

基于注意力与标签相关性的胸部 X 光片疾病分类^①



王焯楠, 程远志, 史 操, 许灿辉

(青岛科技大学 信息科学技术学院, 青岛 266061)
通信作者: 程远志, E-mail: yzcheng2007@163.com

摘 要: 针对传统的胸部辅助诊断系统在胸部 X 光片疾病分类方面图像特征提取效果差、平均准确率低等问题, 提出了一个注意力机制和标签相关性结合的多层次分类网络. 网络的训练分为两个阶段, 在阶段 1 为了提高网络特征提取能力, 引入注意力机制并构建一个双分支特征提取网络, 实现综合特征的提取, 在阶段 2 考虑到多标签分类中标签之间相关性等问题, 利用图卷积神经网络对标签相关关系进行建模, 并与阶段 1 的特征提取结果进行结合, 以实现胸部 X 光片疾病的多标签分类任务. 实验结果表明, 本方法在 ChestX-ray14 数据集上各类疾病的加权平均 AUC 达到 0.827, 有助于辅助医生进行胸部疾病的诊断, 有一定的临床应用价值.

关键词: 胸部疾病; 多标签分类; 标签相关性; 深度学习

引用格式: 王焯楠,程远志,史操,许灿辉.基于注意力与标签相关性的胸部 X 光片疾病分类.计算机系统应用,2023,32(4):104-111. <http://www.c-s-a.org.cn/1003-3254/9028.html>

Chest X-ray Disease Classification Based on Attentional and Label Correlation

WANG Ye-Nan, CHENG Yuan-Zhi, SHI Cao, XU Can-Hui

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Traditional chest-aided diagnosis systems have poor image feature extraction effects and low average accuracy in disease classification based on chest X-ray images. In view of these problems, a multi-level classification network that combines an attention mechanism and label correlation is proposed. The training of the network is divided into two stages. In stage one, in order to improve the feature extraction capability of the network, an attention mechanism is introduced, and a two-branch feature extraction network is constructed to realize the extraction of comprehensive features. In stage two, according to the correlation between labels and other issues in multi-label classification, a graph convolutional neural network is used to model the label correlation, which is then combined with the feature extraction results obtained in stage one, so as to achieve the multi-label classification task of diseases based on chest X-ray images. The experimental results show that the weighted average AUC of diseases by the proposed method on the ChestX-ray14 dataset reaches 0.827. Therefore, the method can assist doctors in diagnosing chest diseases and has certain clinical application value.

Key words: chest disease; multi-label classification; label correlation; deep learning

胸腔是人体疾病的高发部位, 胸部病变十分常见且种类繁多, 严重地威胁着人体的生命健康. 研究表明, 每年有数百万人被诊断出患有胸部疾病. 就肺癌而言, 我国肺癌患者的 5 年生存率不足 20%, 由于多数患者

被检查出患有肺癌时已处于癌症晚期, 难以治愈. 因此, 对于胸部疾病, 早期的诊断和筛查是治疗的关键.

目前, 胸部疾病的检查手段主要有胸部 X 光片和胸部 CT 图像. 相比于 CT 图像, X 光片设备更加普及,

① 基金项目: 国家自然科学基金 (61806107, 61702135)

收稿时间: 2022-09-07; 修改时间: 2022-09-30; 采用时间: 2022-10-19; csa 在线出版时间: 2022-12-23

CNKI 网络首发时间: 2022-12-27

费用更加低廉,同时其对人体的辐射也较低,故X光片更适用于疾病的初步筛查.但X光片需要专业的放射科医生对其进行观察诊断,在我国,放射科医生培养周期长、工作强度大、人手短缺,且随着医生的工作时间增加,误诊、漏诊概率也会大大增加^[1].即使是优秀的放射科医生也会在工作过程中不可避免地出现严重的临床失误^[2].因此,研究一种自动、准确的胸部疾病分类方法对于辅助临床诊断而言有着极高的应用价值.

传统的分类算法主要是通过边缘检测、降维方法对胸片疾病进行分类.其中,Htike等人^[3]提出一种首先用拉普拉斯滤波器对胸片进行处理,然后通过图像共生矩阵提取特征,最后使用旋转森林进行分类的方法.但传统的分类方法往往存在着准确率低、泛化性差的问题,无法应用到实际工作中.近年来,基于深度学习的方法在胸部病变分类上取得了不错的成果,随着ChestX-ray14数据集的公开,Wang等人^[4]评估了目前存在的包括VGGNet-16、ResNet50等几个经典模型在胸部疾病分类中的效果,发现基于ResNet50的分类效果最好.同时,有研究人员将注意力机制融入图像分类模型中.Mnih等人^[5]提出了一种基于注意力机制的递归神经网络模型;张驰名等人^[6]提出使用压缩激励模块实现引入通道注意力机制,来提高网络的细粒度分类能力;Guan等人^[7]提出了一种基于分类残差注意学习的胸片图像分类方法.实验表明,在模型中引入注意力机制能使得模型在图像分类中取得更好的表现.尽管很多网络都取得了不错的效果,但目前的现有算法仍存在漏诊率和误诊率较高等问题,解决这些问题的关键在于,需要训练出效果更好的特征提取器.与此同时,对于多标签分类任务,标签之间共同出现的频率或是某一类别出现的条件概率都可以作为先验知识来辅助提高分类结果,然而在目前的研究中,很少有研究者会对标签的相关关系进行建模,导致很多先验知识没有充分利用.

针对这些问题,本文构建了一个基于注意力机制和标签相关性的多层次分类网络.网络分为两个阶段,分别进行综合特征学习和标签相关性学习.为提高网络特征提取能力,在综合特征学习部分,引入注意力机制,在ViT(vision-Transformer)模型^[8]的基础上构建一个双分支特征提取网络,分别进行细节特征学习和形状特征学习.在标签相关性学习部分,使用图卷积神经网络^[9]对标签相关性进行训练.将两阶段结果进行结合,以实现多标签胸部X光片的准确分类.

1 胸片辅助诊断模型

1.1 模型架构

本文所使用的模型结构如图1所示.将原始X光片经过预处理后得到的图片数据和类别编码数据分别输入到多层次网络中进行训练.模型主要包含两个阶段,阶段1为综合特征学习阶段,阶段2为标签相关性学习阶段.在阶段1中主要通过调整不同大小的感受野尺寸对输入的图像数据分别进行特征学习,然后将特征进行融合从而得到图像的综合特征;在阶段2中,将类别编码送入标签相关性学习网络中以得到带有特征相关性的矩阵信息,并将图片数据送入训练好的阶段1模型中以得到带有综合特征的特征图像,同时对阶段1的模型进行微调.最终得到完整的训练模型,图1中实线表示阶段1实现过程,虚线表示阶段2实现过程.在这里,我们之所以将训练过程分成两个阶段进行而不是联合训练主要有以下两个原因:(1)两个阶段的主要任务是不同的,阶段2的输入一部分来源于阶段1的结果,同时将阶段1的少部分参数进行微调,可以说是将模型结果进行再一次的提高;(2)神经网络的结果会受到批次大小(batchsize)的影响,两个阶段的输入批次大小是不同的,都是根据实验表现单独设置的.

1.2 基于注意力机制的综合特征学习

特征提取对于多标签分类任务而言尤为重要,在胸部疾病分类问题中,不同类别的疾病特征在大小、形状、纹理等多个尺度下均有着不同的表现^[10],单一尺度下的特征提取会存在难以获取到多类别的综合特征的问题.本文针对这一问题,采用双分支特征提取网络,通过调整感受野尺度的不同,分别学习图片的细节特征和形状特征,同时引入注意力机制起到增大重要特征权重的效果.在分别学习到图片的细节特征和形状特征后,最终通过特征融合的方式获取输入图像的综合特征.

由于传统的卷积神经网络存在着感受野固定且局限的问题,近年来,基于注意力机制的网络在图像领域取得了不错的效果.注意力机制可以取代卷积工作更好的提取到图片中的关键信息和关键特征^[11],同时可以有效地从模型中查看注意力分布,增加了模型的可解释性.其中注意力机制在图像领域使用最典型的模型为ViT模型,本文也选取该模型作为特征提取网络的主要结构.基于注意力机制的综合特征学习模型结构如图2所示.

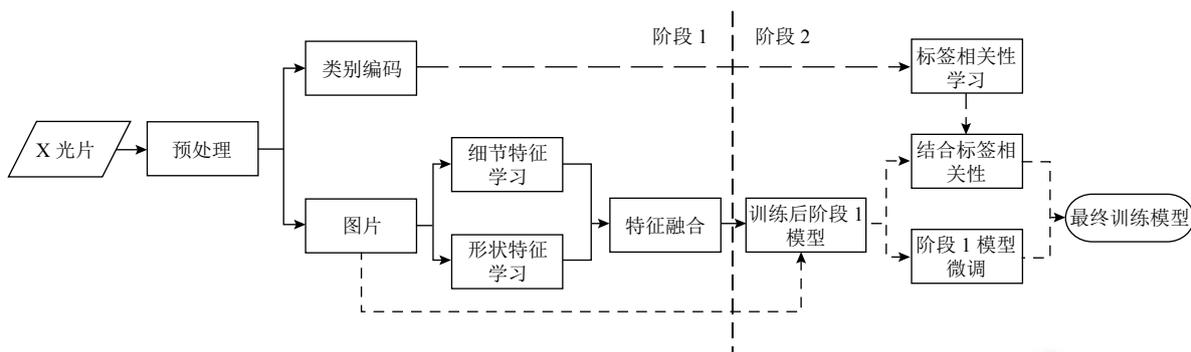


图1 本文网络结构

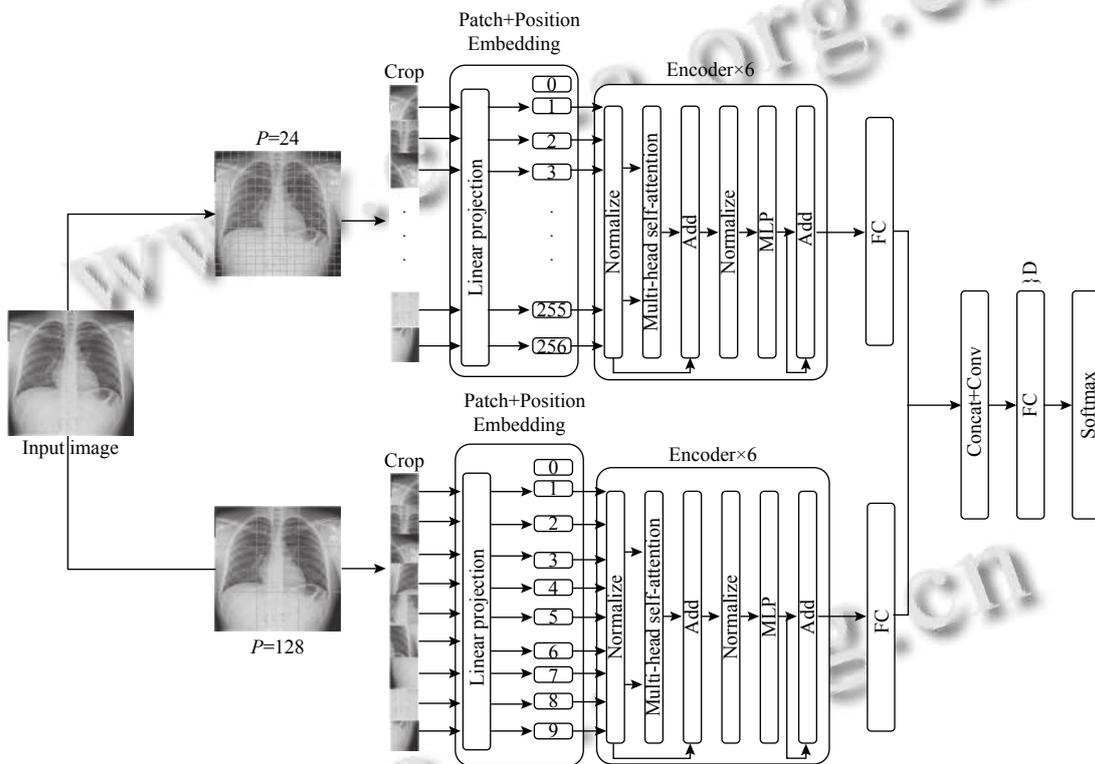


图2 基于注意力机制的综合特征学习模型结构

由图2可知,输入图像送入模型后,有上下两个分支分别进行特征学习,两个分支的区别在于切片patches 的尺寸P的不同.当P取值为24时注意力机制的感受野尺寸很小,模型更加专注于对细节特征的学习,而当P取值为128时感受野尺寸变大,模型更关注对整体形状的特征学习.由于注意力机制的可解释性,通过对权重矩阵的迭代运算,将不同P的取值下的权值矩阵与原图像相乘,得到模型的注意力表现如图3所示.

接下来介绍网络的具体计算.根据不同的P值,对图像进行裁剪,用(H,W,C)依次表示图像的长、宽和通道数,将图像裁剪成N个大小为P×P的patches,其中:

$$N = \frac{H}{P} \times \frac{W}{P} \tag{1}$$

然后将每个 patch 通过 Linear 线性层拉平成一维向量,对每个 patch 进行位置编码,并在初始位置加入代表图片类别的向量.将编码后的向量送入到多个经过串联后的 Encoder 层中,每个 Encoder 层内包含了两个归一化层和两个叠加层外还有一个多头自注意力 (multi-head self-attention, MSA) 层^[12] 和一个多层感知器 (MLP).

注意力机制的主要工作原理是通过计算每个 patch 之间的相关性得分,来达到对关键 patch 保留或增大权

重,对不重要的 patch 减轻权重的效果.对每一个输入 patch,都可以通过线性变换来得到 query (Q)、key (K)、value (V) 这 3 个向量,通过计算 Q 和 K 之间的相似度作为重要性权重,对得到的权重用 *Softmax* 函数进行归一化,然后将结果与 V 相乘得到加权后的 V .自注意力机制 (self-attention)^[13] 则是仅关注自身,此时 $Q = XW_Q$, $K = XW_K$, $V = XW_V$, X 为输入的 patch 所拉平的 1 维向量,即注意力的相关性计算仅在输入特征内部元素之间,不需要其他额外信息.多头自注意力机制则是在不同的子空间进行自注意力的计算,然后将每个空间的结果连到一起.通过多头自注意力机制的计算可以有效防止模型过拟合并且可以更好地捕捉每个输入之间的联系以及更好地获取关键特征.多头自注意力机制公式为:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

$$head_i = Attention(XW_i^Q, XW_i^K, XW_i^V) \quad (3)$$

$$MSA(X) = Concat(head_1, \dots, head_h)W^O \quad (4)$$

其中, $\sqrt{d_k}$ 为缩放因子,用于避免点积带来的方差影响, d_k 的值由 patch 的大小和 h 决定,其中:

$$d_k = \frac{P \times P \times C}{h} \quad (5)$$

式 (3) 中 ($i = 1, \dots, h$), W_i^Q 、 W_i^K 、 W_i^V 、 W^O 是可学习训练的矩阵, X 为输入的 patch 所拉平的 1 维向量, h 为子空间数量.

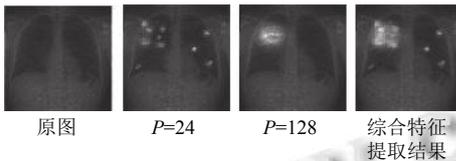


图3 在特定取值下注意力特征表现示意图

在经过一个 MSA 层后通过 MLP 层来对输出进行调整,两个分支都会得到一个维度相同的全连接层 (FC 层) 作为特征提取结果,通过 concat 结构进行连接后,再经过一层卷积操作得到维度为 D 的向量,以此作为特征融合后所得到的综合特征.阶段一网络的预训练过程单独来看是一个完整的模型拟合过程,该过程的样本标签为经过 one-hot 编码后的图像疾病类别标签.在阶段一,获取到综合特征表示,该综合特征大小为 $D \times 1$,故可以直接使用 *Softmax* 做激活函数,即可得到样本在所有类别下的概率,概率和为 1.当两阶段一起结合起来进行训练时,删去阶段 1 中的 *Softmax* 输出

层,将阶段 1 结果直接参与后续运算.

1.3 基于图卷积神经网络的标签相关性学习

对于多标签分类任务而言,有些目标是同时出现的,尤其是胸部疾病分类,胸部疾病具有多发性的特点,并发症是医生诊断的重要考虑因素.本文对不同疾病类别之间的依赖性进行建模,通过引入标签相关性来提高模型的性能.本文使用图卷积神经网络 (GCN) 来对标签相关性进行学习,阶段 2 部分标签相关性模型训练以及与阶段 1 模型结果的结合过程如图 4 所示.

图卷积神经网络的基本思想是通过在各个节点之间的传播来更新节点表示,以达到对标签相关性训练的效果.我们用 $H^l \in \mathbb{R}^{n \times d}$ 来表示某一层的节点特征,用 $A \in \mathbb{R}^{n \times n}$ 来表示互相关矩阵,则下一层的节点特征为 $H^{l+1} \in \mathbb{R}^{n \times d'}$,其中 n 表示节点也就是类别个数, d 表示该层节点维度, d' 表示下一层节点维度,GCN 的主要目的是学习每一层之间的变换函数 $f(\cdot)$,即:

$$H^{l+1} = f(H^l, A) \quad (6)$$

在对该层进行卷积操作后,该变换函数即可表示为:

$$H^{l+1} = h(\hat{A}H^lG^l) \quad (7)$$

其中, G^l 为转移矩阵是通过训练得到的其维度为 $d \times d'$, \hat{A} 为互相关矩阵归一化后矩阵, $h(\cdot)$ 为非线性激活函数.

在 GCN 中,互相关矩阵 A 的通常都是预先定义好的,在本文中使用的条件概率矩阵作为互相关矩阵,条件概率即用 $P(L_j|L_i)$ 表示在疾病 L_i 出现的情况下疾病 L_j 出现的概率.我们首先求得所有标签对之间的对应关系 $M \in \mathbb{R}^{C \times C}$,其中 C 表示类别个数, M_{ij} 表示疾病 L_i 和疾病 L_j 同时出现的情况,然后通过共生矩阵即可求得条件概率矩阵:

$$P_i = \frac{M_i}{N_i} \quad (8)$$

来表示疾病 L_i 出现的情况下每一类疾病出现的概率,其中 P_i 和 M_i 表示一组维度为 j 的向量, N_i 则表示疾病 L_i 在训练集中出现的概率,矩阵 P_i 中:

$$P_{ij} = P(L_j|L_i) \quad (9)$$

在图 4 中, C 表示参与训练的类别,一共有 15 类,包括了 14 种疾病和 1 类未患病情况, d 和 D 分别表示输入向量维度和输出向量维度,为方便展示,我们使用数字 1-15 来对每个节点进行标号.由图 4 可知,首先对待分类的类别进行词嵌入向量的训练,这里我们使用自然语言处理中经典模型 Word2Vec 来得到每个类别的词嵌入向量表示.这里之所以使用包含语义信息的向量表示而不是使用 one-hot 来对类别标签进行编

码, 主要是因为, 这些类别之间在语义上有一定的隐式依赖性, 有助于训练. 经过训练会得到一个由每个类别的词向量组成的矩阵 $Z \in \mathbb{R}^{C \times d}$ 作为 GCN 网络的输入, 而 GCN 网络的初始拓扑结构则是根据互相关矩阵 A 生成的. 在经过多层 GCN 网络训练后, 得到了包含了更新后每个节点特征的输出矩阵 $Z' \in \mathbb{R}^{C \times D}$. 矩阵 Z' 的每

一行代表该行所对应的节点, 其中节点信息已经被更新为其与相邻节点的关系, 节点的维度为 D , 与阶段 1 所得的特征提取网络结果 x 的维度相同. 然后将得到的输出矩阵 Z' 与特征提取网络得到的特征表示 x 进行点积运算, 即可得到带有标签相关性的预测结果 $\hat{y} = Z'x$, 使用 Sigmoid 函数即可得到相应的预测类别.

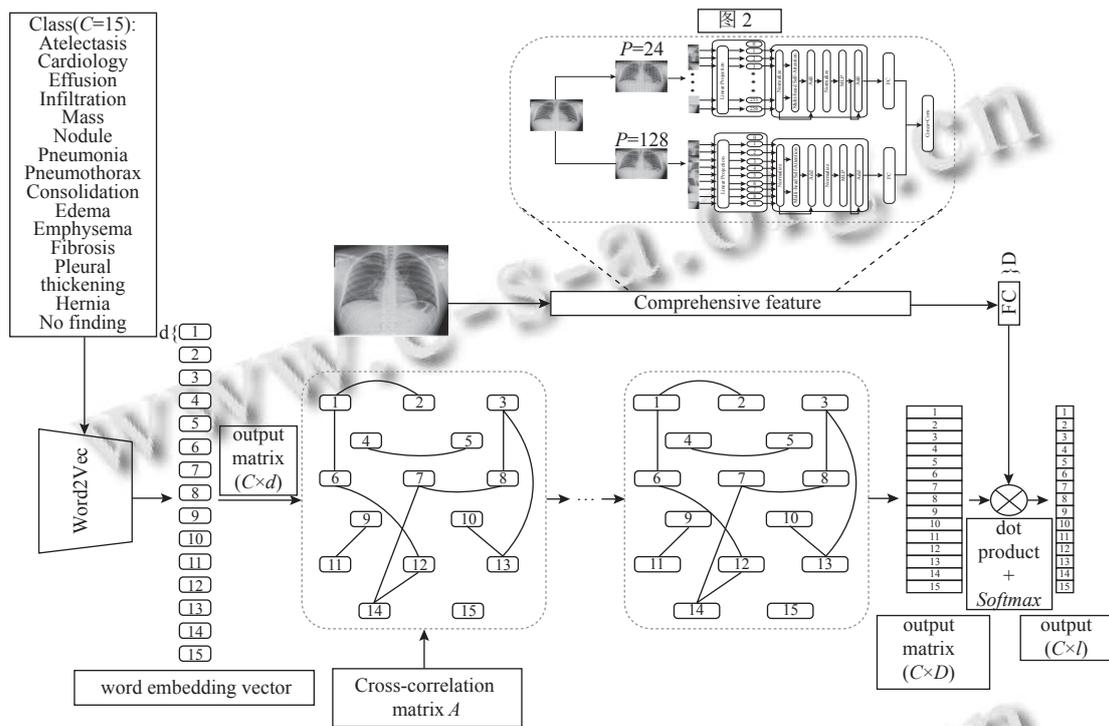


图 4 基于图卷积神经网络的标签相关性学习模型结构

在阶段 2 的训练过程中, 不仅将节点信息更新成由其他相关节点的共同表示, 同时还会根据结果对阶段 1 中的最后的 FC 层参数进行微调.

对于胸部疾病来说, 标签相关关系即为对疾病并发病的探索. 这些依赖关系更像是一种先验经验来辅助特征提取网络的分类准确性. 标签相关性是对样本集标签进行量化的一种方式, 在本文中, 我们将待分类的 15 类疾病类别 (含未患病情况) 进行标签化处理, 然后通过 GCN 网络可训练出不同类别间的依赖关系, 这种依赖关系以矩阵形式输出, 而特征提取网络提取出的结果同样是一个相同维度的矩阵, 经激活函数后, 两个矩阵均为概率矩阵以分别表示图像患各疾病的概率和患该类疾病下其他疾病出现概率, 将上述两种结果做点积运算就意味着在原有特征提取结果的基础上加入了标签之间的关联关系. 由于这是对结果之间的运算, 所以两个阶段的训练过程还是相对独立的, 在第 2 个阶段微调部

分也仅是对特征融合后全连接层部分参数进行微调, 对于前面的 ViT 网络特征提取部分, 是不再修改的.

1.4 模型损失函数

假定样本的真实标签为 $y \in \mathbb{R}^C$, 每一类疾病 $y^i \in \{0, 1\} | 1 \leq i \leq C$, 0 表示没有患病, 1 表示患病, 一共有 C 个类别. 模型的损失函数定义如下:

$$L = \sum_{C=1}^C y^C \log(\sigma(\hat{y}^C)) + (1 - y^C) \log(1 - \sigma(\hat{y}^C)) \quad (10)$$

其中, $\sigma(\cdot)$ 表示 Sigmoid 函数.

2 实验

2.1 实验数据及预处理

本文使用的数据集为 ChestX-ray14 数据集, 该数据集是由美国国立卫生研究院整理并公开的一个大型胸部多标签数据集, 其中包含了 30805 个病例的 112120 张胸部 X 光片图像, 且原始图像的大小为 1024×1024

像素. 数据集中每一个图像都被标注出该患者是否患病, 其中未患病的样本个数为 60 361 张其余均为患有 一种或多种疾病的样本. 若患病则标注出所患的某 一种或多种疾病类别, 其中疾病共 14 类均为常见的胸部 疾病. 这些疾病包括: 疝气、胸膜增厚、纤维化、肺气 肿、水肿、肺实变、气胸、肺炎、肺结节、肿块、浸 润、积液、心脏肿大和肺不张, 该数据集中每一类疾 病具体的数量分布情况和患病率如表 1 所示. 由表 1 可知, 由于疾病之间样本数量的差异较大导致训练 时类别的不平衡, 这将大大增加多标签分类的难度.

表 1 ChestX-ray14 数据集各类疾病样本分布情况

标签	样本数量	患病率 (%)	标签	样本数量	患病率 (%)
疝气	227	0.20	气胸	5 302	4.73
胸膜增厚	3 385	3.02	肺炎	1 431	1.28
纤维化	1 686	1.50	肺结节	6 331	5.65
肺气肿	2 516	2.24	肿块	5 782	5.16
水肿	2 303	2.05	浸润	19 894	17.74
肺实变	4 667	4.16	积液	13 317	11.88
心脏肿大	2 776	2.48	肺不张	11 559	10.31

考虑到数据集中不同类别之间样本数量的不平衡, 本文采用 7:2:1 的划分方式, 对每一类样本单独划分, 这样做既保证了训练、验证和测试过程中各类疾病均有样本参与, 同时又保证了同一图像在 3 部分无交叉. 胸片图像的原始大小为 1024×1024 像素的灰度图, 为了减少计算量且符合网络的输入尺寸, 我们将图片转换为 384×384 像素大小的 3 通道图像, 并对像素归一化将像素值限制在 0–255 之间.

2.2 实验环境及实验细节

本实验的实验环境为 Ubuntu 16.04 操作系统, 硬件为 1 块 NVIDIA RTX3090 显卡, 编程语言为 Python 3.7, 使用的深度学习框架为 PyTorch.

本实验总共训练 100 轮次, 批次大小设置为 64, 初始学习率为 0.001, 采用 StepLR 的学习率衰减策略, 每经过 10 轮训练进行一次衰减, 衰减系数为 0.5, 使用的 Adam 优化器做梯度下降算法.

2.3 实验评价指标

为方便与其他算法进行比较, 同时为体现出网络的整体性能, 本文采用受试者操作特征曲线 (receiver operating characteristic, ROC) 来表现网络对每一类疾病的分类效果, 由于仅使用 ROC 曲线进行对比很难看出不同分类器之间的优劣, 故使用每一类疾病 ROC 曲线下的面积 (area under ROC curve, AUC) 的值来表现

网络的优劣. ROC 曲线所表示的是所有分类结果在不同阈值下真阳性和假阳性之间的博弈, 其横坐标为假阳性率 (false positive rate, FPR), 表示所有负样本中分类器错判为正样本的概率, 纵坐标为真阳性率 (true positive rate, TPR), 表示所有正样本中分类器正确判断为正样本的概率, FPR 和 TPR 的具体计算公式为:

$$FPR = \frac{FP}{FP+TN} \quad (11)$$

$$TPR = \frac{TP}{TP+FN} \quad (12)$$

其中, FP 为假阳例, TN 为真阴例, TP 为真阳例, FN 为假阴例. 一般来说, AUC 的值在 0.5–1 之间, AUC 的值越大则曲线越靠近左上角, 表示算法的分类性能越好. 由于类别样本比例不同, 故本文使用 14 个类别下加权平均 AUC 的值来对网络进行整体评价, 并与其他模型进行对比.

2.4 实验结果和分析

2.4.1 消融实验结果分析

为检验网络中每个部分的具体效果, 本文共进行了 4 组消融实验, 采用加权平均 AUC 的值作为评价指标, 实验结果见表 2. 其中第 1 组为较经典的 ResNet50^[14] 作为骨干网络直接进行特征提取后实现多标签分类的结果, 第 2 组为使用单尺度下 ViT 网络进行特征提取的结果, 前两组主要是用于展示注意力机制对多标签分类结果的提升, 第 3 组为使用单尺度下 ViT 网络进行特征提取并加入标签相关性训练后得到的结果, 第 4 组为多尺度下 ViT 网络进行特征提取并加入标签相关性训练后得到的结果, 其中前两组只有一个训练阶段, 后两组在训练时有两个训练阶段. 由表 2 可以看出, 有两步操作对结果有着显著提高, 一个是引入注意力机制, 另一个是引入标签相关性. 对于多尺度下的特征提取, 相对于单尺度, 由于获取了更多特征信息, 对有些类别的分类效果有明显提升, 但对有些类别分类效果提升不大, 但网络整体效果有一定提升, 本文网络最终得到了加权平均 AUC 值为 82.7% 的结果.

2.4.2 对比实验 1 结果分析

通过对消融实验的结果可以证明, 加入注意力机制后的特征提取相比传统卷积方式更具有优势, 同时, 双分支的下的特征提取结构比单分支下取得了更好的效果. 两个分支的设计分别针对图像表现较小和较大类型的疾病, 我们通过设置 $P=24$ 来学习那些在图像上表现范围较小的疾病, 也就是学习整个图像的细节特

征, 我们通过设置 $P = 128$ 来学习那些在图像上表现范围较大的疾病, 也就是学习整个图像的形状特征. 下面通过几组对比实验来确定两个分支 P 的取值. 根据输入图像的尺寸、输入图像尺寸与 patches 尺寸比例以及疾病在该分辨率下的尺寸, 将 P 的取值分为 $P_s = \{12, 24, 48\}$ 和 $P_l = \{96, 128, 192\}$, 分别表示适应细节特征学习 P 和适应形状特征学习 P 的取值范围. 将分支取值两两组合进行实验, 得到表 3. 由表 3 可以看出当 $P_s = 24$ 且当 $P_l = 128$ 时实验效果最好. 我们在图 3 给出了在该取值下注意力特征表现示意图, 以及双分支结构下表现示意图, 以此来证明双分支结构对不同尺寸疾病的适应能力和综合特征提取效果.

2.4.3 对比实验 2 结果分析

为了展示网络的整体性能, 本文绘制了网络在每一个类别下的 ROC 曲线, 并计算出该类别 AUC 的值, 结果如图 5 所示. 由图可知, 每类疾病的曲线都接近于图像的左上角, 表明网络的整体分类性能良好. 同时为了体现网络在实验结果准确性上的提高, 本文仍采加权平均 AUC 的值作为评价指标, 与同一数据集下, 当前比较新的其他网络进行对比, 对比结果如图 6, 表 4 所示.

表 2 消融实验结果

疾病种类	实验1	实验2	实验3	实验4
肺不张	0.733	0.767	0.799	0.821
心脏肿大	0.848	0.870	0.900	0.915
积液	0.816	0.823	0.856	0.878
浸润	0.660	0.729	0.738	0.757
肿块	0.737	0.801	0.831	0.854
结节	0.778	0.778	0.818	0.834
肺炎	0.690	0.725	0.730	0.738
气胸	0.799	0.856	0.885	0.899
肺实变	0.727	0.743	0.752	0.777
水肿	0.806	0.835	0.850	0.860
肺气肿	0.842	0.895	0.921	0.944
纤维化	0.747	0.822	0.826	0.828
胸膜增厚	0.724	0.755	0.806	0.816
疝气	0.834	0.901	0.922	0.924
加权平均值	0.745	0.784	0.779	0.827

表 3 对比实验 1 结果

P_s	P_l	加权平均AUC
12	96	0.816
	128	0.820
	192	0.818
24	96	0.822
	128	0.827
	192	0.801
48	96	0.811
	128	0.820
	192	0.818

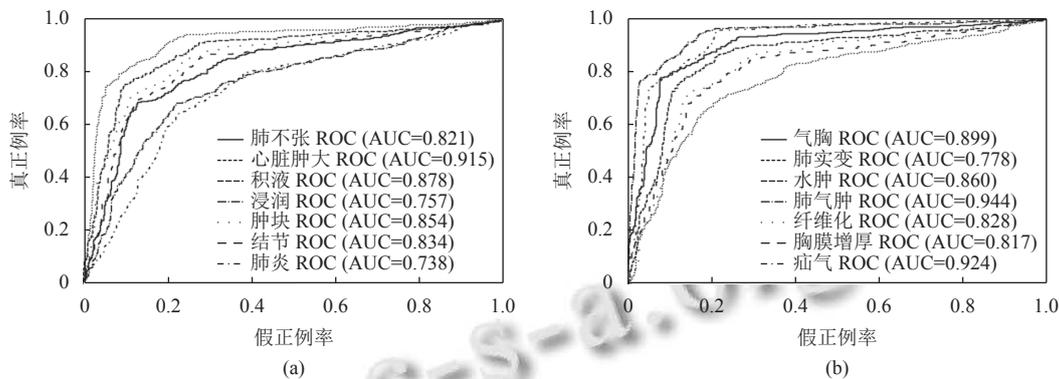


图 5 各类别 ROC 曲线和 AUC 值

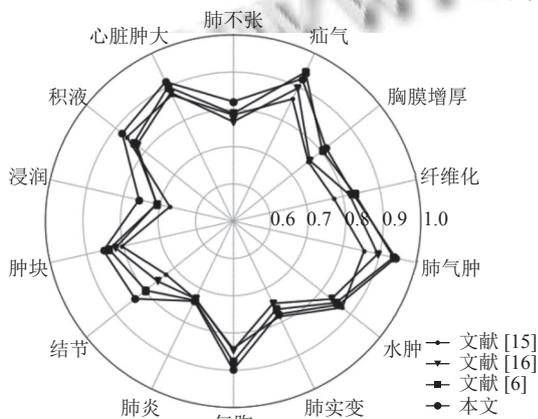


图 6 对比实验各类疾病 AUC 值对比

由表 3 可知, 在绝大多数的疾病类别下本文所提方法都取得了最好的结果, 同时网络在所有类别下的整体表现取得了不错的结果. 文献 [15] 所采用的是密集连接网络作为网络骨干结构, 同时引入了压缩激励模块来实现通道注意力机制的效果; 文献 [16] 所采用的也是密集连接网络的主要结构, 同时引入了位置信息, 将高分辨率下的图像和空间信息结合起来用于分类; 文献 [6] 所采用的是使用 ResNet 网络作为主干结构并在特征提取过程中同样加入了通道注意力机制, 同时, 在特征提取后对特征图进行了全局平均池化和全局最大池化的相结合的操作更大程度的获取特征信

息,相对于前两种方法,该方法提升效果明显.但这些方法均为考虑标签之间的相关性,而本文方法相对于这3种方法均取得了更好地表现.由表3可知,疝气、纤维化的结果不如文献[6]且肺炎的提升效果不高,这是因为该3类疾病的样本数量过少导致网络难以获得足够的疾病特征,几个对比方法都结果都相近.因此综合来看,本文所提出的方法在胸部多标签疾病分类任务上取得了不错的结果,最终达到了各类别平均AUC为82.7%的结果.

表4 对比实验2结果

疾病种类	文献[15]	文献[16]	文献[6]	本文
肺不张	0.785	0.767	0.789	0.821
心脏肿大	0.876	0.883	0.900	0.915
积液	0.862	0.828	0.838	0.878
浸润	0.673	0.709	0.708	0.757
肿块	0.804	0.821	0.839	0.854
结节	0.729	0.758	0.798	0.834
肺炎	0.732	0.731	0.736	0.738
气胸	0.842	0.846	0.877	0.899
肺实变	0.784	0.745	0.763	0.777
水肿	0.872	0.835	0.851	0.860
肺气肿	0.858	0.895	0.938	0.944
纤维化	0.775	0.818	0.832	0.828
胸膜增厚	0.756	0.761	0.804	0.816
疝气	0.864	0.896	0.944	0.924
加权平均值	0.777	0.780	0.797	0.827

3 结语

针对胸部X光片的多标签分类任务,本文提出了一个基于注意力机制和标签相关性的多层次分类网络,网络分为两个阶段,分别进行综合特征学习和标签相关性学习.在阶段1中主要通过ViT模型的基础上构建一个双分支特征提取网络,分别进行细节特征学习和形状特征学习,通过引入注意力机制来更好地获取图片的关键信息,捕捉关键特征;在阶段2中,将类别编码输入送入标签相关性学习网络中以得到带有特征相关性的矩阵信息,同时对阶段1的模型进行微调,最终得到完整的训练模型.本文方法相对于其他多标签分类方法,引入了注意力机制和多标签之间的标签相关性,在绝大多数类别下取得了更高的AUC值,对于医生的辅助诊断有一定的帮助和实际应用价值.

参考文献

- 黄欣,方钰,顾梦丹.基于卷积神经网络的X线胸片疾病分类研究.系统仿真学报,2020,32(6):1188-1194.[doi:10.16182/j.issn1004731x.joss.18-0712]
- 刘露,杨培亮,孙巍巍,等.深度置信网络对孤立性肺结节良恶性的分类.哈尔滨理工大学学报,2018,23(3):9-15.

- [doi:10.15938/j.jhust.2018.03.002]
- Htike ZZ, Naing WYN, Win SL, et al. Computer-aided diagnosis of pulmonary nodules from chest X-rays using rotation forest. Proceedings of the 2014 International Conference on Computer and Communication Engineering. Kuala Lumpur: IEEE, 2014. 96-99.
- Wang XS, Peng YF, Lu L, et al. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 3462-3471.
- Mnih V, Heess N, Graves A. Recurrent models of visual attention. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2204-2212.
- 张驰名,王庆凤,刘志勤,等.基于深度学习的胸部常见病变诊断方法.计算机工程,2020,46(7):306-311,320.[doi:10.19678/j.issn.1000-3428.0055204]
- Guan QJ, Huang YP. Multi-label chest X-ray image classification via category-wise residual attention learning. Pattern Recognition Letters, 2020, 130: 259-266. [doi:10.1016/j.patrec.2018.10.027]
- Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929, 2020.
- Wu ZH, Pan SR, Chen FW, et al. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24. [doi:10.1109/TNNLS.2020.2978386]
- Fei LK, Lu GM, Jia W, et al. Feature extraction methods for palmprint recognition: A survey and evaluation. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 49(2): 346-363. [doi:10.1109/TSMC.2018.2795609]
- Wang HY, Jia HZ, Lu L, et al. Thorax-Net: An attention regularized deep neural network for classification of thoracic diseases on chest radiography. IEEE Journal of Biomedical and Health Informatics, 2020, 24(2): 475-485. [doi:10.1109/JBHI.2019.2928369]
- Tan HC, Liu XP, Yin BC, et al. MHSA-Net: Multi-head self-attention network for occluded person re-identification. arXiv:2008.04015, 2020.
- Yu AW, Dohan D, Luong MT, et al. QANet: Combining local convolution with global self-attention for reading comprehension. arXiv:1804.09541, 2018.
- He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770-778.
- 张智睿,李镡,关欣.密集挤压激励网络的多标签胸部X光片疾病分类.中国图象图形学报,2020,25(10):2238-2248.[doi:10.11834/jig.200232]
- Gündel S, Grbic S, Georgescu B, et al. Learning to recognize abnormalities in chest X-rays with location-aware dense networks. Proceedings of the 23rd Iberoamerican Congress on Pattern Recognition. Madrid: Springer, 2018. 757-765.

(校对责编:牛欣悦)