

基于改进 ResNet50 的抗 VEGF 治疗结果预测研究

陈正雄¹ 黄星宇¹

CHEN Zhengxiong HUANG Xingyu

摘要

糖尿病黄斑病变是一种常见的糖尿病的并发症，该病的发病率较高并且很容易导致失明，而目前治疗方式主要是注射抗血管内皮生长因子（简称 Anti-VEGF）药物治疗，由于有大量患者对这种疗法无反应或反应不充分，于是预测其有效性就变得较为重要。文章使用眼底 OCT（光学相干断层成像）图像作为研究目标，选择了 ResNet50 作为主体网络，通过单张 OCT 图像对网络进行训练，使网络能够在 OCT 图像数据缺失时，通过单张 OCT 图像来实现对治疗结果进行预测。对网络进行了改进，首先在其中加入 SE 注意力机制，使模型能够自适应地调整不同通道之间的权重，使模型能够更加关注与对当前任务更有意义的特征通道，有效提高了预测的准确率。然后，在此基础上以 Ghost 模块替代网络中的部分卷积层，模块通过将部分特征图通过简单的线性变换产生，以这种方式来代替以往的全部由卷积产生特征图的方式，有效的减少模型的参数量，抑制了参数数量的增长。结果表明，改进网络在原网络上仅提升了 27.7% 参数量的情况下提升了 2.2% 的精确度，取得了较好的效果。

关键词

糖尿病性黄斑病变；图像分类；卷积神经网络；深度学习

doi: 10.3969/j.issn.1672-9528.2024.12.034

0 引言

糖尿病作为一种慢性非传染性疾病，其危害性极为严重，成为继心脑血管疾病与肿瘤后，对人体健康造成重大威胁的又一关键病症^[1]。目前，全国有超过 1.1 亿成年人罹患糖尿病，糖尿病患者人数位居世界之首^[2]。众多患者不仅面临着患上致命并发症的风险，而且半数以上处于尚未确诊状态，这无疑进一步加剧了风险程度^[3]。在糖尿病众多并发症里，糖尿病性黄斑病变非最为严重，但因其可致视力受损乃至失明，对患者生活影响极大，成为大部分患者最难以接受的并发症。因视力障碍会严重干扰日常生活的各个方面，极大地降低生活质量并增加生活的不便性与依赖性。据统计，I 型、II 型糖尿病患者因糖尿病视网膜病变而致盲的比例分别为 86%、33%^[4]。目前针对该病的主要治疗方式之一，多采用玻璃体腔内注射抗 VEGF 药物。但因药物较为昂贵，且存在注射后效果不佳的可能，不仅浪费医疗资源，给病人造成较大的经济压力，更会因为治疗方式的错误而导致病情延误。因此，对于糖尿病性黄斑病变的治疗结果的预测极为重要。

随着人工智能的发展，深度学习逐渐进入到医学领域，在皮肤癌检测、乳腺癌检测、血管检测等方面取得突破的同时，在糖尿病视网膜病变的分期和检测中也取得了不错的成

效。丁蓬莉等人^[5]提出的 CompactNet 对眼底图像进行识别分类，但因实验样本不足等问题，导致准确度较低；庞浩^[6]提出了一个多分支结构的卷积神经网络用于彩色眼底图像分类，主要将 GoogleNet、VGGNet 等作为特征提取网络，后使用一个两级的 CNN 网络依靠提取的特征来进行分类，以此实现对彩色眼底图像的检测。这些研究在准确检测的基础上为医疗工作者减轻了不少负担，也有力的验证了卷积神经网络对复杂数据的表达能力非常有助于对眼底图像的分类、预测等等方面的研究。因此，采用深度学习算法来预测糖尿病性黄斑病变患者接受抗 VEGF 治疗的效果，不仅技术上是可行的，而且从医学研究和临床实践的角度来看，具有显著的意义和价值，可以为更精确可靠的诊断带来希望。

1 数据集收集与处理

1.1 数据集简介

本文使用的眼底图像均来源于天池数据集，其包含了 1000 多名患者的 9912 张眼底 OCT 图像，并包含了大量指标作为标签。同时，考虑到目标是对抗 VEGF 药物治疗的结果预测，所以选择根据是否继续注射抗 VEGF 治疗，将数据集分为两类，一是继续注射即为治疗有效类；二是不再继续注射即为治疗无效类。在对数据集进行整理后，得到有效类图像 5275 张，无效类图像 4637 张。

1. 武汉轻工大学电气与电子工程学院 湖北武汉 430023

1.2 数据的预处理

数据集经过严格的质量控制与筛选后，具有较高的准确性和可靠性，因此无需进行复杂的数据清洗和修改工作。值得一提的是，本文所进行的数据预处理主要集中在数据的整理与格式化上。首先，按照研究需求对原始数据进行分类和筛选，剔除部分标签不全的图像。其次，对数据格式进行统一处理，确保所有数据项都符合研究要求，以便后续数据分析和处理。这些预处理工作为后续数据分析奠定了坚实的基础，确保研究的准确性和可靠性。

1.3 数据的归一化

数据归一化是将原始数据转化为特定范围或分布的过程，其目的是消除不同数据之间的尺度差异，使数据更加可比较和可分析。在机器学习和深度学习的数据预处理中，数据归一化是非常重要的步骤。本文采用线性函数转化的方法，实现公式为：

$$y = \frac{x - \min \text{Value}}{\max \text{Value} - \min \text{Value}} \quad (1)$$

式中： x 、 y 分别为转换前后的值； $\max \text{Value}$ 与 $\min \text{Value}$ 分别为样本的最大值和最小值。

2 训练模型设计

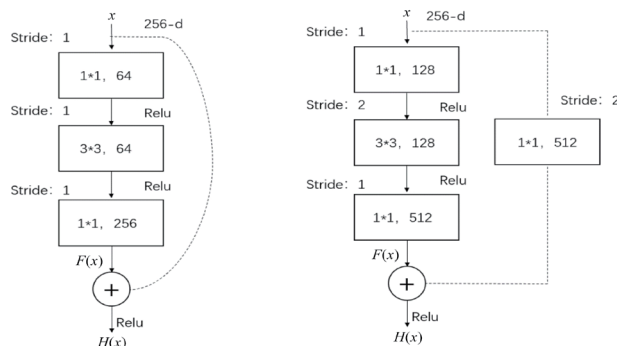
2.1 ResNet 网络

ResNet（深度残差网络）^[7] 是由何凯明等人于 2015 年提出的神经网络结构，该网络在当年 ImageNet 竞赛中取得分类任务、目标检测以及图像分割等多个领域冠军。在诸多任务中性能突出，因此受到广泛关注，并在各个领域引用，均取得较好效果。

其核心思想是通过引入一种创新的结构——残差结构，解决深度神经网络在训练过程中可能出现的退化问题。具体而言，残差结构在原始网络中加入“短路连接”，即直接将输入绕过一层或者多层，与深层的输出相加，简化学习目标，进行有效训练，保持性能，甚至提升性能。

残差结构如图 1 所示，图中虚线部分为 Shortcut，即残差连接。图 1（a）中的残差块输入与输出的矩阵深度维度一致；图 1（b）的展示中，输入矩阵灵活调整其深度维度，以无缝对接至后续层的架构需求。残差块的核心在于其输出 $H(x)$ 的构造方式，将原始输入特征矩阵 x 与经过一系列复杂卷积处理，得到的特征矩阵 $F(x)$ 直接相加的结果。随后，这一和值通过激活函数 Relu 的进一步处理，生成最终的输出矩阵 $H(x)$ ，即数学表达为 $H(x) = F(x) + x$ 。这一过程体现了残差学习的精髓，有助于缓解深层网络训练中的梯度消失问题，并提升网络的整体性能。因此，由于主干上运算所得的 $F(x)$ 和分支上所输入的 x 的特征矩阵的维度需要相同，所以图中

使用了 5 个 1×1 的卷积核，用以升维或者降维，使 x 的特征矩阵的维度与 $F(x)$ 保持一致。其目的不仅在于直接传递信息，还巧妙地实现了与主路径上特征矩阵在空间维度上的对齐，即同样使高度和宽度减半，从而确保两者可以直接相加，实现残差学习的目的。



(a) 同层使用的残差块 (b) 跨层使用的残差块
图 1 残差块结构

2.2 算法的改进

2.2.1 SE 注意力机制

SENet（squeeze-and-excitation networks）即 SE 注意力机制，它的核心思想是在通道维度上增加注意力机制，使模型能够自适应地调整不同通道之间的权重，更加关注与对当前任务更有意义的特征通道^[8]，并对当前任务用处较小的特征通道进行抑制，使模型能够更好地学习，以获得更高的精确度，是一种非常优秀的方法。

如图 2 所示，SE 注意力机制通过一系列操作得到一个 $1 \times 1 \times C$ 地权重矩阵，对原特征图进行重构，从而突出其重要特征，并抑制相对非重要特征，以达到更好的效果，具体实现过程分为四步。

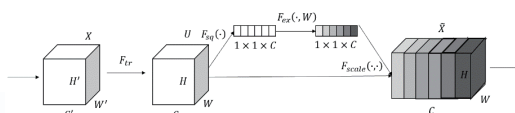


图 2 SE 算法流程图

第一步：给定一个输入特征图 X ，让其经过 F_{ir} 操作生成特征图 U 。

第二步：Squeeze 操作 $F_{sq}(\cdot)$ ：这一步将特征图进行全局平均池化，生成一个 $1 \times 1 \times c$ 的向量，每个数值表示一个通道。公式为：

$$z_c = F_{sq}(u_c) = \frac{1}{H * W} \sum_{i=1}^H \sum_{j=1}^W u_{c(i,j)} \quad (2)$$

第三步：Excitation 操作 $F_{ex}(\cdot)$ ：这是一个精细调整特征通道重要性的步骤，通过精心设计的两层全连接层来实现。首先，利用学习得到的权重矩阵 W ，该权重矩阵是在训练过

程中自动优化以捕捉特征间的相关性，从而生成所需的权重信息。然后，将两个全连接层（ W_1 和 W_2 ）被顺序应用于前一阶段输出的特征向量 z 上，进行信息的压缩与再扩展处理。这一过程不仅精炼了特征表示，还生成了通道权重值 s ，其中每个值都对应着原始特征图中一个通道的权重，反映了该通道对于最终输出贡献的重要性。经过两层全连接层的变换， s 中的不同数值直观地反映出不同通道间的权重差异，实现对通道重要性的动态调整，展示了从输入 z 到输出 s 的转换过程，具体表现为：

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

第四步：Scale 操作 $F_{scale}(\cdot)$ ：从算法流程图的逻辑结构中理解，这一核心操作是将第三步中计算得到的权重向量 s 应用于特征图 U 上，实现了一种基于通道重要性的加权机制。这一步骤的目的是对特征图 U 中的每个通道根据其对应的权重值 s 进行动态调整，从而突出重要特征并抑制次要特征，最终生成特征图 X 。

2.2.2 Ghost 模块

在网络中引入注意力机制，对模型进行改进，虽然网络精确度得到一定的提升，但是网络的参数和计算量也相应增加，即对模型的训练速度和检测速度产生影响。为减少这一影响，选择对模型进行轻量化改进。从而对网络再次进行优化，以达到减少参数数量的目的。

GhostNet 是由华为诺亚方舟实验室与北京大学等单位合作提出的，是一种轻量级的神经网络架构，目的在于解决卷积神经网络部署到嵌入式设备时遇到的内存资源以及计算资源限制的问题^[10]。GhostNet 的核心在于其 Ghost 模块，模块通过生成幻影特征图（ghost feature maps）用来替代以卷积的方式产生特征图，从而有效减少参数量以及提升运算速度。

从图 3 可以看到，传统的卷积计算与 Ghost 操作的区别就在于，Ghost 操作经过少量计算，只产生一部分的特征图而不是全部，再通过这部分特征图进行简单计算，从而得到另一部分的特征图，再将其进行拼接，从而得到完整的特征图，有效减少参数量，提升模型的运算速度。

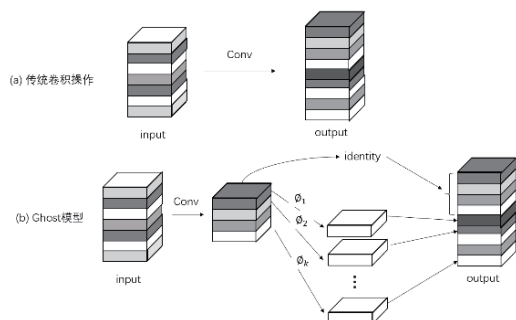


图 3 传统卷积操作和 Ghost 模型处理过程

图 3（a）为传统的卷积操作生成特征图的过程，在这一操作中，假设输入特征是 $X \in \mathbf{R}^{c \times h \times w}$ ，卷积核表示为

$f \in \mathbf{R}^{c \times k \times k \times n}$ ，输出特征是 $Y \in \mathbf{R}^{h' \times w' \times n}$ ，则常规卷积操作可以表示为：

$$Y = f * X + b \quad (4)$$

式中： h' 和 w' 分别是输出数据的高度和宽度； $k \times k$ 分别是卷积核 f 的内核大小，在这个卷积过程中，卷积核数量 n 和通道数 c 通常非常大，这个操作的计算量是： $n \times h' \times w' \times c \times k \times k$ 。

图 3（b）为 Ghost 模块操作生成特征图的过程，由于在传统卷积过程中，输出特征里面有很多特征是相似的“特征”，因此，无需使用大量的 FLOP 和参数来生成冗余的特征图。所以，Ghost 模块选择以一种更廉价的操作来生成所需的特征图，具体分为四个步骤：生成基础特征图、深度可分离卷积、生成幻影特征图、特征图拼接。

假设只有 m 个原始特征图 $Y \in \mathbf{R}^{h' \times w' \times m}$ 是使用一次卷积生成的。即：

$$Y' = X * f' \quad (5)$$

式中： f' 表示一个具有维度 $\mathbf{R}^{c \times k \times k \times m}$ 的卷积核，其中 m 是不大于 n 的一个整数为了保持与传统卷积操作的一致性，设定所有相关的超参数均与普通卷积操作中的相同。这样，输出特征的空间大小 h' 和 w' 保持一致。通过这种方式可以获得 m 个特征图，为将 n 个特征图补齐，需要对已经得到的 m 个特征图中的每一个特征图实施一次线性变换，得到 s 个“幻影”特征图，这样，通过结合原始的特征图和它们的“幻影”特征图，可以够构建出足够数量的特征图来满足后续处理的需求，具体为：

$$y_{ij} = \phi_{ij}(y_i'), \quad \forall i = 1 \cdots m, j = 1 \cdots s \quad (6)$$

式中： y_i' 被指定为序列中的第 i 个原始特征图，当特征图 y_i' 经过特定的线性变换 ϕ_{ij} ，会转变为序列中的第 j 个“幻影”特征图 y_{ij} ，这意味着，对于每一个原始特征图 y_i' 可以通过 s 次不同的线性变换（包括一次恒等映射作为第 s 次变换）来生成 s 个“幻影”特征图 $\{y_{ij}\}_{j=1}^s$ 。而最后的 $\phi_{i,s}$ 是用于保留原始特征图的恒等映射。

通过使用这种简单的方式，获得 $n = m \times s$ 个特征图 $y = [y_{11}, y_{12}, \cdots, y_{ms}]$ 作为 Ghost 模块的输出结果，这一过程中，线性变换 ϕ 被独立应用于每个通道，显著降低了计算需求，相较于传统卷积操作而言，实现了计算量的大幅度减少。

3 实验与分析

在对数据集进行整理后，部分数据存在缺失，在眼科疾病中，医生可以通过一张高质量的 OCT 图像对病情进行初步判断^[11]，也就表明在单张的 OCT 图像中已存在足够的特征信息用作为判断治疗结果的条件。因此，本文通过使用单张图像对网络进行训练，使网络通过单张 OCT 图像进行学习，网络便可以通过单张 OCT 图像对治疗结果进行预测，这有效

地解决了图像数据缺失对预测产生影响，使图像缺失时仍能得到预测的结果。

3.1 实验条件

本实验是在 AutoDL 服务器上进行，采用的 PyTorch 深度学习框架作为开发环境。配置及环境如表 1。

表 1 实验环境配置

参数	配置
操作系统	Linux
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C
GPU	RTX 3080
Torch	1.7.0
CUDA	11.0
Python	3.8
优化算法	SGD
Batch size	32
epochs	600

损失函数为交叉熵损失函数，交叉熵损失函数计算公式为：

$$L = \frac{1}{N} \sum_{i=1}^N - [y_i \log p_i + (1 - y_i) \log (1 - p_i)] \quad (7)$$

式中：在二分类任务中，用 P_i 代表模型预测第 i 类的概率大小； N 代表样本的总个数； y_i 代表样本 i 的分类标签，本次实验设定标签 0 代表治疗无效类，标签 1 代表治疗有效类。

3.2 性能评价指标

本文选取分类精确度（AC）作为模型的主要评判标准^[12]。其计算公式为：

$$P = \frac{TP}{TP + FP} \quad (8)$$

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

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

式中：TP 为真正例，即正确分类为属于该类的样本数；FP 为真负例，即为错误分类为属于该类的样本数；TN 为假正例，为错误分类为不属于该类的样本数；FN 为假负例，即正确分类为不属于该类的样本数。

3.3 网络参数

ResNet 在分类任务上效果较好，考虑到模型深度增加的同时，也会大幅增加参数量，在进行权衡后选择了 ResNet50 作为主体网络进行改进。为获得更高精确度，在网络的残差块的卷积层之后添加了 SE 注意力机制。因添加了注意力机制，则导致模型参数量进一步变大，为解决这一问题，又在网络中加入了 Ghost 模块来对网络进行轻量化改进。

表 2 是改进前后所组成的 3 个模型参数量，可以清晰看出改进前后参数量变化。在加入 SE 模块模块后比原 ResNet50 网络的模型参数量提升了 18.6×10^6 个，而在加入 Ghost 模块后，最终模型的参数量在 ResNet50+SE 模型的基础上减少了 11.5×10^6 个，在 ResNet50 基础上少量的增加了 7.1×10^6 个。由此可见，轻量化的改进达到了该有的效果，有效抑制了参数的膨胀，达到预想的效果。

表 2 改进前后模型参数量

模型	参数量/(10^6 个)
ResNet50	25.6
ResNet50+SE	44.2
ResNet50+SE+Ghost	32.7

3.4 消融实验

为对比改进前后的指标变化情况，分别搭建 3 个网络模型，并在同样条件下进行训练。训练结束后对结果进行分析，图 4 为训练过程中的训练损失的曲线图，三角线条、菱形线条、方形线条分别代表着 ResNet50、ResNet50+SE 以及 ResNet50+SE+Ghost。从图中可以看出，3 个模型的损失均达到稳定状态，说明模型均是收敛的，训练均是有效的。并且从图中还可以看到，加入 SE 注意力机制以后，模型训练过程中收敛速度变得更快，模型的训练过程也更稳定，证明取得较好效果。而在加入 Ghost 模块后，训练损失的曲线的收敛速度是有所下降的，这也是模型轻量化不可避免的问题，虽然出现了一定程度变差，但是其与 ResNet50 对比，依旧获得更好的效果。证明无论是对于注意力机制的加入或 Ghost 模块的加入，都达到了预期效果，证明了其意义和有效性。

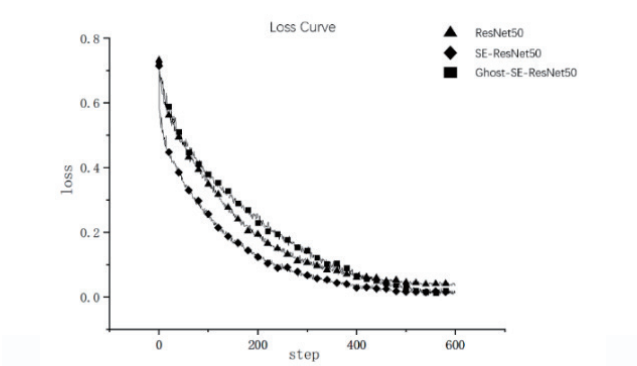


图 4 模型改进前后的损失曲线

图 5 是 3 个模型的精确度的曲线，同样，三角线条、菱形线条、方形线条分别代表着 ResNet50、ResNet50+SE 以及 ResNet50+SE+Ghost。通过图 5 可以清楚地看出 3 个模型都达到了较高的精确度，而 3 条曲线中，仅加入 SE 注意力机制的模型 ResNet50+SE 的精确度是最高的，其次为 ResNet50+SE+Ghost，ResNet50 的精确度稍差。

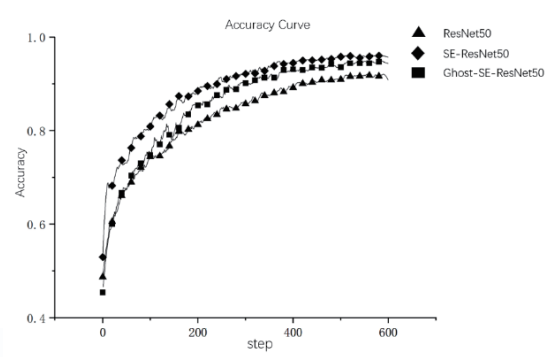


图 5 模型改进前后的精确度曲线

通过表 3 的数据可知 ResNet50+SE 在 ResNet50 的基础上精确度提高了 3.1%，ResNet50+SE+Ghost 则在 ResNet50 的基础上提高了 2.2%，虽然轻量化导致已提升的精确度有所下降，但是对于模型参数来说，是更划算的。由于 ResNet50+SE 在精确度上提升了 3.1%，而参数量上提升了 72.6%，ResNet50+SE+Ghost 在精确度上提升了 2.2%，而在参数量上仅仅提升了 27.7%。简而言之，ResNet50+SE+Ghost 是在更小的参数量增加下达成了更好的效果，以较小的代价获得了精确度提升。

表 3 改进前后模型的精确度

模型	AC/%
ResNet50	92.3
ResNet50+SE	95.4
ResNet50+SE+Ghost	94.5

3.5 对比试验

为验证模型的优势，将其于 ResNet 其他层数的网络以及其他网络进行了对比，分别有 VGG19，AlexNet，ResNet34，ResNet101，性能对比见表 4，可以看出，改进后的模型在预测的精确度高于其他网络，而参数量上也有着极大的优势。

表 4 不同层数网络的对比

检测网络	AC/%	参数量/(10 ⁶ 个)
VGG19	89.7	20.5
AlexNet	90.3	60
ResNet34	91.7	22.7
ResNet101	92.5	44.6
SE+Ghost+ResNet50	94.5	32.7

4 结语

针对糖尿病性黄斑病变的抗 VEGF 治疗结果的预测，本文在 ResNet50 网络的基础上进行了改进，并通过使用单张 OCT 图像对网络进行训练，使得模型能够在 OCT 图像数据不全时的实现治疗结果的有效性进行预测，该算法在 ResNet50 网络中加入了 SE 注意力机制，并且嵌入了 Ghost

模块，在有效提高精确度的同时解决了参数量变得太大的问题，有效抑制了参数量的膨胀，使网络的检测速度得到保证。网络在参数量仅仅增加 27.7% 的同时使精确度提升了 2.2%，使糖尿病性黄斑病变的抗 VEGF 治疗结果的预测有较高的准确度。

参考文献：

[1] 杨文英. 中国糖尿病的流行特点及变化趋势 [J]. 中国科学 (生命科学), 2018,48(8):812-819.

[2] 徐子和, 杨丽, 龚研, 等. 基于深度学习的糖尿病性视网膜病变的分级 [J]. 电脑知识与技术, 2022,18(3):1-3+9.

[3] 陈超, 杨冰洁, 叶珂男, 等. 基于深度学习的糖尿病性视网膜病变的眼底图像辅助诊断系统 [J]. 人工智能, 2021(3): 38-46.

[4] 陈霞琳, 汪迎, 张凯, 等. 糖尿病性视网膜病变相关危险因素分析 [J]. 临床眼科杂志, 2013,21(4):320-322.

[5] 丁蓬莉, 李清勇, 张振, 等. 糖尿病性视网膜图像的深度学习神经网络分类方法 [J]. 计算机应用, 2017,37(3):699-704.

[6] 庞浩. 基于深度卷积神经网络的医学影像诊断关键技术研究 [D]. 北京: 北京邮电大学, 2019.

[7] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE, 2016:770-778.

[8] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018:7132-7141.

[9] 刘树林, 周晓燕, 王玉林. 基于改进 YOLOv5 针对舰船识别的算法研究 [J]. 信息技术与信息化, 2023(8):188-193+198.

[10] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020:1577-1586.

[11] 杨柳茂. 基于 OCT 图像的视网膜疾病分类方法研究 [D]. 长沙: 湖南大学, 2018.

[12] 胡名琪, 刘秋明, 陈辉明, 等. 基于改进残差网络与迁移学习的铜合金金相图分类方法 [J]. 有色金属科学与工程, 2014, 15(5):690-699.

【作者简介】

陈正雄（1999—），男，湖北孝昌人，硕士，研究方向：深度学习、图像分类。

黄星宇（1999—），男，湖北钟祥人，硕士，研究方向：深度学习、目标检测。

（收稿日期：2024-09-12）