

基于卷积神经网络的胸部 X 光图像分类算法

金荷淇^{1,2} 盖绍彦^{1,2} 达飞鹏^{1,2}
JIN Heqi GAI Shaoyan DA Feipeng

摘要

肺部疾病已成为危害人类身体健康的重要因素，胸部 X 光片是医疗实践中广泛应用的肺部疾病检测工具，在协助临床诊断中发挥着重要的作用。针对上述问题设计出一种胸部 X 光片疾病分类算法，应用于医疗场景下辅助医生进行诊断。研究方案在肺炎图像分类领域的权威数据集 Chest X-ray 14 上进行实验，数据集涉及 14 种不同的肺部疾病标签。研究工作基于深度学习算法，使用 ConvNeXt 卷积神经网络为主干网络，设计了该网络的多尺度特征融合模块，以及全局上下文注意力模块，用以提高分类算法的准确性。算法的评价指标使用 AUC (area under curve)，其值越接近 1 则算法效果越好，仅使用 ConvNeXt 主干网络的 AUC 值为 0.836 8，加入多尺度特征融合模块的 AUC 值为 0.844 0，加入全局上下文注意力模块的 AUC 值为 0.839 9，同时加入两个模块的结果为 0.849 3。改进算法的性能优于该领域的常用算法 DenseNet121 (0.830 1)，CheXNet (0.841 4)，MXT (0.830 8)，并且在 4 类病理现象的判断上取得了最优的水平。这两项改进都可提高模型性能，也可共同作用进一步提高模型分类能力。改进算法在综合能力和专项能力之间取得一个平衡，在综合分类性能优秀的同时，对于专项病症的分类能力也取得领先的成绩。

关键词

多标签分类；胸部 X 光片；肺部疾病；人工智能；辅助诊断

doi: 10.3969/j.issn.1672-9528.2024.01.006

0 引言

肺部疾病已成为现代社会中人类身体健康的严重威胁。对于肺部疾病的筛查，常用的医学手段有 CT 扫描和胸部 X 光片 (CXR) 图像诊断，由于成本的问题，CT 扫描仪在多数医疗中心无法获得，因此在许多情况下，CXR 图像成为帮助临床医生做出决策的最具时间和成本效益的工具。然而，由于放射科医生经验和主观方面的差异，CXR 图像的识别结果可能出现一定程度的误差，因此亟需一套准确的肺炎图像计算机辅助诊断算法，帮助医生进行判断，从而降低误诊率。如何准确快速地对 CXR 图像进行分类，从而判断出是否出现肺部异常以及何种异常，是当前肺炎医疗图像辅助诊断算法中迫切需要解决的问题。

在深度学习任务中，数据集不可或缺。近年来，随着美国国立卫生研究院 (national institutes of health, NIH) 发布了公共胸片数据集 Chest X-ray14^[1]，深度学习算法在胸片辅助诊断方面取得了快速发展，目前已有诸多学者如 Wollek 等

人^[2]、Sheu R K 等人^[3]、Chen 等人^[4]围绕该数据集展开研究。目前的深度学习 CXR 图像分类算法主要分为两大类：一类是卷积神经网络算法，另一类是 transformer^[5] 系列的算法。

在卷积神经网络算法中，Seibold C 等人^[6]基于经典网络 ResNet^[7]，提出了一种利用当前弱注释数据进行训练的算法，取得了与直接标签监督相持平的效果。Signoroni A 等人^[8]设计了一个端到端的深度学习架构 BS-Net，采用了弱监督式的学习策略来对 CXR 图像进行多区域的评分，模型准确性优于单一的人类评价方案。Yang 等人^[9]针对有噪声的数据设计了一种噪声标签检测方法，并且用一个三重网络的集成学习框架从主干网络中学习组合特征。Mukesh Mann 等人^[10]设计了一个张量处理单元 (TPU)，使训练后的模型能够准确有效地区分不同肺病。Chen 等人^[11]重新设计了网络的损失函数，经实验验证，在 CXR 图像分类任务中其效果优于原加权二元交叉熵损失。何迪等人^[12]优化 Faster-RCNN 模型，提高模型的识别诊断精度。

在基于 transformer 的算法中，MXT^[13]通过自注意力机制捕捉长短范围的视觉信息，从而提高模型精度。PVT^[14]构建了金字塔式的 transformer 主干，可作为 CXR 分类任务中一个优秀的 backbone 网络。邢素霞等人^[15]通过迁移学习的

1. 东南大学自动化学院 江苏南京 210096
2. 东南大学复杂工程系统测量与控制教育部重点实验室
江苏南京 210096

方法,微调模型结构,提升诊断的准确率和效率。Sun 等人^[16]设计了一种新的 transformer 结构,利用其独特的查询结构来获取图像的全局和局部信息以及标签之间的相关性,同时提出了一个新的损失函数,帮助模型获取 CXR 图像之间的相关性。Ukwuoma 等人^[17]提出了一种混合深度学习方法,基于多层感知的自注意力机制构建 transformer,融合卷积神经网络构建混合工作流,在 CXR 多标签分类中取得了较高性能。CheXViT^[18]通过将 CXR 分类模型 CheXNet^[19]和 Vision transformer^[20]结合起来,构成混合分类器,从而将归纳偏差和长程特征依赖性结合起来,使模型性能得到提高。

两种主流算法在取得长足进展的同时又各自有其不足,多数卷积神经网络在 CXR 分类精度上逊色于 transformer 系列算法,而 transformer 类算法由于模型过于庞大,不适宜在医疗辅助场景下使用。而且无论是卷积算法或 transformer 算法,还是专家医生,对于 CXR 图像的解读都是十分困难的,分析的准确度都难以达到很高的水准。目前普遍适用的方法虽然取得了较好的结果,但其模型性能提升已趋于饱和,而 ConvNeXt 网络^[21]的出现刷新了卷积网络的性能上限。因此,本文设计了一种基于 ConvNeXt 卷积神经网络,结合多尺度特征融合以及全局上下文注意力模块的分类算法,该算法利用轻量级的 ConvNeXt 网络主干,融合了从低级到高级的 4 级特征,再使用全局上下文注意力机制将融合后的特征图进行加权组合,从而充分利用各级语义特征,保留多级特征图之间的内部关系,提升了 CXR 图像分类的准确性。

1 材料和方法

1.1 实验平台

实验平台参数如表 1 所示。

表 1 实验平台参数

Hardware	Description
CPU	Intel® Xeon® Silver 4210 CPU, 64 GB ROM
GPU	NVIDIA GeForce RTX 2080
System	Windows 10
Video memory	8 GB
Language framework	Python 3.7 + PyTorch 1.8.1 + Anaconda

1.2 数据集

本文算法在 Chest X-ray14 数据集上训练并测试,该数据集是美国国立卫生研究院(NIH)发布的公共胸片数据集。Chest X-ray14 数据集包含 30 805 名患者的 112 120 张正面视图胸部 X 射线图像,如图 1 所示,以及 14 类常见的疾病图像标签,每个图像可以对应多个标签,CXR 图像被随机分为训练、验证和测试集,比例为 0.7:0.1:0.2。14 类标签包括 atelectasis(肺不张)、cardiomegaly(心肥大)、effusion(积

液)、infiltration(浸润)、mass(肿块)、nodule(结节)、pneumonia(肺炎)、pneumothorax(气胸)、consolidation(实变)、edema(水肿)、emphysema(肺气肿)、fibrosis(纤维变性)、pleural thickening(胸膜增厚)、hernia(疝气)。各类标签的数量如表 2 所示。



图 1 CXR 图像实例

表 2 Chest X-ray14 数据集中的疾病列表

Pathology	Total positives
atelectasis	11 559
cardiomegaly	2776
effusion	13 317
infiltration	19 894
mass	5782
nodule	6331
pneumonia	1431
pneumothorax	5302
consolidation	4667
edema	2303
emphysema	2516
fibrosis	1686
pleural thickening	3385
hernia	227

1.3 实验过程

实验中使用随机的 78 484 张 CXR 图像作为训练集,首先对训练集进行数据增强,使用随机翻转、镜像、平移和裁剪的方式对训练集进行操作,然后将图像作为输入训练 ConvNeXt 网络,训练 epoch 设置为 60, batchsize 设置为 48,优化器使用 adam^[22],初始学习率设置为 2e-4,在训练过程中使用余弦退火不断调整学习率。结合两个改进模块的操作与 ConvNeXt 主干类似,训练 epoch 设置为 40,其他参数保持不变。

本文损失函数使用多标签的二分类交叉熵损失(BCELoss),对于包含 N 个样本的 batch 数据 $D(x,y)$,每个样本有 $M=14$ 个标签,计算公式为:

$$l = \frac{1}{N} \sum_{n=1}^N l_n \quad (1)$$

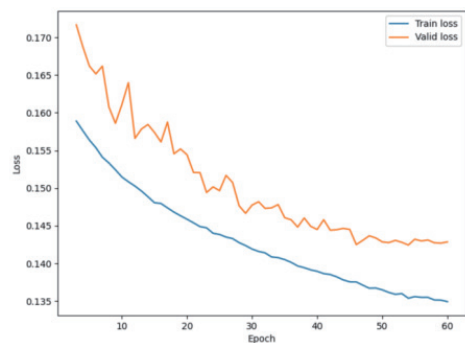
式中: l_n 为第 n 个样本对应的 l , 计算公式为:

$$l_n = \frac{1}{M} \sum_{i=1}^M l_n^i \quad (2)$$

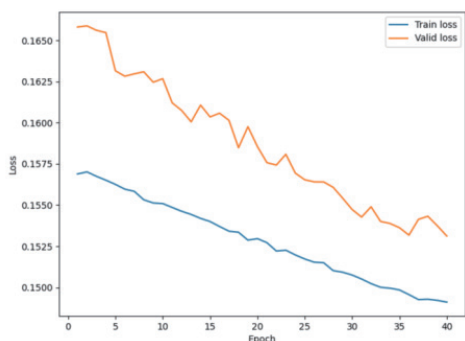
$$l_n^i = -\omega_i \left[y_n^i \cdot \log x_n^i + (1 - y_n^i) \cdot \log (1 - x_n^i) \right] \quad (3)$$

式中： ω_i 为超参数，是 loss 的计算加权，当标签间的样本不均衡时，为出现次数较少的标签赋予更大的权重。

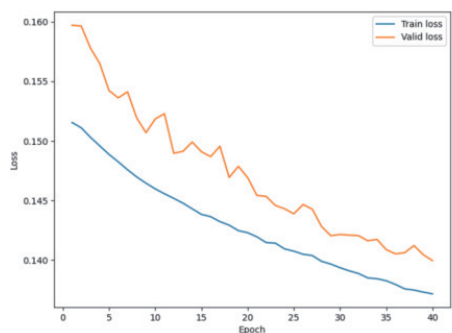
四种网络的训练集和验证集损失曲线如图 2 所示，14 类标签各自的 gradcam 激活图如图 3 所示。



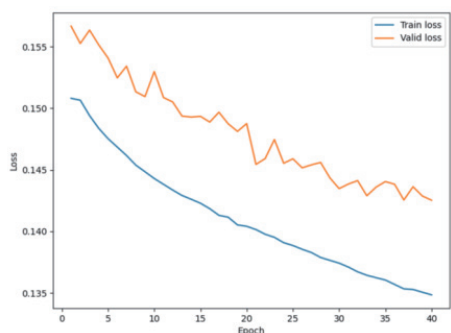
(a) ConvNeXt



(b) ConvNeXt+MsFF



(c) ConvNeXt+GCA



(d) ConvNeXt+MsFF+GCA

图 2 四种网络的训练集、验证集损失曲线

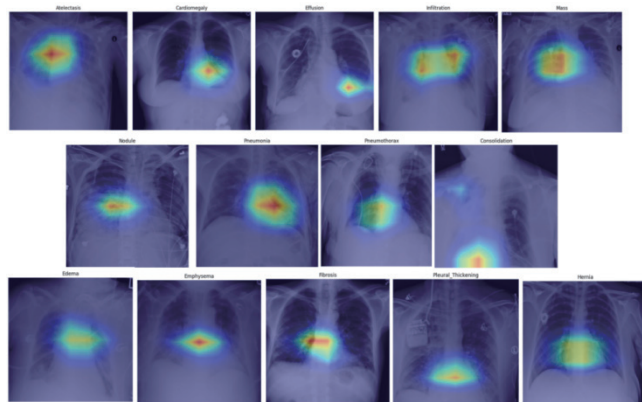


图 3 各标签的 gradcam 激活图

1.4 本文方法

本研究在主干网络的基础上，结合多尺度融合模块和全局上下文注意力模块进行改进，提高 CXR 图像的分类准确度。主干网络 ConvNeXt 是一个纯卷积轻量级网络，对硬件的要求不高，便于在医疗辅助情景下使用。

多尺度融合模块 (multi-scale feature fusion module, MsFF) 对经过主干网络的多级特征起作用，算法在不同尺寸下对图像进行采样，采样得到的特征中，低层特征具有较高的分辨率，可以生成清晰、详细的边界，但上下文信息较少，高层特征具有较抽象的语义信息，擅长进行类别分类，形状和位置信息较弱。本文利用 MsFF 对低层和高层特征进行融合，充分提取上下文信息，提高图像分类的准确性。

全局上下文注意力模块 (global context attention module, GCA) 保留了多级特征图之间的内部关系，从而降低特征融合过程中的信息损失，提高算法准确性。GCA 模块首先从低层特征中提取全局信息，然后对高层特征进行加权计算，保证信息有效性的同时又不增加计算量，既可单独作用于主干网络中，提升模型准确性，又可与 MsFF 模块组合使用，进一步提升 CXR 图像分类的精度。本文算法整体结构简图如图 4 所示。

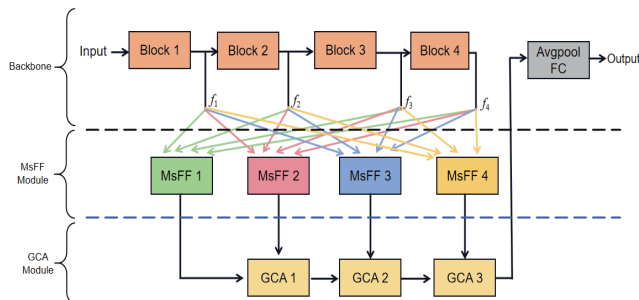


图 4 算法整体结构示意简图

多尺度融合和全局上下文注意力机制曾被用于工业生产过程中的表面缺陷检测任务中^[23]，利用多尺度融合特征图对

工业零件的缺陷边界进行细化,从而提高预测结果。而 CXR 图像中的纹理特征与该任务中工业零件的表面缺陷边界特征具有一定相似性,因此本文设计了针对 CXR 图像分类任务的 MsFF 和 GCA 模块,提高模型的分类准确度。

1.4.1 MsFF 模块

本文设计了主干网络的 MsFF 模块,通过结合从低层到高层的特征图,充分提取原 CXR 图像中的信息,提升整体网络对 CXR 图像分类的准确度。ConvNeXt 主干网络通过逐层抽象的方式来提取目标的特征,根据层数的不同,可相对分为高层和低层。高层网络感受野较大,语义信息表征能力强,但是特征图的分辨率低,缺乏空间几何特征的细节,几何信息的表征能力弱;低层网络的感受野比较小,几何细节信息表征能力强,虽然分辨率高,但是语义信息表征能力弱。因此,MsFF 模块把这些低层和高层特征加在一起,可以有效提升整体网络对 CXR 图像分类的准确度。

MsFF 模块分为多级特征提取和特征融合两个步骤。首先,多级特征提取建立在 ConvNeXt 模型基础之上,用来为 CXR 图像提取多层次特征,其细节结构如表 3 所示。该步骤包括 4 个块,分别对应 ConvNeXt 中的 4 个 block,在图像依次经过 4 个 ConvNeXt Block 后,将每个 block 后面得到的特征图提取出来,为后续的融合做准备。其次,特征融合为该模块的第二个步骤,CXR 图像逐层经过主干网络时,多级特征被依次提取出来,这四个特征用特征集 F 表示,用 f_i 代表特征集 F 中第 i 个元素,即 $F=(f_1, f_2, f_3, f_4)$ 。然后通过多尺度感受野加权 F 生成多尺度上下文信息,并将该信息同时映射到四个不同分辨率的特征映射 T_n 中去。

$$T_n = \left(\frac{W}{2n}, \frac{H}{2n} \right) \quad (4)$$

式中: W 和 H 分别代表输入图像的宽和高。

表 3 多级特征提取结构表

Stage	Type
	$4 \times 4, 96, \text{stride } 4$
Block1	$[\text{ConvNeXt Block, dim}=96] \times 3$
Block2	$[\text{ConvNeXt Block, dim}=192] \times 3$
Block3	$[\text{ConvNeXt Block, dim}=384] \times 9$
Block4	$[\text{ConvNeXt Block, dim}=768] \times 3$

四种特征的映射过程略有不同,因为 f_1 是最先提取出的低层特征,需要将其尺寸缩小才能进行后续的特征拼接,所以对 f_1 使用降采样的方式进行映射。同理 f_4 是最后提取出的高层特征,需要将其尺寸放大以进行拼接,所以对 f_4 使用上采样的方式进行映射。而 f_2 和 f_3 处于中间层级,因此使用上下采样相结合的方式映射。

对于特征 f_i ,将其降采样到 4 种不同分辨率,输出特征映射 M_i^i 为:

$$M_i^i = \sigma[\text{down}(W_{k \times k, s=k} * f_i + b)], i=1,2,3,4 \quad (5)$$

式中: σ 为激活函数 ReLU, down 表示利用卷积核大小为 k ,步长 $s=k$ 的卷积降采样 f_i , b 为偏差, $*$ 表示卷积。对于特征 f_4 ,将其上采样到 4 种不同分辨率,输出特征映射 M_4^i 为:

$$M_4^i = \sigma[\text{up}(f_4)], i=1,2,3,4 \quad (6)$$

式中: up 表示上采样过程,即反卷积。对于特征 f_2 和 f_3 ,采用上采样与下采样相结合的方式映射到 4 种不同分辨率,输出特征映射 M_j^i 为:

$$M_j^i = \sigma\left\{ \sigma[\text{down}(W_{k \times k, s=k} * f_j + b)] \& \sigma[\text{up}(f_j)] \right\} \quad (7)$$

$$i=1,2,3,4; j=2,3$$

式中: $\&$ 表示根据融合特征图的大小对 f_2 和 f_3 进行上下采样操作,上下采样可以以任意顺序进行,只需使调整后的 4 层特征尺寸一致即可。最后,将这些输出的特征进行融合,生成融合特征图,将此操作重复 4 次,由于参数的训练结果不同,最终可得到 4 个不同的融合特征图,再把这 4 个融合特征图相加得到最终特征图,对最终特征图进行后续的分类判断。通过这种方式,该模型便可以有效地获取多尺度的上下文信息,从卷积的不同阶段,实现整体对象感知。特征融合步骤结构如图 5 所示,多尺度融合模块整体结构如图 6 所示。

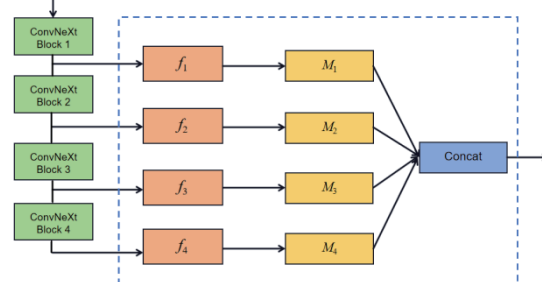


图 5 特征融合结构示意图

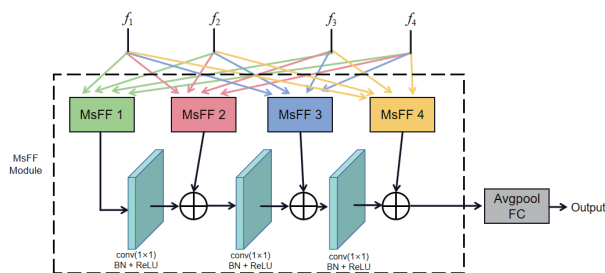


图 6 MsFF 模块整体结构示意图

1.4.2 GCA 模块

本文设计了针对主干网络的 GCA 模块,该模块可单独作用于主干网络,提升模型分类能力,也可与 MsFF 模块相结合,进一步提高模型的准确度。由上文 MsFF 模块产生的多级特征图,都可用于预测 CXR 图像分类的结果,如果采用上采样的方法将融合特征放大到原始图像大小,然后用卷

积将其通道数调整为与类别数一致，这样的方法固然可以预测分类的结果，但不同维度特征图之间的关系就会被破坏掉，过度的上采样还会导致图像信息丢失。因此，针对以上问题，本文设计了基于 ConvNeXt 主干的 GCA 模块，如图 7 所示。

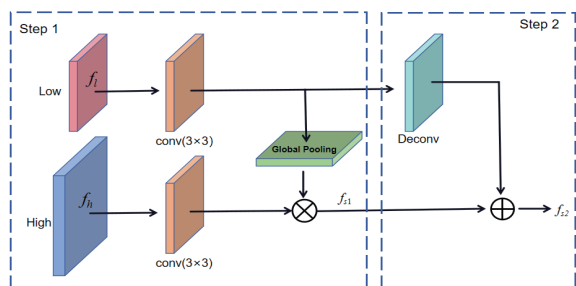


图 7 GCA 模块示意图

该模块可分为两个步骤。第一步，低分辨率特征图和高分辨率特征图分别通过一个 3×3 的卷积，然后低分辨率特征图经过一个全局池化层后再与高分辨率特征图相乘，得到 f_{s1} ：

$$f_{s1} = \sigma(W_{3 \times 3} * f_h + b_1) \otimes \sigma\{G[\sigma(W_{3 \times 3} * f_l + b_2)]\} \quad (8)$$

式中： f_h 、 f_l 分别代表高分辨率特征图和低分辨率特征图， $G(\cdot)$ 表示全局池化， b_1 、 b_2 为训练中学习的偏差参数， $\otimes$ 表示元素相乘， σ 为激活函数 ReLU。

第二步，将低分辨率融合特征图上采样到与高分辨率融合特征图相同的维数，然后与 f_{s1} 相加得到 f_{s2} ：

$$f_{s2} = \sigma[up(f_l) \oplus f_{s1}] \quad (9)$$

式中： $\oplus$ 表示元素相加， up 函数表示上采样操作。与简单地将上采样的低维特征映射添加到高维特征映射中相比，GCA 模块可以利用不同分辨率的融合特征映射来提高上下文信息获取的能力。ConvNeXt 主干网络结合 GCA 模块如图 8 所示，ConvNeXt 主干结合 MsFF 模块和 GCA 模块的结构如图 4 所示。

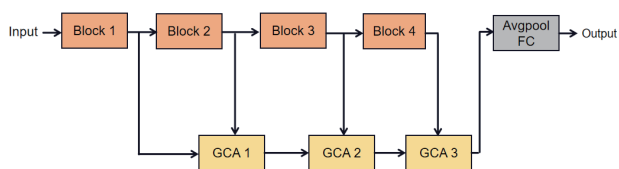


图 8 ConvNeXt 主干网络结合 GCA 模块结构图

表 4 不同模型的 AUC 数值表 /%

Method	Atel	Card	Effu	Infi	Mass	Nodu	Pne1	Pne2	Cons	Edem	Emph	Fibr	PT	Hern	Mean
CheXNet	80.94	92.48	86.38	73.45	86.76	78.02	76.80	88.87	79.01	88.78	93.71	80.47	80.62	91.64	84.14
DenseNet121	82.00	91.20	88.30	72.90	82.10	72.00	74.30	86.70	82.40	89.00	91.20	80.80	74.70	94.50	83.01
ResNet50	76.30	74.10	86.00	66.60	74.80	57.20	69.40	77.90	80.30	87.10	76.00	68.70	71.50	69.70	73.97
EfficientNetB1	77.50	88.40	84.60	64.50	76.10	62.40	71.30	81.00	79.60	83.80	84.10	73.10	71.20	84.90	77.32
Thorax-net	75.10	87.10	81.20	68.10	79.90	71.50	69.40	82.50	74.20	83.50	84.30	80.40	74.60	90.20	78.81
Model A	78.10	88.00	82.90	70.20	83.40	77.30	72.90	85.70	75.40	85.00	90.80	83.00	77.80	91.70	81.59
PVT	78.60	88.10	83.30	70.70	84.90	80.60	74.10	87.60	75.90	84.70	89.70	83.70	77.90	89.80	82.11
MXT	79.80	89.60	84.20	71.90	85.60	80.90	75.80	87.90	75.90	84.90	90.60	84.70	80.00	91.30	83.08
ConvNeXt	90.16	78.66	84.38	84.74	76.02	75.08	89.72	78.82	96.46	79.97	92.01	85.65	87.57	72.33	83.68
Our Model_ab	90.64	79.72	85.20	86.15	77.35	77.37	90.37	79.74	98.18	81.64	92.79	88.16	88.66	73.08	84.93

结合 MsFF 与 GCA 的总体模型的工作流程如下。首先，输入 CXR 图像先依次经过 ConvNeXt 主干的 4 个 block，每经过一个 block 后的特征图被提取出来，称作 f_1 、 f_2 、 f_3 、 f_4 。其次这 4 个特征经过 4 次多尺度融合操作得到 4 个融合特征图。然后将这些融合特征图按顺序依次进行全局上下文注意力操作，得到 1 个最终融合特征图。最后用这个特征图去进行分类判断。

2 结果

2.1 评价指标

在本研究中，网络模型的评价指标使用 AUROC（area under the receiver operator curve）评分和曲线，四种网络的结果都将绘制出 ROC 曲线并计算出 AUC 值。AUC 是 CXR 图像分类任务中常用的性能指标，其值等于 ROC 曲线下围成的面积，ROC 曲线的横坐标为 false positive rate（FPR），表示在所有实际为 negative 的样本中，被错误地判断为 positive 的比率：

$$FPR = \frac{F_p}{F_p + T_n} \quad (10)$$

式中： F_p 表示 False positive， T_n 表示 True negative。

ROC 曲线的纵坐标为 true positive rate（TPR），在所有实际为 positive 的样本中，被正确地判断为 positive 的比率：

$$TPR = \frac{T_p}{T_p + F_n} \quad (11)$$

式中： T_p 表示 True positive， F_n 表示 False negative。

AUC 的值等于 ROC 曲线下围成的面积，其值越接近 1 则模型分类效果越好，表示模型在正、负类之间的差异就越大。

2.2 对比实验

为证明所提方法对于 CXR 图像分类的有效性，本文选取 8 种方法用于对比实验，其中包括 DenseNet121^[24]、ResNet50、EfficientNetB1^[25] 三种 CXR 图像分类经典网络及其变种。CheXNet、Thorax-net^[26]、Model A^[27] 三种针对本任务而设计的专门网络。MXT、PVT 两种基于 transformer 架构的分类算法。对比实验结果如表 4 所示，其中加粗部分表示该项病症的最高 AUC 值，Mean 表示 AUC 均值，ConvNeXt 为仅使用主干网络的结果，Our Model_ab 表示主干网络添加 MsFF 和 GCA 的结果。

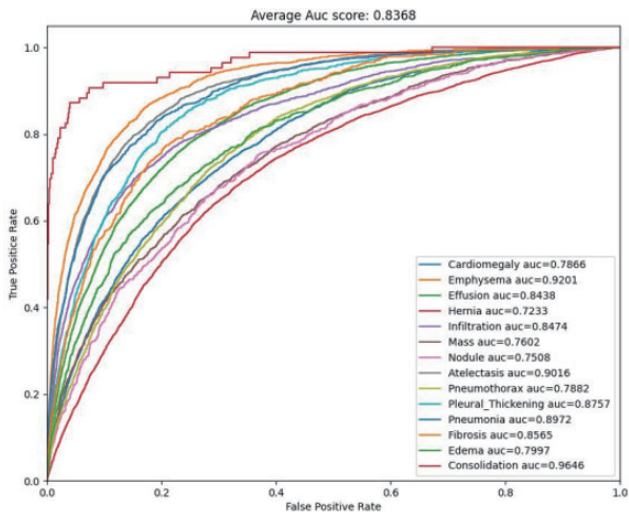
各模型的 AUC 均值与方差如表 5 所示。

表 5 各模型的 AUC 均值与方差

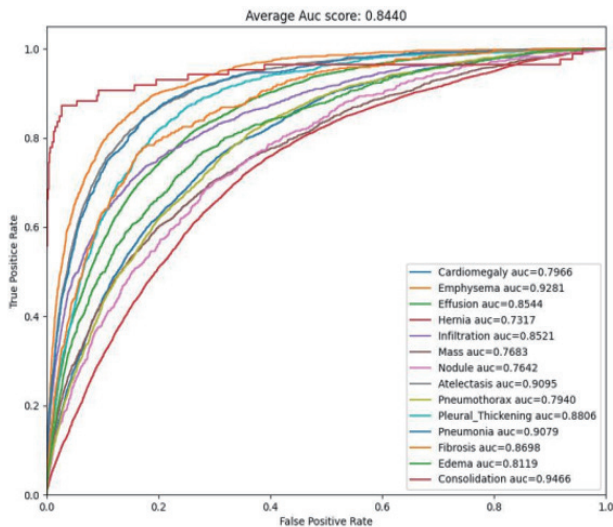
Method	Means/%	Variances
CheXNet	84.14	38.66
DenseNet121	83.01	51.36
ResNet50	73.97	56.53
EfficientNetB1	77.32	58.56
Thorax-net	78.81	41.64
Model A	81.59	39.44
PVT	82.11	33.32
MXT	83.08	32.71
ConvNeXt	83.68	46.85
Our Model_ab	84.93	46.10

2.3 消融实验

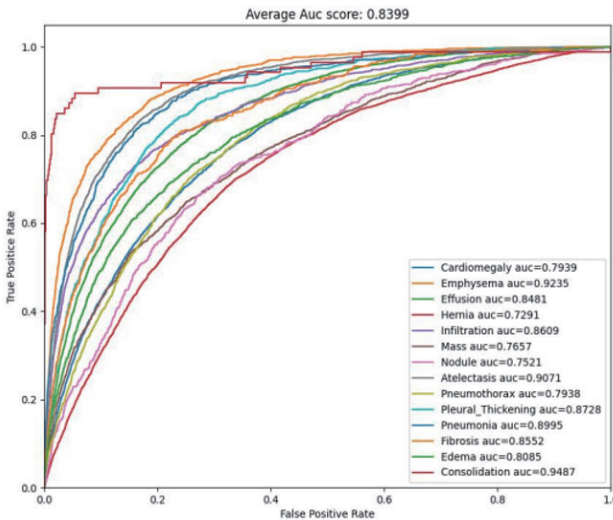
本文的两个改进模块可以分别提高模型的准确性，二者也可以共同作用于主干网络进一步提高模型性能。多尺度融合模块可以结合从低层到高层的特征图，避免池化过程导致的图像信息丢失，全局上下文注意力模块可以用加权融合的方式保留特征图之间的内部关系，提高分类效果。本研究将 14 种不同病症的分类分别绘制曲线、计算结果，ROC 曲线如图 9 所示。其中图 9（a）代表仅使用 ConvNeXt 网络主干的 ROC 曲线；图 9（b）Our Model_a 代表 ConvNeXt 主干网络中加入 MsFF 模块的 ROC 曲线；图 9（c）Our Model_b 代表 ConvNeXt 主干网络中加入 GCA 模块的 ROC 曲线；图 9（d）Our Model_ab 代表主干网络加入 MsFF 模块和 GCA 模块的 ROC 曲线。



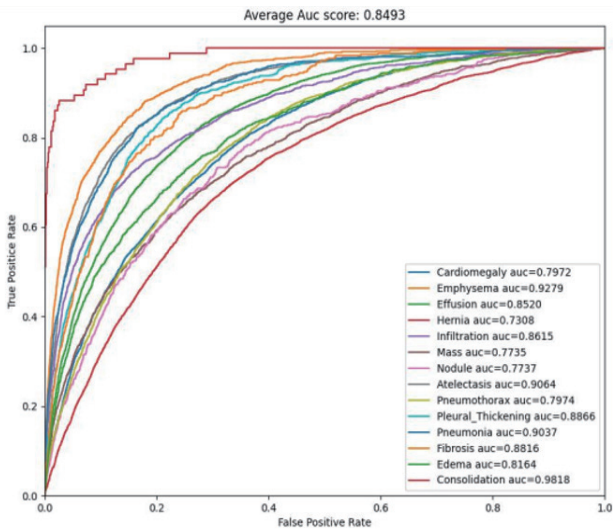
(a) ConvNeXt



(b) Our Model_a



(c) Our Model_b



(d) Our Model_ab

图 9 四种网络 ROC 曲线示意图

加入不同模块的网络结果如表 6 所示。

表 6 不同模块实验结果表

Model	AUC Means/%	AUC Variances
ConvNeXt	83.68	46.85
Our Model_a	84.40	41.88
Our Model_b	83.99	42.46
Our Model_ab	84.93	46.10

3 讨论

由表 4 可知，ConvNeXt 主干是该任务的一个优秀的 base 网络，同时改进算法取得了领先的成绩。在 CXR 图像分类的三种经典网络中，DenseNet121 取得了最优秀的成绩，AUC 均值为 83.01%，领先于 ResNet50（73.97%）和 EfficientNetB1（77.32%）。在 CXR 分类专用网络中，CheXNet 拥有最好的分类效果，AUC 均值达到 84.14%，高于 Thorax-net 的 78.81% 与 Model A 的 81.59%。在两种 transformer 系算法中，MXT 以 83.08% 的 AUC 评分胜于 PVT 的 82.11%。所以三类算法中，最优秀的算法为 DenseNet121、CheXNet、MXT。本文使用的主干网络 ConvNeXt，AUC 均值达到了 83.68%，已经高于 DenseNet121 和 MXT，可以说明 ConvNeXt 在 CXR 图像分类任务中可以作为一个优秀的 base 网络，作为一个纯卷积网络，它既有相比于 transformer 算法更轻便的特点，又有其他卷积神经网络所难以达到的高精度。本文在 ConvNeXt 基础上的改进算法 Our Model_ab，AUC 均值高达 84.93%，高于 DenseNet121、CheXNet、MXT，在 CXR 综合判断能力上达到了领先的水准，说明本文的改进方案行之有效，在该任务上取得了优秀的成绩。

在不同类别的判断效果上，改进算法也同样具有优势。由表 4 可知，CheXNet 在 Cardiomegaly、Mass、Pneumothorax、Emphysema 四项病理现象的判断上领先于其他一众算法。DenseNet121 在 Effusion、Edema、Hernia 三项病理现象的判断处于领先地位。MXT 在 Nodule 这一项上判断效果优异。本研究使用的 ConvNeXt 网络主干，没有任何一项病症的判断取得最优结果，改进后的 Our Model_ab 在 Infiltration、Consolidation、Fibrosis、Pleural Thickening 四项病理现象中取得了领先的判断效果。以上结果说明，在 CXR 图像具体类别的判定效果上，本文的改进算法同样处于最优秀的行列。

本研究的改进算法在综合能力和专项能力之间取得一个平衡。由表 5 可知，CheXNet 与 MXT 的方差较小，分别为 38.66、32.71，说明二者在综合分类能力较优的同时，各

项病症的判断能力也较为均衡。DenseNet121 的方差较大，为 52.36，说明该算法对不同病症判断的能力起伏较大，在分类结果上有些“偏科”，即优势类别的值很高，劣势类别的值很低。而本研究中基于 ConvNeXt 的四种算法，方差在这两者之间，在 AUC 方差和均值之间找到一个平衡点，说明改进算法在 14 类标签分析的综合能力保持优秀的前提下，又有 4 个标签的检测拥有专项能力的领先。在实际的医疗场景下，方差大或方差小的算法都有不同的用武之地，方差小的算法可以首先对病人进行全面肺病检查，确定大致病因后可用相关领域的优势算法进行精确定位，二者相辅相成，没有优劣之分。本文的改进算法可以有效识别 infiltration（浸润）、consolidation（实变）、fibrosis（纤维变性）、pleural thickening（胸膜增厚），在实际医疗场景下可以帮助医生初步诊断该类病人。

由表 6 和图 9 可知，ConvNeXt 加入 MsFF 模块和加入 GCA 模块后，AUC 均值分别提升了 0.72 和 0.31 个百分点；两个模块全部加入主干网络中后，AUC 均值达到了 84.93%，比 base 网络提高 1.25 个百分点，比较结果如图 10 所示，其中纵轴为 AUC 均值（%）；AUC 方差分别为 41.88、42.46、46.10，变化相对较小，略低于 ConvNeXt 主干网络。

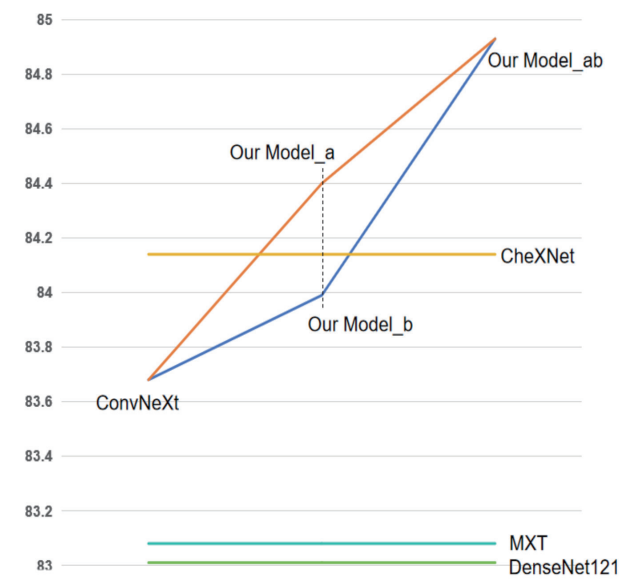


图 10 模型结果对比折线图

由以上结果可知，两个改进模块都可分别提高模型性能，也可以共同作用进一步提高分类能力，其中 Our Model_b 比 Our Model_a 的效果略差，这是因为各级特征未经上下采样映射到 4 种分辨率而直接通过卷积、池化等操作进行融合，会丢失部分图像信息，导致在相同的 epoch 下，模型效果不如 MsFF 模块。因此，更适宜的改进方法是先对主干网络进

行多尺度融合，融合后得到的4级特征按顺序两两进行全局上下文注意力操作，如图11所示，这样输入图像的低级到高级语义特征都会被融合起来，特征图之间的关系信息也会被保留下来。四种网络的方差变化较小，都在综合能力和专项能力之间保持平衡。

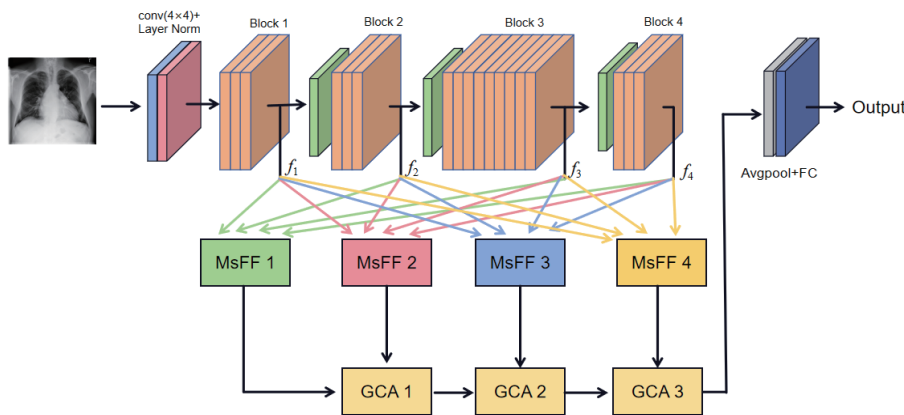


图11 模型整体结构示意图

本文算法在 CXR 图像分类任务上取得了领先于其他诸多常用算法的成绩 ($AUC=0.8493$)，在高 AUC 均值下保持了中等的方差水平 ($s^2=46.10$)，在“准确”和“均衡”中找到一个平衡点，同时兼具综合判断能力和专项判断能力。本文算法在 infiltration (浸润)、consolidation (实变)、fibrosis (纤维变性)、pleural thickening (胸膜增厚) 四种病理现象的判断下达到了目前最优的分类效果，可以在实际医疗场景下，辅助医生对病人进行该类疾病的初步诊断。

4 结论

本文针对 CXR 图像分类任务，提出了一种基于 ConvNeXt 网络主干，结合 MsFF 模块以及 GCA 模块的分类算法。该算法是纯卷积网络，通过结合从低层到高层的特征图，充分提取原 CXR 图像中的信息，提升整体网络对 CXR 图像分类的准确度。MsFF 模块实现了低分辨率与高分辨率特征图的融合，充分利用各级语义特征。GCA 模块实现了多级特征的加权结合，保留多级特征图之间的内部关系信息。实验表明，这二者都可以有效提高主干网络的分类能力，AUC 均值分别提高 0.72 和 0.31 个百分点，二者可以共同作用到主干网络中，实现整体效果的提升，此时 AUC 均值提高 1.25 个百分点。

通过与该任务领域现有方法相比较，本文算法在 CXR 图像综合分类能力上取得了领先的成绩 ($AUC=0.8493$) 和中等的方差水平 ($s^2=46.10$)，在“准确”和“均衡”中找到

一个平衡点，兼具综合判断能力和专项判断能力，在 4 种标签的判断下达到了目前最优的分类效果。

参考文献:

- [1] WANG X, PENG Y, LU L, et al. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 2097-2106.
- [2] WOLLEK A, GRAF R, ČEČATKA S, et al. Attention-based saliency maps improve interpretability of pneumothorax classification[J]. Radiology: artificial intelligence, 2022, 5(2): e220187.
- [3] SHEU R K, PARDESHI M S, PAI K C, et al. Interpretable classification of pneumonia infection using explainable AI (XAI-ICP)[J]. IEEE access, 2023, 11: 28896-28919.
- [4] CHEN F, YOU L, ZHAO W, et al. Centralized contrastive loss with weakly supervised progressive feature extraction for fine-grained common thorax disease retrieval in chest x-ray[J]. Medical physics, 2023, 50(6): 3560-3572.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [6] SEIBOLD C, REIB S, SARFRAZ M S, et al. Breaking with fixed set pathology recognition through report-guided contrastive training[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 690-700.
- [7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [8] SIGNORONI A, SAVARDI M, BENINI S, et al. BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset[J]. Medical image analysis, 2021, 71: 102046.
- [9] YANG M, TANAKA H, ISHIDA T. Performance improvement in multi-label thoracic abnormality

- classification of chest X-rays with noisy labels[J]. International journal of computer assisted radiology and surgery, 2023, 18(1): 181-189.
- [10] MANN M, BADONI R P, SONI H, et al. Utilization of deep convolutional neural networks for accurate chest X-ray diagnosis and disease detection[J]. Interdisciplinary sciences: computational life sciences, 2023(10):1-19.
- [11] CHEN Y, WAN Y, PAN F. Enhancing multi-disease diagnosis of chest X-rays with advanced deep-learning networks in real-world data[J]. Journal of digital imaging, 2023,36: 1332-1347.
- [12] 何迪, 刘立新, 刘玉杰, 等. 基于深度学习的肺炎图像目标检测 [J]. 中国生物医学工程学报, 2022(4):41.
- [13] JIANG X, ZHU Y, CAI G, et al. MXT: a new variant of pyramid vision transformer for multi-label chest X-ray image classification[J]. Cognitive computation, 2022,14(4):1362-1377.
- [14] WANG W, XIE E, LI X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]//Proceedings Of The IEEE/Cvf International Conference On Computer Vision. 2021: 568-578.
- [15] 邢素霞, 鞠子涵, 刘子骄, 等. 视觉 transformer 预训练模型的胸腔 X 线影像多标签分类 [J]. 中国图象图形学报, 2023, 28(4): 1186-1197.
- [16] SUN Z, QU L, LUO J, et al. Label correlation transformer for automated chest X-ray diagnosis with reliable interpretability[J]. La radiologia medica, 2023, 128(6): 726-733.
- [17] UKWUOMA C C, QIN Z, HEYAT M B B, et al. A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images[J]. Journal of advanced research, 2023, 48: 191-211.
- [18] FAISAL M, DARMAWAN J T, BACHROIN N, et al. CheXViT: CheXNet and vision transformer to multi-label chest X-ray image classification[C]//2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA). Piscataway:IEEE,2023:1-6.
- [19] RAJPURKAR P, IRVIN J, ZHU K, et al. Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning[EB/OL]. (2017-11-25)[2023-07-08]. <https://arxiv.org/abs/1711.05225>.
- [20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL].(2020-10-22)[2023-07-09]. <https://arxiv.org/abs/2010.11929>.
- [21] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2022:11976-11986.
- [22] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL].(2014-11-22)[2023-07-12]. <https://arxiv.org/abs/1412.6980>.
- [23] DONG H, SONG K, HE Y, et al. PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection[J]. IEEE transactions on industrial informatics, 2019, 16(12): 7448-7458.
- [24] SEYYED-KALANTARI L, LIU G, MCDERMOTT M, et al. Chexclusion: fairness gaps in deep chest X-ray classifiers[C]//Biocomputing 2021: proceedings of the Pacific symposium. Singapore:World Scientific Publishing Co. Pte. Ltd,2020: 232-243.
- [25] TAN M, LE Q. efficientnet: rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. New York:Curran Associates, 2019: 6105-6114.
- [26] WANG H, JIA H, LU L, et al. Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography[J]. IEEE journal of biomedical and health informatics, 2019, 24(2): 475-485.
- [27] GUAN Q, HUANG Y. Multi-label chest X-ray image classification via category-wise residual attention learning[J]. Pattern recognition letters, 2020, 130: 259-266.

【作者简介】

金荷淇 (1999—), 男, 硕士, 研究方向: 深度学习、图像处理、图像分类与检测。

盖绍彦 (1979—), 通信作者 (email: qxxymm@163.com), 男, 博士, 副教授, 博士生导师, 研究方向: 计算机视觉、模式识别、三维测量等。

达飞鹏 (1968—), 男, 博士, 教授, 博士生导师, 研究方向: 三维精密测量、三维精确识别、三维优化控制理论与技术等。

(收稿日期: 2023-10-23)