STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **x**, Month 202x, pp 0–18.
Published online in International Academic Press (www.IAPress.org)

# Enhancing Echocardiographic Segmentation through KL-Divergence-Based Intensity Distribution Constraints in U-Net Models

Zahir AITMATEN [1], Soraya ALOUI [1], Ahror BELAID [2,*]

[1] *Laboratory of LIMAD, University of A. Mira, Bejaia, Algeria*
[2] *LIMAD, Faculty of Exact Sciences. Data Sciences And Applications research Unit, CERIST, University of A. Mira, Algeria*

**Abstract** Incorporating prior knowledge through appropriate regularization strategies is essential for improving medical image segmentation, particularly in challenging scenarios involving low image quality, poor contrast, and limited training data. In this work, we conduct a systematic study of several regularization approaches for 2D echocardiographic left ventricle segmentation by comparing seven U-Net–based models, ranging from standard loss-based baselines to advanced probabilistic constraints.
The evaluated models include a Dice-only baseline, Dice with weight decay, an autoencoder-based regularization using image reconstruction, and multiple probabilistic formulations that enforce consistency between the intensity distributions of the predicted segmentation and the ground truth. In particular, we propose a density-based constraint relying on the Kullback–Leibler divergence between the intensity distributions of the segmented region of interest and its corresponding label, estimated either through parametric modeling or non-parametric kernel density estimation. Additional variants explore the injection of global and batch-wise reference distribution parameters into the latent space to guide the segmentation process.
Experiments conducted on echocardiographic images of varying quality (poor, medium and good) demonstrate that models incorporating probabilistic density constraints consistently outperform conventional regularization strategies, achieving improvements of up to three Dice points over baseline and autoencoder-based approaches. Beyond quantitative gains, these constraints lead to smoother and more anatomically plausible segmentations, particularly in cases where conventional loss functions struggle.
These results highlight the effectiveness of distribution-aware regularization for ultrasound image segmentation and suggest that enforcing intensity distribution consistency constitutes a robust and principled form of prior knowledge in data-limited and low-quality imaging conditions.

**Keywords** Neural networks, Deep learning, CNN, Auto-encoder, Image segmentation, image processing, KL-Divergence, Echocardiography

**DOI:** 10.19139/soic-2310-5070-2869

## 1. Introduction

Image segmentation is a critical task in medical image analysis and an essential step toward automating medical image processing. It plays a vital role in many medical imaging applications, including computer-aided diagnosis and computer-aided interventions [6]. However, automation can be challenging due to the significant variations in shape and size of anatomical structures between patients [15].

In recent years, significant progress has been made in standard computer vision tasks such as recognition, classification and segmentation, primarily due to the widespread use of Convolutional Neural Networks (CNNs)

*Correspondence to: Zahir AITMATEN (Email: zahir.aitmaten@univ-bejaia.dz). Department of Electrical Engineering, University of Bejaia. Algeria.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

[5]. Deep convolutional neural networks have emerged as the leading approach to semantic segmentation in computer vision and medical image analysis, achieving significant results in fully supervised methods [9]. As a result, CNNs have readily become the preferred approach for medical image processing (segmentation and classification) [20]. While U-Net [26] has become the state-of-the-art method for this task in the biomedical field. Isensee et al. demonstrated in No New Net that a generic U-Net architecture with minor modifications can achieve competitive performance [14]. Other solutions, such as FCN [27], VNet [16], DenseNet have been proposed and have demonstrated significant improvements in comparison to classical methods [13]. Md. Eshmam Rayed et al. [41] give a State-of-the-art advancements and challenges in medical image segmentation. On the other hand, Yan Xu et al [39] give a review of traditional, deep learning and hybrid methods. Transformers that were first used in natural language processing are recently adapted for image processing [35] achieving interesting performances.

Deep Neural Networks (DNNs) have become essential tools in computer vision, but they can only perform optimally when there is sufficient training data available. When there is a lack of training data, approaches such as synthetic data generated by diffusion models [34], [42], data augmentation, transfer learning [38] or imposing constraints can still yield good performance [8].

The semantic segmentation of medical images involves assigning a label to each pixel without requiring human initialization. Success in this task depends heavily on the availability of high-quality imaging data with their labels supplied by specialists [12]. However, the presence of noise, the complexity of the objects, and the low contrast in medical images present significant barriers to achieving an ideal segmentation. In such cases, integrating prior knowledge as constraints has proven to be useful in obtaining precise and feasible segmentation outcomes [6], particularly in cases where the images are of poor quality [7].

Prior information can take numerous forms, such as length boundaries; edge polarity; shape models and moments (area/volume) [6]. In many real-world problems, prior knowledge about the statistics of the desired solution is available. For example, in the foreground–background image segmentation problem, we may have knowledge about the exact shape and size of the object being segmented, and want to find the optimal outcome that has a specific area (foreground pixel number) and boundary length [3]. And the incorporation of prior information in medical imaging has the potential to have a greater impact than its application in natural image analysis. This is because anatomical structures in medical images inherently possess more defined constraints regarding their shape and spatial positioning [7].

Building upon these insights, in the present work we extend the study to a systematic comparison of seven U-Net–based models incorporating different forms of regularization and prior knowledge. These include traditional loss-based baselines, weight decay, autoencoder-based regularization, and probabilistic constraints that enforce consistency between the predicted segmentation and the expected intensity distribution of the region of interest. By softly imposing these constraints on the network and latent space level, particularly through Kullback–Leibler divergence–based density matching, we aim to improve segmentation quality across images of varying quality (poor, medium, and good) and demonstrate the advantages of distribution-aware regularization compared to conventional approaches.

## 2. Related work

In this section, we will discuss the different types of constraints used since the 1980s. In traditional computational vision research, low-level domains such as edge or line detection, were considered to be autonomous bottom-up workflows that proceed by using only the information present in the image itself. However, this rigidly sequential method propagated errors made at a low level without any chance to rectify them. To address this problem, a technique called snake [17] was introduced. The snake basic model is a controlled spline under the influence of image forces and external constraint forces. This model of energy is not able to adapt to changes in the topology of the progressing contour when performing a direct implementation. For this reason, another technique based on active contours evolving in time according to intrinsic geometric measures of the image is proposed in [18]. This approach is based on the relation between the active contours and the computation of minimal distance curves. The evolving contours naturally split and merge, allowing the simultaneous detection of several objects.

In 2011, Klodt et al. [2] showed that shape priors in terms of moment constraints can be enforced. In particular, the lower-order moments, such as the volume, centroid, and variance of the shape, can be effectively imposed in an interactive segmentation process. However, the shape prior suppresses the deviations of the observed shape from the training shapes, which is particularly undesirable in medical image segmentation, where malformations of organs should be detected rather than ignored. It may, therefore, be more appropriate to impose only some coarse-level shape information rather than the exact form of the object. In [3], the authors addressed the problem of minimizing an objective function under multiple constraints. The algorithm proposed in their work can handle a larger class of constraints, such as non-linear equality. The constraints used in the paper are size and object boundary length.

In 2012, Varol et al. [19] proposed a latent variable framework that showed that imposing constraints on the latent variables does not necessarily ensure that equivalent constraints are satisfied in the output space. The paper focused on the problem of learning the mapping from a given latent space to the output space under equality and inequality constraints and proposed to learn this mapping by minimizing the prediction error on labeled examples, while jointly enforcing constraints on unlabeled ones.

In various situations, certain statistics of the ground truth like area and boundary length are accessible. However, enforcing statistics softly does not guarantee that the solution would have statistics that match the desired ones. Nevertheless, the imposition of any constraint improves the accuracy of the segmentation on average. It has been seen that the size and variance constraints are especially powerful for image segmentation [4].

Constraints can be imposed in two ways, hard and soft. Hard constraints enables solving the Lagrangian in different ways that are not computationally tractable. And using soft constraints has two major drawbacks. First, it makes it necessary to wisely choose the relative importance of the different terms in the loss function, which is not easy. Second, this gives no guarantee that the constraints will be satisfied in practice. In 2017, Márquez-Neila et al. [8] tried to prove that imposing hard constraints is computationally feasible, but treating them as soft constraints remains preferable.

Prior information (as constraints) can take many forms, which have been commonly used as a regularization term in traditional segmentation methods based on energy optimization [7]. The majority of the classification and regression models use a pixel-level loss function (e.g. cross-entropy or mean square error) which does not fully take into account the underlying semantic information and dependencies in the output space. Without using high-level prior information about the expected objects, the use of only low-level information such as intensity and texture does not produce the desired segmentation results. Many studies have proved that prior knowledge about the objects' shapes to be segmented can greatly improve the accuracy of the segmentation outcome. However, when dealing with a training set that includes arbitrary prior shapes, there persists an unresolved challenge of determining an appropriate prior shape model to effectively guide the object segmentation process [21]. To overcome these limitations, [21] proposed using a deep Boltzmann machine to learn the arbitrary shape prior. Similarly, [7] used convolutional auto-encoder networks to learn anatomical shape variations from medical images and take into account high-level information by extracting the best distribution of the input. In the ACNN approach [7], an auto-encoder is used to extract the features of the labels instead of the feature of the image input. These features are then used to regularize the segmentation network. In 2017, Ravishankar et al. [22] proposed a method where the predicted mask is introduced in an auto-encoder to imitate the ground truth label over a custom function loss that contains the terms of regularized auto-encoder and segmentation networks. In the same framework, [10] used a variational auto-encoder to regularize the network segmentation, and reconstructed the original image from the code produced by the encoder branch, which is shared between the decoder of the segmentation branch and the variational branch. The motivation for using the auto-encoder branch was to add additional guidance and regularization to the encoder part, since the size of the training dataset was limited. This regularization technique won the BraTS 2018 challenge. Recently, deep learning–based approaches have been extensively investigated for echocardiographic image segmentation. In particular, a feasibility study on transesophageal echocardiography (TEE) was presented in [36], where several CNN architectures, including U-Net, Attention U-Net, U-Net++, and UNeXt, were evaluated for left ventricle segmentation. The results showed that attention-based and multi-scale architectures outperform the standard U-Net, especially in challenging cases characterized by low contrast and poorly defined boundaries, highlighting the importance of architectural design in echocardiographic segmentation tasks.

In cases where the ground truth labels are not available or not sufficient for fully supervised learning, weakly supervised learning can be used with partially labeled images like points, scribbles, bounding boxes or image tags [9]. Numerous weakly supervised segmentation methods have been proposed to improve the segmentation quality of CNNs trained with partial annotations. They can generally be categorized into two groups: pseudo-label methods and regularized loss methods [13]. In 2015, Pathak et al. [5] presented a framework to integrate weakly supervised image-tags into the learning process using some linear constraints. According to the author, the paradigm of weakly supervised learning aims to detect the signal that is common to all the positives but absent from all the negatives. The study demonstrated that constraints provide a natural means of defining the desired output space of a labeling and can decrease the quantity of strong supervision needed for CNNs. In 2019, Kervadec et al. [9] proposed a simple penalty-based method to impose constraints on the network output. Although this method is not optimal, because there is no assurance that the enforced constraints are satisfied, it unexpectedly produces significantly better results than Lagrangian-based constrained CNNs, by decreasing the computational requirements during training. The proposed loss function for weakly supervised image segmentation is novel, simple, and performs significantly better than Lagrangian optimization, achieving results close to full supervision with only a small fraction of annotated pixels and negligible computational overhead. Dalca et al. [30], in 2018, proposed a generative probabilistic model to segment the different regions of the brain in a completely unsupervised setting, which takes into account the anatomical priors.

Constrained-CNN loss is a popular approach for weakly supervised segmentation, which imposes inequality constraints on the network's outputs based on prior knowledge, such as the size and shape of the object of interest. However, describing complex prior knowledge, such as an irregular shape and boundary, in programming language could prove challenging. To address this issue, an adversarial constrained-CNN loss (ACCL) for weakly supervised medical image segmentation is proposed [13]. In this new paradigm, prior knowledge is encoded and represented by reference masks, which are further employed to impose constraints on the segmentation outputs through adversarial learning.

It has been shown that incorporating prior information about the shape improves significantly the performance of segmentation algorithms. However, incorporating such prior knowledge is a practical challenge. In January 2021, Bohlender et al. [23] provided an overview of the efforts using shape priors in deep learning frameworks for medical image segmentation. Very recent works have explored probabilistic and generative modeling paradigms for echocardiographic image segmentation. For instance, Rahman et al. [40] proposed a diffusion-based framework that leverages dual noise modeling to improve the robustness of left ventricle segmentation under severe noise and low-quality imaging conditions. Their results demonstrate that explicitly modeling uncertainty and data distribution can significantly enhance segmentation performance, particularly in challenging echocardiographic scenarios.

Although prior work has explored various forms of constraints, including shape priors and auto-encoder regularization, few or maybe no studies have considered explicitly enforcing intensity distribution constraints to guide segmentation. In the present work, we systematically compare seven U-Net–based models, including probabilistic density constraints and latent-space distribution injection, highlighting their impact on segmentation quality across echocariographic images.

## 3. Proposed Method

In this work, we consider the segmentation of the left ventricle in 2D echocardiographic images. It is known that the region of interest exhibits characteristic intensity distributions, which can serve as prior knowledge. Incorporating this information as a constraint can improve segmentation performance, particularly when training data are limited.

We perform a systematic comparison of seven U-Net–based models, each implementing a distinct strategy for regularization and prior knowledge incorporation:

- **Mode–1: Dice (Dice-only baseline):** Uses the standard Dice loss for segmentation without any additional regularization. Serves as a reference for performance comparison. Corresponding to Branch 1 in Fig. 4.

- **Model–2: W–decay (Dice + weight decay):** Introduces weight decay to regularize the model parameters and reduce overfitting. This regularization is decoupled from the loss function and applied directly during optimization using the AdamW optimizer. Corresponding to Branch 1 in Fig. 4.
- **Model–3: VAE (Variational Autoencoder regularization):** Introduces a second output branch trained to reconstruct the input image from the latent space. This auxiliary task serves as a regularization signal, encouraging the network to preserve image structure while segmenting. Corresponding to the combination of Branch 1 and Branch 3 in Fig. 4.

- **Model–4: KLD–alpha (Dice + KLD, density constraint):** Extends the baseline with a pseudo output branch. The pseudo output is obtained by applying a soft-thresholded layer to the predicted segmentation, enabling recovery of the input image values within the predicted mask and maintaining differentiability, and its intensity distribution is compared to the ground truth distribution via the Kullback–Leibler divergence. This enforces a soft constraint on the output, guiding the segmentation toward intensity distributions consistent with the target region. Corresponding to the combination of Branch 1 and Branch 2 in Fig. 4.
- **Model–5: KLD–kde (KDE-based density constraint):** Similar to Model 4, but the intensity distribution is estimated non-parametrically using kernel density estimation (KDE) from histogram, providing a smoother and more flexible approximation of the target distribution. Corresponding to the combination of Branch 1 and Branch 2 in Fig. 4.
- **Model–6: RD–like–VAE(Reference Distribution injection):** Uses the parameters of a reference distribution computed on the entire training set and injects them into the latent space via fully connected layers. The network is trained to segment the image while aligning its latent representation with the reference distribution. Corresponding to Branch 1 extended with Branch 4, as illustrated in Fig. 4.
- **Model–7: BD–like–VAE (Batch-wise Distribution injection):** Estimates the parameters of the intensity distribution for each batch of images and injects them into the latent space. This enforces adaptive regularization, encouraging the network to learn batch-specific variations in intensity distribution while performing segmentation. Corresponding to Branch 1 extended with Branch 5, as illustrated in Fig. 4.

### *3.1. Loss Function*

The problem can be formulated as follows: Suppose that the region of interest (ROI) in the dataset follows a certain distribution $P_e$ that is calculated over the entire training dataset. Figure 1 shows the histograms of the image and ROI, respectively. The graphs in Fig. 2 show some intensity distributions estimated from the dataset, and we find that the alpha distribution (a) is the best representation of the dataset's distribution, followed by the inverse-gamma (b), log-normal (c), and inverse-Gaussian (d) distributions. However, the normal and Rayleigh distributions do not represent the dataset's distribution, as shown in Fig. 2 (e) and (f).

The graphs in Figure 3 show the alpha distribution for 2, 4, 8, 16, 32, 64, and 128 images chosen randomly.

We observe that the distribution on 2, 4, 8, 16, 32, 64, and 128 or the overall dataset follows the alpha distribution. This gives us the idea to assume that the distribution on each batch during training is the alpha distribution. The ROI's predicted distribution is given by the distribution of the predicted region in the output of the model network, as shown in Fig. 4.

The alpha probability distribution function used from the scipy.stats module corresponds to the Pearson Type III distribution, defined as:

$$f(x;\alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \tag{1}$$

and its generalized form:

$$f(x;\alpha,\text{loc},\text{scale}) = \frac{1}{\text{scale}} f_{standard}\left(\frac{x-\text{loc}}{\text{scale}}\right), \tag{2}$$

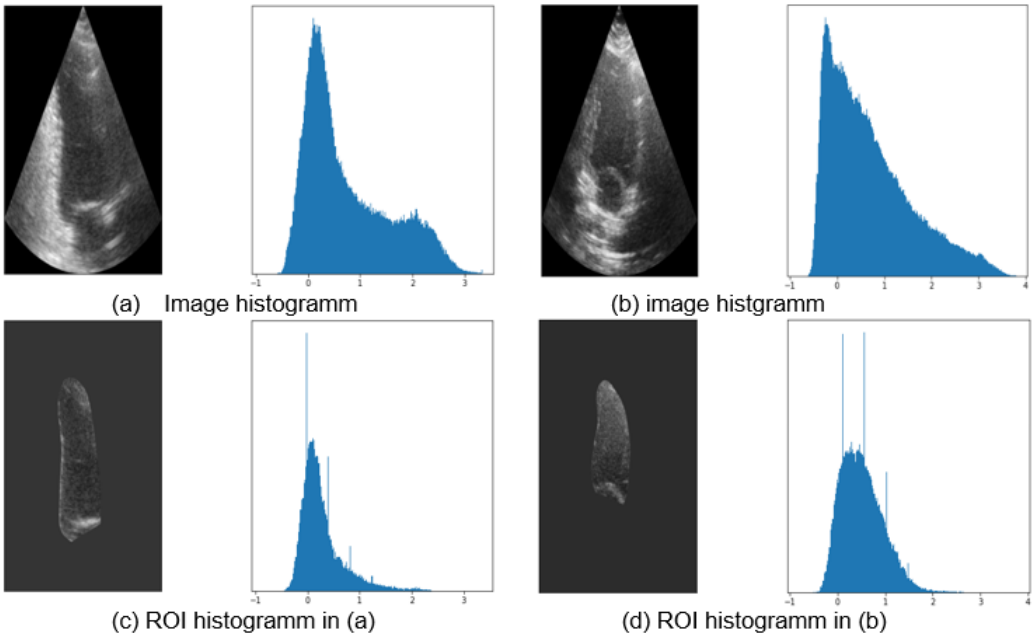


Figure 1. Histogram of image and ROI

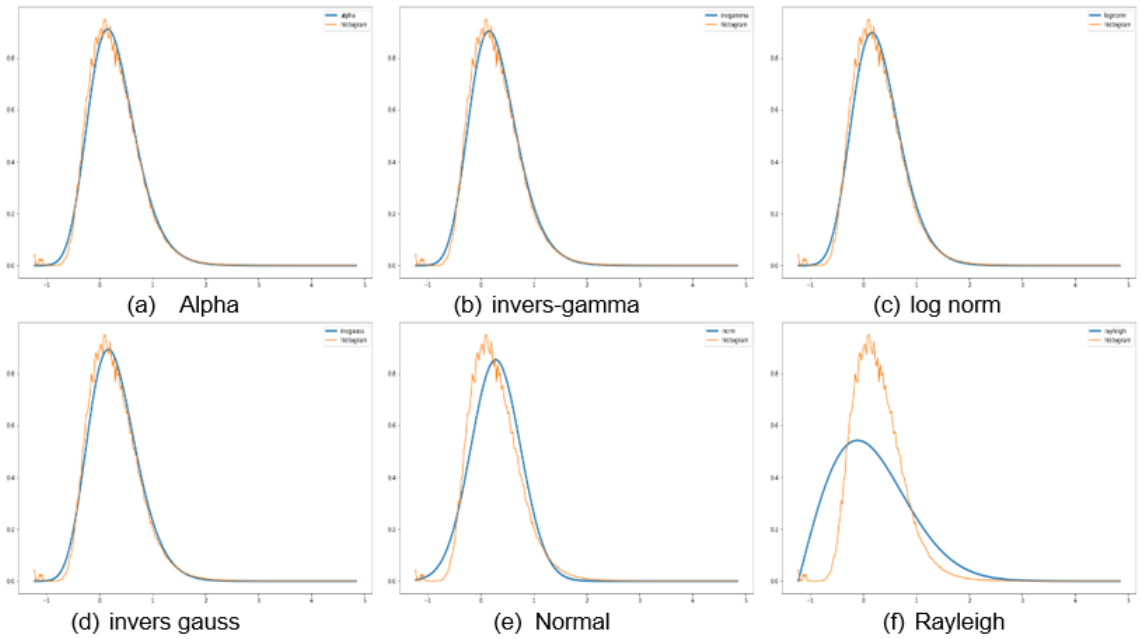


Figure 2. Data set distribution compared to some known distributions

where $\alpha$ is the shape parameter, and $\Gamma(\alpha)$ is the normalization constant. The three parameters $\alpha$, loc, and scale are estimated for both the predicted output and the ground truth on each mini-batch, and compared using the Kullback-Leibler divergence (KLD). Parameters are estimated using the maximum likelihood method.

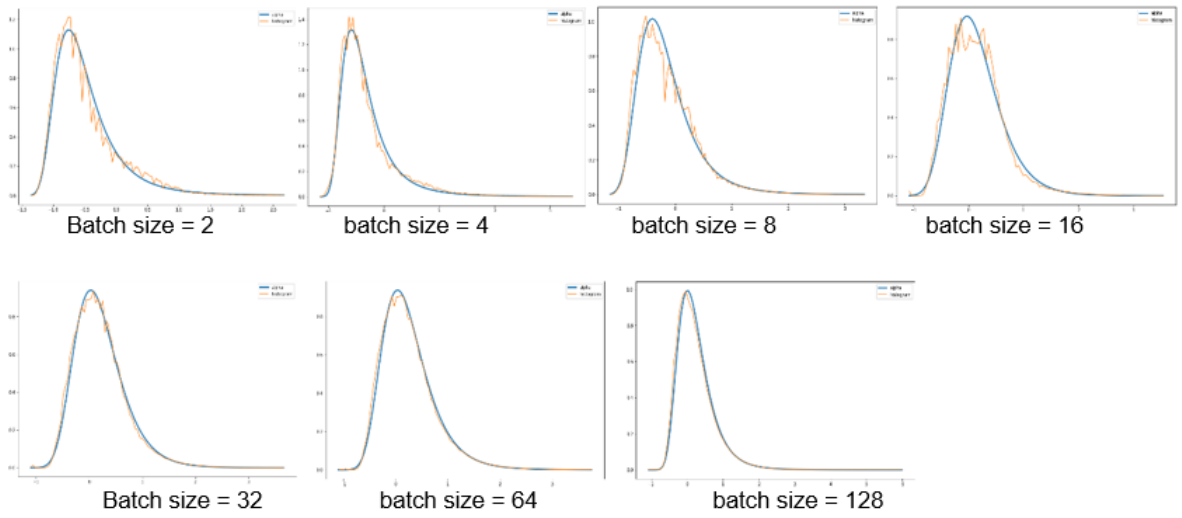


Figure 3. Alpha distribution estimated for different batch sizes

This allows the network to regularize its predictions so that the predicted ROI follows the expected intensity distribution, enhancing segmentation quality especially for small or low-quality datasets.

The overall loss function is given by:

$$L(\theta) = \text{Dice} + \text{KLD}(P_p(Y_w|X;\theta) \parallel P_e(X_w)), \tag{3}$$

where $P_e$ and $P_p$ are the estimated and predicted distributions, $X$ is the input image, $Y$ is the output mask, and $X_w$ and $Y_w$ denote the input and predicted ROI, respectively.

The predicted and estimated distributions are computed as:

$$P_p(Y_w|X;\theta) = P(X \odot Y_{pred}), \tag{4}$$

$$P_e(X_w) = P(X \odot Y_{label}), \tag{5}$$

with $\odot$ representing element-wise multiplication. The KLD and Dice terms are computed as:

$$\text{KLD}(P\|Q) = \sum_x P(x) \ln \frac{P(x)}{Q(x)}, \tag{6}$$

$$\text{Dice} = 1 - \frac{2\sum_i y_{true,i}\, y_{pred,i} + \epsilon}{\sum_i y_{true,i} + \sum_i y_{pred,i} + \epsilon}. \tag{7}$$

To decouple parameter regularization from the gradient-based update, weight decay is incorporated into the optimization process and can be written as:

$$\mathbf{W}_{t+1} = \mathbf{W}_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} - \eta\lambda\mathbf{W}_t \tag{8}$$

where $\eta$ is the learning rate and $\lambda$ denotes the decoupled weight decay coefficient. and

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t \tag{9}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2 \tag{10}$$

where $g_t = \nabla_\theta \mathcal{L}_t$ denotes the gradient of the loss function at iteration $t$

$\beta_1 \in [0, 1)$ and $\beta_2 \in [0, 1)$ are exponential decay coefficients that control the contributions of past gradients to the first-order moment $m_t$ and the second-order moment $v_t$, respectively.

### *3.2. Architecture of the CNN*

The proposed architecture (see Fig. 4) is based on the design introduced by [10]. The ResNet blocks remain unchanged and consist of two convolutional layers with group normalization [29] followed by ReLU activation, combined with identity skip connections (see Fig. 4, bottom). To reduce the model complexity, the number of filters at each spatial level was reduced, decreasing the total number of parameters from approximately 6M to 1.5M.

For the input image resolution, images were automatically cropped to ensure that all regions of interest were visible and that spatial dimensions were divisible by 16. Unlike [10], the variational autoencoder (VAE) regularization branch was removed. Instead, a pseudo second output was introduced to provide an alternative form of regularization, as described in the following sections.

The encoder follows a standard convolutional design in which the spatial resolution is progressively reduced by a factor of 2 while the number of feature maps is doubled at each level. Downsampling is performed using strided convolutions with a stride of 2. All convolutional layers use $3 \times 3$ kernels, starting with 16 filters at the first level. The encoder bottleneck has a feature dimension of $128 \times h \times w$, corresponding to a spatial resolution eight times smaller than the input image.

The decoder mirrors the encoder structure, with a single block per spatial level. Each decoder stage begins with a 2D bilinear upsampling operation that doubles the spatial resolution, followed by a $1 \times 1$ convolution to reduce the number of feature maps by a factor of 2. Skip connections are implemented by summing the upsampled features with the corresponding encoder outputs. The final decoder layer restores the original input resolution and feature size, followed by a $1 \times 1$ convolution that maps the features to a single output channel and a sigmoid activation function.

Based on this common architecture, seven models were developed, which are described in detail in the *Proposed Method* section. All models share the same encoder–decoder backbone, while slight variations in the number of parameters are introduced in specific cases. In particular, Model 3 includes an additional reconstruction branch used for regularization, whereas Models 6 and 7 incorporate a small number of fully connected layers to learn, respectively, a global reference distribution and a batch-wise distribution. Apart from these additions, the core network topology remains unchanged.

### *3.3. Training details of the proposed models*

Regarding the training setup, all models were optimized using the Adam optimizer with a fixed learning rate of $10^{-5}$. A unified training protocol was adopted to ensure a fair comparison between the different models. A learning rate reduction strategy was applied, reducing the learning rate by a factor of $0.5$ when no improvement greater than $0.001$ was observed for 10 consecutive epochs. All models were trained using a batch size of 2.

The base loss function is identical for all models and consists of the Dice loss. Additional regularization terms are introduced depending on the considered model. For Model 3 we adopted the same weighting coefficients as in the original work. Specifically, the loss function combines the Dice loss with a regularization term composed of a Kullback–Leibler divergence assuming a Gaussian prior and an $L_2$ reconstruction loss, both weighted by a factor of $0.1$.

For Models 4, 5, 6, and 7, the loss function is defined as the sum of the Dice loss and a regularization term weighted by a factor of $0.1$. The formulation of the regularization loss differs according to the model and is detailed in the *Proposed Method* section. Early stopping was applied to all models using the validation Dice score as the monitored metric. Training was terminated if no improvement greater than $0.001$ was observed over 30 consecutive epochs. No fixed maximum number of training epochs was imposed; instead, training was stopped

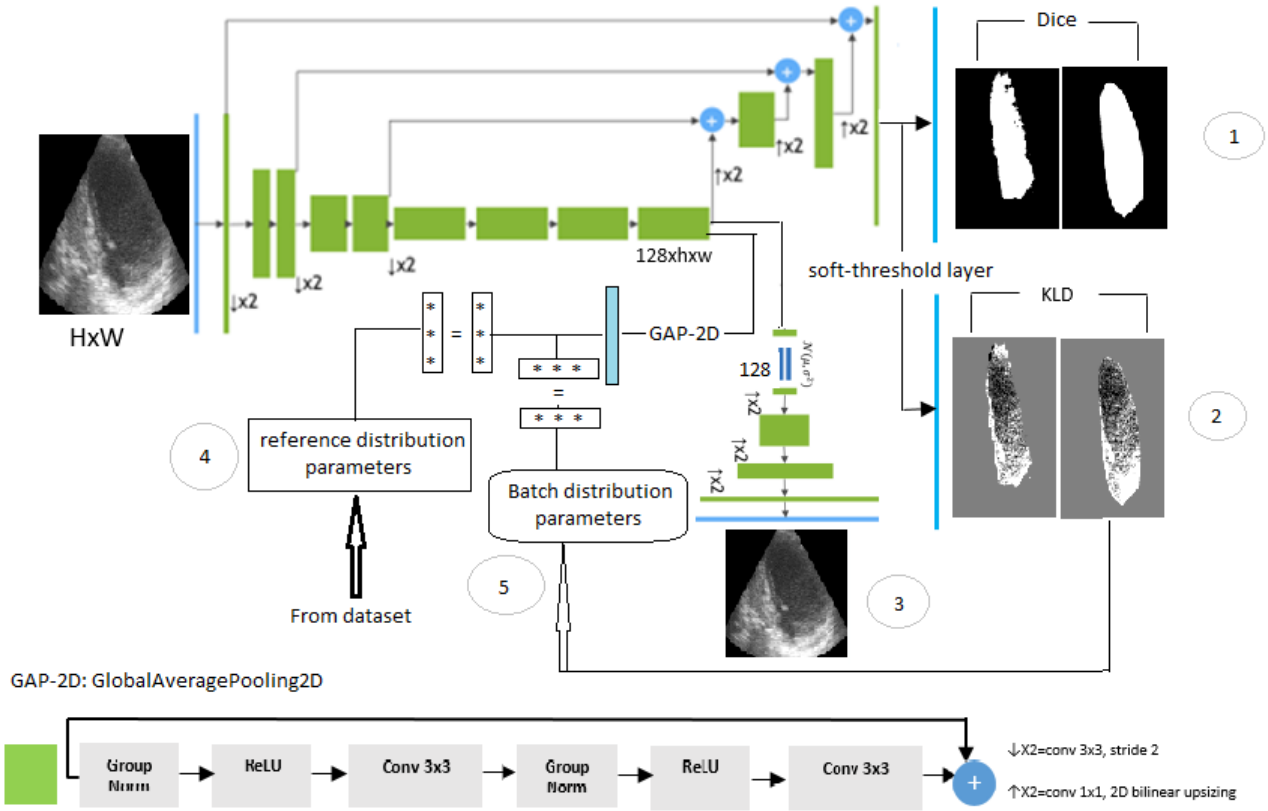


Figure 4. Models network

when the validation Dice score stagnated around low values or when a large discrepancy between training and validation Dice scores was observed, indicating potential overfitting.

All models were trained on an NVIDIA L4 GPU with 22 GB of dedicated GPU memory, ensuring sufficient computational resources for stable training and fair comparison across architectures.

### *3.4. Dataset*

We used the publicly accessible CAMUS dataset [32], which contains fully annotated 2D echocardiographic images. The data were collected at the University Hospital of Saint-Étienne (France) during routine clinical examinations for left ventricular function assessment, using a GE Vivid E95 ultrasound scanner with an M5S probe. For each patient, sequences were recorded in apical two-chamber (A2C) and four-chamber (A4C) views, covering at least one full cardiac cycle. Expert clinicians manually annotated the relevant cardiac structures, leading to a dataset with a wide range of image quality.

For our study, we focused exclusively on the 2CH images. After filtering, we obtained 118 *poor*-quality images with a spatial resolution of $(1152 \times 496 \times 1)$, 386 *medium*-quality images of size $(1184 \times 624 \times 1)$, and 396 *good*-quality images with a resolution of $(1136 \times 704 \times 1)$. We reserved 10 poor, 30 medium, and 36 good images for testing, and split the remaining images into 80% for training and 20% for validation. To highlight the effect of using distribution constraints, segmentation was performed exclusively on the left ventricle, while all other regions were assigned to the background.

## 4. Results and Discussion

The evaluation of all seven models demonstrates clear differences in performance across image qualities. Models 4, 5, 6, and 7 are our proposed architectures, while Models 1, 2, and 3 are existing baseline models. Model 4 consistently achieves the highest validation Dice score across all image types (good=0.883, medium=0.879 and poor=0.791), confirming its robustness. Although Model 4 (Dice+KLD_alpha) generally outperforms Model 5 (KLD_KDE) (good=0.878, medium=0.869 and poor=0.766), the test results show that Model 5 (0.901) slightly surpasses Model 4 (0.897) on good-quality images, while Model 4 remains superior on medium and poor-quality images. Additional experiments conducted on medium and poor-quality test images confirmed the observed trend on good-quality data. Model 5 consistently outperformed Model 4 in terms of Dice score across all image quality levels. Although Model 4 achieved slightly higher performance on the validation set, its test performance was systematically inferior. This behavior suggests that Model 5 exhibits superior generalization capability and robustness to image quality variations.

For medium-quality images, the performance ranking is Model 4 first, followed by Model 5, and then Model 2 (Dice with weight decay). In poor-quality images, Model 4 still attains the highest Dice scores; however, statistical analysis via p-values (Table 4) indicates no significant superiority over the other models, as all models struggle with segmentation under these conditions, with maximum validation and test Dice scores not exceeding 80%. Models 3 and 7 were not evaluated on poor-quality images, as their performance on medium and good-quality images was already suboptimal, making it unlikely for them to yield meaningful results on lower-quality data.

Models 3, which incorporates image reconstruction, and Model 7, using batch-wise distribution, consistently rank at the bottom across all image qualities where they were tested, even when trained for significantly more epochs than the other models. In contrast, Model 6, which leverages the reference distribution estimated from the entire database and injects it into the latent space, achieves a strong third place on good-quality images, demonstrating the advantage of using a global distribution constraint over batch-wise estimation. The baseline models (Models 1 and 2) remain competitive, trailing the best models (4, 5, and 6) by only 2–3 Dice points. Notably, Models 4 and 5 reach their optimal performance in fewer epochs compared to the other models, demonstrating both efficiency and effectiveness.

The quantitative results are summarized in Tables 1, 2, and 3, which report recall, Dice, and modified Hausdorff distance (MHD) for each model across poor, medium, and good-quality images, respectively. Complementary p-values and median statistics for each image category are presented in Tables 4, 5, and 6, providing insight into statistical differences between models. The p-values for medium and good-quality images indicate that Model 4 is significantly different from all other models except Model 5: for good-quality images, the p-value between Model 4 and Model 5 is 0.6655, showing no significant difference, whereas for medium-quality images, Model 4 is statistically superior to Model 5 (p-value = 0.009; median difference = +0.011). These tables highlight the consistent superiority of Model 4 and the competitive performance of our proposed Models 5 and 6, while also revealing the limitations of Models 3 and 7. Regarding the MHD metric, which provides an indication of the alignment between the mask contours and the prediction, we observe that Model 4 consistently outperforms all other models across all image types.

Figures 5, 6, and 7 illustrate example segmentations for poor, medium, and good-quality images, respectively. The visualizations emphasize that Model 4 produces the most accurate and smooth segmentations across most images, whereas Models 3 and 7 exhibit noticeable inconsistencies. Figures 8, 9, and 10 show the evolution of validation Dice during training, highlighting how quickly Models 4 and 5 converge compared to the other models. Additionally, Figures 11, 12, and 13 present the test Dice scores for all models, allowing direct comparison of their final performance. The ranking based on the number of times a model is ranked first in Dice across test images further confirms the superiority of Model 4 over all other models (see Table 7: Ranking distribution of models across image quality categories).

From a qualitative perspective, the superior performance of Model 4 can be attributed to the integration of the KLD term based on the alpha distribution, which effectively regularizes the network and prevents over- or under-segmentation, particularly in medium-quality images. Model 5 benefits from a similar regularization mechanism using KDE, which performs well on good-quality images but is less robust on lower-quality ones. Model 6 benefits

from using the reference distribution from the entire database, enabling it to outperform batch-wise estimation in certain cases, especially for good-quality images. Models 3 and 7 underperform due to the added complexity of the reconstruction branch and batch-wise distribution, which introduces optimization challenges and slows convergence.

Nevertheless, certain failure cases Figures 14 highlight the limits of these approaches: extreme poor-quality images with strong artifacts, low contrast, or highly irregular ventricle shapes lead to substantial under-segmentation or fragmented masks even for the best-performing Model 4. Models 5 and 6 occasionally misidentify background structures as part of the ventricle, resulting in over-segmentation, while Models 3 and 7 fail to produce coherent contours under these challenging scenarios. These observations underscore that, despite the robustness of distribution-based regularization, **segmentation performance can deteriorate sharply under extreme image degradation**, emphasizing the importance of careful preprocessing and potential post-processing strategies when handling low-quality clinical data.

As shown in Table 8, except for Model 3 which requires a longer execution time, all other models exhibit similar execution times.

Overall, these results suggest that appropriately designed distribution-based constraints can improve segmentation accuracy, especially when image quality is variable, while excessive model complexity may be counterproductive. Furthermore, the efficiency observed for Models 4 and 5 indicates that these constraints allow the network to reach stable performance with fewer epochs, which is advantageous for practical deployment. The combination of quantitative metrics, p-values, visual inspections, and ranking distributions provides a comprehensive view of model performance across varying image qualities, with Model 4 emerging as the most reliable and efficient approach across most conditions.

Table 1. Segmentation metrics for poor-quality images Fig. 5

| **Model** | **Image 1** | | | **Image 2** | | | **Image 3** | | | **Image 4** | | | **Image 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ |
| Dice | 0.765 | 0.840 | 0.068 | 0.636 | 0.718 | 0.102 | 0.978 | **0.867** | 0.113 | 0.533 | 0.6824 | 0.104 | 0.717 | 0.804 | 0.077 |
| W–decay | 0.789 | 0.850 | 0.071 | 0.692 | 0.750 | 0.099 | **0.992** | 0.840 | 0.151 | 0.587 | 0.7243 | 0.098 | 0.698 | 0.777 | 0.120 |
| KLD–alpha | **0.814** | **0.867** | 0.056 | **0.758** | **0.783** | **0.089** | 0.990 | 0.824 | 0.103 | 0.629 | **0.7555** | **0.084** | **0.843** | **0.838** | **0.068** |
| KLD–kde | 0.784 | 0.830 | 0.061 | 0.722 | 0.761 | 0.100 | 0.976 | 0.809 | 0.147 | 0.626 | 0.7551 | 0.126 | 0.792 | 0.804 | 0.075 |
| RD–like–VAE | 0.813 | 0.865 | **0.049** | 0.737 | 0.758 | 0.113 | 0.987 | 0.820 | **0.096** | **0.630** | 0.7552 | 0.131 | 0.766 | 0.796 | 0.074 |

Table 2. Segmentation metrics for medium-quality images Fig. 6

| **Model** | **Image 1** | | | **Image 2** | | | **Image 3** | | | **Image 4** | | | **Image 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ |
| Dice | **0.967** | 0.952 | 0.068 | 0.941 | 0.921 | 0.096 | 0.833 | 0.830 | 0.355 | 0.851 | 0.799 | 0.079 | 0.982 | 0.789 | 0.299 |
| W–decay | 0.949 | 0.947 | 0.049 | 0.922 | 0.927 | 0.062 | 0.839 | 0.855 | 0.244 | 0.914 | 0.827 | 0.070 | **0.989** | 0.797 | 0.138 |
| VAE | 0.930 | 0.758 | 0.377 | 0.917 | 0.776 | 0.344 | 0.673 | 0.413 | 0.780 | 0.853 | 0.742 | 0.228 | 0.991 | 0.494 | 0.893 |
| KLD–alpha | 0.966 | 0.951 | **0.021** | **0.949** | **0.929** | **0.031** | **0.897** | **0.908** | **0.051** | 0.956 | **0.870** | 0.052 | 0.9664 | **0.824** | **0.062** |
| KLD–kde | 0.955 | **0.955** | 0.027 | 0.923 | 0.903 | 0.044 | 0.819 | 0.852 | 0.106 | 0.920 | 0.862 | **0.049** | 0.984 | 0.780 | 0.143 |
| RD–like–VAE | 0.956 | 0.948 | 0.023 | 0.889 | 0.916 | 0.038 | 0.867 | 0.871 | 0.201 | **0.959** | 0.862 | 0.053 | 0.980 | 0.793 | 0.315 |
| BD–like–VAE | 0.966 | 0.941 | 0.035 | 0.901 | 0.908 | 0.041 | 0.801 | 0.772 | 0.209 | 0.967 | 0.856 | 0.059 | 0.967 | 0.651 | 0.468 |

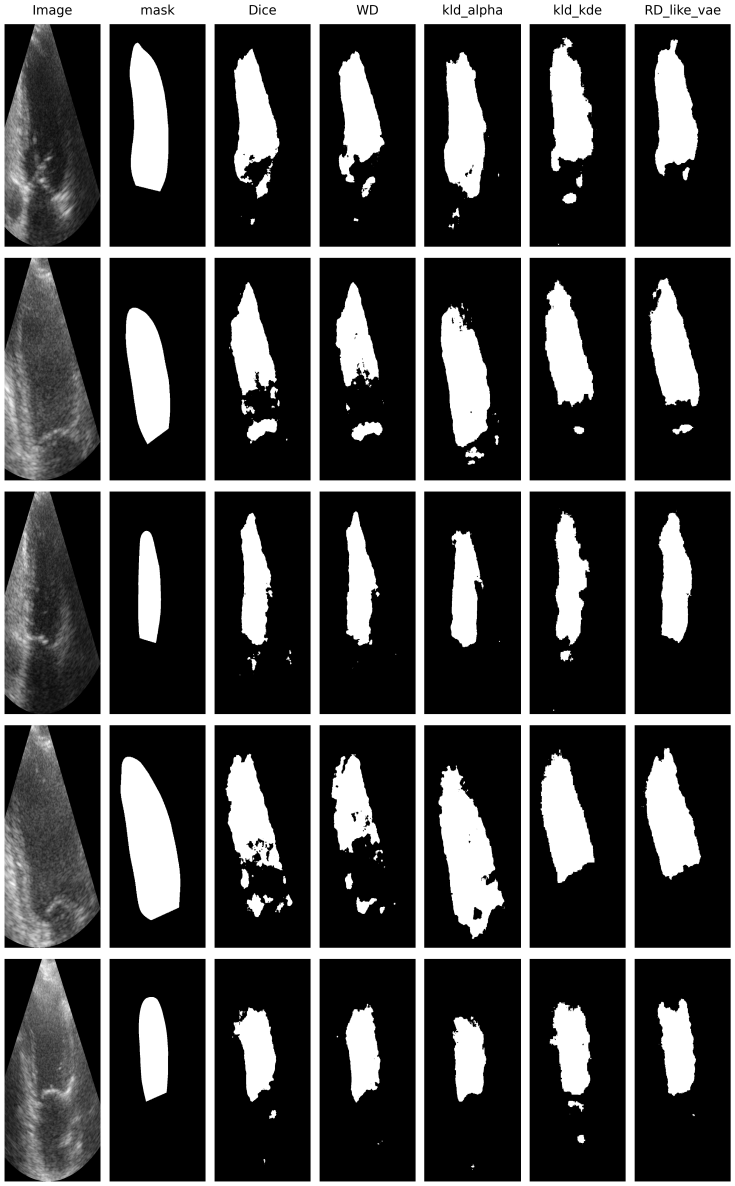


Figure 5. Segmentation results on poor-quality images. From left to right: (1) input image, (2) ground truth, (3) Dice-based model, (4) W-decay, (5) KLD-alpha, (6) KLD-kde, and (7) RD-like VAE.

Table 3. Segmentation metrics for good-quality images Fig. 7

| **Model** | **Image 1** | | | **Image 2** | | | **Image 3** | | | **Image 4** | | | **Image 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ | Rec | Dice | MHD ↓ |
| Dice | 0.921 | 0.948 | 0.025 | 0.897 | 0.930 | 0.064 | **0.984** | 0.940 | 0.104 | 0.956 | 0.930 | 0.094 | 0.937 | 0.894 | 0.264 |
| W–decay | 0.903 | 0.936 | 0.864 | 0.863 | 0.918 | 0.079 | 0.979 | 0.928 | 0.159 | **0.992** | 0.913 | 0.218 | 0.938 | 0.873 | 0.346 |
| VAE | 0.824 | 0.881 | 0.250 | 0.825 | 0.879 | 0.156 | 0.891 | 0.792 | 0.315 | 0.846 | 0.794 | 0.318 | 0.838 | 0.654 | 0.559 |
| KLD–alpha | **0.944** | **0.9559** | 0.026 | **0.933** | **0.944** | 0.046 | 0.972 | **0.944** | **0.025** | 0.982 | **0.939** | 0.068 | **0.961** | **0.921** | **0.035** |
| KLD–kde | 0.941 | 0.9550 | 0.049 | 0.888 | 0.928 | **0.033** | 0.948 | 0.935 | 0.039 | 0.940 | 0.937 | **0.028** | 0.897 | 0.906 | 0.037 |
| RD–like–VAE | 0.939 | 0.949 | **0.023** | 0.897 | 0.933 | 0.043 | 0.963 | 0.918 | 0.046 | 0.967 | 0.930 | 0.035 | 0.886 | 0.909 | 0.038 |
| BD–like–VAE | 0.697 | 0.436 | 0.680 | 0.779 | 0.601 | 0.386 | 0.665 | 0.320 | 0.561 | 0.693 | 0.412 | 0.594 | 0.611 | 0.237 | 0.971 |

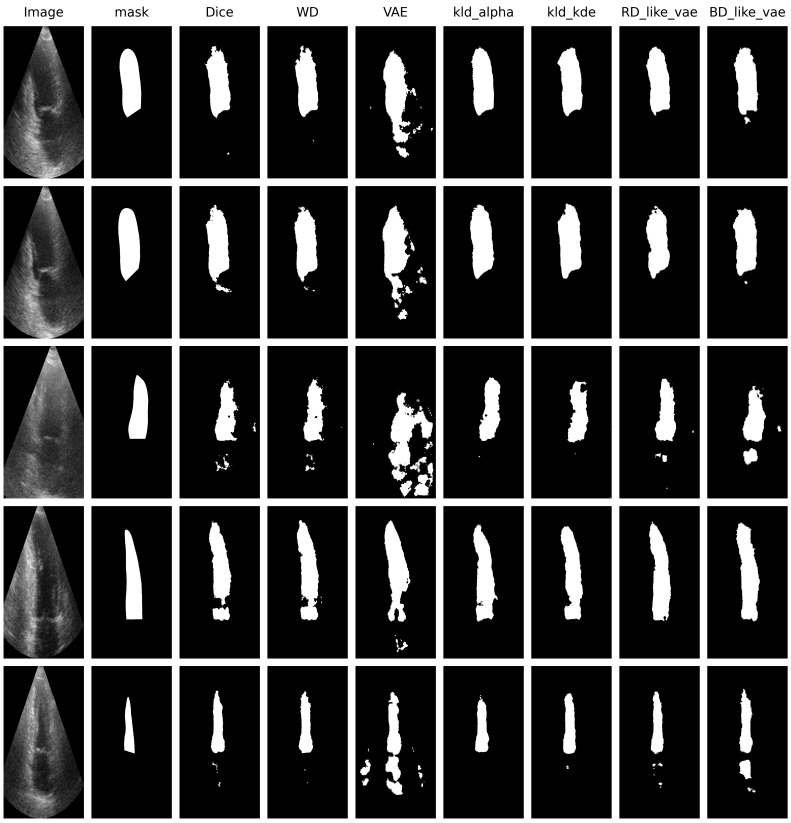


Figure 6. Segmentation results on medium-quality images. From left to right: (1) input image, (2) ground truth, (3) Dice-based model, (4) W-decay, (5) VAE, (6) KLD-alpha, (7) KLD-kde, (8) RD-like VAE, and (9) BD-like VAE.

Table 4. Pairwise comparison between models for poor–quality images. Upper triangular values indicate p-values obtained using the Wilcoxon signed-rank test, while lower triangular values represent the difference in median Dice scores (row, column). Statistically significant p-values ($p < 0.05$) are highlighted in bold.

| | Dice | W–decay | KLD–alpha | KLD–kde | RD–like–VAE |
|---|---|---|---|---|---|
| Dice | – | 0.3750 | 0.1934 | 0.9219 | 0.6953 |
| W–decay | -0.0121 | – | 0.1055 | 0.6953 | 0.4316 |
| KLD–alpha | +0.0307 | +0.0244 | – | **0.0488** | **0.0371** |
| KLD–kde | -0.0008 | +0.0150 | -0.0227 | – | 0.4922 |
| RD–like–VAE | +0.0042 | +0.0114 | -0.0095 | +0.0055 | – |

Table 5. Pairwise comparison between models for medium–quality images. Upper triangular values indicate p-values obtained using the Wilcoxon signed-rank test, while lower triangular values represent the difference in median Dice scores (row, column). Statistically significant p-values ($p < 0.05$) are highlighted in bold.

| | Dice | W–decay | VAE | KLD–alpha | KLD–kde | RD–like–VAE | BD–like–VAE |
|---|---|---|---|---|---|---|---|
| Dice | – | 0.7766 | **0.0000** | **0.0054** | 0.1347 | 0.3599 | **0.0293** |
| W–decay | -0.0011 | – | **0.0000** | **0.0066** | 0.4280 | 0.4771 | **0.0155** |
| VAE | -0.1112 | -0.1142 | – | **0.0000** | **0.0000** | **0.0000** | **0.0000** |
| KLD–alpha | +0.0109 | +0.0110 | +0.1221 | – | **0.0093** | **0.0277** | **0.0001** |
| KLD–kde | +0.0068 | +0.0082 | +0.1201 | -0.0112 | – | 0.5028 | **0.0010** |
| RD–like–VAE | +0.0035 | +0.0014 | +0.1141 | -0.0099 | +0.0007 | – | **0.0137** |
| BD–like–VAE | -0.0106 | -0.0102 | +0.1037 | -0.0200 | -0.0206 | -0.0086 | – |

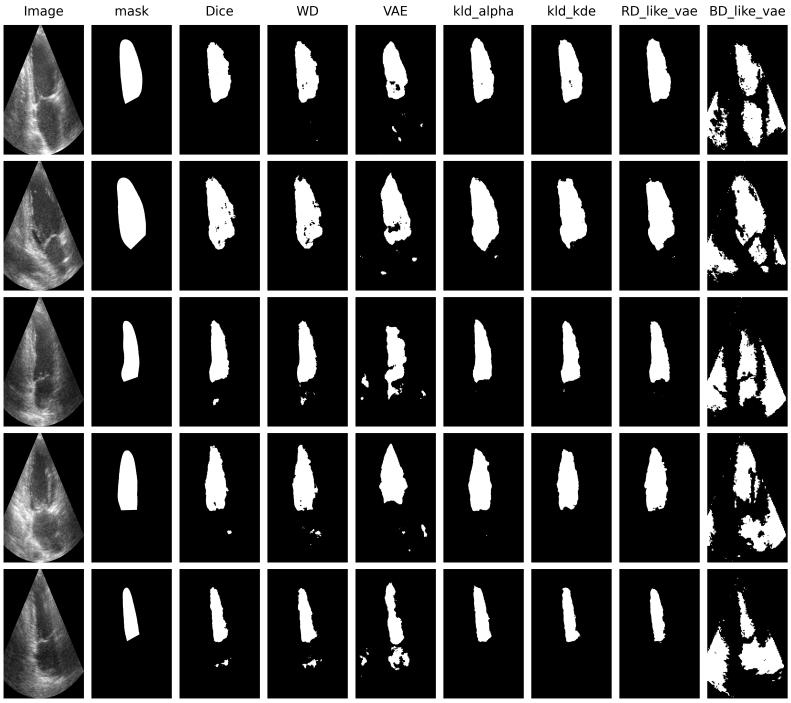


Figure 7. Segmentation results on good-quality images. From left to right: (1) input image, (2) ground truth, (3) Dice-based model, (4) W-decay, (5) VAE, (6) KLD-alpha, (7) KLD-kde, (8) RD-like VAE, and (9) BD-like VAE.

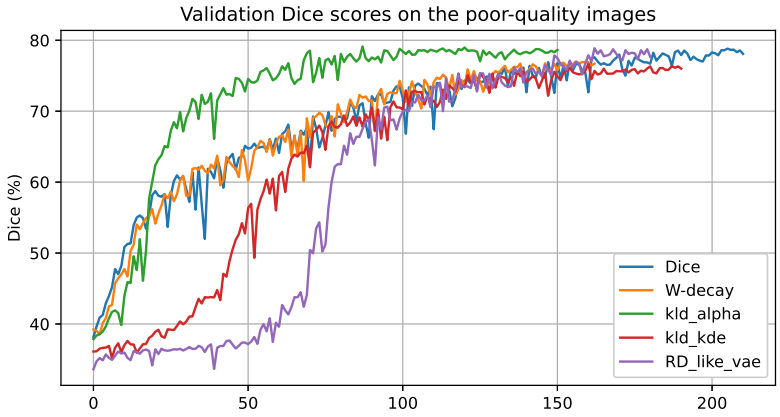


Figure 8. Validation Dice scores on the poor-quality images.

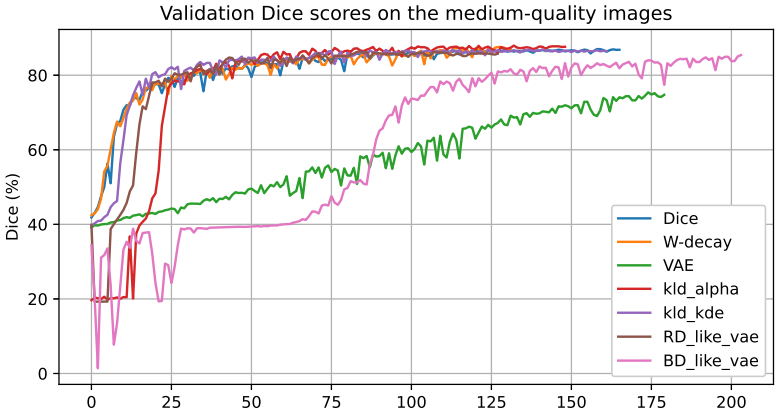


Figure 9. Validation Dice scores on the medium-quality images.

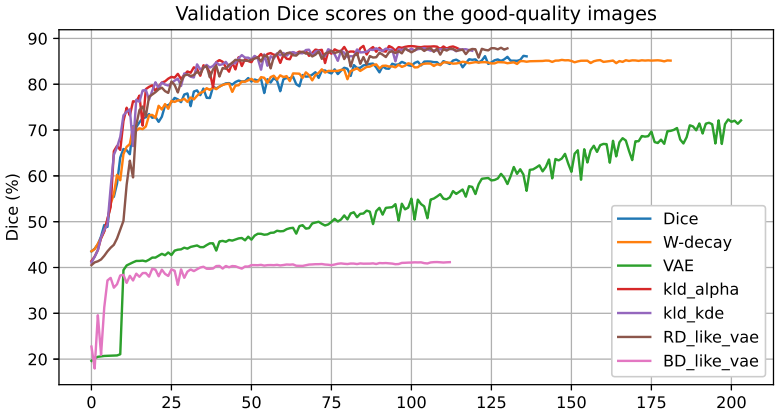


Figure 10. Validation Dice scores on the good-quality images.

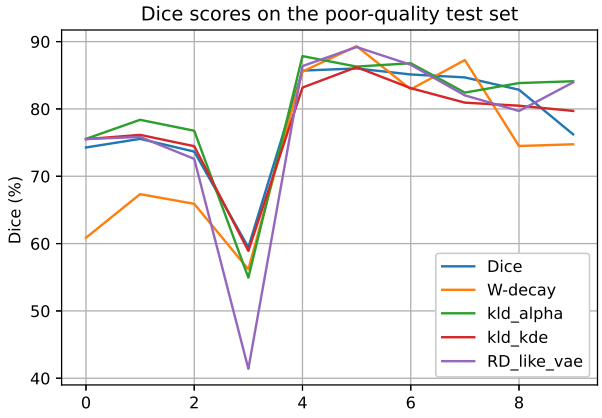


Figure 11. Dice scores on the poor-quality test set.

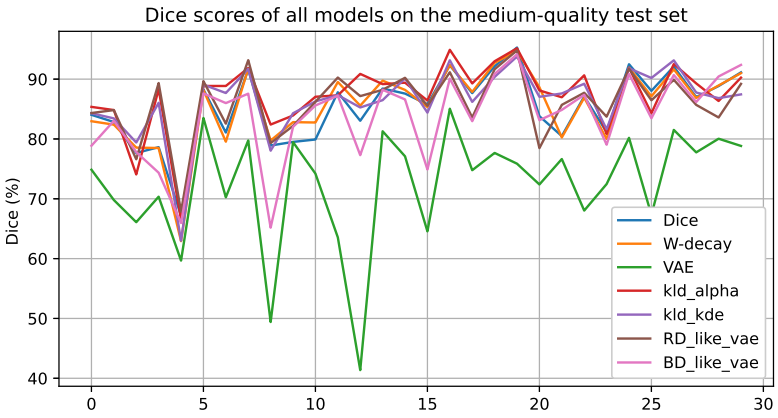


Figure 12. Dice scores on the medium-quality test set.

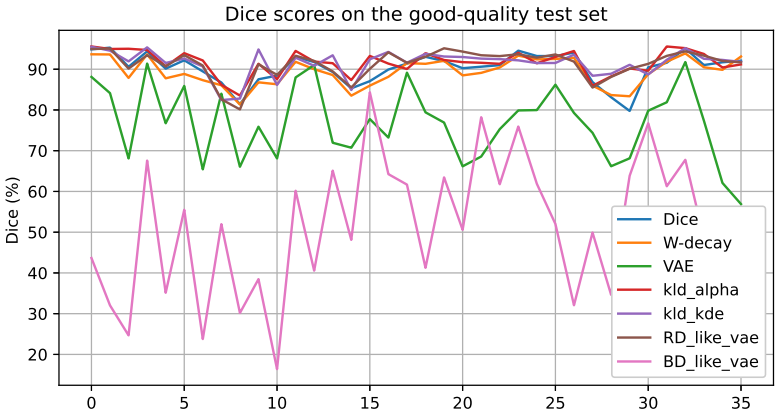


Figure 13. Dice scores on the good-quality test set.

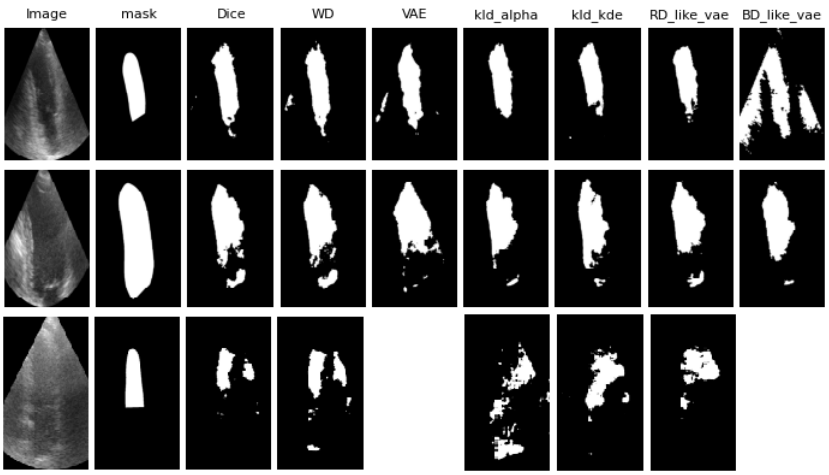


Figure 14. Validation Dice scores on the good-quality images.

Table 6. Pairwise comparison between models for good–quality images. Upper triangular values indicate p-values obtained using the Wilcoxon signed-rank test, while lower triangular values represent the difference in median Dice scores (row, column). Statistically significant p-values ($p < 0.05$) are highlighted in bold.

| | Dice | W–decay | VAE | KLD–alpha | KLD–kde | RD–like–VAE | BD–like–VAE |
|---|---|---|---|---|---|---|---|
| Dice | – | **0.0000** | **0.0000** | **0.0094** | **0.0106** | 0.0747 | **0.0000** |
| W–decay | -0.0121 | – | **0.0000** | **0.0000** | **0.0000** | **0.0001** | **0.0000** |
| VAE | -0.1351 | -0.1221 | – | **0.0000** | **0.0000** | **0.0000** | **0.0000** |
| KLD–alpha | +0.0064 | +0.0206 | +0.1528 | – | 0.6656 | 0.5992 | **0.0000** |
| KLD–kde | +0.0087 | +0.0207 | +0.1473 | +0.0003 | – | 0.4437 | **0.0000** |
| RD–like–VAE | +0.0029 | +0.0187 | +0.1389 | -0.0013 | -0.0006 | – | **0.0000** |
| BD–like–VAE | -0.3706 | -0.3573 | -0.2655 | -0.3883 | -0.3780 | -0.3768 | – |

Table 7. Ranking distribution of models across image quality categories

| Model | Good | | | Medium | | | Poor | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1st | 2nd | 3rd | 1st | 2nd | 3rd | 1st | 2nd | 3rd |
| Dice | 4 | 6 | 8 | 2 | 4 | 8 | 3 | 1 | 1 |
| W–decay | 1 | 0 | 4 | 2 | 7 | 6 | 0 | 2 | **3** |
| VAE | 0 | 0 | 0 | 0 | 0 | 0 | x | x | x |
| KLD–alpha | **12** | 9 | 10 | **12** | 8 | 4 | **7** | 0 | 1 |
| KLD–kde | 11 | 9 | 6 | 5 | **8** | 4 | 0 | **4** | 2 |
| RD–like–VAE | 8 | **12** | **8** | 7 | 3 | 6 | 0 | 3 | **3** |
| BD–like–VAE | 0 | 0 | 0 | 2 | 0 | 2 | x | x | x |
| **Total images** | **36** | | | **30** | | | **10** | | |

Table 8. Execution time, number of parameters, and RAM requirements for each model across three image types.

| **Model** | **Time (s/epoch) poor** | **Time (s/epoch) medium** | **Time (s/epoch) good** | **Params (M)** | **RAM (GB)** |
|---|---|---|---|---|---|
| Model 1 | 61-86 | 244-263 | 261-265 | 1 506 689 | 12.1 |
| Model 2 | 65-69 | 283-298 | 274-279 | 1 506 689 | 12.1 |
| Model 3 | x | 306-312 | 318-328 | 14,906,790 | 16.5 |
| Model 4 | 66-68 | 217-224 | 225-239 | 1,506,721 | 13.7 |
| Model 5 | 69-73 | 233-241 | 241-243 | 1,506,721 | 13.7 |
| Model 6 | 67-68 | 226-232 | 235-237 | 1,519,332 | 16.3 |
| Model 7 | x | 227-231 | 230-235 | 38,088,420 | 16.2 |

## 5. Conclusion

The results of this study demonstrate that integrating distribution-based constraints into CNN training for medical image segmentation can substantially improve performance across varying image qualities. In particular, Model 4, which incorporates the KLD term based on the alpha distribution, consistently achieves the highest accuracy and produces smoother, more anatomically plausible segmentations, especially in medium-quality images. Models 5 and 6, leveraging similar distribution-based regularization mechanisms, also show competitive performance, with Model 6, which uses a global reference distribution in the latent space, consistently outperforming Model 7 that relies on batch-wise estimation.

Nevertheless, segmentation performance can degrade sharply on extreme poor-quality images with strong artifacts, low contrast, or highly irregular ventricle shapes. Even the best-performing models occasionally under-segment or misidentify background structures, highlighting the need for careful preprocessing and potential post-processing strategies in clinical applications. Models with increased architectural complexity, such as those including reconstruction branches or batch-wise distribution estimation (Models 3 and 7), do not necessarily yield better results and may even hinder convergence.

Moreover, analysis of the ranking distribution shows that no single model consistently achieves the best Dice score across all images: some images are best segmented by Model 1, others by Model 2, Model 4, Model 5, or Model 6. This observation suggests that a mixture of experts approach, where each model acts as a specialized expert, could further improve segmentation performance. Incorporating additional experts, such as transformer-based models, may enhance the network's ability to handle diverse image qualities and anatomical variations.

Overall, the study confirms that appropriately designed distribution-based regularization provides a reliable and efficient mechanism to enhance segmentation accuracy, particularly when image quality varies, while excessive complexity or misapplied constraints may be counterproductive. Although estimating the intensity or probability distribution can increase training time, it does not affect inference, and the resulting segmentations more faithfully capture the shape of the target organ. These findings provide a strong foundation for future work exploring adaptive distribution constraints, mixture of experts frameworks, and advanced architectures to achieve even more robust and clinically reliable segmentation results.

## 6. Declarations

Conflict of interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.

REFERENCES


1. Rother, Carsten and V. Kolmogorov and A. Blake, *"GrabCut" interactive foreground extraction using iterated graph cuts*, ACM Transactions on Graphics (TOG), vol. 23, no. 3, pp. 309–314, 2004.
2. Klodt, Maria and D. Cremers, *A Convex Framework for Image Segmentation with Moment Constraints*, International Conference on Computer Vision, pp. 2236–2243, 2011.
3. Lim, Yongsub and K. Jung and P. Kohli, *Constrained discrete optimization via dual space search*, NIPS Workshop on Discrete Optimization on Machine Learning (DISCML), 2011.
4. Lim, Yongsub and K. Jung and P. Kohli, *Efficient energy minimization for enforcing label statistics*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 9, pp. 1893–1899, 2014.
5. Pathak, Deepak P. Krahenbuhl and T. Darrell, *Constrained convolutional neural networks for weakly supervised segmentation*, Proceedings of the IEEE International Conference on Computer Vision, pp. 1796–1804, 2015.
6. Nosrati, S. Masoud and G. Hamarneh, *Incorporating prior knowledge in medical image segmentation: A survey*, arXiv preprint arXiv:1607.01092, 2016.
7. Oktay, Ozan and others, *natomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation*, IEEE Transactions on Medical Imaging, vol. 37,no. 2, pp. 384–395, 2017.
8. Márquez-Neila, Pablo and M. Salzmann and P. Fua, *Imposing hard constraints on deep networks: Promises and limitations*, arXiv preprint arXiv 1706.02025, 2017.
9. Kervadec, Hoel and others, *Constrained-CNN losses for weakly supervised segmentation*, Medical Image Analysis, vol. 54, pp. 88–99, 2019.
10. Myronenko, Andriy, *3D MRI brain tumor segmentation using autoencoder regularization*, Springer, pp. 311–320, 2019.

11. Decencière, Etienne and others, *Dealing with topological information within a fully convolutional neural network*, International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, pp. 462–471, 2018.
12. Simpson, L. Amber and others, *A large annotated medical image dataset for the development and evaluation of segmentation algorithms*, arXiv preprint arXiv:1902.09063, 2019.
13. Zhang, Pengyi and Y. Zhong and X. Li, *ACCL: Adversarial constrained-CNN loss for weakly supervised medical image segmentation*, arXiv preprint arXiv:2005.00328, 2020.
14. Isensee, Fabian and others, *No new-net*, In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, Springer, New York, pp. 234–244, 2018.
15. Roth, R. Holger and others, *Deep learning and its application to medical image segmentation*, Medical Imaging Technology,vol. 36, no. 2, pp. 63–71, 2018.
16. Milletari, Fausto and N. Navab and S. Ahmadi, *V-net: Fully convolutional neural networks for volumetric medical image segmentation*, 2016 Fourth International Conference on 3D Vision. IEEE, pp. 565–571, 2016.
17. Kass, Michael and A. Witkin and D. Terzopoulos, *Snakes: Active contour models*, International Journal of Computer Vision, vol. 1, no. 4, pp. 321–331, 1988.
18. Caselles, Vicent and R. Kimmel and G. Sapiro, *Geodesic active contours*, International Journal of Computer Vision, vol. 22, no. 1, pp. 61–79, 1997.
19. Varol, Aydin and others, *A constrained latent variable model*, IEEE Conference on Computer Vision and Pattern Recognition, pp. 2248–2255, 2012.
20. Litjens, Geert and others, *A survey on deep learning in medical image analysis*, Medical Image Analysis, vol. 42, pp. 60–88, 2017.
21. Chen, Fei and others, *Deep learning shape priors for object segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1870–1877, 2013.
22. Ravishankar, Hariharan and others, *Learning and incorporating shape models for semantic segmentation*, International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, pp. 203–211, 2017.
23. Bohlender, Simon and I. Oksuz and A. Mukhopadhyay, *A survey on shape-constraint deep learning for medical image segmentation*, arXiv preprint arXiv:2101.07721, 2021.
24. Diligenti, Michelangelo and S. Roychowdhury and M. Gori, *Integrating prior knowledge into deep learning*, 16th IEEE International Conference on Machine Learning and Applications, pp. 920–923, 2017.
25. Safar, Simon and M. Yang, *Learning shape priors for object segmentation via neural networks*, IEEE International Conference on Image Processing, pp. 1835–1839, 2015.
26. Ronneberger, O. and P. Fischer and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, vol. 9351, pp. 234–241, 2015.
27. Shelhamer, E. and Long, J. and Darrell, T., *Fully Convolutional Networks for Semantic Segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 39 pp. 640–651, 2017.
28. Guo, Y. and Liu, Y. and Georgiou, T. and Lew, M. S., *A review of semantic segmentation using deep neural networks*, International Journal of Multimedia Information Retrieval, vol. 7, pp. 87–93, 2018.
29. Wu, Y. and He, K., *Group normalization*, European Conference on Computer Vision (ECCV), pp. 3–19, 2018.
30. Dalca, Adrian V. and John Guttag and Mert R. Sabuncu, *Anatomical priors in convolutional networks for unsupervised biomedical segmentation*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9290–9299, 2018.
31. Chen, Xu and others, *Learning active contour models for medical image segmentation*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11632–11640, 2019.
32. Leclerc, Sarah and others, *Deep learning for segmentation using an open large-scale dataset in 2D echocardiography*, IEEE Transactions on Medical Imaging, vol. 38, no. 9, pp. 2198–2210, 2019.
33. Haque, Intisar Rizwan I. and Jeremiah Neubert, *Deep learning approaches to biomedical image segmentation*, Informatics in Medicine,vol. 18, pp. 100297, 2020.
34. Yongyi Shi, Wenjun Xia, Chuang Niu, Christopher Wiedeman, Ge Wang, *Enabling Competitive Performance of Medical Imaging with Diffusion Model-generated Images without Privacy Leakage*, arXiv preprint arXiv:2301.06604, 2023.
35. Yahia Said, Ahmed A. Alsheikhy, Tawfeeq Shawly, Husam Lahza, *Medical Images Segmentation for Lung Cancer Diagnosis Based on Deep Learning Architectures*, Diagnostics, vol. 18, no. 3, pp. 546, 2023.
36. Alsharqi, M. and Al-Mallah, M. and A. Almansour, *Deep Learning Based Automatic Left Ventricle Segmentation from the Transgastric Short-Axis View on Transesophageal Echocardiography: A Feasibility Study*, Diagnostics, vol. 14, no. 15, pp. 1655, 2024.
37. Selahattin GÜÇLÜ, Durmus ÖZDEMIR, Hamdi Melih SARAOG˘ LU, *A new model for anomaly detection in elbow and finger X-ray images: Proposed parallel DenseNet*, BULLETIN OF THE POLISH ACADEMY OF SCIENCES TECHNICAL SCIENCES, pp. 153233–153233, 2025.
38. Hicham Messaoudia, Ahror Belaidb, Douraied Ben Salemd, Henri Conzed, *Cross-dimensional transfer learning in medical image segmentation with deep learning*, Medical image analysis, vol. 88, pp. 102868, 2023.
39. Yan Xu, Rixiang Quan, Weiting Xu, Yi Huang, Xiaolong Chen, Fengyuan Liu, *Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches*, Bioengineering,vol. 11, no. 10, pp. 10341, 2024.
40. Rahman, A. and Balraj, K. and Ramteke, M. and Rathore, A. S., *Echo-DND: A Dual Noise Diffusion Model for Robust and Precise Left Ventricle Segmentation in Echocardiography*, arXiv preprint arXiv:2506.15166, 2025.
41. Md. Eshmam Rayed, S.M. Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, M.F. Mridha, *Deep learning for medical image segmentation: State-of-the-art advancements and challenges*, Informatics in Medicine Unlocked, pp. 101504, 2024.
42. Malek Ben Alaya, Daniel M. Lang, Benedikt Wiestler, Julia A. Schnabe, Cosmin I. Bercea, *MedEdit: Counterfactual Diffusion-based Image Editing on Brain MRI*, International Workshop on Simulation and Synthesis in Medical Imaging, pp. 167–176, 2024.