

# 用使用空间注意力机制和通道注意力机制的深度卷积神经网络从胸 X 光中检测 COVID-19

学生：张瑞阳，复旦大学计算机科学技术学院  
指导老师：张文强教授，复旦大学计算机科学技术学院

## 摘要

COVID-19 疫情持续对全人类的健康和生活产生消极影响。控制疫情的非常关键的一步就是能够有效的检测出感染者。最近，研究者提出了很多基于放射影像，包括胸 X 光和 CT，的使用计算机视觉对 COVID-19 进行检测的方法。有很多方法达到了很高的准确率，但它们不能从空间上和语义上注意到和 COVID-19 有关的信息。因此，在这个研究中，我提出了一个使用空间注意力机制和通道注意力机制的深度卷积神经网络，它能从胸 X 光中检测 COVID-19。我还提出了一个有 COVID-19、非 COVID-19 的肺炎、正常三类的平衡的胸 X 光数据集。通过实验，这个网络和之前的方法相比能够更好的从空间上和语义上注意到 COVID-19 病灶区。代码、数据集、训练好的模型见<https://github.com/Ruiyang-061X/DIP-Project>。

## 1. 介绍

COVID-19 疫情持续对全人类的健康和生活产生很大的消极影响。控制疫情的非常关键的一步就是能够有效的检测出感染者，从而方便这些感染者能够及早接受治疗，增加感染者在轻症阶段就能够得到有效治疗的比例，同时能够使这些感染者尽早的被隔离，防止这些感染者将疾病传染给他人。

目前，COVID-19 主要的检测手段是逆转录聚合酶链反应测试 (RT-PCR)，它能够将 COVID-19 病毒的 RNA 与其他病毒的 RNA 区分开来，从而实现了对 COVID-19 的精确检测。虽然逆转录聚合酶链反应测试非常精确，但是它非常耗时、复杂，而且从全球范围内来看，它是非常紧缺的。

COVID-19 的另外一类检测手段是使用放射影像，包括胸 X 光和 CT，放射科的医生通过仔细观察胸 X 光或者 CT 中肺部区域有没有 COVID-19 病灶区也可以判断是否感染了 COVID-19。在胸 X 光和 CT 中，CT 的结果会更加准确，但胸 X 光有着相当多的优点，例如速度很快，几乎所有的医疗系统中都有，可以移动。

最近，研究者提出了很多使用胸 X 光或者 CT 检测 COVID-19 的方法，其中的有些方法达到了很高的准确率。但是这些方法有时会根据其他的信息做出判断，例如，根据一些和 COVID-19 有关的标记判断是

COVID-19，根据胸 X 光中有很多的治疗仪器判断有病。这些方法的问题就是，它们不能从空间上，也不能从语义上注意到 COVID-19 病灶区。

之前的使用胸 X 光的工作中使用了大量不同的数据集。有些工作使用了 Pediatric Pneumonia Chest X-ray[3]，这些工作会从里面取非 COVID-19 的肺炎和正常数据，但实际上这个数据集是小孩的胸 X 光数据集，用这个数据集很容易导致训练出来的模型只根据胸部的大小来判断是 COVID-19 还是非 COVID-19 的肺炎和正常 [5]。有些工作使用了大量的非 COVID-19 的肺炎和正常数据，例如几千个，而 COVID-19 的数据很少，只有几百个，我认为在一个简单的三分类问题上使用如此不平衡的数据集是不合理的 [15]。有些工作使用的测试集中非 COVID-19 的肺炎和正常数据有几百个，而 COVID-19 只有几十个，这样很容易导致非 COVID-19 的肺炎和正常的准确率很高，COVID-19 的准确率较低，最终的准确率被拉得很高的情况 [8]。

提出的模型的结构见图 1。提出的模型使用的骨架模型是 DenseNet121，DenseNet121 有四个 Dense Block，Dense Block 中的层之间都是密集连接的。每个 Dense Block 之后有一个 Transition，它会对提取出来的特征进行降采样。最后的 Linear 层会将输入分类 COVID-19、非 COVID-19 的肺炎、正常三类。在提出的模型中加入了空间注意力机制和通道注意力机制，每个 Dense Block 之后都有一个 Channel Attention 和一个 Spatial Attention，其中 Channel Attention 实现了通道注意力机制，Spatial Attention 实现了空间注意力机制。其中空间注意力机制有助于模型在空间上更好的注意到 COVID-19 病灶区，通道注意力机制有助于模型在语义上更好的注意到 COVID-19 病灶区。

这个工作的主要贡献如下：

- 提出了一个使用空间注意力机制和通道注意力机制的深度卷积神经网络。
- 提出了一个有 COVID-19、非 COVID-19 的肺炎、正常三类的平衡的胸 X 光数据集。
- 通过实验，提出的模型达到了较高的准确率，而且和之前的模型相比能够更好的从空间上和语义上注意到 COVID-19 病灶区。

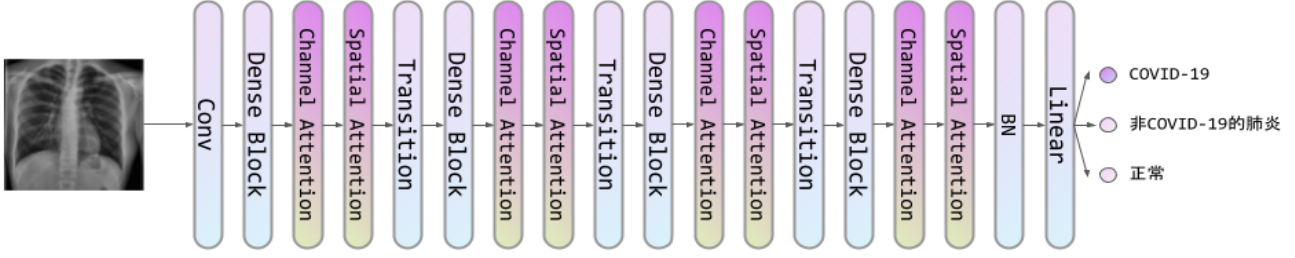


Figure 1. 提出的模型的结构。骨架模型为 DenseNet121[11]，它有四个 Dense Block，每个 Dense Block 之后有一个 Transition，最后的 Linear 层将输入分为 COVID-19、非 COVID-19 的肺炎、正常三类。在其中加入了空间注意力机制和通道注意力机制，每个 Dense Block 之后加上了一个 Channel Attention 和一个 Spatial Attention。

## 2. 相关工作

**使用胸 X 光检测 COVID-19 的相关工作** Wang et al. [17] 提出了 COVID-Net，其中大量使用了 PEPX 模块，这种模块能在计算效率和表示能力之间得到一个很好的平衡。Luz et al. [14] 使用了 EfficientNet、MobileNet 这些轻量级的模型进行 COVID-19 的检测，其目的主要是探索在保持较高的计算效率的同时，取得较好的结果。Rahimzadeh et al. [16] 将 Xception 和 ResNet50 V2 两个模型并行起来进行训练，同时并行来进行测试，最终并行起来的模型的结果超过了 Xception 和 ResNet50 V2 所能达到的结果。Md. Karim et al. [12] 在使用模型进行分类的同时，做了大量可视化和解释性工作。Li et al. [13] 训练了很多轻量级模型，并将它们部署到移动设备上，他还提出了一个使用移动设备拍摄现有数据集形成的数据集，用来模拟移动设备的使用环境。以上工作都没能将注意力机制应用到使用胸 X 光检测 COVID-19 中去。

**计算机视觉中的注意力机制** SENet[10] 中，对提取出来的特征做全局平均池化，然后使用一个线性层将通道数降下来，再使用一个线性层将通道数还原，通过一个 Sigmoid 层，然后将输出和原特征相乘，得到最终结果，从而形成了通道注意力机制。CBAM[19] 在通道注意力机制的基础上，对提取出来的特征做全局平均池化和全局最大池化，并将两个输出并起来，然后通过一个卷积层，通过一个 Sigmoid 层，然后将输出和原特征相乘，得到最终结果，从而形成了空间注意力机制。Non-local Neural Network[18] 提出了非局部注意力机制，这种注意力机制能考虑到全局中任意一个部分对另外一个部分的影响。DANet[7] 在 CBAM[19] 的基础上，引入了 Non-local Neural Network[18]，它将 CBAM[19] 中的注意力机制，替换为 Non-local Neural Network[18] 中的非局部注意力机制。以上工作没能在最近才出现的使用胸 X 光检测 COVID-19 的任务中得到应用。

## 3. 方法

在这部分中，首先会介绍使用的骨架模型，然后会介绍 CBAM[19] 这种注意力机制，最后会描述最终模

型的结构。

### 3.1. 骨架模型

这个工作中使用的骨架模型是 DenseNet121[11]。DenseNet[11] 是在 ResNet[9] 基础上进行改进的模型。对于 ResNet 来说，它的最大特点是，当前层的输入会直接跳过当前层连接到当前层的输出上并于当前层的输出相加，从而形成一种跳连接，这种跳连接会有利于模型的反向传播。而对于 DenseNet，它将 ResNet 的跳连接改成了当前层和当前层之前的所有层的输入会直接连接到当前层的输出上，从而形成一种密集连接，这也是 DenseNet 名字的由来，与此同时，它将 ResNet 的连接时的相加改成了连接时在通道维上并在一起，同时，为了让整个模型有降采样的过程，在密集连接的层之间会有一些操作。

具体来说，密集连接的层被称为密集块，记  $l^{th}$  层的输出为  $x_l$ ， $l^{th}$  层的运算是  $H_l(\cdot)$ ，则有：

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (1)$$

其中  $H_l(\cdot)$  有三层，包括批归一化层、修正线性单元、 $3 \times 3$  的卷积层。密集块之间的被称为过渡层，它有三层，包括批归一化层、 $1 \times 1$  的卷积层、 $2 \times 2$  的平均池化层。

图 2 是一个四层的密集块的结构。其中的连线均为特征和特征在通道维上的并。

### 3.2. CBAM

CBAM[19] 即卷积块注意力模块，是一种计算机视觉中的注意力机制。它既使用了通道注意力机制，也使用了空间注意力机制。对于卷积层提取出的特征，它会先使用通道注意力机制，再使用空间注意力机制。

具体来说，记  $F$  是卷积层提取出的特征， $M_c(F) \in \mathbb{R}^{C \times 1 \times 1}$  是求出的通道注意力图，则：

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \end{aligned} \quad (2)$$

其中  $F_{avg}^c$  为  $F$  在通道维做平均池化的结果， $F_{max}^c$  为  $F$  在通道维做最大池化的结果， $MLP$  有三层，包括线性

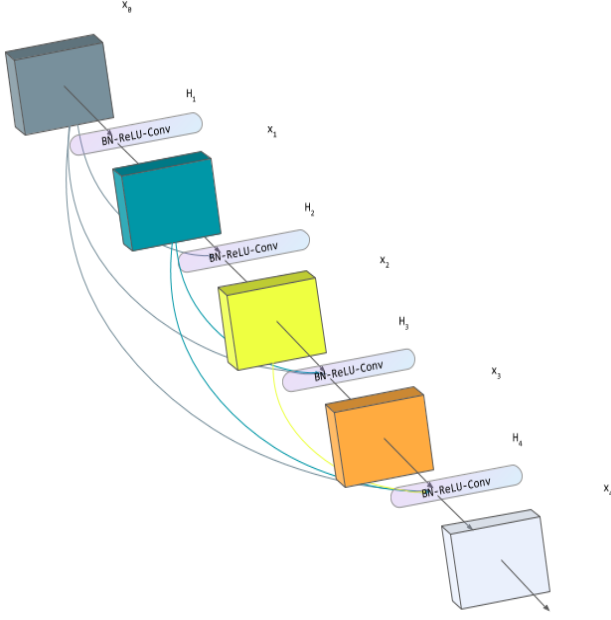


Figure 2. 一个四层的密集块的结构。

层、修正线性单元、线性层，其中第一个线性层的参数为  $W_0 \in \mathbb{R}^{C/r \times C}$ ，第二个线性层的参数为  $W_1 \in \mathbb{R}^{C \times C/r}$ ，其中  $r$  为减少率， $\sigma$  为 sigmoid 函数。

记  $M_s(F) \in \mathbb{R}^{1 \times H \times W}$  为求出的空间注意力图，则：

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (3) \\ &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \end{aligned}$$

其中  $F_{avg}^s$  为  $F$  在高宽维做平均池化的结果， $F_{max}^s$  为  $F$  在高宽维做最大池化的结果， $f^{7 \times 7}$  为一个卷积核大小为  $7 \times 7$  的卷积运算， $\sigma$  为 sigmoid 函数。

图 3 是通道注意力机制和空间注意力机制的结构。

### 3.3. 最终模型

最终模型的结构见图 1。最终模型使用的骨架模型使 DenseNet121，DenseNet121 有四个 Dense Block，Dense Block 中的层之间都是密集连接的。每个 Dense Block 之后有一个 Transition，它有三层，包括批归一化层、 $1 \times 1$  的卷积层、 $2 \times 2$  的平均池化层，它会对提取出来的特征进行降采样。最后的 Linear 层会将输入分为 COVID-19、非 COVID-19 的肺炎、正常三类。每个 Dense Block 之后都有一个 Channel Attention 和一个 Spatial Attention，它们的结构同上所述，从而加入了空间注意力机制和通道注意力机制。其中空间注意力机制有助于模型在空间上更好的注意到 COVID-19 病灶区，通道注意力机制有助于模型在语义上更好的注意到 COVID-19 病灶区。

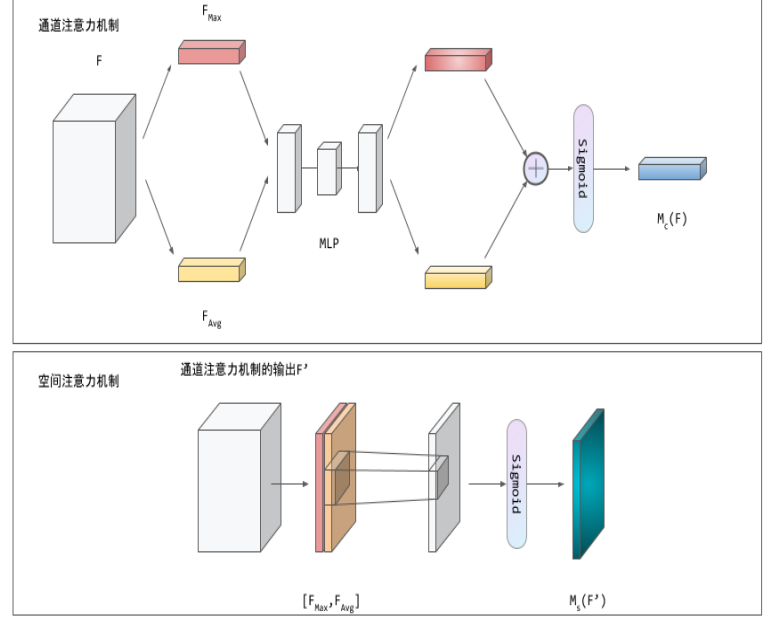


Figure 3. 通道注意力机制和空间注意力机制的结构。

## 4. 实验

### 4.1. 数据集

提出了一个平衡的胸 X 光数据集。这个数据集来自四个数据集，包括 covid-chestxray-dataset[6]、Figure1-COVID-chestxray-dataset[2]、Actualmed-COVID-chestxray-dataset[1]、rsna-pneumonia-detection-challenge[4]。构造这个数据集的时候，取出了 covid-chestxray-dataset[6] 中所有疾病为 COVID-19 并且视角为 PA 的胸 X 光，取出了 Figure1-COVID-chestxray-dataset[2] 中所有的数据，取出了 Actualmed-COVID-chestxray-dataset[1] 中所有的数据，随机取出了 rsna-pneumonia-detection-challenge[4] 中的一部分的非 COVID-19 的肺炎、正常数据。最终的数据集一共有 1425 个数据。有三类数据，其中 COVID-19 有 475 个数据，非 COVID-19 的肺炎有 475 个数据，正常有 475 个数据。这个数据集按照 7:1:2 的比例划分成训练集、验证集、测试集，其中训练集有 993 个数据，验证集有 144 个数据，测试集有 288 个数据。

这个数据集没有从 Pediatric Pneumonia Chest X-ray[3] 中取非 COVID-19 的肺炎和正常的数据，不会使训练出来的模型根据胸部的大小判断是 COVID-19 还是非 COVID-19 的肺炎和正常。这个数据集的测试集是平衡的，不会出现某一类的数据很多，并且准确率很高，进而把整体准确率拉的很高的情况。这个数据集是平衡的，我认为在一个只有三类的分类任务上这是合理的。

## 4.2. 实现细节

训练时使用的批大小为 32。使用的 DenseNet121 有 ImageNet 的预训练参数。损失函数为交叉熵损失函数。使用的优化器为 Adam，学习率设置为  $2e-4$ ， $\beta_s$  设置为 (0.9,0.999)。训练轮数为 50。每轮训练结束后在验证集上进行验证，得到模型在验证集上的准确率。保存下来在验证集上准确率最高的模型，并保存下来对应的准确率。

## 4.3. 结果

为了产生可比性和比较的公平性，在选择进行比较的工作时使用了没有使用 Pediatric Pneumonia Chest X-ray[3] 这个数据集、测试集比较平衡，整个数据集比较平衡的原则。最终只找到一个可以进行比较的工作，Li et al. [13]。

比较的情况见表 1。DenseNet121-CBAM 的准确率比 Li et al. [13] ShuffleNetV2 的准确率低 4.6%，原因应该是 Li et al. [13] ShuffleNetV2 使用了 ChestX-ray8 这个数据集进行了预训练，这个工作没有足够的时间在 ChestX-ray8 这个 42G 的数据集上做预训练，所以准确率会比 Li et al. [13] ShuffleNetV2 的准确率低 4.6%。

表 2 是各类的准确率。其中 COVID-19 的准确率达到了 100.0%，说明模型对 COVID-19 有很高的识别能力。

图 4 是 Li et al. [13] ShuffleNetV2 和 DenseNet121-CBAM 在 COVID-19 胸 X 光上的 CAM 比较，左侧的是 Li et al. [13] ShuffleNetV2 的 CAM，右侧的是 DenseNet121-CBAM 的 CAM。从图中可以看到，Li et al. [13] ShuffleNetV2 注意的是一整块区域，DenseNet121-CBAM 注意的是与 COVID-19 密切相关的地方。说明虽然 Li et al. [13] ShuffleNetV2 的准确率较高，DenseNet121-CBAM 的准确率相对较低，但是

方法	准确率
Li et al. [13] SqueezeNet	82.4
Li et al. [13] MobileNetV2	88.0
Li et al. [13] ShuffleNetV2	90.7
DenseNet121-CBAM	86.1

Table 1. 与 Li et al. [13] 的比较。

COVID-19	非 COVID-19 的肺炎	正常
100.0	75.0	83.3

Table 2. 各类的准确率。

方法	准确率
DenseNet121	83.3
DenseNet121-CBAM	86.1

Table 3. DenseNet121 和 DenseNet121-CBAM 的比较。

DenseNet121-CBAM 能更好的从空间上和语义上注意到 COVID-19 病灶区。

## 4.4. 消融实验

为了研究空间注意力机制和通道注意力机制的有无对准确率的影响，进行了一组 DenseNet121 和 DenseNet121-CBAM 的对比实验。

比较的情况见表 3。DenseNet121-CBAM 的准确率比 DenseNet121 的准确率高 2.8%，说明空间注意力机制和通道注意力机制可以提高模型的准确率。

图 5 是 DenseNet121 和 DenseNet121-CBAM 在 COVID-19 胸 X 光上的 CAM 比较，左侧的是 DenseNet121 的 CAM，右侧的是 DenseNet121-CBAM 的 CAM。从图中可以看到，DenseNet121 注意的区域比较散乱，DenseNet121-CBAM 注意的是与 COVID-19 密切相关的地方。说明空间注意力机制和通道注意力机制不仅可以提高模型的准确率，而且能使模型更好的从空间上和语义上注意到 COVID-19 病灶区。

## 5. 结论

这个工作中，提出了一个使用空间注意力机制和通道注意力机制的深度卷积神经网络，它能将胸 X 光分类为 COVID-19、非 COVID-19 的肺炎、正常。同时，还提出了一个有 COVID-19、非 COVID-19 的肺炎、正常三类的平衡的胸 X 光数据集。通过实验，这个模型达到了较高的准确率，并且和之前的工作相比能够更好的从空间上和语义上注意到 COVID-19 病灶区。通过消融实验，空间注意力机制和通道注意力机制确实可以提高模型的准确率。

## References

- [1] Actualmed-covid-chestxray-dataset. 2020.
- [2] Figure1-covid-chestxray-dataset. 2020.
- [3] Pediatricpneumoniachestx-ray. 2020.
- [4] rsna-pneumonia-detection-challenge. 2020.
- [5] Soumick Chatterjee, Fatima Saad, Chompunuch Sarasaen, Suhita Ghosh, Rupali Khatun, Petia Radeva, Georg Rose, Sebastian Stober, Oliver Speck, and Andreas Nürnberger. Exploration of interpretability techniques for deep covid-19 classification using chest x-ray images. ArXiv, abs/2006.02570, 2020.
- [6] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. arXiv 2003.11597, 2020.
- [7] Jun Fu, Jing Liu, Haijie Tian, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3141–3149, 2019.
- [8] Brian D. Goodwin, Corey Jaskolski, Can Zhong, and Herick Asmani. Intra-model variability in covid-19 classification using chest x-ray images. ArXiv, abs/2005.02167, 2020.

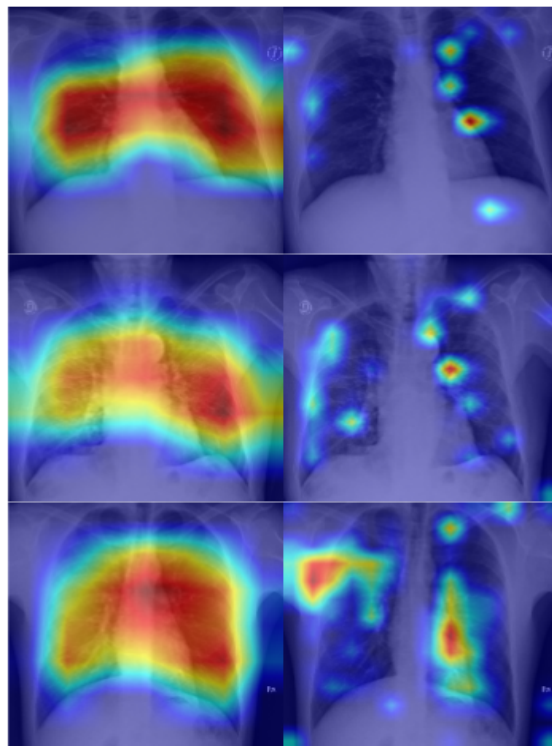


Figure 4. Li et al. [13] ShuffleNetV2 和 DenseNet121-CBAM 在 COVID-19 胸 X 光上的 CAM 比较。左侧的是 Li et al. [13] ShuffleNetV2 的 CAM，右侧的是 DenseNet121-CBAM 的 CAM。

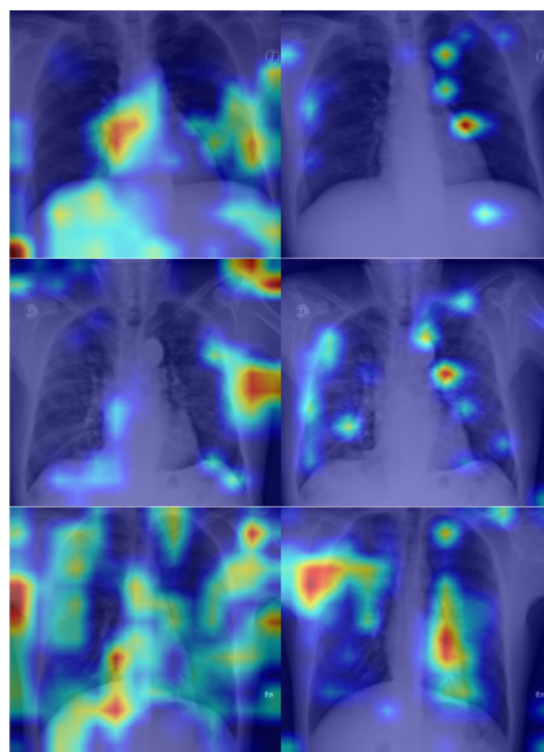


Figure 5. DenseNet121 和 DenseNet121-CBAM 在 COVID-19 胸 X 光上的 CAM 比较。左侧的是 DenseNet121 的 CAM，右侧的是 DenseNet121-CBAM 的 CAM。



- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016.
- [10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. IEEE transactions on pattern analysis and machine intelligence, 2019.
- [11] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2261–2269, 2017.
- [12] Md. Rezaul Karim, Till Döhmen, Dietrich Rebholz-Schuhmann, Stefan Decker, Michael Cochez, and Oya Deniz Beyan. Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images. ArXiv, abs/2004.04582, 2020.
- [13] Xuanying Li, Chengyin Li, and Dongxiao Zhu. Covid-mobilexpert: On-device covid-19 screening using snapshots of chest x-ray. ArXiv, abs/2004.03042, 2020.
- [14] Eduardo Luz, Pedro Henrique Lopes da Silva, Rodrigo Ferreira da Silva, Ludmila Correa de Alkmin Silva, Gladston J. P. Moreira, and David Menotti. Towards an effective and efficient deep learning model for covid-19 patterns detection in x-ray images. ArXiv, abs/2004.05717, 2020.
- [15] Sampa Misra, Seungwan Jeon, Seiyon Lee, Ravi Managuli, and Chulhong Kim. Multi-channel transfer learning of chest x-ray images for screening of covid-19. ArXiv, abs/2005.05576, 2020.
- [16] Mohammad Rahimzadeh and Abolfazl Attar. A new modified deep convolutional neural network for detecting covid-19 from x-ray images. ArXiv, abs/2004.08052, 2020.
- [17] Lechen Wang and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. 2020.
- [18] Xiaolong Wang, Ross B. Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7794–7803, 2018.
- [19] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In-So Kweon. Cbam: Convolutional block attention module. ArXiv, abs/1807.06521, 2018.