

# 基于非局部网络与通道注意力机制的颈动脉狭窄分类模型

李博涛<sup>1, a)</sup>, 郑瀚<sup>2)✉</sup>, 罗继昌<sup>3)</sup>, 张桃红<sup>1, a)✉</sup>, 焦力群<sup>3)✉</sup>, 金

紫薇<sup>4)</sup>

1) 北京科技大学 计算机与通信工程学院, 北京 100083

2) 河池学院 广西高校人工智能与信息处理重点实验室, 广西 546300

3) 首都医科大学 宣武医院, 北京 100053

4) 北京工业大学信息学部计算机学院, 北京 100124

a) 材料领域知识工程北京市重点实验室, 北京 100083

✉ 通信作者, E-mail:张桃红, [zth\\_ustb@163.com](mailto:zth_ustb@163.com); 郑瀚, [05057@hcnu.edu.cn](mailto:05057@hcnu.edu.cn); 焦力群, [894900372@qq.com](mailto:894900372@qq.com)

**摘要** 颈动脉狭窄是缺血性脑卒中的主要成因之一, 目前数字减影造影技术(DSA)被称为颈动脉狭窄诊断的金标准, 但传统的诊断方式需要由病理学家手动筛选分析 DSA 影像, 存在着筛查速度慢、容易出错及对专业诊断人员的依赖等问题。人工智能为我们提供了辅助诊断手段, 但目前的识别往往诊断出一处狭窄就完成识别, 而实际影像有时存在不止一处的问题, 为提高影像对多处狭窄的识别能力, 本文提出了一个非局部通道注意力网络 (Non-Local Channel Attention Net, NLCA Net) 对颈动脉狭窄进行准确分类。该模型主要由两个模块构成: 非局部多尺度特征融合模块 (Non Local Multi-Scale Fusion module, NLMSF) 和通道注意力模块 (Multi-Level Channel Attention module, MLCA)。非局部多尺度特征融合模块 NLMSF 利用非局部网络的思想来模拟空间注意力操作, 同时, 为了更好的提取多尺度特征, 在非局部网络中还加入了多尺度特征融合的模块, 对颈动脉影像分类起到重要作用; 通道注意力模块 MLCA 通过高效的利用影像中的通道特征, 为模型分类提供了更多的语义信息。我们通过使用提取关键帧的技术, 建立颈动脉狭窄数据集, 将本文模型与其他主流的医学影像分类模型在该颈动脉狭窄数据集上进行对比。我们的模型达到了最好的效果, 模型的分​​类准确率要高于其他主流模型至少 2%。

**关键词** 颈动脉狭窄; 数字减影造影术; 非局部网络; 通道注意力; 多尺度特征融合

**分类号** TG142.71

## Carotid Stenosis Classification Model Based on Non-local Networks and Channel Attention Mechanism

LI Botao<sup>1, a)</sup>, ZHENG Han<sup>2) ✉</sup>, LUO Jichang<sup>3)</sup>, ZHANG Taohong<sup>1, a) ✉</sup>, JIAO Liqun<sup>3) ✉</sup>, JIN Ziwei<sup>4)</sup>

1) School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), Beijing 100083, China

2) School of Artificial Intelligence and Information Processing, Hechi University, Key Laboratory of Guangxi Colleges and Universities, Guangxi 546300, China

3) Xuanwu Hospital, Capital Medical University, Beijing 100053, China

4) School of Computer Science, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

a) Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China

✉ Corresponding author, E-mail:ZHANG Taohong,[zth\\_ustb@163.com](mailto:zth_ustb@163.com); ZHENG Han,[05057@hcnu.edu.cn](mailto:05057@hcnu.edu.cn); JIAO Liqun,[894900372@qq.com](mailto:894900372@qq.com)

收稿日期: 2025-02-07

**基金项目：**科技部科技创新 2030 重大项目资助（2020AAA0108703）

**ABSTRACT** Carotid artery stenosis stands as one of the primary culprits behind ischemic stroke, a leading cause of morbidity and mortality worldwide. The precise diagnosis of this condition is crucial for effective patient management and treatment. At present, Digital Subtraction Angiography (DSA) is regarded as the gold standard for diagnosing carotid artery stenosis, offering clear and detailed images that enable clinicians to visualize the degree of narrowing in the artery. However, traditional diagnostic approaches often rely on pathologists to manually scrutinize and interpret these images, a process that is inherently time-consuming, prone to human error, and heavily reliant on the expertise of specialized diagnostic professionals. In this context, the advent of artificial intelligence has brought forth promising auxiliary diagnostic tools, aimed at enhancing the efficiency and accuracy of medical image analysis. Yet, despite significant advancements, many current recognition systems are limited in their capacity to detect only a single area of stenosis within an image. Such limitations could lead to incomplete or inaccurate diagnoses, potentially compromising patient outcomes.

To address these challenges and enhance the diagnostic capabilities for carotid artery stenosis, this paper introduces a novel approach: the Non-Local Channel Attention Network (NLCANet). This advanced network architecture is specifically designed to accurately classify carotid artery stenosis. By leveraging the power of non-local attention mechanisms and channel-wise feature extraction, NLCANet offers a more robust and nuanced approach to image analysis, ultimately ensuring more precise and reliable diagnostic results. Through this innovation, we aim to not only advance the field of medical image classification but also to significantly improve clinical outcomes by enabling faster, more accurate diagnoses for patients suffering from carotid artery stenosis.

The proposed model is built upon two integral components that work in harmony to enhance the performance of carotid artery stenosis classification: the Non-Local Multi-Scale Fusion module (NLMSF) and the Multi-Level Channel Attention module (MLCA). NLMSF draws inspiration from the concept of non-local networks, which are designed to capture long-range dependencies and relationships within an image. By simulating spatial attention operations, this module enables the model to focus on critical regions of the image, regardless of their location. This attention mechanism ensures that the model considers global contextual information rather than just local features, allowing for a more holistic understanding of the image. In addition to spatial attention, the NLMSF also integrates a multi-scale feature fusion strategy. This is particularly important for medical images, where lesions or stenotic regions may vary in size and scale. MLCA, on the other hand, focuses on enhancing the model's ability to prioritize and utilize channel-wise information. Medical images often contain rich, multi-dimensional data across different channels, and each channel may contribute differently to the classification task. The MLCA module is designed to efficiently identify which channels carry the most relevant semantic information for the classification of carotid artery stenosis, enabling the model to weight them accordingly, which in turn boosts classification accuracy and reduces the potential for misdiagnosis. By combining both spatial and channel-level attention mechanisms, the model is empowered to capture a more nuanced understanding of the images, ultimately leading to more accurate and reliable diagnostic predictions. By using key frame extraction techniques, we established a carotid artery stenosis dataset and compared our proposed model with other mainstream medical image classification models on this dataset. Our model achieved the best performance, with a classification accuracy that is at least 2% higher than other mainstream models.

**KEY WORDS** Carotid artery stenosis; Digital Subtraction Angiography; Non-local networks; Channel attention; Multi-scale feature fusion

## 1 引言

颈动脉狭窄是导致缺血性脑卒中的重要病因之一，具有较高的发病率和致残率。数字减影造影技术（Digital Subtraction Angiography, DSA）作为当前诊断颈动脉狭窄的金标准，提供了高分辨率的血管成像，使得医生能够观察到血管中的细微变化。然而，传统的 DSA 影像分析依赖于病理学家的手动筛选和分析，这一过程不仅耗时，还容易受到操作者的经验和疲劳度的影响，导致诊断的准确性和效率受到限制。此外，手动分析 DSA 影像的复杂性和主观性也使得诊断结果在不同病理学家之间存在较大的差异性，这可能影响治疗决策的制定和病患的最终治疗效果。因此，提高颈动脉狭窄的诊断效率和准确性，降低医生的工作负担，已成为当前医学影像领域的一项紧迫任务。

随着计算机视觉和机器学习技术的飞速发展，自动化的医学影像分析技术显现出巨大的潜力。通过引入先进的图像处理和分析算法，可以大幅提升影像的处理速度和分析的准确性，从而为颈动脉狭窄的诊断提供更为可靠和高效的技术支持。这不仅可以优化医疗资源的分配，还能显著提高病患的诊疗体验和治疗结果。因此，开发和应用新的技术方法，以克服传统 DSA 影像分析中存在的问题，是提高颈动脉狭窄诊断质量的关键。

计算机视觉作为一门集成电子工程、计算机科学、和人工智能领域的交叉学科，在图像和视频分析方面已经取得了显著的进展。特别是在图像分类任务上，计算机视觉的发展不仅推动了技术的边界，还为多个行业带来了革命性的变化。在图像分类领域，深度学习模型的发展经历了从初步的卷积神经网络（CNN）[1]到引入空间和通道注意力机制[2][3]的高级架构的演进过程。这一发展不仅显著提高了图像分类的准确性，还增强了模型对关键特征的识别能力。CNN 通过其独特的卷积层、池化层和全连接层结构，在图像分类任务中取得了突破性的进展。这些网络能够自动从图像中学习层次化的特征表示，从而避免了传统机器学习方法中依赖手工特征提取的限制。如 LeNet[4]、AlexNet[5]、VGGNet[6]等模型，它们通过加深网络层次和增加网络宽度，逐步提高了图像分类的性能。随着网络结构的深入，训练深层网络的难度也随之增加。ResNet（残差网络）的提出有效解决了深层网络训练中的退化问题。

在医学图像方面，CNNs 可以从数据中自动学习到丰富的颈动脉血管特征[7][8]，在颈动脉图像分割和分类任务中均展示出比传统方法更好的性能[9][10][11]。在一些特定的心脑血管分类任务上，一些研究表明基于 CNNs 训练的深度学习模型甚至可以实现与人类专家相当的识别性能[12][13]。除此之外，一些最近的研究揭示了利用 CNNs 进一步提升颈动脉血管图像智能识别性能和拓展其应用场景的潜力[14][15][16]。

注意力机制的加入是深度学习在图像分类领域的一大创新。注意力机制能够使网络集中于图像中的重要部分，从而提高分类的准确率。这种机制最初在自然语言处理领域获得成功，后来被迅速引入到计算机视觉中。空间注意力机制通过对特征图的不同区域赋予不同的重要性，强化了模型对颈动脉血管关键区域的识别能力。例如，CBAM

（Convolutional Block Attention Module）[17]首先通过一个空间注意力模块来突出特征图中的关键区域，从而更有效地聚焦于颈动脉图像的重要部分。与空间注意力聚焦于“哪里”不同，通道注意力关注于“什么”是重要的。SENet（Squeeze-and-Excitation Networks）[18]是一个突出的例子，它通过对每个通道的全局信息进行压缩和激励，重新校准特征通道的权重，从而强化了模型对不同特征通道的适应能力。

计算机视觉在图像分类任务上的这些发展不仅提高了分类技术的精度和效率，还为诸如自动驾驶汽车、智能视频监控、医学诊断等多个实际应用领域提供了技术支撑。在医学图像分析中，这些技术的应用尤为关键，它们帮助医生在疾病诊断和治疗计划制定中做出

更准确、更快速的决策。

近年来,视觉领域涌现出了许多创新性的深度学习模型,其中包括 ViT[19] (Vision Transformer)、CaiT[20] (Class-Attention in Image Transformers)、ConvNeXt[21]以及 FixEfficientNet[22]等。这些模型各自具有独特的优势。Vision Transformer (ViT)通过将图像分割为多个块,并采用类似于自然语言处理的编码方式,开创了图像识别的新路径。然而,ViT 对大规模数据集的高度依赖性,使其在面对小规模数据集时显得力不从心。在此基础上发展而来的 CaiT,引入了类注意力机制,显著提升了处理复杂图像任务的能力。尽管如此,CaiT 在处理高分辨率图像时,其高昂的计算成本限制了实时应用的可能性。ConvNeXt 则重新审视了卷积神经网络的设计理念,融合现代架构的最佳实践,同时保留了传统 CNN 的优势。虽然在很多场景下表现优异,但在特定任务上,ConvNeXt 可能不及专门设计的 Transformer 模型那样出色。最后,作为 EfficientNet 家族的一员,FixEfficientNet 专注于在效率和准确性之间找到最佳平衡点。尽管该模型在众多基准测试中成绩斐然,但在极其复杂的医疗影像分析任务中,仍需进一步优化以实现更优性能。

同时,深度学习在医学影像分析领域也取得了显著进展,尤其是在颈动脉狭窄的检测与评估方面。Zhou 等人[23]提出了一种三维卷积深度学习框架,用于自动分割 CTA 图像中的颈动脉分叉,这一步骤对于量化颈动脉狭窄和评估中风风险至关重要。该框架由定位和分割两个阶段组成,有效地提升了诊断精度。值得注意的是,颈动脉斑块是导致颈动脉狭窄的主要原因之一,因此,其检测与识别成为早期预防及临床辅助诊断的关键环节。此外,Geng 等人[24]开发了一种基于深度学习的图像增强技术,专门用于消除大焦斑 DSA 图像的模糊问题,从而获得更清晰的脑血管 DSA 图像,为准确诊断奠定了基础。Kimura 等人[25]则引入了“虚拟 DSA”这一创新概念,利用 U-Net 架构,通过输入单个实时图像生成仅包含提取血管的虚拟 DSA 图像,进一步提高了图像质量和诊断效率。DSA 成像作为颈动脉狭窄诊断的“金标准”,广泛应用于颈动脉狭窄检测及颈动脉、颅内动脉的分割任务中。崔楷文[26]提出了一套结合深度学习与 DSA 图像的颈动脉狭窄度评估流程,首先对颈动脉血管进行图像分割,随后在此基础上执行狭窄检测,并将像素分析范围限制在检测到的目标框内以减少干扰因素的影响。最终,通过传统图像处理方法计算狭窄区域与正常区域的像素数量,进而确定狭窄程度。与此同时,Khankari 等人[27]运用卷积神经网络技术来识别急性卒中患者的动脉标志和血管闭塞情况,旨在精确检测颈动脉末梢,并采用相同的训练和分析策略来识别动脉闭塞,展现了深度学习在复杂医疗影像分析中的巨大潜力。

尽管现有颈动脉狭窄分类算法已经具备良好的性能[28][29],但在不同条件、不同类型的图像分类中仍然具有挑战。现存方法在学习判别性特征时忽略了一些局部细节判别性特征,尤其当分类图像中存在噪声,加大了分型任务的难度。与此同时,针对现有工作在学习细粒度图像细节信息时,忽略了深度分类网络的语义特征,这会让模型导致丢失了大量细节特征信息。受到图像分类模型的启发和医学图像智能化分析的需要[30][31][32],我们吸收了之前模型的优点并克服现有模型存在的问题,提出了非局部通道注意力网络

(NLCANet),来辅助人们对颈动脉狭窄进行准确分类,从而提高诊断效率。NLCANet 主要由两大模块组成,分别为非局部多尺度特征融合模块(NLMSF)和多层通道注意模块(MLCA)。为了克服局部细节丢失的问题,我们设计了 NLMSF 模块,这个模块由三个分支组成,三个分支均来自骨干网络的不同层,通过利用不同多尺度特征融合方法来对特征进行处理,实现了对模型局部细节特征的补充。此外,为了更好的定位颈动脉血管关键区域,还利用非局部的思想来模仿空间注意力的操作,在对颈动脉狭窄分型起到了至关重要的作用。我们提出的另一个高效通道注意力模块,旨在克服分类模型忽略了对语义信息的利用,MLCA 模块通过全局平均池化提取通道的全局信息,然后利用一维卷积捕获通道

间的局部依赖，最后通过 sigmoid 函数得到每个通道的权重并对特征进行重标定。此外，通过对骨干网络每一层的输出进行通道注意力处理，有效地增强了每一层特征的表达能力，这种逐层加强的方法确保了模型可以从多尺度的特征中提取出最有用的信息，适用于颈动脉狭窄分类这种效率和性能要求较高的场景。因此，我们的贡献总结如下：

(1) 我们提出了非局部多尺度特征融合模块 (NLMSF)，该模块通过结合来自骨干网络不同层的三个分支，利用多尺度特征融合技术，有效地增强了模型对局部细节特征的捕捉能力，这对于提高图像中关键细节的识别精度至关重要。

(2) 我们提出了多层通道注意模块 (MLCA)，通过全局平均池化和一维卷积的结合，有效地捕获了通道间的局部依赖关系，并通过权重重标定强化了重要的通道特征，从而提高了模型对语义信息的利用率和分类性能。此外，通过对骨干网络每一层的输出进行通道注意力处理，有效地增强了每一层特征的表达能力。

(3) NLCA Net 在颈动脉狭窄数据集上的实验结果显示，其分类准确率高于其他主流医学影像分类模型至少 2%，这证明了 NLCA Net 在实际应用中的有效性和优越性。

(4) 相比需要病理学家手动分析的传统 DSA 影像诊断方法，NLCA Net 提供了一种自动化、高效且准确的颈动脉狭窄分类方案，解决了传统方法中存在的速度慢和易出错的问题。

## 2 相关工作

### 2.1 非局部网络

“Non-local”网络模型的核心特点是在神经网络中引入非局部操作，旨在捕获长范围的依赖关系。这种方法最初受到图像处理和视频分析领域内自然语言处理技术的启发，尤其是注意力机制的应用。

在进行颈动脉狭窄分类时，捕获长范围的依赖关系非常关键，因为它可以帮助模型全面理解和分析医学影像中的复杂特征。颈动脉狭窄的识别不仅需要识别狭窄的区域，还需要分析与之相邻的血管结构和可能的病变扩散。非局部网络通过捕获图像中远距离的空间关系，可以有效地整合这些关键信息，从而提供更精确的分类结果。此外，这种方法还能增强模型对噪声和图像质量变化的鲁棒性，提高诊断的可靠性和准确性。因此，非局部网络在处理需要高度准确性和细节识别的医学影像分类任务，如颈动脉狭窄分类中，展现出极大的潜力和优势。非局部网络的概念最早由 Wang 等人在 2018 年提出[33]，用于视频分类和对象检测。他们的研究展示了通过增加非局部块来捕捉视频帧之间的时空关系，显著提升了模型的表现。随后，这一概念被扩展到其他领域。例如，在医学影像分析中，非局部网络用于增强图像中的关键特征识别[34][35]，通过关注远距离的生物标记物来提高疾病诊断的准确性。为了提升效率，研究者们探索了将非局部操作与其他网络结构（如 CNN 和循环神经网络）的集成方法。此外，也出现了多种优化算法，如降低其计算和内存成本的方法，使得非局部网络能够在资源受限的设备上运行。非局部网络模型通过其独特的处理长距离信息的能力，在多个领域和任务中提供了新的解决方案，极大地扩展了深度学习的应用范围和深度。然而，过多的关注如何更好的捕获长范围的依赖让他们忽略了不同尺度特征也同样重要，只有长范围依赖而只有单一尺度的特征将导致模型分类效果并不理想。我们提出的 NLMSF 在利用非局部网络的捕获长范围依赖的同时也加入了多尺度特征融合的方法，既收集到了长范围依赖又利用了多尺度特征来进行分类，大大提升了模型对颈动脉分型的性能。

### 2.2 通道注意力

通道注意力模块在进行颈动脉狭窄分类时有很多益处，它能够有效地突出影像中与狭窄相关的关键特征。通过对特征通道进行权重分配，这种模块帮助模型识别和放大那些对

分类决策至关重要的特征，如血管壁的厚度、斑块的形状和大小等。这不仅增强了模型对细微病变的检测能力，也提高了分类的准确性和可靠性。此外，通道注意力机制能够自适应地调整其关注点，这对于处理医学影像中常见的各种变异和噪声尤为重要，从而确保即使在图像质量不一的情况下，也能保持高水平的诊断性能。

通道注意力的概念最早由 Hu 等人在 2017 年提出的 SENet (Squeeze-and-Excitation Networks) 中系统地使用。SENet 通过对传统卷积网络加入简单的 SE 块显著提升了网络的性能，并在多个标准数据集上达到了先进水平。继 SENet 之后，多种基于通道注意力的网络结构被提出。例如 CBAM (Convolutional Block Attention Module) 不仅关注通道注意力，还结合了空间注意力，进一步提升了模型的识别能力。相比于复杂的空间操作，通道注意力机制通常涉及较低的计算成本，因此可以轻松集成到各种网络架构中而不显著增加运算负担。

### 3 模型和方法

本文提出非局部通道注意力网络 NLCANet 模型。模型包括两个模块，实现了非局部网络和多尺度特征的融合，并提出高效的通道注意力模块，该模块将进一步把语义信息整合进特征图中，从而达到更好的颈动脉狭窄分类效果。

#### 3.1 模型结构总览

为应对现有深度学习方法在学习颈动脉狭窄分类时忽视局部细节特征的问题，尤其是在图像中存在噪声时，这种忽视加剧了分类的难度。此外，目前的方法在提取细粒度图像细节时往往忽略了深层分类网络中的语义特征，导致大量细节特征信息的丢失。基于图像分类模型的启示和医学图像智能化分析的需求，我们结合了先前模型的优势并解决了现有模型的缺点，提出了非局部通道注意力网络 Non-Local Channel Attention Net (NLCANet)，如图 1 所示。该网络旨在通过精确分类颈动脉狭窄来提高诊断效率。NLCANet 主要包括两个核心模块：非局部多尺度特征融合模块 (NLMSF) 和多层通道注意模块 (MLCA)。

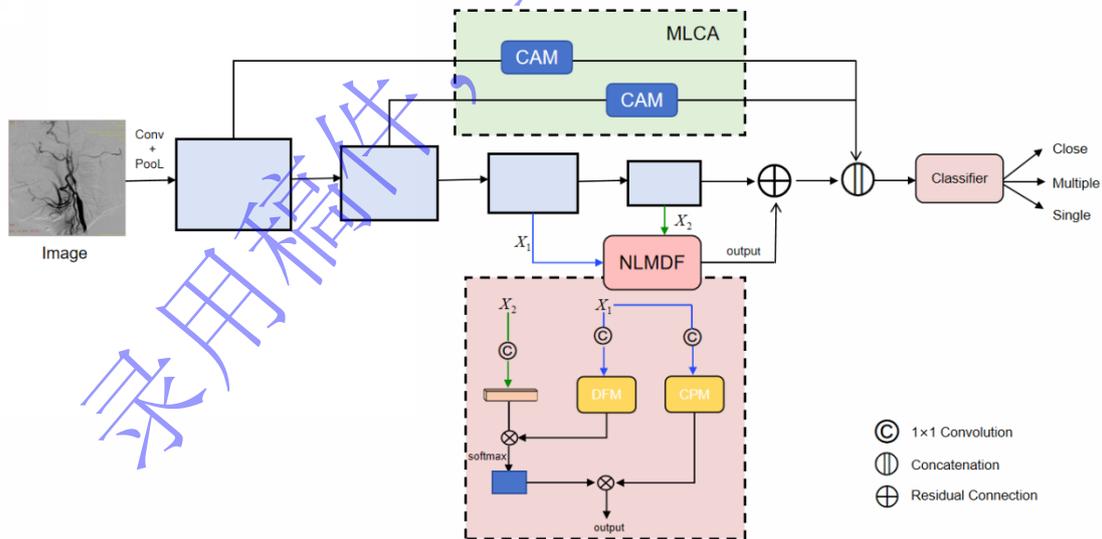


图 1 NLCANet 网络结构图.

Fig.1 NLCANet Network Architecture Diagram.

#### 3.2 非局部多尺度特征融合模块 (NLMSF)

为了克服在对颈动脉狭窄分类时存在局部特征丢失的问题，同时，为了更好的捕捉大范围依赖和利用不同尺度的特征，我们提出了非局部多尺度特征融合模块，如图 1 所示。NLMSF 模块主要利用了非局部网络结构的优势来提升对颈动脉狭窄分类的性能。在处理颈

动脉狭窄分类的医学影像分析中，非局部神经网络表现出显著的优势。这种网络通过其独特的能力在图像中捕捉长距离依赖关系，有效地增强了对颈动脉整体结构及潜在异常区域的全面理解。非局部操作的引入不仅提升了网络对细微变化的敏感性，而且通过与其他多尺度特征融合方法的灵活结合，加强了模型在多尺度上的分析能力。此外，非局部网络能够整合全局信息，有助于在决策过程中考虑整个解剖结构的详细信息，从而提高了分类的准确性。

由于医学影像的特殊性，单单只使用非局部网络并不能达到理想的分类效果。因为，我们将非局部网络和多尺度特征融合方法相结合，在捕获图像中长距离依赖的同时，进一步为特征补充多尺度信息。为了补充多尺度信息，我们设计了扩张融合模块

(DilatedFusionModule) 和组合池化模块 (CombiPoolModule)，分别如图 2 和图 3 所示。

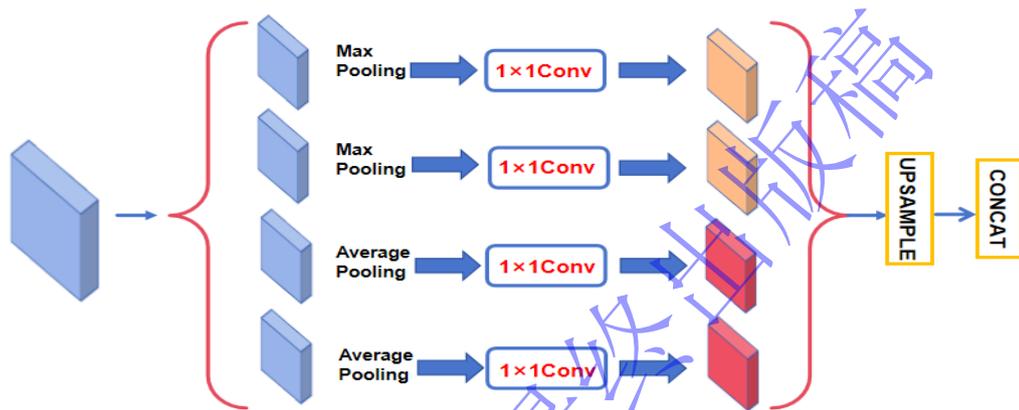


图 2 组合池化模块结构图.

Fig.2 Combined Pooling Module Architecture Diagram.

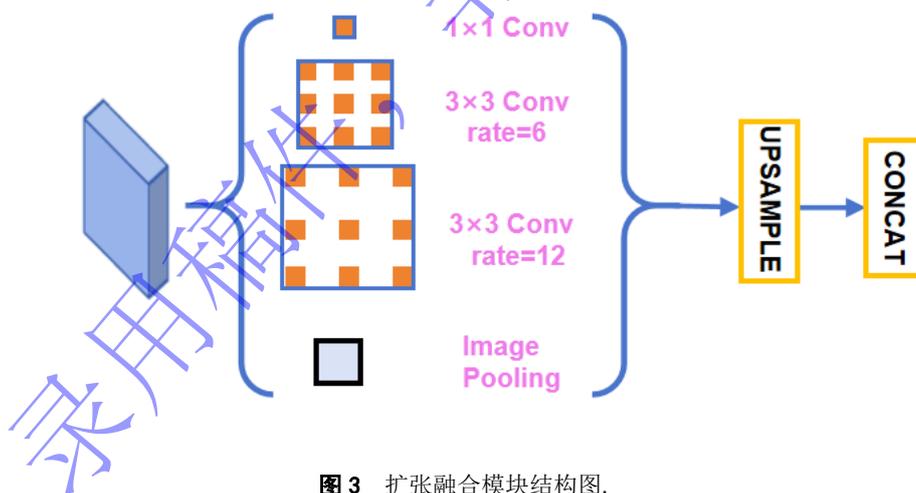


图 3 扩张融合模块结构图.

Fig.3 Dilated Fusion Module Architecture Diagram.

扩张融合模块由三个不同扩张率的空洞卷积、1个全局平均池化和1个 $1 \times 1$ 卷积操作组成。使用空洞卷积可以在增大感受野的同时不会产生丢失局部信息。在颈动脉狭窄分类中，这意味着模型可以更细致地捕捉到颈动脉内壁的细微结构变化，如斑块的形成和狭窄区域的具体特征，这些都是判断狭窄程度的关键因素。通过使用不同学习率的多个空洞卷积，扩张融合模块能够在多个尺度上捕捉图像特征，从局部细节到更广泛的区域特征。这种多尺度特征集成有助于模型在分析颈动脉的整体和局部情况时获得更全面的信息，从而提高分类的准确性和可靠性。池化层在模块中起到降维和特征聚合的作用，有助于减少计

算复杂性同时保持关键信息，确保即使在特征维度较高时，网络也能高效运行，同时避免过拟合。 $1 \times 1$  卷积层作为特征压缩和融合工具，它可以将来自不同空洞卷积层的特征图进行整合，减少特征通道数量，同时保留重要的特征信息。我们通过结合多种类型的卷积和池化操作，增强了模型对不同图像质量和不同病变类型的适应性。这种设计使得模型在处理不同患者的颈动脉图像时，能够更好地泛化。

组合池化模块由四个并行的不同池化率的最大池化层和平均池化层组成。最大池化层通过提取每个特征区域的最大值来突出图像中的显著特征，如边缘和纹理。在颈动脉狭窄分类中，最大池化有助于捕捉颈动脉内壁的突出部分和狭窄区域，这些是识别狭窄非常重要的视觉线索。同时，平均池化层通过计算特征区域的平均值，平均池化有助于提供更平滑的特征表征，减少噪声的影响，从而在保持图像整体信息的同时，有助于捕捉更加细微的差异。这对于辨别颈动脉狭窄程度的轻微变化尤其重要。单一的池化策略可能导致某些类型的信息丢失。例如，最大池化可能忽略掉非最大值的区域信息，而平均池化可能平滑掉重要的局部特征。通过并行使用这两种池化，模型可以在减少信息丢失的同时，更全面地保留关键特征，对于细致判别颈动脉狭窄具有重要意义。此外，并行使用最大池化和平均池化可以从不同角度整合特征，增加模型对输入变化的适应性和鲁棒性。在颈动脉狭窄分类中，这种设计可以确保模型在不同患者、不同成像条件下都能有效工作，提高模型的泛化能力。

具体地说，我们对骨干网络的最后两层做进一步处理，首先将经过骨干网络第三层处理后的特征分别放入扩张融合模块和扩张融合模块，经过两个模块处理后的特征经过一个  $1 \times 1$  卷积来修改通道数。然后将骨干网络输出的特征经过一个  $1 \times 1$  卷积处理后，将其和经过扩张融合模块分支处理后特征进行矩阵乘法，再放入 softmax 输出一个权重矩阵。最后，将权重矩阵和经过扩张融合模块分支处理后的特征进行矩阵乘法操作，得到初步增强特征。通过将非局部网络思想和多尺度特征融合方法相结合，显著地提高了模型对颈动脉狭窄分类的性能。

### 3.3 多层通道注意力模块 (MLCA)

研究发现，仅仅利用了非局部网络和多尺度特征融合的方法并不能保证模型分类性能达到最好效果，于是本文提出了多层通道注意力模块，单个的通道注意力模块结构如图 4 所示。多层通道注意力模块通过对骨干网络每一层的输出进行通道注意力处理，有效地增强了每一层特征的表达能力。通道注意力机制能自动学习并强调对分类任务最重要的通道，从而提升模型对关键特征如血管形态、狭窄区域的大小和位置等的敏感性。这种逐层加强的方法确保了模型可以从多尺度的特征中提取出最有用的信息。将每层处理后的特征拼接在一起，可以实现不同层级的特征融合。这种融合方式允许模型利用从较低层次的细节特征到较高层次的抽象特征的全范围信息，从而获得更全面的特征表示。这对于颈动脉狭窄分类尤为重要，因为狭窄的识别需要同时考虑局部的细节和整体的上下文。通过多层的通道注意力处理，模型在各个层次上都进行了优化和增强，这有助于提高对图像质量波动、成像技术差异等外部变量的适应性。这种鲁棒性是医学影像分析中的一个关键需求，尤其是在多中心临床研究和实际应用中。

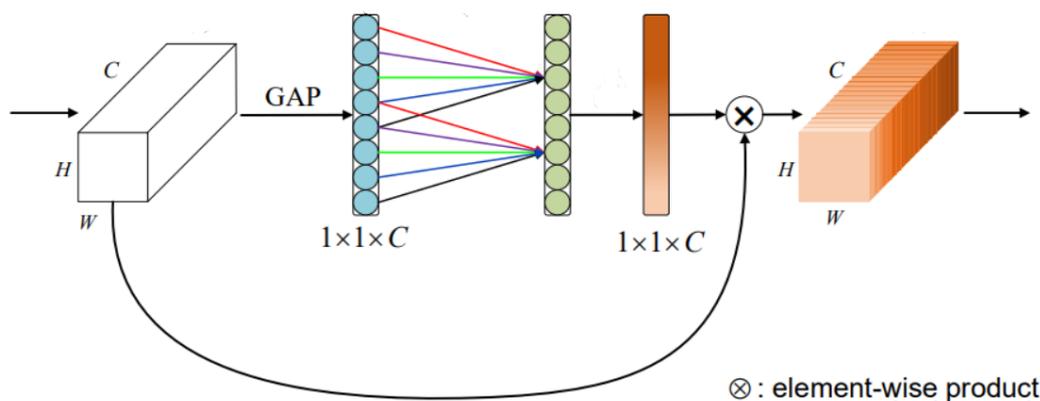


图4 通道注意力模块结构图.

Fig.4 Channel Attention Module Architecture Diagram.

具体的说，将骨干网络的每一层输出特征都放入一个通道注意力模块进行处理，每一个通道注意力模块通过全局平均池化简化输入特征，使用一维卷积捕捉通道间依赖关系，并通过非常小的卷积核来减少参数量。其核心在于使用 sigmoid 激活函数生成的通道权重来调整特征图的通道响应，从而重标定特征，强调重要的通道并抑制不重要的通道。然后，将每层经过处理后的特征和初步增强特征拼接起来，通过  $1 \times 1$  卷积调整通道数，最终输出增强特征。

在颈动脉狭窄分类中，NLCANet 模型通过结合非局部多尺度特征融合模块和多层通道注意力模块，显著提升了特征捕获和处理能力。非局部模块有效整合长距离信息，确保捕捉图像的细节与全局特征，而多层通道注意模块优化通道级特征，提高特征的区分力和诊断准确性。这种结合不仅增强了模型对复杂颈动脉狭窄医学图像的准确解读，还提升了泛化能力和操作的解释性，使得 NLCANet 在实际医疗环境中对颈动脉狭窄的分类更为有效、可靠，并易于应用。

## 4 实验评估

### 4.1 实验数据集

本文使用的数据集来源于颈动脉 DSA（数字减影血管造影）原始视频数据中的视角最佳、最清晰的关键帧。数据集中患者的性别和年龄分布较为均匀，涵盖的就诊时间范围为 2010 年至 2022 年，总计约有 2000 张颈动脉 DSA 影像。具体来说，本研究的数据集包含总计 1958 张图像，分别涵盖轻度狭窄（800 张）、中度狭窄（601 张）和重度狭窄（557 张）三类。轻度狭窄（Mild Stenosis）是指狭窄程度一般在  $<50\%$ ，即血管腔的直径减少不到一半。此时血流仍然较为通畅，通常无明显症状，部分患者可能会出现轻微头晕或短暂性视力模糊。中度狭窄（Moderate Stenosis）是指狭窄程度在  $50\% \sim 69\%$  之间，意味着血管腔缩小超过一半，但仍有一定的血流通。患者可能会出现一过性脑缺血发作（TIA），表现为短暂性肢体无力、麻木、言语不清或短暂性视力丧失。重度狭窄（Severe Stenosis）是指狭窄程度  $\geq 70\%$ ，甚至可能接近完全闭塞（99%）。此时血流严重受阻，极易引发缺血性脑卒中（中风）。患者可能会出现持续性神经功能障碍，如偏瘫、语言障碍、意识丧失等。

为了评估模型的泛化能力，我们将数据集划分为训练集、验证集和测试集。具体划分方式如下：训练集占数据集的 70%，用于训练模型，共计约 1371 张图像。验证集占数据集的 15%，用于调优模型超参数，共计约 294 张图像。测试集占数据集的 15%，用于评估模型性能，共计约 294 张图像。每一类（轻度狭窄、中度狭窄、重度狭窄）的图像在划分过程中保持了相同比例的分配，确保各类别的代表性。为了确保数据的可靠性和准确

性，我们对数据集进行了严格的数据清洗，排除了低质量的图像。在数据处理过程中，始终遵循严格的隐私保护措施，数据集中不包含任何患者的个人隐私信息，所有数据仅用于本次研究。颈动脉 DSA 影像数据集详细情况如下表 1 所示：

表 1 颈动脉 DSA 影像数据集统计表

Table 1 Carotid Artery DSA Image Dataset Statistical Table

<i>Dataset</i>	<i>Number of pictures, format</i>
Total number of pictures	1958
Mild stenosis	800
Moderate stenosis	601
Severe stenosis	557
Storage format	JPG

#### 4.2 数据增强技术

由于单处狭窄患者数量较多，单处狭窄样本的数量相较于多处狭窄和完全闭塞的样本数量也更多一些。为了增加数据集的规模并解决类别不均衡问题，采用了数据增强技术[17]生成更多样本，如图 5 所示，同时也为了防止模型过拟合。数据增强方法为有监督的数据增强，例如随机旋转，裁剪，添加噪声等。具体来说，旋转可以使模型更加全面的学习到物体的形状和结构特征；裁剪可以使模型具有对于不同尺度、位置的目标检测能力；添加噪声可以增强模型对于噪声的抵抗能力，提高其在复杂环境下的泛化性能。这些数据增强方法不仅有效扩大了数据集的规模，还增强了模型的鲁棒性和准确性。

#### 4.3 实验设置

实验使用小批量 SGD 进行训练。对于分类任务，数据集的初始学习率被设置为  $1e-2$ 。遵循之前的工作，我们采用了多步骤学习率策略，使用了 0.9 的动量和 0.0001 的权重衰减。同时，将批处理大小设置为 16，得以更高效地更新模型参数。接着，将训练轮次设定为 25。为了保持公平并取得良好的比较效果，所有的模型均在同样的实验设置下进行对比。

#### 4.4 模型消融实验

通过消融实验对 NLCANet 中的非局部多尺度特征融合模块 (NLMSF) 和多层通道注意力模块 (MLCA) 的有效性进行了全面的分析。为进行对比分析，采用 resnet18、resnet50、mobilenetv3 和 Inceptionv4 进行实验，实验结果如图 6 所示。可以看出，由于后三者的骨干网络参数量较多，网络较深，导致出现了过拟合的现象，最终选择 resnet18 作为颈动脉狭窄数据集上的消融实验骨干网络。

通过消融实验来验证 NLCANet 中每个组成部分的作用，结果如图 7 所示。结果表明，仅使用基础的 ResNet18 模型其性能相对较低，在处理复杂或细粒度特征时存在局限性。接着，通过引入多层通道注意力模块 (MLCA)，模型性能得到明显提升。MLCA 模块通过强化重要通道的特征并抑制不重要的通道，增强了模型对关键信息的捕捉能力。进一步地，添加非局部多尺度特征融合模块 (NLMSF) 使性能获得更大的提升。NLMSF 模块通过在多个尺度上捕获长距离依赖性，有效地整合了全局和局部信息，从而提升了特征的表达能力和鲁棒性。最终，当结合 MLCA 和 NLMSF 两个模块时，模型展示出最佳的性能，这表明这两个模块互补地增强了模型对复杂数据结构的理解和分类能力。

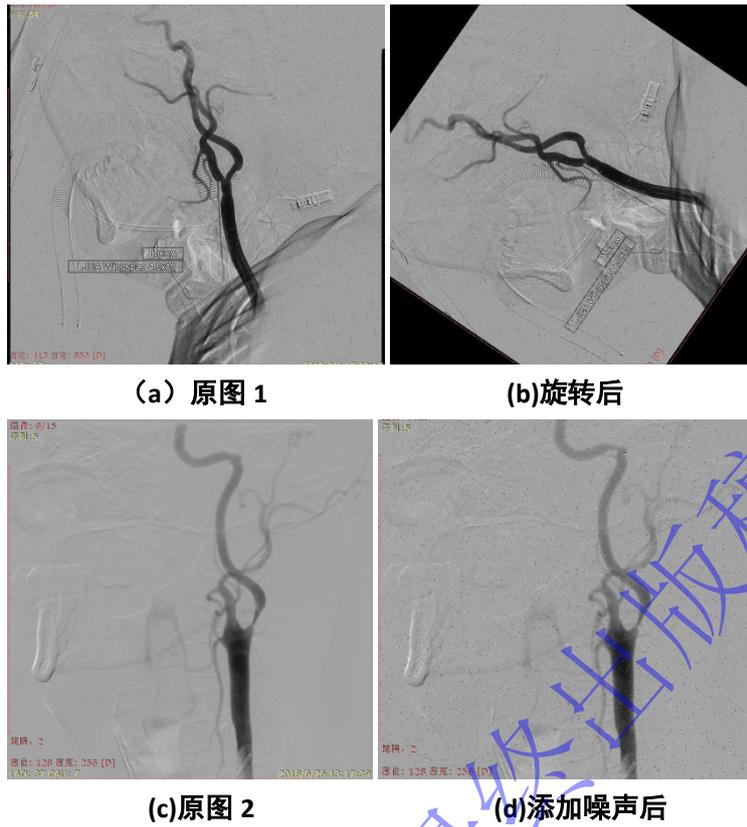


图 5 数据增强对比图. (a) 原图 1; (b) 经过旋转后的图 1; (c) 原图 2; (d) 添加噪声后的图 2

Fig.5 Data augmentation comparison chart. (a) Original image 1; (b) Rotated image 1; (c) Original image 2; (d) Noise-added image 2

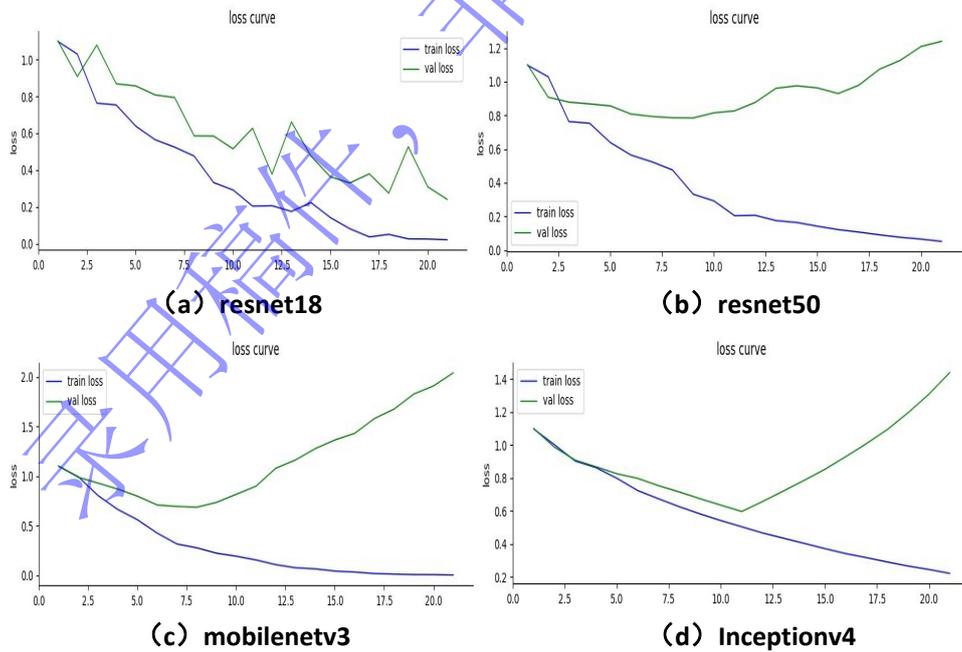


图 6 数据增强对比图. (a) 原图 1; (b) 经过旋转后的图 1; (c) 原图 2; (d) 添加噪声后的图 2

Fig.6 Data augmentation comparison chart. (a) Original image 1; (b) Rotated image 1; (c) Original image 2; (d) Noise-added image 2

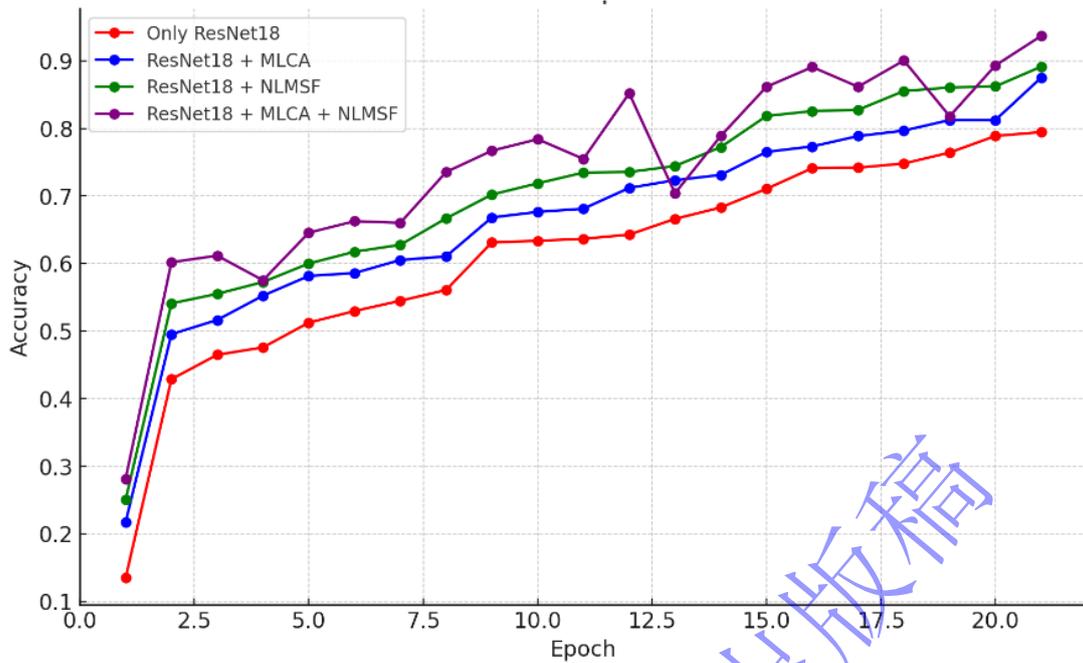


图7 消融实验结果图.

Fig.7 The results of the ablation experiment diagram.

#### 4.5 模型对比实验

将模型 NLCANet 与当前主流的几种分类模型进行了细致的比较和分析，以评估其在颈动脉狭窄分类任务上的性能。具体比较的模型包括 DFL-CNN[36]、TASN[37]、MC-LOSS[38]和 FE-Net[39]。通过一系列实验，我们证实了 NLCANet 在多个关键性能指标上均优于其他模型，对比结果如图 8 所示。我们提出的 NLCANet 在颈动脉狭窄数据集上取得优异的性能。其主要原因是非局部多尺度特征融合模块提取了局部细节判别性特征，显著特征重建阶段具有去噪和对浅层视觉信息和高级语义信息进行融合的优点，有助于网络进一步学习完整的判别性特征。在信息融合阶段，将去噪后的显著性特征与局部细节判别性特征进行融合，实现信息互补。在组合通道注意力模块进一步学习和增强互补信息夏，获得了最佳的分类性能。

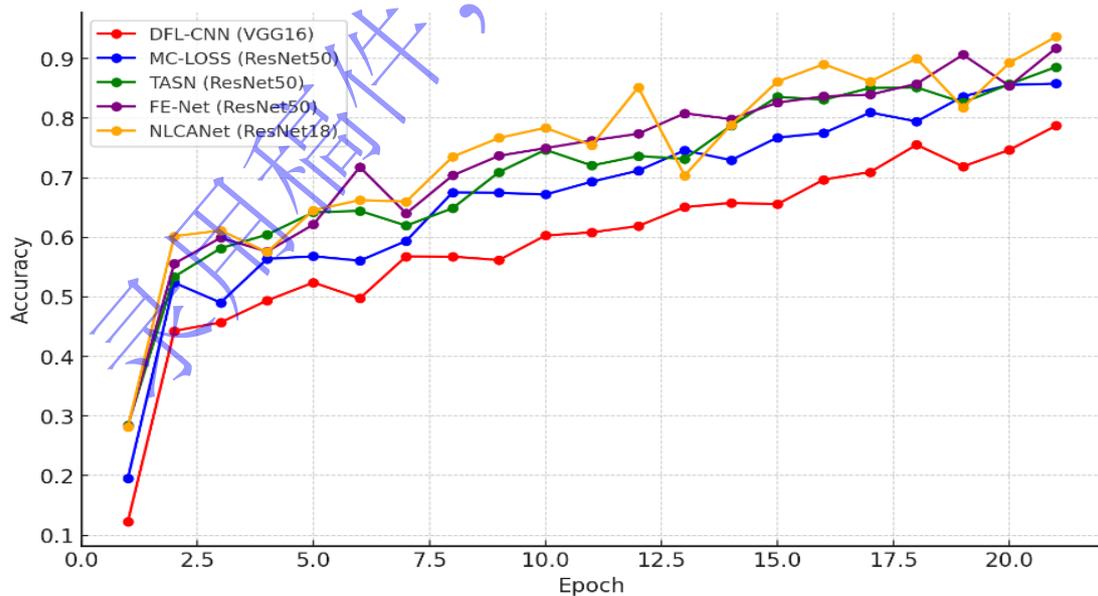


图8 对比实验结果图.

Fig.8 Comparative experimental results diagram.

在和主流模型对比中，我们的模型取得了最好的效果。但是，为了进一步验证非局部多尺度特征融合模块（NLMSF）的性能，我们分别采用不同的组合方式来进行对比实验，分别使用 FPN、PPM、用传统注意力替代非局部网络、只使用扩张融合模块、只使用组合池化模块和 U-Net 的结构进行实验，实验结果如图 9 所示。实验结果表明，具体而言，FPN 和 PPM 虽然通过层级结构增强了特征的语义丰富性，但在具体的颈动脉狭窄分类任务中，它们的性能仍然较低。传统注意力模块通过关注关键区域来改善结果，但与专门针对此类医学图像设计的模型相比，其效果有限。U-Net 表现出了较好的性能，这得益于其有效的编解码结构和跳跃连接，这在医学图像分割领域已被广泛验证。然而，仅使用组合池化模块（Only CP）和仅使用扩张融合模块（Only DF）的方法通过关注局部和全局上下文进一步提高了性能。特别是，只使用扩张融合模块的方法通过深层特征的精细化处理，展现了较高的分类能力。最终，我们的 NLMSF 模型通过融合这些技术的优势，并结合针对颈动脉狭窄特有的特征提取策略，达到了所有对比模型中最优的性能。这一成果验证了 NLMSF 模块在处理复杂和多变的医学图像数据方面的有效性和优越性。

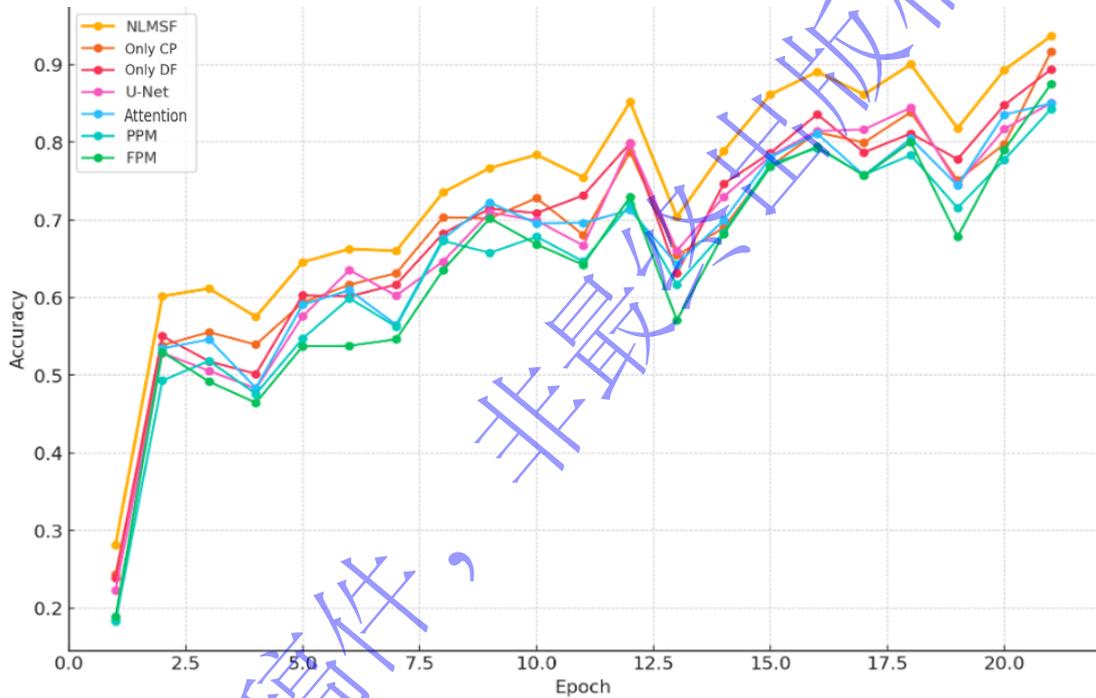


图 9 NLMSF 模块对比实验结果图。

Fig.9 NLMSF module comparative experimental results diagram.

在我们的研究中，我们专注于通过多级联通道注意力（MLCA）模块来提升颈动脉狭窄类型分类的准确性。为了验证 MLCA 模块的效能，我们将其与几种常用的注意力机制进行了对比实验，包括 SENet、CBAM 和新型注意力机制 NAM，以及传统的单级特征提取策略。实验结果如表 2 所示，显示了各种结构在颈动脉狭窄分类任务上的平均准确率。

从实验结果中可以看出，SENet 的平均准确率为 0.841，而 CBAM 稍好，为 0.852。这表明在通道注意力方面的改进对于性能的提升具有一定的影响。NAM 和单级特征提取策略的表现更为出色，分别达到了 0.916 和 0.912 的平均准确率，说明更为复杂或定制的注意力机制能进一步提升分类性能。最引人注目的是，我们提出的 MLCA 模块实现了最高的平均准确率 0.937，显著优于其他比较模型。MLCA 模块的设计通过整合多个层级的注意力机制，允许模型不仅关注单一层级的特征表达，而是全面考虑多尺度的特征融合。这种方法显著增强了模型对于不同类型颈动脉狭窄的识别能力，尤其是在处理复杂和细微的图像特征时。通过这些对比实验，可以清楚地看到 MLCA 模块在提升颈动脉狭窄分类任务中的有

效性，这为未来的医学图像分析工作提供了有价值的参考，并展示了多级联通道注意力机制在提高深度学习模型性能方面的巨大潜力。

表 2 MLCA 模块对比实验结果表

Table 2 MLCA module comparative experimental results table.

Structure	Mean-accuracy
SENet	0.841
CBAM	0.852
NAM	0.916
Single-Level	0.912
MLCA	0.937

## 5 总结

本文提出的非局部通道注意力网络（NLCA<sub>Net</sub>），旨在提高颈动脉狭窄的分类准确性，并为医学影像分析领域带来创新的视角。其主要的两大模块分别为非局部多尺度特征融合模块和多层通道注意模块。NLMSF 模块通过结合不同层级的特征并利用非局部网络原理，有效地融合了多尺度信息，这不仅增强了模型对颈动脉狭窄图像中局部细节的捕捉能力，而且提高了关键细节的识别精度。LCA 模块通过全局平均池化和一维卷积技术有效捕获通道间的依赖关系，实现了通道特征的重标定，从而提升了模型对图像语义信息的处理效率和分类性能。此外，该模块通过增强网络每层的特征表达能力，为整体模型的性能提供了显著提升。与传统的 DSA 影像分析方法相比，NLCA<sub>Net</sub> 提供了一种更为高效和准确的解决方案。这对于医疗实践尤其重要，可以显著减少病理学家的工作负担，减少人为错误，加快诊断速度，从而在保证诊断质量的同时提高患者的治疗效果。我们相信，NLCA<sub>Net</sub> 及其核心模块的进一步研究和优化，将为未来医学影像分析领域的发展开辟新的道路。

## 致谢

本论文由科技部科技创新 2030 重大项目（2020AAA0108703）资助。

## 文献

- [1] Bighetti S, Bettolini L, Rovaris S, et al. Cutaneous angiosarcoma: Harnessing LC-OCT for early identification and optimal patient outcomes.[J]. Journal of the European Academy of Dermatology and Venereology : JEADV,2024.
- [2] Chunyang L, Zhigang L, Weikang L. TDCAU-Net: retinal vessel segmentation using transformer dilated convolutional attention-based U-Net method.[J]. Physics in medicine and biology,2023,69(1).
- [3] Wang, H., Liu, J., Li, B., et al. (2024). Lightweight fundus image segmentation network integrating structured convolution and dual attention mechanisms. Journal of Computer-Aided Design & Computer Graphics, 1-16. Retrieved from <http://kns.cnki.net/kcms/detail/11.2925.TP.20240521.1608.008.html>. (汪华登,刘金,黎兵兵,等.融合结构化卷积和双重注意力机制的轻量级眼底图像分割网络[J/OL].计算机辅助设计与图形学学报:1-16[2024-06-03].<http://kns.cnki.net/kcms/detail/11.2925.TP.20240521.1608.008.html>.)
- [4] Lecun, Y., et al. "Gradient-Based Learning Applied to Document Recognition." Proceedings of the IEEE, Jan.

- 1998, pp. 2278–324, <https://doi.org/10.1109/5.726791>.
- [5]Krizhevsky, Alex, et al. “ImageNet Classification with Deep Convolutional Neural Networks.” *Communications of the ACM*, May 2017, pp. 84–90, <https://doi.org/10.1145/3065386>.
- [6]Simonyan, Karen, and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” *International Conference on Learning Representations, International Conference on Learning Representations*, Jan. 2015.
- [7]Sivapriya G., et al. "Automated diagnostic classification of diabetic retinopathy with microvascular structure of fundus images using deep learning method." 88.PA(2024).
- [8]Nader Rafic, Bourcier Romain, and Autrusseau Florent. "Using deep learning for an automatic detection and classification of the vascular bifurcations along the Circle of Willis." 89.(2023):102919-102919.
- [9]K. Susheel Kumar, Nagendra Pratap Singh, and Shekhar Yadav. "Screening Retinal Images and Extraction of the Retinal Blood Vessel for Identifying Diseases and Classification of Arteries and Veins by Using Deep Learning." 16.8(2023):790-804.
- [10]Appaji Abhishek, et al. "Deep learning model using retinal vascular images for classifying schizophrenia." 241.(2022):238-243.
- [11]Uema Ryotaro, et al. "Use of a convolutional neural network for classifying microvessels of superficial esophageal squamous cell carcinomas." 36.8(2021):2239-2246.
- [12]Hacilar Hilal, et al. "Classification of Cardiovascular Artery Diseases Using Artificial Neural Network." 10.1(2020):34-41.
- [13]Setiawan Wahyudi, Utoyo Moh. Imam, and Rulaningtyas Riries. "Classification of neovascularization using convolutional neural network model." 17.1(2019):463-463.
- [14]Yin, P. (2023). Research on eyeball image segmentation technology based on deep learning (PhD dissertation). South China University of Technology. (尹鹏帅.基于深度学习的眼球图像分割技术研究.2023.华南理工大学, PhD dissertation.)
- [15]Chen, C. (2024). Research on fundus retinal artery and vein classification method based on deep learning (MA thesis). Beijing University of Civil Engineering and Architecture. (陈畴羽.基于深度学习的眼底视网膜动静脉分类方法研究.2024.北京建筑大学, MA thesis.)
- [16]Huang, S. (2024). Research on diabetic retinopathy classification method based on self-attention learning network (MA thesis). Tianjin University of Technology. (黄丝雨.基于自注意力学习网络的糖尿病视网膜病变分类方法研究.2024.天津理工大学, MA thesis.)
- [17]Woo, Sanghyun, et al. “CBAM: Convolutional Block Attention Module.” *Computer Vision – ECCV 2018, Lecture Notes in Computer Science*, 2018, pp. 3–19, [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [18]Hu, Jie, et al. “Squeeze-and-Excitation Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Aug. 2020, pp. 2011–23, <https://doi.org/10.1109/tpami.2019.2913372>.
- [19]Alexey D. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv: 2010.11929, 2020.
- [20]Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [21]Liu Z, Mao H, Wu C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 11976-11986.
- [22]Tan M, Le Q. Efficientnetv2: Smaller models and faster training[C]//International conference on machine learning. PMLR, 2021: 10096-10106.
- [23]ZHOU T, TAN T, PAN X, et al. Fully automatic deep learning trained on limited data for carotid artery segmentation from large image volumes [J]. *Quantitative Imaging in Medicine and Surgery*, 2021, 11(1): 67.

- [24]GENG J, ZHANG P, XU Y, et al. Application of deblur technology for improving the clarity of digital subtractive angiography [J]. *Interventional Neuroradiology*, 2022: 15910199221143168.
- [25]KIMURA R, TERAMOTO A, OHNO T, et al. Virtual digital subtraction angiography using multizone patch-based U-Net [J]. *Physical and Engineering Sciences in Medicine*, 2020, 43: 1305-1315.
- [26]崔楷文. 基于深度学习的颈动脉狭窄辅助诊断 [D], 2020.
- [27]KHANKARI J, YU Y, OUYANG J, et al. Automated detection of arterial landmarks and vascular occlusions in patients with acute stroke receiving digital subtraction angiography using deep learning [J]. *Journal of NeuroInterventional Surgery*, 2023, 15(6): 521-525.
- [28]Ma, W. (2024). Risk stratification of carotid atherosclerotic plaques based on ultrasound images (PhD dissertation). Huazhong University of Science and Technology. (马卫.基于超声图像的颈动脉粥样硬化斑块风险分级.2024.华中科技大学,PhD dissertation.)
- [29]Yan, R. (2022). Research on carotid plaque segmentation and analysis based on CTA images (MA thesis). Beijing Jiaotong University. (颜荣耀.基于CTA影像的颈动脉斑块分割及分析研究.2022.北京交通大学,MA thesis.)
- [30]Ilyasova N. Yu.,and Demin N. S.."Systems for Recognition and Intelligent Analysis of Biomedical Images." 33.4(2024):1142-1167.
- [31]Bao Junxiao, et al."Deep Learning Algorithm in Biomedical Engineering in Intelligent Automatic Processing and Analysis of Sports Images." 2022.(2022).
- [32]Sollee John, et al."Artificial Intelligence for Medical Image Analysis in Epilepsy." 182.prepublish(2022):106861-.
- [33]Wang, Xiaolong, et al. "Non-Local Neural Networks." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, <https://doi.org/10.1109/cvpr.2018.00813>.
- [34]Amini Amirkolae Hamed,and Amini Amirkolae Hamid."Medical image translation using an edge-guided generative adversarial network with global-to-local feature fusion.." 36.6(2022):11-14.
- [35]Ningning Lv, et al."Non-local sparse attention based swin transformer V2 for image super-resolution." 222.(2024):109542-.
- [36]Wang Y, Morariu V I, Davis L S. Learning a discriminative filter bank within a CNN for fine-grained recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4148-4157.
- [37]Zheng H, Fu J, Zha Z J, et al. Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5012-5021.
- [38]Chang D, Ding Y, Xie J, et al. The devil is in the channels: Mutual-channel loss for fine-grained image classification[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4683-4695.
- [39]Chen W, Xie D, Zhang Y, et al. All you need is a few shifts: Designing efficient convolutional neural networks for image classification[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 7241-7250.