

# 基于 ConvNeXt 网络的新冠肺炎 X 射线 图像诊断方法

张帅<sup>1</sup>, 张俊忠<sup>2</sup>, 曹慧<sup>1\*</sup>, 仇大伟<sup>1\*\*</sup>, 吉旭瑞<sup>1</sup>

<sup>1</sup>山东中医药大学智能与信息工程学院, 山东 济南 250355;

<sup>2</sup>山东中医药大学第一临床医学院, 山东 济南 250355

**摘要** 卷积神经网络(CNN)在新冠肺炎 X 射线图像分类中取得诸多成果,但卷积结构只能学习到特征图相邻空间位置的上下文信息。为了更好地结合胸部 X 射线图像全局信息,除了提高 CNN 的深度,所提网络通过设计主干网络 ConvNeXt、聚合注意力模块、长短期记忆网络,更加关注全局与局部信息的交互性。对公开访问的 COVID-19 Radiography Database 数据集图像分类,所提网络在三分类实验中相较于基础模型 ConvNeXt 在准确率、精确率、召回率上分别提高 1.60 个百分点、1.23 个百分点、1.76 个百分点,且其在多项实验指标上优于 Vision Transformer、Swin-Transformer,准确率、精确率、召回率、特异度分别提高到 95.6%、96.03%、95.76%、97.53%。另外选用 Chest X-ray 数据集进一步验证所提网络泛化能力,采用 Score-CAM 算法验证其有效性。实验结果表明,所提网络对新冠肺炎 X 射线图像具有较好的分类效果。

**关键词** 图像处理; X 射线图像; 新冠肺炎; 注意力机制; 长短期记忆网络

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP222126

## Coronavirus Disease X-Ray Image Diagnosis Method Based on ConvNeXt Network

Zhang Shuai<sup>1</sup>, Zhang Junzhong<sup>2</sup>, Cao Hui<sup>1\*</sup>, Qiu Dawei<sup>1\*\*</sup>, Ji Xurui<sup>1</sup>

<sup>1</sup>College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine,  
Jinan 250355, Shandong, China;

<sup>2</sup>First Clinical Medical College, Shandong University of Traditional Chinese Medicine, Jinan 250355,  
Shandong, China

**Abstract** The convolutional neural network (CNN) has made immense progress in the classification of coronavirus disease (COVID-19) X-ray images; however, the convolution structure can only learn the context information of the adjacent spatial positions of the feature map. Hence, to realize a better combination of the global information of chest X-ray images, we propose a network that pays more attention to the interaction of the global and local information by designing the backbone network, ConvNeXt, a convergent attention module, and a long short-term memory network while improving the CNN depth as well. Herein, this experiment classified the images of the COVID-19 Radiography Database dataset, which can be publicly accessed. Compared with the basic model of ConvNeXt, the proposed network displays an improvement in the accuracy, accuracy, and recall by 1.60, 1.23, and 1.76 percentage points, respectively, in the three classification experiments, and it is superior to Vision Transformer and Swin-Transformer in many experimental indicators, with the accuracy, accuracy, recall, and specificity increased to 95.6%, 96.03%, 95.76%, and 97.53%, respectively. Furthermore, the Chest X-ray dataset was selected to further verify the proposed network generalization capability, and the score-CAM algorithm was used to verify its effectiveness. The experimental results show that the proposed network has a high potential for application in the classification of COVID-19 X-ray images.

**Key words** image processing; X-ray image; coronavirus disease; attentional mechanism; long short-term memory

收稿日期: 2022-07-21; 修回日期: 2022-08-16; 录用日期: 2022-08-20; 网络首发日期: 2022-08-29

基金项目: 国家自然科学基金(82074579, 81973981, 82174528)

通信作者: \*caohui63@163.com; \*\*dwqiu@foxmail.com

# 1 引言

新冠肺炎流行初期,诊断新冠肺炎的主要技术是逆转录聚合酶链反应(RT-PCR)检测试剂盒,但采样技术、试剂盒质量和疾病演变方法均会导致核酸检测的假阴性率很高<sup>[1]</sup>。由于医学影像的客观性和简便性,胸部X射线图像诊断已成为临床医生认可的新冠肺炎早期发现和随访的重要组成部分。近年来,许多科研人员、医务工作者致力于新冠肺炎患者胸部X射线图像数据的公开,使得深度学习在医学影像上的应用更加广泛<sup>[2-5]</sup>。

研究者设计了多种有关新冠肺炎的深度学习诊断模型,并在医学图像分类上达到较高的准确率。Dastider等<sup>[6]</sup>考虑肺部超声(LUS)帧的空间和时间特征,引入具有长短期记忆网络(LSTM)的卷积神经网络(CNN)改善分类性能。Islam等<sup>[7]</sup>开发组合的深度CNN-LSTM网络,采用3种类型胸部X射线图像自动诊断新冠肺炎病例,其中,CNN从图像中提取复杂特征,LSTM用作分类器,取得较好检测结果。Naeem等<sup>[8]</sup>提出多级特征提取方法,减轻了CNN的训练复杂性,有助于准确、稳定识别新冠肺炎病灶区域。Hassanien等<sup>[9]</sup>应用ConvNeXt网络于恶性乳腺肿瘤研究,采用视觉解释模块生成超声图像上的热力图,帮助解释深度学习模型,并取得较好结果。Ketu等<sup>[10]</sup>使用卷积层提取特征信息,并从时间序列数据中训练学习,

该网络丰富了LSTM层功能,且实验证明了在LSTM层中使用额外的卷积层会提高检测模型的性能。

受上述研究启发,结合全局与局部信息可提高网络对新冠肺炎X射线图像分类性能,本文设计两个模块改进ConvNeXt网络。首先,提出聚合注意力模块(CGMB),通过通道注意力结合全局信息,调整通道关系;其次,空间注意力采用多个小卷积替换大卷积,具有更少的参数和更多的非线性,从而增强从特征图中学习信息的能力。CGMB的设计使得所提GlstNet结合全局与局部信息的能力更优。最后,在网络中加入LSTM层,LSTM记忆先前信息并将其关系到当前信息,更加关注全局与局部信息的交互性。消融实验结果表明,两个模块的嵌入有利于提高网络区分covid、normal、lung\_opacity等3类胸部X射线图像的能力,具有更好的泛化性能。

# 2 网络设计

## 2.1 GlstNet框架

GlstNet使用ConvNeXt作为主干网络,主干结构含4个比例为3:3:9:3的stage,每个stage的通道数为96、192、384、768,特征图大小 $H \times W$ 为 $56 \times 56$ 、 $28 \times 28$ 、 $14 \times 14$ 、 $7 \times 7$ 。网络依旧保持逆瓶颈层(中间大、两头小)结构,可以有效避免信息流失<sup>[11]</sup>。GlstNet的总体框架结构如图1所示,网络被划分为A~D等4个部分:

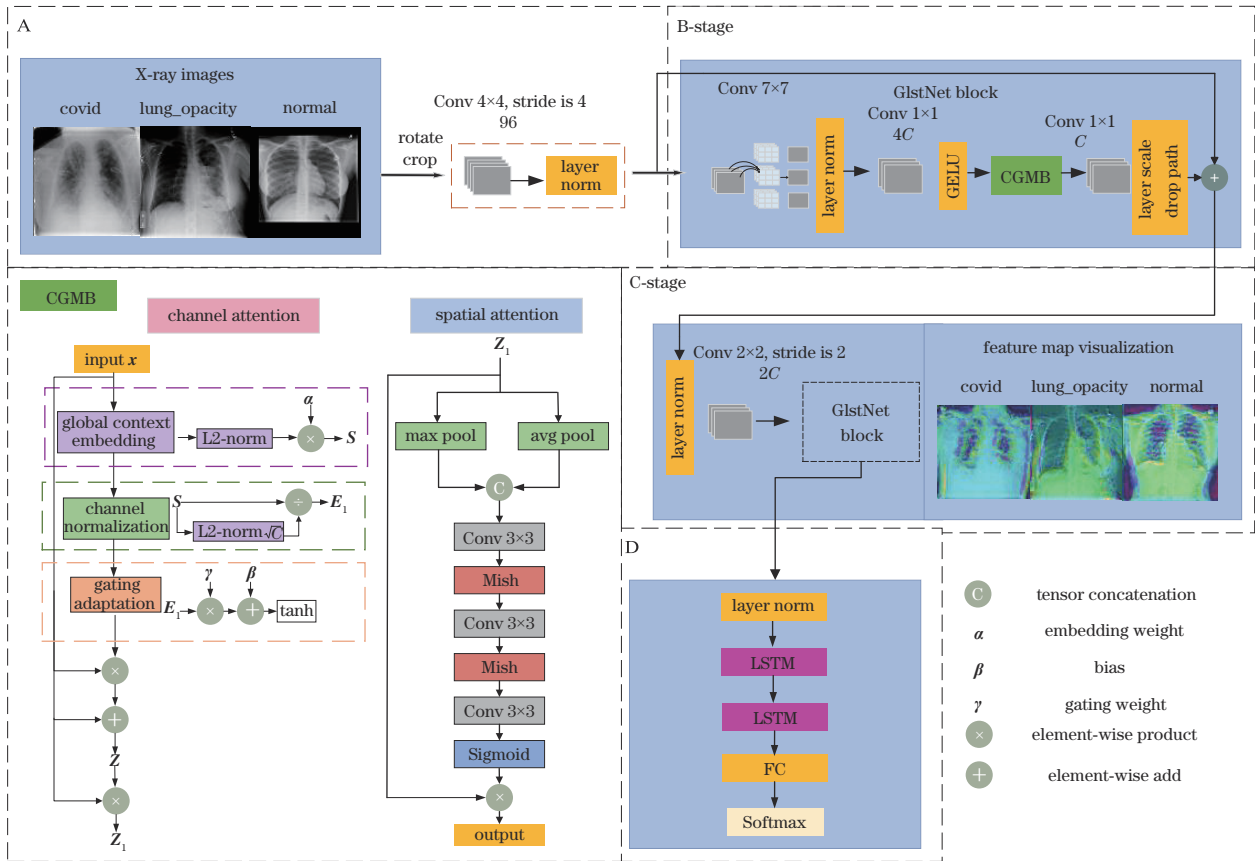


图1 GlstNet框架结构

Fig. 1 GlstNet framework structure

A 部分:旋转输入图像、裁剪到  $224 \times 224$  大小,不使用其他数据增强方法。经卷积运算减小图像高宽并增加通道大小,得到特征图  $\mathbf{F}_1 \in \mathbf{R}^{56 \times 56 \times 96}$ 。

B 部分:特征图进入 stage 结构,先经过深度卷积(DW),采用 layer normalization(LN)代替 batch normalization(BN),后通过  $1 \times 1$  卷积增加通道大小与 GELU 激活函数进入 CGMB,最终通过  $1 \times 1$  卷积还原输入通道大小,并通过 layer scale 层<sup>[12]</sup>、drop path 层<sup>[13]</sup>融合输出  $\mathbf{F}_{B-out}$ 。

C 部分: $\mathbf{F}_{B-out}$  经 LN 层,通过  $2 \times 2$  卷积减小特征图尺寸,增加通道大小。进入 block 结构,经 4 个 stage 提取特征信息后,最终获得输出  $\mathbf{F}_{out} \in \mathbf{R}^{7 \times 7 \times 768}$ 。B、C 部分 stage 比例设置为 3:3:9:3。

D 部分: $\mathbf{F}_{out}$  经 LN 层后,通过 2 层 LSTM 网络,特征图空间稠密性信息得到进一步重利用并传递到全连接层,最终经 Softmax 分类。LSTM 层考虑先前信息并关系到当前信息,可增强通道信息的交互性,同时避免网络过拟合。

## 2.2 聚合注意力模块

人类可以自然有效地在复杂场景中发现显著、感兴趣区域,在模仿人类视觉系统的推动下,注意力机制被引入神经网络,注意力机制可以看作基于输入图像特征的动态权重调整过程<sup>[14-15]</sup>。注意力机制在许多视觉任务中取得了巨大成功,包括图像分类、目标检测、语义分割等方面<sup>[16-19]</sup>。

convolutional block attention module(CBAM)<sup>[20]</sup>详细讨论了注意力模块位置相互关系,沿通道与空间维度顺序推断注意力图,且参数量较少。CGMB 模块由 CBAM 的网络拓扑结构启发并改进得到。CGMB 由通道注意力(GCT)<sup>[21]</sup>和空间注意力(MCSA)<sup>[22]</sup>两个模块组成。CGMB 模块位于每个 stage 升维卷积之后,可有效地结合全局与局部信息,关注胸部 X 射线图像中的病灶区域和抑制无用信息,以提高网络分类性能。通道注意力包含 3 个模块,通过引入 L2 normalization 取代全连接层对通道间的特征关系建模,参数与卷积权重联合优化决定神经元的竞争与合作状态。模块关注不同通道间关系,最后采用 tanh 激活控制注意向量以实现更精确的识别。空间注意力专注于局部空间信息提取,3 个小感受野的卷积叠加,以 Mish 激活函数代替 ReLU,训练稳定性与准确性均得到较高提升。消融实验结果表明,CGMB 模块中通道注意力、空间注意力具有互补性,两者相互配合可更加有效地关注新冠肺炎患者肺部病灶区域。

结合全局信息模块(global context embedding)设计全局上下文信息嵌入模块,嵌入全局信息并在归一化之前控制每个通道的权重,具有大感受野的信息有助于避免由小感受野信息带来的局部模糊问题<sup>[23]</sup>,该模块相比全连接层计算效率更高。给定输入

$\mathbf{x} \in \mathbf{R}^{H \times W \times 4C}$ ,  $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_C]$ ,  $\mathbf{x}_C = [x_c^{i,j}] \in \mathbf{R}^{H \times W}$ ,  $C \in \{1, 2, \dots, C\}$ ,  $\mathbf{x}_C$  对应  $\mathbf{x}$  的每个通道,自适应嵌入权重  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_C]$  负责控制每个通道的权重,  $\epsilon$  为极小常数,避免零点处求导问题。该模块的输出  $\mathbf{S}$  可表示为

$$S_C = \alpha_C \|\mathbf{x}_C\|_2 = \alpha_C \left\{ \left[ \sum_{i=1}^H \sum_{j=1}^W (x_c^{i,j})^2 \right] + \epsilon \right\}^{\frac{1}{2}}. \quad (1)$$

通道归一化模块(channel normalization):设  $\mathbf{S} = [S_1, \dots, S_C]$ , 用  $\sqrt{C}$  规范化  $E$ , 避免  $C$  较大时  $E$  太小:

$$E = \frac{\sqrt{C} S_C}{\|\mathbf{S}\|_2} = \frac{\sqrt{C} S_C}{\left[ \left( \sum_{c=1}^C S_c^2 \right) + \epsilon \right]^{\frac{1}{2}}}. \quad (2)$$

门控机制(gating adaptation):门控机制适应原始特征,当通道的门控权重被正向激活时,注意力机制促进该通道与其他通道竞争;反之,注意力机制鼓励该通道与其他通道合作。门控机制通过引入  $1 + \tanh(x)$  门,避免如 Sigmoid 函数引发梯度消失等问题,使得训练过程更加稳定。门控权重  $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_C]$ 、门控偏置  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_C]$  负责调整输入特征通道的尺度,通道注意力输出为  $\mathbf{Z}$ ,  $\mathbf{Z}$  与输入  $\mathbf{x}$  通过逐元素乘法得到  $\mathbf{Z}_1$ , 并送入空间注意力机制:

$$\mathbf{Z} = \mathbf{x}_C [1 + \tanh(\gamma_C E + \beta_C)], \quad (3)$$

$$\mathbf{Z}_1 = \mathbf{x} \otimes \mathbf{Z}. \quad (4)$$

空间注意力采用双分支空间注意力机制,首先使用平均池化层和最大池化层提取空间信息,得到特征图  $\mathbf{F}_{max} \in \mathbf{R}^{H \times W \times 1}$ 、 $\mathbf{F}_{avg} \in \mathbf{R}^{H \times W \times 1}$ , 然后在通道维度上拼接  $\mathbf{F}_{max}$  与  $\mathbf{F}_{avg}$  得到特征图  $\mathbf{F}_{cat}$ :

$$\mathbf{F}_{max} = \max[\mathbf{u}_i(k, p)], \quad (5)$$

$$\mathbf{F}_{avg} = \frac{1}{C} \sum_{i=1}^C \mathbf{u}_i(k, p), \quad (6)$$

$$\mathbf{F}_{cat} = \text{Concat}(\mathbf{F}_{max}, \mathbf{F}_{avg}), \quad (7)$$

式中:  $k, p$  表示特征图的高和宽,  $k \in \{1, 2, \dots, H\}$ ,  $p \in \{1, 2, \dots, W\}$ ;  $\mathbf{u}_i$  是输入特征图的第  $i$  个通道。空间注意力机制采用 3 个  $3 \times 3$  卷积,使用 Mish 作为激活函数替换原有  $7 \times 7$  卷积与 ReLU 函数,保持相同感受野情况下具有更多的非线性、更少的参数量。Mish 相比 ReLU 梯度更加平滑、稳定性更高,使得网络精度得到进一步提升,最终得到输出  $\mathbf{F}_{CGMB-out}$ 。

$$\text{Mish}(x) = x \times \tanh\{\ln[1 + \exp(x)]\}, \quad (8)$$

$$\mathbf{F}_{CGMB-out} = \text{Sigmoid}[\text{Conv3}(\mathbf{F}_{cat})] \otimes \mathbf{X}_0. \quad (9)$$

## 2.3 长短期记忆网络

LSTM<sup>[24-26]</sup>通过添加单元状态改进传统循环神经网络(RNN)的长序列信息丢失问题。LSTM 记忆先前信息并将其融合到当前信息进行学习,检测长短期

依赖关系,可有效地解决梯度消失与爆炸问题。LSTM 包括 3 个门控单元:遗忘门、更新门和输出门,实验设置 2 层 LSTM,其中,输入神经元个数为

768,隐藏神经元个数为 2048,输出神经元个数为 768, batch size 设置为输出的第一维。LSTM 层的结构如图 2 所示。

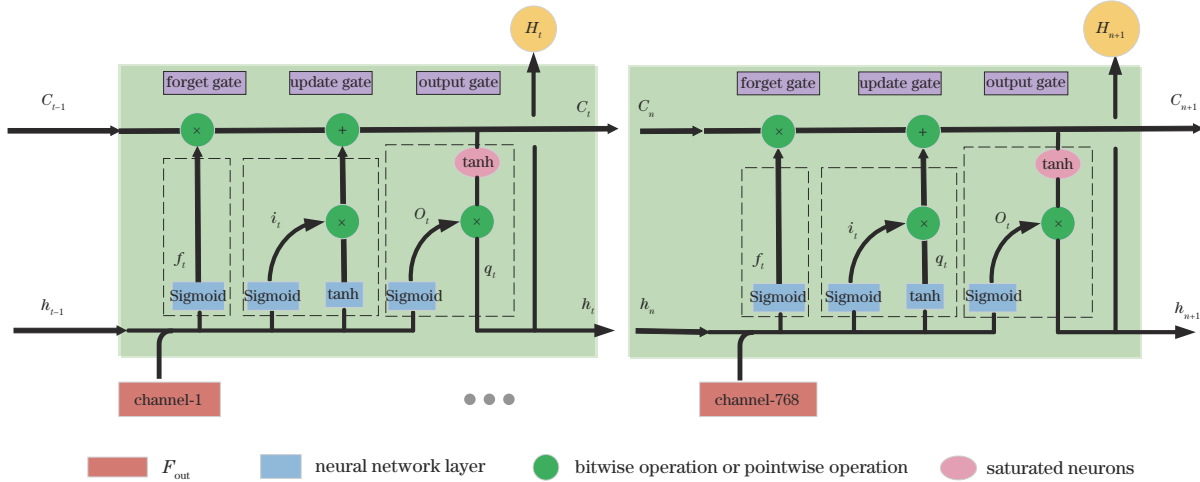


图 2 LSTM 网络结构

Fig. 2 LSTM network structure

$C_t$  表示细胞状态 (cell state),  $C_{t-1}$  表示上一时刻的细胞状态,  $H_t$  表示输出信号,  $h_{t-1}$  表示隐藏层状态为上一时刻的输出信号, 箭头方向为信息传递方向, 两箭头相交表示两个信号的连接, 箭头分开表示信号经复制后传递到两个不同方向。CNN 有效识别图像在像素序列中的局部和全局结构并输出  $F_{out}$ ,  $F_{out}$  中 768 个通道分别传入 LSTM 进一步提取深层次特征, 并关联通道信息。LSTM 详细结构如下:

遗忘门:  $F_{out}$  与  $t-1$  时刻 LSTM 层输出相拼接进入神经网络得到  $f_t$ , 由  $f_t$  与  $C_{t-1}$  相乘决定关键信息的保留。

$$f_t = \text{Sigmoid}(W_f \times [h_{t-1}, F_{out}] + b_f), \quad (10)$$

式中:  $[\cdot, \cdot]$  表示拼接操作。

更新门: 更新门包含 Sigmoid 神经网络层 (神经网络参数为  $W_i, b_i$ , 输出为  $i_t$ ) 与 tanh 神经网络层 (神经网络参数为  $W_c, b_c$ , 输出为  $q_t$ ), 由两输出共同决定将特定信息更新到  $t$  时刻的  $C_t$  中。

$$i_t = \text{Sigmoid}(W_i \times [h_{t-1}, F_{out}] + b_i), \quad (11)$$

$$q_t = \tanh(W_c \times [h_{t-1}, F_{out}] + b_c), \quad (12)$$

$$C_t = f_t \times C_{t-1} + i_t \times q_t. \quad (13)$$

输出门: 输出门包含 Sigmoid 神经网络层 (神经网络参数为  $W_o, b_o$ ) 输出  $O_t$ ,  $C_t$  经 tanh 激活函数与  $O_t$  相乘得到输出信号  $h_t$ 。

$$O_t = \text{Sigmoid}(W_o \times [h_{t-1}, F_{out}] + b_o), \quad (14)$$

$$h_t = O_t \times \tanh C_t. \quad (15)$$

$h_t$  再次作为输入信号传递到下一层 LSTM, 信息进一步展平并传递到全连接层, 最后经 Softmax 分类。CNN-LSTM 网络结构如图 3 所示。

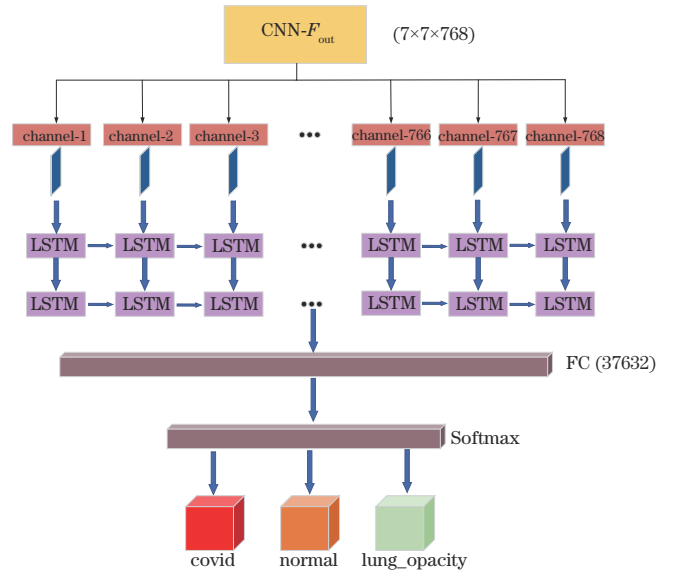


图 3 CNN-LSTM 网络结构

Fig. 3 CNN-LSTM network structure

### 3 实验方法

#### 3.1 数据集

数据集 I 选自 COVID-19 Radiography Database 数据集 Version 5<sup>[27-28]</sup>, 选用 3616 张新冠肺炎图像, 6012 张肺浑浊图像, 8851 张正常图像。实验中 80% 的数据用于训练, 20% 用于测试, 训练集中, 20% 的子集用于验证以避免过度拟合, 具体如表 1 所示。

数据集 II 选自 kaggle 公开 Chest X-ray 新冠肺炎数据集 (<https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia>), 包含 6432 张胸部 X 射线图像, 具体如表 2 所示。

表 1 数据集 I 划分情况

Table 1 Division of dataset I

Dataset	Type	Total	Training image	Validation image	Test image
COVID-19 Radiography Database	covid	3616	2314	578	724
	normal	8851	5664	1416	1771
	lung_opacity	6012	3847	962	1203

表 2 数据集 II 划分情况

Table 2 Division of dataset II

Dataset	Type	Total	Training image	Validation image
Chest X-ray	covid	576	460	116
	normal	1583	1266	317
	pneumonia	4273	3418	855

### 3.2 评价标准

为了全面衡量所提 GlstNet 的有效性,实验采用准确率( $R_{accuracy}$ )、精确率( $R_{precision}$ )、召回率( $R_{recall}$ )、特异度( $R_{specificity}$ )作为评价指标:

$$R_{accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FP} + N_{FN} + N_{TN}}, \quad (16)$$

$$R_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (17)$$

$$R_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (18)$$

$$R_{specificity} = \frac{N_{TN}}{N_{TN} + N_{FP}}, \quad (19)$$

式中: $N_{TP}$ 表示真阳性的数量; $N_{FP}$ 表示假阳性的数量; $N_{FN}$ 表示假阴性的数量; $N_{TN}$ 表示真阴性的数量。

### 3.3 消融结果

消融实验中,模型训练时使用 AdamW 优化器和交叉熵损失函数,使用预训练权重,采用 warm up 方法,最大 epoch 为 200, batch size 为 32。采用 PyTorch

1.5 库,实验在 NVIDIA GTX 2080Ti 环境下运行。

为验证所提添加模块的有效性,表 3 为在 COVID-19 Radiography Database 数据集上的消融实验指标。Baseline 为 ConvNeXt 网络<sup>[29]</sup>,CBAM 注意力机制中通道注意力简称 ST,空间注意力简称 SA。根据消融实验结果可得,模型 3 在 ConvNeXt 网络中引入通道注意力(GCT),结合小卷积感受野之外信息并通过门控机制促进训练过程中的竞争与合作,但结果相比 ConvNeXt 网络准确率、精确率、召回率下降许多,可知单通道注意力改进不利于注意力模块的学习。模型 4 将空间注意力(MCSA)引入 ConvNeXt 网络,准确率、精确率、召回率同样有所下降,即单空间注意力改进同样不利于注意力模块的学习。模型 5 引入 CGMB 模块,准确率、精确率、召回率分别提高 1 个百分点、0.9 个百分点、1.0 个百分点。消融实验结果表明,CGMB 模块中通道注意力与空间注意力具有互补性,CGMB 结合全局与局部信息可以有效地促进网络对于胸部 X 射线图像病灶区域的学习。LSTM 层的添加使得准确率、精确率、召回率分别提高 1.60 个百分点、1.23 个百分点、1.76 个百分点,验证了 LSTM 层的有效性。消融实验结果表明,主干网络 ConvNeXt、CGMB 模块、LSTM 层在全局与局部信息提取、特征空间关系学习上具有较好表现,可提高胸部 X 射线图像病灶的识别性能。消融实验各阶段测试集混淆矩阵如图 4 所示。

表 3 在数据集 I 上的测试集指标

Table 3 Test set metrics on dataset I

No.	Model	$R_{accuracy} / \%$	$R_{precision} / \%$	$R_{recall} / \%$	$\Delta R_{accuracy} / \%$
1	Baseline	94.00	94.80	94.00	
2	Baseline+CBAM	92.00	92.80	91.50	↓ 2.0
3	Baseline+GCT+SA	90.50	91.20	89.90	↓ 4.0
4	Baseline+ST+MCSA	92.80	93.20	92.63	↓ 1.2
5	Baseline+CGMB	95.00	95.70	95.00	↑ 1.0
6	<b>GlstNet</b>	<b>95.60</b>	<b>96.03</b>	<b>95.76</b>	<b>↑ 1.6</b>

### 3.4 数据集 I 实验结果

为了直观地展示实验结果与性能,绘制了 GlstNet 网络的准确率变化曲线图与测试集混淆矩阵,如图 5 所示。由图 5 可知,开始训练的 50 轮准确率曲线迅速上升,75 轮时曲线逐渐稳定,直到 200 轮。表 4 显示了 GlstNet 与主流算法的对比,从表 4 可看出,GlstNet 在融合全局与局部信息后,表现出更好的网络特征识别能力。表 5 展示了不同网络在数据集 I 的 3 种类别性能指标,可以看出,GlstNet 在 COVID-19 Radiography Database 数据集上表现出良好的分类性能,证明了所

提网络的有效性。

### 3.5 数据集 II 实验结果

表 6 比较了 GlstNet 与主流算法在 Chest X-ray 数据集上的性能,GlstNet 的准确率、精确率、召回率分别达到 97.20%、97.03%、97.60%。与更深层次的 Densenet、ResNet 相比,GlstNet 结合全局与局部信息,可有效地识别肺部病灶区域,具有较强的泛化能力。图 6 展示了 GlstNet 在 Chest X-ray 数据集上的混淆矩阵,表 7 展示了在数据集 II 上的 3 种类别性能指标。

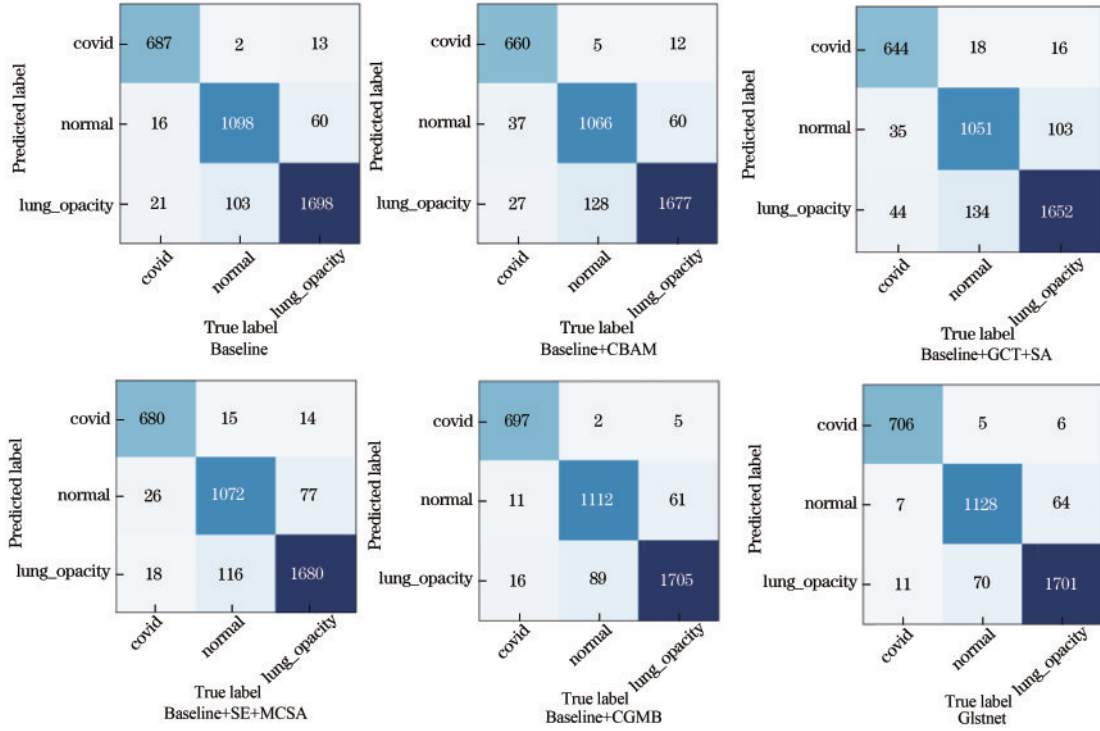


图 4 在数据集 I 上的测试集混淆矩阵  
Fig. 4 Confusion matrix for the test set on dataset I

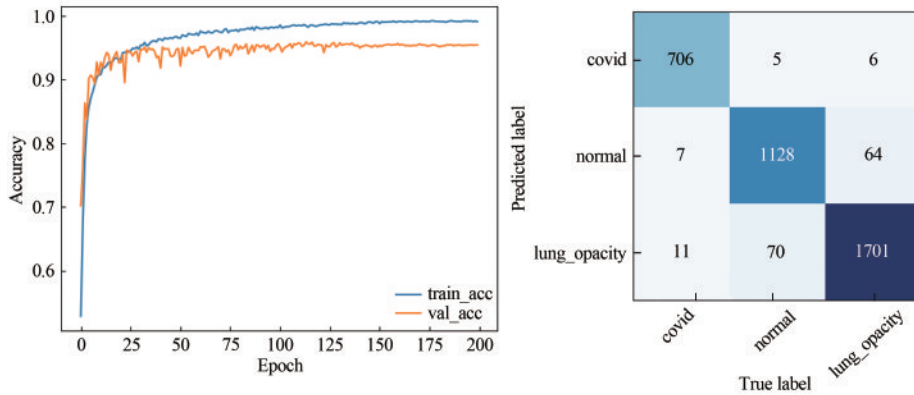


图 5 GlstNet 准确率变化曲线与混淆矩阵  
Fig. 5 GlstNet accuracy change curve with confusion matrix

表 4 GlstNet 与主流算法的对比

Table 4 Comparison of GlstNet network with mainstream algorithms

Model	$R_{accuracy} / \%$	$R_{precision} / \%$	$R_{recall} / \%$
ResNet18 <sup>[27]</sup>	93.43	93.43	93.43
ResNet50 <sup>[27]</sup>	93.01	93.12	93.02
ResNet101 <sup>[27]</sup>	93.01	93.04	93.01
ChexNet <sup>[27]</sup>	93.21	93.28	93.21
DenseNet201 <sup>[27]</sup>	92.70	92.78	92.70
InceptionV3 <sup>[27]</sup>	93.46	93.49	93.47
ConvNeXt	94.00	94.80	94.00
Vision Transformer	94.64	95.23	94.63
Swim Transformer	95.18	95.66	95.23
<b>GlstNet</b>	<b>95.60</b>	<b>96.03</b>	<b>95.76</b>

表 5 在数据集 I 测试集上的 3 种类别性能指标

Table 5 Performance metrics for three categories on the validation set of dataset I

Type	$R_{accuracy} / \%$	$R_{precision} / \%$	$R_{recall} / \%$
covid	98.50	97.50	99.60
normal	94.10	93.80	97.20
lung_opacity	95.50	96.00	95.80

### 3.6 GlstNet 可视化

为了验证所提网络的有效性,使用了置信分数可视化技术(Score-CAM)<sup>[33]</sup>,结果如图 7 所示。可视化可突出模型在网络学习过程中非常关注的区域。图 7 第 1 列为 COVID-19 Radiography Database 数据集的 3 类胸部 X 射线图像,第 2~4 列分别为 ConvNeXt、ConvNeXt+GCT+SA、ConvNeXt+MCSA+ST 可

表 6 GlstNet 与主流算法的对比

Table 6 Comparison of GlstNet with mainstream algorithms

Model	$R_{accuracy} / \%$	$R_{precision} / \%$	$R_{recall} / \%$
ALEXNET CNN <sup>[30]</sup>	94.70	94.00	94.00
LENET CNN <sup>[30]</sup>	93.20	93.00	93.00
DenseNet-169 <sup>[30]</sup>	95.70		
ResNet-50 <sup>[31]</sup>	93.30		
DenseNet <sup>[32]</sup>	95.03	94.91	94.24
EfficientNet <sup>[32]</sup>	93.40	93.62	91.07
VGG-16 <sup>[32]</sup>	92.08	92.02	92.77
<b>GlstNet</b>	<b>97.20</b>	<b>97.03</b>	<b>97.60</b>

表 7 在数据集 II 验证集上的 3 种类别性能指标

Table 7 Performance metrics for three categories on the validation set of dataset II

Type	$R_{accuracy} / \%$	$R_{precision} / \%$	$R_{recall} / \%$
covid	100.00	99.10	100.00
normal	92.40	96.50	97.40
pneumonia	98.70	97.20	97.50

可视化结果, GlstNet 的可视化结果显示在第 5 列。可以看出, 所提网络相比于 ConvNeXt、ConvNeXt+GCT+

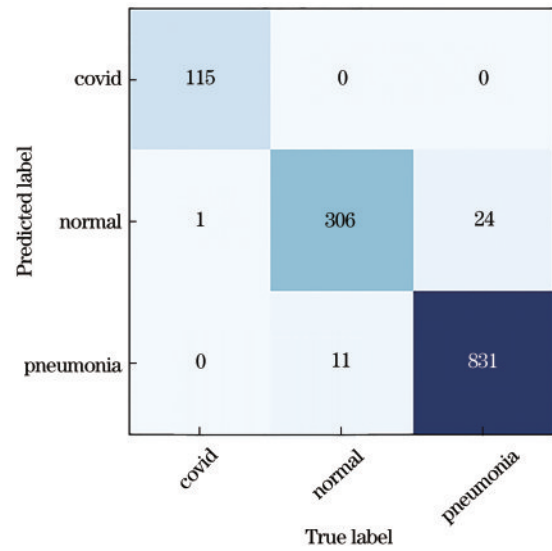


图 6 GlstNet 在 Chest X-ray 验证集上的混淆矩阵

Fig. 6 GlstNet network confusion matrix on Chest X-ray validation set

SA、ConvNeXt+MCSA+ST 可更有效地学习胸部 X 射线图像中的病灶区域。

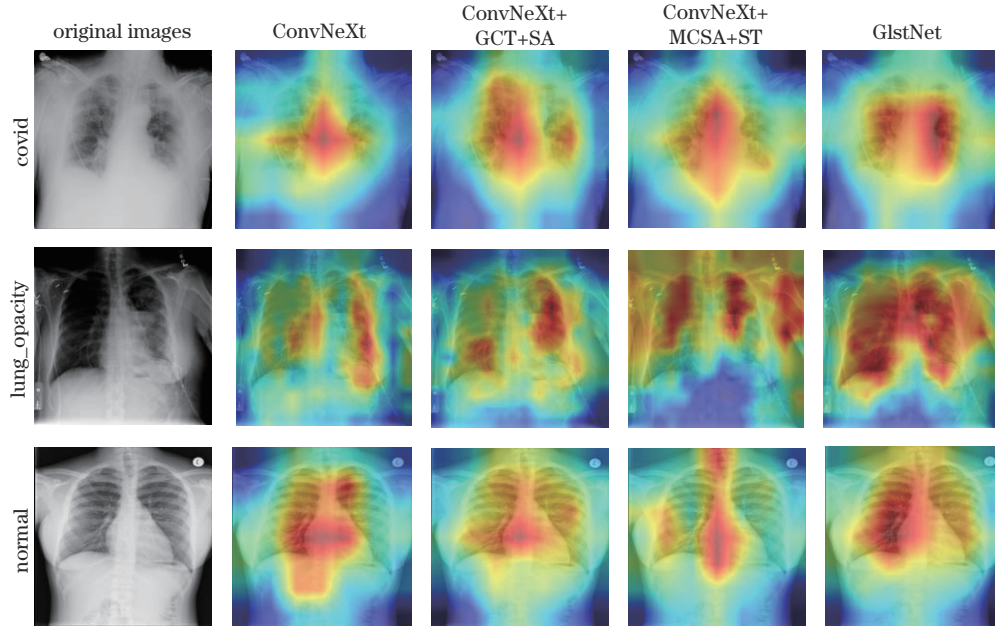


图 7 Score-CAM 可视化结果

Fig. 7 Score-CAM visualization results

## 4 结 论

针对 X 射线图像全局与局部信息的交互性问题, 提出 GlstNet 进行新冠肺炎 X 射线图像诊断。GlstNet 包含 3 个模块: ConvNeXt 主干网络、聚合注意力模块与 LSTM 网络。与传统 CNN 不同, 通道注意力不再采用卷积层与池化层, 通过对通道间的特征关系建模, 参数与卷积权重联合优化决定神经元的竞争与合作状态, 避免由小感受野卷积带来的部分信息丢失与局部

模糊。空间注意力保持相同感受野大小的同时, 通过改进卷积核大小与激活函数为模块带来更少参数量与更多的非线性。消融实验结果证明, 聚合注意力模块中通道注意力与空间注意力功能互补, 可更高效地结合全局与局部信息, 学习胸部 X 射线图像病灶区域。CNN 后接 LSTM 层, LSTM 可进一步提取深层次特征, 融合通道间高层次特征信息, 提高网络学习能力。实验结果表明, 聚合注意力模块、LSTM 网络与 ConvNeXt 可以较好地配合, 有效地利用胸部 X 射线图

像诊断新冠肺炎,具有较好的泛化性能。在两个新冠肺炎胸部 X 射线图像数据集上进行的实验结果表明了所提 G1stNet 的有效性,对新冠肺炎 X 射线图像具有较好的分类效果。

## 参 考 文 献

- [1] Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results[J]. *Expert Review of Molecular Diagnostics*, 2020, 20(5): 453-454.
- [2] Dino H I, Zeebaree S R M, Hasan D A, et al. COVID-19 diagnosis systems based on deep convolutional neural networks techniques: a review[C]//2020 International Conference on Advanced Science and Engineering (ICOASE), December 23-24, 2020, Duhok, Iraq. New York: IEEE Press, 2020: 184-189.
- [3] 冯毅博, 仇大伟, 曹慧, 等. 基于深度可分离稠密网络的新型冠状病毒肺炎 X 线图像检测方法研究[J]. *生物医学工程学杂志*, 2020, 37(4): 557-565.  
Feng Y B, Qiu D W, Cao H, et al. Research on coronavirus disease 2019 (COVID-19) detection method based on depthwise separable DenseNet in chest X-ray images[J]. *Journal of Biomedical Engineering*, 2020, 37(4): 557-565.
- [4] 杨锋, 魏国辉, 曹慧, 等. 基于内容的医学图像检索研究进展[J]. *激光与光电子学进展*, 2020, 57(6): 060003.  
Yang F, Wei G H, Cao H, et al. Research progress on content-based medical image retrieval[J]. *Laser & Optoelectronics Progress*, 2020, 57(6): 060003.
- [5] 朱江平, 王睿珂, 段智涓, 等. 基于多尺度注意力机制相位展开的三维人脸建模[J]. *光学学报*, 2022, 42(1): 0112005.  
Zhu J P, Wang R K, Duan Z J, et al. Three-dimensional face modeling based on multi-scale attention phase unwrapping[J]. *Acta Optica Sinica*, 2022, 42(1): 0112005.
- [6] Dastider A G, Sadik F, Fattah S A. An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound[J]. *Computers in Biology and Medicine*, 2021, 132: 104296.
- [7] Islam M Z, Islam M M, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images[J]. *Informatics in Medicine Unlocked*, 2020, 20: 100412.
- [8] Naeem H, Bin-Salem A A. A CNN-LSTM network with multi-level feature extraction-based approach for automated detection of coronavirus from CT scan and X-ray images[J]. *Applied Soft Computing*, 2021, 113: 107918.
- [9] Hassanien M A, Singh V K, Puig D, et al. Predicting breast tumor malignancy using deep ConvNeXt radiomics and quality-based score pooling in ultrasound sequences [J]. *Diagnostics*, 2022, 12(5): 1053.
- [10] Ketu S, Mishra P K. India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability[J]. *Soft Computing*, 2022, 26(2): 645-664.
- [11] Tan M X, Le Q V. Efficientnet: rethinking model scaling for convolutional neural networks[C]//Proceedings of the 36th International Conference on Machine Learning, June 9-15, 2019, Long Beach, California, USA. Copenhagen: MLR Press, 2019: 6105-6114.
- [12] Touvron H, Cord M, Sablayrolles A, et al. Going deeper with image transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 32-42.
- [13] Huang G, Sun Y, Liu Z, et al. Deep networks with stochastic depth[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9908: 646-661.
- [14] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: a survey[J]. *Computational Visual Media*, 2022, 8(3): 331-368.
- [15] 孙瑞, 孙晓兵, 刘晓, 等. 基于注意力机制的偏振成像目标分类方法[J]. *光学学报*, 2021, 41(16): 1611004.  
Sun R, Sun X B, Liu X, et al. Polarimetric imaging target classification method based on attention mechanism [J]. *Acta Optica Sinica*, 2021, 41(16): 1611004.
- [16] 徐聪, 王丽. 基于改进DeepLabv3+网络的图像语义分割方法[J]. *激光与光电子学进展*, 2021, 58(16): 1610008.  
Xu C, Wang L. Image semantic segmentation method based on improved DeepLabv3+ network[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 1610008.
- [17] 王选齐, 杨锋, 曹斌, 等. 卷积神经网络在甲状腺结节诊断中的应用[J]. *激光与光电子学进展*, 2022, 59(8): 0800002.  
Wang X Q, Yang F, Cao B, et al. Application of convolution neural network in diagnosis of thyroid nodules [J]. *Laser & Optoelectronics Progress*, 2022, 59(8): 0800002.
- [18] 李智唯, 曹慧, 杨锋, 等. 基于卷积神经网络的脑肿瘤分割的研究进展[J]. *激光与光电子学进展*, 2021, 58(24): 2400003.  
Li Z W, Cao H, Yang F, et al. Research progress of brain tumor segmentation based on convolutional neural network[J]. *Laser & Optoelectronics Progress*, 2021, 58(24): 2400003.
- [19] Tian C W, Xu Y, Li Z Y, et al. Attention-guided CNN for image denoising[J]. *Neural Networks*, 2020, 124: 117-129.
- [20] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer Vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 3-19.
- [21] Yang Z X, Zhu L C, Wu Y, et al. Gated channel transformation for visual recognition[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11791-11800.
- [22] Feng Y B, Yang X, Qiu D W, et al. PCXRNet: condense attention block and Multiconvolution spatial attention block for Pneumonia Chest X-Ray detection



- [EB/OL]. (2021-09-07) [2022-02-04]. [https://www.techrxiv.org/articles/preprint/PCXRNet\\_Condense\\_attention\\_block\\_and\\_Multiconvolution\\_spatial\\_attention\\_block\\_for\\_Pneumonia\\_Chest\\_X-Ray\\_detection/14904837](https://www.techrxiv.org/articles/preprint/PCXRNet_Condense_attention_block_and_Multiconvolution_spatial_attention_block_for_Pneumonia_Chest_X-Ray_detection/14904837).
- [23] Hu J, Shen L, Albanie S, et al. Gather-excite: exploiting feature context in convolutional neural networks[C]//NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, December 3-8, 2018, Montreal, Canada. New York: ACM Press, 2018: 9423-9433.
- [24] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [25] 刘东远, 张耀, 刘洋, 等. 基于 LSTM 循环神经网络的 fNIRS 脑功能成像滤波方法[J]. *中国激光*, 2021, 48(19): 1918007.  
Liu D Y, Zhang Y, Liu Y, et al. LSTM-based recurrent neural network for noise suppression in fNIRS neuroimaging: network design and pilot validation[J]. *Chinese Journal of Lasers*, 2021, 48(19): 1918007.
- [26] 潘信亮, 杨仁辉, 江涛, 等. 基于 Bi-LSTM 的近岸水体深度反演[J]. *光学学报*, 2021, 41(10): 1012003.  
Pan X L, Yang R H, Jiang T, et al. Depth inversion of coastal waters based on Bi-LSTM[J]. *Acta Optica Sinica*, 2021, 41(10): 1012003.
- [27] Rahman T, Khandakar A, Qiblawey Y, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images[J]. *Computers in Biology and Medicine*, 2021, 132: 104319.
- [28] Chowdhury M E H, Rahman T, Khandakar A, et al. Can AI help in screening viral and COVID-19 pneumonia? [J]. *IEEE Access*, 2020, 8: 132665-132676.
- [29] Liu Z, Mao H Z, Wu C Y, et al. A convnet for the 2020s [EB/OL]. (2022-01-10)[2022-03-04]. <https://arxiv.org/abs/2201.03545>.
- [30] Kumaraguru T, Abirami P, Darshan K M, et al. Smart access development for classifying lung disease with chest X-ray images using deep learning[J]. *Materials Today: Proceedings*, 2021, 47: 76-79.
- [31] Tabrizchi H, Mosavi A, Vamossy Z, et al. Densely connected convolutional networks (DenseNet) for diagnosing coronavirus disease (COVID-19) from chest X-ray imaging[C]//2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA), June 23-25, 2021, Lausanne, Switzerland. New York: IEEE Press, 2021.
- [32] Nedumkunnel I M, George L E, Sowmya K S, et al. Explainable deep neural models for COVID-19 prediction from chest X-rays with region of interest visualization [C]//2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), May 21-23, 2021, Jalandhar, India. New York: IEEE Press, 2021: 96-101.
- [33] Wang H F, Wang Z F, Du M N, et al. Score-CAM: score-weighted visual explanations for convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 111-119.