on deep learning in medical image analysis*. Springer. 2018, pp. 3–11.
17. Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. "Road extraction by deep residual u-net". In: *IEEE Geoscience and Remote Sensing Letters* 15.5 (2018), pp. 749–753.
18. Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pp. 203–211.
19. Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).
20. Alexey Dosovitskiy. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint* arXiv:2010.11929 (2020).
21. Satoshi Takahashi et al. "Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review". In: *Journal of Medical Systems* 48.1 (2024), p. 84.
22. Ze Liu et al. "Swin transformer: Hierarchical vision transformer using shifted windows". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 10012–10022.
23. Enze Xie et al. "SegFormer: Simple and efficient design for semantic segmentation with transformers". In: *Advances in neural information processing systems* 34 (2021), pp. 12077–12090.
24. Hu Cao et al. "Swin-unet: Unet-like pure transformer for medical image segmentation". In: *European conference on computer vision*. Springer. 2022, pp. 205–218.
25. Jieneng Chen et al. "Transunet: Transformers make strong encoders for medical image segmentation". In: *arXiv preprint* arXiv:2102.04306 (2021).
26. Alexander Kirillov et al. "Segment anything". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 4015–4026.
27. Nikhila Ravi et al. "Sam 2: Segment anything in images and videos". In: *arXiv preprint* arXiv:2408.00714 (2024).
28. Ian J Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).
29. Jascha Sohl-Dickstein et al. "Deep unsupervised learning using nonequilibrium thermodynamics". In: *International conference on machine learning*. pmlr. 2015, pp. 2256–2265.
30. Phillip Isola et al. "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.
31. Zi Ye et al. "Marrying Perona Malik diffusion with Mamba for efficient pediatric echocardiographic left ventricular segmentation". In: *Scientific Reports* 15.1 (2025), p. 32152.
32. Ali Karimi et al. "Improving spleen segmentation in ultrasound images using a hybrid deep learning framework". In: *Scientific Reports* 15.1 (2025), p. 1670.
33. Zhaozheng Chen, Mingjie Jiang, and Bernard Chiu. "Unsupervised shape-and-texture-based generative adversarial tuning of pre-trained networks for carotid segmentation from 3D ultrasound images". In: *Medical Physics* 51.10 (2024), pp. 7240–7256.
34. John Kalkhof et al. "MED-NCA: Bio-inspired medical image segmentation". In: *Medical Image Analysis* (2025), p. 103601.
35. Boheng Zhang et al. "MM-UKAN++: A Novel Kolmogorov-Arnold Network Based U-shaped Network for Ultrasound Image Segmentation". In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* (2025).
36. Guangju Li et al. "Human visual perception-inspired medical image segmentation network with multi-feature compression". In: *Artificial Intelligence in Medicine* (2025), p. 103133.
37. Wenliang Lu et al. "Echo-ODE: A dynamics modeling network with neural ODE for temporally consistent segmentation of video echocardiograms". In: *Frontiers in Physiology* 16 (2025), p. 1629121.
38. Boyuan Peng et al. "Bidirectional Copy–Paste Mamba for Enhanced Semi-Supervised Segmentation of Transvaginal Uterine Ultrasound Images". In: *Diagnostics* 14.13 (2024), p. 1423.
39. Wen Wen et al. "Multimodal model enhances qualitative diagnosis of hypervascular thyroid nodules: integrating radiomics and deep learning features based on B-mode and PDI images". In: *Gland Surgery* 14.8 (2025), pp. 1558–1571.
40. Sigurd Vangen Wifstad et al. "EasyPISA: Automatic Integrated PISA Measurements of Mitral Regurgitation From 2-D Color-Doppler Using Deep Learning". In: *Ultrasound in Medicine & Biology* 50.11 (2024), pp. 1628–1637.
41. Walid Al-Dhabyani et al. "Dataset of breast ultrasound images". In: *Data in brief* 28 (2020), p. 104863.
42. Moi Hoon Yap et al. "Automated breast ultrasound lesions detection using convolutional neural networks". In: *IEEE journal of biomedical and health informatics* 22.4 (2017), pp. 1218–1226.
43. Hanna Piotrzkowska-Wróblewska et al. "Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions". In: *Medical physics* 44.11 (2017), pp. 6105–6109.
44. Lina Pedraza et al. "An open access thyroid ultrasound image database". In: *10th International symposium on medical information processing and analysis*. Vol. 9287. SPIE. 2015, pp. 188–193.
45. Haifan Gong et al. "Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules". In: *Computers in biology and medicine* 155 (2023), p. 106389.
46. Sarah Leclerc et al. "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography". In: *IEEE transactions on medical imaging* 38.9 (2019), pp. 2198–2210.
47. Olivier Bernard et al. "Challenge on endocardial three-dimensional ultrasound segmentation (CETUS)". In: *Proceedings MICCAI challenge on echocardiographic three-dimensional ultrasound segmentation (CETUS)* (2014), pp. 1–8.
48. David Ouyang et al. "Video-based AI for beat-to-beat assessment of cardiac function". In: *Nature* 580.7802 (2020), pp. 252–256.
49. Rohit Singla et al. "The open kidney ultrasound data set". In: *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer. 2023, pp. 155–164.
50. Zachary Baum et al. "MR to ultrasound registration for prostate challenge-dataset". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI*. Vol. 2023. 2023.
51. Jialu Li et al. "Liver lesion segmentation in ultrasound: A benchmark and a baseline network". In: *Computerized Medical Imaging and Graphics* 123 (2025), p. 102523.
52. Hannah Strohm et al. "Contraction assessment of abdominal muscles using automated segmentation designed for wearable ultrasound applications". In: *International journal of computer assisted radiology and surgery* 19.8 (2024), pp. 1607–1614.
53. Adrien Meyer et al. "Ultrasam: a foundation model for ultrasound using large open-access segmentation datasets". In: *International Journal of Computer Assisted Radiology and Surgery* (2025), pp. 1–10.
54. Payel Pramanik et al. "DAU-Net: Dual attention-aided U-Net for segmenting tumor in breast ultrasound images". In: *Plos one* 19.5 (2024), e0303670.
55. Golla Madhu et al. "UCapsNet: A two-stage deep learning model using U-Net and Capsule Network for breast cancer segmentation and classification in ultrasound imaging". In: *Cancers* 16.22 (2024), p. 3777.
56. Xian-Ya Zhang et al. "Automatic tumor segmentation and lymph node metastasis prediction in papillary thyroid carcinoma using ultrasound keyframes". In: *Medical Physics* 52.1 (2025), pp. 257–273.

57. Pan Pan et al. "Attention decoupled contrastive learning for semi-supervised segmentation method based on data augmentation". In: *Physics in Medicine & Biology* 69.12 (2024), p. 125017.
58. Sonja MC de Zwarte et al. "Automated Segmentation of Fetal Intracranial Volume in Three-Dimensional Ultrasound Using Deep Learning: Identifying Sex Differences in Prenatal Brain Development". In: *Human Brain Mapping* 45.17 (2024), e70058.
59. Bahram Jafrasteh et al. "MGA-Net: A novel mask-guided attention neural network for precision neonatal brain imaging". In: *NeuroImage* 300 (2024), p. 120872.
60. Zheng Yue et al. "Prior-knowledge embedded u-net based fully automatic vessel wall volume measurement of the carotid artery in 3d ultrasound image". In: *IEEE Transactions on Medical Imaging* (2024).
61. Luyao Li et al. "An efficient deep learning strategy for accurate and automated detection of breast tumors in ultrasound image datasets". In: *Frontiers in Oncology* 14 (2025), p. 1461542.
62. Yan Pang et al. "Efficient Breast Lesion Segmentation from Ultrasound Videos Across Multiple Source-limited Platforms". In: *IEEE Journal of Biomedical and Health Informatics* (2025).
63. Oona Rainio, Jarmo Teuho, and Riku Klén. "Evaluation metrics and statistical tests for machine learning". In: *Scientific Reports* 14.1 (2024), p. 6086.
64. Zahra Hamedi et al. "AI-Based Analysis of Abdominal Ultrasound Images to Support Medical Diagnosis in Emergency Departments". In: *Healthcare of the Future 2025*. IOS Press, 2025, pp. 16–21.
65. Sajal Kumar Babu Degala et al. "Segmentation and Estimation of Fetal Biometric Parameters using an Attention Gate Double U-Net with Guided Decoder Architecture". In: *Computers in Biology and Medicine 180* (2024), p. 109000.
66. Min Hu et al. "Mamba-and ResNet-based dual-branch network for ultrasound thyroid nodule segmentation". In: *Bioengineering* 11.10 (2024), p. 1047.
67. Sonlin Shi, Palisha Alimu, and Pazilai Mahemut. "The Study of Echocardiography of Left Ventricle Segmentation Combining Transformer and Convolutional Neural Networks". In: *International Heart Journal* 65.5 (2024), pp. 889–897.
68. Po-Wei Chen et al. "Deep learning model for diagnosis of venous thrombosis from lower extremity peripheral ultrasound imaging". In: *iScience* 27.12 (2024).
69. Benjamin Hers et al. "SegFormer3D: Improving the Robustness of Deep Learning Model-Based Image Segmentation in Ultrasound Volumes of the Pediatric Hip". In: *Ultrasound in Medicine & Biology* 51.5 (2025), pp. 751–758.
70. Shushan Wei, Zhaoting Hu, and Lu Tan. "Res-ECA-UNet++: an automatic segmentation model for ovarian tumor ultrasound images based on residual networks and channel attention mechanism". In: *Frontiers in Medicine* 12 (2025), p. 1589356.
71. Kelsey A Stewart et al. "Trends in ultrasound use in low and middle income countries: a systematic review". In: *International Journal of Maternal and Child Health and AIDS* 9.1 (2020), p. 103.
72. Amy Sarah Ginsburg et al. "A survey of barriers and facilitators to ultrasound use in low-and middle-income countries". In: *Scientific Reports* 13.1 (2023), p. 3322.
73. Masaaki Komatsu et al. "Towards clinical application of artificial intelligence in ultrasound imaging". In: *Biomedicines* 9.7 (2021), p. 720.