Hans汉斯

基于注意力机制的小儿肾图分割网络

李志力

上海理工大学,光电信息与计算机工程学院,上海

收稿日期: 2024年8月2日; 录用日期: 2024年8月26日; 发布日期: 2024年8月30日

摘要

本文提出了一种全新的网络架构,解决了U-Net在小儿肾图分割中对全局信息的部分损失问题。由于 U-Net的编码器通过多次下采样降低了输入图像的空间分辨率,导致模型无法完全利用图像的全局上下 文信息,且下采样过程丢失了大量重要信息。因此,将Transformer架构引入U-Net的深层学习目标位置 信息及局部与全局的关系。此外,设计了多尺度通道和空间注意力模块,该模块融合了多尺度特征以弥 补网络池化过程中丢失的信息,能有效地捕捉了特征图的全局上下文信息,抑制背景噪声。最后,利用 新华医院收集的小儿肾图数据,进行了对比实验与消融实验,验证了所提方法的有效性。

关键词

肾图分割,深度学习,语义分割,注意力机制

Segmentation Network of Paediatric Kidney Map Based on Attention Mechanism

Zhili Li

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Aug. 2nd, 2024; accepted: Aug. 26th, 2024; published: Aug. 30th, 2024

Abstract

In this paper, a novel network architecture is proposed to solve the problem of partial loss of global information by U-Net in paediatric renogram segmentation. Since the encoder of U-Net reduces the spatial resolution of the input image by multiple downsampling, it results in the model not being able to fully utilise the global contextual information of the image, and the downsampling process loses a large amount of important information. Therefore, the Transformer architecture is introduced into U-Net's deep learning of target location information and the relationship

between local and global. In addition, the Multiscale Channel and Spatial Attention (MSCA) module is designed, which incorporates multiscale features to compensate for the information lost in the network pooling process, and can effectively capture the global context information of the feature map and suppress the background noise. Finally, using the paediatric nephrogram data collected from Xinhua Hospital, comparative experiments and ablation experiments were conducted to verify the effectiveness of the proposed method.

Keywords

Kidney Image Segmentation, Deep Learning, Semantic Segmentation, Attention Mechanism

Copyright © 2024 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <u>http://creativecommons.org/licenses/by/4.0/</u> Open Access

1. 引言

肾脏是人体重要的排泄器官,主要工作是产生尿液,将身体新陈代谢产生的废物和有害物质排出人体。因此,肾脏的功能状况对人体健康至关重要。肾图是反映肾脏功能状态及尿路排泄通畅情况的重要检查手段。该检查过程需静脉注射同位素示踪剂,通过肾图仪持续20分钟监测,获取左右两肾的原始影像数据。之后,医生会根据影像对左右肾区进行手工标注,绘制出相应的时间放射性曲线,即常称的肾图曲线。然而,在医学影像分割领域,肾脏分割研究相对薄弱。目前临床上,影像科医生仍需首先用手工标注肾脏区域,然后才能出诊断结果,这一做法存在效率低下、主观性强、劳动强度大和操作繁琐等缺陷,并且长期作业可能影响诊断水平。因此,设计一个肾图自动分割网络模型是有必要的,以协助医生完成标注,减少人工标注误差,提高分割精度和效率。

图像分割是一种像素级别的分类任务,旨在为图像中的每个像素分配正确的类别标签。目前,在医学图像分割领域,U-Net 是一个表现出色的经典模型[1]。它的结构包括编码器和解码器。编码器通过卷积和池化操作进行下采样,从而提取图像的特征表示。解码器则通过反卷积操作进行上采样,将特征映射还原到原始图像大小。但是卷积操作的局限性导致难以学习全局和远程依赖关系,这个问题限制了传统 U-Net 模型在图像分割任务中的表现,尤其是在处理对比度较低的医学图像时。为了克服这种局限性,2017年,谷歌团队提出了Transformer [2],最初用于解决自然语言处理问题,并在该领域取得了巨大成功。Transformer 的成功应用不仅限于自然语言领域,还扩展到计算机视觉领域,包括医学图像检测、分类和分割等任务。Vision Transformer (ViT)在 ImageNet 分类任务上取得了 state-of-the-art 的成果[3],通过将自注意力机制直接应用于全尺寸图像,实现了全局和长距离的语义信息交互。受到 Transformer 架构在图像分割任务中成功应用的启发,本文在 U-Net 网络的基础上,整合了 Transformer 编码器。

尽管 U-Net 网络采用跳跃连接来融合浅层和深层特征以缓解下采样导致的空间信息丢失问题,但同一尺度特征图包含的信息有限。为了增加特征信息,需要融合不同尺度的信息,然而融合后可能存在噪声,导致信息冗余,从而影响分割准确性。为了解决这一问题,注意力机制被广泛应用于计算机视觉领域。2017 年提出的 Squeeze-and-Excitation Networks (SENet)通过全局平均池化将每个通道的二维特征压缩为一个实数[4],并通过两个全连接层构建通道之间的依赖关系,最后通过 Sigmoid 函数进行归一化,加权到每个通道,以关注有用的特征通道。2019 年,提出了 Convolutional Block Attention Module (CBAM) [5],包括通道注意力和空间注意力,通道注意力对每个通道分配权值,而空间注意力对每个像素点分配

权值,在通道和空间两个维度上建立依赖关系,重点关注关键特征并限制无用的特征。CBAM 可以轻松 融入任何 CNN 架构中,并且参数量可以忽略不计。受到这些启发,本文引入了卷积块注意力模块(CBAM),并采用了多尺度方法,提出了一种多尺度通道和空间注意力模块(MSCA)。

本文的主要贡献如下:

1) 本文在 U-Net 编码器后引入了 Transformer 编码架构, Transformer 的多头自注意力可以有效构建 全局关系和局部特征,提取丰富的全局上下文信息。

2) 设计了多尺度通道和空间注意力模块(MSCA),不仅能够有效地融合多尺度的特征信息,还能够 保留传统网络池化操作所丢失的信息和捕捉边缘信息。同时将融合后的多尺度特征通过通道注意力和空 间注意力,捕捉特征图丰富的上下文信息,并且关注感兴趣区域,抑制背景噪声。

3) 将提出的基于注意力机制的分割网络模型与多种分割模型本文方法进行比较,均取得了较好结果, 证明了该模型在小儿肾脏分割任务上的有效性。同时,本文进行了消融实验,验证了所提模块的优化性能。

2. 小儿肾图分割网络结构设计

本文提出的网络架构如图 1 所示,解码器部分由卷积构成,在编码器的最后一层引入 Transformer 架构,以捕获全局特征,建立全局和远程依赖关系,Transformer 层的输出进行上采样与多尺度通道和空 间注意力模块(MSCA)输出进行拼接,再上采样,再拼接,逐渐恢复到原来大小。多尺度通道和空间注意 力模块(MSCA)融合了不同尺度的浅层特征,避免了网络池化信息的丢失,并且在通道和空间维度分别赋 予不同权值,能够捕捉到不同维度的上下文信息,实现通道和空间信息的交互,聚焦关键区域,抑制多 余的信息,优化了网络性能。



为了使 Transformer 也适用于计算机视觉任务,许多 modifications 的工作已被发表。Vision Transformer (ViT)在 ImageNet 分类任务上达到 state-of-the-art,通过直接应用 Transformers with global self-attention 到 full-sized images。受此启发,提出将三层 Transformer block 引入 U-Net 的深层结构,更好地捕获全局特征,建立依赖关系。

如图 2 所示,输入图像 $x \in R^{H \times W \times C}$,其空间分辨率为 $H \times W$,通道数为 C。首先将图片序列化,输入 x 被重塑为 2D 的 patches,每个 patch 大小为 p^2 共有 HW/p^2 个 patch。使用可训练的线性投影将矢量化 的 patches 映射到一个潜在的 D 维 embedding 空间,为了对 patches 空间信息进行编码,引入位置 embedding 以保留位置信息。

$$Z_0 = \left[x_p^1 E; x_p^2 E; \cdots; x_p^N E \right] + E_{pos}$$
⁽¹⁾

其中 $E \in \Box^{(p^2 \cdot c) \times D}$ 为 patch 的线性投影, $E_{pos} \in \Box^{N \times D}$ 为位置 embedding。Transformer 编码器由多头自注意 力和多层感知机组成。

$$MSA(q,k,v) = softmax \left(q \cdot k / \sqrt{d}\right) \cdot v$$
⁽²⁾

$$z'_{l} = MSA(LN(z_{l-1})) + z_{l-1}$$
(3)

$$z_{l} = MLP(LN(z_{l}')) + z_{l}'$$

$$\tag{4}$$

其中 $LN(\cdot)$ 为层归一化, z_l 为第 L 层编码的图像表示, q,k,v分别为 token 的线性映射。



Figure 2. Transformer encoders 图 2. Transformer 编码器

Vision Transformer (ViT)是一种基于 Transformer 架构的模型,它将图像分割成小的图块,然后通过 一个 Transformer 编码器对这些 patch 进行处理。其中自注意力是 Transformer 模型的核心组件之一,它允 许模型在处理序列数据时在不同位置之间建立关联。如图 3 所示,自注意力机制被用来捕捉图像中不同 区域之间的关系。对于每个图块,计算与其他所有 patch 的相似度,利用点积操作得到一个注意力权重向 量。这个向量指示了每个 patch 与其他 patch 的关联程度,即哪些 patch 对当前 patch 更重要。使用注意力 权重对所有 patch 的特征向量进行加权求和,以获取每个 patch 的上下文表示。这一步类似于加权平均, 但是权重是由注意力机制计算得到的。为了增强模型的表示能力,可以使用多头注意力机制。在这种情 况下,会并行地运行多个自注意力子层,每个子层产生不同的注意力权重。最后将这些子层的输出合并 起来。



在多尺度通道和空间注意力模块中(MSCA),分别将深层特征送入卷积核大小为3的卷积和卷积核大小为1的卷积,并将卷积后的特征图相加,再通过上采样与浅层特征图拼接得到新的特征图。新的特征 图作为 CBAM 的输入,每个通道的二维特征通过全局平均池化压缩成一个实数。输出权重的数量与特征 图中输入通道的数量相同。最后,将得到的归一化权重对每个通道特征和位置特征进行加权。提出的融 合多尺度特征的注意力网络如图4所示,其中X1为浅层输入特征图,X2为深层输入特征图,X2经过卷积 和上采样与 X2 拼接得到拼接特征图 X3,再经过通道注意力机制与空间注意力机制得到最终输出特征图 Output。表1 描述了模型的每一层结构。



Figure 4. Multi-scale channels and spatial attention modules 图 4. 多尺度通道和空间注意力模块

3. 数据集和实验设计

本文所使用的肾图数据源自上海交通大学医学院附属新华医院核医学科提供的原始肾图 DICOM 文件,经解析后共得到 700 套肾图数据,并在此基础上进行了预处理。肾图是通过肾图仪监控人体内放射

性同位素的位置和含量,根据这些信息绘制的灰度图。监控时间为 20 分钟,前 1 分钟每秒绘制一张灰度 图(共 60 张),之后的 19 分钟每 15 秒绘制一张灰度图(共 76 张),因此每个人得到一套共 136 张图像。肾 图数据集的制作在专业影像科医生的指导下进行。在医疗肾图的分析诊断过程中,由于第一分钟人体内 放射性物质尚未到达肾图需要的指定位置,故其诊断效果无效,因此我们去除了第一分钟所采集的 60 张 灰度图。传统肾图标注通常集中在第二到第三分钟所采集的灰度图上,因此我们按时间顺序每 7 张图片 进行一次图像融合,得到共 11 张图。随后,利用 OpenCV 库对这 11 张图进行伪彩色处理,增强图像对 比度,以利于后续的标注和训练。

编码器	阶段	操作	图像大小	通道数
	E0	2D 卷积	H imes W	3→32
	E1	2D 卷积 + 最大池化	$H \times W \rightarrow H/2 \times W/2$	32→64
	E2	2D 卷积 + 最大池化	$H/2 \times W/2 \rightarrow H/4 \times W/4$	64→128
	E3	2D 卷积 + 最大池化	$H/4 \times W/4 \rightarrow H/8 \times W/8$	128→256
	E4	2D 卷积 + 最大池化	$H/8 \times W/8 \rightarrow H/16 \times W/16$	256→256
Transformer	C1	Transformer 编码器 ×3	<i>H</i> /16 × <i>W</i> /16	256
	D1	双线性插值 + 2D 卷积	$H/16 \times W/16 \rightarrow H/8 \times W/8$	256→128
解码器	D2	双线性插值 + 2D 卷积	$H/8 \times W/8 \rightarrow H/4 \times W/4$	128→64
	D3	双线性插值 + 2D 卷积	$H/4 \times W/4 \rightarrow H/2 \times W/2$	64→32
	D4	2D 卷积	$H/2 \times W/2 \rightarrow H \times W$	32→3

Table 1. Modelling each layer of the structure 表 1. 模型每一层结构

如图 5、图 6 所示,伪彩色处理会将灰度图像中的灰度值(0~255)映射到一系列彩色值上。这些彩色 值通常从浅色(如蓝色或绿色)渐变到深色(如红色或紫色)。在伪彩色处理中,颜色的深浅通常用于表示图 像中的不同灰度值的强度或密度。颜色越浅的区域通常代表着低强度或低密度的特征,而颜色越深的区 域通常代表着高强度或高密度的特征。这种渐变的颜色表示方式有助于突出图像中的不同特征,并且在 可视化和分析数据时提供了更丰富的信息。例如,在医学图像处理中,伪彩色处理经常用于突出不同密 度的组织或病变,浅色可能代表低密度的正常组织,而深色可能代表高密度的异常区域。



Figure 5. Kidney map processing effect 图 5. 肾图处理效果图



Figure 6. Labelme image annotation 图 6. Labelme 图像标注

我们选取其中的第2或第3张图片,利用 labelme 软件进行标注,并将原图和标签整合到我们的数据 集中。通过这一系列操作,我们得到了包含700张肾图的数据集。具体过程如图7所示。图5是处理完 成后的效果图。

在完成肾图数据集的预处理后,我们在医生的指导下进行了图片标签的标注工作。我们使用了专业 的图片标注软件 labelme 进行标注,如图 7 所示。标注完成后,会生成与原始图像对应的 jpg 格式的标签 图片。图 8 展示了标注结果。通过以上步骤,我们完成了肾图数据集的预处理和标签制作工作。



原始图按时间顺序叠加

Figure 7. Image pre-processing process 图 7. 图像预处理过程



Figure 8. Labelme image annotation results 图 8. Labelme 图像标注结果

李志力

为了客观地评价模型的性能,本文用准确率(Accuracy)、平均交并比(mIou)、交并比(IoU)和 Dice 系数(Dice)四个指标对模型分割性能进行定量评估。

首先引入混淆矩阵,如图 9 所示,A 代表真实值,B 代表预测值,混淆矩阵(Confusion Matrix)是用于 评估分类模型性能的一种表格,通常用于监督学习问题中的分类任务。它对模型的预测结果和真实标签 之间的关系进行了可视化和总结。混淆矩阵的构建基于以下概念:

1) 真阳性(True Positive, TP): 模型将正类样本正确地预测为正类。

2) 真阴性(True Negative, TN): 模型将负类样本正确地预测为负类。

3) 假阳性(False Positive, FP): 模型将负类样本错误地预测为正类。

4) 假阴性(False Negative, FN): 模型将正类样本错误地预测为负类。

混淆矩阵以实际类别(真实标签)和预测类别(模型预测结果)为基础,将样本划分为四个类别,并统计 每个类别的数量。



Figure 9. Confusion matrix schematic 图 9. 混淆矩阵示意图

准确率评价指标定义为:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(5)

其中 TP 代表前景被预测正确的部分, FP 代表背景被预测错误的部分, TN 代表背景被预测正确的部分, FN 代表前景被预测错误的部分。交并比评价指标定义为:

$$IoU = \frac{TP}{FN + TP + FP}$$
(6)

此外,平均交并比对应的公式定义为:

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{jj} + \sum_{j=0}^{k} p_{ji} - p_{ii}}$$
(7)

其中i代表实际值,j代表预测值, p_{ii} 表示实际值i被预测为预测值j。Dice 系数对应的公式定义为:

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{8}$$

其中X代表实际值,Y代表预测值。

在训练前对数据集进行预处理,用 Imagenet 的平均值和标准差减去数据集的平均值进行归一化,然 后减去数据集相应维度的统计平均值,以消除共同部分,突出个体之间的差异和特征。对于小儿肾图医 学图像数据集,实验以原始分辨率训练,训练过程持续了 400 轮,学习率从 0.0001 开始。实验中将图像 分辨率缩小至 224 × 224,从而提高训练与测试的效率。将实验数据按 8:2 的比例划分为训练集和测试集。 本文进行实验所使用的实验平台如表 2 所示。

实验软硬件环境	环境配置
操作系统	Ubuntu16.04
处理器	Intel Core i7-9800X @3.8GHz
内存	32g
显卡	NVIDIA RTX2080Ti
编程语言	Python 3.6.4
深度学习框架	PyTorch 1.6.0

Table 2. Experimental platforms 表 2. 实验平台

实验中采用的肾图数据集来源于上海新华医院提供的小儿肾图原 DICOM 文件,经过第 3.3.1 小节描述的预处理步骤后得到。实验的流程首先是读取预处理后的数据,然后将数据输入网络进行训练。在实验过程中,根据所得数据调整网络参数,以获取性能最优的网络模型。训练完成后,将测试集数据输入网络,获得最终的测试结果,并对结果进行分析。具体流程如图 10 所示。



Figure 10. Data processing and experimental procedures 图 10. 数据处理和实验流程

在小儿肾图数据中,将提出的方法与几个最先进的图像分割方法进行了比较,包括 U-Net、Deeplab-v3、SegNet、CENe、PSP Net、Attention U-Net 和 FCN8s。模型在小儿肾脏图的性能比较如表 3 所示。由于保留了 CNN 提取短程信息的优点,同时构建了 Transformer 模块,从而解决了 CNN 在全局特征方面的缺陷,

分割性能得到很大程度地提升。与性能第二的 Attention U-Net 相比, mIoU 在肾脏医学图像中提升了 2.94%, Dice 系数提升了 0.12%。与目前主流的图像分割模型相比,该模型在多个评价指标有了很大提高。多种对比方法的可视化分割结果如图 11 所示。提出的方法相较于以往方法对目标边缘的分割有更好的效果,证明了提出的 MSCA 能够融合多尺度信息从而更好地捕捉全局和细节信息。提出的方法分割结果与真实标注更为接近,表现优于其他网络,即使在肾脏与周围组织界限不明显的情况下也能保持良好的分割效果,由分割可视化图进一步证明了模型在肾脏区域分割上的优越性。

Table 3. Comparison of the effect of different network segmentation 表 3. 不同网络分割效果对比

	Dice	Accuracy	IoU	mIoU
FCN8s [6]	0.8560	0.9210	0.7875	0.8215
SegNet [7]	0.8643	0.9344	0.7933	0.8398
U-Net [1]	0.8744	0.9422	0.8087	0.8453
PSP Net [8]	0.8977	0.9424	0.8292	0.8641
Deeplab-v3 [9]	0.9082	0.9465	0.8343	0.8703
CENet [10]	0.9317	0.9677	0.8568	0.8858
Attention U-Net [11]	0.9330	0.9680	0.8789	0.8894
Ours	0.9341	0.9765	0.9056	0.9152

2	••	••	ħ	P.	R .•
0)	4	•	• •	••	8
0	••	••	••	4 N	Ø 10
Ø 0	••	••	••	••	••
(a) Figure 11 Com	(b)	(c)	(d)	(e)	(f)

图 11. 不同网络分割效果对比图

为了验证提出的 Transformer 模块、多尺度通道和空间注意力(MSCA)模块的有效性,在小儿肾图上

分别对这三个模块进行了消融实验,实验结果如表 4 所示。由于 Transformer 层能够较充分地提取目标的 位置关系、部分与整体之间的关系信息,与 U-Net 相比,嵌入 Transformer 层后,Dice 提高 1.1%。由于 利用提出的注意力机制强调信息量大的特征,抑制无效的特征,并融合了多尺度特征。与原始 U-Net 相 比,嵌入多尺度通道和空间注意力模块(MSCA)后,Dice 提高了 1.1%。与原始 U-Net 相比,提出的方法 Dice 提升了 3.8%。因此,消融实验验证了提出的每个模块都能够提高模型的精度。由图 12 所示,提出 的方法对目标边缘的分割效果最好。



Figure 12. Comparison chart of ablation experiments 图 12. 消融实验对比图

Table	. Comparison of segmentation effect of different module	s
表 4.	下同网络分割效果对比	

Transformer	MSCA	Dice	Accuracy	mIoU	IoU
		0.8963	0.9345	0.8865	0.8932
	\checkmark	0.9074	0.9383	0.8893	0.9020
\checkmark		0.8993	0.9651	0.8972	0.9054
\checkmark	\checkmark	0.9341	0.9765	0.9056	0.9152

4. 总结

本文提出了一种全新的网络架构,解决了 U-Net 在小儿肾图分割中对全局信息的部分损失问题。 U-Net 的编码器通过多次下采样降低了输入图像的空间分辨率,导致模型无法完全利用图像的全局上下 文信息,且下采样过程丢失了大量重要信息。因此,将 Transformer 架构引入 U-Net 的深层学习目标位置 信息及局部与全局的关系。同时,设计了多尺度通道和空间注意力模块(MSCA),该模块融合了多尺度特征以弥补网络池化过程中丢失的信息,并且有效地捕捉了特征图的全局上下文信息,抑制了背景噪声。最后,本文通过对新华医院收集的小儿肾图数据进行大量实验,包括对比实验与消融实验,验证了所提方法的有效性。总之,提出的模型通过多尺度通道和空间注意力模块(MSCA)和 Transformer 层来自动识别和分割肾轮廓,在准确性上超越了现有模型。但是,复杂的网络会增加更多的训练参数,训练和推理的时间增长,这将在下一步的工作中进行改进。

参考文献

- Zhang, Z., Wu, C., Coleman, S. and Kerr, D. (2020) Dense-inception U-Net for Medical Image Segmentation. *Computer Methods and Programs in Biomedicine*, **192**, Article ID: 105395. <u>https://doi.org/10.1016/j.cmpb.2020.105395</u>
- [2] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. arXiv: 1706.03762.
- [3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2020) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [4] Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E. (2020) Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 2011-2023. <u>https://doi.org/10.1109/tpami.2019.2913372</u>
- [5] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer International Publishing, 3-19. <u>https://doi.org/10.1007/978-3-030-01234-2_1</u>
- [6] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/cvpr.2015.7298965
- [7] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <u>https://doi.org/10.1109/tpami.2016.2644615</u>
- [8] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017). Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 21-26 July 2017, 6230-6239. <u>https://doi.org/10.1109/cvpr.2017.660</u>
- [9] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., et al. (2019) Ce-net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 38, 2281-2292. <u>https://doi.org/10.1109/tmi.2019.2903562</u>
- [10] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020*, Springer International Publishing, 213-229. <u>https://doi.org/10.1007/978-3-030-58452-8_13</u>
- [11] Oktay, O., Schlemper, J., Folgoc, L., et al. (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999.