

基于深度学习的青少年手腕骨骨龄评价

李睿¹, 张世杰², 黄奥云³, 陈虎¹

(1. 四川大学 计算机学院, 四川 成都 610065;

2. 四川大学 视觉合成图形图像技术国防重点学科实验室, 四川 成都 610065;

3. 四川川大智胜软件股份有限公司, 四川 成都 610045)

摘要:利用人工智能中的深度学习方法自动检测并评价西南地区青少年左手腕关节X线片的骨龄。在四川大学华西第二医院共收集2 426例1-18岁青少年左手腕X线片,利用YOLOv3框架和少部分数据进行标定、训练以检测X线片上传传统骨龄评价方法需要的区域,将关键区域截图并调整姿态组成新的图片。再利用caffe框架将扩展后的数据集分成训练集、验证集、测试集,以骨龄为标签对不同性别数据分别进行训练以获得男性和女性骨龄预测的模型,并计算误差在 ± 1 岁以内的准确率。选择caffe框架训练出来的最好模型,测试出测试集中 ± 1 岁的准确率为男性81.06%,女性85.08%。利用深度学习中简单的神经网络训练少量数据即可得到不错的骨龄评价准确率,表明了深度学习方法在西南地区青少年骨龄评价的可行性以及在数据增加和网络优化之后准确率存在的极大提升空间。

关键词:人工智能;深度学习;骨龄;放射学

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2020)01-0124-05

doi:10.3969/j.issn.1673-629X.2020.01.022

Bone Age Assessment of Hand Bone for Adolescents Based on Deep Learning

LI Rui¹, ZHANG Shi-jie², HUANG Ao-yun³, CHEN Hu¹

(1. School of Computer Science, Sichuan University, Chengdu 610065, China;

2. National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China;

3. Wisisoft Co., Ltd., Chengdu 610045, China)

Abstract: Deep learning method in artificial intelligence is used to automatically detect and evaluate the bone age of X-rays of wrist bones in young children from Southwest China. A total of 2 426 cases of left wrist X-rays of 1-18 years old are collected from West China Second Hospital of Sichuan University. The YOLOv3 framework and a small amount of data are used for calibration and training to detect the areas needed by traditional bone age assessment methods on the X-rays, and the key areas are captured and adjusted to form a new picture. Then the extended data set is divided into training set, verification set and test set by the caffe framework. With bone age as a label, the models for predicting bone age of men and women are obtained by training different sex data, and the accuracy of the error within ± 1 year is calculated. The best model trained by the caffe framework is selected to test the accuracy of ± 1 year old in the test set, which is 81.06% for males and 85.08% for females. A simple neural network in deep learning can be used to train a small amount of data to obtain a ideal accuracy rate of bone age evaluation, which indicates the feasibility of deep learning method in bone age evaluation of adolescents in southwest China and the great room for improvement of accuracy after data increase and network optimization.

Key words: artificial intelligence; deep learning; bone age; radiology

0 引言

机器学习^[1]是人工智能^[2]的一个分支,它通过计算机的一些算法从已有数据中学习特征参数,然后对具有类似特征的新样本做出识别或对未来的变化做出

预测。机器学习的发展从20世纪80年代末期开始大致经历了两次推进:浅层学习(shallow learning)和深度学习(deep learning)^[3]。浅层学习主要指支持向量机(support vector machine, SVM)^[4-5]、Boosting^[6]、最

收稿日期:2019-02-27

修回日期:2019-06-28

网络出版时间:2019-09-25

基金项目:国家重点研发计划(2016YFC0801100)

作者简介:李睿(1994-),男,硕士研究生,研究方向为计算机视觉;陈虎,讲师,硕士生导师,研究方向为模式识别、图像处理等。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190925.1523.058.html>

大熵方法^[7] (logistic regression, LR^[8]) 等经典的机器学习方法。而相较于只有极少隐层节点的浅层学习方法,具有更多隐层的深度学习方法还包含了更海量的训练数据,降低过拟合可能性的同时也可以学习到更有用的特征,从而提升了分类或预测的准确性。常用深度学习这个术语来指训练神经网络^[9] (一种由大量节点函数相互联结构成的运算模型) 的过程,通过输入数据在神经网络的训练下不断优化改变权重参数得出适用的模型,并利用神经网络模型实现各种功能应用。无人驾驶汽车的出现、人脸识别身份验证以及其他方面的诸多图像识别都来源于这项技术的应用,所以深度学习几乎成为人工智能的代名词。

在骨龄评价的过程中,青少年骨骼生长伴随着手部腕关节一系列的变化,因此医学上根据这些变化制定了标准的骨龄评分系统,需要通过拍摄个体的 X 线片,并根据其特征评价手腕骨的成熟度即骨龄,其中最常用的方法是 G-P 图谱法^[10] 和 TW2^[11] 评分法及中华 05 法^[12]。骨龄评价是用来判断青少年儿童生长发育情况、辅助运动员选拔,以及司法鉴定中确定年龄的重要方法,还对一些儿科内分泌疾病的诊断有很大帮助,具有重要的临床意义。然而,传统方法主要利用左手腕骨、掌骨、指骨骨化中心大小,形态结构相互关系来确定发育程度。其中涉及到人工观察骨骺等发育情况并对应图谱的评分以及等级划分,最后综合评分来确定 X 线片所属骨龄范围,方法步骤比较繁琐,耗时较长。而采用深度学习自动提取相关部位的特征^[13] 进行骨龄的识别及判断,十分快捷方便。所以这里提出一种简单的深度学习方法来训练医师评价的骨龄数据,使得通过深度学习得到的模型可以让计算机通过识别 X 线片来自动评价骨龄,并获得与医师骨龄评价间误差范围在±1 岁以内的评估结果。

在 RSNA(北美放射学会)举办的一次 X 线片骨龄预测大赛中,最好的深度学习模型使得计算机得出的骨龄与人工评价的骨龄误差在 6 个月左右。但是他们使用的是来自美国斯坦福儿童医院和科罗拉多儿童医院的公开数据集。为了得到适用于西南地区青少年儿童的骨龄评价模型,文中采集了来自四川大学华西第二医院的数据,并构建了一个简单的骨龄评价方法验证深度学习在西南地区青少年儿童(1-18 岁)骨龄评价的可行性和可靠性。

1 材料与方法

1.1 研究材料

该研究采集了四川大学华西第二医院 2017 年底至 2018 年期间拍摄的汉族青少年儿童左手腕 X 线片,并由医师评估了骨龄(为了与其他类似研究具有

可比性,该院医师评价骨龄采用了较为传统的 TW2 法),其中包含骨龄为 1-18 岁以及成年骨龄的 X 片。且该数据纳入的均为身体健康无影响骨骼生长的内分泌疾病病史人群,排除了图像显示骨骼存在畸形改变的数据,该研究符合有关的医学伦理及法律条款规定。数据中的 X 片被制作成了带标签的数据集,其中男性 1 242 例,女性 1 163 例,男性按照年龄分为 1 到 18 岁共 18 个类别,女性分为 1-16 岁以及 18 岁 17 个类别(其中 17 岁数据极少,暂时未做分类),除了三岁及以下年龄段数据,每一类数据量在 80 例左右。样本分布情况如表 1 所示。

表 1 不同性别、年龄数据分布

年龄/岁	男性/例	女性/例
0.0-1.0	10	22
1.0-2.0	29	42
2.0-3.0	74	76
3.0-4.0	77	78
4.0-5.0	78	79
5.0-6.0	80	79
6.0-7.0	80	78
7.0-8.0	80	80
8.0-9.0	80	78
9.0-10.0	80	80
10.0-11.0	80	78
11.0-12.0	80	79
12.0-13.0	79	80
13.0-14.0	80	80
14.0-15.0	65	67
15.0-16.0	78	7
16.0-17.0	32	0
17.0-18.0	80	80
总计	1 242	1 163

上述数据中,90% 为训练数据(其中训练集与验证集比例为 8 : 1),余下 10% 为测试集(不参与训练)。且以上数据均为排除了位于每个年龄段边界之后的数据(例如 15.1 岁或者 11.9 岁),只取 X.2-X.8 的年龄段数据来训练,以达到更好的分类效果。

1.2 方 法

1.2.1 预处理

在使用深度学习方法的过程中,数据预处理是所有后续步骤的重要基石。而该方案的第一步则是对目标检测的部分数据进行标注并训练,目标检测的准确性关乎后续 caffe^[14] (深度学习框架) 分类训练所需数据的完整性,在目标检测网络训练开始之前需要耗费一定的时间去标注数据。

(1) 目标检测。

由于该数据集采集质量较高,可以在不进行图像质量优化的情况下直接检测,文中的深度学习方法中只需要检出骨龄评价所需的关键部位,并排除不同姿态带来的影响及误差即可。由人工评价骨龄的先验知识可知,在骨龄评估中腕部和指关节起了决定性的作用,所以可以利用 YOLOv3^[15] (基于深度学习的目标检测框架)来自动定位手掌、关节以及腕部。其中数据标注如图 1 所示。

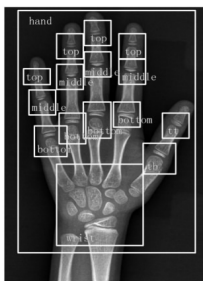


图 1 数据标注

每张手骨 X 线片被标注了 16 个框体,分为 hand、wrist、top、mid、bottom、tt、tb 七个类,在标定并训练目标检测网络模型之后,利用 YOLOv3 接口与训练的模型自动定位到 X 片上手掌各个部位,并利用 OpenCV^[16]单独截出需要的部位再通过仿射变换^[17]做相应的姿态调整(对空出的图片区域填充灰度值为 0 的像素)。

(2) 手掌姿态矫正。

利用 YOLOv3 定位的中指指关节和腕骨中心坐标确定倾斜角度,然后对图片做仿射变换旋转矫正姿态(见图 2)。旋转的角度公式为:

$$\theta = \arctan \frac{y_1 - y_2}{x_1 - x_2} \quad (1)$$

其中, θ 为旋转的角度; x_1 和 y_1 为中指指关节的中心横纵坐标; x_2 和 y_2 为腕骨中心横纵坐标。手掌姿态矫正后有利于后续 YOLOv3 框架对手掌各部位的二次定位。

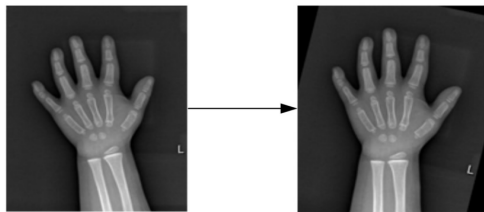


图 2 手掌姿态旋转

(3) 图片组合。

由 TW2 评分法可知,小指、中指、大拇指以及腕骨在骨龄评价中占有较大比重,故该训练采用了组合第一、三、五掌指关节以及腕骨图片的方式来进行训练(分别对以上部位进行姿态矫正后再组合,以减小误

差)。在此之前尝试过多种关节组合方案,选择了效果最好的一种,而且加入更多的部位组合起来可以提高准确率,但是由于 YOLOv3 并不是 100% 检测到所有部位,所以这里只选取了四个可检测到的部位来组合成一张图片训练。同样利用 OpenCV 库新建一个空的图像容器 Mat 对象,将读取的四个部位截图拷贝到容器 Mat 对应的四个感兴趣区域即完成组合,如图 3 所示。

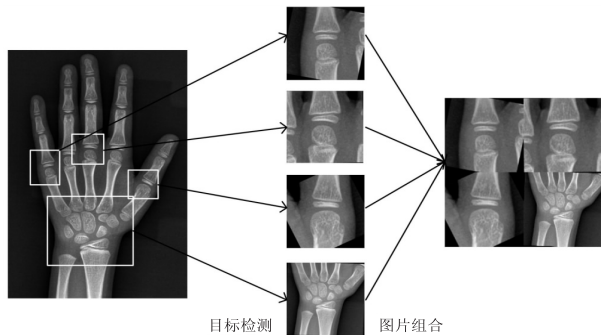


图 3 图片组合

(4) 数据扩展。

由于某些年龄段数据较少,不利于深度学习分类网络的训练,所以需要利用图像增强算法^[18]改变原图的亮度及对比度扩展出更多的训练数据。图像增强算法通过 OpenCV 访问图像像素改变其对比度亮度生成新的扩展图像,具体调整公式为:

$$g(i, j) = a * f(i, j) + b \quad (2)$$

其中, i 和 j 表示 i 行 j 列, $f(i, j)$ 为源图像像素, $g(i, j)$ 为输出图像像素, a 为控制对比度参数, b 为控制亮度的参数。

扩展效果如图 4 所示。

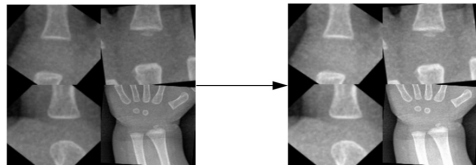


图 4 图像扩展

1.2.2 深度学习网络及框架

(1) 网络及框架: 目标检测的框架为 YOLOv3 (YOLO 的最新版本), 此框架为纯 C 语言编写, 在效率和可移植性上都相对较高。此外该框架内置了作者自训练的类似于残差网络的 darknet-53 训练网络, 该方案中只需按照官方的训练方法进行训练无需调整额外的参数。训练时随机标注了数据集中 760 例 X 线片, 由于该网络为端到端的训练网络, 直接以 X 线片原始尺寸作为输入数据即可, 按照 5 : 1 的比例划分训练集和验证集迭代 9 000 次训练。骨龄评价的框架为 caffe, 此框架可移植性较高。改进了 AlexNet^[19] (2012 年 ImageNet 竞赛冠军获得者 Hinton 和他的学生设计)

作为骨龄评价训练的网络,训练过程中微调了网络中的参数,修改了其中的 Relu 激活函数(在人工神经网络的神经元上运行的函数,负责将神经元的输入映射到输出端)为 PReLU(PReLU 是针对 ReLU 的一个改进型,在负数区域内,PReLU 有一个很小的斜率),验证集准确率提高了两个百分点。此外,上述深度学习的框架都包含已封装的 C++接口,十分便于集成到自动化的桌面应用中。改进后的骨龄评价训练网络的结构如图 5 所示。

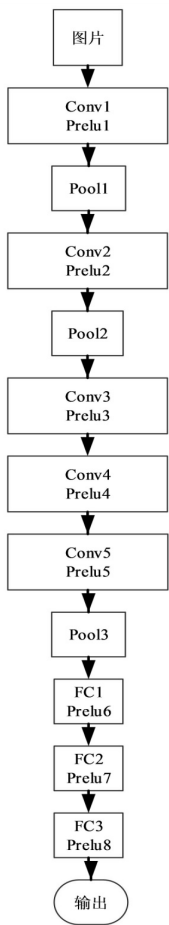


图 5 骨龄评价训练网络结构

如图 5 所示,该网络有五个带有 Prelu 激活函数的卷积层(Conv1-Conv5),卷积层可利用卷积核根据步长在输入图像的像素点上滑动,同时与该像素点以及相邻的像素点做内积,通过这样的卷积操作即可提取图像的特征图,三个最大池化层(Pool1-Pool3)(池化层通过下采样降低特征图维度,减少训练过拟合的可能性),以及三个带有 Prelu 的全连接层(FC1-FC3)(全连接层即可综合出卷积和池化层的可分类信息)。其中输入图片为训练集中 500×500 像素尺寸的灰度图,输出为 18 个分类(代表 0-18 岁中的每一个年龄段)。输入数据集归一化到 500×500 像素尺寸的灰度图,随机裁剪尺寸为 480×480 像素输入神经网络进行训练,选择了男性迭代 10 000 次训练的权重模型及女

性迭代 30 000 次训练的权重模型,分别在测试集上达到了不错的准确率。

(2)骨龄评价网络超参数(在机器学习的上下文中,超参数是在开始学习过程之前设置值的参数)设置:骨龄评价训练网络初始学习率设置为 0.001,在训练的过程中,由于输入图片数据尺寸较大,设置较大的初始学习率会使训练过程发生梯度爆炸^[20]而无法继续,所以训练时将初始学习率(base_lr=0.01)减小了一个数量级。使迭代 4 000 次下降一次学习率(gamma 值为 0.9),梯度更新权重为 0.9,权重参数以 0.000 5 的衰退值下降。

1.2.3 评估方法

将预测结果按照年龄段分类,每一岁为一类,按性别分别统计测试集所有数据误差在±1 岁以内的准确率,并计算平均绝对误差(MAE)。公式如下:

$$MAE(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i|$$

(3)

其中, y_i 为每一例输入样本的实际骨龄; \hat{y} 为每一例输入样本的预测骨龄; n_{samples} 为总的样本数据量。

平均绝对误差即为每一例数据由深度学习模型评估出的骨龄与人工评价的骨龄的误差取绝对值并累加求和最后除以数据的总数量求得的值,可以作为评价模型偏差的一个标准。

2 结 果

目标检测:选择了迭代 9 000 次训练的一个模型, IOU(区域覆盖率,即在原始图像上预测出的目标框体与人工在原始图像上标注的目标框体重叠区域面积占两个框体合并面积的比例)平均为 80%。公式如下:

$$IOU = \frac{\text{area}(c) \cap \text{area}(g)}{\text{area}(c) \cup \text{area}(g)}$$

(4)

其中, area(c) 表示预测出的候选框(candidate bound)面积; area(g) 表示原标记框(ground truth bound)面积。

目标框体漏检率(未预测出原始图像上实际应该自动标注的框体占框体总数的比率)为 3%,使用 YOLOv3 的 C++接口截图定位手腕骨 X 线片各个关键部位结果如图 6 所示(由于部分标签重叠,故分开展示)。



图 6 YOLOv3 自动检测结果

骨龄评价:男性数据迭代训练了 10 000 次,修改 caffe 源码封装的 C++测试接口以进行批量测试,测试集上 ± 1 岁的准确率为 81.06%,女性数据迭代训练了 30 000 次,用同样的方法在女性数据测试集上测得准确率为 85.08%。具体如表 2 所示。

表 2 不同性别训练结果分布

性别	测试集数量/例	± 1 岁准确率/%	MAE/月
男性	132	81.06	10.6
女性	114	85.08	11.1

3 讨 论

由于该实验数据集仅仅为国外公开数据集数量的六分之一,所以平均绝对误差还没达到最低,但 1 岁以内的误差准确性几乎能够达到目前人工水平且高于胡婷鸿^[21]等人所做的类似研究。在国内,针对西南地区 1-18 岁青少年儿童的手腕骨 X 线片骨龄数据集尚不完整,也少有针对性做深度学习评价骨龄的研究,实验结果证明仅仅利用矫正了姿态的第一、三、五掌指关节以及腕骨就可以达到 ± 1 岁误差内 80% 以上的准确率。

与其他方法相比,该研究具有以下优势:(1)用计算机代替了人工阅片,且该研究中的两种深度学习框架都具有较好的可移植性,应用性极强;(2)利用了 YOLOv3 作为图片预处理的框架进行目标检测、姿态调整操作,效果提升明显;(3)利用了覆盖西南地区 1-18 岁青少年儿童的数据,数据覆盖年龄段较为完整,且对研究西南片区的青少年儿童骨龄发育有一定帮助^[22];(4)仅仅使用了两千多例数据以及改进的 AlexNet 网络,部署较为简单,效果较好。

4 结束语

针对该实验可以预见人工智能对放射学及医学的一些传统方法将起到良好的改进与促进作用^[23],并辅助放射科医师更好更快的工作;除了该方案中使用的矫正了姿态的第一、三、五掌指关节以及腕骨区域图片,加入更多评价骨龄时关注的区域到训练数据中也可以更全面地促进计算机学习评价骨龄所需特征并进一步提高准确度。考虑到全部区域的加入会增加图像姿态不同带来的误差暂时没有比较,减小这种差异需要在组合全部部位之前进行图像姿态的矫正归一化。对于该文章中使用的深度学习模型,如果继续增加数据集以及改进深度学习网络为更多层更深的网络并使用更好的设备进行训练模型,会得到 MAE 更低、准确率更高的深度学习模型。

由于人工的评价骨龄在耗时上高于全自动的软件

评价,而且使用深度学习的方法可以在数据继续扩充的情况下不断优化训练模型并提升准确度,可以说促进模型优化的很大一部分因素在于数据量的多少;此外一般的计算机支持就可以达到远远高于人工评价的速度甚至更高的准确度,所以目前看来基于深度学习的骨龄评价很有应用价值和发展前景。另一方面,放射科医师对未来医学人工智能的发展仍然起到了指导性的作用,在人工智能、深度学习、计算机视觉技术不断发展的时代,计算机技术将会成为辅助广大医师工作的强大力量。

参考文献:

[1] MITCHELL T, BUCHANAN B, DEJONG G, et al. Machine learning[M]. [s. l.]: McGraw-Hill, 2003.

[2] RUSSELL S J, NORVIG P. Artificial intelligence: a modern approach[M]. 北京:人民邮电出版社, 2002.

[3] 郭丽丽, 丁世飞. 深度学习研究进展[J]. 计算机科学, 2015, 42(5): 28-33.

[4] 张浩然, 韩正之, 李昌刚. 支持向量机[J]. 计算机科学, 2002, 29(12): 135-137.

[5] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and their Applications, 2002, 13(4): 18-28.

[6] FREUND Y, SCHAPIRE R E. A desicion-theoretic generalization of on-line learning and an application to boosting [C]//European conference on computational learning theory. [s. l.]: Springer, 1995: 23-37.

[7] SHORE J E, JOHNSON R W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy[J]. IEEE Transactions on Information Theory, 1980, 26(1): 26-37.

[8] CUCCHIARA A. Applied logistic regression[J]. Technometrics, 1992, 44(1): 81-82.

[9] LANGE N. Pattern recognition and neural networks[M]//Pattern recognition and neural networks. [s. l.]: Cambridge University Press, 1996: 233-234.

[10] 张绍岩, 王姿欢, 蒋竞雄. 骨龄评价方法的发展及应用[J]. 中国妇幼卫生杂志, 2012, 3(6): 345-348.

[11] TANNER J M, WHITEHOUSE R H, et al. Assessment of skeletal maturity and prediction of adult height (TW2 method)[M]. 2nd ed. London: Academic Press, 1983.

[12] 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.

[13] 陈 珍, 夏靖波, 柏 骏, 等. 基于进化深度学习的特征提取算法[J]. 计算机科学, 2015, 42(11): 288-292.

[14] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]//ACM international conference on multimedia. [s. l.]: ACM, 2014: