

Bone Age Recognition Using Convolution Neural Network

Yongcan Wang^{1,2}, Yong Hu³, Yanyan Shen¹, Shuqiang Wang^{1*}, Zuhui Wang¹

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen

²University of Science and Technology of China, Hefei Anhui

³The University of Hong Kong, Hong Kong

Email: *sq.wang@siat.ac.cn

Received: Dec. 20th, 2017; accepted: Jan. 2nd, 2018; published: Jan. 9th, 2018

Abstract

Patients' bone maturity analysis plays an important role in determining clinical management. At present, skeletal maturity assessment is still highly depended on physicians' evaluation. This process is time-consuming and often contains interrater variability. In this paper, based on the skeletal maturity evaluation criteria and the clinical applications, the authors propose an automatic bone age recognition scheme using convolutional neural network models. The scheme will first accurately extract distal radius and ulna regions from hand radiographs, then it will help us classify the bone age phases automatically. After parameters optimization and data augmentation, our deep learning model finally achieves 90% and 92% recognition accuracies on radius and ulna datasets respectively.

Keywords

Convolution Neural Network, Bone Age, X-Ray Image, Classification

基于卷积神经网络的骨龄阶段识别研究

王永灿^{1,2}, 胡 勇³, 申妍燕¹, 王书强^{1*}, 王祖辉¹

¹中国科学院深圳先进技术研究院, 深圳

²中国科学技术大学, 安徽 合肥

³香港大学, 香港

Email: *sq.wang@siat.ac.cn

收稿日期: 2017年12月20日; 录用日期: 2018年1月2日; 发布日期: 2018年1月9日

*通讯作者。

文章引用: 王永灿, 胡勇, 申妍燕, 王书强, 王祖辉. 基于卷积神经网络的骨龄阶段识别研究[J]. 图像与信号处理, 2018, 7(1): 1-15. DOI: 10.12677/jisp.2018.71001

摘要

骨骼成熟度分析对于指导临床管理等具有重要作用,而目前评定骨骼成熟度还主要靠医师进行人工读片。本文根据骨骼成熟度评价指标与临床应用实际,提出一种基于卷积神经网络的骨龄阶段自动识别方法,实现从手部X光片中自动检测出尺骨、桡骨末端区域,并对其进行骨龄阶段分类的自动化识别过程。最终,通过卷积神经网络建模训练,参数优化、数据增强等技术的运用,尺骨与桡骨分别获得了90%与92%的识别精度。

关键词

卷积神经网络, 骨龄, X光图像, 分类

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

骨龄作为生长发育程度的一项重要指标,在医学、体育及司法鉴定等领域中发挥着重要作用,特别是在对青少年脊柱侧凸等病人的临床管理中,进行骨骼成熟度分析以理解其生长的高峰期与停止期对于决定临床观察间隔、定时开始和终止支撑治疗至关重要。而由于手腕部骨块数量较多、包含的信息量大,且采集更加方便,故通常用于评价骨骼成熟度(骨龄)。

目前国际上较为常用的骨骼发育成熟度评价方法有 G-P 图谱法和 TW 计分法[1]。而由于各国人骨骼发育情况存在较大差异,我国也先后制定了李果珍法、CHN 法、中华-05 标准等评价方法[1]。针对目前常用的临床和放射学方法评定骨骼成熟度在预测生长高峰期与停止期仍然不足,且我国儿童的生长发育出现了显著加速的长期趋势,中国儿童青少年生长发育整体提前。按照 TW3 (Tanner Whitehouse 3)方法,峰值高度速率和手指骨骺成熟度不同阶段紧密相关,尺骨与桡骨末端(distal radius and ulna, DRU)生长过程贯穿骨骼生长的整个周期,且是最后一个结束。2013 年香港大学的 Luk 等基于国际通用的 TW3 方法,依据当代中国儿童为样本,提出了一种通过桡骨远端和尺骨(DRU)分类的评定标准来评定骨骼成熟度[2],相对于之前选用的中华-05 等骨龄评价标准更加符合目前中国青少年骨骼发育情况。但这些都是通过放射科医师根据标准进行人工判断分类,耗时耗力且存在较大主观性。

随着深度学习研究的火热发展,深度神经网络模型逐渐被引入各个领域,并取得了不错的效果。在医疗诊断过程中,精准地诊断通常要借助于高质量的医学影像检查。随着近年来医学成像技术的不断提升,医院具有众多高端的成像设备以更快地获取更高质量的医学图像。但对于图像的解释判断一般还是由医师完成,不但费时费力,而且存在较多主观因素。计算机辅助检测在临床实践与研究中是一个重要工具,可以利用机器学习与图像处理等技术进行自动化诊断。但传统方法效果还不够理想,近年来一些利用深度学习技术的研究取得了较好的效果[3] [4] [5] [6] [7],体现出了其在该方面的优越性能。采用深度学习方法分析桡骨和尺骨 X 射线图片来自动分类评估骨骼成熟度,从而理解其生长的高峰期与停止期对于决定青少年脊柱侧凸等病人的临床管理有重要意义。

2. 相关工作

2.1. DRU 骨骼成熟度评定标准

Luk 等提出的一种使用桡骨和尺骨末端 X 光片进行骨骼成熟度评定的分类架构[2]中, 定义桡骨和尺骨骨骺成熟的不同阶段, 桡骨分级为 R1~R11, 尺骨分级为 U1~U9, 各阶段桡骨、尺骨图片如图 1 所示。

该分类架构中给出的桡骨各阶段特征如下: R1 阶段: 桡骨末端有一个或多个点状骨骺; R2 阶段: 桡骨末端出现不规则椭圆形骨骺; R3 阶段: 骨骺最大直径超过桡骨末端宽度的一半; R4 阶段: 在骨骺的远端边缘出现双线, 表示手掌和背面; R5 阶段: 在骨骺的近端边缘可见不规则的增厚的白线, 在内侧更明显, 呈现凹形, 骨骺宽度小于干骨后端宽度; R6 阶段: 骨骺内外侧都超过了干骨后端, 骨骺宽度大于干骨后端宽度, 与尺骨交接处关节开始发育; R7 阶段: 骨骺在内侧覆盖桡骨末端, 在外侧呈现出半球形; R8 阶段: 帽出现在近端外侧, 生长板清晰可见, 骨骺边缘比中间部分宽; R9 阶段: 骨骺几乎完全覆盖了干骨后端, 并伴随着生长板空间的硬化, 生长板可见但是模糊, 骨骺内侧和外侧几乎与干骨后端相接; R10 阶段: 生长板完全消失, 形成硬化线, 有时出现虚线; 在生长板的内侧和外侧端部可以看到凹口; R11 阶段: (骨骺)与桡骨末端完全融合, 生长板瘢痕可能仍然可见。

该分类架构中给出的尺骨各阶段特征如下: U1 阶段: 尺骨末端有一个或多个点状骨骺; U2 阶段: 尺骨末端出现不规则椭圆形骨骺; U3 阶段: 骨骺最大直径超过尺骨末端宽度的一半; U4 阶段: 在骨骺的内侧末端可见茎状突起; U5 阶段: 尺骨的头部完全显现, 且密度比茎状突起部分大, 毗邻端骨骺的边界是扁平状的; U6 阶段: 骨骺宽度等于干骨后端宽度, 骨骺的末端边缘覆盖了干骨后端中间 1/3; U7 阶段: 生长板变窄, 骨骺的末端边缘与干骨后端形成平滑的曲线, 曲线中间半部分可以看到有融合; U8 阶段: 茎状突起末端可见未融合的生长板, 骨骺的末端边缘与干骨后端几乎融合, 但茎状突起下方可见生长板; U9 阶段: 完全融合。

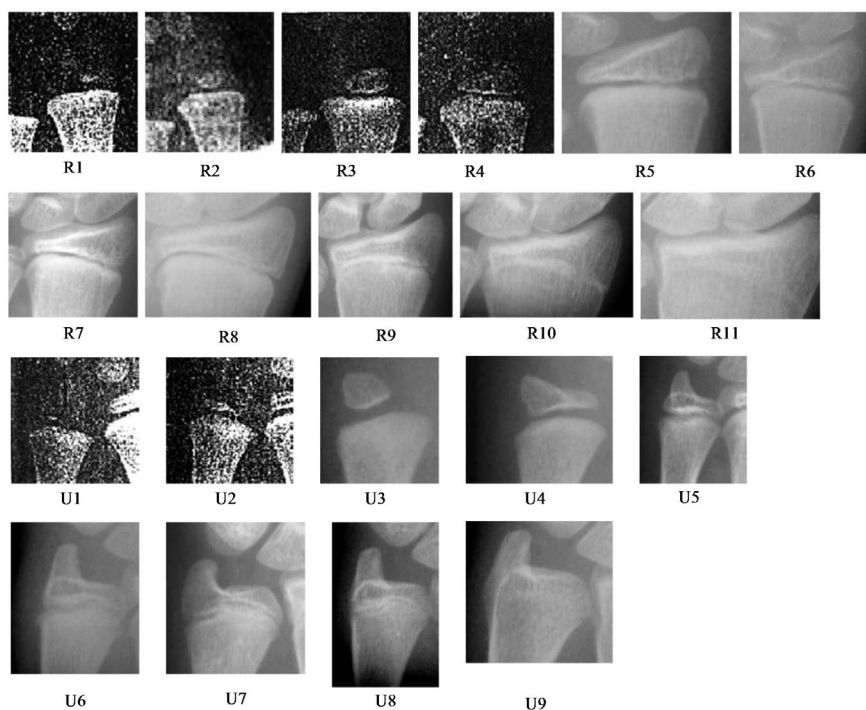


Figure 1. Radius and ulna at various stages

图 1. 各阶段桡骨与尺骨

该架构提供了青少年生长陡增和停止的紧密关系，对于改善临床决策有极大利用价值。该架构研究了这些骨骼每一阶段改变对应的骨龄和性征、站高、坐高、臂距、长骨长度的发展，发现站高、坐高、和臂距生长在 R7 (平均, 11.4 岁)阶段和 U5 (平均, 11.0 岁)阶段达到高峰，长骨生长高峰也是在 R7 和 U5，高度和臂距生长停止是在 R10 (平均, 15.6 岁)阶段和 U9 (平均, 17.3 岁)阶段。表 1 为与桡骨及尺骨阶段联系的生长改变。

与桡骨阶段联系的生长改变：两阶段实际年龄平均间隔是 1.1 年，而骨龄平均间隔是 1.5 年。站高、坐高、臂距三者高峰期都在 R7 阶段，站高、坐高生长停止期在 R9，臂距停止期在 R11。长骨生长高峰期在 R7，停止在 R11。

与尺骨阶段联系的生长改变：两阶段间平均实际年龄间隔是 1.1 年，平均骨龄间隔是 1.3 年。站高和臂距生长高峰在 U4，坐高在 U5，生长停止都在 U9。长骨生长高峰在 U5，停止在 U9。

在实际应用中往往只需要知道其骨骼生长发育所处阶段，特别是对于生长高峰期与生长停止期等关键阶段的判断。特别是在分析青少年脊柱侧凸等病人骨骼成熟度以用于指导临床管理中，理解其生长的高峰期与停止期，对于决定临床观察间隔、定时开始和终止支撑治疗至关重要。由表 1 可知，生长高峰期时，桡骨一般处在 R7 阶段，尺骨一般处在 U4、U5 阶段，而生长停止期时，桡骨一般在 R9 或 R11 阶段，尺骨处在 U9 阶段。以该分类架构用于青少年脊柱侧凸病人临床管理，最小化支柱支撑时间时，通常在 R7/U5 时开始使用支撑，R8 或 R9 时结束支撑。R11 和 U9 时骨骼成熟，可确定是否需要手术融合。

一般来讲，桡骨 R7 阶段各项生长发育速率达到最高峰，随后的桡骨 R8、R9 阶段生长速率逐渐下降，到 R10、R11 阶段，生长发育速率基本接近于 0，进入成熟期，停止生长。相应的在尺骨 U4、U5 阶段生长速率达到最高峰，随后的 U6、U7 阶段速率逐渐下降，到 U8、U9 阶段时，生长速率接近于 0，进入成熟期，停止生长。基于以上分析，本文将桡骨与尺骨对应生长发育阶段划分为 4 个阶段，即生长发育前期、高峰期、停滞期、成熟期。桡骨 R7、尺骨 U4、U5 阶段对应高峰期、高峰期之前为前期，即桡骨 R1~R6 阶段、尺骨 U1~U3 阶段，桡骨 R8、R9 阶段、尺骨 U6、U7 阶段为发育停滞期，桡骨 R10、R11 阶段、尺骨 U8、U9 阶段为成熟期，如图 2 所示。

在本文中桡骨与尺骨按上述生长发育阶段分别进行分类，由于前期中桡骨 R1~R4 阶段及尺骨 U1~U3 阶段数据缺乏，桡骨阶段分类中前期仅使用 R5、R6 数据、尺骨阶段分类仅做高峰期、停滞期、成熟期三分类，如图 3 所示。

Table 1. The relationship between growth change and radius and ulna stage

表 1. 成长改变与桡骨及尺骨阶段对应表

成长改变		桡骨阶段	尺骨阶段
两阶段实际年龄间隔		1.1 年(0.3~1.8)	1.1 年(0.2~1.6)
平均骨龄间隔		1.5 年(1.2~1.8)	1.3 年(0.6~2.2)
生长高峰期	站高	R7	U4
	坐高	R7	U5
	臂距	R7	U4
	长骨	R7	U5
生长停止期	站高	R9	U9
	坐高	R9	U9
	臂距	R11	U9
	长骨	R11	U9

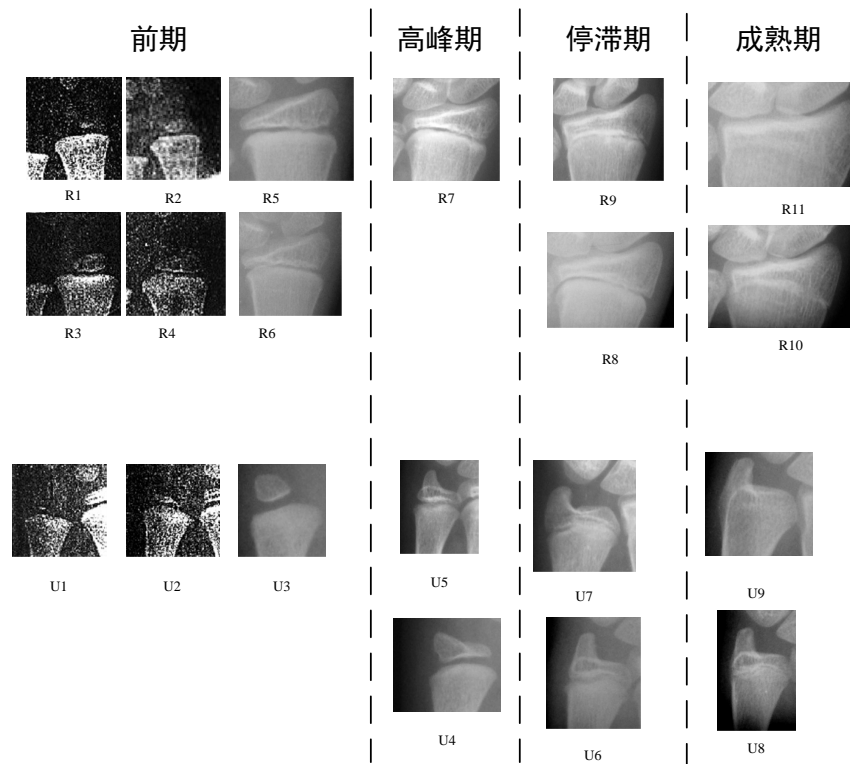


Figure 2. The corresponding relationship between the stages of development and radius and ulna

图 2. 各生长发育阶段与桡骨与尺骨阶段对应关系

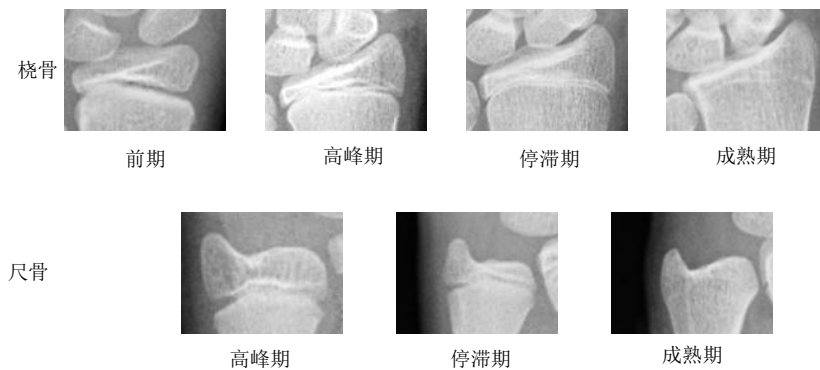


Figure 3. Radius and ulna at various stages of classification in this paper

图 3. 本文中各分类阶段桡骨与尺骨

用于青少年脊柱侧凸病人临床管理，最小化支柱支撑时间时，则当两者同时为发育高峰期时，开始支撑，当桡骨进入发育停滞期，结束支撑，两者都为成熟期时，确定是否需要手术融合。

2.2. 骨龄自动识别

通过影像医师人工分析耗时耗力，且存在一定主观性，所以研究骨龄阶段自动识别方法得到重视。

国外对骨龄自动识别方面的研究最早可追溯到 1989 年，Michael 和 Nelson 设计并实现的计算机辅助的骨龄自动识别系统[8]。2001 年，加州大学圣迭戈分校小组通过指骨区域分割方法实现了基于 TW2 评价方法的骨龄自动识别[9]。2003 年，Bocchi 等人首先通过 Gabor 变换与多尺度滤波预处理 X 光片中各

类骨骼，再利用神经网络实现了基于 TW2 法的骨龄自动识别[10]。2007 年，Gertych 等人通过腕骨自动分割、形态学特征提取与模糊分类三步识别骨龄[11]。2017 年，Lee 等人提出了一种基于深度学习方法的骨龄预测方法，首先分割出整个手部区域，再进行分类[12]。但由于各地区人种差异，我国青少年骨骼发育情况与欧美等国家不能使用统一标准与方法。

国内对骨龄阶段自动识别的研究开始于 20 世纪末。1998 年第四军医大学开发了可实现自动查表与计算的计算机辅助评价系统[13]。2010 年，董娜等通过分割手指区域，提取骨骼特征，使用 SVM 分类器实现了基于 CHN 的骨龄自动识别[14]。2011 年，王勇对桡骨进行自适应分割，然后联合骨块局部特征与全局特征，使用改进的基于混沌优化的分类器进行骨龄自动评估[15]。2014 年，安徽大学赵娟提出使用 k-余弦曲率算法分割指骨特征区域，提取几何形状特征，并用 SVM 多分类算法实现基于 CHN 的阶段划分[16]。2016 年，朱翔宇等提出以区域生长法等分割桡骨，然后提取灰度、纹理等特征，使用 k 近邻分类算法对中华-05 标准中桡骨骨龄 80 级~84 级进行自动识别[17]。但这些方法基本都是通过先手工提取特征，再使用传统分类器进行分类，未实现完全自动化识别。

2.3. 卷积神经网络

卷积神经网络作为最重要的神经网络模型之一，一直以来在深度学习领域发挥着重要作用。其作为最早在深度模型中表现出优秀性能的神经网络模型，早在上世纪 90 年代就被运用于商业应用，解决支票中的手写体识别问题。而如今随着技术的不断改进，与应用领域的不断拓展，卷积神经网络更是被运用到了图像理解、文本处理、语音识别等众多领域。同时，卷积神经网络在各大比赛中被广泛应用，并取得优异成绩，特别是在 2012 年的 ImageNet 图像识别比赛中以显著优势领先其他方法赢得比赛[18]，将深度学习技术带人新一轮高潮。

卷积神经网络也是一种前馈神经网络，与普通神经网络类似，都有输入层、隐藏层和输出层，由输入经逐层参数运算，最终得到输出，根据理想输出与实际输出间的误差，再反向传播调整网络参数，从而获取一个合适的模型。而卷积神经网络与普通神经网络不同之处在于，普通神经网络的隐层都是全连接层，而卷积神经网络隐层通常为卷积层与池化层交替，然后连接全连接层。卷积神经网络利用多个卷积核对原始数据进行卷积，每个卷积核提取图像的一类底层特征，对应的生成一张特征图，下一层各卷积核再对这些特征图进行卷积，然后经过逐层组合得到更高级的特征，实现特征自动提取。而池化操作可以降低数据的维度，同时使得对小范围的平移和形变保持一定的特征不变性。

卷积神经网络开始主要用于图像识别，用来处理图片这类由多维矩阵组成的数据，由于卷积层这种局部连接、权值共享的结构，极大地缩减了网络参数的同时，方便了对图像这类多维数据的处理，同时保留了图像的空间结构特征信息，使得不需要额外的网络参数学习网络的空间特征信息。

3. 数据与方法

3.1. 数据

本实验所使用的桡骨和尺骨末端 X 光片图像数据来源于香港大学深圳医院，经多批次获取，单个样本原始图片如图 4 所示，各图片大小尺寸不一。同时提供有 xls 格式保存的标签文件，文件与标签经日期与序号匹配，后期处理中，可匹配到有效标签图片共 1101 张。标签为放射科医师按 Luk 等提出的 DRU 骨骼成熟度评定标准标注，桡骨分级为 R1~R11 共 11 类，尺骨分级为 U1~U9，共 9 类。

取原数据样本中对应尺骨高峰期阶段数据 200 张，停滞期阶段图片 250 张图片，成熟期阶段样本 150 张。三阶段共计 600 张样本，随机从各阶段中取 50 张，共计 150 张作为测试集，剩余的 450 张作为训练集，如表 2 所示。



Figure 4. Data sample

图 4. 数据样本

Table 2. Sample size of ulna three classification at different stages

表 2. 尺骨三分类各阶段样本量

	高峰期	停滞期	成熟期	总计
训练集	150	200	100	450
测试集	50	50	50	150

取原数据样本中 R5、R6 阶段的图片共 75 张，构成桡骨生长前期样本。从原 R7 阶段 96 张图片中选取 75 张图片，构成桡骨生长高峰期样本。从原 R8、R9 阶段图片中选取 175 张图片，构成桡骨生长停滞期样本。取原 R10、R11 阶段中 75 张图片构成桡骨生长成熟期样本。四阶段共计 400 张样本，随机从各阶段中取 25 张，共计 100 张作为测试集，剩余的 300 张作为训练集，如表 3 所示。

3.2. 方法

由于原始图像为拍摄的整个手部的 X 光片，如图 4 所示，整张图片尺寸较大，且大小不一，不利于作为卷积神经网络输入，如典型的两个尺寸为 1400×900 、 989×1302 。而对于我们此处分类识别的有用信息尺骨、桡骨末端区域仅为原始图片中位置相对固定的一小部分，且尺骨、桡骨末端区域尺寸变化不大，大小相对固定。为降低其他无关区域的影响，且降低需识别区域的大小，我们先使用目标检测方法，检测出尺骨、桡骨末端区域对应位置，再根据检测得到的位置坐标剪切出尺骨、桡骨图片，调整成统一大小。

通过目标检测算法从原始图片中检测并剪裁出尺骨与桡骨末端区域的图片后，再按照重划分方案，为图片匹配相应的标签。再训练相应的分类器，以运用于最终的骨骼发育阶段检测。从而达到输入一张图片，经训练好的卷积神经网络分类器自动判断其对应的骨骼发育阶段，并可可视化输出，如图 5 所示。

所以整个骨骼发育阶段识别模型主要包括从原始的整个手部 X 光片中检测出所需要的尺骨与桡骨末端区域的目标检测模型和通过尺骨与桡骨末端区域识别骨骼发育阶段的分类器模型以及前期的预处理、中间的由检测结果进行自动剪裁与大小调整和最后的结果可视化输出。整体流程如图 6 所示。

3.2.1. 检测

为将尺骨、桡骨作为独立个体检测，并剪裁输出，分别进行阶段分类。为保持图像的原始分辨率，剪裁出的图片尺寸尽量保持与在原始图片中的尺寸相近，且为方便后续卷积神经网络的设计，尺骨、桡骨分别输出的图片统一调整大小为 128×96 。

Table 3. Sample size of each stage of radius of four classifications

表 3. 桡骨四分类各阶段样本量

	前期	高峰期	停滞期	成熟期	总计
训练集	50	50	50	50	300
测试集	25	25	25	25	100

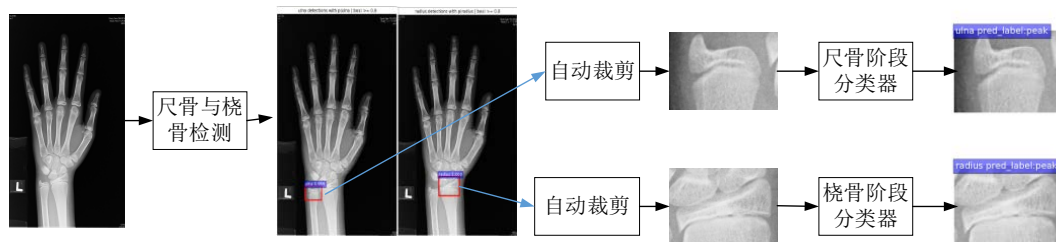


Figure 5. Overall architecture

图 5. 整体架构

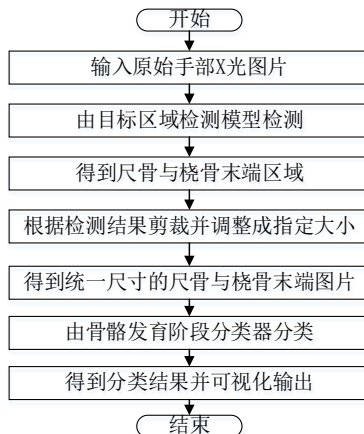


Figure 6. Overall flow chart

图 6. 整体流程图

其中目标区域检测算法使用 Faster R-CNN 算法[19]，基础网络选用 ZFNet [20]，即共享卷积层为 5 层卷积，第一层卷积为 96 个 7×7 卷积，接一个 3×3 的最大池化，第二层卷积为 256 个 5×5 卷积，接一个 3×3 的最大池化，步长都为 2，后面接 3 层 3×3 卷积，特征图个数分别为 384、384、256，步长都为 1。在共享卷积层上加一个 3×3 卷积层，并连接 2 个 1×1 的卷积分支，分别输出初选区域分类得分与边界框，构成 RPN 网络提取感兴趣区域，在共享卷积层上通过感兴趣区域池化层连接两层 4096 的全连接层，然后由两个分支分别输出分类得分与边界框坐标，构成 Fast R-CNN 网络[21]预测最终的边界框位置与分类。其中网络输出类别数设为要检测的类别数加背景类，即为 3。网络结构如图 7 所示。

训练过程中使用小批量数据梯度下降，初始学习率为 0.01，权重初始化使用高斯分布。第一步 RPN 训练迭代 8000 次，第二步 Fast R-CNN 训练迭代 4000 次，第三步 RPN 训练迭代 8000 次，第四步 Fast R-CNN 训练迭代 4000 次。

3.2.2. 分类

卷积神经网络作为一种多层神经网络模型，一般包含输入层、隐藏层和输出层，而其中隐藏层通常由一个或多个的卷积层与池化层交替结构连接一个或多个全连接层构成。图片等数据源由输入层输入，

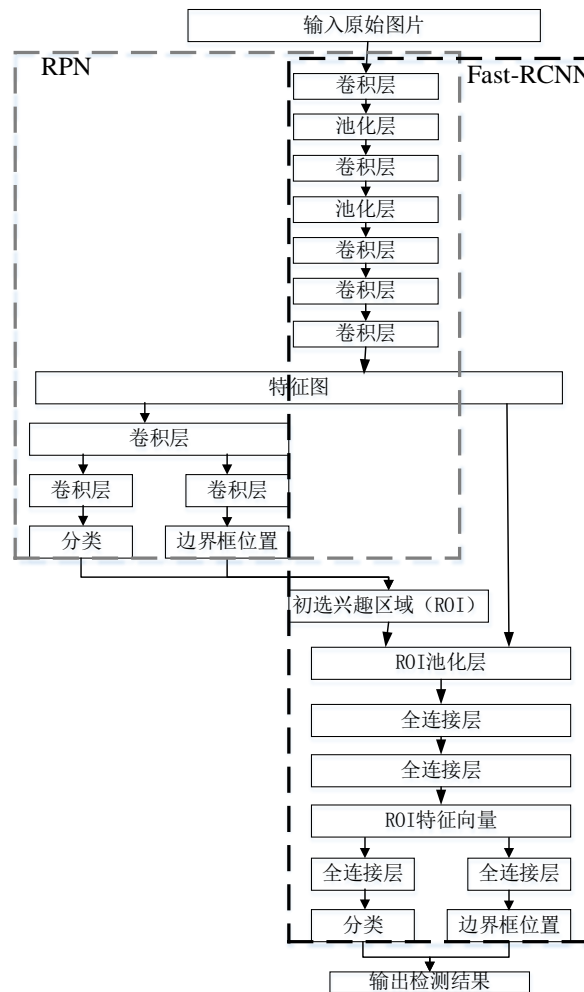


Figure 7. Network structure diagram

图 7. 网络结构图

经卷积层提取特征，池化层降低数据维度并提高特征不变性，多层卷积池化组合特征后，由全连接层选取其中的有效特征构建与输出的映射关系。卷积神经网络基本结构示意图如图 8 所示。

网络架构为 128×96 的尺骨或桡骨图片输入，卷积层为 4 层或 5 层，卷积核大小选用 ZFNet 网络的 7×7 、 5×5 、 3×3 或 VGGNet [22] 的全 3×3 结构，卷积核个数在 32、64、128 间选取，每层卷积层后接一个 2×2 的最大池化层，外加两层全连接层，全连接层结点数为 4096 或 1024，全连接层可选用 Dropout [23]，具体选用根据后续实验比较确定，激活函数均使用 ReLU，以加快网络训练过程。

4. 实验与结果

4.1. 实验设置

实验环境为 Ubuntu16.04 操作系统，CPU 为 Intel® Xeon(R) CPU E5-1620 v3，主频为 3.50 GHz，GPU 为 NVIDIA Quadro M4000，CUDA8.0，cuDNN5.0。分类实验使用 Theano、Lasagne 深度学习框架。训练过程中，使用 Adam 优化算法[24]最小化多分类交叉熵损失函数，Adam 是一种自适应参数更新方法，包括 3 个参数值，分别为学习率和两个衰减率参数，取 0.0001 与 0.9、0.999。权重参数 W 使用 Xavier 均匀分布初始化，偏置 b 使用常数 0 初始化。参数更新按小批量进行，每批样本量为 16。

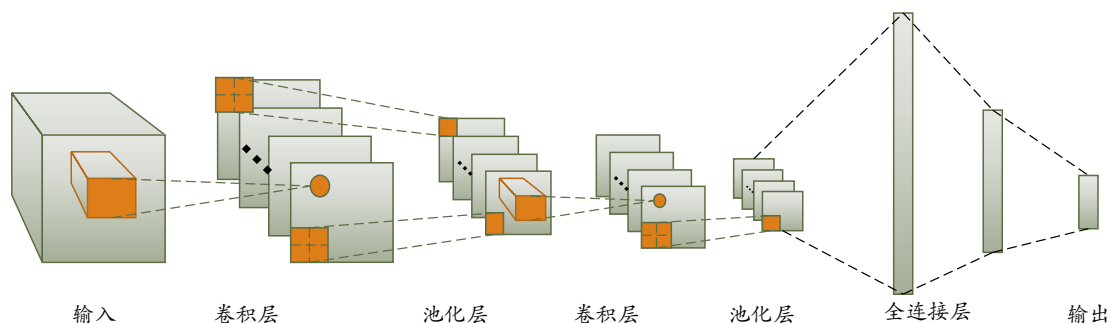


Figure 8. The basic structure of convolution neural network

图 8. 卷积神经网络基本结构示意图

4.2. 分类结果

4.2.1. 样本均衡

根据研究表明, 当数据样本不均衡时, 若少数类不能充分学习到有效特征, 则会根据样本分布, 以大概率判别为多数类[25]。为解决该问题, 通常对少数类样本使用过采样方法以达到与多数类平衡。此处尺骨分类使用的训练样本量原始分布为 150、200、100, 存在轻微不平衡, 分别对两少数类样本进行简单过采样, 使得三个阶段类的样本量都为 200。桡骨使用的训练样本量原始分布为 50、50、150、50, 分别对少数类样本进行简单过采样, 使得三个阶段类的样本量都为 150。

使用 4 层卷积、2 层全连接的基本网络结构, 进行对比试验, 结果见表 4 所示。具体网络结构为第一层卷积为 32 个 7×7 卷积核, 第二层 64 个 5×5 卷积核, 第三层 64 个 3×3 卷积核, 第四层为 32 个 3×3 卷积核, 全连接层节点分别为 4096 和 1024 个, 再接输出层 3 个节点。卷积均为有效卷积, 不改变大小, 每层卷积后接一层 2×2 的最大池化。

由表 4 实验结果知, 对不平衡样本进行简单过采样, 使各类样本均衡后, 分类精度略有提高, 但由于原数据样本不均衡性并不严重, 所以提升不明显。但训练过程中, 均衡样本收敛速度明显快于不平衡样本。如图 9(a)所示, 在非均衡样本下网络在训练迭代至 40 代左右才收敛, 而在均衡样本下, 如图 9(b)所示, 网络在训练迭代至 20 代左右就已收敛。

4.2.2. 数据增强

在训练神经网络模型过程中, 为了避免过拟合, 通常根据不同任务针对性的选取位置平移变换、尺度变换、旋转、反射变换、颜色变换、对比度变换、遮挡、噪声扰动等变换方式中的几种进行数据增强从而间接获取更多数据[26]。此处再使用位置平移变换、尺度缩放变换等, 将尺骨与桡骨训练集分别扩增至 5 倍, 即尺骨每类 1000 张样本, 共 3000 张训练样本, 桡骨每类 750 张, 共 3000 张训练样本。

使用 4 层卷积、2 层全连接的基本网络结构, 进行对比试验, 结果见表 5 所示。

由上述实验结果, 可以看出数据增强对分类精度的提升效果明显。可见数据增强对于数据量不足的数据, 提升效果显著。下面在数据增强的样本上, 进行实验, 调整网络结构。

4.2.3. 网络结构

对于不同的任务, 不同的网络结构往往能够得到不同效果, 正如“没有免费的午餐”理论指出的, 不存在一个模型对于所有的任务都是最优的。所以针对特定的任务应尽量找出最适合该任务的网络模型。

以数据增强 5 倍后的数据集作为样本, 修改全连接层结点数、调整卷积核大小、增加卷积层数、使用 Dropout 来调整网络结构, 以表 6 中 9 种网络进行对比实验, 通过实验结果确定其中最合适该任务的网络模型结构。

Table 4. Classification accuracy under different sample distribution of basic network
表 4. 基本网络各样本分布下分类精度

	原始样本	均衡样本
尺骨分类精度	68.7%	70.7%
桡骨分类精度	83%	84%

Table 5. Classification accuracy before and after data augment
表 5. 基本网络数据增强前后分类精度

	原始样本	数据增强
尺骨分类精度	68.7%	78.7%
桡骨分类精度	83%	88%

Table 6. Different network structures
表 6. 不同网络结构

网络结构		网络
卷积层	全连接层	
	1024-256	1
32c7-64c5-64c3-32c3	1024-1024	2
	4096-1024	3
32c3-64c3-64c3-32c3	1024-1024	4
	4096-1024	5
32c3-64c3-64c3-128c3-128c3	1024-1024	6
	4096-1024	7
32c7-64c5-64c3-128c3-128c3	1024-1024	8
	4096-1024	9

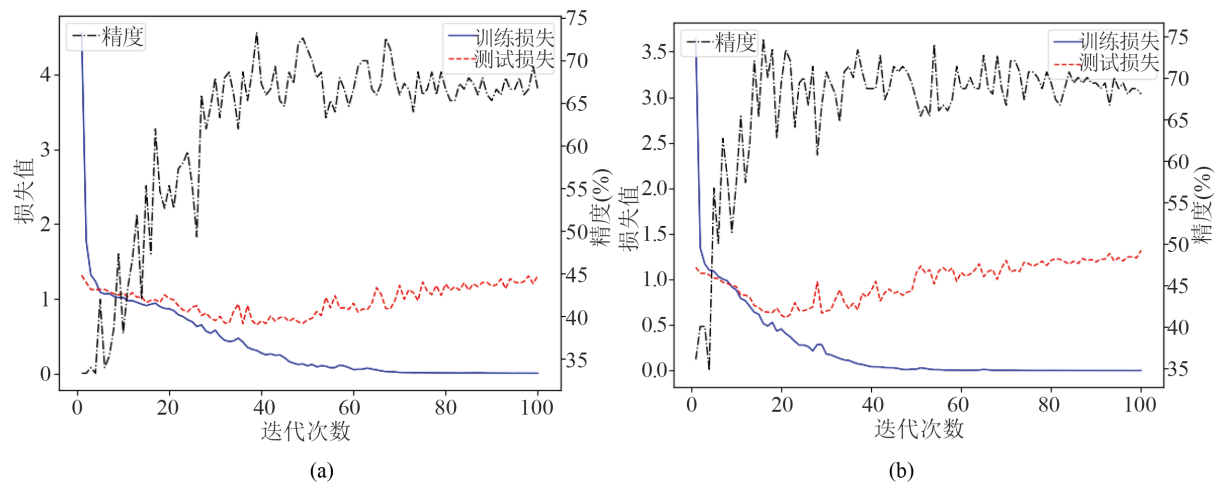


Figure 9. Comparison of convergence rates between imbalanced samples and balanced samples
图 9. 非均衡样本与均衡样本收敛速度比较

表 6 中网络结构分为卷积层与全连接层两列给出，卷积层以“32c7-64c5-64c3-32c3”格式给出，该例表示有 4 层卷积层，其中第一层卷积层“32c7”代表一个 7×7 的卷积层，卷积核数量为 32，步长均为 1，每个卷积层连接一个 2×2 的最大池化层。全连接层以“1024-1024”格式给出，该例表示两层全连接，每层结点数量为 1024 个。

在这 9 种网络结构下，分别训练尺骨骨龄阶段分类器与桡骨骨龄阶段分类器，训练过程中，统一使用 Adam 优化算法最小化多分类交叉熵损失函数，Adam 是一种自适应参数更新方法，包括 3 个参数值，分别为学习率和两个衰减率参数，取 0.0001 与 0.9、0.999。权重参数 W 使用 Xavier 均匀分布初始化，偏差 b 使用常数 0 初始化。分类效果如图 10 所示，(a)为尺骨分类情况，(b)为桡骨分类情况。

由图 10 可知，尺骨骨龄阶段分类在网络 8 上获得最好结果 0.88，桡骨骨龄阶段分类在网络 7 上获得最好结果 0.92。两者均在较深的网络下获得最优分类结果，但由于尺骨图片相对于原始尺寸有所放大，尺骨分类前两层使用了更大的卷积核，以获得更大的感受野下能获得更好的精度。在全连接层使用 Dropout 明显比未使用 Dropout 的网络获得更好的结果。

4.2.4. 网络参数优化

以数据增强 5 倍后的数据集作为样本，尺骨骨龄阶段分类器使用网络 8 结构，桡骨使用网络 7 结构分别使用 Adam 与 SGD 优化方法，及不同学习率参数，比较分类精度与网络收敛效果。图 11(a)为尺骨分类结果，图 11(b)为桡骨分类结果。从图 11 中可以看出使用 Adam 优化方法比 SGD 优化方法能获得更好的分类结果，使用 Dropout 比不使用 Dropout 能获得更好的分类结果，且 Adam 受学习率等参数影响更小，具有更强的适应性。

图 12(a)为相同优化方式下，是否使用 Dropout 情况下损失函数随迭代次数的收敛情况，从图中可以看出两者具有相同的收敛速度，但使用 Dropout 时，损失值能降到更低。且随着迭代次数的上升，没有使用 Dropout 时损失值上升明显，过拟合情况明显，而使用了 Dropout 后则有效的抑制了过拟合，网络获得了更好的收敛效果。图 12(b)为使用两种不同的优化方式下，测试精度随迭代次数变化情况，Adam 方法通常能获得更快的收敛速度与收敛效果。

综合各种网络结构、优化方式、学习率参数等实验结果最终尺骨与桡骨骨龄阶段分类结果如表 7 所示。

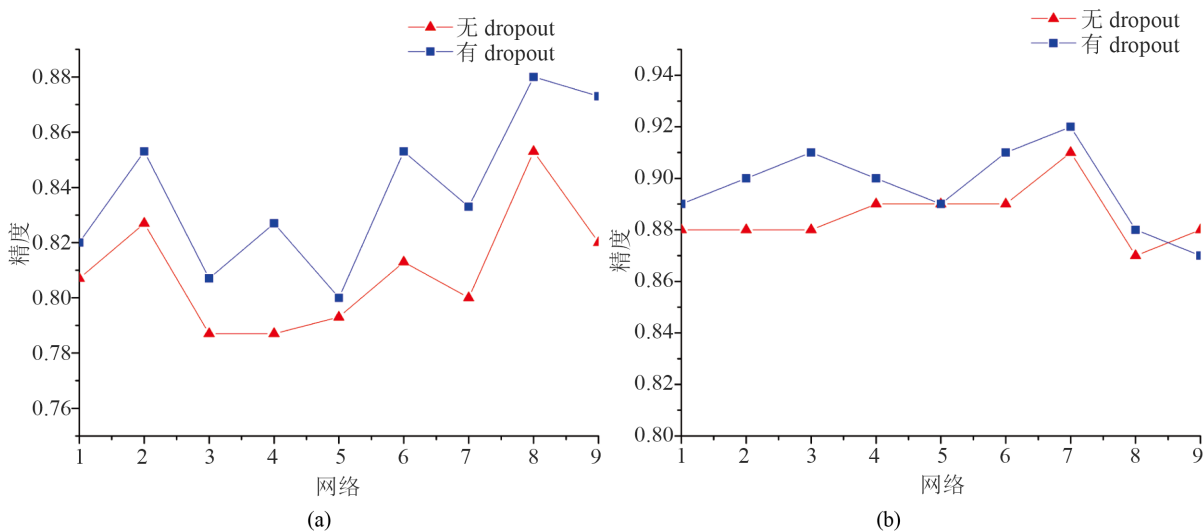


Figure 10. Curve: classification accuracy of ulna and radius under different network structures

图 10. 不同网络结构下尺骨与桡骨分类精度

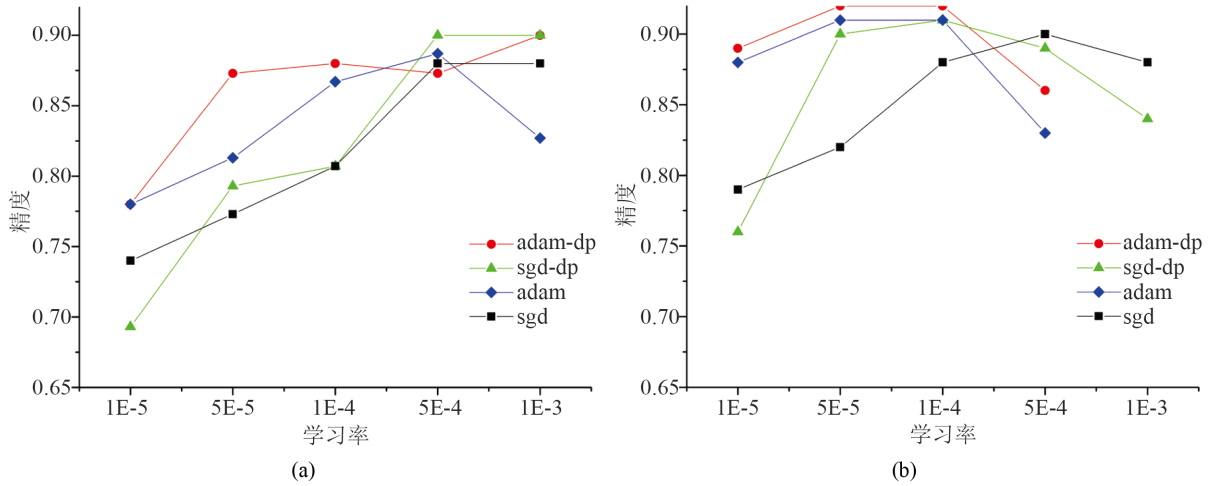


Figure 11. Curve: classification accuracy of ulna and radius under different learning rates and optimizer
图 11. 不同学习率与优化器下尺骨与桡骨分类精度

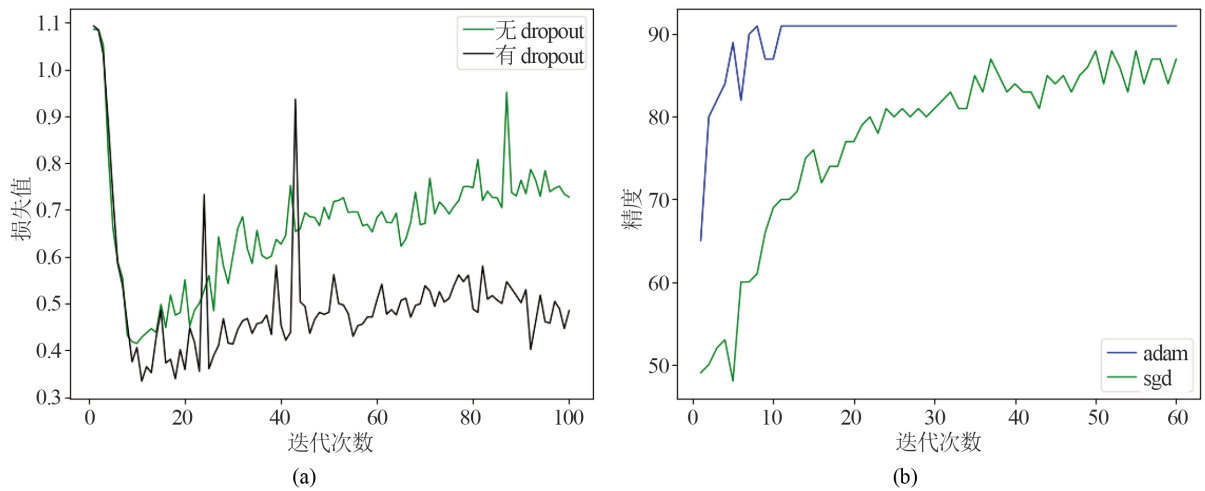


Figure 12. Curve: comparison of convergence effects
图 12. 收敛效果比较

Table 7. Classification accuracy of ulna and radius under optimal conditions
表 7. 尺骨与桡骨最优情况分类精度

	网络结构	优化方式	学习率	Dropout	分类精度
尺骨	网络 8	Adam	0.001	使用	90%
桡骨	网络 7	Adam	0.0001	使用	92%

5. 总结

由于骨骼成熟度分析对于指导临床管理至关重要，特别是理解其生长的高峰期与停止期。本文根据手部 X 光图片及对应尺骨与桡骨末端阶段标签构建了基于卷积神经网络的骨龄阶段自动识别模型，为医生提供参考，提高了识别效率，减少医生的工作量。本文主要工作总结有以下几个方面：

1) 在 Luk 等提出的骨骼成熟度分类架构基础上，结合临床应用实际，将骨骼成熟度重新进行阶段划分，数据预处理并构建相应的卷积神经网络模型，自动完成目标区域检测与分类即可可视化输出结果的全

过程。

2) 进行网络参数调优, 并运用数据增强、Dropout、样本均衡等技术提升网络性能, 优化网络模型。

最终, 通过卷积神经网络建模训练、参数优化、数据增强等技术的运用, 获得尺骨骨龄阶段识别 90% 的分类精度及桡骨骨龄阶段识别 92% 的分类精度。通过保存训练后的网络模型参数, 构建了骨骼发育阶段自动识别模型, 可实现输入一张手部 X 光片图片后, 自动完成检测尺骨与桡骨末端区域, 并进行发育阶段分类, 并可可视化输出结果的全过程。但由于时间、数据、能力所限, 还存在诸多不足, 需做进一步研究, 如对骨龄阶段的划分不够精细, 分类粒度较粗, 缺乏尺骨发育前期样本, 后续可考虑设计一个在阶段更精细划分下仍然能获得较好分类性能的分类器模型, 同时由于区域检测后仍有部分噪声, 可考虑精细分割出尺骨与桡骨末端后再进行分类。

参考文献 (References)

- [1] 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.
- [2] Luk, K.D.K., Saw, L.B., Grozman, S., *et al.* (2014) Assessment of Skeletal Maturity in Scoliosis Patients to Determine Clinical Management: A New Classification Scheme Using Distal Radius and Ulna Radiographs. *Spine Journal Official Journal of the North American Spine Society*, **14**, 315-325. <https://doi.org/10.1016/j.spinee.2013.10.045>
- [3] Roth, H., Lu, L., Liu, J., *et al.* (2015) Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation. *IEEE Transactions on Medical Imaging*, **35**, 1170-1181. <https://doi.org/10.1109/TMI.2015.2482920>
- [4] Anthimopoulos, M., Christodoulidis, S., Ebner, L., *et al.* (2016) Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, **35**, 1-1. <https://doi.org/10.1109/TMI.2016.2535865>
- [5] Dou, Q., Chen, H., Yu, L., *et al.* (2016) Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks. *IEEE Transactions on Medical Imaging*, **35**, 1182. <https://doi.org/10.1109/TMI.2016.2528129>
- [6] Greenspan, H., Ginneken, B.V. and Summers, R.M. (2016) Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*, **35**, 1153-1159. <https://doi.org/10.1109/TMI.2016.2553401>
- [7] Shin, H.C., Roth, H.R., Gao, M., *et al.* (2016) Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, **35**, 1285-1298. <https://doi.org/10.1109/TMI.2016.2528162>
- [8] Michael, D.J. and Nelson, A.C. (1989) HANDX: A Model-Based System for Automatic Segmentation of Bones from Digital Hand Radiographs. *IEEE Transactions on Medical Imaging*, **8**, 64-69. <https://doi.org/10.1109/42.20363>
- [9] Pietka, E., Gertych, A., Pospiech, S., *et al.* (2001) Computer-Assisted Bone Age Assessment: Image Preprocessing and Epiphyseal/Metaphyseal ROI Extraction. *IEEE Transactions on Medical Imaging*, **20**, 715-729. <https://doi.org/10.1109/42.938240>
- [10] Bocchi, L., Ferrara, F., Nicoletti, I., *et al.* (2003) An Artificial Neural Network Architecture for Skeletal Age Assessment. *International Conference on Image Processing*, Barcelona, 14-17 September 2003, 1-1077-80. <https://doi.org/10.1109/ICIP.2003.1247153>
- [11] Gertych, A., Zhang, A., Sayre, J., *et al.* (2007) Bone Age Assessment of Children Using a Digital Hand Atlas. *Computerized Medical Imaging and Graphics*, **31**, 322-331. <https://doi.org/10.1016/j.compmedimag.2007.02.012>
- [12] Lee, H., Tajmir, S., Lee, J., *et al.* (2017) Fully Automated Deep Learning System for Bone Age Assessment. *Journal of Digital Imaging*, **30**, 1-15. <https://doi.org/10.1007/s10278-017-9955-8>
- [13] 罗家燕, 卢瑞祥. 计算机化骨龄评价系统的研制[J]. 第四军医大学学报, 1998, 19(1): 115-116.
- [14] 董娜, 王瑶, 聂磊, 等. 基于 X 光图像的骨龄评估系统设计与实现[J]. 计算技术与自动化, 2010, 29(1): 67-71.
- [15] 王勇. 骨龄自动识别中的关键技术研究[D]: [硕士学位论文]. 宜昌: 三峡大学, 2011.
- [16] 赵娟. 基于 CHN 法的骨龄识别方法的研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2014.
- [17] 朱翔宇, 欧阳斌, 罗冬梅, 等. 基于中华 05 骨龄标准的桡骨骨龄等级计算机评定系统的设计[J]. 体育科学, 2016, 36(9): 59-65.
- [18] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Net-

- works. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, 3-6 December 2012, 1097-1105.
- [19] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.
- [20] Zeiler, M.D. and Fergus, R. (2014) Visualizing and Understanding Convolutional Networks. In: *European Conference on Computer Vision*, Springer, Cham, 818-833.
- [21] Girshick, R. (2015) Fast R-CNN. *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [22] Srivastava, N., Hinton, G.E., Krizhevsky, A., *et al.* (2014) Dropout: A Simple Way to Prevent Neural Networks from Over-Fitting. *Journal of Machine Learning Research*, **15**, 1929-1958.
- [23] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition.
- [24] Kingma, D. and Ba, J. (2014) Adam: A Method for Stochastic Optimization.
- [25] 杨明, 尹军梅, 吉根林. 不平衡数据分类方法综述[J]. 南京师范大学学报: 工程技术版, 2008, 8(4): 7-12.
- [26] Bastien, F., Bengio, Y., Bergeron, A., *et al.* (2010) Deep Self-Taught Learning for Handwritten Character Recognition.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-6753, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jisp@hanspub.org