

# JCS: 一种基于联合分类分割的可解释性新冠肺炎筛查系统

吴宇寰, 高尚华, 梅杰, 徐君, 范登平, 张荣国, 程明明

**Abstract**—近来, 新冠肺炎 (COVID-19) 在 200 多个国家中流行, 影响了上亿人。为了控制肺炎传播, 识别和分离感染人群是重要的一步。诊断新冠肺炎最主要的工具是逆转录聚合酶链式反应测试 (RT-PCR)。尽管如此, RT-PCR 检测的灵敏度还不足以有效地预防流感大流行。胸部 CT 检查为 RT-PCR 检测提供了一个有价值的补充工具, 对早期诊断具有较高的敏感性。然而, 胸部 CT 检测常常消耗大量时间, 需要平均 21.5 分钟来诊断一个病例。本文开发了一个新的联合分类与分割系统 (JCS) 来进行实时、可解释的 COVID-19 胸部 CT 诊断。为了训练我们的 JCS 系统, 我们建立了一个大型 COVID-19 分类与分割数据集 (COVID-CS), 其中包含 400 个 COVID-19 病人和 350 个未感染者的 144,167 张胸部 CT 图像。3,855 对 200 例患者的胸部 CT 图像进行了精细的像素级阴影标记, 这些阴影是肺实质的密度增加。我们还对病变计数、混浊区域和位置进行了注释, 从而有利于各种诊断方面。大量实验表明, 本文提出的 JCS 诊断系统对 COVID-19 的分类和分割是非常有效的。它在分类测试集上获得了 95.0% 的平均敏感性和 93.0% 的特异性, 在我们的 COVID-CS 数据集的分割测试集上获得了 78.5% 的 Dice 分数。COVID-CS 数据集与代码公开在 <https://github.com/yuhuan-wu/JCS>。

**Index Terms**—新冠肺炎, 联合诊断, 新冠肺炎胸部 CT 数据集。

## I. 介绍

新冠肺炎或称 COVID-19, 是一种由 SARS-CoV-2 引起的流行病。它在全球范围内短时间内爆发, 截至 2020 年 4 月 15 日已造成 1914916 例确诊病例和 123010 例确诊死亡。由于缺乏医疗用品和工作人员, COVID-19 将 200 多个国家的卫生系统推向崩溃的边缘, 因此被世界卫生组织宣布为大流行 [2]。目前对 COVID-19 的主要诊断手段是通过逆转录聚合酶链式反应 (RT-PCR) 检测 [3]。然而, RT-PCR 检测不够准确, 无法很好地预防新冠肺炎大流行 [4], [5]。因此, RT-PCR 检测的假阴性 (false-negative) 病例是对公众健康的潜在威胁, 遗漏任何 COVID-19 病例都可能导致大面积的继发感染。

为了阻止 COVID-19 的迅速传播, 医学影像被用作 RT-PCR 检测的重要补充工具 [6]。这是因为大多数

本文为 [1] 的翻译版。如对该翻译版有疑问请邮件咨询吴宇寰 (wuyuhuan@mail.nankai.edu.cn)。

(通讯作者: 程明明, cmm@nankai.edu.cn)

吴宇寰、高尚华、梅杰、徐君、范登平和程明明来自南开大学。

张荣国来自北京推想科技。

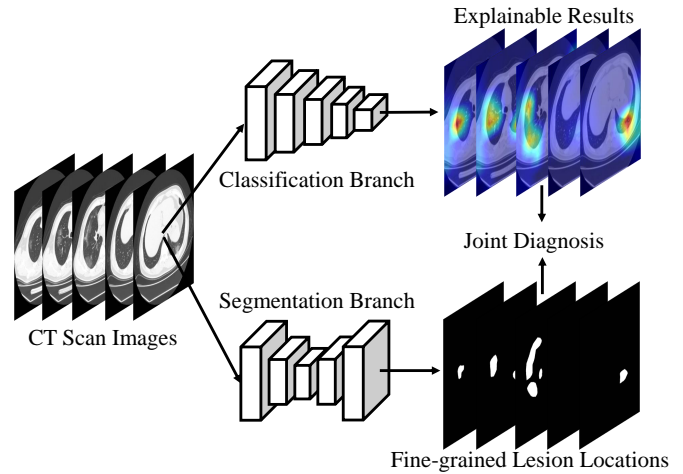


Figure 1. 我们针对 COVID-19 的 JCS 诊断系统的图示。我们的 JCS 系统将只在分类分支报告阳性 COVID-19 预测时执行分割诊断。

COVID-19 患者的胸部 X 射线 (X-ray) 影像临床表现出肺部感染 [7]。[4], [5] 的工作显示了 CT 检测具有较高敏感度。此外, CT 影像扫描测试是监测疾病的严重性的必要工具 [8]。然而, CT 影像扫描测试的诊断时间是其主要的限制: 经调查, 即使是有充分经验的放射科医生也需要 21.5 分钟 [9] 来分析每个病例的检测结果。而在传染病大流行爆发期间, 经验丰富的放射科医生严重缺乏, 难以及时识别潜在感染患者。因此, 自动诊断系统的需求十分迫切。

由于深度卷积神经网络 (CNNs) 强大的识别能力, 人工智能 (AI) 技术正在改变医学影像处理界。深层 CNN 通常在大规模数据集上进行训练以证明其能力。然而, 由于大多数现有的 COVID-19 的 CT 影像扫描数据集 [10]–[13] 仅包含几十个病例的百余张 CT 图像, 故这些数据集不能满足上述需求。而且, 大多数数据集仅提供患者级别的标签 (即类标签) 表明此人是否被感染, 但这些数据集缺少像素级别的精细标注。因此, 用这些数据集训练的 CNN 模型通常忽略了解释最终预测的有价值的信息。尽管已经建立了多个 CT 影像扫描诊断系统 [5], [14]–[18] 来检测疑似 COVID-19 病例, 但大多数都有两个缺点 1) 它们是在小规模的数据集上训练的, 因此对于多样的 COVID-19 感染检测来说不够鲁棒。2) 它们基于黑箱 CNNs 进行分类, 但缺

乏可解释的透明性来辅助医生进行诊断。

为了缓解上述缺点,在这篇论文中,我们 1) 建立了一个包含标签级别的和像素级别的大规模数据集 *COVID-CS* 并 2) 提出了一种基于诊断系统的联合分类分割模型 (*JCS*) 来为与 COVID-19 战斗的医护人员提供可解释的诊断结果。具体来说,我们利用我们收集的包含数百个病例的上千张图像的 *COVID-CS* 数据集训练我们的 *JCS* 系统来获取更好的诊断性能。如图 1 所示,我们的 *JCS* 诊断模型首先通过分类分支辨识 COVID-19 的疑似病例然后通过激活图技术 [19] 解释我们 *JCS* 诊断模型是根据哪些位置判断这些病例可能是 COVID-19 病例。此外,我们的系统可以通过细粒度的图像分割技术,在肺部造影中发现 COVID-19 感染的位置和区域。利用可解释的分类结果和相应的细粒度病变分割,我们的 *JCS* 系统大大简化和加速了放射科医生或其他医学专家的诊断过程。如表 Table II 中所示,我们的系统仅需 22s 来判断一个感染病例,仅需 1s 来排除一个未感染病例,比经验丰富的放射学家的 RT-PCR 和 CT 影像扫描分析快得多。在我们 *JCS* 系统的协助下,有经验的放射科医生仅需要 55.4s (32.4s 用于医生诊断, 22.0s 用于 *JCS* 推理) 判断一个感染病例,需 1s 来排除一个未感染病例,并保持与纯人工相同的特异性和敏感性。因此,基于 *JCS* 辅助的诊断速度提升和诊断精度的相同表明了我们 *JCS* 系统的优越性。

总而言之,我们的主要贡献有以下三点:

- 我们建立了一个大规模 COVID-19 数据集,名为 *COVID-CS*, 其中包含了从 200 个新冠病人得到的 CT 图像的 3,855 张精细的像素级别标签,从另外 200 个新冠病人的 75,541 张 CT 图片,以及 350 个未感染病例的 75,541 个病人级别的标注。
- 我们开发了一种新型 COVID-19 的诊断系统,其提供了可解释的联合分类和准确的病变分割,清晰展示了我们系统相较之前系统的优越性。
- 在我们的 *COVID-CS* 数据集上,我们的 *JCS* 系统对于新冠肺炎达到了 95.0% 的敏感度、93.0% 的特异性 和 78.5% 的 Dice 得分,超越了其他 SOTA 的分割方法。

论文的后续部分如下。在 §II, 我们简要总结了相关工作。在 §III, 我们介绍了现有的用于识别与分析新冠病例的诊断系统。在 §IV, 我们提供了我们 *COVID-CS* 数据集与其标注过程的细节。广泛的实验在 §V 中, 实验评估了我们系统在识别 COVIN-19 的性能, 并做了深入分析。§VI 对工作进行了总结。

Table I  
不同数据集汇总 (截至 2020/4/10)

Dataset	Date	Link	Type	#Images	#Cases
PLXR [12]	2020/03/23	Link	X-rays	98	70
8023Dataset [10]	2020/03/25	Link	X-rays	229*	-
CTSeg [13]	2020/03/28	Link	CT	110	60
COVID-CT [11]	2020/03/30	Link	CT	746*	-
<b>COVID-CS (Ours)</b>	2020/04/12	-	CT	>144K <sup>†</sup>	<b>750</b>

\*: 这个数字是从作者的 GitHub 存储库报告的。

<sup>†</sup>: 在我们数据集中, 200 个阳性样本的 3,855 张图像被在像素级别进行了标注。64,771 张另外 200 个阳性样本的图像被在病人级别进行了标注, 另有 75,541 张从 350 个阴性样本得到的图像。

Table II

不同方法诊断 COVID-19 的平均时间。“CT R.” 代表 CT 专家单独诊断, “CT R. + *JCS*” 表示我们的 *JCS* 系统与 CT 专家共同诊断。

Method	RT-PCR	CT R.	CT R. + <i>JCS</i>	<i>JCS</i>
Time	~4h [20]	21.5min [9]	1s <sup>†</sup> /54.4s	1s <sup>†</sup> /22.0s

<sup>†</sup>: 诊断未感染病例。

## II. 相关工作

### A. 现有的 COVID-19 数据集

截至 4.15, 累计 191 4916 人感染新冠肺炎。然而 CT 检测通常是私有的, 不能被无法公开访问。为了促进诊断系统的发展, 世界各地的研究人员公开发布了一些与 COVID-19 相关的数据集。这些数据的汇总在表 I 中。

Cohen 等人的 X 射线数据集 [10] 包含了总共 122 张正面视图 X 射线, 包括 100 张 COVID-19 病例的图像、11 张 SARS 图像和 11 张其他肺炎图像。[11] 的 COVID-CT 数据集含有 746 张 CT 影像, 其中 349 张来自新冠患者 397 张来自非新冠病例。这些数据集中的所有图像都是从公共网站和/或在 medRxiv、bioRxiv 和期刊中 COVID-19 相关论文中收集的, 通过阅读论文中的图注, 选择包含 COVID-19 异常的 CT。COVID-19 数据集的其他一些资源有 PLXR [12] 和 CTSeg [13], 分别包含 98 和 110 个 CT 影像扫描图像病例。这些数据规模小, 缺乏多样性, 因为它们通常包含几十个病例的不到数百张图像。为了充分发挥深层 CNNs 的强大功能, 构建一个大规模的数据集来训练 COVID-19 系统中的深层 CNNs 是非常必要的。

### B. 人工 COVID-19 诊断

防止 COVID-19 扩散的最关键步骤是立即将每个患者与正常人区分开来。错过任何一个病人都可能导致大面积的继发性 COVID-19 感染。目前, 主要的手工诊断工具是 RT-PCR 检测 [21]。然而, RT-PCR 检测的敏感性还不足以有效地预防大流行 [4], [5]。CT 影像扫描在许多医院广

泛使用, 是 RT-PCR 检测的补充工具。但部分经 RT-PCR 检测呈阳性的特殊病例 CTs 正常 [22]–[24]。结合这两种测试可以最大限度地识别潜在的感染者, 因为它可以高灵敏度地识别早期的 COVID-19 患者 [4], [5], [25]。CT 影像扫描对于监测疾病的严重程度也是必要的 [8]。在大流行爆发期间, 经验丰富的医务人员严重缺乏, 难以及时确定可能感染的病人。因此, 人们迫切需求自动诊断系统。

### C. COVID-19 自动检测系统

目前大多数医学成像系统都是针对存在多年的常见疾病开发的, *e.g.*, 结核病 [26]。这些已开发的系统可以直接修改, 以缓解 COVID-19 爆发。医生们发现 COVID-19 患者的胸部 X 光片显示出某些异常。在 ResNet-50 [27] 的基础上, [28] 提出了 COVID-ResNet 来鉴别三种类型的 COVID-19 感染与正常肺炎患者。在 1531 例胸部 X 线片上, 张 *et al.* 提出了一种用于 COVID-19 筛查的深度异常检测系统, 灵敏度为 96.0%, 特异性为 70.65%。Yang *et al.* [29] 提出了一个评价 102 例志愿者影像的系统, 其敏感性为 83.3%, 特异性为 94.0%。Li *et al.* [30] 设计的系统利用肺 CT 轴位和冠状位严重度指数 (CTSI) 识别 78 例 COVID-19 患者, 敏感性为 82.6%, 特异性为 100.0%。Chung *et al.* [15] 通过收集 21 例患者的资料证实, 目视检查有助于识别 COVID-19 病例, 并通过整体肺部严重程度评分 (LTSS) 预测严重程度。Bernheim *et al.* [16] 分析 121 例 COVID-19 患者, 由经验丰富的放射科医师进行目测, 将其分为早、中、晚期病例。Wang *et al.* [17] 通过对 90 例患者 366 次 CT 影像扫描的研究发现, COVID-19 疾病在感染后 6-11 天内会变得很严重。Shi *et al.* [18] 开发了一种影像学辅助诊断程序来检测 COVID-19 引起的肺炎。Fang *et al.* [5] 用基于 CTSI 的方法检测了 81 例患者, 灵敏度为 98.0%, 相比之下, RT-PCR 法仅为 71.0%。Zhou *et al.* [31] 对 62 例 COVID-19 患者进行了非对比 CTSI 检查, 证实 CT 辅助评价在进展期至早期诊断中有较好的检出准确性。尽管在一小部分样本上取得了成功, 但这些 COVID-19 诊断系统还没有经过大规模样本的测试。他们在诊断推理过程中不能提供有用的诊断证据。更多的工作可以参 [32]–[34] 的评论。

据我们所知, 只有两种方法通过像素级分割来提取感染区域。Rajinikanth *et al.* [35] 通过分水岭变换技术 [36] 进行分割, 其结果粗糙, 精度有限。Zhou *et al.* [37] 开发了一个具有注意力机制的 U-Net, 在 CTseg [13] 数据集上 Dice 得分为 69.1% 但它的训练和测试分割只有 88 和 22 张图片在这项工作中, 我们提出了一个诊断系统, 通过整合基于学习的分类和分割网络, 为医生提供可解释的诊断证据, 并改善诊断过程的用户交互方面。

### D. 深度分类分割方法

自从 ImageNet 数据集 [38] 放出, 深度卷积神经网络 (CNNs) 已经成为图像分类任务的主力军, 其性能也在不断提高。代表性的深度分类器, *e.g.*, AlexNet [39], VGGNet [40], ResNet [27], DenseNet [41], and Res2Net [42], 已经被广泛应用于其他计算机视觉任务的特征提取, 诸如图像分割 [43]–[45], 显著性物体检测 [46], 人脸识别 [47], 航空图像分析 [48], 风格迁移 [49], 特征匹配 [50], 人群计数 [49] 与图像恢复 [51], *etc.* 尽管这些分类器的表现能力令人印象深刻, 但分类过程并不能清楚地解释预测结果。

图像分割解决了像素级预测的问题。语义分割的目的是对自然图像中每个像素的语义标签进行分类 [52]。这方面的代表性工作有 CN [53] 和 DeepLab [54]。实例分割的重点是在图像中用一个唯一的实例标签和像素级掩码来区分每个语义实例 [55]–[57]。全景分割 [58] 融合了语义分割和实例分割, 对非对象进行语义分割 (天空、水面、草地等) 并对对象进行实例分割 (猫、狗、公共汽车等)。U-Net [59] 在医学图像分割中被广泛使用。其进一步扩展到三维 U-Net [60], TerausNet [61] 和 U-Net++ [62], 在多功能图像分割任务中带来了更好的性能。本文将图像分类与分割技术相结合, 开发了一种新的 covid-19 诊断系统。

## III. 我们的 COVID-19 诊断系统

混浊是 COVID-19 患者的基本 CT 表现 [63], 它被定义为肺实质的密度增高 [64]。我们的系统包含一个用于识别 COVID-19 混浊的可解释分类分支和一个用于发现混浊区域的分割分支。分类器是在许多图像上训练的, 这些图像具有低成本的患者级注释, 以及一些带有像素级注释的图像, 以获得更好的激活映射。利用精确标注的 CT 图像训练分割分支, 对病变进行细粒度分割。通过整合这两个模型, 我们的 JCS 系统为 COVID-19 提供了信息丰富的诊断结果。

### A. 可解释的分类器

由于 CNNs 具有很强的表示能力, COVID-19 感染能只通过患者水平的监督训练来预测。为此, 我们构建了一个由所提出的分类模型组成的分类分支, 使我们的 JCS 诊断系统具有对 COVID-19 患者的鉴别能力。

1) 通过分类器诊断 COVID-19: 根据患者的 CT 影像扫描图像, 预测其是否为 COVID-19 阳性是一项二分类任务。由于设计新的分类模型不是我们的重点, 因此我们构建了基于 Res2Net 网络 [42] 的分类器。Res2Net [42] 作为一个强大的网络, 具有比 ResNet 更强的多尺度表示能力。最后一层被修改为具有两个通道的完全连接层, 以输

出 COVID-19 感染与否的概率。如果感染通道的概率大于未感染通道的概率，则患者被诊断为 COVID-19 阳性，反之亦然。对于每一个病人，CT 图像被逐个送入分类模型。如果感染的 CT 图像数量高于阈值，则诊断为 COVID-19 阳性。

2) 类激活映射带来的可解释性: 由于 CNN 分类的诊断过程是在一个黑盒中进行的, 因此我们采用激活映射 [19] 来增加 COVID-19 诊断系统对其预测的解释透明度。分类网络的最后一个卷积层紧跟着全局平均池 (GAP) 层和完全连接层。通过 GAP 层, 我们的分类模型将特征大小从  $H \times W$  降到  $1 \times 1$ , 从而丧失了空间表示能力。通过激活映射 [19], 我们的系统找到了预测结果的响应区域。假设在 GAP 层之前特征区域的梯度与预测一致在 GAP 层之前的特征图包含了高层语义和位置信息。每个通道对应于不同语义线索的激活。通过特征映射预测概率的梯度得到激活映射。具体地说, 给定 COVID-19 分支  $y_p$  的预测和 GAP 之前的特征映射  $X$ , 第  $k$  个通道  $X$  的权重计算如下:

$$w_k = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \frac{\partial y_p}{\partial X_{i,j}^k}, \quad (1)$$

这里  $X_{i,j}^k$  是在特征图  $X$  的第  $k$  个通道中位置  $(i, j)$  的值。where  $X_{i,j}^k$  is the value at position  $(i, j)$  in the  $k$ -th channel of feature map  $X$ . 在式 (1) 中更大的梯度为某个通道生成更大的激活映射权重。COVID-19 病例的激活映射计算如下:

$$AM_p = \sum_k ReLU(w_k X^k). \quad (2)$$

如图 Fig. 9 所示, 激活图准确定位了 COVID-19 患者的混浊区域, 为我们的 JCS 系统预测提供了可解释的结果。

3) 通过图像混合缓和数据偏差: 通过使用我们的可解释分类模型, 我们的系统能只通过患者级别的注释进行训练。然而, 由于 CT 图像来自多个来源, 通过激活图观察到, 分类器可能会被训练以过度拟合不需要的区域 (e.g., 病变外的区域)。因此, 我们提出, 利用图像混合技术 [65] 帮助分类器关注 COVID-19 病例的病变区域。在训练过程中, 不同来源的 CT 图像和相应的病人级别注释是混合的。具体地, 对于两个随机取样的 CT 图像  $x_i$  和  $x_j$  ( $i \neq j$ ) 以及相应的标签  $\hat{y}_i$  和  $\hat{y}_j$ , 新混合的样本和相应的标签被写为:

$$\begin{aligned} x_{ij}^m &= \lambda x_i + (1 - \lambda) x_j, \\ \hat{y}_{ij}^m &= \lambda \hat{y}_i + (1 - \lambda) \hat{y}_j, \end{aligned} \quad (3)$$

这里  $\lambda \in [0, 1]$  是在 Beta 分布中生成的随机数, i.e.,  $\lambda \sim \text{Beta}(\alpha, \alpha)$ 。对于混合样本, 我们的分类模型被训练成更关注 COVID-19 病例的决定性病变区域, 而不是数据源中的

偏差。此外, 混合过程削弱了标签的置信度, 从而减轻了我们的系统从过度拟合。

4) 激活映射的像素级监督: 传统的分类模型只利用图像标签进行训练。它们的激活映射可能不准确, 因为这些模型会自动学习不同类别图像的差异。在我们提出的数据集中, 有成千上万的图像与像素级的注释为特定的混浊区域, 他们可以直接监督激活映射。出于上述观察和 [66] 的工作, 在训练阶段, 我们将分割损失  $L_{seg}$  应用于 COVID-19 类通道的激活映射:

$$L_{seg} = \frac{1}{HW} \|AM_{p,c}^{norm} - S\|_2, \quad (4)$$

这里  $AM_{p,c}^{norm}$  是归一化为  $(0, 1)$  的 COVID-19 类通道的激活映射,  $S$  是二值的客观事实的像素级注释图,  $\|\cdot\|_2$  表示  $l_2$  范数。如果图像没有客观事实的像素级注释,  $L_{seg}$  将不会被计算。在应用分割损失  $L_{seg}$  之后, 图 Fig. 9 的结果表明, 激活映射显著提高了定位混浊的能力。

## B. 精准分割

我们的分割分支旨在从 COVID-19 患者的 CT 图像中发现确切的病变区域。Fig. 2 显示了我们的结合了/未结合分割和分类模型分割分支的结构。这种组合的细节如图 Fig. 4 所示。

1) 编码-解码结构: 我们的分割模型由一个编码器和一个解码器组成。

**编码器.** 编码器基于 VGG-16 [40] 主干, 去掉了最后 2 层全连接层。5 个 VGG 块分别定义为  $\{E_1, E_2, E_3, E_4, E_5\}$  VGG-16 主干首先输入 CT 图像, 并从五个 VGG 块的最后一层生成多尺度特征图。为了将输入特征图缩小一半, 每个块的前面 (第一个块除外) 是一个步长为 2 的 *max pooling* 函数。块  $E_4$  生成的特征图包含分辨率最高的最精确特征, 而块  $E_5$  的特征图则最粗糙, 分辨率最低。为了获取更好的性能, 我们提出了增强特征模块 (EFM, 其将在 §III-B2 中介绍) 以提高我们编码器的表征能力。EFM 模块在块  $E_5$  的最后一层 *conv5\_3* 后加入。其包含 2 个组膨胀卷积 (GAM) 以在更大的感受野上提取更强力的特征图。GAM 模块生成一个更小的特征图, 与 VGG-16 主干的最粗糙特征图相比, 大小为一半。它还增强了块  $E_5$  生成的特征图的表征能力。因此我们的编码器输出 6 个等级的特征图  $\{M_E^1, M_E^2, M_E^3, M_E^4, M_E^5, M_E^6\}$ , 其步长分别为  $\{1, 2, 4, 8, 16, 32\}$ 。因为我们采用了 U 形编解码器架构 [67], 所有这六个特征映射将在解码器中使用, 稍后将介绍。

**解码器.** 我们的解码器有 5 个不同大小的输出端。这里, 我们不预测最小特征图的子输出, 步长为 32, 因此没有

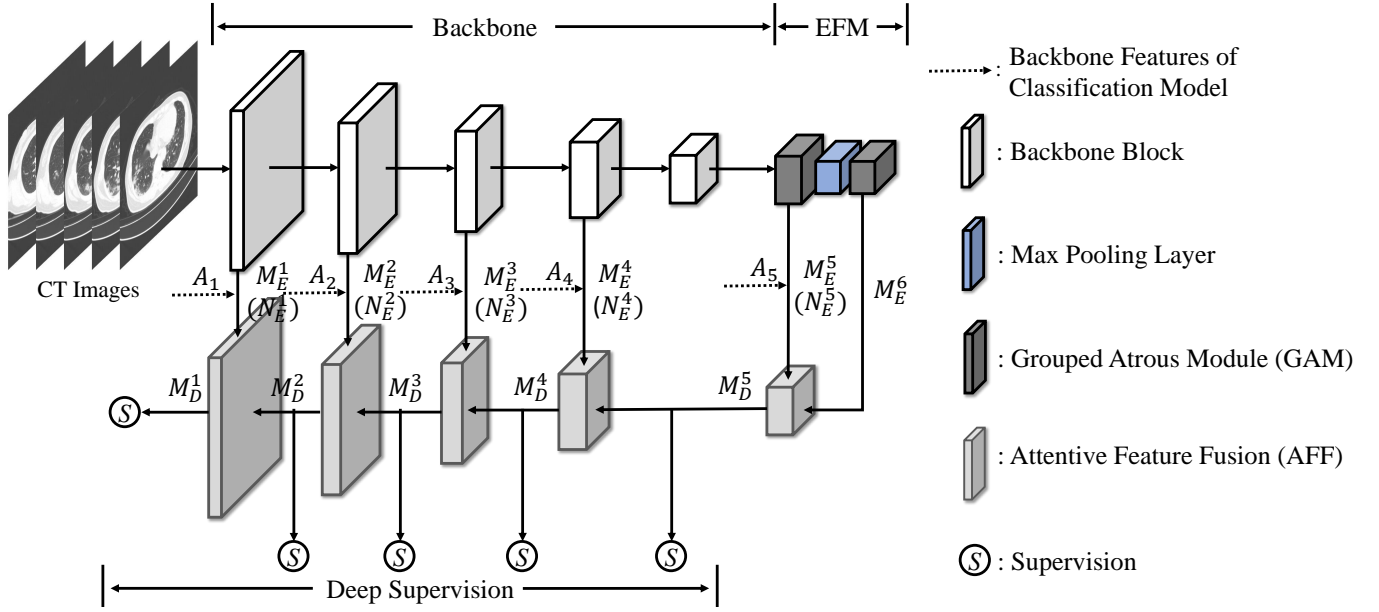


Figure 2. 分割分支的结构. EFM 代表增强特征模块 (§III-B2). AFF 代表注意特征融合策略 (§III-B3). 如果不与分类模型相结合,  $M_E^1 \sim M_E^5$  将被输入解码器; 否则, 结合后的  $N_E^1 \sim N_E^5$  将被输入解码器 (Fig. 4, §III-B4). 我们运用深层次的监督来训练我们的分支 (§III-B5).

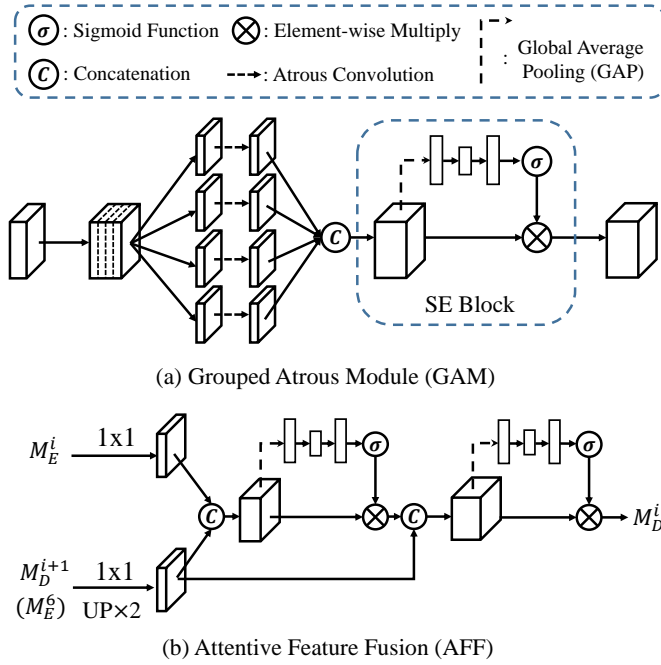


Figure 3. 我们为分割网络提出的 (a) GAM 和 (b) AFF. 在 AFF 中, 若  $i = 5$ ,  $M_D^{i+1}$  将被  $M_E^6$  替代. 立方体表示三维特征图, 矩形表示特征向量.

子输出与最小特征图  $M_E^6$  的大小匹配在我们的解码器中, 我们提出了一种注意特征融合 (AFF, 该算法将在本文中 §III-B3 介绍) 算法, 将不同阶段的特征图进行聚合, 预测各阶段的子输出. 我们的 AFF 强调了顶层特征图的重要性, 并利用注意机制从底层特征图中过滤出有用的特征. 最后

一次输出与输入的 CT 图像具有相同的分辨率, 作为最终的预测.

2) 增强特征模块: 提出的 EFM 模块在 VGG-16 编码器的最后一层  $E_5$  之后添加. 它由两个连续的 GAM 模块和它们之间的 *max pooling* 函数组成. 如图 Fig. 3 (a) 所示, GAM 模块的第一层是一个  $1 \times 1$  卷积层, 用于扩展特征图的通道. 然后将特征图平均分为 4 组. 与普通组卷积不同的是, 我们将具有不同膨胀率的膨胀卷积 [54] 使用在 4 个组中, 以获得具有不同感受野的更丰富的特征图. 膨胀卷积可以极大地扩大卷积滤波器的感受野并与正常卷积保持相同的计算成本. 在 2D 情况下, 卷积核大小为  $3 \times 3$  的膨胀卷积可以简单地表示如下:

$$q[i, j] = bias + \sum_{k=-1}^{+1} \sum_{l=-1}^{+1} (x[i+k \cdot n, j+l \cdot n] \cdot w[k+1, l+1]), \quad (5)$$

其中  $n$  表示膨胀率,  $w$  是卷积权重, 其大小为  $3 \times 3$ ,  $q$  和  $x$  分别是输出和输入特征图,  $i$  和  $j$  是特征图位置. 注意,  $n = 1$  的特例即是普通卷积. 为了充分利用有用的特性, 我们在网络中采用了 Squeeze-Excitation (SE) 块 [68], 也就是说, 使用注意机制重新校准信道卷积特征响应. 更具体地说, 输入特征图的每个通道将乘以由 SE 块计算的通道权重. SE 块由两个线性层组成, 后跟一个 sigmoid 函数. 全局平均池后的输入特征映射将被输入到这个块中, 我们可以为每个输入特征通道导出一个范围为  $0, 1$  的通道权重. 我们将 SE 块中的缩减率设置为 4, 这意味着我们将

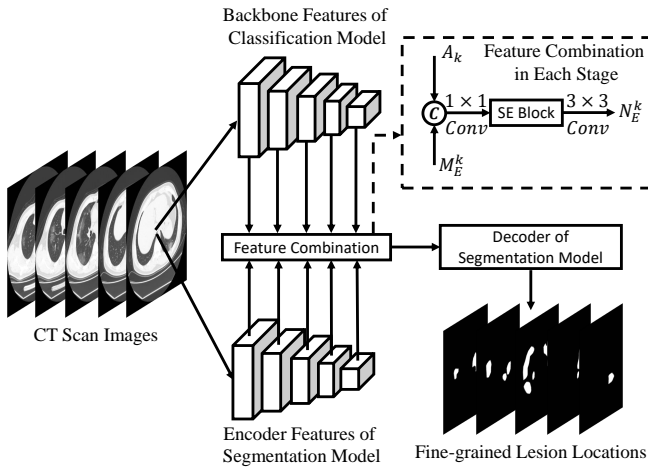


Figure 4. 分割和分类模型的结合. 我们将分割模型的编码器特征与分类模型的主干特征相结合。

第一个线性层的输出数设置为输入通道的  $1/4$ 。为了将输出通道减少一半，我们在 SE 块之后添加了一个  $1 \times 1$  的卷积层。

最后，我们使用一个  $3 \times 3$  卷积层，其中通道数等于输入特征图的通道数，作为到下一个模块的过渡层。

3) 注意特征融合：自顶向下解码器的传统融合策略 [67], [69] 平等对待输入特征映射。为了更好地聚集特征图，我们提出了一种注意特征融合 (AFF) 策略。在我们的 AFF 融合策略中，较小尺寸的特征图更有价值。如题 Fig. 3 (b) 所示，当前阶段中的输入特征图  $M_E^i$  和  $M_D^{i+1}$  通过  $1 \times 1$  的卷积层缩小到一半大小。然后通过双线性插值对缩减后的  $md^{i+1}$  进行上采样，输出一个双倍大小的特征图。我们将两个输出连接起来，并应用 SE 块（也用于 GAM）来生成增强的特征图。后将该特征映射与前一阶段的双上采样输出的特征映射拼接起来。在拼接之后，我们使用另一个 SE 块来再次增强特征图。在每个 SE 块之后，我们使用一个  $3 \times 3$  卷积层作为过渡层，该层具有与输入相同的通道数。一个神经元的一个  $1 \times 1$  卷积层将被用来预测一个特征映射作为当前阶段的子输出。

4) 与分类模型的结合：如上所述，我们设计了两个模型，一个用于 COVID-19 分类，另一个用于 COVID-19 混浊分割。但是，它们是分开工作在诊断系统上的，可能有一种方法可以将它们结合在一起以获得更好的性能。受此启发，我们利用分类模型的特征来增强分割模型的特征。如题 Fig. 4 所示，将分割模型的编码器和分类模型的主干点的各个阶段的特征映射合并在一起。分割模型编码器的特征图如 §III-B1 中，定义为  $M_E^1, M_E^2, M_E^3, M_E^4, M_E^5$ 。Res2Net [42] 分类模型主干有 5 个阶段，我们使用  $k \in [1, 5]$  阶段的最后一个特征图  $A_k$  进行特征结合。在合并阶段  $k$  的特征时，我

们有两个用于合并的特征图  $A_k, M_E^k$ 。我们首先调整较小的  $A_k$  的大小，使其与较大的  $M_E^k$  大小相同，然后将它们拼接在一起。然后，我们应用一个简单的  $1 \times 1$  卷积层来减少特征通道，使输出特征图的通道数与  $M_E^k$  相同。这样的  $1 \times 1$  卷积层后面是一个还原率为 4 的 SE 块。最后，我们使用了一个输入和输出通道数相同的  $3 \times 3$  卷积层作为过渡层。输出  $N_E^k$  将被视为增强的编码器特征，并被输入到分割模型的解码器中 (Fig. 2)。然后预测结果如 §III-B1 中所述。

5) 深度监督损失：虽然最终的预测仅来自最后一个子输出，但我们将深度监督策略 [70] 应用于不同大小的所有子输出。对于每侧输出，我们将其采样到客观事实图像的大小，并计算标准二进制交叉熵损失和 Dice 损失 [71] 之和，如下所示：

$$L = BCE(\mathbf{P}, \mathbf{G}) + 1 - \frac{\mathbf{P} \cdot \mathbf{G}}{\|\mathbf{P}\|_1 + \|\mathbf{G}\|_1}, \quad (6)$$

其中二分类交叉熵 (BCE) 取所有  $H \times W$  个像素的平均值， $p_{i,j}$  为通过 *sigmoid* 函数计算的在  $(i, j)$  位置像素的置信得分，“ $\cdot$ ”代表点乘。 $\mathbf{P}$  和  $\mathbf{G}$  为分别预测图和客观事实图， $\|\mathbf{P}\|_1$  和  $\|\mathbf{G}\|_1$  代表对应的  $\ell_1$  范数。

### C. 联合诊断

一个可解释的分类器或精确的分割模型本身不能完全实现 COVID-19 诊断的综合功能。与分割模型相比，我们的分类器是用 COVID-19 感染和未感染病例的 CT 图像进行训练的，以较低的标注代价获得更多的训练数据。尽管我们的分类器可以通过激活映射技术提供 COVID-19 的可解释的病变位置，但它不能进行准确和完整的病变分割。为此，我们的分割模型进一步提供了补充分析，通过发现肺部完整的病变并估计 COVID-19 患者的严重程度。但是由经验丰富的放射科医生注释大量的分割标签是非常昂贵的。为了更好地发挥它们的优势，我们通过联合解释分类和分割模型开发了 COVID-19 诊断系统。在实践中，我们的分类模型将首先预测可疑病例的 CT 图像是否为 COVID-19 阳性。如果预测为阳性，疑似病例极有可能被 COVID-19 感染。我们的分割模型将在 CT 图像上进行深入分析，并在每个 CT 图像中发现整个混浊区域。

## IV. COVID-CS 数据集

数据在基于深度学习的人工智能诊断系统中起着至关重要的作用。目前，很少有公开的 COVID-19 数据集具有大规模样本或细粒度像素级标记。为了填补这个空白，我们构建了一个新的 COVID-19 分类和分割 (COVID-CS) 数据集。在本节中，我们将介绍数据集的数据收集、专业标记和统计。Fig. 5 展示了我们 COVID-CS 数据集 (普通样本

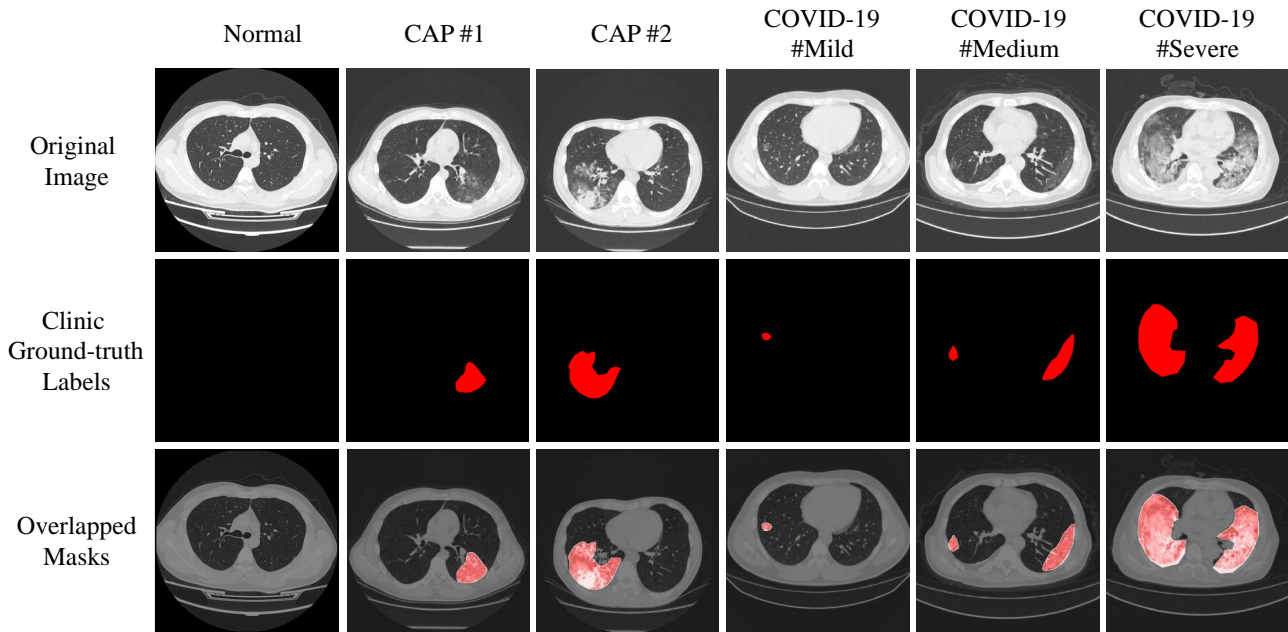


Figure 5. 我们的 *COVID-CS* 数据集示例，包括 1 例正常人（第 1 列）、2 例社区获得性肺炎（CAP）病例（第 2 和 3 列）和 3 例 COVID-19 患者的 CT 影像扫描图像和标签（第 4 ~ 6 列）。

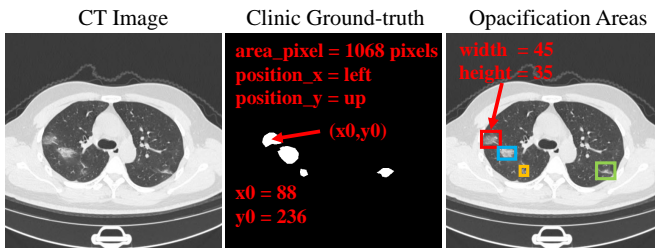


Figure 6. 在我们 *COVID-CS* 数据集中关于浑浊的区域（像素级）、位置  $(x_0, y_0)$ 、方向（左，上）与宽度/高度的各种信息的说明

和 COVID-19 样本) 的一些样例，同时也展示了一些 CAP 患者的样例。Fig. 6 展示了从我们 *COVID-CS* 数据集的分割集中提供的多样信息。

#### A. 数据收集

为了保护患者的隐私，我们在构建数据集时省略了患者的个人信息。我们收集了 750 例患者的 144167 张 CT 影像扫描图像，其中 400 例为 COVID-19 阳性，其余 350 例为阴性，这些样例均经 RT-PCR 检测证实。如以往工作所作 [72]，我们不考虑社区获得性肺炎（CAP）患者（详情见图 Fig. 5）。尽管 CAP 患者的 CT 图像也有相似的混浊，我们提出的诊断系统可能会诊断为 COVID-19 阳性，但 CAP 的威胁远小于 COVID-19。我们的目的是快速开发一个自动诊断系统，并尽快诊断出疑似病例。此外，利

Table III  
THE CT SCANNERS AND NUMBERS OF POSITIVE CASES.

Manufacturer	Product Name	#Cases
GE Medical Systems	Revolution CT	1
GE Medical Systems	LightSpeed VCT	6
GE Medical Systems	Discovery CT750 HD	12
GE Medical Systems	BrightSpeed	12
Toshiba	Aquilion ONE	33
GE Medical Systems	LightSpeed16	64
United Imaging Healthcare	uCT 780	272

用 CAP/COVID-19 分类器 [72]、RT-PCR 测试与医师的经验，CAP 患者可简单诊断为 COVID-19 阴性。

所有患者均行标准胸部 CT 影像扫描。每个病例有 250~400 个 CT 图像，每个病例的 CT 图像数量仅由 CT 影像扫描仪的类型及其扫描设置决定。CT 影像扫描仪包括 BrightSpeed、Discovery CT750 HD、LightSpeed VCT、LightSpeed16、GE Medical Systems 的 Revolution CT、东芝的 Aquilion ONE 和 United Imaging Healthcare 的 uCT 780。不同扫描器的病例数总结在表 Table III 中。CT 重建片的厚度为 0.75~1.25mm（详情见图 Fig. 7）。

#### B. 专家标注

我们在我们的 *COVID-CS* 数据集中为收集的 CT 影像扫描图像提供两个方面的标签，以便实现联合分类和分割任务。如上所述，我们的数据集分为 400 个 COVID-19

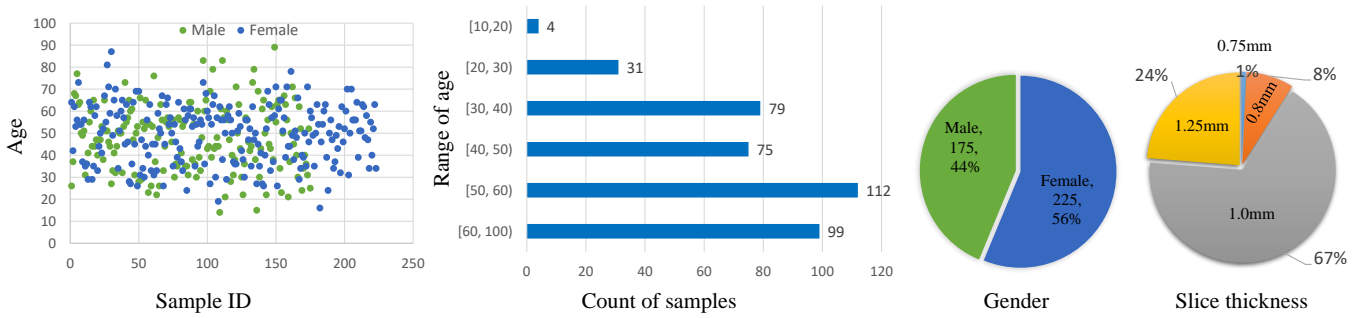


Figure 7. The age, gender, and slice thickness distribution of the COVID-19 patients in our *COVID-CS* dataset. Zoom in for details.

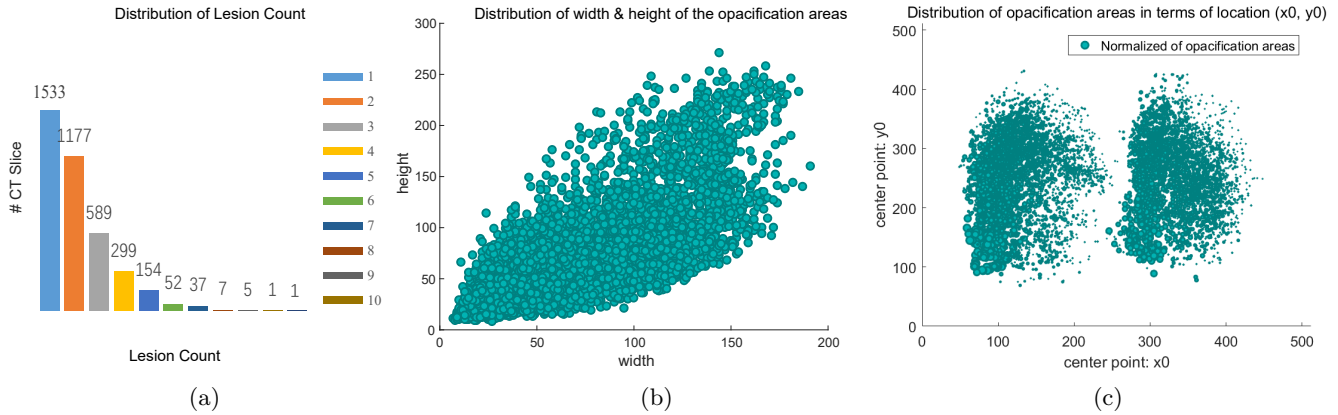


Figure 8. 在我们的 *COVID-CS* 数据集中的分割集 (200 个 COVID-19 病例) 的统计数据. (a) 病变计数分布. (b) 浑浊区域宽度和高度的分布. (c) 浑浊区域与位置的关系

病例和 350 个未感染病例。对于细分任务，我们通过以下策略进行专家标注：

- 为了节省标记时间，放射科医生只能为每个病人选择最多 30 个离散的 CT 影像扫描图像，在这些图像中可以观察到感染并作进一步的注释。在这一步中，我们的目标是用像素级别的注释标记每个浑浊区域。
- 为了生成高质量的标注，我们首先邀请放射科医生根据其临床经验标记尽可能多的浑浊区域。然后，我们邀请另一位资深放射科医生对标记进行多次细化，以进行交叉验证。在此步骤之后，将修复一些不准确的标签。

通过实现上述标记过程，我们最终得到 200 位 COVID-19 患者的 3,855 张像素级标注的 CT 影像扫描图像，图像分辨率为  $512 \times 512$ 。由于缺乏放射科医师，其他 200 例 COVID-19 患者的 64,771 张 CT 图像没有像素级标注，但这些数据将用于分类试验。从 Fig. 5 的最后三栏可以看出，我们的数据集覆盖了轻微、普通、严重的 COVID-19 病例。

### C. 数据统计

**年龄.** 400 例 COVID-19 患者 (男 175 例，女 225 例) 年龄 14-89 岁，平均 48.9 岁。Fig. 7 显示了年龄分布、年龄范围内的样本计数和性别百分比。

**病变计数.** 我们在图 Fig. 8 (a) 中展示了病变计数的分布。我们观察到，病变计数分布在 1 至 10 在每个 CT 影像扫描图像。

**浑浊区域.** 我们在图 Fig. 8 (b) 中绘制了浑浊区域的宽度和高度。宽度和高度的范围分别为 7 ~ 191 和 8 ~ 271，显示出不同的分布。

**位置.** 我们还在 Fig. 8 (c) 中显示了每个浑浊区域和相应的中心位置 ( $x_0, y_0$ ) 之间的关系。可以看出，归一化的浑浊区域的范围从最小尺寸 (35/28452 像素) 到最大尺寸 (28452/28452 像素)。它还显示，在我们的 *COVID-CS* 数据集中，浑浊区域均匀分布在不同的位置，这些区域也均匀分布在肺部。



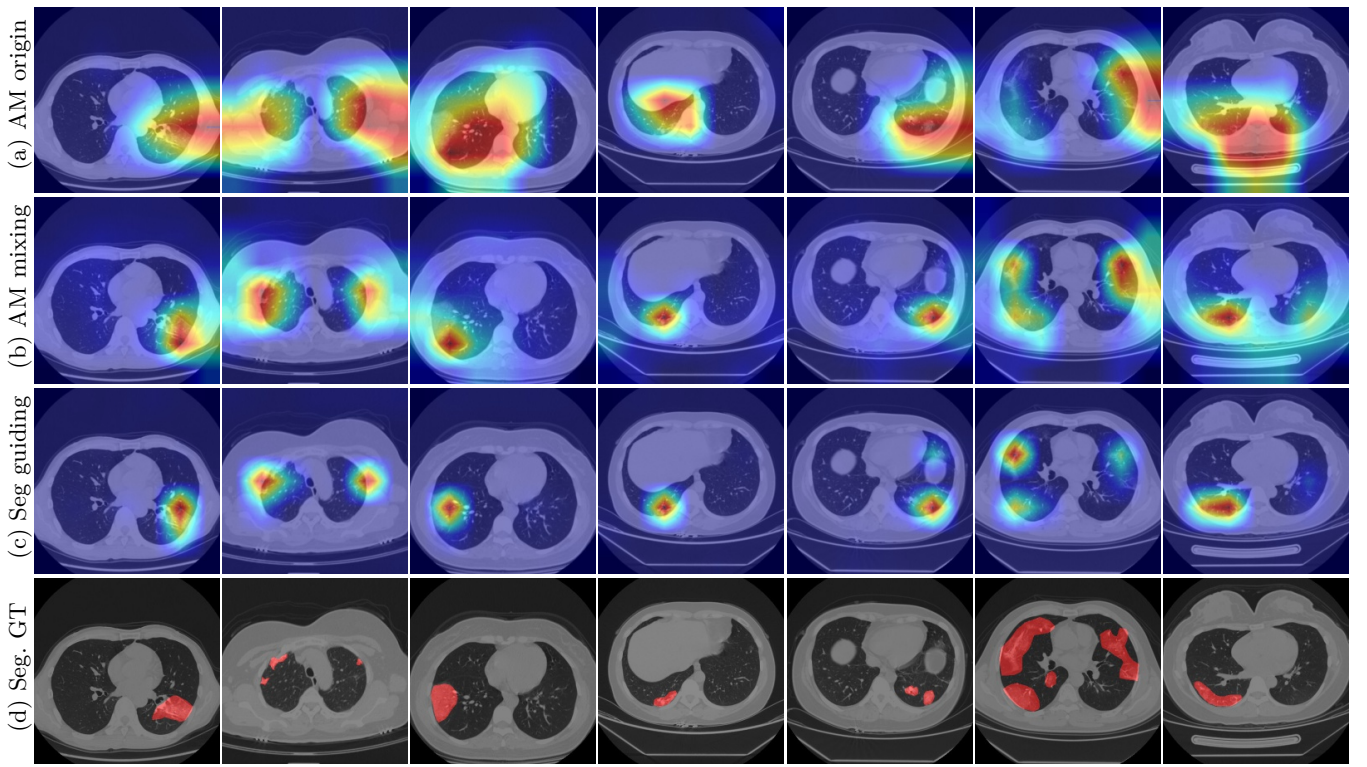


Figure 9. 激活映射 (AM) 可视化. AM origin (mixing) 代表没有 (使用) 图像混合技术 [65] 训练的模型。Seg guiding 代表使用分割损失  $L_{seg}$  训练的模型的 AM。

## V. 实验

### B. 实现细节

#### A. 实验设置

**训练/测试规定.** 对于分割任务, 我们的训练集包含来自 150 名 COVID-19 患者的 2794 张图像, 测试集包含来自其他 50 名 COVID-19 患者的 1061 张图像。对于分类任务, 训练集包含分割集中 150 个 COVID-19 感染病例的 2794 张图像。另外, 我们随机选取 150 例未感染病例, 7500 张 CT 图像作为阴性病例进行训练。测试集包含随机选取的 200 例感染病例的 64711 张图像和 200 例未感染病例的 68041 张图像。

**评估指标.** 对于分类任务, 我们采用了 [26] 建议的被广泛使用的特异性和敏感性指标对于分割任务, 我们使用两个标准度量, *i.e.*, Dice 分数 [73] 与 IoU。为了提供更全面的评估, 我们进一步使用了广泛使用的度量——增强对齐度量 ( $E_\phi$ ) [74]。

**方法对比.** 在分类任务上, 我们比较了有无图像混合技术的分类模型 [65]。在分割任务上, 为了深入评估我们的 JCS 模型, 我们将其与多功能的尖端模型进行比较, *i.e.*, 用于医学图像的 U-Net [67], DSS [75], PoolNet [76], 和用于显著性检测的 EGNNet [77]。

在我们的 JCS 中, 分类模型和分割模型分别训练。对于分类模型, 我们在 4 个 gpu 上以 256 的批量大小对其进行训练。为了提高计算效率, 将 CT 图像的大小调整为  $224 \times 224$ 。我们采用 SGD 优化器, 初始学习率为 0.1, 每 30 个周期除以 10。分类器训练 100 个周期。在数据增强方面, 我们采用了随机水平翻转和随机裁剪技术, 并采用了图像混合技术 [65] 来消除数据的偏差。图像混合的 Beta 分布中的  $\alpha$  设置为 0.5。对于分割模型, 每个小批量中的 CT 图像数始终为 4, 并且输入 CT 图像的大小不变, 为  $512 \times 512$ 。我们的分割模型的主干是在 ImageNet 上预先训练的 [38]。两个序列 GAM 中四个空洞卷积的膨胀率分别为  $\{1, 3, 6, 9\}$  和  $\{1, 2, 3, 4\}$ 。初始学习率为  $2.5 \times 10^{-5}$ 。The initial learning rate is  $2.5 \times 10^{-5}$ 。我们采用 *poly* 学习率策略, 实际学习率将乘以因子  $(1 - \frac{cur\_iter}{max\_iter})^{power}$ , 其中 *power* 为 0.9。分割模型进行了 21000 次迭代训练。我们使用 Adam [78] 优化器, 将  $\beta_1, \beta_2$  分别设置为 0.9 和 0.999。对于数据增强, 我们使用随机水平翻转和随机裁剪。结合分类模型, 对分类训练集进行了像素级标注预处理。

Table IV

不同阈值下分类模型的敏感性和特异性。在最终设置中，我们将阈值设置为 25（灰色行）。

No.	Threshold	Sensitivity	Specificity
1	15	96.0%	91.5%
2	20	95.0%	92.0%
3	25	95.0%	93.0%
4	30	94.5%	93.5%

### C. 结果

**可解释分类的激活映射.** Fig. 9显示了在有或没有图像混合的情况下训练的分类分支的激活映射 (AM) 的可视化 [65]。首先，我们训练了我们的分类模型，并在敏感性和特异性方面取得了良好的性能。但是我们发现，我们的分类模型中的 AM 最初是用随机水平翻转和随机裁剪 (Fig. 9 (a)) 训练的，不仅覆盖了病变区域，而且呈现出不相关的区域。如果这个问题不解决，分类网络过拟合的自动诊断系统对临床诊断是非常有害的。为了解决这个问题，我们研究并确定了图像混合技术可以解决这个问题。通过引入图像混合技术 [65]，我们的分类模型的 AM 提供了更准确的混浊区域位置，如 Fig. 9 (b) 所示。另外 Fig. 9 (c) 表示在像素级监督的帮助下训练的模型的 AM (分割损失  $L_{seg}$  在 §III-A4中介绍) 模型的 AM 对混浊的定位更加准确和特异。然而，在分类性能上增加分割损失  $L_{seg}$  的改进可以忽略，这可能是由于饱和和分类精度的原因 (No.3, Table IV)。

**可解释分类的性能.** 在推理过程中，AM 辅助医学专家使用 JCS 系统判断预测是否正确。对于每一个病人，在一些 CT 图像中可以发现混浊，许多图像可能没有混浊。所以我们为分类设置了一个阈值。当可疑患者的 CT 图像数大于阈值时，患者被诊断为 COVID-19 阳性。通过改变阈值，我们的模型可以在敏感性和特异性之间进行权衡。Table IV 显示了在我们的 COVID-CS 数据集的测试集上不同阈值下分类模型的结果。可以看出，该模型对阈值的变化具有很强的鲁棒性，当阈值为 25 时，该模型的灵敏度达到 95.0%，特异性为 93.0%。然而，AM 不能提供准确的混浊区域分割（如果有的话）。随后，我们进一步利用我们的分割模型来发现 COVID-19 患者的 CT 图像中的混浊区域。

**分割分支中 EFM 和 AFF 消融实验.** In §III-B 我们引入了两个新的分割模块 EFM 和 AFF。EFM 的目的是提高编码器在分割分支中的表示能力。在特征融合阶段，采用 AFF 算法，较小尺寸的特征图更受重视，而传统的融合策略对输入的特征图一视同仁。EFM 和 AFF 的消融研究如 Table V 所示。第一个结果是没有 EFM 和 AFF 的基准性

Table V

分割模型中 EFM 和 AFF 的消融实验。基线是基于 VGG16 的分割模型，没有 EFM&AFF (No. 1)。我们分别添加了 EFM 和 AFF，并展示了它们的有效性 (No.2 和 No.3)。第四个结果是分割模型的完整版本。

No.	EFM	AFF	Dice	IoU	$E_\phi$
1			71.0%	57.7%	88.0%
2	✓		74.3%	61.4%	88.9%
3		✓	75.9%	63.4%	90.9%
4	✓	✓	<b>77.5%</b>	<b>65.4%</b>	<b>92.0%</b>

Table VI

分割模型与分类模型相结合的消融研究。基准分割结果仅使用分割模型 (No.1) 生成。在添加分类模型的特征之后，我们在 Dice 度量 (No.2) 方面实现了 1.0% 的改进。

No.	SEG	+CLS	Dice	IoU	$E_\phi$
1	✓		77.5%	65.4%	92.0%
2	✓	✓	<b>78.5%</b>	<b>66.4%</b>	<b>92.7%</b>

能。在将所提出的 EFM 和 AFF 分别应用于基准之后，性能在 Dice 度量方面分别提高了 3.3 因此，EFM 和 AFF 对分割分支都有很大的帮助。当结合 EFM 和 AFF 时，我们在 Dice 度量方面获得了 6.5 在 IoU 和 E-measure [74] 度量方面的改进与 Dice 度量类似。因此，本文提出的 EFM 和 AFF 对分割模型是非常有益的。

**分割模型与分类模型相结合的消融实验.** 如 §III-B4中所述，将分类模型与分割模型相结合，得到更丰富的特征。为了验证这样的选择，我们运行了如 Table VI 中所示的实验。基准是单一切割模型 (No.1, Table VI)。但我们也观察到分类模型和分割模型的组合选择 (No.2, Table VI) 在 Dice 度量方面有 1.0% 的改进，并且表明分类模型的特征可以帮助分割模型预测更好的结果。

**分割性能比较.** Table VII 列出了 4 种最新的分割方法和我们的分割模型的定量比较。可以看出，该模型在所有三个指标上都取得了最好的结果。与第二名的 PoolNet [76] 相比，它在 Dice 得分、IoU 和  $E_\phi$  上分别提高了 8.8%、10.5% 和 8.8%。另外，PoolNet [76] 和 EGNet [77] 在 3 个指标上获得了相似的结果。U-Net 在 IoU 上比 DSS [75] 的表现好，尽管它们在 Dice 得分上相当。Fig. 10 显示了比较方法的定性结果。可以看出，其他竞争者在轻度、中度和重度 COVID-19 感染的 CT 图像中对病变区域的预测不准确甚至错误。但是我们的分割模型正确地发现了所有 COVID-19 感染水平上的整个病变区域。

为了进一步研究它的稳定性，我们在分割测试集上对我们的分割模型进行了统计分析。图 Fig. 11 (a) 显示了本

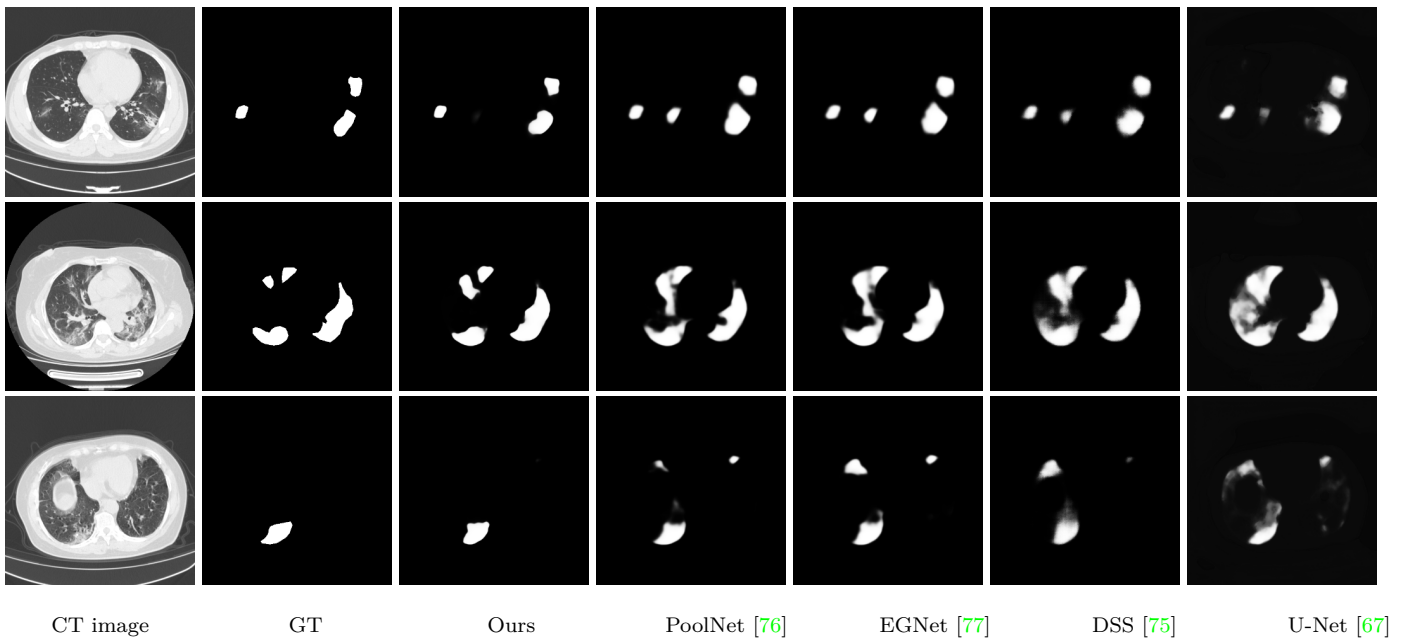


Figure 10. 不同方法在分割测试集上的定性比较。第一行、第二行和第三行分别显示轻度、中度和重度 COVID-19 患者不同病变区域的 CT 图像对比结果。

Table VII  
分割测试集的定量结果。

Methods	Publication	Dice	IoU	$E_\phi$
U-Net [67]	MICCAI'15	65.1%	54.1%	79.7%
DSS [75]	TPAMI'19	65.7%	51.7%	79.9%
EGNet [77]	ICCV'19	69.3%	55.4%	83.6%
PoolNet [76]	CVPR'19	69.7%	55.9%	83.9%
<i>JCS (Ours)</i>	Submit'20	<b>78.5%</b>	<b>66.4%</b>	<b>92.7%</b>

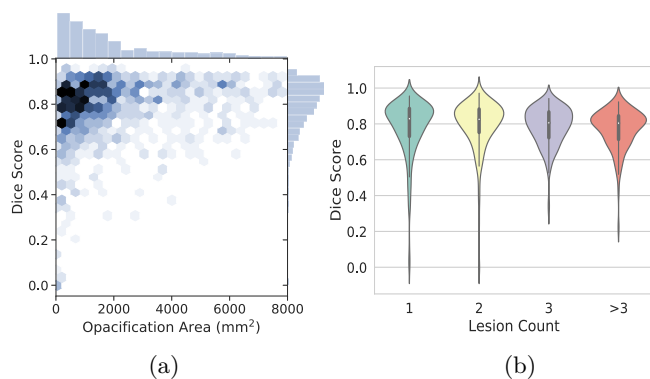


Figure 11. 在分割测试集上对我们的分割模型进行统计分析。(a) 每个 CT 图像的混浊区域与相应 Dice 评分之间的关系。(b) 病损计数与 Dice 得分概率分布的关系。

模型的 Dice 评分与 CT 图像混浊区域的相关性。请注意，由于  $\geq 8000\text{mm}^2$  不透明区域的 CT 图像只占有所有 CT 图像数量的 1.0%，它们未绘制在 Fig. 11 (a) 中。我们观察到 95.9% 的 CT 图像的 Dice 评分为  $[0.6, 1]$ ，而其他 3.3% 的

CT 图像的 Dice 评分在  $[0.1, 0.6]$  之间，被认为是不良病例。CT 图像中只有 0.8% 的图像 Dice 得分低于 0.1，以失败病例为例。我们还从不同的角度探讨了每个切片的病变计数与 Dice 评分之间的关系。如图 Fig. 11 (b) 所示，CT 图像中的病灶计数对 Dice 评分的概率分布影响不大。对于 4 种不同的病变计数，中位 Dice 得分高于 0.8，95.0% 置信区间为  $[0.5, 1]$ 。我们还观察到失败病例的病变计数  $\leq 2$ 。在分割病变区域方面的持续良好性能和低失败概率 (0.8%) 证实了我们的分割模型的稳定性。

**诊断用时.** 我们 *JCS* 系统在单张 RTX 2080Ti 上进行速度测试。假设每个疑似病例都有 300 张 CT 图像，我们 *JCS* 系统的分类模型只需 1.0 秒左右的时间就可以确定是否被感染。如果被感染，分割模型将花费 21.0 秒进行细粒度病变分割。因此，在 *JCS* 系统中，每个感染病例的系统成本为 22.0 秒，每个未感染病例的系统成本为 1.0 秒。值得注意的是，无论病例是否感染，完整的 RT-PCR 检测和放射科医生的 CT 诊断分别需要 4 小时和 21.5 分钟。

## VI. 结论

为了训练用于 COVID-19 诊断的强 CNN 模型，本文系统地构建了一个大规模的 COVID-19 分类与分割 (*COVID-CS*) 数据集。我们还开发了一个用于 COVID-19 诊断的联合分类和分割 (*JCS*) 系统。在我们的系统中，分类模型确定了疑似患者是否为 COVID-19 阳性，并提供令人信服的视觉解释。在 *COVID-CS* 数据集的分类测试集

上, 该方法具有 95.0% 的灵敏度和 93.0% 的特异性。为了提供互补的像素级预测, 我们实现了一个分割模型来发现 COVID-19 患者 CT 图像中的细粒度病变区域。与竞争方法相比, e.g., PoolNet [76], 我们的分割模型获得了 8.8% 的 Dice 得分增益。我们的模型依然十分稳定。在我们的分割测试集上, 其仅在 0.8% 的图像上失效, 在 95.9% 的图像中获得了 [0.6, 1] 的 Dice 得分。我们为 COVID-19 设计的 JCS 诊断系统的在线演示将很快提供。

#### ACKNOWLEDGMENT

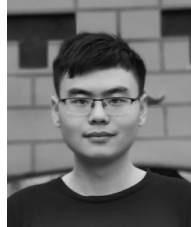
This research was supported by Major Project for New Generation of AI under Grant No. 2018AAA0100400, NSFC (61922046), and Tianjin Natural Science Foundation (17JCJQJC43700, 18ZXZNGX00110).

#### REFERENCES

- [1] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, and M.-M. Cheng, "Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation," *IEEE Transactions on Image Processing*, vol. 30, p. 3113–3126, 2021. [Online]. Available: <http://dx.doi.org/10.1109/TIP.2021.3058783>
- [2] WHO, "Coronavirus disease (covid-19) outbreak situation," <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, 2020.
- [3] H. B. Jenssen, "China national health commission diagnosis and treatment of pneumonitis caused by new coronavirus (trial version 6)," <http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml>, accessed 08 09, 2020.
- [4] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, 2020.
- [5] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, 2020.
- [6] Y. Wang, H. Hou, W. Wang, and W. Wang, "Combination of ct and rt-pcr in the screening or diagnosis of covid-19," *Journal of Global Health*, vol. 10, no. 1, 2020.
- [7] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," 2020.
- [8] S. Inui, A. Fujikawa, M. Jitsu, N. Kunishima, S. Watanabe, Y. Suzuki, S. Umeda, and Y. Uwabe, "Chest ct findings in cases from the cruise ship "diamond princess" with coronavirus disease 2019 (covid-19)," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200110, 2020.
- [9] Z. Huang, S. Zhao, Z. Li, W. Chen, L. Zhao, L. Deng, and B. Song, "The battle against coronavirus disease 2019 (covid-19): Emergency management and infection control in a radiology department," *Journal of the American College of Radiology*, 2020.
- [10] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," *arXiv 2003.11597*, 2020. [Online]. Available: <https://github.com/ieee8023/covid-chestxray-dataset>
- [11] J. Zhao, Y. Zhang, X. He, and P. Xie, "Covid-ct-dataset: a ct scan dataset about covid-19," *arXiv preprint arXiv:2003.13865*, 2020.
- [12] N. Sajid, "Covid-19 patients lungs x ray images 10000," <https://www.kaggle.com/nabeelsajid917/covid-19-x-ray-10000-images>, accessed 04 10, 2020.
- [13] H. B. Jenssen, "Covid-19 ct segmentation dataset," <http://medicalsegmentation.com/covid19/>, accessed 04 10, 2020.
- [14] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.
- [15] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad, A. Jacobi, K. Li, S. Li, and H. Shan, "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.
- [16] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li, S. Li, H. Shan, A. Jacobi, and M. Chung, "Chest ct findings in coronavirus disease-19 (covid-19): Relationship to duration of infection," *Radiology*, p. 200463, 2020.
- [17] Y. Wang, C. Dong, Y. Hu, C. Li, Q. Ren, X. Zhang, H. Shi, and M. Zhou, "Temporal changes of ct findings in 90 patients with covid-19 pneumonia: A longitudinal study," *Radiology*, p. 200843, 2020.
- [18] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, and C. Zheng, "Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study," *The Lancet Infect Disease*, vol. 20, no. 4, pp. 425–434, 2020.
- [19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Int. Conf. Comput. Vis.*, Oct 2017.
- [20] J. Won, S. Lee, M. Park, T. Kim, M. Park, B. Choi, D. Kim, H. Chang, V. Kim, and C. Lee, "Development of a laboratory-safe and low-cost detection protocol for sars-cov-2 of the coronavirus disease 2019 (covid-19)." *Experimental neurobiology*, 2020.
- [21] B. Udugama, P. Kadhiresan, H. N. Kozlowski, A. Malekjahani, M. Osborne, V. Y. Li, H. Chen, S. Mubareka, J. B. Gubbay, and W. C. Chan, "Diagnosing covid-19: the disease and tools for detection," *ACS nano*, vol. 14, no. 4, pp. 3822–3835, 2020.
- [22] W. Yang and F. Yan, "Patients with rt-pcr-confirmed covid-19 and normal chest ct," *Radiology*, vol. 295, no. 2, pp. E3–E3, 2020.
- [23] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng *et al.*, "Clinical characteristics of 24 asymptomatic infections with covid-19 screened among close contacts in nanjing, china," *Science China Life Sciences*, vol. 63, no. 5, pp. 706–711, 2020.
- [24] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li *et al.*, "Chest ct findings in coronavirus disease-19 (covid-19): relationship to duration of infection," *Radiology*, p. 200463, 2020.
- [25] C. Long, H. Xu, Q. Shen, X. Zhang, B. Fan, C. Wang, B. Zeng, Z. Li, X. Li, and H. Li, "Diagnosis of the coronavirus disease (covid-19): rrt-pcr or ct?" *European journal of radiology*, p. 108961, 2020.

- [26] Y. Liu, Y.-H. Wu, Y. Ban, H. Wang, and M.-M. Cheng, "Rethinking computer-aided tuberculosis diagnosis," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016.
- [28] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," *ArXiv*, 2020.
- [29] R. Yang, X. Li, H. Liu, Y. Zhen, X. Zhang, Q. Xiong, Y. Luo, C. Gao, and W. Zeng, "Chest ct severity score: An imaging tool for assessing severe covid-19," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200047, 2020.
- [30] K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, "Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19)," *European Radiology*, 2020.
- [31] Z. Zhou, D. Guo, C. Li, Z. Fang, L. Chen, R. Yang, X. Li, and W. Zeng, "Coronavirus disease 2019: initial chest ct findings," *European Radiology*, 2020.
- [32] L. Wynants, B. Van Calster, G. S. Collins, R. D. Riley, G. Heinze, E. Schuit, M. M. Bonten, D. L. Dahly, J. A. Damen, T. P. Debray *et al.*, "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal," *bmj*, vol. 369, 2020.
- [33] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE Trans. Medical Imaging*, 2020.
- [34] Y. Qiu, Y. Liu, and J. Xu, "MiniSeg: An extremely minimum network for efficient covid-19 segmentation," *arXiv preprint arXiv:2004.09750*, 2020.
- [35] V. Rajinikanth, N. Dey, A. N. J. Raj, A. E. Hassanien, K. C. Santosh, and N. S. M. Raja, "Harmony-search and otsu based system for coronavirus disease (covid-19) detection using lung ct scan images," *arXiv*, 2020.
- [36] J. B. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fundamenta Informaticae*, no. 1,2, pp. 187–228, 2000.
- [37] T. Zhou, S. Canu, and S. Ruan, "An automatic covid-19 ct segmentation based on u-net with attention mechanism," *arXiv preprint arXiv:2004.06673*, 2020.
- [38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 248–255.
- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.
- [40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. Learn. Represent.*, 2015.
- [41] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017.
- [42] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, 2021.
- [43] M.-M. Cheng, Y. Liu, Q. Hou, J. Bian, P. Torr, S.-M. Hu, and Z. Tu, "Hfs: Hierarchical feature selection for efficient image segmentation," in *European conference on computer vision*. Springer, 2016, pp. 867–882.
- [44] Y. Liu, P.-T. Jiang, V. Petrosyan, S.-J. Li, J. Bian, L. Zhang, and M.-M. Cheng, "Del: Deep embedding learning for efficient image segmentation," in *IJCAI*, vol. 864, 2018, p. 870.
- [45] J. Zhao, R. Bo, Q. Hou, M.-M. Cheng, and P. Rosin, "Flic: Fast linear iterative clustering with active search," *Computational Visual Media*, vol. 4, no. 4, pp. 333–348, 2018.
- [46] Y. Liu, Y. C. Gu, X. Y. Zhang, W. Wang, and M. M. Cheng, "Lightweight salient object detection via hierarchical visual perception learning," *IEEE Trans. Cybernetics*, pp. 1–11, 2020.
- [47] K. Zhao, J. Xu, and M.-M. Cheng, "Regularface: Deep face recognition via exclusive regularization," in *IEEE Conf. Comput. Vis. Pattern Recog.*, June 2019, pp. 1136–1144.
- [48] Y.-Q. Tan, S.-H. Gao, X.-Y. Li, M.-M. Cheng, and B. Ren, "Vecroad: Point-based iterative graph exploration for road graphs extraction," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020.
- [49] L. Zhang, Z. Shi, M.-M. Cheng, Y. Liu, J.-W. Bian, J. T. Zhou, G. Zheng, and Z. Zeng, "Nonlinear regression via deep negative correlation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.
- [50] J.-W. Bian, Y.-H. Wu, J. Zhao, Y. Liu, L. Zhang, M.-M. Cheng, and I. Reid, "An evaluation of feature matchers for fundamental matrix estimation," in *Brit. Mach. Vis. Conf.*, 2019.
- [51] J. Xu, Y. Huang, M. M. Cheng, L. Liu, F. Zhu, Z. Xu, and L. Shao, "Noisy-as-clean: Learning self-supervised denoising from corrupted image," *IEEE Trans. Image Process.*, vol. 29, pp. 9316–9329, 2020.
- [52] Q. Wang, J. Gao, and X. Li, "Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4376–4386, 2019.
- [53] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [54] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.
- [55] Y. Liu, Y.-H. Wu, P. Wen, Y. Shi, Y. Qiu, and M.-M. Cheng, "Leveraging instance-, image- and dataset-level information for weakly supervised instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021.
- [56] Y.-H. Wu, Y. Liu, L. Zhang, W. Gao, and M.-M. Cheng, "Regularized densely-connected pyramid network for salient instance segmentation," *IEEE Trans. Image Process.*, 2021.
- [57] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, 2020.
- [58] K. Sofiuk, O. Barinova, and A. Konushin, "Adaptis: Adaptive instance selection network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7355–7363.
- [59] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image. Comput. Comput. Assist. Interv.*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 2015, pp. 234–241.
- [60] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation

- from sparse annotation,” in *Med. Image. Comput. Comput. Assist. Interv.*, Oct 2016, pp. 424–432.
- [61] V. Iglovikov and A. Shvets, “Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation,” *ArXiv e-prints*, 2018.
- [62] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Trans. Medical Imaging*, 2019.
- [63] Y. Xiong, D. Sun, Y. Liu, Y. Fan, L. Zhao, X. Li, and W. Zhu, “Clinical and high-resolution ct features of the covid-19 infection: comparison of the initial and follow-up changes,” *Investigative radiology*, 2020.
- [64] A. Leung, R. Miller, and N. Müller, “Parenchymal opacification in chronic infiltrative lung diseases: Ct-pathologic correlation.” *Radiology*, vol. 188, no. 1, pp. 209–214, 1993.
- [65] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopezpaz, “mixup: Beyond empirical risk minimization,” in *Int. Conf. Learn. Represent.*, 2018.
- [66] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu, “Guided attention inference network,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019.
- [67] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2015, pp. 234–241.
- [68] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 7132–7141.
- [69] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 2117–2125.
- [70] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, “Deeply-supervised nets,” in *Artificial intelligence and statistics*, 2015, pp. 562–570.
- [71] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Int. Conf. 3D Vision.* IEEE, 2016, pp. 565–571.
- [72] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, and D. Shen, “Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification,” *arXiv preprint arXiv:2003.09860*, 2020.
- [73] F. Shan+, Y. Gao+, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, “Lung infection quantification of covid-19 in ct images with deep learning,” *arXiv preprint arXiv:2003.04655*, 2020.
- [74] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, “Enhanced-alignment Measure for Binary Foreground Map Evaluation,” in *Int. Joint Conf. Artif. Intell.*, 2018.
- [75] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, “Deeply supervised salient object detection with short connections,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, 2019.
- [76] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, “A simple pooling-based design for real-time salient object detection,” in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 3917–3926.
- [77] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, “Egnet: Edge guidance network for salient object detection,” in *Int. Conf. Comput. Vis.*, 2019, pp. 8779–8788.
- [78] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Int. Conf. Learn. Represent.*, 2015.



**Yu-Huan Wu** is currently a Ph.D. candidate with College of Computer Science at Nankai University, supervised by Prof. Ming-Ming Cheng. He received his bachelor's degree from Xidian University in 2018. His research interests include computer vision and machine learning.



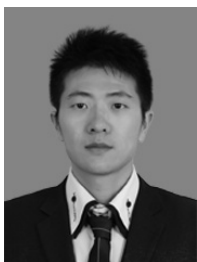
**Shang-Hua Gao** is a Ph.D. student in Media Computing Lab at Nankai University. He is supervised via Prof. Ming-Ming Cheng. His research interests include computer vision, machine learning, and radio vortex wireless communications.



**Jie Mei** is a Ph.D. student in College of Computer Science, Nankai University, Tianjin, China. His research interests include computer vision, machine learning, and remote sensing image processing.



**Jun Xu** received his B.Sc. and M.Sc. degrees from School of Mathematics Science, Nankai University, Tianjin, China, in 2011 and 2014, respectively, and the Ph.D. degree from the Department of Computing, Hong Kong Polytechnic University, in 2018. He worked as a Research Scientist at IIAI, Abu Dhabi, UAE. He is currently a Lecturer with School of Statistics and Data Science, Nankai University. More information can be found at <https://csjunxu.github.io/>.



**Deng-Ping Fan** received his Ph.D. degree from the Nankai University in 2019. He joined Inception Institute of Artificial Intelligence (IIAI) in 2019. He has published about 20 top journal and conference papers such as CVPR, ICCV, etc. His research interests include computer vision, deep learning, and saliency detection, especially on co-salient object detection, RGB salient object detection, RGB-D salient object detection, and video salient object detection.



**Rong-Guo Zhang** received his Ph.D. degree majoring in pattern recognition and intelligent systems from Institute of Automation, Chinese Academy of Sciences in 2012. Now he serves as head of Institute of Advanced Research, Beijing Infervision Technology Co Ltd. His research interests include computer vision, deep learning, and medical image processing.



**Ming-Ming Cheng** received his Ph.D. degree from Tsinghua University in 2012. Then he did two years research fellow with Prof. Philip Torr in Oxford. He is now a professor at Nankai University, leading the Media Computing Lab. His research interests include computer graphics, computer vision, and image processing. He received research awards, including ACM China Rising Star Award, IBM Global SUR Award, and CCF-Intel Young Faculty Researcher Program. He is on the editorial boards of IEEE TIP.