

一种基于课程学习的胚胎图像语义分割方法

汤红忠^{1,2}, 王 蔚¹, 王 涛¹, 陆旺达³, 黄向红⁴, 章 兢^{1,3}

(1. 湘潭大学自动化与电子信息学院, 湖南湘潭 411105; 2. 湘潭大学智能计算与信息处理教育部重点实验室, 湖南湘潭 411105;
3. 湖南大学电气与信息工程学院, 湖南长沙 410082; 4. 湘潭市中心医院, 湖南湘潭 411199)

摘要: 胚胎植入前形态学特征是人类体外受精胚胎质量评估的重要依据。目前, 胚胎学家主要利用胚胎时差成像(Time-Lapse Imaging, TLI)技术观察胚胎图像形态变化, 从而筛选最有发育潜能的胚胎进行移植或冷冻保存。然而, 人工评估不仅费时费力, 且需要较强的专业知识, 并存在一定的主观性等。针对这一问题, 该文提出一种基于课程学习的胚胎图像语义分割方法, 实现了胚胎细胞、细胞质与雌雄原核的分割, 为后续胚胎质量评估提供定量的形态特征参数。首先, 利用评估语义分割算法性能的IoU指标(Intersection over Union, IoU)构建课程学习的难度评分函数(Scoring Function, SF), 根据SF评分将所有样本从易到难进行排序; 再结合SF评分与目标类别数定义课程学习的步调函数(Pacing Functions, PF), 构建了难易程度递增的样本子集; 最后, 设计多阶段渐进式U-net语义分割(Multi-Stage Progressive U-net, MSPU)模型, 根据课程难度顺序依次训练不同阶段的网络, 从而实现胚胎图像的语义分割。相关实验结果表明, 本文提出的MSPU模型在胚胎图像语义分割任务上获得较好的性能, 与基准模型相比, IoU值提高1.4%; 特别是在较易与较难的分割任务上具有不错的表现, 如细胞与雌雄原核的分割IoU值分别提升了4.6%与1.2%。

关键词: 课程学习; 胚胎时差图像; 多阶段渐进式U-net网络; 语义分割; 评分函数; 步调函数

基金项目: 湖南省出生缺陷协同防治科技重大专项子项目(No.2019SK1010); 湖南省自然科学基金(No.2023JJ50242); 国家自然科学基金区域创新发展联合基金子课题(No.U19A2083); 湖南省教育厅重点项目(No.22A0115); 湘潭大学智能计算与信息处理教育部重点实验室(No.2020ICIP06)

中图分类号: TP301.6; TP391.41 **文献标识码:** A

文章编号: 0372-2112(2023)11-3365-12

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230097

A Semantic Segmentation Method of Embryo Image Based on Curriculum Learning

TANG Hong-zhong^{1,2}, WANG Wei¹, WANG Tao¹, LU Wang-da³, HUANG Xiang-hong⁴, ZHANG Jin^{1,3}

(1. College of Automation and Electronic Information, Xiangtan University, Xiangtan, Hunan 411105, China;

2. Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education,
Xiangtan University, Xiangtan, Hunan 411105, China;

3. College of Electrical and Information Engineering, Hunan University, Changsha, Hunan 410082, China;

4. Xiangtan Central Hospital, Xiangtan, Hunan 411199, China)

Abstract: Morphological features of embryo play an important role to evaluate the quality of human embryos in vitro fertilization. Embryologists mainly use time-lapse imaging (TLI) technology to observe the morphological variation of embryo images and select the most potential embryos for subsequent transfer or cryopreservation. However, manual evaluation is not only time-consuming and laborious, but it also demands specialty-oriented skills and has some subjectivities. To solve this problem, we propose a semantic segmentation method of embryo images based on curriculum learning to segment embryo cell, cytoplasm and pronucleus, which can provide the quantitative parameters of morphological feature for the subsequent evaluation of embryo quality. First, IoU (Intersection over Union) index that is usually used to evaluate the performance of semantic segmentation algorithms is adopted to construct a difficulty scoring function (SF) of curriculum learning. All samples are sorted from easy to difficult according to SF value. Second, we define a pace function (PF) by combining the SF and the number of target categories, and sample subsets is established with increasing sequence of difficulty. Lastly, we design a multi-stage progressive U-net (MSPU) model to segment embryo cell, cytoplasm and pronucleus em-

bryo images at pronuclear-stage, in which the network at different stages are trained using the sample subsets with increasing sequence of difficulty. Experimental results demonstrate that our proposed MSPU model obtains a satisfactory performance on the semantic segmentation of embryo images, and IoU is improved by 1.4% compared with Vanilla Baseline. Our proposed MSPU model shows a pleasing consistency on the easy and difficult segmentation tasks, for example the improvement of 4.6% and 1.2% can be obtained for the segmentation of embryonic cells and pronucleus, respectively.

Key words: curriculum learning; embryo time-lapse image; multi-stage progressive U-net; semantic segmentation; scoring function; pacing function

Foundation Item(s): Major Project of Cooperative Prevention and Control of Birth Defects in Hunan Province (No.2019SK1010); Joint Fund of Natural Science of Hunan province (No.2023JJ50242); Joint Fund for Regional Innovation and Development of National Natural Science Foundation in China (No.U19A2083); Key Projects of Education Department of Hunan Province (No.22A0115); Open Project of Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University (No.2020ICIP06)

1 引言

体外受精-胚胎移植(In-Vitro Fertilization, IVF)作为一种辅助生殖技术,为人类不孕不育提供了一种有效的解决方法^[1,2]. 胚胎质量评估已成为辅助生殖领域的热点研究问题,其中胚胎植入前形态学评估成为人类体外受精胚胎质量评估的关键方法^[3,4]. 胚胎学家利用时差成像(Time-Lapse Imaging, TLI)技术观察胚胎动态发育图像,并手工计算胚胎图像的空间形态参数评价胚胎质量好坏,从而选择最有发育潜力的胚胎进行移植^[5,6]. 这种人工评估不仅费时费力,且具有一定主观性. 因此,在胚胎发育的特定时期如何实现空间形态参数智能测量已成为现阶段辅助生殖技术领域中的热点研究问题之一^[7,8].

近年来,随着深度学习理论在自然图像上取得突破性的进展,涌现出一系列的基于深度学习的胚胎发育智能评估技术成果^[9-12]. 这些工作大概可分为以下两类:(1)基于序列图像的胚胎质量评估^[9,13-15]; (2)基于非序列图像的胚胎质量评估^[10-12,16,17]. 上述工作大多没有关注胚胎发育原核期图像特征. 因此,本文围绕受精后20小时内处于原核期的胚胎图像开展研究,实现原核期形态特征参数的智能测定.

2009年, Bengio 等人^[18]第一次提出了课程学习(Curriculum Learning, CL)的范式. 课程学习将机器学习模型分为多个阶段训练,每个阶段采用指定难度的样本子集训练模型. 按课程的设计方式,课程学习方法主要分为两大类^[19,20]:(1)预定义的课程学习^[21,22,26],如采用样本复杂度^[21-23]、数据分布^[24]、信号噪声^[25]等先验知识来设计样本难度评分函数从而满足不同应用领域的需求. 这类方法的局限性在于忽略了模型的状态和反馈,导致模型无法根据数据的难易程度动态调整. (2)自动的课程学习^[27-29],这类方法以数据驱动方式来设计课程先后顺序^[27-30]. 例如自步学习方法(Self-Paced Learning, SPL)^[28,29]采用损失函数同时决定模型

损失与课程学习难度评分函数. Jiang 等人^[30]提出一种自步课程学习方法(Self-Paced Curriculum Learning, SPCL),使得模型对于样本难度的判断更加准确. 为了适应不同应用领域,基于教师网络的自动课程学习方法^[31,32]提出了更多的难度评分准则. 可见,样本的难度挖掘与训练调度上仍然有诸多问题亟需解决^[33].

目前,课程学习已成功应用于自然图像^[27,28]或医学图像的分类^[26,34,35]领域,但是,结合课程学习策略实现图像的语义分割工作并未多见. Zhang 与 Dai 等人^[36,37]结合特定领域知识合成街景图像,开展了街景语义分割的研究. 然而,受特定领域知识的影响,这些策略与方法并不能直接将其应用于医学图像. 因此,根据胚胎发育原核期图像特点,本文首次尝试将课程学习应用于胚胎图像,设计一个多阶段胚胎图像语义分割网络模型,实现胚胎细胞、细胞质、原核的自动分割.

2 本文提出的课程学习算法

2.1 问题描述

TLI拍摄的胚胎发育图像,每5分钟采一帧,主要存在的问题:(1)TLI图像动态变化,存在重叠、伪影,甚至会失真;(2)灰度变化极其平缓,胚胎中各个目标形状不规则且边界模糊;(3)胚胎细胞中雌雄原核(下文简称原核)体积小,且一个胚胎视频中具有原核的图像帧数也非常少,导致图像类别极不均衡. 这些问题给语义分割任务带来一定的挑战.

相关符号说明如下:带标注的训练样本集 $T = \{T_i\}_{i=1}^N = \{(x_i, y_i)\}_{i=1}^N$, x_i 表示第 i 幅图像, y_i 表示第 i 幅图像的标注. 若采用 T 训练语义分割网络 $H(\cdot)$, 则 $\hat{y}_i = H(x_i; \omega)$, $H(\cdot)$ 代表样本与相应的标注之间的映射关系, \hat{y}_i 为网络模型预测结果; ω 表示模型参数. 相应的, $L_\omega(T) = L(y_i, \hat{y}_i)$ 定义为损失函数(Loss), 表示预测值 \hat{y}_i 与标注 y_i 之间的损失. 则目标函数为 $\arg\min_{\omega} \sum_{i=1}^N L_\omega(T_i)$;

采用小批次梯度下降方法优化网络模型参数 ω . 引入课程学习后, 基于课程难度对样本集 T 进行排序, 排序后新的样本集为 T' , 划分的样本子集定义为 $\psi_1, \psi_2, \dots, \psi_{C(T)}$, 其中, $C(T)$ 表示样本类别数.

2.2 自步学习

传统机器学习方法大都通过随机策略选择样本训练网络模型. 然而, 这样的方式可能无法挖掘训练样本中潜在的常识推理与因果逻辑关系^[38,39]. 2009年 Bengio 等人^[18]提出了一种课程学习策略. 即根据先验知识定义样本学习难度, 利用样本难易程度依次训练机器学习模型. 这种方式对先验领域知识的要求较高. 为解决这一问题, SPL 将样本学习的难度评分嵌入到模型优化的目标函数中, 定义如下^[28]:

$$\underset{\omega, v}{\operatorname{argmin}} L(\omega, v) = \sum_{i=1}^N v_i L_{\omega}(T_i) - \gamma \sum_{i=1}^N v_i, \quad (1)$$

$$v_i = \begin{cases} 1, & L_{\omega}(T_i) < \gamma \\ 0, & \text{otherwise} \end{cases}$$

式(1)中第一项是每个 Loss 的加权和; 第二项为权重 v 的正则项. SPL 采用交替凸搜索算法 (Alternative Convex Search, ACS)^[29]来优化式(1). 优化步骤为: (1) 固定模型参数 ω , 计算样本的损失, 然后更新 v_i ; 当样本损失小于 γ 时, 该样本作为“容易”样本, $v_i = 1$; 否则, $v_i = 0$, 该样本作为“更难”的样本; (2) 固定 v , 更新模型参数

ω ; 由此可见, SPL 倾向于从所有样本中选择具有更小训练误差的样本进行模型学习, 从而更新模型参数. 然而, 网络训练与课程学习策略都由模型的损失函数来决定, 容易造成模型的过拟合问题^[30].

2.3 本文提出的 MSPU 模型

本文提出了一种多阶段渐进式胚胎图像语义分割 (Multi-Stage Progressive U-net, MSPU) 模型, 基于从易到难的课程学习策略, 分阶段实现胚胎细胞、细胞质和原核的分割, 具体流程如图 1 所示.

(1) 随机选择训练样本, 完成单阶段的 U-net 网络训练, 获得每个样本的 IoU 值, 将 IoU 值作为课程学习的性能度量 (Performance Measure, PM);

(2) 基于样本的 IoU 值构建课程学习的难度评分函数 (Scoring Function, SF), 根据 SF 值将所有训练样本以单调递增方式进行排序;

(3) 结合 SF 值与目标类别数构建课程学习的步调函数 (Pacing Functions, PF), 根据 PF 值, 建立难易程度递增的样本子集, 包括容易样本子集 (Easy Samples, ES)、中等难度样本子集 (Medium Samples, MS)、困难样本子集 (Hard Samples, HS);

(4) 最后, 设计一个多阶段渐进式 (MSPU) 胚胎图像语义分割模型, 根据从易到难的顺序, 采用步骤 3 中获得的样本子集对不同阶段的网络逐步训练与优化.

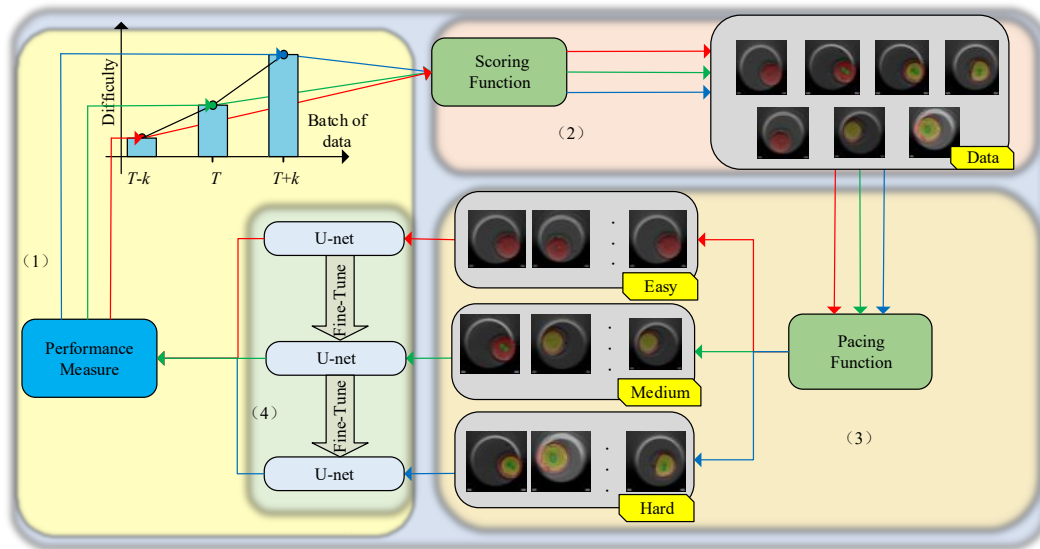


图 1 本文提出的 MSPU 模型

2.3.1 本文的课程学习策略

(1) 难度评分函数

如图 2 所示, 胚胎学家勾勒胚胎图像目标区域时任务难易程度不一, 从左到右, 胚胎细胞边界清晰, 勾勒容易, 细胞质图像边界模糊, 勾勒难, 而两个原核边界

模糊且易重叠, 勾勒最为困难.

设计合适的学习顺序是课程学习中的关键问题. 本文在模型损失与难度评分上采用了不同的损失函数. 医学图像分割包括两种损失函数形式^[40]: (a) 基于像素类别的损失函数, 如交叉熵损失; (b) 基于区域的

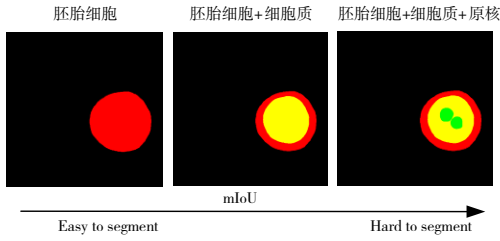


图2 样本难度变化

损失函数,如IoU损失.这两者之间存在着一定的正相关关系.更重要的是,IoU与交叉熵之间并不是简单的线性正比关系.

为了构造课程学习的SF,首先随机选择训练样本,实现单阶段U-net网络的训练,从而获得评估语义分割算法性能的IoU.本文采用样本IoU值来构建课程学习的样本难度的SF,定义如下:

$$f(T_i) = f(x_i, y_i) = e^{-\lambda \frac{|t_m \cap p_m|}{|t_m \cup p_m|}} \quad (2)$$

其中, t_m 与 p_m 分别定义为图像 x_i 中第 m 个像素的标注与预测结果, $|t_m \cap p_m|$ 与 $|t_m \cup p_m|$ 分别表示预测结果与标注的交集与并集. $\lambda > 0$, 为设定的超参数.若 $f(x_i, y_i) > f(x_j, y_j)$, 则表示样本 x_i 分割难度大于样本 y_j . 基于式(2), 根据SF值对初始训练样本集 T 从小到大排序, 则排序后的样本集定义为 $T' = \text{rank}(T, \text{SF})$.

(2) 步调函数

如何对不同学习难度的样本进行调度是课程学习又一关键步骤.步调太大造成对当前样本学习不充分,太小则会造成样本学习冗余.因此本文结合语义分割目标类别数构建PF,具体定义如下:

$$\psi_{C(T)} = g(T, C(T)) \quad (3)$$

其中, $C(T)$ 表示训练样本的目标类别数.如图3所示, $C(T) = \{1, 2, 3\}$. 因此,PF可将 T' 分为难易程度不同的三个样本子集(ES、MS、HS),且样本子集中的小批次数据也以从易到难的顺序进行排列.

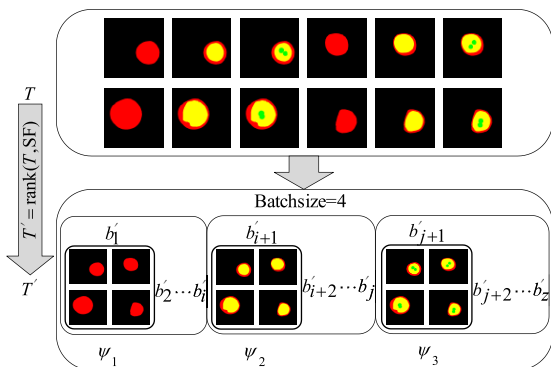


图3 基于步调函数划分样本子集

(3) 目标函数

结合式(2)与式(3),本文目标函数定义为

$$\underset{\omega, \nu}{\operatorname{argmin}} L'(\omega, \nu, \psi) = \frac{1}{N} \sum_{i=1}^N \nu_i L_{\omega}(T_i) - \delta \frac{1}{N} \sum_{i=1}^N \nu_i, \quad (4)$$

$$\nu_i = \begin{cases} 1, & \text{if } f(T_i) < \delta \text{ and } T_i \in \psi_{C(T)} \\ 0, & \text{otherwise} \end{cases}$$

其中, $L_{\omega}(T_i) = \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i))$ 为本文采用的交叉熵损失函数, y_i 与 \hat{y}_i 分别为第 i 幅图像的标注与预测值.与式(1)不同,式(4)中 ν 取值由SF与PF共同决定,其中 $f(T_i)$ 为难度评分函数SF, ψ 为步调函数PF.可见,向量 ν_i 用于控制网络训练过程中的样本序列选择.当样本的难度 $f(T_i)$ 小于 δ , 且它属于PF划分的对应子集时, $\nu_i = 1$, 表示该样本为容易的样本, 且被选择加入模型学习, 更新网络权重; 否则为 $\nu_i = 0$, 样本不被选择加入网络学习, 这种方式可以有效避免模型的过拟合现象.

(4) 本文提出的课程学习算法流程

算法1给出本文提出的课程学习算法流程.首先用初始训练样本集 T 训练语义分割模型 $H(\cdot)$, 将IoU指标作为 $H(\cdot)$ 的PM; 计算SF, 根据SF对 T 进行排序获得数据集 T' ; 将 T' 划分为样本子集ES、MS、HS三个不同的样本子集.最后逐步训练网络模型 $H(\cdot)$, 每次加入的样本子集经过 p 轮训练后, 待模型收敛, 再继续增加更难样本进入训练集, 不断优化模型直至最后收敛.

算法1 本文提出的课程学习算法

输入: 初始训练样本集 T

输出: 最优的网络模型参数 $\tilde{\omega}$

1: 初始化 深度模型 $H(\cdot)$, 网络参数 ω , 网络训练次数 $p, T^{\text{train}} = \phi$

2: $\text{PM} \leftarrow H(\text{IoU})$

3: $\text{SF} \leftarrow f(\text{IoU})$

4: $T' = \text{rank}(T, \text{SF})$

5: $\{\psi_1, \psi_2, \dots, \psi_{C(T)}\} \leftarrow \text{PF}(T')$

6: for $s = 1, 2, \dots, C(T)$ do

7: $T^{\text{train}} = T^{\text{train}} \cup \psi_s$

8: while not converged for p epochs do

9: $h_{\omega} \leftarrow \text{train}(H, T^{\text{train}})$

10: end while

11: end for

2.3.2 MSPU 模型

(1) MSPU 模型

本文提出的MSPU模型采用U-net网络^[41]作为骨干网络,共包括三个结构完全相同的子网络,如图4所示.每个子网络主由编码网络与解码网络组,并且在编码

和解码之间引入跳跃连接,使得两者能够共享网络权重,以达到更好的分割效果.

(2) MSPU 模型的优化步骤

采用算法 1 实现不同阶段子网络模型的训练与优化,共包括三个步骤:

(a) 如图 4 中的第一个框图,首先训练基准网络 (Vanilla Baseline), 即利用所有的训练样本, 随机选择样本方式实现单阶段 U-net 网络模型的训练.

(b) 根据 Vanilla Baseline 的 IoU 指标, 计算样本的

SF 与 PF, 按照从易到难的顺序将样本划分为子集 ES、MS 与 HS.

(c) 如图 4 的第二个框图, 多阶段胚胎图像语义分割网络训练分为三个阶段: Stage 1: 利用样本子集 ES 训练 U-net 网络, 直至网络收敛; Stage 2: 将 Stage 1 收敛后的模型作为预训练模型, 采用样本子集 MS 进行该阶段的模型训练, 微调网络参数, 直至模型收敛; Stage 3: 将 Stage 2 模型作为预训练模型, 并采用样本子集 HS 完成该阶段的模型训练, 微调网络参数, 直至网络模型最终收敛.

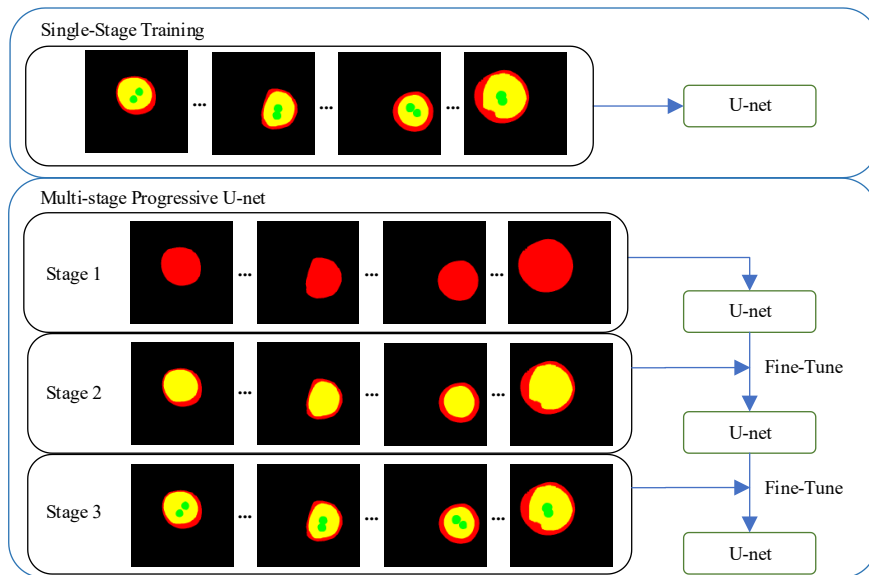


图 4 本文提出的 MSPU 模型及优化过程

3 实验

3.1 实验配置

实验平台配置如下: 处理器为 Intel (R) Core (TM) i7-8700 CPU @ 3.2 GHz; RAM 为 32 GB; GPU 为 GeForce GTX 2080 Titan 11 GB 显存; 计算机操作系统为 Windows 10; 开发工具为 Python, Pytorch 框架. 网络模型训练时, 采用 Adam 优化器, 学习率设置为 0.000 1, Batch Size 为 4.

3.2 数据集

本文制作了胚胎图像语义分割数据集, 数据来源于湘潭市中心医院生殖遗传中心, 总共包含 18 组胚胎的 Time-Lapse 序列图像, 每段序列图像拍摄时长 5~6 天, 每隔 5 分钟采样一帧图像. 表 2 给出了数据集分布, 交叉验证依据胚胎细胞划分, 即训练集、验证集与测试集不会存在同一个胚胎细胞. 训练集、验证集与测试集按图像帧数量比例划分约为 8:1:1. 表中数据表示图像数量以及胚胎数量.

表 1 胚胎数据集分布

	Train	Validation	Test	Total
细胞	3 953(10)	564(4)	565(4)	5 082(18)
细胞+细胞质	1 818(10)	304(4)	303(4)	2 425(18)
细胞+细胞质+原核	1 778(10)	223(4)	222(4)	2 223(18)

3.3 评价指标

为了定量评估语义分割模型的性能, 本文采用 IoU 和 Dice 相似系数 (Dice Similarity Coefficient, DSC) 作为评价指标. IoU 的定义如下:

$$\text{IoU} = \frac{|t_m \cap p_m|}{|t_m \cup p_m|} \quad (5)$$

DSC 的定义如下:

$$\text{DSC} = \frac{2|t_m \cap p_m|}{|t_m| + |p_m|} \quad (6)$$

其中, $|t_m| + |p_m|$ 为标注与模型预测的面积之和. 由 IoU 和 DSC 值区间范围为 $[0, 1]$, 值越大表示模型分割性能越好.

为进一步说明问题,本文引入精确率(Precision)与召回率(Recall)综合评估模型性能. Precision与Recall的定义如下:

$$\text{Precision} = \frac{|t_m \cap p_m|}{|p_m|} \quad (7)$$

$$\text{Recall} = \frac{|t_m \cap p_m|}{|g_m|} \quad (8)$$

Precision与Recall分别表示预测结果准确率与找全率.

3.4 实验结果分析

(1) 骨干网络的选取

为了选择合适的骨干网络,本节对比了U-net、FCN^[42]、Segnet^[43]与Deeplabv3+^[44]在胚胎图像语义分割上的性能,如表3所示.由表2可知,与其他骨干网络相比,U-net获得了最好的分割性能,IoU值最高,且网络参数量也较少.因此,本文选取U-net为胚胎语义分割的主干网络.

表2 不同网络的分割性能与模型参数量

Model	U-net	FCN	Segnet	Deeplabv3+
IoU/%	93.2	93.1	91.9	92.1
Params/M	31.1	69.7	3.8	41.1

(2) 样本难度评分函数的构建

为了构建样本难度评分函数,本文以随机方式选择样本实现单阶段U-net网络模型的训练.根据网络预测结果,可以获得不同目标的IoU、Dice、Precision与Recall的平均值,如图5所示.

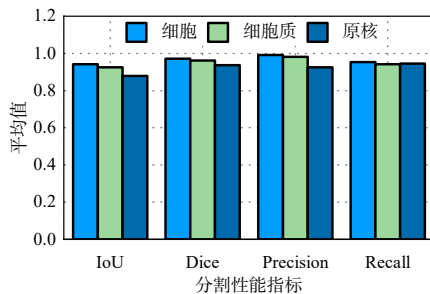


图5 不同类别的分割性能指标

在这四个指标中,IoU值不仅严格单调递减,而且不同类别之间差异度相当.由此可知,IoU值可以有效量化胚胎细胞、细胞质、原核分割的难易程度,因此,本文采用IoU值来构建了样本难度评分函数.

(3) 不同课程学习策略对模型性能的影响

为了验证本文提出的MSPU模型的有效性,将其它与无课程学习策略的单阶段网络模型、引入课程学习策略的多阶段网络模型进行对比,主要包括:

(a) 无课程学习策略的单阶段网络模型(Vanilla

Baseline):以随机方式选择训练全部样本完成单阶段U-net网络训练;同时探索了不同的单阶段网络模型,即分别采用样本子集ES、MS或者HS分别完成单阶段U-net网络训练;

(b)引入课程学习策略的多阶段网络模型,具体包括反课程学习策略(Anti-Curriculum),即SF为 $f'=-f$,即越困难的样本被优先加入样本队列训练网络模型;随机课程学习策略(Random-Curriculum),即随机定义SF,随机定义实质上有多种排列方式;Our Curriculum是本文提出的从易到难的课程学习策略.

如表3所示,给出了不同训练策略下模型分割获得的IoU.与Vanilla Baseline相比,采用样本子集ES训练的网络,在较容易样本分割时虽然呈现出更好的性能;然而,仅仅采用样本子集HS训练的单阶段网络,在原核分割任务上IoU值下降很快.可知,单阶段网络模型无法根据样本难易程度进行模型的在线调整与优化,在不同难度的目标分割任务上无法同时获得令人满意的结果.

Anti-Curriculum尽管在最难的原核分割任务上获得了最好的性能,但IoU平均值最低,其原因在于反课程学习策略使得网络模型更加关注更难样本的学习,在一定程度上忽视了容易样本的学习,使得模型整体分割性能不够理想,如在最容易的细胞分割任务上性能反而最差.

与Vanilla Baseline相比,MSPU模型在细胞的分割性能上有较大提升,IoU提升了4.6%,在原核分割任务上,也获得了不错的分割结果,IoU提升了1.2%.可以看出,本文提出的MSPU模型采用了从易到难的课程学习策略,使得网络模型整体分割性能上表现最佳.由此可知,本文提出的MSPU模型,通过不同阶段选择不同难易程度的样本子集逐步优化网络模型的学习,本质上是对网络学习施加了引导性的偏好约束,促使网络模型挖掘不同数据子集中潜在特征与更多的判别信息,逐步提升了网络模型的分割性能,也大大降低了背景噪声的干扰.

图6给出了不同模型分割结果的可视化图.图6中的第二行为第一行标记区域的细节放大图.由图可知,单阶段网络模型分割时容易出现错分与误分,如容易将气泡等噪声误分为细胞或者原核,如图6(c)中的第二行所示.通过细节图也可以发现,在细胞质区域,Anti-Curriculum与Random-Curriculum分割的细胞质边界(黄色)明显向外扩张,原核边界(绿色)没有保持与金标准一致的边缘弧度.而本文提出的模型在更难的分割任务上(如细胞质与原核),不仅获得了更为清晰的边界,而且保持了目标区域的平滑性,获得了与金标准更为一致的分割结果,验证本文提出方法的有效性,

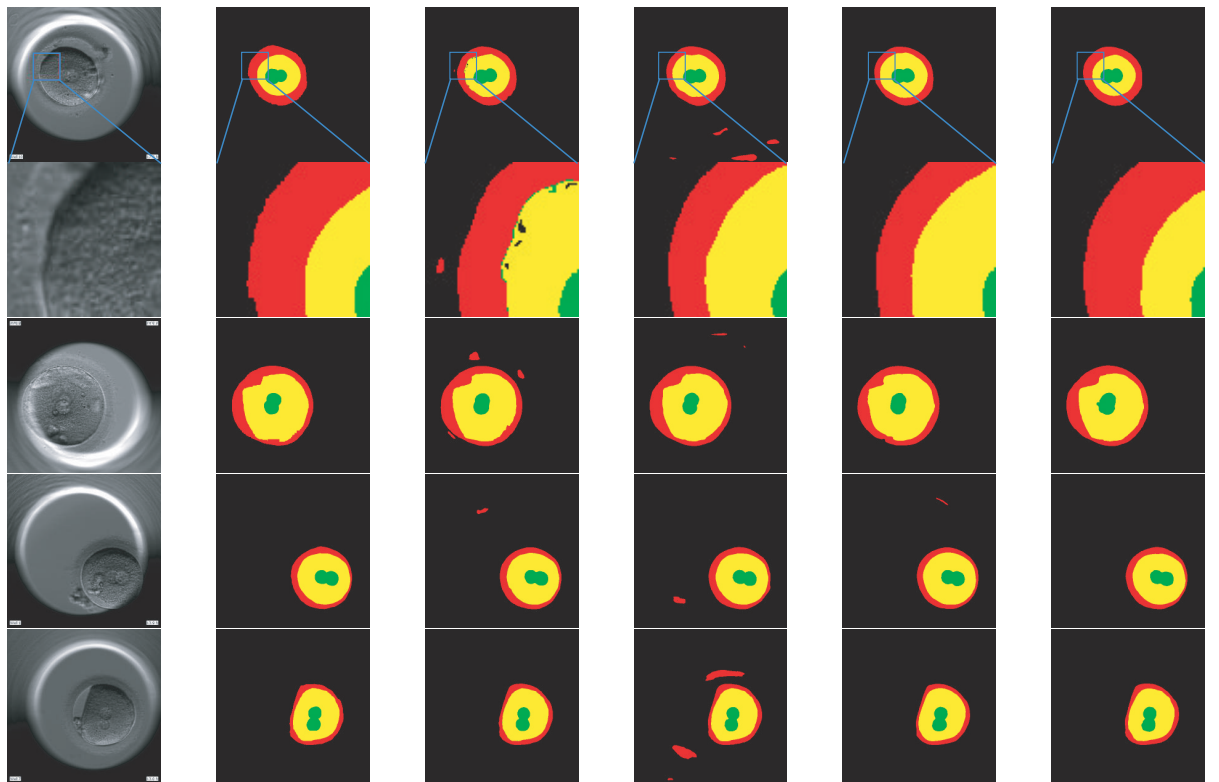
表 3 各类训练策略在测试集上的分割性能

Model	CL	训练阶段	细胞	细胞质	原核	overall
Vanilla Baseline	w/o	细胞+细胞质+原核	0.865	0.966	0.903	0.932
Easy images only	w/o	细胞	0.941	—	—	0.915
Medium images only		细胞质	—	0.925	—	
Hard images only		原核	—	—	0.879	
ANTI-Curriculum	w/	原核	—	—	0.879	0.908
		原核+细胞质	—	0.951	0.914	
		原核+细胞质+细胞	0.758	0.964	0.960	
Random-Curriculum	w/	细胞	0.941	—	—	0.923
		细胞+原核	0.966	—	0.912	
		细胞+原核+细胞质	0.847	0.935	0.916	
Our Curriculum	w/	细胞	0.941	—	—	0.946
		细胞+细胞质	0.900	0.978	0.912	
		细胞+细胞质+原核	0.911	0.962	0.915	

特别能在更困难的分割任务上取得一定提升.

结合表3与图6,发现模型在细胞质和原核分割性能较显弱势,其原因可能在于:(1)在一个胚胎序列图像中,具有细胞的图像帧数(约为350帧)远远多于具有原核与细胞质的图像帧数(约为100帧).而且在单幅图像中,原核目标极小,在整个图像中像素占比极

少,造成类别极不均衡的现象;(2)从图6的细节放大图可以看出,细胞与背景边界分明,分割难度最低;细胞与细胞质边界模糊,灰度平滑,分割难度加大;核仁边界与细胞质边界的更为模糊,且目标更小,很明显分割难度最大.上述因素给胚胎图像语义分割的带来了一定的挑战.



(a) 胚胎图像 (b) 金标准 (c) Vanilla Baseline (d) ANTI-Curriculum (e) Random-Curriculum (f) Our Curriculum

图6 不同训练策略在胚胎图像上的分割结果

为了进一步验证不同课程学习策略对模型性能的影响,本文对比分析了18组不同胚胎图像的IoU值分布,如图7所示.横坐标代表不同课程学习策略,纵坐标代表分割的IoU值.可以看出,本文提出的方法获得了最高精度,方差最小,且离散点数量最少,说明本文提出的课程学习策略有利于提高胚胎图像语义分割模型鲁棒性.

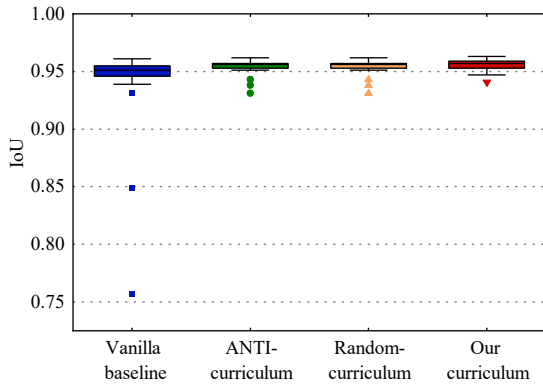


图7 18个胚胎上的分割结果分布箱线图

图8给出了不同课程学习策略对网络模型收敛性能和精度的影响,可以看出,引入课程学习后可以大大提高网络模型的收敛速度,尽管ANTI-Curriculum模型收敛速度更快,而本文的课程学习策略能同时平衡网络收敛性能与分割精度.

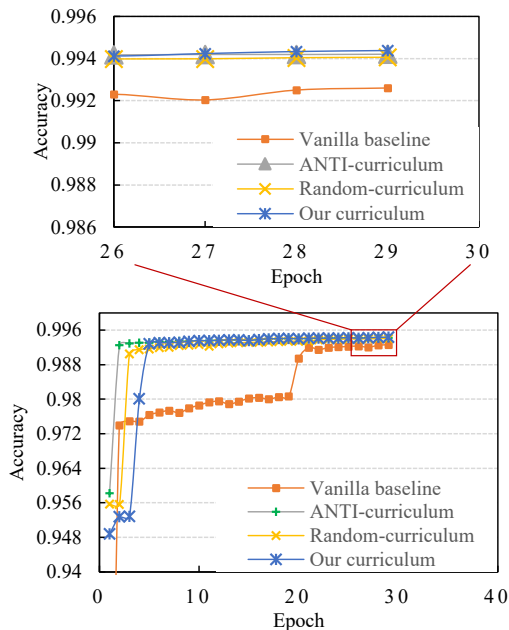


图8 不同课程学习策略下的网络收敛性能

3.5 公开数据集SegTHOR上的实验

为了进一步对比分析本文提出课程学习策略在语义分割模型上的有效性,本文在公开数据集胸腔危机器官分割SegTHOR^[45]上进行了对比实验.

SegTHOR胸腔危及器官数据集是一种计算机断层扫描(Computed Tomography, CT)图像. SegTHOR数据集与胚胎数据集具有一定的相似性,具体包括:(1)均由序列图像构成;(2)两种数据集中均为灰度图像,对比度低,目标边界模糊与不清晰;(3)不同目标均存在一定的难度梯度,不同器官之间外形尺寸存在较大的差异,如与心脏和主动脉相比,而气管和食道的分割难度更高;(4)不同病人的器官也有程度大小不一的形变,这一点正与本文中TIL技术采集到的人类胚胎图像类似,胚胎细胞、细胞质、雌雄原核尺寸大小不一,随着胚胎发育产生一定的形变,导致了分割模型在不同目标上呈现出较大差异的分割性能.

首先,将40个病人的胸腔图像以8:1:1的比例划分为训练集、验证集与测试集.为了度量不同网络模型在该数据集上的分割性能,采用DSC作为分割的度量指标^[46].其原因在于胸腔CT图像中存在极小的目标分割(气管与食道),像素占比约为整个切片图像的3.7%与2.6%.与IoU相比,采用DSC作为评价指标,更能体现不同类别的分割性能.

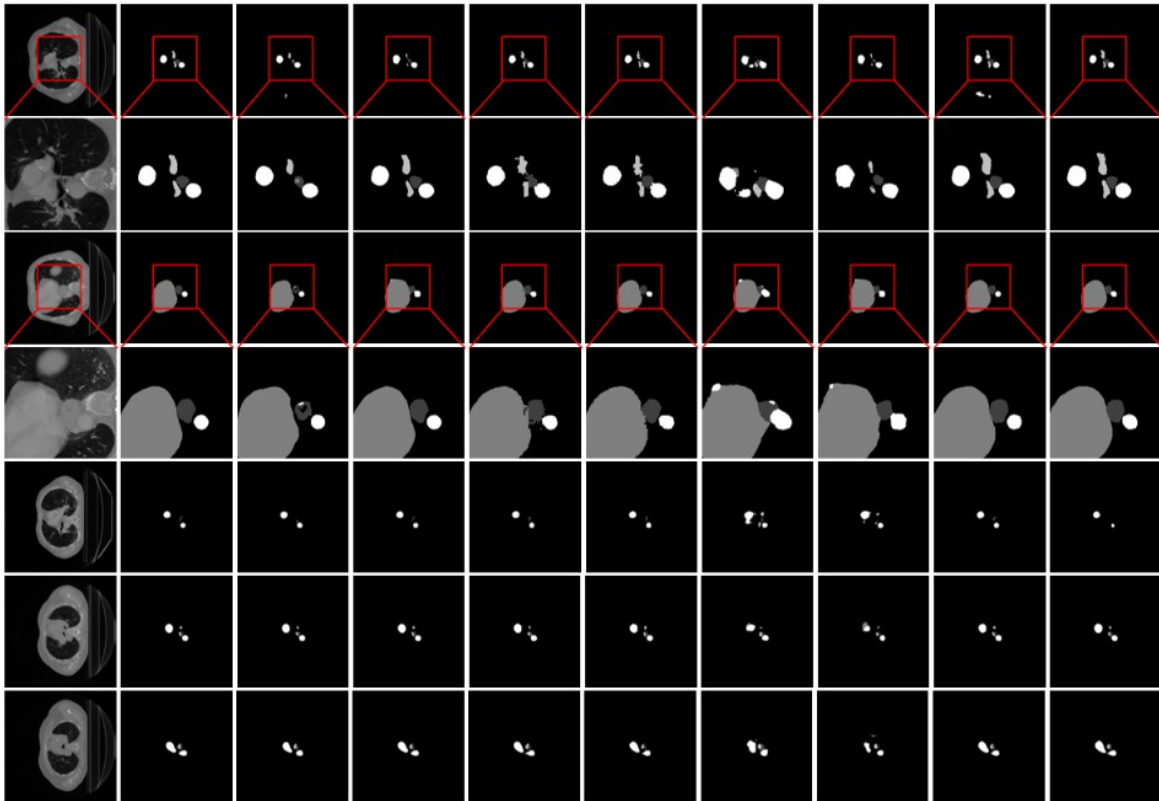
如表4所示, MSPU、MSPFCn、MSPSegnet、MSPDeepLabv3+表示分别以U-net、FCN、Segnet与DeepLabv3+作为骨干网络训练的多阶段模型.表4的结果表明,将本文提出的多阶段渐进式学习策略应用于不同骨干网络的优化,均能提升网络的分割性能,特别在样本分割难度更大的目标(食道、气管)上获得了更好的分割性能.同时,在各不同骨干网络方法改进上,通过横向对比可以观察到本文提出的MSPU模型获得更佳的整体分割性能.

表4 SegTHOR数据集上分割DSC值

模型	食道	心脏	气管	主动脉	Over-all
U-net	0.601	0.922	0.848	0.867	0.761
MSPU	0.936	0.988	0.882	0.933	0.893
Fcn	0.825	0.986	0.813	0.931	0.846
MSPFCn	0.888	0.988	0.891	0.918	0.886
Segnet	0.553	0.796	0.705	0.752	0.637
MSPSegnet	0.580	0.794	0.789	0.812	0.680
DeepLabv3+	0.870	0.972	0.779	0.892	0.832
MSPDeepLabv3+	0.864	0.962	0.868	0.901	0.855

图9给出了本文提出的课程学习策略对不同骨干网络的优化分割结果可视化图. 第二行与第四行是第一行和第三行的红色框区域的细节放大图. 由第四行的第三列与第四列可知,在较大目标(心脏)分割上,Unet与MSPU性能相当,但对于较小目标分割上(食道),MSPU获得了更好的分割性能,错分

率更小. 从图9中第二行与第四行可知,与单阶段网络模型相比,本文提出的MSPU获得了与金标准更为一致的分割结果. 这也说明本文提出的课程学习方法极大降低了困难目标(食道与气管)的错分率,提升了网络模型整体分割性能,分割的目标区域边界更为准确与清晰.



(a) CT图像 (b) 金标准 (c) U-net (d) MSPU (e) FCN (f) MSPFCN (g) Segnet (h) MSPSegnet (i) Deeplabv3+ (j) MSPDeeplabv3+

图9 不同方法在SegTHOR数据集上分割结果

4 总结与讨论

本文结合课程学习策略,提出了一种多阶段渐进式胚胎图像语义分割模型. 该模型利用评估语义分割算法性能指标IoU构建课程学习的难度评分函数,并结合目标类别数构建课程学习的步调函数,实现不同难度的样本子集划分,利用从易到难课程学习策略渐进式提升胚胎图像语义分割模型的性能,且使深度神经网络模型获得了更好的鲁棒性. 在未来工作中,将本文提出的模型应用于其他类型医学图像分割任务,从而验证模型的泛化性是未来的重点工作.

参考文献

- [1] HANEVIK H I, HESSEN D O. IVF and human evolution [J]. Human Reproduction Update, 2022, 28(4): 457-479.
- [2] PUISSANT F, VAN R M, BARLOW P, et al. Embryo scoring as a prognostic tool in IVF treatment[J]. Human Reproduction, 1987, 2(8): 705-708.
- [3] 耿蒙慧, 张璨, 邢阿英, 等. 胚胎质量评估方法的研究进展[J]. 中华妇幼临床医学杂志(电子版), 2018, 14(5): 612-616.
GENG M H, ZHANG C, XING A Y, et al. Research progress of methods on the assessment of embryonic quality[J]. Chinese Journal of Obstetrics & Gynecology and Pediatrics (Electronic Edition), 2018, 14(5): 612-616. (in Chinese)
- [4] MENG Q X, XU Y Y, ZHENG A Y, et al. Noninvasive embryo evaluation and selection by time-lapse monitoring vs. conventional morphologic assessment in women under-

- going in vitro fertilization/intracytoplasmic sperm injection: A single-center randomized controlled study[J]. *Fertility and Sterility*, 2022, 117(6): 1203-1212.
- [5] ARMSTRONG S, ARROLL N, CREE L M, et al. Time-lapse systems for embryo incubation and assessment in assisted reproduction[J]. *Cochrane Database of Systematic Reviews*, 2014(5): CD011320.
- [6] 黄国宁. Time-lapse 是否能预测胚胎染色体整倍性?[J]. *生殖医学杂志*, 2019, 28(11): 1249-1250.
HUANG G N. Could time-lapse predict euploidy embryos? [J]. *Journal of Reproductive Medicine*, 2019, 28(11): 1249-1250. (in Chinese)
- [7] ADAMSON G D, ABUSIEF M E, PALAO L, et al. Improved implantation rates of day 3 embryo transfers with the use of an automated time-lapse-enabled test to aid in embryo selection[J]. *Fertility and Sterility*, 2016, 105(2): 369-375.
- [8] WATANABE H, FUKUNAGA N, SANAMI S, et al. Development of an automatic pronuclear detection system for human embryos using deep learning technology[J]. *Fertility and Sterility*, 2019, 112(3): e429-e430.
- [9] LIAO Q Y, ZHANG Q, FENG X E, et al. Development of deep learning algorithms for predicting blastocyst formation and quality by time-lapse monitoring[J]. *Communications Biology*, 2021, 4(1): 415.
- [10] BORMANN C L, KANAKASABAPATHY M K, THIRUMALARAJU P, et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation[J]. *eLife*, 2020, 9: 55301.
- [11] LEAHY B D, JANG W D, YANG H Y, et al. Automated measurements of key morphological features of human embryos for IVF[C]//*Medical Image Computing and Computer Assisted Intervention — MICCAI 2020*. Cham: Springer International Publishing, 2020: 25-35.
- [12] ZHAO M P, XU M R, LI H H, et al. Application of convolutional neural network on early human embryo segmentation during in vitro fertilization[J]. *Journal of Cellular and Molecular Medicine*, 2021, 25(5): 2633-2644.
- [13] SAWADA Y, SATO T, NAGAYA M, et al. Evaluation of artificial intelligence using time-lapse images of IVF embryos to predict live birth[J]. *Reproductive Biomedicine Online*, 2021, 43(5): 843-852.
- [14] KRAGH M F, RIMESTAD J, BERNTSEN J, et al. Automatic grading of human blastocysts from time-lapse imaging[J]. *Computers in Biology and Medicine*, 2019, 115: 103494.
- [15] GARGEE V, SHREYA C, RUCHI G, et al. Time series prediction of viable embryo and automatic grading in IVF using deep learning[J]. *The Open Biomedical Engineering Journal*, 2021, 15(Suppl-2): 190-203.
- [16] VERMILYEA M, HALL J M M, DIAKIW S M, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF [J]. *Human Reproduction (Oxford, England)*, 2020, 35(4): 770-784.
- [17] THIRUMALARAJU P, KANAKASABAPATHY M K, BORMANN C L, et al. Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality[J]. *Heliyon*, 2021, 7(2): e06298.
- [18] BENGIO Y, LOURADOUR J, COLLOBERT R, et al. Curriculum learning[C]//*Proceedings of the 26th Annual International Conference on Machine Learning*. New York: ACM, 2009: 41-48.
- [19] WANG X, CHEN Y D, ZHU W W. A survey on curriculum learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(9): 4555-4576.
- [20] SOVIANY P, IONESCU R T, ROTA P, et al. Curriculum learning: A survey[J]. *International Journal of Computer Vision*, 2022, 130(6): 1526-1565.
- [21] PLATANIOS E A, STRETCU O, NEUBIG G, et al. Competence-based curriculum learning for neural machine translation[C]//*Proceedings of the 2019 Conference of the North American Association for Computational Linguistics*, 2019: 1162-1172.
- [22] SPITKOVSKY V I, ALSHAWI H, JURAFSKY D. From baby steps to Leapfrog: How “Less is more” in unsupervised dependency parsing[C]//*HLT 10: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. New York: ACM, 2010: 751-759.
- [23] WEI Y C, LIANG X D, CHEN Y P, et al. STC: A simple to complex framework for weakly-supervised semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(11): 2314-2320.
- [24] KOCMI T, BOJAR O. Curriculum learning and mini-batch bucketing in neural machine translation[C]//*RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning*. Incoma Ltd. Shoumen, Bulgaria

- ia, 2017: 379-386.
- [25] RANJAN S, HANSEN J H L. Curriculum learning based approaches for noise robust speaker recognition[J]. *ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(1): 197-210.
- [26] WEI J, SURIAWINATA A, REN B, et al. Learn like a pathologist: Curriculum learning by annotator agreement for histopathology image classification[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 2473-2483.
- [27] GRAVES A, BELLEMARE M G, MENICK J, et al. Automated curriculum learning for neural networks[C]//Proceedings of the 34th International Conference on Machine Learning - Volume 70. New York: ACM, 2017: 1311-1320.
- [28] KUMAR M P, PACKER B, KOLLER D. Self-paced learning for latent variable models[C]//The 23rd International Conference on Neural Information Processing Systems. New York: ACM, 2010: 1189-1197.
- [29] TULLIS J G, BENJAMIN A S. On the effectiveness of self-paced learning[J]. *Journal of Memory and Language*, 2011, 64(2): 109-118.
- [30] JIANG L, MENG D, ZHAO Q, et al. Self-paced curriculum learning[C]//The 29th AAAI Conference on Artificial Intelligence (AAAI). Menlo Park: AAAI Press, 2015: 2694-2700.
- [31] MATIISEN T, OLIVER A, COHEN T, et al. Teacher-student curriculum learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(9): 3732-3740.
- [32] FAN Y, TIAN F, QIN T, et al. Learning to teach[C]//The 6th International Conference on Learning Representations (ICLR). Vancouver: ICLR, 2018: 1-16.
- [33] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 761-769.
- [34] JIMÉNEZ-SÁNCHEZ A, MATEUS D, KIRCHHOFF S, et al. Medical-based deep curriculum learning for improved fracture classification[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 694-702.
- [35] ZHAO R C, CHEN X L, CHEN Z L, et al. Diagnosing glaucoma on imbalanced data with self-ensemble dual-curriculum learning[J]. *Medical Image Analysis*, 2022, 75: 102295.
- [36] ZHANG Y, DAVID P, GONG B Q. Curriculum domain adaptation for semantic segmentation of urban scenes[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2020-2030.
- [37] DAI D X, SAKARIDIS C, HECKER S, et al. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding[J]. *International Journal of Computer Vision*, 2020, 128(5): 1182-1204.
- [38] 文载道, 王佳蕊, 王小旭, 等. 解耦表征学习综述[J]. *自动化学报*, 2022, 48(2): 351-374.
- WEN Z D, WANG J R, WANG X X, et al. A review of disentangled representation learning[J]. *Acta Automatica Sinica*, 2022, 48(2): 351-374. (in Chinese)
- [39] 任莎莎, 刘琼. 小目标特征增强图像分割算法[J]. *电子学报*, 2022, 50(8): 1894-1904.
- REN S S, LIU Q. A tiny target feature enhancement algorithm for semantic segmentation[J]. *Acta Electronica Sinica*, 2022, 50(8): 1894-1904. (in Chinese)
- [40] MA J, CHEN J N, NG M, et al. Loss odyssey in medical image segmentation[J]. *Medical Image Analysis*, 2021, 71: 102035.
- [41] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [42] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 3431-3440.
- [43] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [44] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision — ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [45] LAMBERT Z, PETITJEAN C, DUBRAY B, et al. SegTHOR: Segmentation of thoracic organs at risk in CT images[C]//2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). Piscataway: IEEE, 2020: 1-6.

- [46] ISENSEE F, JAEGER P F, KOHL S A A, et al. NNU-Net: A self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2021, 18(2): 203-211.

作者简介



汤红忠 女, 1979年出生于湖南衡山. 现为湘潭大学自动化与电子信息学院教授. 主要研究方向为机器学习与计算机视觉、医学图像处理与智能分析.

E-mail: diandiant@126.com



王 蔚 男, 1999年出生于湖北武汉. 现为湘潭大学自动化与电子信息学院硕士研究生. 主要研究方向为模式识别与人工智能, 医学图像处理研究.

E-mail: 1365287417@qq.com



王 涛 男, 1999年出生于河北唐山. 现为湘潭大学自动化与电子信息学院硕士研究生. 主要研究方向为模式识别与人工智能.

E-mail: 1969622489@qq.com



陆旺达 男, 2000年出生于湖南长沙. 现为湖南大学电气与信息学院硕士研究生. 主要研究方向为计算机视觉与机器学习.

E-mail: 1969622489@qq.com



黄向红 女, 1963年出生于河南郑州. 现为湘潭市中心医院生殖与遗传中心临床主任, 湖南省医学会生殖医学委员会. 主要研究方向为不孕不育、各项辅助生殖技术的应用.

E-mail: hxh@163.com



章 兢 男, 1957年出生于湖南韶山. 现为湖南大学电气与信息学院博士生导师. 主要研究方向为工业自动化、复杂系统计算机控制.

E-mail: zhangj@hnu.cn