Robust segmentation of arterial walls in intravascular ultrasound images using Dual Path U-Net

Keywords: Dual Path UNet DPU-Net Real-time augmentor Intravascular Ultrasound IVUS Segmentation

关键词：双路径 UNet DPU-Net 实时增强器血管内超声 IVUS 分割

A Fully Convolutional Network (FCN) based deep architecture called Dual Path U-Net (DPU-Net) is proposed for automatic segmentation of the lumen and media-adventitia in IntraVascular UltraSound (IVUS) frames, which is crucial for diagnosis of many cardiovascular diseases and also for facilitating 3D reconstructions of human arteries.

One of the most prevalent problems in medical image analysis is the lack of training data. To overcome this limitation, we propose a twofold solution.

First, we introduce a deep architecture that is able to learn using a small number of training images and still achieves a high degree of generalization ability.

Second, we strengthen the proposed DPU-Net by having a real-time augmentor control the image augmentation process. Our real-time augmentor contains specially-designed operations that simulate three types of IVUS artifacts and integrate them into the training images.

We exhaustively assessed our twofold contribution over Balocco’s standard publicly available IVUS 20 MHz and 40 MHz B-mode dataset, which contain 109 training image, 326 test images and 19 training images, 59 test images, respectively.

Models are trained from scratch with the training images provided and evaluated with two commonly used metrics in the IVUS segmentation literature, namely Jaccard Measure (JM) and Hausdorﬀ Distance (HD).

Experimental results show that DPU-Net achieves 0.87 JM, 0.82 mm HD and 0.86 JM, 1.07 mm HD over 40 MHz dataset for segmenting the lumen and the media, respectively.

Also, DPU-Net achieves 0.90 JM, 0.25 mm HD and 0.92 JM, 0.30 mm HD over 20 MHz images for segmenting the lumen and the media, respectively.

In addition, DPU-Net outperforms existing methods by 8–15% in terms of HD distance. DPU- Net also shows a strong generalization property for predicting images in the test sets that contain a signiﬁcant amount of major artifacts such as bifurcations, shadows, and side branches that are not common in the training set.

Furthermore, DPU-Net runs within 0.03 s to segment each frame with a single modern GPU (Nvidia GTX 1080).

The proposed work leverages modern deep learning-based method for segmentation of lumen and the media vessel walls in both 20 MHz and 40 MHz IVUS B-mode images and achieves state-of-the-art results without any manual intervention.

The code is available online at <https://github.com/Kulbear/IVUS-Ultrasonic>.

提出了一种基于全卷积网络 (FCN) 的深度架构，称为双路径 U-Net (DPU-Net)，用于自动分割血管内超声 (IVUS) 帧中的管腔和中外膜，这对许多心血管疾病的诊断至关重要以及促进人体动脉的 3D 重建。

医学图像分析中最普遍的问题之一是缺乏训练数据。为了克服这个限制，我们提出了双重解决方案。

首先，我们引入了一种深度架构，该架构能够使用少量训练图像进行学习，并且仍然具有高度的泛化能力。

其次，我们通过实时增强器控制图像增强过程来增强所提出的 DPU-Net。我们的实时增强器包含专门设计的操作，可模拟三种类型的 IVUS 伪影并将它们集成到训练图像中。

我们详尽地评估了我们对 Balocco 标准公开可用的 IVUS 20 MHz 和 40 MHz B 模式数据集的双重贡献，这些数据集分别包含 109 个训练图像、326 个测试图像和 19 个训练图像、59 个测试图像。

使用提供的训练图像从头开始训练模型，并使用 IVUS 分割文献中的两个常用指标进行评估，即 Jaccard Measure (JM) 和 Hausdorff Distance (HD)。

实验结果表明，DPU-Net 在 40 MHz 数据集上分别实现了 0.87 JM、0.82 mm HD 和 0.86 JM、1.07 mm HD，用于分割流明和媒体。

此外，DPU-Net 在 20 MHz 图像上分别实现了 0.90 JM、0.25 mm HD 和 0.92 JM、0.30 mm HD，用于分割流明和媒体。

此外，DPU-Net 在 HD 距离方面比现有方法高出 8-15%。 DPU-Net 还显示出强大的泛化特性，可用于预测测试集中的图像，这些图像包含大量主要伪像，例如训练集中不常见的分叉、阴影和侧分支。

此外，DPU-Net 在 0.03 秒内运行以使用单个现代 GPU（Nvidia GTX 1080）分割每一帧。

拟议的工作利用基于现代深度学习的方法对 20 MHz 和 40 MHz IVUS B 模式图像中的管腔和介质血管壁进行分割，并在没有任何人工干预的情况下实现了最先进的结果。

该代码可在 <https://github.com/Kulbear/IVUS-Ultrasonic> 在线获得。

1. Introduction

Cardiovascular diseases, due to their high incidence, high mortality and irreversible after-eﬀect, require accurate/fast detection and treat- ment. IntraVascular UltraSound (IVUS) is one of the imaging modalities that is commonly used to assist medical staﬀ and helps them diagnose cardiovascular diseases. IVUS provides the medical experts with an inside-out view of the coronary artery. To acquire the IVUS frames, a catheter that carries an ultrasound emitter is inserted into the coronary artery to provide a cross-sectional tomographic view of the artery. Although IVUS frames allow assessing the vessel morphology [1], the coronary images acquired are not easy for the human eyes to interpret.

No doubt that automatically and accurately labeling (segment) vessel walls from IVUS frames would be beneﬁcial for many applications. For instance, medical staﬀ can beneﬁt from the automatic labeling of the boundaries of the circular layers to diagnose cardiovascular diseases. Therefore, segmentation of the acquired IVUS images has important clinical implications even though it has always been a challenging task since IVUS images usually contain signiﬁcant imaging artifacts. In particular, an accurate separation of the interior (lumen) and exterior (media) vessel walls in IVUS images plays a critical role in creating precise 3D reconstructions of the artery and also diagnose cardiovas- cular diseases where such a quantitative measurement of the coronary artery boundary can aﬀect clinical decisions.

心血管疾病由于其高发病率、高死亡率和不可逆的后遗症，需要准确/快速的检测和治疗。 血管内超声 (IVUS) 是一种成像方式，常用于协助医务人员并帮助他们诊断心血管疾病。 IVUS 为医学专家提供了一个由内而外的冠状动脉视图。 为了获取 IVUS 帧，将携带超声发射器的导管插入冠状动脉，以提供动脉的横截面断层扫描视图。 尽管 IVUS 帧允许评估血管形态 [1]，但获取的冠状动脉图像对于人眼来说并不容易解释。

毫无疑问，从 IVUS 框架自动准确地标记（分段）血管壁对许多应用都是有益的。 例如，医务人员可以受益于圆形层边界的自动标记来诊断心血管疾病。 因此，获得的 IVUS 图像的分割具有重要的临床意义，尽管它一直是一项具有挑战性的任务，因为 IVUS 图像通常包含显着的成像伪影。 特别是，IVUS 图像中内部（腔）和外部（介质）血管壁的准确分离在创建精确的动脉 3D 重建和诊断心血管疾病方面起着至关重要的作用，其中冠状动脉的这种定量测量 边界会影响临床决策。

The segmentation of IVUS images is a well-investigated problem from a conventional perspective where numerous ideas and approaches of computer vision and image processing have been used [2–8]. Older methods used various strategies such as shape and intensity priors[6], gradient vector ﬂow in a nonparametric energy function [5], para- metric deformable models with probabilistic cost functions [4,2],or even the radio frequency signals [3] to segment lumen and media. In recent years, a signiﬁcant amount of research has been conducted and supported by successful experimental results. In [7,8], the authors leverage a type of region detector called Extremal Regions of Extremum Level (EREL) [9,10] to cluster the regions of interest, namely, the lumen and media. In order to segment the arterial walls in IVUS frames, Yang et al. [11] and Kim et al. [12], proposed two derivatives of the well- known U-Net [13] that are based on the concept of deep convolutional neural networks. Convolutional Neural Networks (CNNs) play an im- portant role in visual image recognition since 2012 following the suc- cess of AlexNet in the ImageNet competition [14]. Semantic segmen- tation is one of the most active research ﬁelds in computer vision and image processing that has been a subject of a large volume of research, such as in[15–19]. A very popular architecture was proposed in [16] that attempted to transfer the knowledge learned from image classiﬁ- cation tasks to semantic segmentation by making the network archi- tecture in a fully convolutional fashion. Since then, Fully Convolutional Networks (FCNs) have been widely used to solve the problem of making pixel-level dense predictions.

从传统的角度来看，IVUS 图像的分割是一个经过充分研究的问题，其中使用了许多计算机视觉和图像处理的想法和方法 [2-8]。较旧的方法使用各种策略，例如形状和强度先验[6]、非参数能量函数中的梯度矢量流 [5]、具有概率成本函数的参数可变形模型 [4,2]，甚至射频信号 [3] ] 来分割管腔和介质。近年来，大量的研究得到了成功的实验结果的支持。在 [7,8] 中，作者利用一种称为极值水平极值区域 (EREL) [9,10] 的区域检测器来聚类感兴趣的区域，即管腔和介质。为了在 IVUS 框架中分割动脉壁，Yang 等人。 [11] 和 Kim 等人。 [12]，提出了众所周知的 U-Net [13] 的两个衍生物，它们基于深度卷积神经网络的概念。自 2012 年 AlexNet 在 ImageNet 竞赛 [14] 中取得成功之后，卷积神经网络 (CNN) 在视觉图像识别中发挥着重要作用。语义分割是计算机视觉和图像处理中最活跃的研究领域之一，已成为大量研究的主题，例如 [15-19]。 [16] 中提出了一种非常流行的架构，它试图通过以完全卷积的方式构建网络架构，将从图像分类任务中学到的知识转移到语义分割中。从那时起，全卷积网络（FCN）被广泛用于解决像素级密集预测的问题。

Several years have passed after the emergence of deep learning [14] and although numerous deep learning based methods have dominated almost every ﬁeld of study related to computer vision and medical image analysis, not many deep learning based approaches have been used in segmentation of intravascular ultrasound frames. One of the main reasons why only a small number of studies have been conducted in this ﬁeld is the lack of suﬃcient training data. Many deep archi- tectures require a very large number of training examples in order to achieve high quality generalization abilities. However, the available IVUS datasets contain a very small number of training images. For example, Balocco’s dataset [1] A contains only 19 training images. Having a small number of training images signiﬁcantly decreases the perfor- mance of a deep model especially in IVUS segmentation and since most of the state-of-the-art pre-trained models trained on natural photo images [14,20], transfer learning cannot be a valid option for IVUS segmentation and hence, the model needs to be trained from scratch.

深度学习 [14] 出现后几年过去了，尽管许多基于深度学习的方法几乎主导了与计算机视觉和医学图像分析相关的每个研究领域，但基于深度学习的方法用于血管内超声分割的并不多帧。在这个领域只进行了少量研究的主要原因之一是缺乏足够的训练数据。许多深层架构需要非常大量的训练示例才能实现高质量的泛化能力。然而，可用的 IVUS 数据集包含非常少量的训练图像。例如，Balocco 的数据集 [1] A 仅包含 19 个训练图像。拥有少量训练图像会显着降低深度模型的性能，尤其是在 IVUS 分割中，并且由于大多数最先进的预训练模型都是在自然照片图像上训练的 [14,20]，迁移学习不能作为 IVUS 分割的有效选项，因此需要从头开始训练模型。

Several years have passed after the emergence of deep learning [14] and although numerous deep learning based methods have dominated almost every ﬁeld of study related to computer vision and medical image analysis, not many deep learning based approaches have been used in segmentation of intravascular ultrasound frames. One of the main reasons why only a small number of studies have been conducted in this ﬁeld is the lack of suﬃcient training data. Many deep archi- tectures require a very large number of training examples in order to achieve high quality generalization abilities. However, the available IVUS datasets contain a very small number of training images. For example, Balocco’s dataset [1] A contains only 19 training images. Having a small number of training images signiﬁcantly decreases the perfor- mance of a deep model especially in IVUS segmentation and since most of the state-of-the-art pre-trained models trained on natural photo images [14,20], transfer learning cannot be a valid option for IVUS segmentation and hence, the model needs to be trained from scratch.

为了克服上述问题，我们在本文中提出了一种深度架构，它不仅能够比当前的深度模型更好地泛化（在少量图像上训练后），而且不需要初始化一个预训练的模型。 DPU-Net，我们提出的全卷积网络是基于 UNet 架构构建的，是 IVUSNet [11] 的改进扩展，可以自动描绘管腔和媒体血管壁的边界。我们通过将所提出的 DPU-Net 与在相同数量的帧上训练相同时间的其他两种最先进的架构进行比较来评估其泛化能力，即 SegNet [21] 和 UNet [ 13]。我们的结果表明，DPU-Net 在由 Jaccard Measure (JM) 表示的分割精度方面优于它们两者，或者称为联合交集 (IoU) 和 Hausdorff 距离 (HD)。 DPU-Net 可以使用少量图像进行训练并实现更好的泛化，这一事实使其成为各种医学图像分析问题的合适选择。

In addition to the proposed DPU-Net, we introduce several aug- mentation operations that are specially designed to alleviate the eﬀects of three common IVUS artifacts, namely shadow, side vessel and bifurcations. These operations are included in our real-time augmentation framework that is able to generate augmented images for training as quickly as requested.

除了提议的 DPU-Net 之外，我们还引入了几种增强操作，这些操作专门用于减轻三种常见 IVUS 伪影的影响，即阴影、侧血管和分叉。 这些操作包含在我们的实时增强框架中，该框架能够根据要求尽快生成用于训练的增强图像。

We evaluated the proposed work on the two test sets of a publicly available IVUS B-mode benchmark dataset [1]. These sets not only contain a small number of training images but also have test sets generated based on diﬀerent distributions of the artifacts than the training set. The evaluation results reveal that, the proposed DPU-Net enhanced by the real-time augmentor outperforms every existing state-of-the-art approach. Also, due to our IVUS-based augmentation operations, the accuracy results for images contaminated with artifacts are far superior to other automatic methods. Our work is an end-to-end method that requires no human intervention.

我们在公开可用的 IVUS B 模式基准数据集的两个测试集上评估了拟议的工作 [1]。 这些集不仅包含少量的训练图像，而且还具有基于与训练集不同的工件分布生成的测试集。 评估结果表明，实时增强器增强的拟议 DPU-Net 优于所有现有的最先进方法。 此外，由于我们基于 IVUS 的增强操作，被伪影污染的图像的准确性结果远远优于其他自动方法。 我们的工作是一种端到端的方法，不需要人工干预。

The rest of the paper is organized as the follows: Section 3 provides a detailed explanation of our proposed architecture. Section 2.1 states how we build the augmentation pipeline. We also demonstrate multiple experiments that reinforce our contribution in Section 4. Finally, we conclude our study in Section 6.

论文的其余部分组织如下：第 3 节详细解释了我们提出的架构。 2.1 节说明了我们如何构建增强管道。 我们还展示了多个实验，这些实验加强了我们在第 4 节中的贡献。最后，我们在第 6 节中总结了我们的研究。

1. Datasets

We exploits a publicly available IVUS dataset [1] to validate DPU- Net. This dataset is designed to be useful in diﬀerent approaches that might need a single frame or a multi-frame dataset [1]. Therefore, there are two datasets available which were obtained with diﬀerent two ultrasound frequencies including 20 MHz and 40 MHz. Note that the two datasets are obtained and stored to diﬀerent resolutions, which are 384- by-384 for the 20 MHz dataset and 512-by-512 for the 40 MHz dataset. The 20 MHz dataset contains two sets (train and test) of IVUS gated frames using a full pullback at the end-diastolic cardiac phase from 10 patients. Dataset frames were manually annotated by four clinical ex- perts. Speciﬁcally, two of them repeated the task one week after the ﬁrst marking [1]. There are 109 and 326 IVUS frames in the training and testing sets, respectively. Also, the test set contains a large number of IVUS artifacts including bifurcation (44 frames), side vessel (93 frames), and shadow (96 frames) artifacts. The remaining 143 frames do not contain any artifacts except for plaque.

我们利用公开可用的 IVUS 数据集 [1] 来验证 DPU-Net。该数据集旨在用于可能需要单帧或多帧数据集的不同方法 [1]。因此，有两个可用的数据集是用不同的两个超声频率（包括 20 MHz 和 40 MHz）获得的。请注意，这两个数据集以不同的分辨率获得并存储，20 MHz 数据集为 384×384，40 MHz 数据集为 512×512。 20 MHz 数据集包含两组（训练和测试）IVUS 门控帧，使用来自 10 名患者在心脏舒张末期完全回拉。数据集框架由四位临床专家手动注释。具体而言，其中两人在第一次标记后一周重复该任务 [1]。训练和测试集中分别有 109 和 326 个 IVUS 帧。此外，测试集包含大量 IVUS 伪影，包括分叉（44 帧）、侧血管（93 帧）和阴影（96 帧）伪影。剩余的 143 帧不包含除斑块之外的任何伪影。

The 40 MHz dataset also comes with two predeﬁned sets (train and test) both contain annotated ground truth. The train and test sets consist of 19 and 59 IVUS frames, respectively. Having small number of training images signiﬁcantly increases the diﬃculty of training a well-generalized deep model to produce results at a reasonable level. We address the problem of having a limited training examples with DPU-Net along with major data augmentations as mentioned in Section 2.1.

40 MHz 数据集还带有两个预定义集（训练和测试），它们都包含带注释的地面实况。 训练集和测试集分别由 19 个和 59 个 IVUS 帧组成。 训练图像数量较少会显着增加训练泛化深度模型以产生合理水平的结果的难度。 我们解决了使用 DPU-Net 以及主要数据增强的有限训练示例的问题，如第 2.1 节所述。

* 1. Augmentation

The idea of data augmentation is to make an eﬀort to ﬁll the missing and unseen values by augmenting the observed data, and has been studied in machine learning for several decades [22]. The ‘craving for data’ however was felt more strongly after the emergence of large deep models, especially in computer vision and image analysis where the great dimensionality of the problem’s input domain creates signiﬁcant variability in the latent space. Therefore, once one decides to employ a deep model, augmenting the training set becomes an inevitable necessity, especially when we do not have a big and diverse training set which is always the case in medical image analysis applications. Most of medical datasets contain a relatively small number of images due to a range of diﬃculties in acquiring medical images. For example, the Balocco datasets A and B [1] contain only 19 and 109 training images, respectively, which highlights the need for artiﬁcially augmenting the training sets. In order to best utilize the available information in the training set, we designed a real-time augmenter class capable of si- multaneously transforming and warping the training images and also several new operations speciﬁcally designed for IVUS images.

数据增强的想法是通过增强观察到的数据来努力填充缺失和看不见的值，并且在机器学习中已经研究了几十年 [22]。然而，在大型深度模型出现后，“对数据的渴望”变得更加强烈，尤其是在计算机视觉和图像分析中，问题的输入域的大维度会在潜在空间中产生显着的可变性。因此，一旦决定采用深度模型，扩充训练集就成为必然，尤其是当我们没有庞大而多样的训练集时，这在医学图像分析应用中总是如此。由于获取医学图像的一系列困难，大多数医学数据集包含的图像数量相对较少。例如，Balocco 数据集 A 和 B [1] 分别仅包含 19 和 109 个训练图像，这突出了人为增加训练集的必要性。为了最好地利用训练集中的可用信息，我们设计了一个实时增强类，能够同时转换和扭曲训练图像，以及一些专门为 IVUS 图像设计的新操作。

* + 1. Augmentation operations

Various types (operations) of augmentations have been tried in the literature to create new training images: Elastic distortions are used in visual document analysis [23], RGB color shifting, translation, scale, horizontal shearing, horizontal ﬂipping, and rotation transformations are examples that have been considered in [24–26,14,20,27,17,13].In medical applications, on the other hand, not all kinds of augmentations might be useful. In fact, the type of medical image augmentation (transformation) strongly depends on the application. Performing redundant augmentation thus increases the training time without fulﬁlling noticeable accuracy gains [28]. For example translating (shifting) IVUS frames does not sound plausible since the catheter is always located at the center of the images. Thus, the model will never see an IVUS frame with a catheter located at somewhere else rather than the center of the frame. In contrast, rotations and ﬂipping seem reasonable because it is highly probable that the catheter is rotated or moved inside the vessels. Therefore, in this section we explore the effects of various types of the existing transformations on the accuracy of the trained models. We also introduce three new ﬁlters that have been exclusively designed for augmenting IVUS frames. We show that applying a combination of the existing transformations and our designed masks on the training frames can grant the model a greater ability to generalize and thus improve the accuracy of the ﬁnal segmentation.

文献中已经尝试了各种类型的增强（操作）来创建新的训练图像：弹性失真用于视觉文档分析 [23]，RGB 色移、平移、缩放、水平剪切、水平翻转和旋转变换是示例[24-26,14,20,27,17,13] 中已经考虑过这些。另一方面，在医学应用中，并非所有类型的增强都可能有用。事实上，医学图像增强（转换）的类型在很大程度上取决于应用。因此，执行冗余增强会增加训练时间，而不会实现显着的准确度增益 [28]。例如，平移（移动）IVUS 帧听起来并不合理，因为导管始终位于图像的中心。因此，模型永远不会看到导管位于其他地方而不是框架中心的 IVUS 框架。相比之下，旋转和翻转似乎是合理的，因为导管很有可能在血管内旋转或移动。因此，在本节中，我们将探讨各种类型的现有转换对训练模型准确性的影响。我们还介绍了三个专为增强 IVUS 帧而设计的新过滤器。我们表明，在训练帧上应用现有转换和我们设计的掩码的组合可以赋予模型更大的泛化能力，从而提高最终分割的准确性。

* + 1. Proposed augmentation operations

Designing particular strategies to alleviate the eﬀects of ultrasound artifacts has been practiced many times in various conventional studies. In [29], a circular cursor was dragged to cover the catheter artifact. In [30], a preprocessing step was added to the method to clear away the motion artifacts. Many other studies has also made an eﬀort to include the artifact detection steps [6,31–33]. In the proposed study, not only do we use the aforementioned common augmentation operations such as rotation and rescaling, but we have also devised three diﬀerent types of augmentations to mimic the common IVUS artifacts, namely bifurcation, side vessel, and shadow. These operations are included into our augmentation pipeline to increase the robustness of the DPU-Net architecture in dealing with several types of IVUS artifacts.

在各种传统研究中已经多次实践了设计特定策略以减轻超声伪影的影响。 在 [29] 中，拖动圆形光标以覆盖导管伪影。 在[30]中，该方法添加了一个预处理步骤以清除运动伪影。 许多其他研究也努力包括伪影检测步骤 [6,31-33]。 在拟议的研究中，我们不仅使用上述常见的增强操作，如旋转和重新缩放，而且我们还设计了三种不同类型的增强来模拟常见的 IVUS 伪影，即分叉、侧血管和阴影。 这些操作包含在我们的增强管道中，以提高 DPU-Net 架构在处理多种类型的 IVUS 工件时的稳健性。

In order to mimic bifurcation, side vessel and shadow artifacts, we have designed three diﬀerent masks with dimensions equal to the size of the training images. Each of these masks were rotated oﬄine 360 times.2 Fig. 1 illustrates all the masks. The masks designed can then be multiplied by the training images at the augmentation time. One advantage of using masks during augmentation is its low computational cost, since applying the masks only needs the multiplication operation and no warping or convolution is required. Another beneﬁt of using this type of augmentation is that we can apply several operations to one image at the same time. Considering that the masks designed to mimic shadow, side vessel, and bifurcation are called shadow mask, side vessel mask, and bifurcation mask and are denoted by SH,SV, BF a new augmented image (I∗) can be obtained as follows: 

为了模拟分叉、侧血管和阴影伪影，我们设计了三个不同的掩码，其尺寸等于训练图像的大小。 这些掩码中的每一个都旋转了 360 次。2 图 1 说明了所有掩码。 然后可以在增强时将设计的掩码与训练图像相乘。 在增强期间使用掩码的一个优点是其计算成本低，因为应用掩码只需要乘法运算，不需要变形或卷积。 使用这种类型的增强的另一个好处是我们可以同时对一个图像应用多个操作。 考虑到模拟阴影、侧血管和分叉的掩膜被称为阴影掩膜、侧血管掩膜和分叉掩膜，并用 SH、SV、BF 表示，一个新的增强图像 (I\*) 可以得到如下： 

where I denotes an original training image and \* represents an ele-ment-wise multiplication. It should be noted that other operations based on each specific mask can be defined as well such as  for only applying the shadow mask, for only applying the side vessel mask,for only applying the bifurcation mask, for applying the side vessel and bifurcation mask together and so forth. Fig. 2 shows three of these masks randomly selected and the resulting images after applying them to one of the 20 MHz training frames of the Balocco’s dataset [1].

其中 I 表示原始训练图像，\* 表示元素乘法。 需要注意的是，也可以定义基于每个特定掩膜的其他操作，例如  仅应用阴影掩膜，  仅应用侧脉掩膜，  仅应用分叉掩膜， 应用侧脉 和分叉掩码一起等等。 图 2 显示了随机选择的三个掩码以及将它们应用于 Balocco 数据集 [1] 的 20 MHz 训练帧之一后的结果图像。

* + 1. Real-time augmenter

From a practical point of view, most types of augmentations are done by applying a specific transformation (such as similarity or affine) on the image. This shows that the image augmentation is inherently a computationally expensive task since a costly warping procedure (that includes interpolation) needs to be performed after applying the transformation on the training images. The cumbersome process of augmenting new images signiﬁcantly increases the training time. Thus, in order to prevent the training process from being delayed, the new images are augmented in either an oﬄine or online form. Basically, if the augmented training set consists of a small number of images and enough disk space is available, the training set can be augmented and saved on disk before the training process is started. This way of creating new images is called oﬄine augmentation. However, due to the high disk space requirement, oﬄine augmentation is not practical in many applications that tend to train a large model over large datasets such as ImageNet [34] or Microsoft coco [35]. A simple remedy for this is to let the CPU augment the images while the training process is being performed on GPU which is called online augmentation. This technique has been adopted during training in many of the state-of-the-art deep learning architectures [14,26]. But recent advances in GPU hardware and software has made it possible to feed forward and back-propagate the images incredibly fast and hence even a small computation on CPU will delay the training process. Therefore, in our study, we propose a real-time image augmenter module, a process that runs on the background and controls a number of parallel processes designed to augment the required images. The training process (runs on GPU) can instantly read an image from a buﬀer that is always full of augmented images. As soon as an image is read from the buﬀer, the processes that are running on the background in each CPU core, augment another image and place it into an empty location. Fig. 3 illustrates how our proposed real-time augmentation pipeline works.

从实用的角度来看，大多数类型的增强都是通过对图像应用特定的变换（例如相似性或仿射）来完成的。这表明图像增强本质上是一项计算成本很高的任务，因为在对训练图像应用变换之后需要执行昂贵的变形过程（包括插值）。增加新图像的繁琐过程显着增加了训练时间。因此，为了防止训练过程被延迟，新图像以离线或在线形式增强。基本上，如果增强训练集由少量图像组成并且有足够的可用磁盘空间，则可以在训练过程开始之前增强训练集并将其保存在磁盘上。这种创建新图像的方式称为 oﬄine 增强。然而，由于高磁盘空间要求，离线增强在许多倾向于在大型数据集（如 ImageNet [34] 或 Microsoft coco [35]）上训练大型模型的应用中并不实用。一个简单的补救措施是让 CPU 增强图像，同时在 GPU 上执行训练过程，这称为在线增强。该技术已在许多最先进的深度学习架构的训练过程中采用 [14,26]。但是 GPU 硬件和软件的最新进展使得图像的前馈和后向传播速度非常快，因此即使在 CPU 上进行少量计算也会延迟训练过程。因此，在我们的研究中，我们提出了一个实时图像增强器模块，该过程在后台运行并控制许多旨在增强所需图像的并行进程。训练过程（在 GPU 上运行）可以立即从始终充满增强图像的缓冲区中读取图像。一旦从缓冲区中读取图像，在每个 CPU 内核的后台运行的进程就会扩充另一个图像并将其放置到一个空位置。图 3 说明了我们提出的实时增强管道是如何工作的。

1. Proposed method

Since [14], deep learning (DL) approaches have been shown to achieve superior performance on many vision recognition tasks. In particular, Convolutional Neural Networks (CNN) and its derivatives outperform conventional methods in almost all visual recognition tasks that can be treated as supervised learning tasks [20,18,21,27]. In this section, we ﬁrst introduce the popular IVUS B-mode dataset we used to train and validate the proposed work. Then, we propose a Dual Path UNet (DPU-Net), which produces a binary prediction mask for either the lumen or media area to delineate the vessel wall, with detailed explanation in terms of its intuition and design.

自 [14] 以来，深度学习 (DL) 方法已被证明可以在许多视觉识别任务上实现卓越的性能。 特别是，卷积神经网络 (CNN) 及其衍生物在几乎所有可以被视为监督学习任务的视觉识别任务中都优于传统方法 [20,18,21,27]。 在本节中，我们首先介绍用于训练和验证建议工作的流行 IVUS B 模式数据集。 然后，我们提出了一个双路径 UNet (DPU-Net)，它为管腔或介质区域生成二进制预测掩码以描绘血管壁，并在其直觉和设计方面进行了详细解释。

* 1. Dual Path UNet

Dual Path UNet (DPU-Net) is designed incorporating intuitions from human perception and lessons learned from existing popular fully convolutional network (FCN) [16,21,36] and segmentation-purposed refinements for FCNs. We adopt U-Net [13] as the base architecture of our proposed work according to the overall architecture design. As in popular 1-stage architectures, there are two major components for DPU-Net:

Dual Path UNet (DPU-Net) 的设计结合了人类感知的直觉和从现有流行的完全卷积网络 (FCN) [16,21,36] 中吸取的经验教训以及 FCN 的分段改进。 根据整体架构设计，我们采用 U-Net [13] 作为我们提出的工作的基础架构。 与流行的 1-stage 架构一样，DPU-Net 有两个主要组件：

1. An encoder network that can downsample and process the input to produce a low-resolution deep feature map.

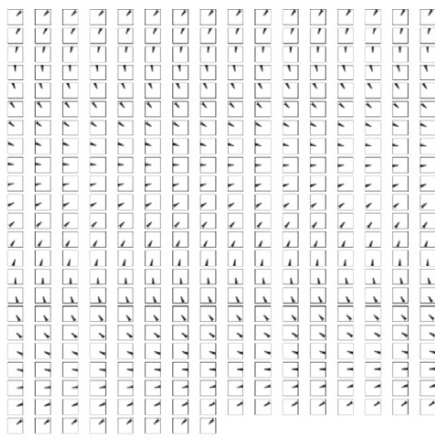
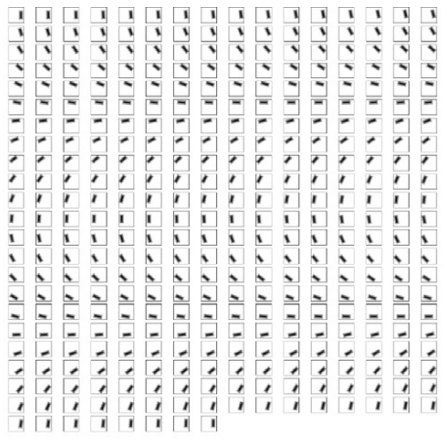
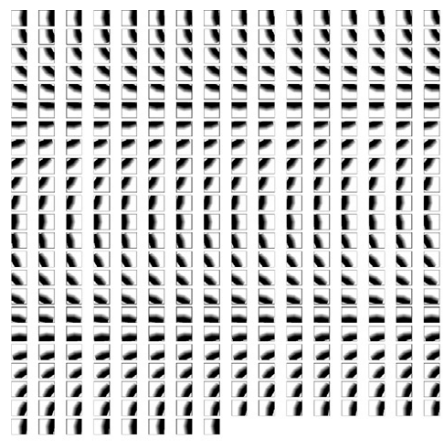
2. A decoder network that can restore the resolution of the deep feature map output by the encoder network towards the original size.

1. 一个编码器网络，可以对输入进行下采样和处理以生成低分辨率的深度特征图。

2. 一个解码器网络，可以将编码器网络输出的深度特征图的分辨率恢复到原始大小。

Due to the limited computational power and memory we had, we ﬁrst downsample the input images before we feed them to the deep model and upsample the prediction result at the end of the model. This procedure will be explained later in this section. The upsampled output feature map is sent to one more convolutional layer followed by a sigmoid activation to produce the ﬁnal result.

由于计算能力和内存有限，我们首先对输入图像进行下采样，然后再将它们提供给深度模型，并在模型结束时对预测结果进行上采样。 本节稍后将解释此过程。 上采样的输出特征图被发送到另一个卷积层，然后是 sigmoid 激活以产生最终结果。



1. (b) (c)

Fig. 1. 360 designed masks for emulating the common artifacts of IVUS frames.

1. Designed shadow masks (SH).
2. Designed side vessel masks (SV).
3. Designed bifurcation masks (BF)

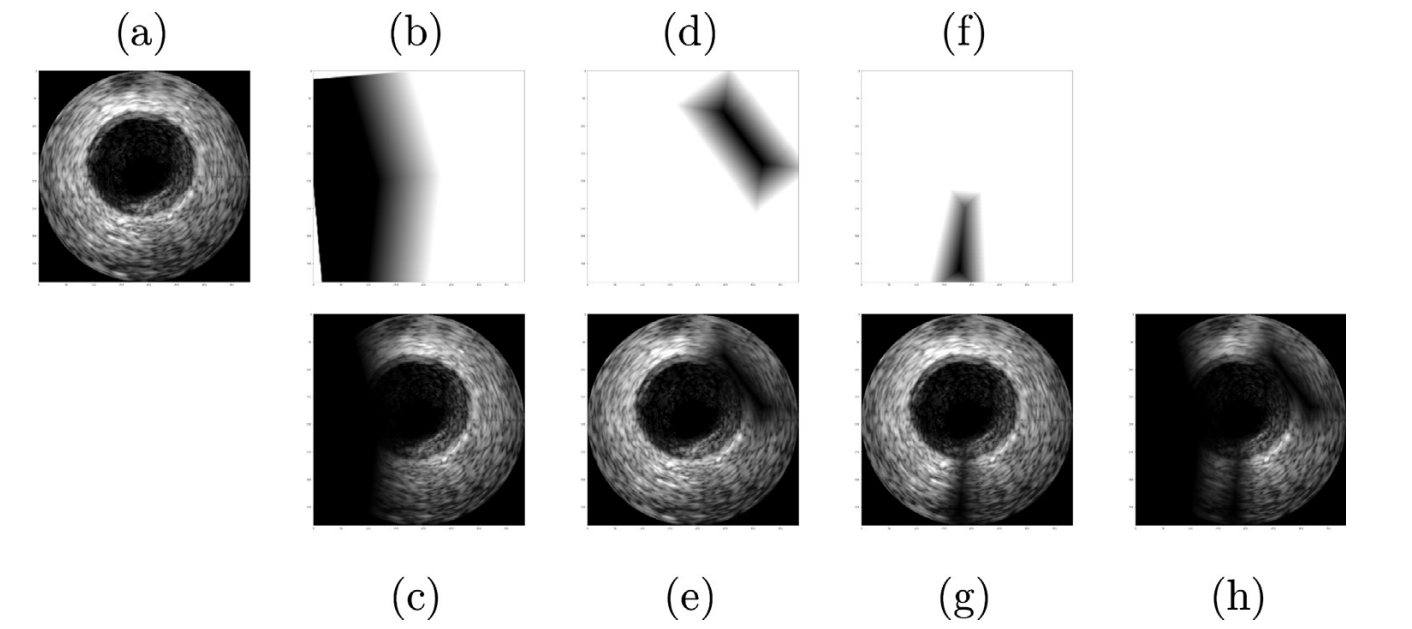


Fig. 2. The proposed artifact simulation operations to mimic the three common IVUS artifacts.

1. An original training Image .
2. Designed SHadow mask .
3. The image obtained after multiplying the shadow mask (b) by the image .
4. Designed Side Vessel mask .
5. The result of multiplying the image by the side vessel mask .
6. Designed BiFurcation mask.
7. The result of multiplying the image by the bifurcation mask.
8. The image obtained after applying all of the masks designed to the original image at the same time .

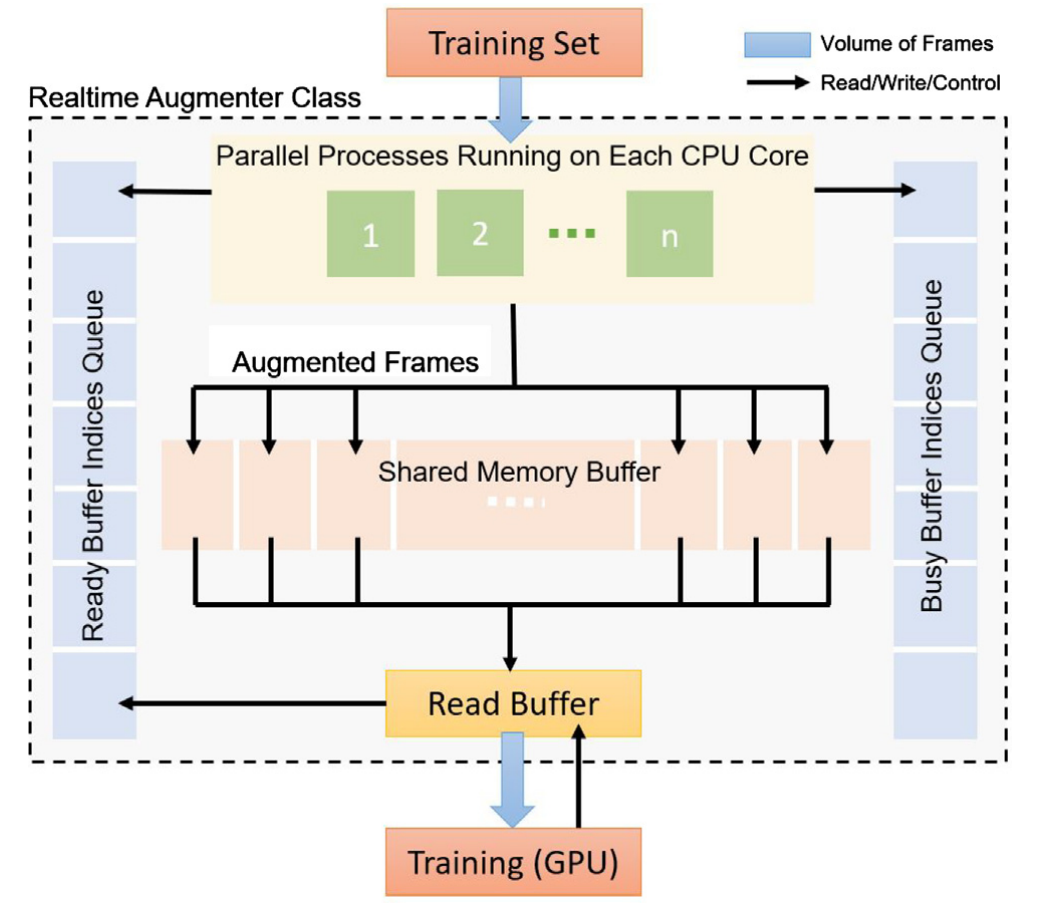


Fig. 3. The proposed real-time augmente

The encoder network contains 6 encoding blocks whereas the de- coder network contains 5 decoding blocks. Starting from the second block in the network, each block receives the feature map from its previous block; and speciﬁcally for each decoder block, there is an extra skip connection that can help forward information from the encoder network. Skip connections from the encoder network to the decoder network give additional information to help restore the feature map to the original size. Particularly, spatial relations among pixels are preserved as the skip connections actually connect corresponding blocks between the encoder and the decoder. Note that these skip connections can also help reinforce the gradient ﬂow in deep models therefore avoiding the common gradient vanishing problem and speed up the training process [37]. The entire architecture is therefore symmetric as shown in Fig. 4. There are minor diﬀerences among the blocks in the architecture.

编码器网络包含 6 个编码块，而解码器网络包含 5 个解码块。 从网络中的第二个块开始，每个块从其前一个块接收特征图； 特别是对于每个解码器块，有一个额外的跳过连接可以帮助从编码器网络转发信息。 跳过从编码器网络到解码器网络的连接提供额外的信息，以帮助将特征图恢复到原始大小。 特别是，像素之间的空间关系被保留，因为跳跃连接实际上连接了编码器和解码器之间的相应块。 请注意，这些跳跃连接还可以帮助加强深度模型中的梯度流，从而避免常见的梯度消失问题并加快训练过程 [37]。 因此，整个架构是对称的，如图 4 所示。架构中的块之间存在细微差别。

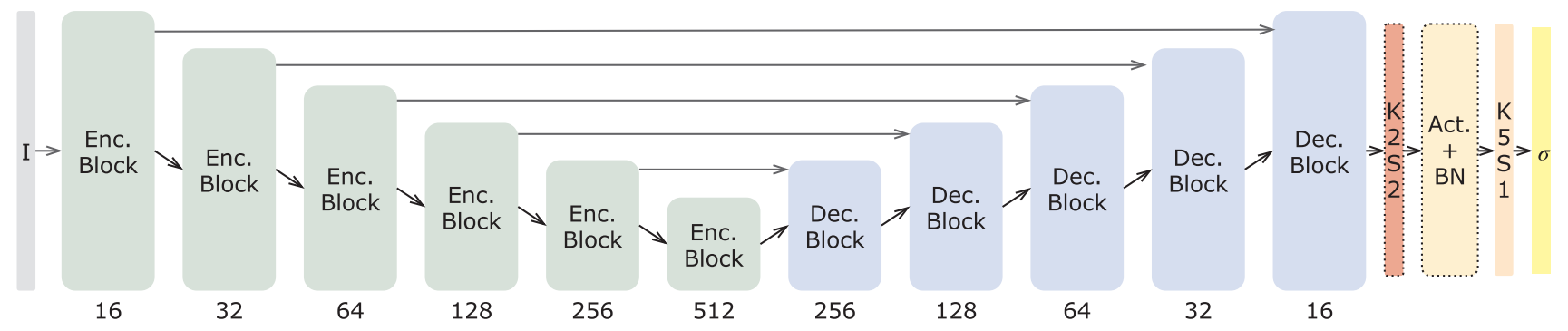


Fig. 4. The DPU-Net architecture. Every convolutional layer in the same block has the same output depth as labeled at the bottom of the block. At the end, the layer in red represents a 2-by-2 transposed convolution (deconvolution) with a stride size of 2, where the layer right before the sigmoid output (labeled with yellow) is a 5-by5 convolution layer. The abbreviations used in the figure mean as follows: “K2S2” means “kernel size 2 and stride size 2”, “BN” means “batch normalization”. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

图 4. DPU-Net 架构。 同一块中的每个卷积层都具有与块底部标记的相同的输出深度。 最后，红色层表示步长为 2 的 2×2 转置卷积（反卷积），其中 sig moid 输出（标有黄色）之前的层是 5-by 5 卷积层。 图中使用的缩写含义如下：“K2S2”表示“kernel size 2 and stride size 2”，“BN”表示“batch normalization”。 （为了解释这个图例中对颜色的引用，读者可以参考本文的网络版本。）

There are two common options for the pooling layer, namely average pooling and max pooling. Max pooling is widely used for downsampling the input feature map obtained from the previous layer. We choose max pooling over average pooling as we consider the IVUS images to be relatively blurry with low resolution. Max pooling forces the network to capture the information in the most activated neuron in a sub-region of the kernel size, but drops other non-signiﬁcant information.

池化层有两种常见的选择，即平均池化和最大池化。 最大池化广泛用于对从前一层获得的输入特征图进行下采样。 我们选择最大池化而不是平均池化，因为我们认为 IVUS 图像在低分辨率下相对模糊。 最大池化迫使网络在内核大小的子区域中捕获最活跃的神经元中的信息，但丢弃其他不重要的信息。

Except for the ﬁrst encoding block, each encoding block contains a downsampling branch that downsamples the received feature map, followed by a two-branch convolution path, as shown in Fig. 5(a). To avoid losing information caused by using max pooling and also to reduce the spatial resolution of the input, we build and expand the downsampling branch by using both a 2-by-2 average pooling layer and a 3-by-3 convolutional layer with a stride of 2. Finally, we concatenate the two outputs together at the depth dimension. This aggregation idea is similar to [27,36].

除了第一个编码块，每个编码块都包含一个下采样分支，对接收到的特征图进行下采样，然后是一个两分支的卷积路径，如图 5（a）所示。 为了避免使用最大池化导致的信息丢失并降低输入的空间分辨率，我们通过使用 2×2 平均池化层和 3×3 卷积层来构建和扩展下采样分支 步幅为 2。最后，我们在深度维度将两个输出连接在一起。 这种聚合思想类似于[27,36]。

After downsampling, the aggregated feature map output by the downsampling branch is passed to two subsequent branches, namely the reﬁning branch and the main branch. First, we follow the design in [21,13] to include a branch with consecutive convolutional layer followed by activation and batch normalization, here we call it main branch.

下采样后，下采样分支输出的聚合特征图被传递到两个后续分支，即精炼分支和主分支。 首先，我们按照 [21,13] 中的设计包含一个具有连续卷积层的分支，然后是激活和批量归一化，这里我们称之为主分支。

A recent trend is to use small kernel size for the feature map re- ﬁnement [17,19], also the concept of networks in network [27] is widely used in the literature. Therefore, we introduce a reﬁning branch that has one convolutional layer with a 3-by-3 kernel size followed by a convolutional layer with a 1-by-1 kernel size to produce similar but reﬁned feature map. A 1-by-1 convolution is able to reﬁne or trim a feature map since it covers only a single pixel without inﬂuence from its neighbors. But over all the depth, this idea is similar to the global average pooling [38] with more learning capacity. In addition, as the ability to capture features at diﬀerent scales are usually desired, we set convolutional layers with a kernel size of 5 in the main branch, compared to the kernel sizes of 3 and 1 in the reﬁning branch. The outputs from the main and reﬁning branches are summed up and passed to the next block and its corresponding decoding block. A critical issue is that deep networks are hard to train due to the gradient vanishing problem. The multi-branch and local networks-in-network architecture not only provide a good local topology but also reinforce the gradient ﬂow to accelerate the training.

最近的一个趋势是使用小内核大小进行特征图细化 [17,19]，网络 [27] 中的网络概念也在文献中广泛使用。因此，我们引入了一个精炼分支，它具有一个 3×3 内核大小的卷积层，后跟一个内核大小为 1×1 的卷积层，以生成相似但经过改进的特征图。 1×1 卷积能够细化或修剪特征图，因为它仅覆盖单个像素而不受其邻居的影响。但在所有深度上，这个想法类似于全局平均池化 [38]，具有更多的学习能力。此外，由于通常需要在不同尺度上捕获特征的能力，我们在主分支中设置了内核大小为 5 的卷积层，相比之下，在精炼分支中的内核大小为 3 和 1。主分支和精炼分支的输出相加并传递到下一个块及其相应的解码块。一个关键问题是，由于梯度消失问题，深度网络难以训练。多分支和本地网络中网络架构不仅提供了良好的本地拓扑，而且还加强了梯度流以加速训练。

Decoding blocks needs a slightly diﬀerent conﬁguration, as shown in Fig. 5(b). Every decoding block receives the feature map from both its previous block and its corresponding encoding block. Only the feature map received from the previous block is upsampled by a 2-by-2 transposed convolution and then concatenated with the feature map from its corresponding encoding block. Note that this concatenated feature map will only be passed to the main branch, where the reﬁning branch handles the upsampled feature map only.

解码块需要稍微不同的配置，如图5（b）所示。 每个解码块都从其前一个块和相应的编码块接收特征图。 只有从前一个块接收到的特征图通过 2×2 转置卷积进行上采样，然后与来自其相应编码块的特征图连接。 请注意，此连接的特征图只会传递到主分支，其中精炼分支仅处理上采样的特征图。

The activation used in the DPU-Net is the Parametric Rectiﬁed Linear Unit (PReLU) [39].



DPU-Net 中使用的激活是参数整流线性单元 (PReLU) [39]。

Compared to the ordinary ReLU activation, PReLU [39] allows a part of the gradients ﬂow through the neuron when it is not activated, whereas ReLU only passes gradients when the neuron is active. As suggested in [39,40], PReLU outperforms ReLU in many benchmarks and also has a more stable performance.

与普通的 ReLU 激活相比，PReLU [39] 允许一部分梯度在未激活时流过神经元，而 ReLU 仅在神经元处于活动状态时通过梯度。 正如 [39,40] 中所建议的，PReLU 在许多基准测试中都优于 ReLU，并且性能也更稳定。

To reduce the computation cost involved during training, we initially downsampled the input image by a factor of 2 (i.e., a 384-by-384 image in the 20 MHz dataset will be resized to a 192-by-192 image). However, the ground truth masks are kept in the original size. The reason why we do not downsample the ground truth masks is to make the model predict a smoother map since if we downsample the binary ground truth masks, errors (especially for pixels around boundary) will be added to the ground truth. Therefore, by not changing the ground truth dimension, we achieve a smoother predictions especially around boundaries. The feature map in all of the available architectures should be in the downsampled size in terms of the width and height dimensions, while we add an extra resize branch that helps restore the feature map to its original size, which is the same as the ground truth mask in width and height.

为了减少训练过程中涉及的计算成本，我们最初将输入图像下采样了 2 倍（即，20 MHz 数据集中的 384 x 384 图像将被调整为 192 x 192 图像）。 但是，ground truth 掩码保持原始大小。 我们不对地面实况掩码进行下采样的原因是为了使模型预测更平滑的地图，因为如果我们对二元地面实况掩码进行下采样，错误（尤其是边界周围的像素）将添加到地面实况中。 因此，通过不改变地面实况维度，我们实现了更平滑的预测，尤其是在边界附近。 所有可用架构中的特征图在宽度和高度维度上都应该是下采样的尺寸，同时我们添加了一个额外的调整大小分支，有助于将特征图恢复到其原始尺寸，这与地面实况相同 掩码的宽度和高度。

Finally, the output feature map from the last decoding block is ﬁrst upsampled by a 2-by-2 transposed convolution layer with a stride size of 2 then reﬁned by a 5-by-5 convolution layer, which is experimentally proved to help improve performance [14]. Thanks to the extra 2-by-2 transposed convolution layer, the obtained ﬁnal outputs now have the exact same size as the original dataset images. As we want DPU-Net to produce binary masks, the activation used after the last convolutional layer is a sigmoid function. Also, it is worth mentioning that the skipconnections between corresponding encoding and decoding blocks add context information for the decoder network and also provide extra gradient ﬂow to the current architecture.

最后，最后一个解码块的输出特征图首先由步长为 2 的 2×2 转置卷积层上采样，然后由 5×5 卷积层细化，实验证明这有助于提高性能 [14]。 由于额外的 2×2 转置卷积层，获得的最终输出现在具有与原始数据集图像完全相同的大小。 由于我们希望 DPU-Net 生成二进制掩码，因此在最后一个卷积层之后使用的激活是一个 sigmoid 函数。 此外，值得一提的是，相应的编码和解码块之间的跳过连接为解码器网络添加了上下文信息，并为当前架构提供了额外的梯度流。

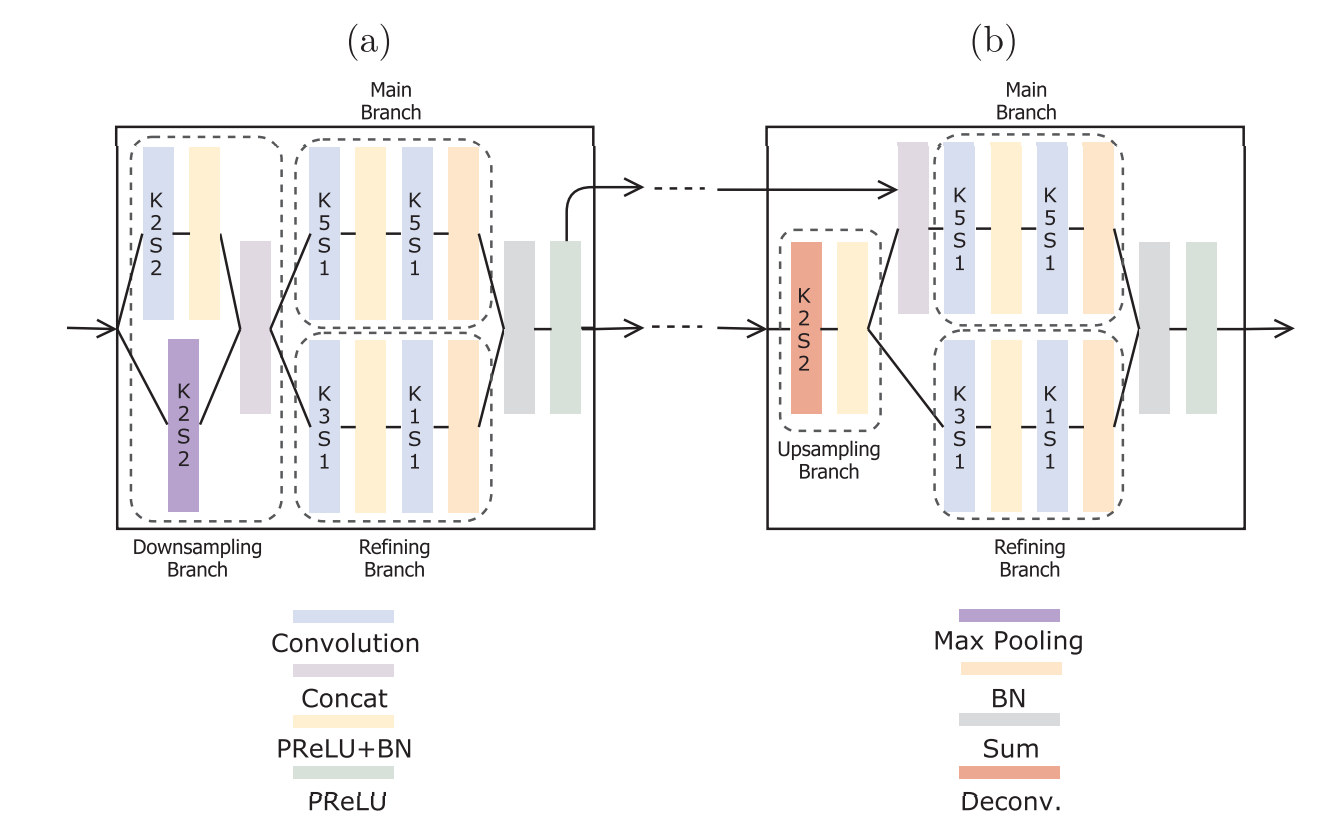


Fig. 5. A detailed illustration on the encoding block and the decoding block. Note that the ﬁrst encoding block does not have the downsampling branch; therefore the main branch and reﬁning branch will directly accept the raw image as the input. (a) An encoding block with downsampling branch, followed by the main branch and the reﬁning branch. (b) A typical decoding block that accepts feature maps from both the previous block and the skip-connection. The abbreviations used in the ﬁgure mean as follows: “K2S2” means “kernel size 2 and stride size 2”, “BN” means “batch normalization.”.

图 5. 编码块和解码块的详细说明。 注意第一个编码块没有下采样分支； 因此主分支和精炼分支将直接接受原始图像作为输入。 (a) 具有下采样分支的编码块，其后是主分支和精炼分支。 (b) 一个典型的解码块，它接受来自前一个块和跳过连接的特征图。 图中使用的缩写表示如下：“K2S2”表示“内核大小2和步幅大小2”，“BN”表示“批量归一化”。

1. Experiments

In this section, we ﬁrst introduce how we setup and train the DPU- Net. Then, we show a detailed investigation on the eﬀectiveness of diﬀerent augmentation techniques for training IVUS segmentation models. In addition, we provide a group of comparison experiments that shows our architecture surpassed the performance of some of the best previous work, namely U-Net [13] and SegNet [21]. Finally, we report the accuracy of the segmentation results by DPU-Net and compare it with existing IVUS segmentation literature.

在本节中，我们首先介绍我们如何设置和训练 DPU-Net。 然后，我们详细研究了用于训练 IVUS 分割模型的不同增强技术的有效性。 此外，我们提供了一组比较实验，表明我们的架构超越了一些以前最好的工作，即 U-Net [13] 和 SegNet [21] 的性能。 最后，我们报告了 DPU-Net 分割结果的准确性，并将其与现有的 IVUS 分割文献进行了比较。

* 1. Experiments setup

The training and evaluation is based on two publicly available IVUS B-mode datasets [1]. Both datasets have been widely used in the IVUS segmentation literature [7,41,29,8]. The two datasets are acquired with diﬀerent ultrasound frequencies, namely, 20 MHz and 40 MHz. The pattern and texture are similar in these two groups of IVUS images. As no oﬃcial validation set was provided in [1], we randomly select a small subset from the augmented images as our validation set during the training to do the 5-fold cross validation. As the result of the 5-fold cross validation, we need to train 5 models based on the 5 diﬀerent training/validation splits. We then use the 5 models to preform inference on their corresponding validations set. After doing this, we should end up with the completed out-of-fold inferenced training set. We evaluated this out-of-fold training set to select the best training conﬁguration of the architecture, namely, the batch size, learning rate, and the depth of each blocks in the architecture. All models are trained end-to-end over only the training set of the two given datasets without involving any other external resources such as extra training images and pretrained model weights.

训练和评估基于两个公开可用的 IVUS B 模式数据集 [1]。这两个数据集都已广泛用于 IVUS 分割文献 [7,41,29,8]。这两个数据集是用不同的超声频率获得的，即 20 MHz 和 40 MHz。这两组IVUS图像的图案和纹理相似。由于 [1] 中没有提供官方验证集，我们在训练期间从增强图像中随机选择一个小的子集作为我们的验证集进行 5 折交叉验证。作为 5 折交叉验证的结果，我们需要基于 5 个不同的训练/验证拆分来训练 5 个模型。然后我们使用 5 个模型对其相应的验证集进行推理。这样做之后，我们应该最终得到完整的折叠推理训练集。我们评估了这个 out-of-fold 训练集以选择架构的最佳训练配置，即批量大小、学习率和架构中每个块的深度。所有模型仅在两个给定数据集的训练集上进行端到端训练，不涉及任何其他外部资源，例如额外的训练图像和预训练的模型权重。

Metrics used for evaluating segmentation results have been calculated by the provided function in dataset which needs the contour of the predicted segment and the ground truth as input argument. Jaccard Measure (JM) and Hausdorﬀ Distance (HD) are two popular segmentation performance metrics that have been calculated using the aforementioned function. The Jaccard Measure, sometimes called Intersection over Union (IoU), is calculated based on the comparison of the automatic segmentation from the deep model ( ypred) and the manual segmentation delineated by experts ( ytrue).

用于评估分割结果的指标已由数据集中提供的函数计算得出，该函数需要预测段的轮廓和地面实况作为输入参数。 Jaccard Measure (JM) 和 Hausdorff Distance (HD) 是使用上述函数计算的两个流行的分割性能指标。 Jaccard Measure，有时也称为联合交集 (IoU)，它是根据深度模型的自动分割 (ypred) 与专家描绘的手动分割 (ytrue) 的比较计算得出的。

The Hausdorﬀ Distance between the automatic (Cpred) and manual (Ctrue) curves is the greatest distance of all points belonging to Cpred to the closest point [7] in Ctrue and is deﬁned as follows: 

自动 (Cpred) 和手动 (Ctrue) 曲线之间的 Hausdorff 距离是属于 Cpred 的所有点到 Ctrue 中最近点 [7] 的最大距离，定义如下：

* 1. Training the model

All the models are trained and evaluated on a computer with a Core i7-8700 K processor, 16 GB of RAM, and a GTX 1080 8 GB graphics card. Training a model from scratch generally takes less than 40 min to complete. To make the training faster and enjoy a relatively large batch size, we downsized every frame of the dataset by a factor of 2 as mentioned earlier.

所有模型都在配备 Core i7-8700 K 处理器、16 GB RAM 和 GTX 1080 8 GB 显卡的计算机上进行训练和评估。 从头开始训练模型通常需要不到 40 分钟即可完成。 为了使训练更快并享受相对较大的批量大小，如前所述，我们将数据集的每一帧缩小了 2 倍。

We implement DPU-Net with TensorFlow [42]. The weights in the model are all initialized using He initialization [39] with the default setting in TensorFlow. Then, we train the model with Adam optimization [43]. The learning rate is set to 0.0001 with no decay scheme. The real-time augmentor is used through training each model for 96 epochs, with a batch size of 6 and 144 iterations in total for each epoch. Note that we need two separate groups of models to predict the lumen area and the media area since the output activation is a sigmoid function: 

我们使用 TensorFlow [42] 实现了 DPU-Net。 模型中的权重均使用 He 初始化 [39] 进行初始化，使用 TensorFlow 中的默认设置。 然后，我们用 Adam 优化训练模型 [43]。 学习率设置为 0.0001，没有衰减方案。 通过将每个模型训练 96 个 epoch 来使用实时增强器，每个 epoch 的批次大小为 6 次，总共 144 次迭代。 请注意，由于输出激活是一个 sigmoid 函数，我们需要两组独立的模型来预测流明面积和介质面积：

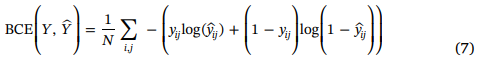
4.2.1. Loss function

We propose a heuristically designed loss function, that helps the model beneﬁt from both Binary Cross Entropy (BCE) and the Soft Dice Loss (SD). We minimize this objective function which is in fact a weighted average of the aforementioned loss functions.

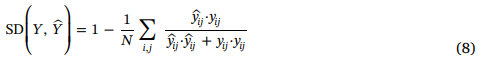


我们提出了一种启发式设计的损失函数，它有助于模型从二元交叉熵 (BCE) 和软骰子损失 (SD) 中受益。 我们最小化这个目标函数，它实际上是上述损失函数的加权平均值。

where λ = 0.8 denotes the weight of each term, the pixel-wise binary cross entropy is



and the soft dice loss which is a diﬀerentiable form of intersection over union is deﬁned as:



where N represents the total number of pixels in each image, yij is the ground truth value of ith row and jth column in the image annotation (either 1 or 0) and ŷ is the predicted probability of being a foreground pixel (either the lumen or the media).

For training each model, we monitor the average Jaccard Measure without extracting contours. The prediction given by a single model is a probability map of the input image size. We follow the same simple average ensemble practice in [15] to produce the ﬁnal result. The binarization is performed by using a searched threshold value where the threshold is searched on the out-of-fold version of the training set. Once the prediction maps are generated, ensembled, and binarized, we ﬁll any hole inside the binary region using the ‘ﬁllhole’ algorithm explained in [44], extract and trace the boundary using the algorithm described in [45], smooth the contour coordinates using ‘rloess’ [46] method and report it as the ﬁnal segmentation output.

为了训练每个模型，我们在不提取轮廓的情况下监控平均 Jaccard Measure。 单个模型给出的预测是输入图像大小的概率图。 我们遵循 [15] 中相同的简单平均集成实践来产生最终结果。 二值化是通过使用搜索阈值来执行的，其中阈值是在训练集的折叠版本上搜索的。 一旦生成、集成和二值化预测图，我们使用 [44] 中解释的“填充孔”算法填充二值区域内的任何孔，使用 [45] 中描述的算法提取和跟踪边界，平滑轮廓坐标 使用“rloess”[46] 方法并将其报告为最终的分割输出。

4.3. Results

In this section, we report the results of our thorough experiments over various augmented sets. Particularly, we present and compare the segmentation output of DPU-Net over 10 augmented sets for both 20 MHz and 40 MHz frames as well as segmentation results of UNet and SegNet trained over the same set. Also, we compare the segmentation output of DPU-Net trained over the best augmented set with existing state-of-the-art IVUS segmentation approaches.

在本节中，我们报告了我们对各种增强集的彻底实验的结果。 特别是，我们展示并比较了 DPU-Net 在 20 MHz 和 40 MHz 帧的 10 个增强集上的分割输出，以及在同一组上训练的 UNet 和 SegNet 的分割结果。 此外，我们将在最佳增强集上训练的 DPU-Net 的分割输出与现有最先进的 IVUS 分割方法进行比较。

4.3.1. Augmentation results

In order to have a ﬁrm grasp on the eﬀects of various augmentation operations, we have trained our proposed DPU-Net over augmented images generated based on various combinations of the proposed op- erations. The results are reported and compared together in Table 1 over the original training sets of Balocco’s dataset [1] which contains only 19 and 109 images for 40 MHz and 20 MHz IVUS frames (without any augmentation), respectively. The performance of the DPU-Net can be seen in Table 2 where all the evaluations results are obtained over the oﬃcial test set.

4.3.1. 增强结果

为了牢牢掌握各种增强操作的影响，我们已经在基于提议操作的各种组合生成的增强图像上训练了我们提出的 DPU-Net。 结果在表 1 中报告并与 Balocco 数据集 [1] 的原始训练集进行了比较，该数据集仅包含 19 和 109 个图像，分别用于 40 MHz 和 20 MHz IVUS 帧（没有任何增强）。 DPU-Net 的性能可以在表 2 中看到，其中所有评估结果都是在官方测试集上获得的。

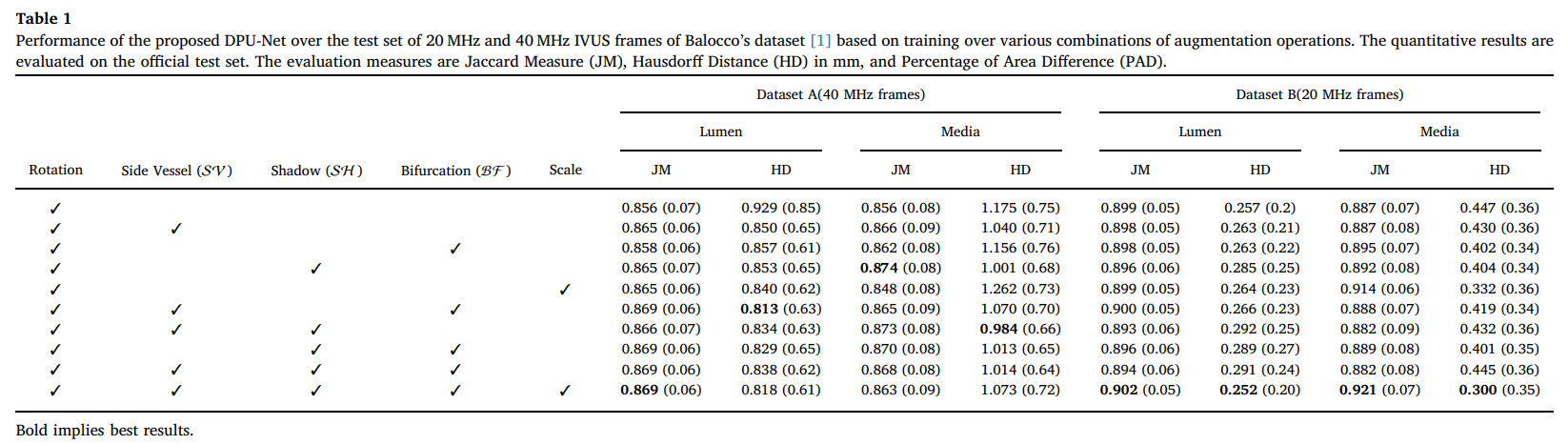


Table 1 Performance of the proposed DPU-Net over the test set of 20 MHz and 40 MHz IVUS frames of Balocco’s dataset [1] based on training over various combinations of augmentation operations. The quantitative results are evaluated on the oﬃcial test set. The evaluation measures are Jaccard Measure (JM), Hausdorﬀ Distance (HD) in mm, and Percentage of Area Diﬀerence (PAD).

表 1 基于对各种增强操作组合的训练，建议的 DPU-Net 在 Balocco 数据集 [1] 的 20 MHz 和 40 MHz IVUS 帧的测试集上的性能。 定量结果在官方测试集上进行评估。 评估量度是 Jaccard 量度 (JM)、以毫米为单位的豪斯多夫距离 (HD) 和面积差异百分比 (PAD)。

4.3.2. DPU-Net vs SegNet and U-Net

In the image segmentation literature there are two well-known ar- chitectures, namely the SegNet [21] for street scene segmentation and the U-Net [13] for segmentation of neuronal structures in electron microscope stacks. As both architectures can output a prediction map at the original input size (for U-Net, we need to change the valid padding to the same padding for all convolutional layers). To keep the comparison fair, we add the last upsampling transposed convolution layer to the end of both architectures. For illustration purposes, let the original size of the training images be W × H, as we mentioned in Section 3, we downsample the training images to a size of  . We train these two networks and our DPU-Net with the re-scaled low-resolution images and masks at the original size with no augmentation applied. The comparison results are shown in Table 2.

在图像分割文献中，有两种著名的架构，即用于街道场景分割的 SegNet [21] 和用于电子显微镜堆栈中神经元结构分割的 U-Net [13]。 由于两种架构都可以输出原始输入大小的预测图（对于 U-Net，我们需要将所有卷积层的有效填充更改为相同的填充）。 为了保持比较公平，我们将最后一个上采样转置卷积层添加到两种架构的末尾。 出于说明目的，让训练图像的原始大小为 W × H，正如我们在第 3 节中提到的，我们将训练图像下采样到  的大小。 我们使用重新缩放的低分辨率图像和原始大小的掩码训练这两个网络和我们的 DPU-Net，而没有应用增强。 比较结果如表2所示。

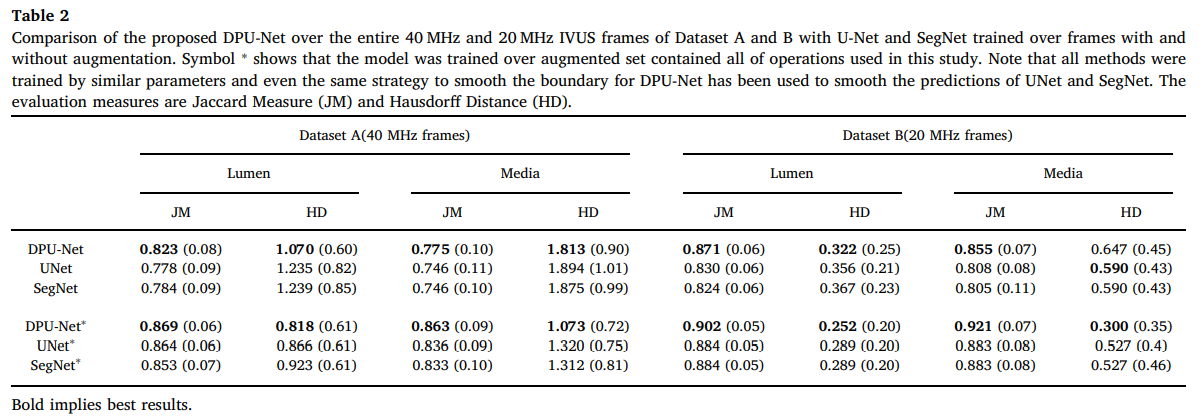


Table 2 Comparison of the proposed DPU-Net over the entire 40 MHz and 20 MHz IVUS frames of Dataset A and B with U-Net and SegNet trained over frames with and without augmentation. Symbol ∗ shows that the model was trained over augmented set contained all of operations used in this study. Note that all methods were trained by similar parameters and even the same strategy to smooth the boundary for DPU-Net has been used to smooth the predictions of UNet and SegNet. The evaluation measures are Jaccard Measure (JM) and Hausdorﬀ Distance (HD).

表 2 建议的 DPU-Net 在数据集 A 和 B 的整个 40 MHz 和 20 MHz IVUS 帧上与在有和没有增强的帧上训练的 U-Net 和 SegNet 的比较。 符号 \* 表明模型是在包含本研究中使用的所有操作的增强集上训练的。 请注意，所有方法都使用相似的参数进行训练，甚至用于平滑 DPU-Net 边界的相同策略也已用于平滑 UNet 和 SegNet 的预测。 评估措施是 Jaccard 措施 (JM) 和 Hausdorff 距离 (HD)。

4.3.3. Comparison with existing methods

In this section, we present experimental results over the test set of 20 MHz and 40 MHz IVUS B-mode datasets [1]. We obtained 5 models according to the 5-fold cross validation for each dataset with the con- ﬁguration mentioned in Section 4.2 and ensemble the predicted maps by the simple average voting. The quantitative results are shown in Table 3.

4.3.3. 与现有方法的比较

在本节中，我们展示了 20 MHz 和 40 MHz IVUS B 模式数据集测试集的实验结果 [1]。 我们根据每个数据集的 5 倍交叉验证获得了 5 个模型，使用第 4.2 节中提到的配置，并通过简单的平均投票来集成预测图。 定量结果见表3。

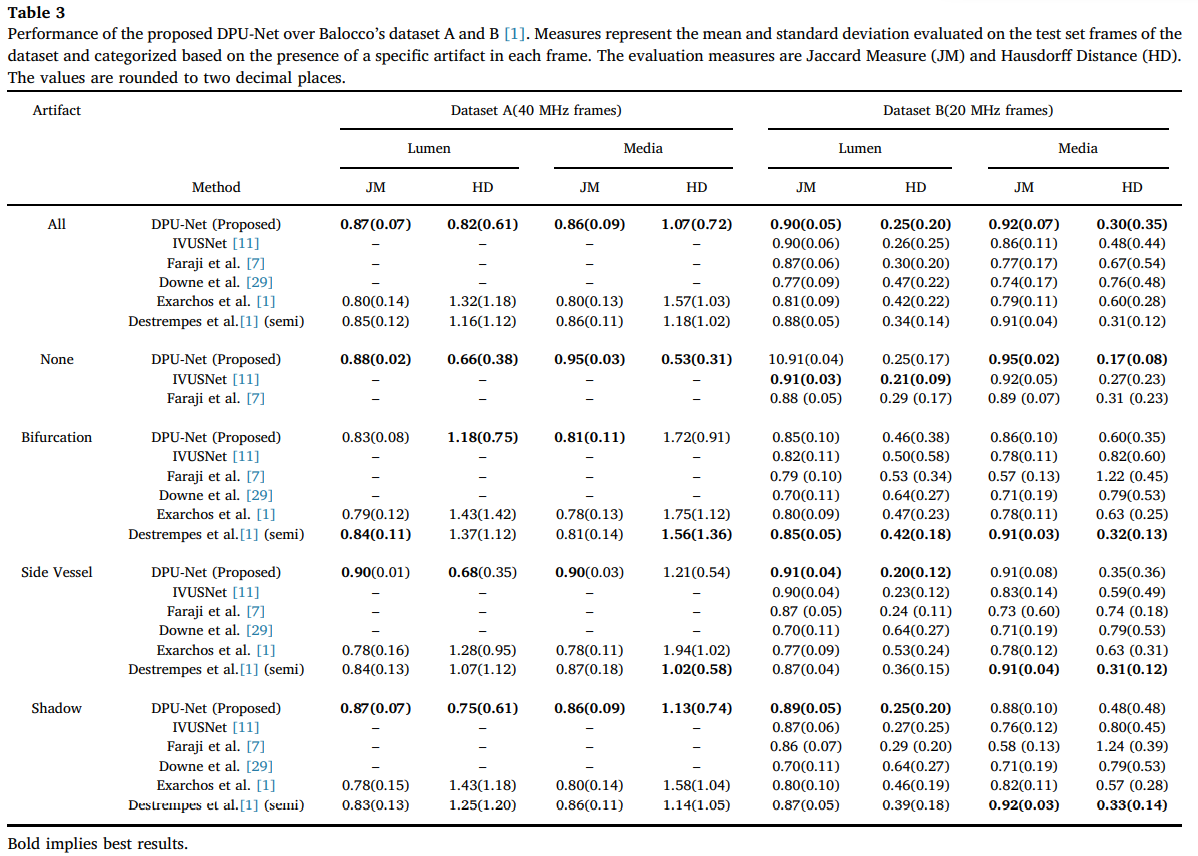


Table 3 Performance of the proposed DPU-Net over Balocco’s dataset A and B [1]. Measures represent the mean and standard deviation evaluated on the test set frames of the dataset and categorized based on the presence of a speciﬁc artifact in each frame. The evaluation measures are Jaccard Measure (JM) and Hausdorﬀ Distance (HD). The values are rounded to two decimal places.

表 3 提议的 DPU-Net 在 Balocco 的数据集 A 和 B [1] 上的性能。 度量表示在数据集的测试集帧上评估的平均值和标准偏差，并根据每个帧中特定工件的存在进行分类。 评估措施是 Jaccard 措施 (JM) 和 Hausdorff 距离 (HD)。 这些值四舍五入到两位小数。

5. Discussion

In this study, we performed several experiments to evaluate our proposed DPU-Net along with showing the improvement achieved by employing the newly proposed augmentation operations. We drew a comparison between DPU-Net, SegNet and UNet over original IVUS frames and also augmented frames. From the quantitative results we can see that DPU-Net performance is signiﬁcantly better than the other two existing architectures. We discuss three empirical reasons that help achieve such a result. First, DPU-Net has more convolutional layers but lower depth for each convolutional layer. This is similar to having several layers with a relatively small amount of neurons instead of putting many neurons in a single layer. A multi-layer architecture has stronger ability of learning representation [20,14] as it provides more possible ways of intersecting features and more non-linearity can be introduced by multiple non-linear activations. Since a deeper layer in a neural network usually has a larger receptive ﬁeld, it can capture features at multiple scales. However, there is always the limitation that a kernel of a ﬁxed size cannot be a universal solution. In the two IVUS datasets, especially the 40 MHz one, the shape and size of the lumen or media region are signiﬁcantly diﬀerent among images. In both U-Net and SegNet, there is no particular design to handle or improve the multi-scale segmentation. Our main branch and reﬁning branch handle this problem naturally by having convolutional layers with diﬀerent kernel sizes. Other improvements of DPU-Net over the original SegNet and U-Net consist of employing two downsampling methods simulta- neously, namely the pooling and strided convolution. This modiﬁcation over these two base architectures can help to ensure that we can leverage the information in multiple ways to increase the diversity of features.

5. 讨论

在这项研究中，我们进行了几次实验来评估我们提出的 DPU-Net，同时展示了通过采用新提出的增强操作所实现的改进。我们在原始 IVUS 帧和增强帧上对 DPU-Net、SegNet 和 UNet 进行了比较。从定量结果我们可以看出，DPU-Net 的性能明显优于其他两种现有架构。我们讨论了有助于实现这一结果的三个经验原因。首先，DPU-Net 具有更多的卷积层，但每个卷积层的深度较低。这类似于具有相对少量神经元的多层，而不是将许多神经元放在单个层中。多层架构具有更强的学习表示能力 [20,14]，因为它提供了更多可能的交叉特征方式，并且可以通过多个非线性激活引入更多非线性。由于神经网络中更深的层通常具有更大的感受野，因此它可以捕获多个尺度的特征。但是，始终存在固定大小的内核不能是通用解决方案的限制。在两个 IVUS 数据集中，尤其是 40 MHz 的数据集中，管腔或介质区域的形状和大小在图像之间存在显着差异。在 U-Net 和 SegNet 中，都没有特殊的设计来处理或改进多尺度分割。我们的主分支和精炼分支通过具有不同内核大小的卷积层自然地处理这个问题。 DPU-Net 对原始 SegNet 和 U-Net 的其他改进包括同时采用两种下采样方法，即池化和跨步卷积。对这两个基础架构的这种修改有助于确保我们能够以多种方式利用信息来增加特征的多样性。

Having shown the superiority of DPU-Net to UNet and SegNet that all were trained over non-augmented IVUS frames, we compared var- ious combinations of our proposed augmentation operations in Table 1 to ﬁgure out the best combination of the operations. Although, it was almost obvious that a training set contains all of the operations will achieve the best results, we did the cumbersome processes of training over diﬀerent combinations of operations to support and prove our hypothesis as it is shown in Table 1 so. Looking at the reported results in Table 2, we see that training UNet and SegNet over these operations signiﬁcantly improved their resulted segmentation, though DPU-Net still outperforms them. Training over a set created using our proposed augmentation operations can dramatically increases the accuracy of the segmentation. More speciﬁcally, it can increase the lumen JM from 0.823 to 0.869 for 40 MHz images and from 0.871 to 0.902 for 20 MHz frames. The accuracy of media segmentation is also increased from 0.775 to 0.863 for 40 MHz images and from 0.855 to 0.921 for 20 MHz images.

展示了 DPU-Net 相对于 UNet 和 SegNet 的优越性，所有这些都在非增强的 IVUS 框架上进行了训练，我们比较了表 1 中我们提出的增强操作的各种组合，以找出操作的最佳组合。尽管几乎很明显包含所有操作的训练集将获得最佳结果，但我们对不同的操作组合进行了繁琐的训练过程，以支持和证明我们的假设，如表 1 所示。查看表 2 中报告的结果，我们看到在这些操作上训练 UNet 和 SegNet 显着改善了它们的分割结果，尽管 DPU-Net 仍然优于它们。对使用我们提出的增强操作创建的集合进行训练可以显着提高分割的准确性。更具体地说，它可以将 40 MHz 图像的流明 JM 从 0.823 增加到 0.869，将 20 MHz 帧的流明 JM 从 0.871 增加到 0.902。 40 MHz 图像的媒体分割精度也从 0.775 增加到 0.863，20 MHz 图像从 0.855 增加到 0.921。

that DPU-Net signiﬁcantly improves the result for segmenting the media region, from an JM score 0.79 [1] to 0.92, and HD from 0.60 to 0.30 on the 20 MHz dataset. The reasons behind why DPU-Net does not exceed all the methods in every category of [1] can be addressed from two perspectives. First, the training set is too small to capture all the common artifacts in the real world and even the test set. However, the architecture is still considerably eﬀective as the training set contains only 1 image with side vessel artifact while the test set contains 93 frames with side vessel artifacts. Secondly, the shadow artifacts are generally overlapped with parts of the media area that makes the segmentation becomes much more challenging since the media regions leak to the background. Some predictions are illustrated in Fig. 6.

在 20 MHz 数据集上，DPU-Net 显着改善了分割媒体区域的结果，从 JM 得分 0.79 [1] 到 0.92，HD 从 0.60 到 0.30。 DPU-Net 没有超过 [1] 中每个类别中的所有方法的原因可以从两个角度解决。 首先，训练集太小，无法捕获现实世界中的所有常见工件，甚至是测试集。 然而，该架构仍然相当有效，因为训练集仅包含 1 个带有侧血管伪影的图像，而测试集包含 93 个具有侧脉伪影的帧。 其次，阴影伪影通常与媒体区域的部分重叠，这使得分割变得更具挑战性，因为媒体区域泄漏到背景中。 一些预测如图 6 所示。

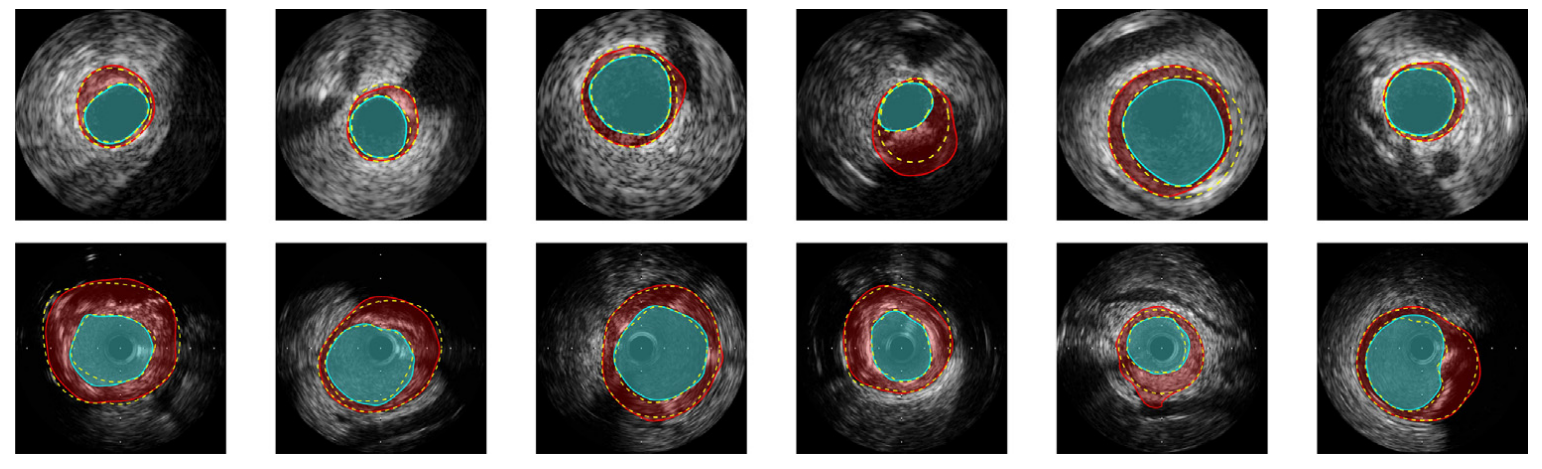


Fig. 6. Some of the lumen and media segmentation results for images from dataset B (ﬁrst row) and dataset A (second row). Segmented lumen and media have been highlighted by cyan and red colors, respectively. The yellow dashed lines illustrate the gold standard that have been delineated by four clinical experts [1]. (For interpretation of the references to colour in this ﬁgure legend, the reader is referred to the web version of this article.)

图 6. 来自数据集 B（第一行）和数据集 A（第二行）的图像的一些流明和媒体分割结果。 分段的流明和媒体分别用青色和红色突出显示。 黄色虚线表示由四位临床专家描绘的金标准 [1]。 （为了解释这个图例中对颜色的引用，读者可以参考本文的网络版本。）

6. Conclusion

In this paper, we proposed DPU-Net, a fully convolutional deep network, that is able to generalize even if there is a small number of training images for the segmentation of arterial walls in IVUS images. We evaluated the generalization ability of our proposed DPU-Net by comparing it with two existing general-purposed segmentation architectures, namely SegNet and UNet, that were trained over the same number of images for the same amount of time without doing any augmentation. The results indicate a signiﬁcant improvement for DPU- Net over SegNet and UNet. Speciﬁcally, DPU-Net achieved more than 4% and 5% higher accuracy in terms of JM for 40 MHz and 20 MHz datasets, respectively. These emprical outcomes express a higher generalization ability than SegNet and UNet. The contributions of this paper can be summarized as follows:

• We introduced a domain-speciﬁc design for image augmentation that can:

– Generate various types of augmented images in real-time.

– Add various combinations of three common IVUS artifacts into the training images. We empirically proved that we can produce a signiﬁcant number of eﬀective augmented images. This can be counted an eﬀective augmentation pipeline and can be generalized for diﬀerent deep architectures and tasks.

• We proposed DPU-Net that outperforms existing approaches over a publicly available IVUS benchmark dataset [1] which contains IVUS images with a signiﬁcant number of artifacts. We also compare it with several existing inﬂuential architectures in the deep learning literature, namely SegNet [21] and U-Net [13]. This shows that the proposed work has a potential to be used in solving other segmentation problems as well.

在本文中，我们提出了 DPU-Net，一种完全卷积的深度网络，即使在 IVUS 图像中用于分割动脉壁的训练图像数量很少，它也能够进行泛化。 我们通过将我们提出的 DPU-Net 与两个现有的通用分割架构（即 SegNet 和 UNet）进行比较来评估我们提出的 DPU-Net 的泛化能力，这些架构在相同数量的图像上训练相同的时间而不进行任何增强。 结果表明 DPU-Net 比 SegNet 和 UNet 有显着改进。 具体来说，DPU-Net 在 40 MHz 和 20 MHz 数据集的 JM 方面的准确度分别提高了 4% 和 5% 以上。 这些实证结果表现出比 SegNet 和 UNet 更高的泛化能力。 本文的贡献可以总结如下：

• 我们为图像增强引入了一个特定领域的设计，它可以：

– 实时生成各种类型的增强图像。

– 将三种常见 IVUS 伪影的各种组合添加到训练图像中。 我们凭经验证明我们可以产生大量有效的增强图像。 这可以算作一个有效的增强管道，并且可以推广到不同的深层架构和任务。

• 我们提出的 DPU-Net 在公开可用的 IVUS 基准数据集 [1] 上优于现有方法，该数据集包含具有大量伪影的 IVUS 图像。 我们还将其与深度学习文献中现有的几个有影响力的架构进行了比较，即 SegNet [21] 和 U-Net [13]。 这表明所提出的工作也有可能用于解决其他分割问题。

To further improve the segmentation performance, we devised our own augmentation framework called real-time augmentor. Our real- time augmentor not only generates augmented images in a way that does not interrupt the training process on GPU(s), but also contains our proposed IVUS artifact-based augmentation operations that include three common IVUS artifacts into the training data to simulate the frames with artifacts. We thoroughly investigated how various augmentation operations aﬀect the ﬁnal accuracy of the model. Consequently, the experimental results reveal that a DPU-Net model trained using these augmented data outperforms every available state- of-the-art automatic and semi-automatic IVUS segmentation methods by a large margin.

为了进一步提高分割性能，我们设计了自己的增强框架，称为实时增强器。 我们的实时增强器不仅以不中断 GPU 训练过程的方式生成增强图像，而且还包含我们提出的基于 IVUS 伪影的增强操作，将三种常见的 IVUS 伪影包含在训练数据中以模拟 带有伪像的帧。 我们彻底研究了各种增强操作如何影响模型的最终精度。 因此，实验结果表明，使用这些增强数据训练的 DPU-Net 模型在很大程度上优于所有最先进的自动和半自动 IVUS 分割方法。

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ultras.2019.03.014>.

与本文相关的补充数据可在在线版本中找到，网址为 <https://doi.org/10.1016/j.ultras.2019.03.014>。

References

1. S. Balocco, C. Gatta, F. Ciompi, A. Wahle, P. Radeva, S. Carlier, G. Unal, E. Sanidas, J. Mauri, X. Carillo, et al., Standardized evaluation methodology and reference database for evaluating IVUS image segmentation, Comput. Med. Imag. Graph. 38 (2) (2014) 70–90.
2. E.G. Mendizabal-Ruiz, M. Rivera, I.A. Kakadiaris, Segmentation of the luminal border in intravascular ultrasound b-mode images using a probabilistic approach, Med. Image Anal. 17 (6) (2013) 649–670.
3. G. Mendizabal-Ruiz, I.A. Kakadiaris, A physics-based intravascular ultrasound image reconstruction method for lumen segmentation, Comput. Biol. Med. 75 (2016) 19–29.
4. A. Taki, Z. Najaﬁ, A. Roodaki, S.K. Setarehdan, R.A. Zorooﬁ, A. Konig, N. Navab, Automatic segmentation of calciﬁed plaques and vessel borders in IVUS images, Int. J. Comput. Assis. Radiol. Surg. 3 (3–4) (2008) 347–354.
5. X. Zhu, P. Zhang, J. Shao, Y. Cheng, Y. Zhang, J. Bai, A snake-based method for segmentation of intravascular ultrasound images and its in vivo validation, Ultrasonics 51 (2) (2011) 181–189.
6. G. Unal, S. Bucher, S. Carlier, G. Slabaugh, T. Fang, K. Tanaka, Shape-driven seg- mentation of the arterial wall in intravascular ultrasound images, IEEE Trans. Inform. Technol. Biomed. 12 (3) (2008) 335–347.
7. M. Faraji, I. Cheng, I. Naudin, A. Basu, Segmentation of arterial walls in in- travascular ultrasound cross-sectional images using extremal region selection, Ultrasonics 84 (2018) 356–365.
8. Y. Li, M. Faraji, Erel Selection Using Morphological Relation. Available from: arXiv preprint arXiv:1806.03580.
9. M. Faraji, J. Shanbehzadeh, K. Nasrollahi, T. Moeslund, Extremal regions detection guided by maxima of gradient magnitude, IEEE Trans. Image Process.
10. M. Faraji, J. Shanbehzadeh, K. Nasrollahi, T.B. Moeslund, Erel: Extremal regions of extremum levels, 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 681–685.
11. J. Yang, L. Tong, M. Faraji, A. Basu, Ivus-net: An Intravascular Ultrasound Segmentation Network. Available from: arXiv preprint arXiv:1806.03583.
12. S. Kim, Y. Jang, B. Jeon, Y. Hong, H. Shim, H. Chang, Fully automatic segmentation of coronary arteries based on deep neural network in intravascular ultrasound images, Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, 2018, pp. 161–168.
13. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
14. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classiﬁcation with deep con- volutional neural networks, Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
15. D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.
16. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic seg- mentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
17. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFS, 2016. Available from: arXiv preprint arXiv:1606.00915.
18. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning. Available from: arXiv preprint arXiv:1711.05225.
19. C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large Kernel Matters–Improve Semantic Segmentation by Global Convolutional Network. Available from: arXiv preprint arXiv:1703.02719.
20. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
21. V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder- decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
22. M.A. Tanner, W.H. Wong, The calculation of posterior distributions by data aug- mentation, J. Am. Stat. Assoc. 82 (398) (1987) 528–540.
23. P.Y. Simard, D. Steinkraus, J.C. Platt, Best practices for convolutional neural net- works applied to visual document analysis, Null, IEEE, 2003, p. 958.
24. D. Cireşan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classiﬁcation. Available from: arXiv preprint arXiv:1202.2745.
25. D.C. Cireşan, U. Meier, J. Masci, L.M. Gambardella, J. Schmidhuber, High-perfor- mance neural networks for visual object classiﬁcation. Available from: arXiv pre- print arXiv:1102.0183.
26. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. Available from: arXiv preprint arXiv:1409.1556.
27. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
28. M. Paulin, J. Revaud, Z. Harchaoui, F. Perronnin, C. Schmid, Transformation pur- suit for image classiﬁcation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3646–3653.
29. R. Downe, A. Wahle, T. Kovarnik, H. Skalicka, J. Lopez, J. Horak, M. Sonka, Segmentation of intravascular ultrasound images using graph search and a novel cost function, in: Proc. 2nd MICCAI Workshop on Computer Vision for Intravascular and Intracardiac Imaging, Citeseer, 2008, pp. 71–79.
30. M.-H. Cardinal, J. Meunier, G. Soulez, R.L. Maurice, É. Therasse, G. Cloutier, Intravascular ultrasound image segmentation: a three-dimensional fast-marching method based on gray level distributions, IEEE Trans. Med. Imag. 25 (5) (2006) 590–601.
31. J. Dijkstra, G. Koning, J. Tuinenburg, P. Oemrawsingh, J. Reiber, Automatic border detection in intravascular iltrasound images for quantitative measurements of the vessel, lumen and stent parameters, Computers in Cardiology 2001, IEEE, 2001, pp. 25–28.
32. M.E. Plissiti, D.I. Fotiadis, L.K. Michalis, G.E. Bozios, An automated method for lumen and media-adventitia border detection in a sequence of IVUS frames, IEEE Trans. Inform. Technol. Biomed. 8 (2) (2004) 131–141.
33. J.D. Klingensmith, R. Shekhar, D.G. Vince, Evaluation of three-dimensional seg- mentation algorithms for the identiﬁcation of luminal and medial-adventitial bor- ders in intravascular ultrasound images, IEEE Trans. Med. Imag. 19 (10) (2000) 996–1011.
34. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009, IEEE, 2009, pp. 248–255.
35. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: common objects in context, European Conference on Computer Vision, Springer, 2014, pp. 740–755.
36. S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 5987–5995.
37. M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, C. Pal, The importance of skip connections in biomedical image segmentation, Deep Learning and Data Labeling for Medical Applications, Springer, 2016, pp. 179–187.
38. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
39. K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectiﬁers: surpassing human-level performance on imagenet classiﬁcation, Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.
40. B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectiﬁed activations in convolutional network. Available from: arXiv preprint arXiv:1505.00853.
41. L. Lo Vercio, J.I. Orlando, M. del Fresno, I. Larrabide, Assessment of image features for vessel wall segmentation in intravascular ultrasound images, Int. J. Comput. Assis. Radiol. Surg. 11 (8) (2016) 1397–1407.
42. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Software Available from tensorﬂow.org, 2015. < https://www.tensorﬂow.org/ >.
43. D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization. Available from: arXiv preprint arXiv:1412.6980.
44. P. Soille, Morphological Image Analysis: Principles and Applications, Springer Science & Business Media, 2013.
45. R.C. Gonzalez, R.E. Woods, S. Eddins, Digital Image Processing using Matlab: Pearson Prentice Hall, Upper Saddle River, New Jersey.
46. S.J. Orfanidis, Introduction to Signal Processing, Prentice-Hall Inc., 1995.