

DOI:10.16136/j.joel.2023.05.0304

基于 TW3-C RUS 法的骨龄评估方法研究

谷 静*, 孙启雷, 张明选

(西安邮电大学 电子工程学院, 陕西 西安 710121)

摘要:经典 TW3-C RUS (Tanner and Whitehouse 3-Chinese RUS) 法将手骨的关键骨骺区域严格划分为 9 个等级, 未充分考虑骨骺发育的连续性, 导致骨龄评估存在一定误差。针对该问题, 本文提出一种基于 TW3-C RUS 法的改进骨龄评估方法。采用阈值法的思想, 动态选择网络模型输出的前 N 个等级概率值, 并将前 N 个概率值作为权值计算手骨的加权得分, 降低由于手骨单一等级判定引起的误差。针对网络模型冗余问题, 采用跨阶段局部网络 (cross stage partial network, CSP-Net) 轻量化深度残差网络 (residual network 50, Resnext50)。实验表明, 改进后的方法对男性骨龄评估的平均绝对误差 (mean absolute error, MAE) 为 0.4214 岁, 女性 MAE 为 0.4128 岁, 相比于经典 TW3-C RUS 法, 骨龄评估准确率有明显提升。轻量化后的网络模型参数量为 46.28 MB, 相比 Resnext50 网络模型有明显降低。

关键词:骨龄评估; TW3-C RUS; 卷积神经网络; 模型轻量化

中图分类号: TP391 **文献标识码:** A **文章编号:** 1005-0086(2023)05-0554-07

Research on bone age assessment method based on TW3-C RUS method

GU Jing*, SUN Qilei, ZHANG Mingxuan

(School of Electronic Engineering, Xi'an University of Posts & Telecommunications, Xi'an, Shaanxi, 710121, China)

Abstract: The classic Tanner and Whitehouse 3-Chinese RUS (TW3-C RUS) method strictly divides the key epiphyseal regions of hand bone development into 9 grades, and does not fully consider the continuity of epiphyseal development, leading in certain errors in bone age assessment. To solve this problem, a modified method for assessing bone age based on TW3-C RUS is proposed. The idea of threshold method is adopted to dynamically select the first N level probability values output by the network model, and use the first N probability values as weights to calculate the hand bone weighted score, so as to reduce the error caused by the single level judgment of the hand bone. Aiming at the problem of network model redundancy, the cross stage partial network (CSP-Net) is used to lighten the deep residual network 50 (Resnext50). Experiments show that the mean absolute error (MAE) of the improved method for evaluating bone age is 0.4214 years old for men and 0.4128 years old for women. Compared with the classic TW3-C RUS method, the bone age evaluation efficiency and accuracy are significantly higher. The amount of network parameters after the lightweight is 46.28 MB, which is significantly lower than that of the Resnext50 network model.

Key words: bone age assessment; Tanner and Whitehouse 3-Chinese RUS (TW3-C RUS); convolutional neural network; model lightweight

0 引言

骨龄是人体生长发育过程中骨骼的真实年龄, 是一种生物年龄, 可以真实反映出人体骨骼的

发育状况^[1]。近年来, 骨龄已广泛应用于评估儿童预期成年身高、监测儿童骨骼发育状况等方面^[2]。通常骨龄是借助人体左手的 X 光图像进行人工评估, 专业医生通过观察手骨 X 光图像的骨

* E-mail: guj@xupt.edu.cn

收稿日期: 2022-07-11 修订日期: 2022-07-11

基金项目: 陕西省自然科学基金基础研究计划(2020SF-370)资助项目

化中心结构、大小来判定各骨化中心的发育等级,并统计分析出骨龄^[3]。常用的骨龄评估方法有 G-P (Greulich and Pyle) 图谱法、TW3-C 法、CHN (Chinese) 法等^[4-6]。其中 TW3-C 计分法需要对 13 块骨骺区域分等级,每一块骨骺区域均有 8 个等级,需要医生对大量数据指标进行长时间的分析对比,才能得出骨龄评估结果。采用人工骨龄评估的方法存在受主观因素影响大、骨龄评估耗时较长等问题^[7]。

目前已有基于深度学习的自动骨龄评估方法的研究,文献[8]提出一种基于随机森林算法^[9]的桡骨等级分类器,实现了桡骨区域的自动分割,进一步提高了骨骺区域分割的准确率。文献[10]提出基于亚历克斯·克里热夫斯基网络 (alex krizhevsky network, Alex-Net)^[11]的 Top2 CHN (top2 Chinese) 法骨龄评估算法,该方法利用网络预测的两个最高等级的概率来计算骨头的加权得分,得出最终的骨龄。但是,该方法不能灵活地选择哪些等级概率需要参与计算,对于原本预测正确的骨骺等级会带来一定的误差。文献[12]提出一种基于深度学习的端到端儿童手骨 X 光图像骨龄评估框架,采用掩膜区域卷积神经网络 (mask region-based convolutional neural network, Mask-RCNN) 分割出手骨区域,然后将分割好的手骨区域直接输入改进的极致的开端网络 (extreme inception, Xception) 网络进行骨龄预测,但是该方法未考虑手骨发育连续性的特点。

本文针对以上问题提出一种基于 TW3-C RUS (Tanner and Whitehouse 3-Chinese RUS) 法的改进骨龄评估算法 (topN Tanner and Whitehouse 3-Chinese RUS, TopN TW3-C RUS)。该算法采用阈值法的原理,使网络模型输出满足阈值的前 N 个概率等级,计算手骨加权得分,对发育程度处于相应等级上的骨骺直接输出该等级分值;对于发育程度处于不同等级之间的骨骺区域,使用挑选出的等级概率计算其加权得分。实验证明,该算法可以有效地减小骨龄评估误差。针对网络模型冗余问题,使用 CSP-Net 网络结构轻量化 Resnext50 网络模型,进一步降低网络参数量。

1 改进 TW3-C RUS 的骨龄评估方法

1.1 改进 TW3-C RUS 计分规则

经典 TW3-C RUS 计分法^[13]将手骨的发育过程简化为 8 个等级,并赋予相应的分值。同时将骨龄评估划分为 3 个步骤:手骨划分、等级预测、骨龄映射。首先将手骨图像划分出 13 块骨骺区域(桡骨、尺骨、掌骨 I、掌骨 III、掌骨 V、近节指骨 I、近节指骨

III、中节指骨 I、中节指骨 III、中节指骨 V、远节指骨 I、远节指骨 III、远节指骨 V),然后根据每一块骨骺区域的发育程度,将其归为最相近的一个发育等级,并赋予相应分值,最终将 13 块骨骺区域分值相加获得手骨发育总分,对比骨龄分值映射曲线得到骨龄。

然而骨头发育是一个连续的过程,在进行等级判定时,同一个骨骺区域发育程度可能处于不同的等级之间,而分值表中部分手骨区域等级间分值差异明显,此时若将手骨判定为其中一个等级会产生较大误差。另一方面,网络模型在预测发育程度处于不同等级之间的手骨时,预测手骨处于某一个等级的概率值较低,容易造成等级误判。表 1 为男性桡骨和尺骨的分值表。

表 1 男性桡骨、尺骨分值表
Tab. 1 Male radius and ulna scores

Bone area	1	2	3	4	5	6	7	8
Radius	16	21	30	39	59	87	138	213
Ulna	27	30	32	40	58	107	181	—

表 2 为数据集中处于等级 4 的桡骨区域采用 Resnext50 网络模型预测的结果。其中第一行为桡骨的 8 个等级,第二行为网络输出每个等级对应的概率值。由第二行可以看出等级 5 的概率值最高,因此网络最终将其判定为等级 5,参照表 1,其最终得分为 59,而该桡骨区域的真实等级为 4,得分应为 39,因此带来了 20 分的误差。

针对这种情况采用阈值法的思想提出 TopN TW3-C RUS 法。通过设定等级概率阈值,对于发育程度处于不同等级之间的手骨区域,网络模型预测准确率较低(如表 2 所示),则根据阈值,动态选择前 N 个概率之和满足阈值的等级,将对应等级概率作为权值,计算该手骨区域的加权得分,减小该手骨区域的分值误差。以表 2 数据为例,采用 TopN TW3-C RUS 法,设阈值为 0.61,满足阈值的概率等级为等级 4 和等级 5,通过概率加权计算得分为 49.08,误差为 10.08,相比于原来方法,误差降低了 9.92 分。

TopN TW3-C RUS 计分法流程如图 1 所示。首先将分割出的桡骨区域图片输入 Resnext50 网络^[14]进行训练,输出对应 8 个等级的预测概率值,并通过归一化指数函数 (softmax) 对概率值排序。首先取出最大概率值与阈值比较,若大于阈值,则直接进行下一步的计算加权得分,若小于阈值,则对最大概率值和第二概率值求和并再次与阈值进行比较,循环此步骤即可挑选出满足阈值的所有等级对应的概率值用于计算手骨加权得分。

表 2 某桡骨等级预测概率实例

Tab. 2 Predicted probability results of radius grade

Level	1	2	3	4	5	6	7	8
Probability	0.0001	0.0075	0.0624	0.4152	0.4223	0.0466	0.0397	0.0062

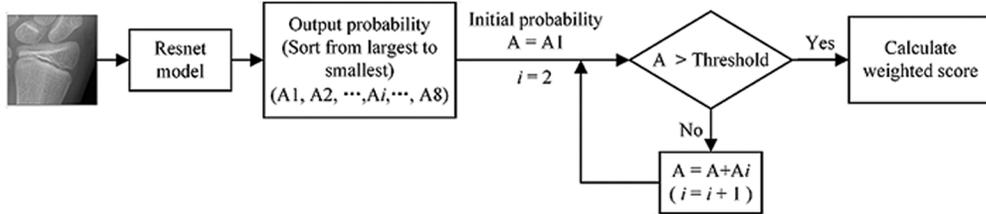


图 1 TopN TW3-C RUS 计分法流程图

Fig. 1 Flow chart of TopN TW3-C RUS scoring method

加权分值计算式如下所示：

$$S = \frac{\sum_{n=1}^N S_n P_n}{\sum_{n=1}^N P_n}, (\sum_{n=1}^N P_n > P), \quad (1)$$

式中, S 表示该骨骺区域最终总分, S_n 表示第 n 个等级对应的分值, P_n 表示第 n 个等级的预测概率, P 表示概率阈值, N 表示满足阈值的概率等级个数。实验结果表明使用该方法可明显降低骨龄预测误差。

1.2 概率阈值计算

TopN TW3-C RUS 法中阈值通过在包含 6—15 岁的男性、女性儿童手骨测试集中统计得出, 选择骨龄平均绝对误差(mean absolute error, MAE) 最低的概率值作为阈值。分别基于 Resnext50 网络模型和下一步优化后的跨阶段局部残差网络(cross stage partial residual network 50, Csp-Resnext50) 进行统计, 实验阈值曲线如图 2、图 3 所示。

图 2 和图 3 阈值曲线中横坐标为阈值分布, 纵坐

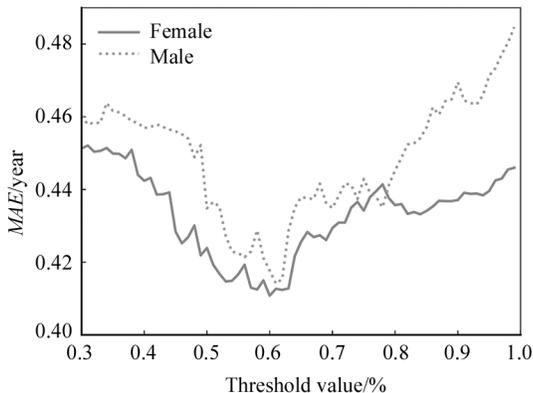


图 2 Resnext50 阈值曲线

Fig. 2 Resnext50 threshold curve

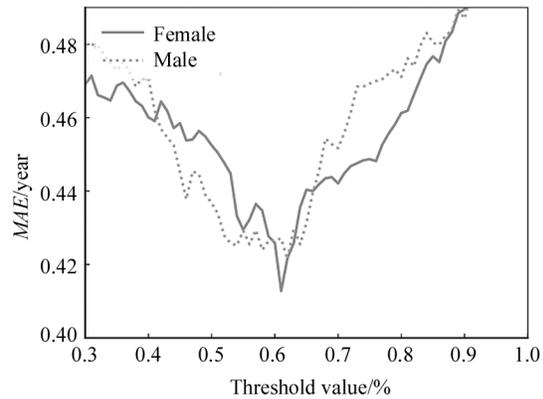


图 3 Csp-Resnext50 阈值曲线

Fig. 3 Csp-Resnext50 threshold curve

标为误差值。对于男性和女性, 两种不同网络的阈值曲线均呈相同变化趋势, 随着阈值的升高, 曲线先下降后上升, 且均在 0.61 附近降到最低, 因此采用本实验方法统计阈值具有较高可行性。

2 优化网络模型

通过实验测试, 由于 TopN TW3-C RUS 法相较于原始方法增加了一定的计算量, 使得网络模型计算量有所增加。因此针对这一问题, 采用 CSP-Net 网络结构^[15] 优化 Resnext50 网络模型。原始 Resnext50 网络模型如图 4 所示。

Resnext50 网络模型包含 4 个残差层和一个全连接层, 输入为固定尺寸和标签的骨骺区域图片, 输出为该骨骺区域对应的概率。输入图片为 $224 \times 224 \times 3$ 的彩色(RGB)图像, 经由网络的 7×7 卷积层和最大池化层(maximum pooling, MaxPool) 转化为 $56 \times 56 \times 64$ 的特征图, 而后经过批归一化层(batch-no

rmalization, BN) 和 Leaky_ReLU 激活函数。此后图像经过 4 个残差层进行特征提取, 然后全连接层 (fully connected layers, FC) 将特征图展平为 $1 \times 1 \times 2048$ 的特征图并通过 softmax 输出为各骨骺区域的等级概率。

优化 Resnext50 网络模型的核心思想是将 Resnext50 网络的 4 个残差层的输入特征图平均分

为两路, 一路进入残差层进行特征提取, 另一路通过平均池化层 (average pooling, AvgPool) 和 1×1 卷积核进行下采样, 与残差层的输出进行拼接并使用固定参数线性修正单元函数 (leaky rectified linear unit, Leaky-ReLU) 进行激活。这样既可以保留原始特征图部分特征, 又降低了残差层的计算量。优化后的 Csp-Resnext50 网络结构如图 5 所示。

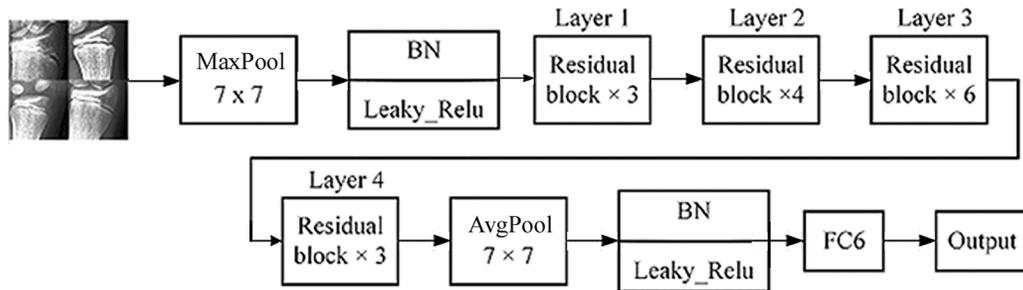


图 4 Resnext50 网络模型结构图

Fig. 4 Diagram of Resnext50 network model structure

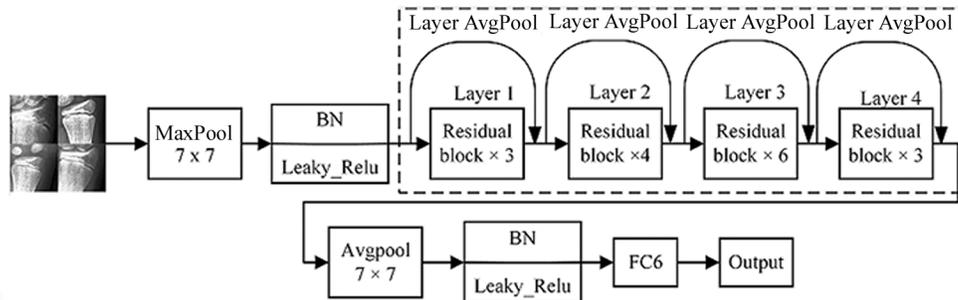


图 5 Csp-Resnext50 网络模型结构

Fig. 5 Csp-Resnext50 network model structure

3 实验及结果分析

3.1 实验环境

本文实验的模型训练基于 Linux 系统的服务器完成, 配备有 4 块 Intel (R) Xeon (R) Gold 5120 CPU 和 4 块 TITAN RTX 24 GB 的 GPU。网络模型基于 Pytorch 框架编写, 初始学习率设为 0.0001, 批次大小为 16, 男性和女性各训练 5000 个批次 (Epoch)。

3.2 数据集收集

本文采用的数据集由国内某儿童医院提供, 均为自身体状况良好的 6—15 岁儿童左手骨 X 光图像, 数据集共有图像 4255 张, 其中男性 2315 张, 女性 1940 张。训练数据集与测试数据集以 9:1 的比例随机划分。

由于数据集中部分手骨图像较模糊, 特征不够

明显, 不利于网络模型训练。因此, 基于跨平台的计算机视觉库 (open source computer vision library, OpenCV), 采用对其直方图均衡化、边缘特征提取等图像增强方法对数据集进行图像预处理, 增强手骨边缘特征和纹理特征, 便于训练网络模型。以桡骨为例, 直方图均衡化处理前后如图 6 所示。



图 6 图像增强结果对比: (a) 原图; (b) 结果图
Fig. 6 Comparison of image enhancement results: (a) Original image; (b) Result image

图6中使用直方图均衡化方法处理后的图片与原图相比,其目标骨骼区域的对比度更高、特征更加明显,便于神经网络对特征的提取。

3.3 评价指标

本实验采用的评价指标为 MAE。将测试集按照手骨图片的年龄段、性别标签进行分类,通过 MAE 计算式计算出预测骨龄和真实年龄之间的误差,并统计出±1岁以内的骨龄预测准确率。MAE 计算式如下所示:

$$MAE = \frac{1}{K} \sum_{i=1}^K |(F_i - Y_i)|, \quad (2)$$

式中,MAE 表示最终计算的误差, i 表示每张手骨图片, K 表示测试集中手骨图片总数, F_i 表示测试集中每个手骨图片的预测年龄, Y_i 表示每个手骨图片的真实年龄。

3.4 实验结果和分析

为了验证统计阈值的方法对 CSP-Resnext50 网络的适用性,采用改进前后的网络进行方法验证。验证结果如表3所示。该方法对 Resnext50 网络和 Csp-Resnext50 网络得出的最优阈值差异为 0.01,在相同网络的男性、女性之间最优阈值差异也仅为 0.01,实验结果表明,该方法对改进前后的网络以及性别差异具备一定泛化性,证明本方法适用于改进后的网络。

表3 阈值统计结果

Tab. 3 Threshold statistical results

Network	Male MAE /year	Female MAE /year
Resnext50	0.61	0.60
Csp-Resnext50	0.62	0.61

本文中男性数据集和女性数据集均在相同网络模型和实验环境下进行训练。实验分别采用经典 TW3-C RUS 法、Top2 TW3-C RUS 法(利用网络输出的固定前 2 个等级概率进行计算手骨加权得分)以及 TopN TW3-C RUS 法(以下简称 3 种方法),并统计出 3 种方法对男性、女性的 MAE 和±1岁以内的准确率。

表4对比了3种方法分别在男性和女性测试集上的 MAE。由表4可知,改进的 TopN TW3-C RUS 法相比经典 TW3-C RUS 法和同类型的 Top2 TW3-C RUS 法的 MAE 更低,其中男性 MAE 降低至 0.4141 岁,女性 MAE 降低至 0.4108 岁。实验结果

表明,本文方法在骨龄评估中的 MAE 有明显降低,在医学骨龄评估领域有一定应用价值。

表4 3种方法的 MAE 对比

Tab. 4 MAE comparison of three methods

Method	Male MAE /year	Female MAE /year
TW3-C RUS	0.4630	0.4557
Top2 TW3-C RUS	0.4409	0.4343
TopN TW3-C RUS	0.4141	0.4108

图7对比了3种方法的±1岁准确率,其中横坐标为男性、女性,分别对应3种不同方法,纵坐标为±1岁以内的准确率。由图中可以看出,改进后的 TopN TW3-C RUS 法在男性和女性测试集上的准确率均比其他两种方法更优。相比于未改进前的 TW3-C RUS 法,改进后的方法男性±1岁以内的准确率提高 5.18%,女性±1岁以内的准确率提高 5.46%,实验结果表明改进后的方法具有更高的准确率。

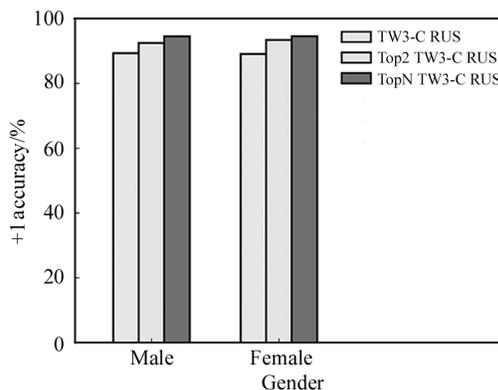


图7 3种方法的±1岁准确率对比

Fig. 7 Comparison of the ±1-year-old accuracy of the three methods

表5展示了改进前后的网络模型参数量对比以及男性、女性的 MAE 对比。由表5可知,基于改进后的 TopN TW3-C RUS 法,使用 Csp-Resnext50 网络模型相比于原始 Resnext50 网络模型,在保证骨龄评估精度的前提下,网络大小由原来的 88.07 MB 降低到 46.28 MB,其网络参数量大幅下降。实验结果表明采用 Csp-Resnext50 网络模型可以在保证相同误差的前提下明显降低网络的计算量。

图8对比了改进前后网络模型的±1岁准确率,其中横坐标为男性、女性,分别对应 Resnext50 网络和 CSP-Resnext50 网络,纵坐标为±1岁以内的准确率。由图中可以看出,Resnext50 网络和 CSP-

Resnext50网络的±1岁以内准确率几乎相同,说明对Resnext50网络轻量化之后不会对准确率造成较大影响,证明了改进后的网络的可行性。

表5 不同网络的MAE对比
Tab.5 MAE comparison of different networks

Network	Network size	Male MAE /year	Female MAE /year
Resnext50	88.07 MB	0.4141	0.4108
Csp-Resnext50	46.28 MB	0.4214	0.4128

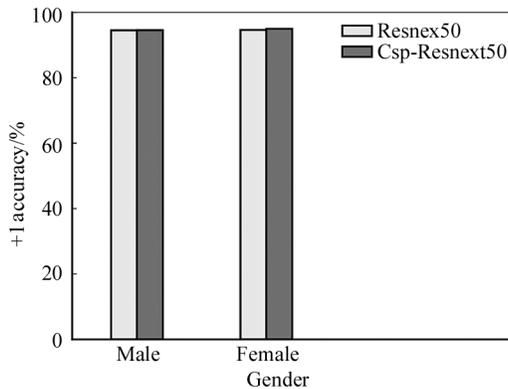


图8 不同网络的±1岁准确率对比

Fig.8 Comparison of ±1-year-old accuracy of different networks

4 结论

本文在TW3-C RUS法的基础上,结合儿童手骨发育特征和网络模型输出的特点,提出TopN TW3-C RUS法,并采用CSP-Net网络结构轻量化骨龄评估网络模型。实验结果表明,该方法充分利用骨骺发育连续性的特点,能够准确地评估处于不同等级之间的手骨区域,同时在保持骨龄评估精度的情况下,大幅降低了评估网络的计算量。相比于经典TW3-C RUS法和固定输出前两个等级的Top2 TW3-C RUS法,均达到了更好的结果,可以作为辅助医生评估骨龄的参考依据。

参考文献:

[1] WANG Y P, CAO Y P, FU G K, et al. Classification of microscopic images of human peripheral blood leukocytes based on deep convolutional neural networks[J]. Journal of Optoelectronics • Laser, 2019, 30(5): 546-555.
王亚品, 曹益平, 付光凯, 等. 基于深度卷积神经网络的人体外周血白细胞显微图像分类[J]. 光电子·激光, 2019, 30(5): 546-555.

[2] YAN R. Application analysis of bone age test in children's

growth and development[J]. China Rural Health, 2021, 13(10): 36-38.

闫荣. 骨龄测试在儿童生长发育中的应用分析[J]. 中国农村卫生, 2021, 13(10): 36-38.

[3] WU Q S. Research on bone age recognition method of wrist based on deep learning [D]. Chengdu: Sichuan University, 2021: 1-77.

吴青松. 基于深度学习的手腕部骨龄识别方法研究[D]. 成都: 四川大学, 2021: 1-77.

[4] TANG W, WU G, SHEN G. Improved automatic radiographic bone age prediction with deep transfer learning [C]//2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, October 19-21, 2019, Suzhou, China. New York: IEEE, 2019: 1-6.

[5] LIU X H, WANG G Y, XU Y M, et al. Bone age assessment by deep convolutional neural networks combined with clinical TW3-RUS [C]//2019 IEEE International Conference on Bioinformatics and Biomedicine, November 18-21, 2019, San Diego, CA, USA. New York: IEEE, 2019: 948-952.

[6] ATSUMI M, KAWANO S, MORIOKA T, et al. Deep learning based ancient asian character recognition [C]//2020 International Conference on Advanced Mechatronic Systems, December 10-13, 2020, Hanoi, Vietnam. New York: IEEE, 2020: 296-301.

[7] ZHANG S J, LI R, ZHAN M J, et al. Research on bone age evaluation based on deep learning regional fusion [J]. Modern Computer, 2020, 37(9): 54-59.

张世杰, 李睿, 占梦军, 等. 基于深度学习区域融合的骨龄评价研究[J]. 现代计算机, 2020, 37(9): 54-59.

[8] LIU J L. Computer-aided assessment of radial bone age [D]. Beijing: Beijing Jiaotong University, 2017: 1-64.

刘洁琳. 计算机辅助桡骨骨龄等级评估[D]. 北京: 北京交通大学, 2017: 1-64.

[9] VANSTEENKISTE T, RUYSSINCK J, JANSSENS O, et al. Automated assessment of bone age using deep learning and Gaussian process regression [C]//2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 18-21, 2018, Honolulu, HI, USA. New York: IEEE, 2018: 674-677.

[10] CHI K K, CAI R H, DING W L, et al. Intelligent bone age assessment based on deep learning [J]. Computer Science, 2019, 46(S2): 89-93.

池凯凯,蔡荣辉,丁维龙,等.基于深度学习的智能骨龄评估[J].计算机科学,2019,46(S2):89-93.

- [11] PALANISWAMY T. Deep learning empowered automatic bone age assessment[C]//2021 Third International Conference on Inventive Research in Computing Applications, September 02-04, 2021, Coimbatore, India. New York:IEEE,2021:1821-1827.
- [12] ZHANG S, ZHANG J H. Bone age assessment method based on deep learning in children's hand bone X-ray images[J]. Aerospace Medicine and Medical Engineering, 2021,34(3):252-259.
- 张帅,张俊华.基于深度学习的儿童手骨 X 光图像骨龄评估方法[J].航天医学与医学工程,2021,34(3):252-259.
- [13] LIU H Q, ZHANG Y. Leukocyte segmentation based on superpixel and graph convolutional neural network[J]. Jour-

nal of Optoelectronics • Laser, 2021,32(10):1074-1082.

刘汉强,张元.基于超像素与图卷积神经网络的白细胞分割[J].光电子·激光,2021,32(10):1074-1082.

- [14] MIAO J Y, YUE H S, WU X M, et al. Bone age assessment based on super point features[C]//2020 Chinese Automation Congress, November 6-8, 2020, Shanghai, China. New York:IEEE,2020:3921-3926.
- [15] ZHANG G S, ZHAO Y, MA H Y. Gesture recognition based on convolutional neural network [J]. Journal of Optoelectronics • Laser, 2019,30(12):1317-1322.
- 张国山,赵阳,马红悦.基于卷积神经网络的手势识别[J].光电子·激光,2019,30(12):1317-1322.

作者简介:

谷 静 (1975—),女,硕士,副教授,硕士生导师,主要从事通信与信息系统、图像处理方面的研究.