

# 基于多尺度超像素融合网络的脑 CT 图像分类方法

冀俊忠, 张梦隆, 宋晓, 张晓丹\*

(北京工业大学信息学部计算机学院, 北京 100124)

**摘要:** 近年来, 基于深度学习的医学图像分类方法已成为一个研究热点。其中, 脑 CT 图像缺少亮度、颜色、纹理等易于识别的自然视觉特征, 且病灶形态位置多变, 而被广泛用于图像特征提取的卷积神经网络无法有效地挖掘病灶特征, 从而影响了分类性能。本文针对脑 CT 图像的视觉特性, 提出了一种基于多尺度超像素融合网络的脑 CT 图像分类方法。该方法基于多尺度超像素图, 从图像融合和特征融合两个层面辅助卷积神经网络提取更具有表达性的分类特征。首先通过多尺度超像素对脑 CT 图像进行病灶区域增强, 获得优化的融合图像; 然后将融合图像的高层次特征和多尺度超像素低层次特征进行多层融合, 从而获得更有判别性的融合特征。实验结果表明, 本文提出的方法有效提升了脑 CT 图像分类性能。

**关键词:** 图像分类; 脑 CT 图像; 超像素; 特征融合

**中图分类号:** TP391 **文献标志码:** A

## Brain CT Image Classification Based on Multi-scale Superpixel Fusion Network

Ji Junzhong, ZHANG Menglong, SONG Xiao, ZHANG Xiaodan

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

**Abstract:** In recent years, the medical image classification based on deep learning has become a new hot topic. Brain CT images lack natural visual features such as brightness, color and texture that can be easily recognized, and the morphology and location of lesions are variable. However, the convolutional neural network, which is widely used for image feature extraction, cannot effectively mine the lesion features of Brain CT images, thus affecting the classification performance. According to the visual characteristics of Brain CT images, this paper proposes a novel Brain CT image classification method based on multi-scale superpixel fusion network (MSFN). MSFN is able to extract more expressive classification features through image fusion and feature fusion. Firstly, the original Brain CT image is enhanced by multi-scale superpixel to obtain the optimized fusion image. Then, we combine the high-level features of the fusion image with the multi-scale superpixel low-level features to obtain more discriminative features for the classification of Brain CT images. Experimental results validate the effectiveness of the proposed method.

**Keywords:** Image Classification; Brain CT image; Superpixel; Feature Fusion

临床急诊中脑疾病的诊断是极其紧迫的, 即使短时间的延误也可能导致患者病情恶化。电子计算机断层扫描(Computed Tomography, CT)是最常用的脑疾病诊断工具之一, 具有成像快、成本低、适用范围广、病变检查率高等特点。但是传统的脑 CT 诊断方法通常需要放射科医生目测出血面积大小、估计出中线偏移等信息, 这个过程是相对耗时的。随着医学影像技术的进步和发展, 脑 CT 图像数量呈现出几何形式的增长, 能够辅助放射科医生工作的脑 CT 图像自动分类诊断技术, 对于提升诊断效率、减少误诊漏诊率, 具有十分重要的现实意义。

近年来, 深度学习在计算机视觉领域的巨大成功, 也促进了医疗图像分析技术的快速发展, 基于深度学习的医学图像自动分类方法引起了研究学者的广泛关注<sup>[1-4]</sup>。其中, 卷积神经网络<sup>[5]</sup> (Convolutional Neural Networks, CNN)是一种传统的深度学习模型, 能够捕获图像的局部区域信息、提取高层语义特征, 被广泛用于图像的特征提取与分类任务<sup>[6-9]</sup>, 基于 CNN 的脑 CT 图像分类方法<sup>[10-17]</sup>也获得了广泛的应用。Garland 等人<sup>[10]</sup>提出通

收稿日期: 2022-06-06

基金项目: 国家自然科学基金资助项目(61906007); 北京市教委科技计划一般项目(KM202110005022)

第一作者: 冀俊忠(1969), 男, 教授, 主要研究方向为脑影像分析、数据挖掘、模式识别  
zhangxiaodan@bjut.edu.cn

过 CNN 完成对脑 CT 图像的分类；Gautam 等人<sup>[11]</sup>提出新的 CNN 模型对经过预处理的脑 CT 图像进行分类；Romany<sup>[12]</sup>和 Anupamal 等人<sup>[13]</sup>基于图像分割和 CNN 提取脑 CT 特征实现颅内出血诊断；Grewal 等人<sup>[14]</sup>使用 CNN 提取图像特征，并基于双向 LSTM 挖掘空间关系，实现脑出血的自动检测；Li 等人<sup>[15]</sup>利用 CNN 提取特征，采用循环神经网络学习脑 CT 图像间的依赖关系并实现了脑疾病多标签分类；惠瑞等人<sup>[16]</sup>将 CNN 提取的图像特征和 3D CNN 提取的空间关系特征进行融合，实现了基于脑 CT 的阿尔茨海默病等功能性疾病诊断。

然而，现有工作在使用传统 CNN 提取特征时，并没有考虑脑 CT 图像与自然光学图像之间的差异：脑 CT 图像缺少亮度、颜色、纹理等易于识别的自然视觉特征；空间分辨率、对比度低；病灶形态位置多变；存在噪声、伪影等不稳定因素。这些视觉特性导致了对脑 CT 病灶区域的特征表达更加困难。Blaivas 等人<sup>[17]</sup>通过定量图像质量评价技术对比了这两种图像的差异，验证了自然图像预训练的 CNN 模型在医学图像处理任务上的缺陷。为了更有效的表达脑 CT 图像特征，特别是挖掘病灶区域特征，本文首次将超像素分割<sup>[18]</sup>技术引入 CNN 特征模型，以更好的辅助 CNN 进行脑 CT 图像特征表达。

超像素是由一系列位置相邻且颜色、亮度、纹理等特性相似的像素点组成的小区域，其有利于图像高层次特征的提取与表达，已经被广泛应用于各种自然图像视觉任务<sup>[19-21]</sup>。结合脑 CT 图像的视觉特性，传统的超像素分割技术在病灶区域分割上具备天然优势。单一的分割结果往往不能准确地描述目标边界，容易获得过分割或欠分割的结果。因此，本文采用 Wei 等人<sup>[22]</sup>提出的超层次分割算法(Super Hierarchy, SH)来构造超像素层次结构，在多个尺度下对脑 CT 图像进行分割以获得多尺度超像素图。多尺度超像素包含的信息能够表达复杂多变的病灶区域，能够辅助 CNN 加强对病灶特征的表达能力，并提升脑 CT 图像分类的准确率。

综上，本文提出一种基于多尺度超像素融合网络的脑 CT 图像分类方法。该方法基于多尺度超像素图，从图像融合和特征融合两个层面辅助 CNN 提取更具有表达性的特征：1) 在图像融合中，将多尺度超像素图与脑 CT 图像相结合，以此减少图像中存在的噪声和冗余、划分病灶区域、降低病灶和周围脑组织像素的灰度相似性，达到数据增强的目的。2) 在特征融合中，将 CNN 特征和多尺度超像素特征进行多层融合。CNN 能够提取语义信息丰富、感受野大的高层次特征，但其对细节感知能力较差，容易忽略局部病灶信息；超像素能够表达图像低层次视觉特征，可以有效分割小面积病灶，但其语义信息比较有限。因此，在分类时综合利用各层次特征有望生成更具判别性的融合特征，进一步提升分类准确率。在脑 CT 数据集上的实验结果表明，本文所提方法能够有效地提升脑 CT 图像分类准确率。

## 1 基于多尺度超像素融合网络的脑 CT 分类方法

多尺度超像素可以有效减少图像复杂度和噪声冗余、准确表达复杂多变的病灶区域，能够辅助 CNN 进行更有效的脑 CT 图像特征表达和分类。基于上述思想，本文提出了基于多尺度超像素融合网络(multi-scale superpixel fusion network, MSFN)的脑 CT 图像分类方法，整体框架如图 1 所示。该分类模型主要由三部分构成：第一部分如图 1(a)所示，为多尺度超像素图像融合模型，实现了多尺度超像素对原脑 CT 图像的数据增强，获得融合图像；第二部分如图 1(b)所示，为多尺度超像素特征编码模型，用来提取多尺度超像素中包含边界和区域信息的低层次特征；第三部分如图 1(c)所示，为多尺度超像素特征融合分类模型，将 CNN 提取的融合图像高层次特征和多尺度超像素低层次特征进行多层融合，得到更具有判别性的融合特征，并用于预测脑 CT 图像的分类标签。

### 1.1 多尺度超像素图像融合模型

针对脑 CT 图像不同于自然图像的视觉特性，为了减少图像中位移伪影、容积效应

误差、设备噪声等诸多不稳定因素，进一步划分病灶区域、降低病灶和周围脑组织像素的灰度相似性，本文构建了基于多尺度超像素的脑 CT 图像融合模型。

超像素区域内的像素值相同，区域间的像素差异更加显著。对脑 CT 图像进行超像素分割，相当于在超像素区域内进行平滑、在超像素区域间进行锐化，能够抑制噪声并

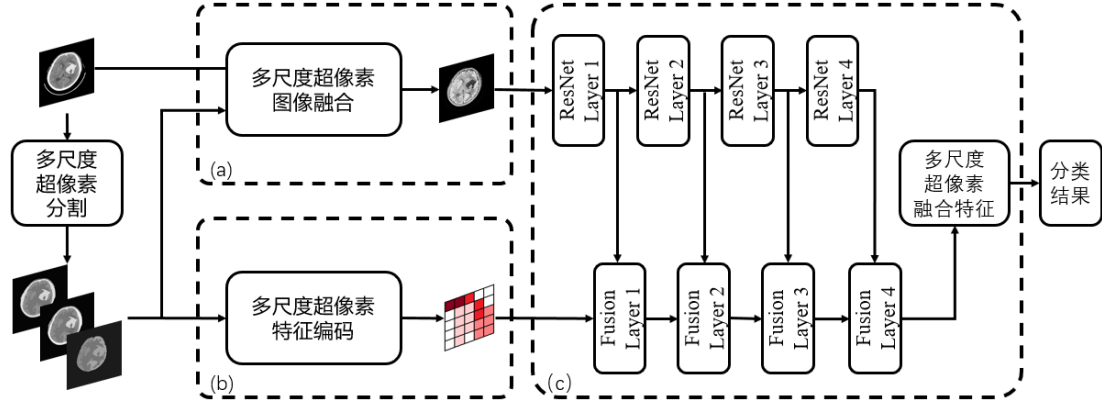


图 1 基于多尺度超像素融合网络的脑 CT 图像分类方法框架图

突出有效病灶区域。但是，由于脑 CT 图像病灶区域的像素值、形状和位置有着极大的不确定性，过低的分割尺度会导致欠分割，将病灶和周围脑组织划分为一个超像素，不能准确分割病灶区域；过高的分割尺度不能有效清除图像中的噪声，并且病灶内部可能会因过分割而产生新的冗余。因此，本文利用超层次分割算法(Super Hierarchy, SH)生成多个分割尺度的超像素图，以达到有效去噪并准确分割病灶区域的目的。

SH 采用最小生成树 MST(Minimum Spanning Tree, MST)算法对图像像素进行迭代聚类，从而实现超像素分割。首先，将图像  $I \in \mathbb{R}^{N \times N}$  表示为含有  $N^2$  个顶点的无向图。然后，使用基于边缘收缩的 Boruvka 算法<sup>[23]</sup>求解 MST，对顶点进行聚类，并将顶点连接而成的子树压缩为一个顶点继续进行聚类，以加速超像素生成。最后，通过迭代的方式连接所有顶点后，就有效地构建了一个超像素层次结构，可以在该结构中动态生成 1 到  $N^2$  间任意数量的超像素，以此获得不同尺度超像素图。

为了保留图像像素信息，本文将多尺度超像素图和脑 CT 图像进行融合，以去除噪声和冗余信息，降低病灶和周围脑组织像素的灰度相似性，获得更加有效的融合图像。对给定的脑 CT 图像  $I \in \mathbb{R}^{N \times N}$ ，首先，基于超层次分割算法 SH<sup>[25]</sup>，和超像素的分割尺度数量  $S$ ，生成第  $s$  个分割尺度下的超像素图  $P^s = \{p_{ij}^s\}_{N \times N}$ ：

$$P^s = SH(I, scale^s) \quad (1)$$

其中， $s \in \{1, 2 \dots S\}$ ， $scale^s$  为第  $s$  个分割尺度， $\mathcal{P} = \{P^1, P^2, \dots, P^S\}$  表示包含  $S$  个不同尺度超像素图的超像素图集合。

然后，将集合  $\mathcal{P}$  中超像素图与脑 CT 图像  $I$  进行像素加权得融合图像  $I'$ ：

$$I' = I \odot \sum_{s=1}^S [f(W)]_s \cdot P^s \quad (2)$$

其中， $f(\cdot)$  为 Softmax 函数， $W$  为表示训练的权重，可以实现自适应分配各个尺度的比重，能够更好的融合多尺度超像素信息。

## 1.2 多尺度超像素特征编码模型

为了有效提取多尺度超像素图底层信息，辅助图像高层次特征完成复杂脑 CT 图像的分类任务，本文设计了一种多尺度超像素特征编码算法。

在脑 CT 超像素图中，病灶往往表现为高密度影或低密度影的小面积区域，而背景、颅骨、脑组织等部位往往被分割为大面积区域。因此，为了更加关注小面积的病灶区域，

本文提出基于区域和边界信息的多尺度超像素特征编码算法，生成能够有效表达病灶区域的低层次特征。

首先，对分割尺度为 $scale^s$ 的超像素图 $P^s$ ，及其像素值集合 $\mathcal{G}^s = \{g_1^s, g_2^s, \dots, g_{scale^s}^s\}$ ，为每个像素生成映射矩阵集合 $M^s = \{M^{s,1}, M^{s,2}, \dots, M^{s, scale^s}\}$ ，其中，第 $k$ 个映射矩阵 $M^{s,k}$ 中第 $i, j$ 个元素 $M_{ij}^{s,k}$ 的计算方式如下：

$$M_{ij}^{s,k} = \begin{cases} 1 & p_{ij}^s = g_k^s \\ 0 & others \end{cases} \quad (3)$$

如图 2 所示， $k \in \{1, 2, \dots, scale^s\}$ ， $M^{s,k}$ 表示像素值为 $g_k^s$ 的超像素区域映射， $M^s$ 表达了超像素的边界信息。

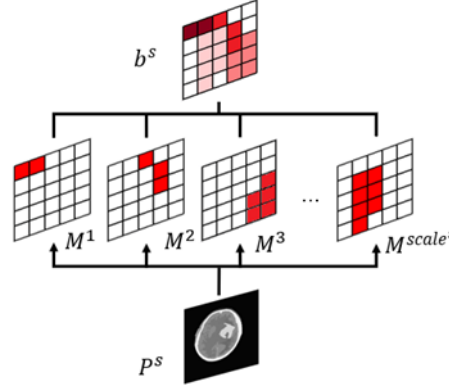


图 2 多尺度超像素特征编码算法

然后，基于面积和边界信息对 $M^s$ 中映射矩阵进行编码，得到当前分割尺度 $scale^s$ 下的超像素图编码结果 $b^s$ ：

$$b^s = \sum_{k=1}^{scale^s} g_k^s \cdot \frac{N^2 - s_k}{N^2} \odot M^{s,k} \quad (4)$$

其中 $g_k^s$ 表示第 $k$ 个超像素的灰度值， $s_k$ 表示第 $k$ 个超像素的像素点个数，可以代表对应超像素的面积， $N^2$ 表示超像素图所含像素点个数。超像素面积越小， $b^s$ 越大，这意味着其更加关注面积较小的病灶区域。最后将 $\mathcal{P}$ 中每个超像素图的编码结果通过矩阵拼接成 $B = [b^1, b^2, \dots, b^S]$ ，即为多尺度超像素编码低层次特征，能够更好的表达小面积病灶区域。

### 1.3 多尺度超像素特征融合分类模型

经过多尺度超像素的特征编码算法得到的特征为图像尺度的低层次特征，具有病灶信息丰富、语义信息少的特点，而神经网络可以提取更抽象、语义信息丰富的高层次特征。充分利用高层次特征和低层次特征，可以实现信息的互补，有利于提升分类的准确率。因此，本文提出了一种多尺度超像素特征融合分类模型。该模型首先使用神经网络提取融合图像的多个高层次特征，然后将其和 1.2 小节得到的低层次特征依次融合，最后进行脑 CT 图像分类，如图 3 所示。

首先，本文选择传统 CNN 中的残差神经网络<sup>[9]</sup>（Residual Network, ResNet）中的 ResNet-34 作为主干网络来提取图像特征，使用前文提取的脑 CT 融合图像 $I'$ 作为网络输入。随着网络层次升高，特征尺寸越来越小、感受野越来越大、语义信息越来越丰富。在同一尺寸特征中，网络越深层输出的特征往往具有越丰富的语义信息。由于 ResNet-34 中残差结构(Basic Block)输出的通道数共有四种，分别是 64、128、256 和 512。一般将输出通道数相同的残差结构归为同一层网络，以此可以将 ResNet-34 划分为 Layer1、Layer2、Layer3、Layer4 四个层级网络，本文按顺序选择每个层级网络中最后一个残差

结构的特征激活输出 $l_1$ 、 $l_2$ 、 $l_3$ 、 $l_4$ 作为主干网络的高层次特征，用来和本文所提低层次特征进行融合。

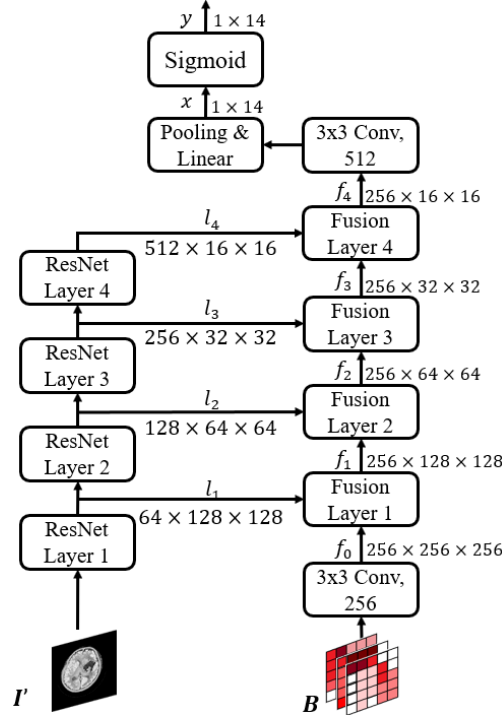


图3 多尺度超像素特征融合分类模型

然后，对上一节通过多尺度超像素编码算法提取的低层次特征 $B$ 进行降维处理，通过 256 个  $3 \times 3$  卷积核构成的卷积层生成特征 $f_0$ 。虽然 $f_0$ 作为浅层特征具有非常少的语义信息，但是由于没有经过过多的下采样，感受野较小，依旧保留了准确的病灶区域和边界信息。将这一特征自下而上依次和 ResNet-34 中不同尺寸的高语义特征 $l_1$ 、 $l_2$ 、 $l_3$ 、 $l_4$ 进行多层融合，生成融合特征 $f_1$ 、 $f_2$ 、 $f_3$ 、 $f_4$ ，其中不同尺寸特征融合模型如图 4 所示。

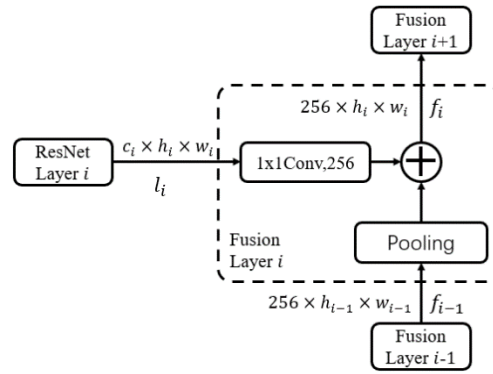


图4 第 $i$ 层特征融合模型

图 4 虚线框中展示的是第 $i$ 层融合模型 Fusion Layer  $i$  ( $i \in \{1,2,3,4\}$ )。先采用池化和卷积操作分别将特征 $f_{i-1}$ 和高层次特征 $l_i$ 转换为通道为 256，大小为 $h_i \times w_i$ 的特征矩阵。之后通过矩阵相加进行特征融合，得到第 $i$ 层融合特征 $f_i$ 。最终得到经过四次融合的特征 $f_4$ 。

最后，将 $f_4$ 输入 512 个  $3 \times 3$  卷积核构成的卷积层、池化层和全连接层得到分类特征，

并进行标签预测。卷积层目的是为了消除 ResNet-34 生成的高层次特征多次叠加计算中产生的冗余信息。池化和全连接层进一步将融合特征变换为和类别标签等长的向量 $x$ ，将其通过 Sigmoid 线性回归生成预测值向量 $y$ ，其中 $x$ 第 $i$ 个元素对应标签为正例的概率 $y_i$ 可以表示为：

$$y_i = \text{Sigmoid}(x_i) \quad (5)$$

在分类模型训练过程中，对于给定的样本标签 $Y$ 和生成的标签预测 $y$ ，我们通过二分类交叉熵损失来计算样本中每个标签的分类误差，之后通过对所有标签分类误差求均值得到样本误差，公式如下：

$$\text{Loss} = -\frac{1}{T} \sum_{i=1}^T (Y_i \log(y_i) + (1 - Y_i) \log(1 - y_i)) \quad (6)$$

其中 $T$ 表示样本中标签的个数，本文使用的数据集中 $T = 14$ 。

## 2 实验结果与分析

### 2.1 实验数据与预处理

由于脑 CT 图像公开数据集的缺失，绝大多数分类研究都使用私有数据集<sup>[10, 14, 16]</sup>。CQ500<sup>[24]</sup>是近年来唯一公开的脑 CT 临床数据集，由印度新德里神经科学和基因学的成像高级研究中心提供，此数据集是在新德里的各个放射中心进行的头部 CT 扫描数据集的子集。其中标签信息包含的 14 种脑部疾病分类：颅内出血(Intracranial Hemorrhage)、脑实质出血(Intraparenchymal Hemorrhage)、脑室出血(Intraventricular Hemorrhage)、硬膜下出血(Subdural Hemorrhage)、硬膜外出血(Extradural Hemorrhage)、蛛网膜下出血(Subarachnoid Hemorrhage, SAH)、左侧脑出血(Left)、右侧脑出血(Right)、慢性出血(Chronic Bleed)、骨折(Fracture)、颅骨骨折(Calvarial Fracture)、其他骨折(Other Fracture)、中线偏移(Midline Shift)、质量效应(Mass Effect)。

CQ500 仅为公开的测试用数据集，我们基于实际获得的 451 例患者扫描数据构建了一个包含训练集、验证集、测试集的脑 CT 数据集，验证本文所提方法的有效性。构建数据集的过程如下：首先，考虑到三名放射学专家对同一个标签可能有不同标注，针对不统一的标注，我们选择多数的专家的选择作为真实的标签；然后，根据确定的颅内出血标签信息将数据集分为确诊患者 204 例，未确诊患者 247 例；最后，从确诊患者病例中根据标签信息选取相关病灶位置的脑 CT 图像共 742 张，从未确诊患者病例中选取和标签包含病灶相同位置的脑 CT 图像共 1045 张，合计 1787 张脑 CT 图像作为本文的数据集。

### 2.2 实验设置和评价指标

在本文实验中，训练集、验证集和测试集的划分基于提取的 1787 张脑 CT 图像，划分比例为 8:1:1，在划分过程中我们首先将每一名患者的图像通过患者 id 绑定为一组图像来处理，保证同一名患者的图像仅出现在训练集、验证集和测试集其中之一，确保实验更接近实际情况。

本文基于 PyTorch 开源框架，使用残差神经网络 ResNet-34 作为主干网络，每次迭代训练时的批大小为 16，使用 Adam 梯度下降算法最小化损失函数，更新参数时学习率  $Lr$  的大小的计算为：

$$Lr = 0.95^{\text{epoch}} \cdot lr \quad (7)$$

其中  $\text{epoch}$  为训练集迭代训练次数， $lr$  为初始学习率，大小为  $1 \times 10^{-4}$ 。训练共迭代数据集 25 次，每次完整迭代后测试模型在验证集上的分类准确率，并记录准确率最高的一次模型作为最优模型。

实验采用目前广泛使用的准确率(Accuracy, ACC)，灵敏度(Sensitivity, SEN)，特异

度 (Specificity, SPE), 正预测率 (Positive predictive value, PPV), 负预测率 (Negative predictive value, NPV) 和 F 评估值 (F-score, F) 综合评价模型的分类性能, 六种指标的计算公式如下:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$SEN = \frac{TP}{TP + FN} \quad (9)$$

$$SPE = \frac{TN}{TN + FP} \quad (10)$$

$$PPV = \frac{TP}{TP + FP} \quad (11)$$

$$NPV = \frac{TN}{TN + FN} \quad (12)$$

$$F = \frac{2TP}{2TP + FP + FN} \quad (13)$$

其中, TP 表示实际标签为正, 预测标签为正的样本数量; TN 表示实际标签为负, 预测标签为负的样本数量; FP 表示实际标签为负, 预测标签为正的样本数量; FN 表示实际标签为正, 预测标签为负的样本数量。

### 2.3 多尺度超像素脑 CT 图像融合模型对分类性能的影响

在测试融合模型对分类性能的影响之前, 需要确定超像素图的尺度。我们首先使用超层次分割算法生成脑 CT 图像在  $scale \in \{5, 10, 15, 20\}$  分割尺度下的超像素图。在  $scale = 5$  时, 超像素图的复杂度最低, 冗余信息最少, 但是会出现不能有效分割病灶位置的情况; 在  $scale = 20$  时, 超像素图都可以明显的分割出病灶区域, 但是会对病灶内部过度分割增加复杂度, 且对外部冗余信息处理效果不佳。因此本文实验使用以上四个尺度的超像素图, 不再对更低或更高尺度的超像素图进行讨论。

我们测试了仅使用脑 CT 图像、超像素图、多尺度超像素图、融合图像通过残差神经网络。ResNet-34 提取特征进行分类的效果。本文 Baseline 方法为基准模型, 基于 ResNet-34 网络实现脑 CT 图像分类。由于输入图像的尺寸与模型输入不同, 因此将原模型中平均池化层改为自适应池化层, 将最后一个全连接层的输出维度改为 14, 以完成本文的脑 CT 图像多标签分类任务; SH<sub>5</sub>、SH<sub>10</sub>、SH<sub>15</sub>、SH<sub>20</sub> 分别表示仅使用 5、10、15、20 尺度超像素图进行分类; SH<sub>5,10</sub>、SH<sub>10,15</sub>、SH<sub>15,20</sub>、SH<sub>5,10,15</sub>、SH<sub>10,15,20</sub> 和 SH<sub>5,10,15,20</sub> 表示将对应超像素图经过矩阵拼接后进行分类; Add<sub>5,10</sub>、Add<sub>10,15</sub>、Add<sub>15,20</sub>、Add<sub>5,10,15</sub>、Add<sub>10,15,20</sub> 和 Add<sub>5,10,15,20</sub> 表示将对应超像素图拼接后通过矩阵对位相加和脑 CT 原图像融合并进行分类; Fusion<sub>5,10</sub>、Fusion<sub>10,15</sub>、Fusion<sub>15,20</sub>、Fusion<sub>5,10,15</sub>、Fusion<sub>10,15,20</sub> 和 Fusion<sub>5,10,15,20</sub> 表示使用本文提出的算法将对应的超像素图 and 原图像融合并进行分类。实验结果如表 1 所示。接下来我们依次分析超像素尺度选择、融合模型对分类性能的影响。

(1) 超像素尺度选择对性能的影响。从表 1 中 SH<sub>5</sub>、SH<sub>10</sub>、SH<sub>15</sub>、SH<sub>20</sub> 的结果可以看出, 使用单一尺度超像素图进行分类任务时, 随着超像素尺度的增加分类效果有所上升, 但准确率达不到使用脑 CT 原图像时 Baseline 的效果。其主要原因是超像素图相比于脑 CT 图像缺少了细节信息, 影响了分类性能; 高尺度超像素图最接近原图, 拥有更多的像素关键信息, 因此相对低尺度像素图分类性能有所提升。SH<sub>5,10</sub>、SH<sub>10,15</sub>、SH<sub>15,20</sub>、SH<sub>5,10,15</sub>、SH<sub>10,15,20</sub> 和 SH<sub>5,10,15,20</sub> 采用了多个尺度超像素图进行分类, 但分类效果也低于



多个尺度中最高尺度超像素图及原图的分类结果。这是因为多个尺度超像素图依然缺少了像素细节信息，加入的低尺度超像素图并未提供有效的辅助特征。因此需要将原图像和多尺度超像素图相结合，在保留图像像素信息的同时，去除噪声和冗余信息，降低病灶和周围脑组织像素的灰度相似性，获得更加有效的融合图像，提升脑 CT 图像分类效果。

表 1 脑 CT 图像、超像素图、多尺度超像素图与融合图像对分类性能的影响

	Method	ACC %	SEN %	SPE %	PPV %	NPV %	F%
脑 CT 图像	Baseline	90.61	56.82	<b>96.99</b>	<b>78.16</b>	92.24	65.8
超像素图	SH <sub>5</sub>	86.25	37.14	95.54	61.17	88.93	46.2
	SH <sub>10</sub>	89.14	59.88	94.67	67.98	92.58	63.7
	SH <sub>15</sub>	89.74	61.79	95.03	70.14	92.93	65.7
	SH <sub>20</sub>	89.83	62.04	95.09	70.49	92.98	66.0
多尺度超像素图	SH <sub>5,10</sub>	87.87	46.07	95.78	67.35	90.38	54.7
	SH <sub>10,15</sub>	88.26	48.06	95.86	68.68	90.71	56.5
	SH <sub>15,20</sub>	88.19	50.77	95.08	65.53	91.29	57.2
	SH <sub>5,10,15</sub>	89.06	62.61	94.06	66.58	93.01	64.5
	SH <sub>10,15,20</sub>	89.07	54.59	95.59	70.06	91.76	61.4
	SH <sub>5,10,15,20</sub>	88.06	42.18	96.73	70.93	89.85	52.9
融合图像	Add <sub>5,10</sub>	85.66	30.44	95.83	57.37	88.21	39.8
	Add <sub>10,15</sub>	86.39	47.19	93.61	57.63	90.59	51.9
	Add <sub>15,20</sub>	88.41	59.52	93.73	63.64	92.63	61.5
	Add <sub>5,10,15</sub>	88.70	54.59	95.59	70.06	90.31	55.9
	Add <sub>10,15,20</sub>	90.54	71.80	94.09	69.66	94.64	70.7
	Add <sub>5,10,15,20</sub>	90.65	63.03	95.87	74.27	93.20	68.2
	Fusion <sub>5,10</sub>	89.59	67.18	93.72	66.33	93.94	66.8
	Fusion <sub>10,15</sub>	89.84	66.33	94.17	67.71	93.82	67.0
	Fusion <sub>15,20</sub>	89.70	69.64	93.39	66.00	94.35	67.8
	Fusion <sub>5,10,15</sub>	91.76	72.71	95.34	74.94	94.84	73.8
	Fusion <sub>10,15,20</sub>	<b>91.82</b>	<b>72.79</b>	95.42	75.02	<b>94.88</b>	<b>73.9</b>
	Fusion <sub>5,10,15,20</sub>	91.52	68.16	95.93	76.02	94.09	71.9

(2) 图像融合模型对性能的影响。从表 1 中 Add 等实验结果可以看出，将多尺度超像素图和脑 CT 图像对位相加的传统融合方法效果不理想，为了对不同尺度的超像素图进行一定的取舍后再进行融合，本文提出的多尺度超像素的脑 CT 图像融合方法 Fusion 训练了一个可以实现自适应分配各个尺度的比重，能够更好的融合多尺度超像素信息，降低图像复杂度和冗余信息的同时保证了病灶区域的准确分割，提升了脑 CT 图像分类任务的效果。相比于 Add 方法，Fusion 方法在使用相同多尺度超像素的情况下均获得了更好的分类性能。

本文所提 Fusion 方法在分割尺度数量  $S=3$  时，Fusion<sub>5,10,15</sub> 和 Fusion<sub>10,15,20</sub> 方法在 ACC, SEN, NPV 和 F 指标上均获得很好的表现，然而在描述标签为负的样本分类效果的指标 SPE，和描述分类预测标签为正样本的准确率指标 PPV 上表现相对一般，不如对比实验的结果。原因如下：首先，本文使用的数据集中标签分布相对不均，负标签占有比例较大，约为总标签数的 83%；然后，对比方法的模型分类能力不佳，难以实现正标签的正确预测，在训练过程中为了获得更低的损失值，致使模型预测结果更倾向于占比更多的负标签；最后，在统计对比方法分类结果时，TN 值偏大、FP 值偏小，从而导致 SPE 和 PPV 指标虚高。虽然这代表了对比方法的负样本分类准确率和预测正样本准确率较高，但与此同时正样本分类准确率 SEN 相对较低，这说明模型正样本分类准确率欠佳，对于本文脑 CT 图像分类任务而言不能有效识别患病患者，漏诊率偏高。因此，需要结合多种指标来共同评价分类模型的效果。本文模型虽然 SPE 和 PPV 表现相对一般，但综合考虑 ACC, SEN, NPV 和 F 指标，本文提出的模型性能更优。

Fusion 模型在分割尺度数量  $S=3$  时的分类效果最优，分割尺度数量更高时 ( $S=4$ )



的分类效果有所下降，原因是四个尺度的超像素图存在信息冗余，影响了融合效果。观察图 5 可见，最低尺度  $scale=5$  与最高尺度  $scale=20$  的超像素图间差异较大：在低分割尺度超像素图中，病灶区域被欠分割；而在高尺度超像素图中，非病灶区域被过分割。采用图中四种尺度超像素进行图像融合时，模型生成的融合图像同时含有区域欠分割和过分割信息，冗余信息影响了融合图像质量，从而导致模型收敛效果欠佳。

综上所述，本文在后续实验中使用脑 CT 图像在分割尺度  $scale \in \{5,10,15,20\}$  下超像素图，多尺度超像素的分割尺度数量  $S=3$ ，即多尺度超像素由 5、10、15 和 10、15、20 两种尺度超像素图组成。

(3) 可视化展示。我们对  $Fusion_{5,10,15}$  融合过程中多尺度超像素与权重的加权图和多尺度超像素融合图像  $I'$  进行了可视化展示，如图 5 所示。可以清晰的看出，加权图很好的去除了颅脑外的冗余信息，准确的划分表达病灶信息，有助于提升分类性能。

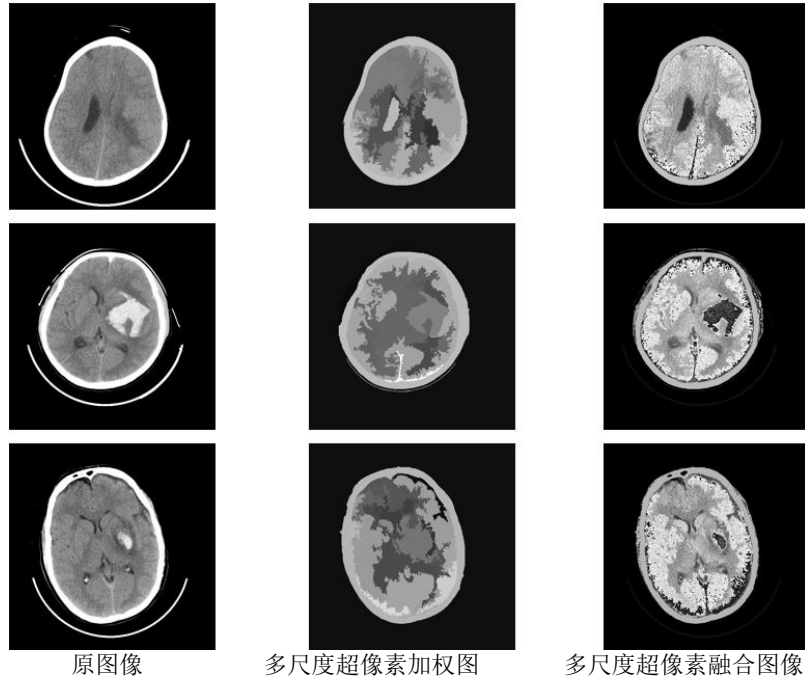


图 5 多尺度超像素可视化图

## 2.4 基于多尺度超像素融合网络的脑 CT 分类方法

本节测试了基于多尺度超像素融合网络的脑 CT 图像分类方法的消融实验及相关对比实验结果。其中，多尺度超像素选择使用 5、10、15 和 10、15、20 两种尺度的超像素图的组合。MSFN 表示本文提出的多尺度超像素融合网络，其主要包含以下三部分：图像融合、特征编码算法和特征融合模型。

(1) 为了验证各个部分对分类任务的影响，本文进行了如下的消融实验来验证有效性：MSFN<sub>100</sub> 表示在 Baseline 基准方法基础上使用融合图像  $I'$  作为输入；MSFN<sub>001</sub> 表示在 Baseline 方法基础上，加入了不同层次特征融合模型；MSFN<sub>011</sub> 方法表示在 MSFN<sub>001</sub> 基础上使用了多尺度超像素特征编码算法；MSFN<sub>101</sub> 表示在 MSFN<sub>001</sub> 的基础上使用融合图像  $I'$  作为输入。实验结果如表 2 所示。

a) 多尺度超像素脑 CT 图像融合模型的有效性。通过对比 MSFN<sub>011</sub> 和 MSFN 的分类效果可以看出：不使用多尺度超像素脑 CT 图像融合的情况下分类效果对比 Baseline 有了少许提升，但仍不及本文的 MSFN 方法，使用脑 CT 原图像相比多尺度超像素图像融合，缺少了冗余信息去除、病灶区域划分、病灶区域灰度相似性降低等优势，最终限制了分类效果。

表 2 基于多尺度超像素融合网络的脑 CT 图像分类结果

Method	图像融合	特征编码器	特征融合	SH Scale	ACC %	SEN %	SPE %	PPV %	NPV %	F%	Paras (10 <sup>7</sup> )
Baseline	0	0	0	—	90.61	56.82	96.99	78.16	92.24	65.8	2.129
MSFN <sub>001</sub>	0	0	1	—	91.57	59.64	97.61	82.49	92.75	69.2	2.272
MSFN <sub>011</sub>	0	1	1	5,10,15	91.36	60.88	97.12	80.00	92.92	69.1	2.272
				10,15,20	91.61	59.55	<b>97.67</b>	<b>82.85</b>	92.74	69.3	
MSFN <sub>101</sub>	1	0	1	5,10,15	91.16	64.35	96.23	76.35	93.45	69.8	2.272
				10,15,20	91.90	68.07	96.40	78.16	94.11	72.8	
MSFN <sub>100</sub>	1	0	0	5,10,15	91.40	73.04	94.89	73.10	94.88	73.1	2.129
				10,15,20	91.82	72.79	95.42	75.02	94.88	73.9	
MSFN	1	1	1	5,10,15	92.54	73.71	96.48	79.40	94.75	75.4	2.272
				10,15,20	<b>92.66</b>	<b>74.44</b>	96.11	78.33	<b>95.21</b>	<b>76.3</b>	

b) 超像素特征编码算法的有效性。通过对比 MSFN<sub>101</sub> 和 MSFN 的分类结果可以看出：不使用超像素特征编码算法的情况下分类效果虽然对比 Baseline 有了显著提升，但由于直接使用多尺度超像素图没有参考区域面积和边界信息，表达不佳，不能直接作为低层次特征，导致和高层特征的融合特征效果不好，和 MSFN 分类效果有较大的差距。

c) 特征融合模型的有效性。本文测试了输入为脑CT图像和融合图像两种情况下，不同层次特征融合模型对分类结果的影响。由于对比实验取消了不同层次特征融合模型后，不能处理来自多尺度超像素图的低层次特征，因此MSFN<sub>001</sub>和MSFN<sub>101</sub>中也均未使用超像素特征编码模型，特征融合模型仅对残差神经网络中不同层次的中间层特征进行多层融合。通过MSFN<sub>001</sub>和Baseline，MSFN<sub>100</sub>和MSFN<sub>101</sub>两组对比实验可以看出，不论输入是原图像还是经过多尺度超像素融合后的图像，与仅使用ResNet提取特征相比特征融合模型的使用都显著的提升了分类的效果。本文提出的方法可以综合利用ResNet中得到的多种尺寸高层次特征，通过融合各个尺寸下的富含语义信息的特征，针对脑CT图像中区域面积差异大的病灶，利用多种尺寸下的特征能更好的表达病灶信息，为分类任务提供了更具有判别性的特征。另外，表 2 中参数量（Paras）表明，本文实验中，模型参数数量的增加和模型准确率的提升没有必然联系。

(2) 为了验证本文提出的多尺度超像素融合网络在脑CT分类上的有效性，我们参考了最新的医学图像分类工作<sup>[25]</sup>，使用几种常用的深度网络模型<sup>[26-28]</sup>进行了对比实验，如表 3 所示。其中ResNet-101<sup>[9]</sup>在 ResNet-34 的基础上，将残差结构中卷积层由两层增加到三层，并使用了更多数量残差结构加深网络；Inception v4<sup>[26]</sup>通过增加网络宽度，利用不同尺寸的小卷积核的结果进行拼接，实现不同尺度的特征提取，通过丰富每层的信息提高网络性能；Inception-Resnet-v2<sup>[26]</sup>在Inception模型基础上引入了残差连接，以适应深层网络的训练。Residual Attn<sup>[27]</sup>通过堆叠多个注意力模块生成注意力感知特征，来自不同注意力模块的感知特征会随着神经网络层级的深入而发生适应性变化。SENet<sup>[28]</sup>通过对特征通道间的相关性进行建模，把重要的特征进行强化来提升准确率。MSFN表示本文提出的多尺度超像素融合网络在使用 10、15、20 尺度超像素图下的模型。

由表 3 可以看出本文提出的算法在对比实验中获得了更高的准确率。相比于其他方法，本文提出的多尺度超像素特征融合算法以ResNet-34 为主干网络，以多尺度超像素图辅助特征提取，获得了更具有判别性的融合特征，能更有效地表达脑CT图像中病灶区域信息，从而提升了分类效果。

通过消融实验和相关对比实验，验证了本文提出的MSFN方法在脑CT图像分类任务中的有效性。基于多尺度超像素脑CT图像模型融合生成的融合图像有效的降低了图像的噪声，并且准确地划分了病灶区域；多尺度超像素编码器提取准确的低层次特征，能更好的关注小面积病灶；特征融合模型通过融合两种不同层次的特征生成了更具判别性的融合特征，对形态位置多变的病灶区域有更有效的表达。因此，合理利用多尺度超像素的信息并同时应用不同层次特征有利于提升脑CT图像分类性能。

表 3 和其他方法的对比实验结果

Method	ACC %	F%
Baseline(ResNet-34)[9]	90.61	65.63
ResNet-101[9]	89.38	49.48
Inception v4[26]	91.41	72.89
Inception-Resnet-v2[26]	90.52	67.69
Residual Attn[27]	90.52	65.10
SENet[28]	86.17	32.79
MSFN	<b>92.66</b>	<b>76.30</b>

### 3 结 论

本文提出了一种基于多尺度超像素融合网络的脑 CT 疾病诊断分类方法。该方法首先使用基于多尺度超像素的脑 CT 图像融合模型实现脑 CT 图像数据增强，得到融合图像；然后利用多尺度超像素特征编码模型提取区域和边界信息，得到低层次特征；最后利用残差神经网络提取融合图像的高层次特征，并通过多尺度超像素特征融合分类模型将其和低层次特征进行多层融合，使用融合后特征进行脑 CT 图像分类任务来验证方法的有效性。相关实验表明本文所提方法能够有效提取并利用多尺度超像素中的信息，获得的融合特征更具有判别性，在脑 CT 疾病诊断分类任务上优于传统 CNN 方法。尽管我们只在脑 CT 图像分类任务上进行了实验，但是所提方法对应脑 CT 图像的特征提取是通用的，可以广泛用于基于脑 CT 图像的各类任务。

#### [参考文献] (References)

- [1] Gao X H W, Hui R. A Deep Learning Based Approach to Classification of CT Brain Images [C]. Proceedings of the 2016 Sai Computing Conference, 2016, 28-31.
- [2] Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs [C]. Jama-J Am Med Assoc, 2016, 316(22): 2402-2410.
- [3] Van Grinsven M J J P, Van Ginneken B, Hoyng C B, et al. Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images [C]. IEEE T Med Imaging, 2016, 35(5): 1273-1284.
- [4] 邢新颖, 冀俊忠, 姚垚. 基于自适应多任务卷积神经网络的脑网络分类方法[J]. 计算机研究与发展, 2020, 57(7): 1449-1459.  
Xing X Y, Ji J Z, Yao Y. Brain Networks Classification Based on an Adaptive Multi-Task Convolutional Neural Networks[J]. Journal of Computer Research and Development, 2020, 57(7): 1449-1459. (in Chinese)
- [5] Lecun Y, Boser B, Denker J S, et al. Backpropagation Applied to Handwritten Zip Code Recognition [J]. Neural Comput, 1989, 1(4): 541-551.
- [6] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks [C]. Commun Acn, 2017, 60(6): 84-90.
- [8] Szegedy C, Liu W, Jia Y Q, et al. Going Deeper with Convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 1-9.
- [9] He K M, Zhang X Y, Ren S Q, et al. Deep Residual Learning for Image Recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770-778.
- [10] Garland J, Ondruschka B, Stables S, et al. Identifying Fatal Head Injuries on Postmortem Computed Tomography Using Convolutional Neural Network/Deep Learning: A Feasibility

- Study [J]. *J Forensic Sci*, 2020(65): 2019-2022.
- [11] Gautam A, Raman B. Towards effective classification of Brain hemorrhagic and ischemic stroke using CNN [J]. *Biomed Signal Proces*, 2021, 63.
  - [12] Mansour R F, Aljehane N O. An optimal segmentation with deep learning based inception network model for intracranial hemorrhage diagnosis [J]. *Neural Comput Appl*, 2021, 1-13.
  - [13] Anupama C S S, Sivaram M, Lydia E L, et al. Synergic deep learning model-based automated detection and classification of Brain intracranial hemorrhage images in wearable networks [J]. *Pers Ubiquit Comput*, 2020, 1-10.
  - [14] Grewal M, Srivastava M M, Kumar P, et al. Radnet: Radiologist Level Accuracy Using Deep Learning for Hemorrhage Detection in Ct Scans [C]. *IEEE 15th International Symposium on Biomedical Imaging*, 2018, 281-284.
  - [15] Li J Q, Fu G H, Chen Y D, et al. A multi-label classification model for full slice Brain computerised tomography image [J]. *Bmc Bioinformatics*, 2020, 21(Suppl 6):200.
  - [16] 惠瑞, 高小红, 田增民. 基于深度学习的 CT 脑影像分类方法用于阿尔茨海默病的初步筛查 [J]. *中国医疗设备*, 2017, 32(12): 5.  
Hui R, Gao X H, Tian Z M. CT Brain Image Classification Based on Deep Learning in Application of Screening of Alzheimer Disease[J]. *China Medical Devices*, 2017, 32(12): 5. (in Chinese)
  - [17] Blaivas L, Blaivas M. Are Convolutional Neural Networks Trained on ImageNet Images Wearing Rose Colored Glasses? A Quantitative Comparison of ImageNet, Computed Tomographic, Magnetic Resonance, Chest X-Ray, and Point-of-Care Ultrasound Images for Quality [C]. *J Ultras Med*, 2021(40), 377-383.
  - [18] Ren X F, Malik J. Learning a classification model for segmentation [C]. *Ninth IEEE International Conference on Computer Vision, Vols I and II, Proceedings*, 2003, 10-17.
  - [19] Tighe J, Lazebnik S. SuperParsing: Scalable Nonparametric Image Parsing with Superpixels [J]. *Lect Notes Comput Sc*, 2010, 6315:352-365.
  - [20] Wang S, Lu H C, Yang F, et al. Superpixel Tracking [C]. *IEEE International Conference on Computer Vision (ICCV)*, 2011, 1323-1330.
  - [21] Zhang J, Kan C, Schwing A G, et al. Estimating the 3D Layout of Indoor Scenes and its Clutter from Depth Sensors [C]. *IEEE International Conference on Computer Vision (ICCV)*, 2013, 1273-1280.
  - [22] Wei X, Yang Q, Gong Y, et al. Superpixel Hierarchy [C]. *IEEE Trans Image Process*, 2018, 27(10): 4838-4849.
  - [23] West D B. *Introduction to Graph Theory* [M]. Prentice Hall, 1996.
  - [24] Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study [C]. *Lancet*, 2018, 392(10162): 2388-2396.
  - [25] Liao Z, Liao K, Shen H, et al. CNN Attention Guidance for Improved Orthopedics Radiographic Fracture Classification [J]. *IEEE Journal of Biomedical and Health Informatics*, 2022.
  - [26] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning [C]. *Thirty-First AAAI Conference on Artificial Intelligence*, 2017, 4278-4284.
  - [27] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* [C], 2017: 6450-6458.
  - [28] Hu J, Shen L, Sun G. Squeeze-and-excitation networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* [C], 2018: 7132-7141.