

优化参考图谱发育指征的CHN智能骨龄评估方法研究

毛科技^{*①} 武坤秀^① 陆伟^① 陈立建^{①②} 毛家发^①

^①(浙江工业大学计算机科学与技术学院 杭州 310032)

^②(浙江广播电视大学萧山学院 杭州 311200)

摘要: 骨龄(BA)是评估儿童生长发育是否正常的重要指标之一。中国人手腕骨发育标准-CHN计分法是目前中国儿童生长发育中骨龄评估(BAA)最常用的方法之一。但是在CHN计分法中,某些参照骨图谱的发育指征跨度较大,导致专家依据个人经验主观判断它的发育分期而影响评估准确度。在利用深度学习对该类图谱的发育分期进行评估时,会导致它的评估结果产生随机性。该文基于专家评估过的2万余张儿童手腕部X线片,在CHN计分法的基础上,在相邻发育分期间隔跨度较大的参照骨标准图谱之间勾绘新的成熟度指征,产生细化图谱,并利用层次分析法为其分配对应的成熟度得分,提高骨龄评价的准确率。该文在AlexNet网络的基础上融合Harris特征和卷积注意力模块,对各参照骨的发育分期进行评估。在自制的年龄分布为5-11岁的数据集上,采用优化后的CHN法得到的骨龄在容忍度为0.5岁和1岁时的准确率分别达到了94.6%和99.13%。实验结果表明所提方法可以更加精细地分辨儿童手腕骨发育程度,大幅提高骨龄评估的准确率,辅助临床应用。

关键词: 骨龄智能评估; CHN计分法; 卷积神经网络; 注意力机制; 层次分析法

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2023)03-0958-10

DOI: [10.11999/JEIT211577](https://doi.org/10.11999/JEIT211577)

A Study of the CHN Intelligent Bone Age Assessment Method with Reference to Atlas Developmental Indications

MAO Keji^① WU Kunxiu^① LU Wei^① CHEN Lijian^{①②} MAO Jiafa^①

^①(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310032, China)

^②(College of Xiaoshan, Zhejiang Radio Broadcasting and Television University, Hangzhou 311200, China)

Abstract: Bone Age (BA) is one of the most important indicators in evaluating children's growth. The Bone Age Assessment (BAA) based on Chinese wrist bone development standard-CHN (CHN) scoring method is widely used in the evaluation of children's growth and development and height prediction. However, the adjacent developmental levels of some reference bones last longer, leading to the subjective judgment of developmental levels by experts based on personal experience, which affects the accuracy of predictions. When deep learning is used to evaluate the developmental levels of these atlases, the prediction results will be random. In this paper, based on more than 20000 X-ray images evaluated by experts, a new mature indicator with a large interval with a large interval is drawn to generate exquisite atlas to perform some reference bones. Additionally, the corresponding maturity score is determined by analyzing the level structure process to maximize the impact of error -level prediction on BAA. Combining Harris features and convolutional blocks of the convolutional neural network of the attention module is designed to evaluate automatically the level of bone maturity. In addition, an annotated database with an age distribution of 5-11 years is built to train and evaluate the method. The accuracy of predictions obtained by adding a new standard atlas to the CHN method reaches 94.6% and 99.13% when the tolerance is 0.5 years and 1 year, respectively. The experimental results show that the method proposed in this paper can distinguish the development degree of reference bones more precisely, and improve greatly the accuracy of BAA, proving the potential for practical clinical application.

收稿日期: 2021-12-27; 改回日期: 2022-07-26; 网络出版: 2022-08-04

*通信作者: 毛科技 maokeji@zjut.edu.cn

基金项目: 国家自然科学基金(62072410), 浙江省基础公益研究计划(LGG22F020014)

Foundation Items: The National Natural Science Foundation of China(62072410), The Basic Public Welfare Research Project of Zhejiang Province (LGG22F020014)

Key words: Bone age intelligence assessment; Chinese wrist bone development standard (CHN); Convolutional neural network; Attention mechanism; Analytic hierarchy process

1 引言

骨龄是评价儿童青少年发育程度的重要依据^[1], 可以协助诊断与治疗内分泌相关疾病。除此之外, 骨龄的主要作用是预测儿童青少年的成年身高^[2], 成年身高预测有助于了解儿童身高增长潜力^[3]。由于手腕部最能代表骨骼发育程度与生长潜力, 所以手腕部在骨龄测评中最为常用^[4]。目前, 世界上最常用的骨龄评估方法^[5]分别是Greulich-Pyle图谱法(GP法)^[6]和Tanner Whitehouse计分法(TW法)^[7]。GP图谱法将待测X线片与标准图谱逐个比较, 取最相似的为其骨龄, 这种方法简便、直观, 但主观性强, 偏差大。TW计分法中最广泛使用的是TW3法^[8]。由于TW系列骨龄评估方法更适用于欧洲人, 所以我国针对中国儿童青少年提出中华05法^[9]和中国人手腕骨发育标准-CHN计分法^[10]。中华05法中的RUS-CHN仍然沿用白种人样本计算的权重分配方案和TW2法手腕骨发育分期得分表。标准制定人称采用中华05法对3-18岁儿童青少年进行评测的结果比采用CHN法评测的结果在男性中偏小0.3-1.1岁, 女性中偏小0.2-1岁。而且CHN法是最早应用于运动员选材领域的骨龄标准, 相比于其他骨龄评估方法在身高预测方面更有优势, 根据CHN法骨龄预测身高是目前预测我国儿童身高最准确的方法之一。

随着医学图像处理和智慧医疗的发展, 越来越多的学者对骨龄的智能评估进行研究。2009年, Thodberg等人^[11]提出了BoneXpert骨龄评估系统, 在欧洲等地开始投入商业使用。该系统基于形状驱动的主动外观模型, 使用3000多张图像对应的参照骨形态、骨密度、纹理等特征训练建模, 进行骨龄评估。但是该系统要求输入的X线片必须清晰完整, 系统的泛化能力较低。Lee等人^[12]运用迁移学习^[13]的方法开发了一个基于深度学习的骨龄智能评估系统。该系统对GoogleNet网络进行微调, 实验结果表明男性和女性的评估结果与专家结果的误差在1岁范围内的准确率分别为94.18%和90.39%, 均方根误差分别为0.82岁和0.93岁。2019年, Ren等人^[14]先用注意力模块对图像进行预处理, 将生成的注意力图作为回归网络的输入, 通过注意力损失监督学习训练骨龄评估模型, 评估的骨龄与专家结果的平均绝对误差达到了5.2~5.3个月。Han等人^[15]利用高斯金字塔算法对图像缩放, 并使用中值滤波器对其预处理, 然后采用形态学分水岭算法对图像

的感兴趣区域和背景区域进行分割, 利用resnet网络提取图像特征。在自制的数据集集中模型效果与专家结果的平均绝对误差为0.45岁。

此外, 不同的学者采用不同的骨龄评价标准对骨龄智能评估进行研究。Liu等人^[16]基于GP图谱法, 采用VGG-U-Net将X线片的手腕骨部分切割出来, 然后构造对抗网络对骨龄进行智能评估。Son等人^[17]采用TW3骨龄标准, 训练了一个神经网络对X线片的13个ROI区域进行提取并对其发育成熟等级进行分类从而预测骨龄。刘宗才等人^[18]依据中华05骨龄标准, 利用深度学习模型和专家对照组阅片来探讨贵州省儿童青少年骨龄智能评估的准确性。

上述方法均是通过改变X线片的预处理方式, 改进已有的图像分类算法并应用到骨龄评估领域, 来提高智能骨龄评估的准确率。这些方法在分析模型性能时所用的指标为平均绝对误差或容忍度为1岁时骨龄评估的准确率。但在实际临床应用中, 平均绝对误差只能代表模型的平均效果, 并不能保证骨龄评估值的可靠性, 且当评估值与标准值的误差为1岁时, 骨龄评估结果并不能辅助临床应用。

CHN法将整张手腕部X线片中的14块参照骨的发育分期等级分为7级或8级, 并为每个等级分配得分。之后计算全部参照骨的发育分期得分总和, 并将其转化为对应的骨龄。由于CHN法中某些参照骨图谱的发育指征跨度较大, 专家依据个人经验主观判断它的发育分期而影响评估准确度。而利用深度学习对该类参照骨的发育分期进行评估时, 会导致该类参照骨的评估结果产生随机性。如果这两个相邻分期对应的得分相差较大, 那么会使最终的骨龄评价结果产生较大误差。

本文针对这一特定问题提出了一个优化参考图谱发育指征的方法, 在CHN法的基础上, 在相邻等级跨度较大的参照骨图谱之间勾绘新的成熟度指征, 产生细化图谱, 使发育指征处于相邻发育等级之间的参照骨有一个确切的成熟度等级, 并利用层次分析法对新选择的成熟度指征分配得分, 提高参照骨发育分期的分辨能力和骨龄评估的准确率。本文提出的骨龄智能评估方法的结构图如图1所示, 分为3个阶段。第1阶段中利用Faster-RCNN技术将X线片中的14个ROI区域进行定位并分割; 第2阶段中设计一个卷积神经网络预测参照骨的发育分期等级(包括本文新选择的发育分期); 在第3阶段, 计算14块参照骨对应的成熟度得分总和, 并根据性别

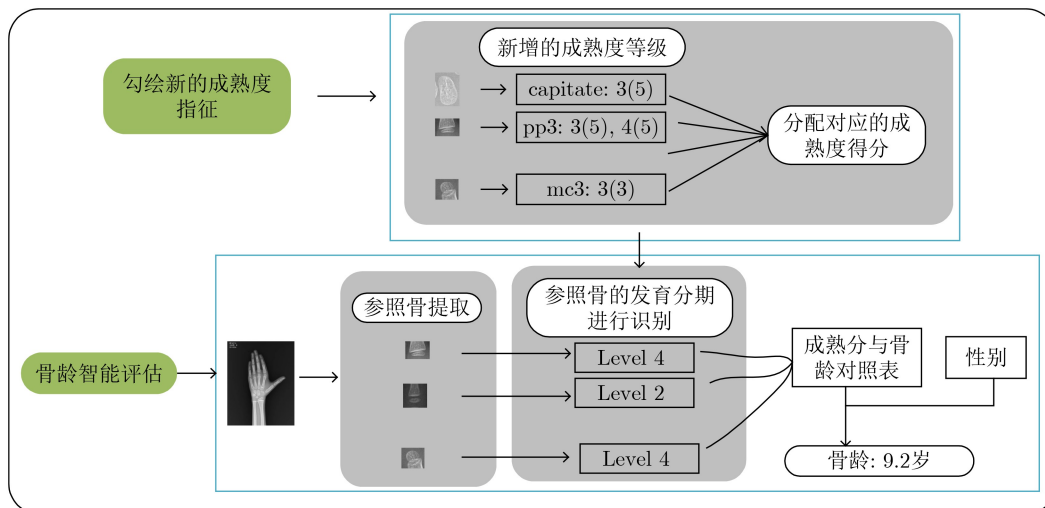


图1 基于CHN法的骨龄智能评估系统结构图

信息在CHN法成熟分与骨龄对照表中得到对应的骨龄。此外，本文采用骨龄预测值与金标准值的误差在0.5岁范围内的准确率作为评价指标。

2 数据分析

2.1 数据获取

由于小学阶段是生长发育的重要阶段，为了对该阶段的骨龄评估进行有效研究，需要有充足的数据。本项目在浙江省“校园公益行”项目的支持下曾多次从浙江省某市内多个小学采集到10万余张手腕部X线片，年龄分布在5-11岁。所有X线片均由同一台DR设备拍摄，为了方便对图像的应用与研究，将拍摄到的DICOM格式图像转换为大小为 1626×2032 的JPG格式图像，并邀请多位骨龄专家依据CHN法对其中的2万余张X线片中的各个参照骨进行发育分期评定，之后计算得到每张X线片对应的骨龄。

在利用CHN法对骨龄进行评价时需要14块参照骨参与评估，分别为桡骨(radius)，掌骨I (metacarpal I, mc1)，掌骨III (metacarpal III, mc3)，掌骨V (metacarpal V, mc5)，近节指骨I (proximal phalanges I, pp1)，近节指骨III (proximal phalanges III, pp3)，近节指骨V (proximal phalanges V, pp5)，中节指骨III (intermediate phalanges III, mp3)，中节指骨V (intermediate phalanges V, mp5)，远节指骨I (distal phalanges I, dp1)，远节指骨III (distal phalanges III, dp3)，远节指骨V (distal phalanges V, dp5)，头状骨(capitate)，钩骨(hamate)。为了减少一些无关信息对骨龄评估造成的影响，提高骨龄识别精度，需要将这些ROI区域提取出来。

本文使用Faster-RCNN^[19]算法对X线片的

ROI区域进行切割。先将图片输入VGG16网络提取特征生成特征图，然后输入区域建议网络得到每个锚点的目标性得分和边界框位置，保留一定数量的比较精确的ROI区域，将这些区域候选框投影到特征图上生成特征矩阵，之后通过池化层缩放到统一的大小，然后通过一系列全连接层得到该ROI区域的预测概率和边界框位置。我们把专家评估过的X线片输入模型中，将提取到的各类参照骨按照发育分期分类，便于后续对骨龄智能评估的分析与研究。

2.2 骨发育的连续性和评价方法的阶段性

骨头在发育过程中的形态变化是连续的，而计分法将骨头的发育过程划分为若干个阶段，每个阶段为一个发育分期。当采用计分法判定参照骨发育分期时，尽管参照骨在持续发育，参照骨的分期却没有发生变化。以近节指骨V为例，在发育分期为3期时，其对应的发育指征为骨骺最大横径为干骺端的 $1/2$ 及以上，如图2(a)所示。在发育分期为4期时，其对应的发育指征为骨骺近侧缘凹、明显致密，如图2(b)所示。但在判定参照骨发育分期过程中，有部分参照骨的发育指征已经超过前1等级但还没有到达后1等级，介于相邻两个参考图谱的发育指征之间，此时等级评价较为困难。例如图2(c)所示的参照骨的发育程度比图2(a)所示的3期参照骨标准图谱的发育程度更加成熟，但依旧没有达到图2(b)中4期标准图谱的发育程度，其对应的发育指征与3期和4期的参考图谱均不一致，导致专家依据个人经验主观判断它的发育分期而影响评估准确度。

2.3 评估结果的不一致性

本文对数据集中同时经过多位专家评估过的参照骨图谱进行分析。以近节指骨V为例，由于本文中用到的X线片的年龄在5-11岁，这期间近节指骨

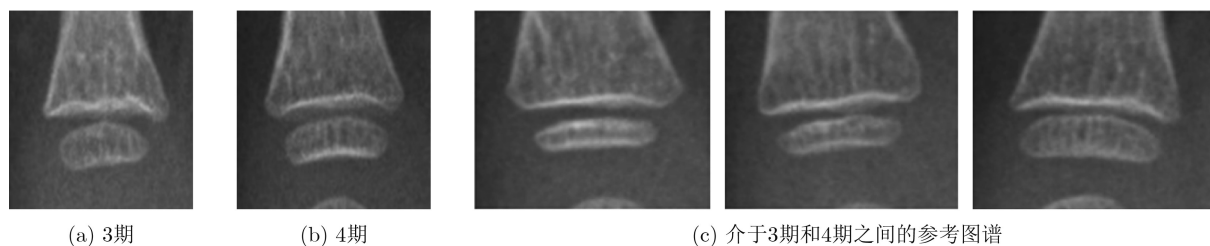


图2 近节指骨V介于3期与4期之间的图谱

V的发育分期主要集中在3期、4期和5期之间。其中多位专家评估全部一致为3期的图谱有431张，全部一致为4期的图谱有830张，而有1276张图谱的发育分期意见由于评估专家的不同而不一致，对于同一张图谱，部分专家评估为3期，部分专家评估为4期。对于这些评估意见不一致的图谱，更多专家更认可其为3期的有699张，占比54.78%，更多专家认可其为4期的有577张，占比45.21%。

表1描述了同时经过多位专家评估过的近节指骨V按照多位专家的评估意见是否全部一致的数量分布。在评估结果为3期的参照骨图谱中，多位专家评估意见全部一致的图像占38.14%，不全部一致的图像占61.86%。在评估结果为4期的图谱中，意见全部一致的占41%，不全部一致的占59%。可以看出每个发育分期的参照骨图谱中都有一部分图像由于专家的个人经验和主观判断而使它的发育分期评估准确度受到影响。而且这些专家评估意见不一致的参照骨图谱对应的成熟度指征往往介于3期和4期之间，与3期和4期标准图谱的骨发育程度均不一致，无法更加精细地分辨骨发育成熟度，此时专家会根据个人知识和经验来确定该参照骨的发育等级，会导致不同的专家评估的结果有所不同。在利用深度学习对该类图谱的发育分期进行评估时，会导致它的评估结果产生随机性。

表2描述了男生、女生的近节指骨V的每个发育分期所对应的成熟度得分。如表2所示，当一张男生的X线片的其他参照骨发育分期全部一致，只有近节指骨V的发育分期评估结果不一致，分别为3期和4期时，其对应的成熟度得分会相差11，那么对应的骨龄会相差0.3岁，可以看出即使一块参照骨图谱的发育分期在相邻等级中发生误差也会对最终的骨龄评价造成影响。

2.4 深度学习算法的影响

以本文第4节提到的算法作为评估模型，本文在经过多位专家评估过的近节指骨V图谱中，选择2560张骨发育程度在3期到4期之间的图谱作为测试集来预测图谱的发育分期。其中实验结果与专家评估结果产生误差的图像有308张。而在这些产生误

表1 近节指骨V各发育分期等级数量分布表

发育分期等级	多位专家意见是否全部一致	图片数量	在该分期图像中的占比(%)
3	一致	431	38.14
	不一致	699	61.86
4	一致	830	59
	不一致	577	41

表2 CHN法近节指骨V不同发育分期得分表

	1	2	3	4	5	6	7	8
男	19	24	34	45	57	64	67	71
女	21	29	43	53	65	73	77	80

差的图像中，多位专家对每块参照骨的发育分期评估意见不全部一致的图像占比为100%。由此可以看出在利用深度学习对参照骨进行发育分期评估时，预测结果与专家意见产生误差都是由于其发育指征介于相邻两个发育分期之间。即利用深度学习评估参照骨发育分期时，如果某块参照骨的预测结果产生误差，那么该参照骨在不同专家评估时也会出现不一致的结果。分析可得，导致自动骨龄评估模型与专家评估结果产生误差的主要原因是某块参照骨的发育程度介于相邻发育等级之间，使模型评估时无法精细地判别骨发育程度。

3 优化参考图谱

根据第2节所述，本文按照参照骨的发育规律，在不改变该阶段开始和结束时对应的发育指征及其得分的条件下，在相邻等级间隔跨度较大的参照骨图谱之间选择新的成熟度指征，产生一个细化图谱。如果在相邻等级之间新增多个标准图谱，新增的图谱会与相邻图谱的发育指征相似度极高，从而增加评估复杂度。增加一个新的标准图谱，既可以使介于相邻等级中间的图谱有一个确切的成熟度指征，提高骨发育成熟度的分辨能力，又不使相邻发育分期的发育指征极其相似而降低发育分期的分辨能力。选择的新的成熟度指征由介于相邻发育分期之间的参照骨图谱所处的发育状态来决定，改进后的CHN法在下文中称为Re_CHN法。

图3(a)和图3(c)分别为近节指骨V的3期和4期对应的标准图谱,而介于3期和4期之间的参照骨图谱的发育指征基本为骨骺近侧缘有凹起的趋势、开始出现致密白线。所以将其选择为新的成熟度指征,增加一个新的标准图谱,如图3(b)所示。在标示新选择的发育分期时仍然以原CHN等级为编码,但在其后增加后缀来与原CHN等级相区别。括号里的数字为新增的标准图谱与相邻两等级中前1等级对应的标准图谱之间的距离 $dis \times 10$,其中 dis 的计算如式(3)所示。例如图3(b)中近节指骨V新增的标准图谱的 dis 为0.5,所以近节指骨V新增的标准图谱对应的发育分期标示为3(5)期。

除此之外,还需要确定新增的标准图谱对应的成熟度得分,其计算公式为

$$s = s_1 + (s_2 - s_1) \times (1 - dis) \quad (1)$$

即计算该得分时最需要考虑的是 dis 的取值。本文在计算 dis 时需要考虑3种因素对其造成的影响,分别为专家对该类参照骨的评估意见,采用深度学习对其进行预测的评估结果,以及该类图像与上1等级图像的相似度,在下文中分别用 W_1 , W_2 , W_3 表示这3种因素在计算 dis 时所占的权重。本文利用层次分析法通过以下几个步骤来确定 W_1 , W_2 , W_3 的取值:

(1) 填写准则层判断矩阵,将3个因素两两进行比较,其中 a_{ij} 表示要素 i 与要素 j 相对重要度之比。其中 $a_{ij}=1, 3, 5$ 时分别表示要素 i 相比要素 j 的重要性相同、略重要、重要。当 $a_{ij}=2, 4$ 时是以上相邻判断的中值。由于 W_1 相比 W_2 更重要一些,所以 a_{12} 为2。又因为该判断矩阵是一个成对判断矩阵,所以 a_{21} 为1/2。同理, a_{13} 为3, a_{31} 为1/3, a_{23} 为1, a_{32} 为1。

(2) 计算该矩阵的特征向量,得到的向量即为各因素的权重,分别为54.85%, 24.09%, 21.06%。

(3) 对矩阵进行一致性检验。如果一致性比率 $CR < 0.1$,则认为判断矩阵的一致性可以接受;否则需要对判断矩阵进行修改, CR 为 CI 与 RI 的比值,其中 CI 为一致性指标,计算公式如式(2)所示。

该矩阵的特征根 λ_{max} 为3.01829, n 为矩阵阶数,所以 CI 为0.009。 RI 为平均随机一致性指标,是一个常量,根据平均随机一致性指标表查询得到,取值为0.52,所以一致性指标 CR 为0.017。因为 $0.017 < 0.1$,所以本文的判断矩阵通过了一致性检验标准。因此, W_1 , W_2 , W_3 的值分别为54.85%, 24.09%, 21.06%。

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (2)$$

$$dis = W_1 \times E_1 + W_2 \times A_1 + W_3 \times F_1 \quad (3)$$

dis 的计算公式如式(3)所示,其中 E_1 表示在发育分期为新选择成熟度等级的参照骨图谱中,按照最初CHN法被专家评估为前1等级的概率。例如发育分期为3(5)期的近节指骨V图谱中,按照最初CHN标准被专家评估为3期的图像占比为54.78%,所以 E_1 为54.78%。 A_1 表示利用深度学习预测该等级的图谱的发育分期时将其判定为前1等级的概率。例如发育分期为3(5)期的近节指骨V图谱中,被模型评估为3期的图像占比42.48%,所以 A_1 为42.48%。 F_1 表示在发育分期为新选择成熟度等级的图谱中,与相邻的上1等级标准图谱之间的图像相似度更高的图像所占的比例。以近节指骨V为例,发育分期为3(5)期的图谱中,与3级标准图谱之间基于直方图的相似度更高的图谱占比58.46%,所以 F_1 为58.46%。此外,式(1)中 s 为待求的新增的成熟度等级对应的得分, s_1 和 s_2 分别表示新增成熟度等级的相邻等级对应的得分。以男生近节指骨V为例, s_1 为3期对应的得分34, s_2 为4期对应的得分45,那么近节指骨V新增的成熟度等级3(5)期的对应的成熟度得分为40。

从表3可以看出,增加一个新的标准图谱可以更加精细地分辨骨发育成熟度,提高骨龄评价标准对儿童骨发育成熟程度的分辨能力,使发育指征介于相邻发育等级之间的参照骨图谱有一个确切、精细的成熟度等级,同时减小发育分期评估产生误差时对骨龄评价造成的影响。例如男生的近节指骨V在原CHN标准中3期和4期对应得分相差11,对



(a) 3期

(b) 3(5)期

(c) 4期

图3 近节指骨V的3期、3(5)期和4期对应的标准图谱

在通道注意力模块中,将上一层的输出 $F^{C \times H \times W}$ 作为特征图分别通过最大池化层和平均池化层得到 F^{GAP} 和 F^{GMP} 两个 $1 \times 1 \times C$ 的特征图。由于SENet^[26]已经充分证明了避免降维和适当地跨通道交互可以有效地提高通道注意力学习的性能,所以本文将得到的 F^{GAP} 和 F^{GMP} 分别通过一个大小为 K 的1维卷积,在不降维的情况下学习每个通道及其周围 K 个邻居特征的局部跨通道交互。 K 的大小决定了每个通道有多少个邻居参与该通道的注意力预测,该参数会直接影响模型的效率。本文运用ECA-net^[27]的 K 值选择公式,如式(5)所示,其中 b 和 γ 分别为2和1。其中的 $|t|_{\text{odd}}$ 表示与 t 最近的奇数。通过1维卷积得到的 F^{GAP} 和 F^{GMP} 两个特征向量基于element-wise进行加和操作,再经过sigmoid激活函数得到通道注意力权重 M_c ,并与输入 F 做乘法操作,得到通道注意力特征图 F' 。

$$K = \lfloor (\log_2(C) + b) / \gamma \rfloor_{\text{odd}} \quad (4)$$

在空间维度上,将前一模块的输出 $F^{C \times H \times W}$ 作为空间注意力模块的输入,基于通道对输入特征图做最大池化和平均池化得到 $F_{\text{max}}^{1 \times H \times W}$ 和 $F_{\text{avg}}^{1 \times H \times W}$ 两种大小为 $1 \times H \times W$ 的通道描述,然后基于通道做连接操作,将两个特征图合并为一个2通道特征图,然后进行 7×7 的卷积操作降维为单通道。再通过sigmoid激活函数,得到空间注意力权重 M_s ,与该模块的输入 F' 做乘法操作,得到空间注意力特征 F'' 。

两个全连接层之后加一个隐层 H 层, H 层是一个全连接层,借鉴特征哈希的思想,使用初始化随机映射,在训练过程中由标签有监督地来控制激活情况。在尽量不损失原始特征表达能力的同时,把原始的4096位的高维特征向量压缩成较低维的特征向量,把原始特征均匀哈希到新特征向量上,减少开销和参数的数量。 H 层的哈希位的长度过大容易导致模型过拟合^[28]。所以 H 层的长度应该较小,表4描述了远节指骨III的发育分期分类准确率在不同的 H 层长度下的变化,从表中可以看到,当 H 层的长度为32位时,分类模型的效果最好。同理,本文为每块参照骨选取了最合适的哈希码长度,其中近节

表4 不同长度哈希码下的模型的准确率(%)

	None	12	24	32	48	128
准确率	94.37	71.83	98.35	98.59	97.18	96.20

表5 14块参照骨的发育分期分类准确率(%)

	mp3	mp5	pp5	pp3	pp1	dp1	dp3	dp5	Capitate	Hamate	Radius	mc1	mc3	mc5
准确率	89.00	96.70	88.60	86.00	97.87	85.30	98.02	96.42	83.66	93.10	83.72	97.43	91.60	92.38

指骨V,远节指骨I,掌骨V的长度为128,其他参照骨的长度均为32。

5 实验结果与分析

5.1 ROI发育分期分类

5.1.1 数据集

为了评估每块参照骨的发育分期,在参照骨发育分期分类模型的训练过程中,我们针对每类参照骨,在其每个发育分期中选择1000张参照骨图谱,按照4:1的比例分为测试集和训练集。例如:近节指骨III在5-11岁内的参照骨发育分期包括3期、3(5)期、4期、4(5)期、5期,每个发育分期选取1000张图谱,其中80%的图像作为训练集,20%的图像作为测试集。将参照骨图像统一归一化为 127×127 的大小,并采用随机旋转,中心裁剪的数据增强技术来避免过拟合,提升模型性能。

5.1.2 实验环境与训练参数

本文采用Pytorch框架进行实验,用到的显示适配器是GeForce RTX 2080 TI,CPU为i7-8700。此外,本文使用梯度下降法作为优化器来训练模型,初始学习率为0.01,动量参数和权重衰减参数分别为0.9和0.0005,训练周期数为300,采用L2正则化控制模型复杂度,减小过拟合。

5.1.3 分类结果

图5展示了部分参照骨包括pp3, Capitate, mp3, dp1, Radius, hamate, mc3在不同的发育分期分类模型下的准确率。可以看出,本文中的神经网络分类模型效果高于AlexNet和其他骨龄评估方法常用的Resnet模型分类准确率。

表5描述了14块参照骨的发育分期分类准确率。可以看出由于部分参照骨在其相邻发育分期之间又增加了新的成熟度指征,使参照骨在不同发育

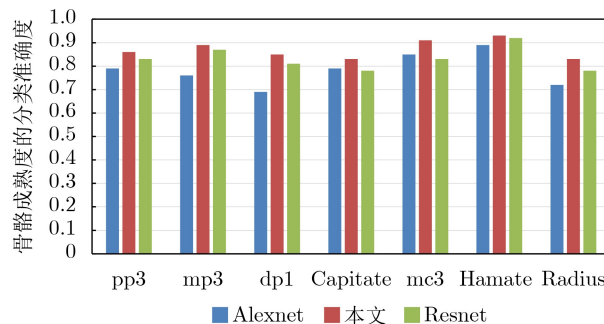


图5 不同方法下各ROI区域的发育分期等级分类准确率

分期下的特征变化并不十分明显, 所以其发育分期分类的准确率比其他参照骨的发育分期分类准确率低。

5.2 骨龄评估

5.2.1 数据集

在评价骨龄评估模型的性能时, 为了保证骨龄标签的可靠性和准确性, 降低专家因为个人知识经验、主观性、专业能力等因素对骨龄值的影响, 我们随机选取2 500多张经过多位专家评估过的儿童X线片作为验证本文方法性能的数据集, 其年龄分布范围在5-11岁。

5.2.2 评价指标

骨龄是一种主观性很强的评价指标, 会由于不同的专家知识与经验、专业水平导致评估结果不一致, 针对一张X线片很难确定一个真正的金标准。因此, 本文骨龄标签值采用两种不同的标准: (1) 将经过多位专家评估的骨龄结果的平均值作为该X线片的骨龄标签值, 可以在一定程度上中和不同专家对于同一块参照骨发育分期评估的不同意见, 在下文中用label1表示。(2) 采用少数服从多数原则, 将多数专家认同的参照骨发育分期所计算的骨龄值作为另一个骨龄标签值, 在下文中用label2表示。因为这两个标签值在骨龄评估领域都可以作为参考依据, 所以将预测值与这两个标签值的误差在一定范围内的准确率作为本文的评价指标。

5.2.3 结果分析

表6展示了在不同年龄段中, 本文的骨龄智能评估方法在label1和label2两个不同标准下的准确率。表7展示了在相同的数据集和算法模型的情况下, 分别采用原始的CHN法和Re_CHN法得到的骨龄值在label1和label2下的准确率。从表中可以看出采用Re_CHN法评估的骨龄与label1的误差在 ± 0.5 岁和 ± 1 岁范围内的准确率分别为87.93%和98.67%。与label2的误差在 ± 0.5 岁和 ± 1 岁范围内的准确率分别为85.62%和96.20%。由于label1和label2在骨龄评估中都可以作为临床应用的评价依

表6 不同年龄组在不同性能指标下的准确率(%)

年龄(岁)	Label1		label2	
	± 0.5 岁	± 1 岁	± 0.5 岁	± 1 岁
5-6	86.88	99.17	86.58	96.76
6-7	85.78	98.95	84.03	94.18
7-8	86.36	98.73	83.45	97.82
8-9	80.70	97.69	81.38	94.02
9-10	89.92	98.32	88.24	97.70
10-11	88.79	97.43	88.48	98.87

据, 所以只要骨龄预测结果与其中一个标签的误差在 ± 0.5 岁范围内, 那么该预测值就是可以被专家认可的。而本文方法的骨龄智能评估结果与专家认可的骨龄值的误差在 ± 0.5 岁范围内的精确率达到了94.6%。相比于最初的CHN法得到的骨龄的准确率更高, 与骨龄标准值之间的误差更小。充分说明Re_CHN法为一些与标准图谱的发育指征皆不相同的参照骨赋予新的成熟度指征会大大地提高骨龄评估的准确率, 可以在一定程度上辅助临床应用, 对提高骨龄评估的效率有一定意义。

5.2.4 与其他方法的对比

在本数据集中, 将本文提出的方法和现有的骨龄评估方法进行对比。Ren等人^[14]先将整张X线片进行预处理, 分别提取图像的粗粒度注意力图和细粒度注意力图, 回归网络采用Inception-V3结构提取图像特征, 结合性别信息直接对骨龄进行评估。Wu等人^[29]先将X线片的手部区域分割出来生成手部掩模, 再利用residual attention network网络对骨龄进行评估。表8描述了这几种方法和本文方法在本文数据集下预测的骨龄与专家认可的骨龄值的误差在 ± 0.5 岁和 ± 1 岁范围内的准确率, 可以看出本文所提出的骨龄智能评估方法的准确率更加有优势。

6 结束语

在CHN法中由于部分参考图谱的发育指征跨度较大, 忽略了发育过程中的一些重要指征, 导致发育指征介于相邻发育分期之间的参照骨无法更加精细地判别骨发育成熟度, 使最终的骨龄评价产生误差。本文以CHN法为基础, 在相邻等级间隔跨度较大的参照骨图谱之间勾绘新的成熟度指征, 并利用层次分析法为其分配成熟度得分, 提高骨龄评

表7 分别采用CHN法和Re_CHN法得到的骨龄评估准确率(%)

标准	BAA方法	± 0.5 岁	± 1 岁
label1	Re_CHN	87.93	98.67
	CHN	78.27	97.54
label2	Re_CHN	85.62	96.20
	CHN	72.78	95.81
label1 or label2	Re_CHN	94.60	99.13
	CHN	85.18	98.50

表8 不同方法在本文数据集下的性能指标(%)

方法	± 0.5 岁	± 1 岁
Regression CNN ^[14]	69.60	93.05
Residual Attention Network ^[29]	76.73	91.40
本文方法	94.60	99.13

估的准确率。此外,本文融合了Harris角点检测、注意力机制和特征哈希技术设计一个模型对参照骨的关键区域进行特征提取,评估参照骨的成熟度发育分期。在验证实验中,本文提出的方法评估的骨龄结果与专家认可的骨龄值的误差在 ± 0.5 岁范围内的准确率为94.6%,可以辅助临床应用,对提高骨龄评估的效率有一定的意义。未来的研究将专注于11~18岁的儿童青少年的骨龄智能评估。

参考文献

- [1] 潘复平, 张国栋. 骨龄与青春期发育关系的追踪观察[J]. 中华预防医学杂志, 1985, 19(2): 79-82.
PAN Fuping and ZHANG Guodong. Follow-up observation on the relationship between bone age and puberty development[J]. *Chinese Journal of Preventive Medicine*, 1985, 19(2): 79-82.
- [2] 邵伟东, 金春华, 潘慧, 等. 中国儿童手腕部骨龄评测标准 CHN法与参考图谱[M]. 北京: 中国协和医科大学出版社, 2018: 15-20.
SHAO Weidong, JIN Chunhua, PAN Hui, et al. Chinese Children's Wrist Bone Age Evaluation Standard[M]. Beijing: Peking Union Medical College Press, 2018: 15-20.
- [3] 张烨城. 骨龄在体育教学及训练中的应用[J]. 青少年体育, 2018(9): 58-59. doi: 10.3969/j.issn.2095-4581.2018.09.037.
ZHANG Yecheng. Application of bone age in physical education teaching and training[J]. *Youth Sport*, 2018(9): 58-59. doi: 10.3969/j.issn.2095-4581.2018.09.037.
- [4] 沈勋章. 手腕部骨龄鉴定方法的研究进展[J]. 中国医药科学, 2011, 1(12): 9-12.
SHEN Xunzhang. Identification of wrist skeletal age of the research progress[J]. *China Medicine and Pharmacy*, 2011, 1(12): 9-12.
- [5] SPAMPINATO C, PALAZZO S, GIORDANO D, et al. Deep learning for automated skeletal bone age assessment in X-ray images[J]. *Medical Image Analysis*, 2017, 36: 41-51. doi: 10.1016/j.media.2016.10.010.
- [6] GREULICH W W and IDELL PYLE S. Radiographic atlas of skeletal development of the hand and wrist[J]. *The American Journal of the Medical Sciences*, 1959, 238(3): 393.
- [7] TANNER J M and WHITEHOUSE R H. Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty[J]. *Archives of Disease in Childhood*, 1976, 51(3): 170-179. doi: 10.1136/adc.51.3.170.
- [8] 叶义言. 新版骨龄评分法概述[J]. 中华儿科杂志, 2004, 42(1): 30-32. doi: 10.3760/j.issn:0578-1310.2004.01.009.
YE Yiyang. Overview of new version of bone age scoring method[J]. *Chinese Journal of Pediatrics*, 2004, 42(1): 30-32. doi: 10.3760/j.issn:0578-1310.2004.01.009.
- [9] 张绍岩, 花纪青, 刘丽娟, 等. 中国人手腕骨发育标准—中华05. III. 中国儿童骨发育的长期趋势[J]. 中国运动医学杂志, 2007, 26(2): 149-153. doi: 10.16038/j.1000-6710.2007.02.004.
ZHANG Shaoyan, HUA Jiqing, LIU Lijuan, et al. The standards of skeletal maturity of hand and wrist for Chinese-China 05. III. The secular trend of skeletal development in Chinese children[J]. *Chinese Journal of Sports Medicine*, 2007, 26(2): 149-153. doi: 10.16038/j.1000-6710.2007.02.004.
- [10] 张绍岩, 杨士增, 邵伟东, 等. 中国人手腕骨发育标准—CHN法[J]. 体育科学, 1993, 13(6): 33-39.
ZHANG Shaoyan, YANG Shizeng, SHAO Weidong, et al. The standards of skeletal development of hand and wrist for Chinese-CHN method[J]. *China Sports Science*, 1993, 13(6): 33-39.
- [11] THODBERG H H, KREIBORG S, JUUL A, et al. The BoneXpert method for automated determination of skeletal maturity[J]. *IEEE Transactions on Medical Imaging*, 2009, 28(1): 52-66. doi: 10.1109/tmi.2008.926067.
- [12] LEE H, TAJMIR S, LEE J, et al. Fully automated deep learning system for bone age assessment[J]. *Journal of Digital Imaging*, 2017, 30(4): 427-441. doi: 10.1007/s10278-017-9955-8.
- [13] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 1717-1724. doi: 10.1109/CVPR.2014.222.
- [14] REN Xuhua, LI Tingting, YANG Xiujuan, et al. Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph[J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(5): 2030-2038. doi: 10.1109/jbhi.2018.2876916.
- [15] HAN Yaxin and WANG Guangbin. Skeletal bone age prediction based on a deep residual network with spatial transformer[J]. *Computer Methods and Programs in Biomedicine*, 2020, 197: 105754. doi: 10.1016/j.cmpb.2020.105754.
- [16] LIU Bo, ZHANG Yu, CHU Meicheng, et al. Bone age assessment based on rank-monotonicity enhanced ranking CNN[J]. *IEEE Access*, 2019, 7: 120976-120983. doi: 10.1109/access.2019.2937341.
- [17] SON S J, SONG Y, KIM N, et al. TW3-based fully automated bone age assessment system using deep neural networks[J]. *IEEE Access*, 2019, 7: 33346-33358. doi: 10.1109/access.2019.2903131.
- [18] 刘宗才, 吴锦华, 王荣品, 等. 深度学习骨龄评测系统对贵州省儿童及青少年骨龄测评的准确性[J]. 中国医学影像技术, 2019,

- 35(12): 1799–1803. doi: [10.13929/j.1003-3289.201907037](https://doi.org/10.13929/j.1003-3289.201907037).
- LIU Zongcai, WU Jinhua, WANG Rongpin, *et al.* Accuracy of deep learning based bone age assessment system of children and adolescents in Guizhou[J]. *Chinese Journal of Medical Imaging Technology*, 2019, 35(12): 1799–1803. doi: [10.13929/j.1003-3289.201907037](https://doi.org/10.13929/j.1003-3289.201907037).
- [19] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: [10.1109/tpami.2016.2577031](https://doi.org/10.1109/tpami.2016.2577031).
- [20] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90. doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [21] LIU Jun, WANG Guang, DUAN Lingyu, *et al.* Skeleton-based human action recognition with global context-aware attention LSTM networks[J]. *IEEE Transactions on Image Processing*, 2018, 27(4): 1586–1599. doi: [10.1109/tip.2017.2785279](https://doi.org/10.1109/tip.2017.2785279).
- [22] WANG Jinrui, LI Shunming, AN Zenghui, *et al.* Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines[J]. *Neurocomputing*, 2019, 329: 53–65. doi: [10.1016/j.neucom.2018.10.049](https://doi.org/10.1016/j.neucom.2018.10.049).
- [23] DAHL G E, SAINATH T N, and HINTON G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, 2013: 8609–8613. doi: [10.1109/ICASSP.2013.6639346](https://doi.org/10.1109/ICASSP.2013.6639346).
- [24] KO B, KIM H G, OH K J, *et al.* Controlled dropout: A different approach to using dropout on deep neural network[C]. 2017 IEEE International Conference on Big Data and Smart Computing, Jeju, Korea (South), 2017: 358–362. doi: [10.1109/BIGCOMP.2017.7881693](https://doi.org/10.1109/BIGCOMP.2017.7881693).
- [25] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional block attention module[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: [10.1007/978-3-030-01234-2_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [26] HU Jie, SHEN Li, ALBANIE S, *et al.* Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011–2023. doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372).
- [27] WANG Qilong, WU Banggu, ZHU Pengfei, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11531–11539. doi: [10.1109/CVPR42600.2020.01155](https://doi.org/10.1109/CVPR42600.2020.01155).
- [28] LIU Haomiao, WANG Ruiping, SHAN Shiguang, *et al.* Deep supervised hashing for fast image retrieval[J]. *International Journal of Computer Vision*, 2019, 127(9): 1217–1234. doi: [10.1007/s11263-019-01174-4](https://doi.org/10.1007/s11263-019-01174-4).
- [29] WU E, KONG Bin, WANG Xin, *et al.* Residual attention based network for hand bone age assessment[C]. 2019 IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 2018: 1158–1161. doi: [10.1109/ISBI.2019.8759332](https://doi.org/10.1109/ISBI.2019.8759332).
- 毛科技: 男, 副教授, 研究方向为医学图像处理、智能计算、大数据分析。
- 武坤秀: 女, 硕士生, 研究方向为医学图像处理。
- 陆伟: 男, 硕士生, 研究方向为医学图像处理。
- 陈立建: 男, 博士生, 研究方向为医学图像处理。
- 毛家发: 男, 教授, 研究方向为医学图像处理。

责任编辑: 马秀强