Computer Science and Application 计算机科学与应用, 2026, 16(5), 33-40
Published Online May 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.165161

Hans汉斯

基于注意力融合的OCT视网膜病变分类方法

郝永栋

温州大学计算机与人工智能学院，浙江 温州

收稿日期：2026年4月1日；录用日期：2026年5月2日；发布日期：2026年5月12日

摘　要

光学相干断层扫描(Optical Coherence Tomography, OCT)是一种非侵入、高分辨率的医学成像技术，能够清晰呈现视网膜层状结构及病变特征，在眼科疾病辅助诊断中具有重要应用价值。针对OCT图像分类任务，现有方法多采用单流卷积神经网络对原始图像进行建模，虽具备一定性能，但在全局结构信息与局部细节特征的协同表征方面仍存在不足，难以充分捕捉病灶边缘、纹理异常及层间结构扰动等关键判别信息。为此，文章提出一种基于双流与双向注意力融合的OCT视网膜病变分类方法。首先构建双流输入：一支为原始OCT图像，用于保留整体结构信息；另一支为辅助表征图像，用于突出边缘与局部异常特征，从而形成互补表示。其次，通过双分支特征提取网络分别对两路输入进行编码，并引入双向注意力融合模块，实现跨分支信息交互，使不同模态特征相互引导，增强对关键病灶区域的表达能力。最后，将融合特征输入分类头完成病变类别预测。在OCT2017数据集上的实验结果表明，所提方法在准确率、召回率及F1值等指标上均优于多种对比方法，证实了双流结构与双向注意力机制的有效性。该方法为OCT图像病变分类提供了一种有效的多特征融合建模思路。

关键词

光学相干断层扫描图像，深度学习，卷积神经网络

Attention Fusion-Based Method for OCT Retinal Disease Classification

Yongdong Hao

College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou Zhejiang

Received: April 1, 2026; accepted: May 2, 2026; published: May 12, 2026

Abstract

Optical Coherence Tomography (OCT) is a non-invasive, high-resolution medical imaging technique that can clearly visualize retinal layer structures and pathological features, and has significant value

文章引用: 郝永栋. 基于注意力融合的 OCT 视网膜病变分类方法[J]. 计算机科学与应用, 2026, 16(5): 33-40.
DOI: 10.12677/csa.2026.165161

in assisting the diagnosis of ophthalmic diseases. For OCT image classification tasks, existing methods mostly adopt single-stream convolutional neural networks to model raw images. Although they achieve certain performance, they still have limitations in jointly representing global structural information and local detailed features, making it difficult to fully capture key discriminative cues such as lesion boundaries, texture abnormalities, and inter-layer structural disruptions. To address this issue, this paper proposes an OCT retinal disease classification method based on dual-stream input and bidirectional attention fusion. First, a dual-stream input is constructed: one stream consists of raw OCT images to preserve global structural information, while the other stream consists of auxiliary representation images to emphasize edges and local abnormal features, forming complementary representations. Second, a dual-branch feature extraction network is employed to encode the two inputs separately, and a bidirectional attention fusion module is introduced to enable cross-branch information interaction, allowing features from different modalities to guide each other and enhance the representation of key lesion regions. Finally, the fused features are fed into a classification head to predict disease categories. Experimental results on the OCT2017 dataset demonstrate that the proposed method outperforms multiple baseline approaches in terms of accuracy, recall, and F1-Score, confirming the effectiveness of the dual-stream structure and the bidirectional attention mechanism. This method provides an effective multi-feature fusion modeling approach for OCT image-based disease classification.

Keywords

Optical Coherence Tomography Image, Deep Learning, Convolutional Neural Network

1. 引言

视网膜疾病是导致视力损伤甚至失明的重要原因之一，常见病变包括脉络膜新生血管(CNV)、糖尿病性黄斑水肿(DME)以及玻璃膜疣(Drusen)等。上述疾病在早期阶段往往已经伴随视网膜层状结构改变、液体积聚、局部隆起或组织紊乱等病理特征，因此，实现对相关病变的及时识别与准确分类，对于提高临床诊断效率和辅助制定治疗方案具有重要意义[1]。

光学相干断层扫描(Optical Coherence Tomography, OCT) [2]是一种非侵入、高分辨率的层析成像技术，能够对视网膜组织进行高精度横截面成像，并清晰反映不同层结构及其异常变化。与普通眼底图像相比，OCT 能够更直观地呈现病灶在纵向层次上的位置、形态及结构扰动，已成为眼科疾病辅助诊断的重要工具。随着深度学习技术在医学图像分析领域的迅速发展，基于 OCT 图像的自动分类方法逐渐成为研究热点[3]。

然而，现有 OCT 分类方法仍面临一些问题。多数方法仅将原始图像作为单一输入，通过单分支网络完成特征提取与分类，这类方法虽然能够学习整体语义表示，但在病灶边缘、局部纹理异常以及层结构细微变化等方面的建模能力仍有限[4]。事实上，OCT 图像中的病理信息不仅体现在全局结构分布上，也体现在局部异常细节中。如果仅依赖单一表征形式，容易导致对某些细粒度病灶特征利用不足，从而限制分类性能的进一步提升[5]。

基于此，本文围绕 OCT 视网膜病变分类任务，研究双流输入与特征交互融合问题，尝试从不同表征角度对图像信息进行联合建模。通过引入双流网络结构与双向注意力融合机制，使模型能够同时关注整体层状结构和局部细节异常，并在跨流交互中实现互补信息增强，从而提升病变分类效果。该研究对于

提高 OCT 图像智能分析能力、推动眼科疾病辅助诊断系统的发展具有一定的理论意义和应用价值。

2. 实验方法

2.1. 网络总体框架

针对 OCT 病变分类任务中单一输入表征不足、双流特征融合不充分的问题，本文提出一种基于双流双向注意力与门控融合的 OCT 病变分类网络(见图 1、图 2)。该网络整体由双流输入模块、双流特征提取模块、双向注意力交互模块、门控自适应融合模块以及分类头五部分组成。

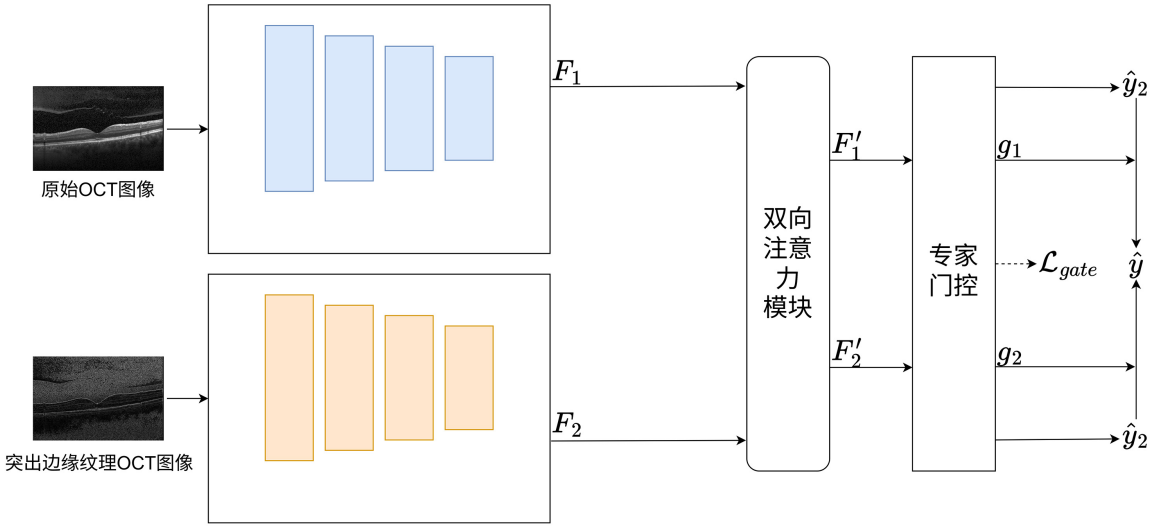

Figure 1. Method framework diagram
图 1. 方法框架图

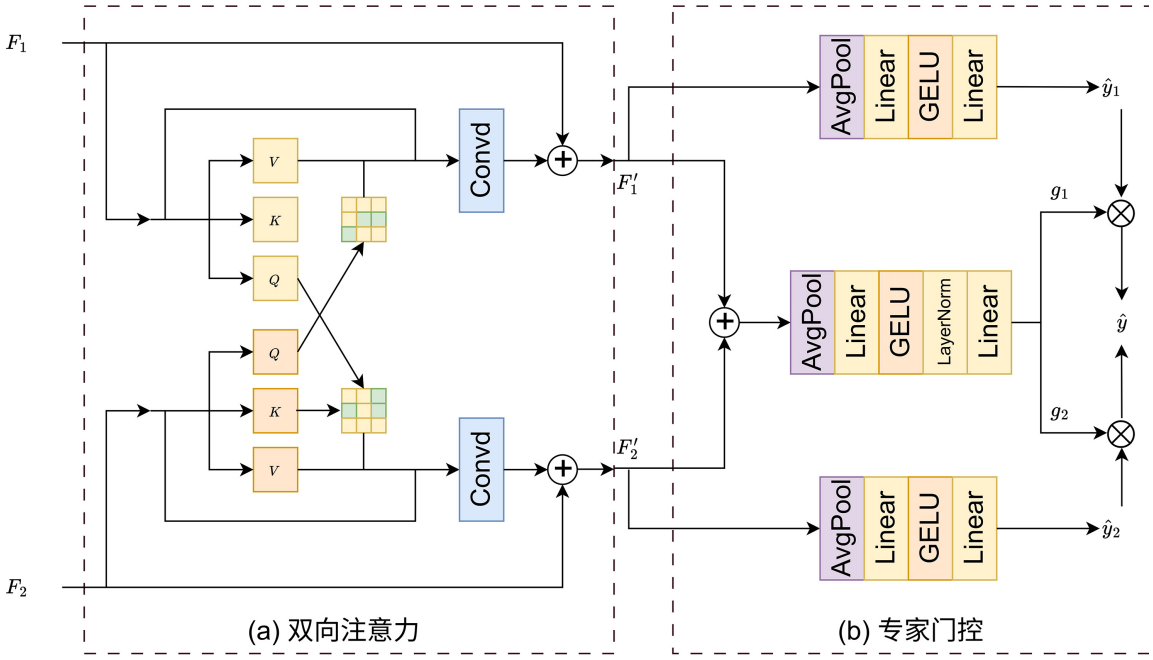

Figure 2. Schematic diagram of bidirectional attention and gated fusion structure
图 2. 双向注意力与门控融合结构示意图

对于输入的 OCT 图像，首先构建两路具有互补性的输入表征。其中，一路输入原始 OCT 图像，用于保留视网膜整体层状结构、灰度分布及病灶宏观形态信息；另一路输入辅助表征图像，用于突出边缘、纹理或局部异常区域等细粒度特征。随后，双流输入分别进入两个特征提取分支，获得两组高层特征表示。

在特征融合阶段，本文首先利用双向注意力交互模块建立双流特征之间的跨分支关联，使一条分支能够感知并吸收另一条分支中的关键信息，从而实现特征层面的相互引导。进一步地，为了避免简单融合带来的信息冗余和无效叠加，本文引入门控自适应融合机制，根据两路特征的内容自适应生成融合权重，对双流特征进行动态调节与选择性保留。最终，将融合后的特征输入分类头，完成病变类别预测。

2.2. 双流特征提取

为了分别建模原始图像与辅助表征图像中的特征信息，本文采用双分支特征提取结构。对于输入 X_1 和 X_2，分别通过两个结构相同但参数独立的特征提取分支，得到两组高维特征表示：

$$F_1 = E_1(X_1), F_2 = E_2(X_2)$$

其中，$E_1(\cdot)$ 和 $E_2(\cdot)$ 表示两个特征提取分支，F_1 和 F_2 分别表示原始图像流和辅助表征流的特征图。

2.3. 双向注意力融合模块

为了克服传统特征拼接方式在跨流交互建模方面的不足，本文设计双向注意力融合模块[6]，对双流特征进行深层交互。给定双流特征 F_1 和 F_2，首先以 F_1 作为查询特征，对 F_2 中的关键信息进行选择性聚合，从而得到由辅助表征流增强后的特征表示；随后，再以 F_2 作为查询特征，对 F_1 中的有效信息进行反向引导与补充。通过这种双向建模方式，两路特征不再是静态拼接关系，而是形成动态交互和相互增强关系。双向交互后的特征可表示为：

$$\widetilde{F_1} = A(F_1, F_2) + F_1$$

$$\widetilde{F_2} = A(F_2, F_1) + F_2$$

其中，$A(\cdot)$ 表示跨流注意力操作，残差连接用于保留原始特征并提升训练稳定性。

2.4. 门控自适应融合模块

经过跨模态注意力增强后，两类特征之间已经进行了较充分的信息交互。不过，这两类特征对最终预测的贡献并不完全一致。为此，本文进一步引入门控自适应融合模块[7]，对双流交互后的特征进行动态加权控制。其预测结果可表示为

$$\hat{y}_1 = h_1(\tilde{F}_1)$$

$$\hat{y}_2 = h_2(\tilde{F}_2)$$

其中 $h_1(\cdot)$ 和 $h_2(\cdot)$ 表示对应的分类头，由全局平均池化和多层感知机构成。

门控网络根据两种模态的全局统计特征生成自适应融合权重。首先计算两种模态的全局特征表示：

$$u_1 = \mathrm{GAP}(\tilde{F}_1),$$

$$u_2 = \mathrm{GAP}(\tilde{F}_2),$$

随后，将两种模态特征进行拼接并输入门控网络：

$$g = \sigma\left(\phi\left(\left[u_1, u_2\right]\right)\right),$$

其中$[\cdot]$表示特征拼接操作，$\phi(\cdot)$表示门控多层感知机，$\sigma(\cdot)$表示 Sigmoid 函数，$g \in (0,1)$表示幅值模态的融合权重。

最终预测结果通过加权融合得到：

$$\hat{y} = g \cdot \hat{y}_1 + (1-g) \cdot \hat{y}_2.$$

3. 实验配置与结果分析

3.1. 数据集与预处理

本文在 OCT2017 数据集[1]上对所提方法进行实验验证。该数据集包含多种常见视网膜病变类别，主要包括糖尿病性黄斑水肿(DME)以及正常样本(Normal)。不同类别样本具有较为明显的层状结构差异和病灶表现特征，见表 1。

Table 1. OCT2017 dataset description
表 1. OCT2017 数据集描述

	英文缩写	训练集	测试集	总数
糖尿病性黄斑水肿	DME	11,348	242	11,590
正常眼底	Normal	12,696	242	11,938

在数据预处理阶段，首先将输入图像统一缩放到固定尺寸，以满足网络输入要求。随后，对训练集采用随机翻转、随机旋转等数据增强方式，以提升模型的泛化能力。对于双流输入，第一路采用原始 OCT 图像，第二路采用由原始图像构造的辅助表征图像，用于突出边缘、纹理细节。为避免数据泄漏，训练集、验证集与测试集严格独立划分，结果如图 3 所示。

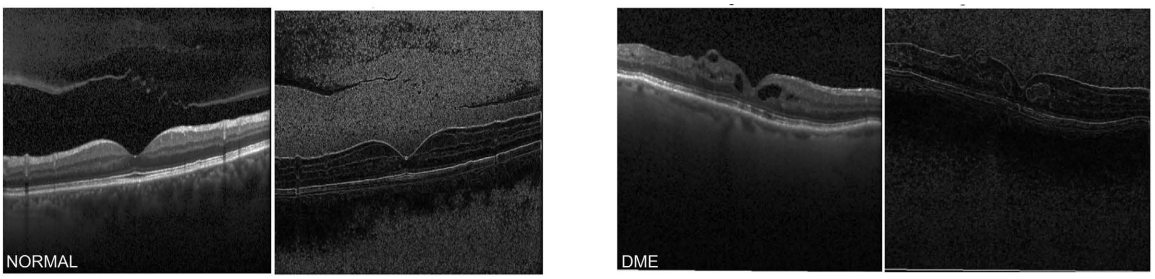

Figure 3. Samples of two different retinal conditions in the OCT2017 dataset and their edge-enhanced versions
图 3. OCT2017 数据集中两种不同视网膜状况的样本展示及边缘增强后样本展示

3.2. 实验设置

本文实验基于 PyTorch 深度学习框架实现，在 GPU 环境下完成训练与测试。优化器采用 AdamW，初始学习率设置为1×10^{-4}，批大小设置为 32，训练轮数设置为 100。损失函数采用交叉熵损失函数。为了全面评估模型性能。本文采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)以及 F1 值作为评价指标。

3.3. 损失函数

本文采用交叉熵损失函数作为分类目标函数，其表达式为：

$$L_{cls} = -\sum_{i=1}^{C} y_i \log\left(\hat{y}_i\right)$$

其中，C 表示类别数，y_i 表示真实标签的 one-hot 编码，$\hat{y}_i$ 表示模型对第 i 类的预测概率。通过最小化该损失函数，促使模型学习更具区分性的特征表示，从而提高病变分类精度。

4. 实验结果及分析

4.1. 模型性能评估

本文对测试集结果进行了详细的分类报告分析，如表 2 所示，给出了各类别在 Precision、Recall 和 F1-Score 等指标上的表现。

从整体结果来看，模型在测试集上取得了 99.38%的分类准确率，宏平均(macro avg)和加权平均(weighted avg)的 Precision、Recall 和 F1-Score 均达到 0.9938 以上，表明模型在不同类别之间具有较好的整体性能和稳定性。

从具体类别来看，对于 DME 类别，模型的 Recall 达到 1.0000，说明所有 DME 样本均被正确识别，未出现漏检情况；其 Precision 为 0.9878，表明存在少量其他类别样本被误判为 DME。对应的 F1-Score 为 0.9938，说明模型在该类别上取得了较好的精确率与召回率平衡。

对于 NORMAL (正常)类别，模型的 Precision 达到 1.0000，说明所有被预测为 NORMAL 的样本均为真实正常样本，即不存在误检情况；其 Recall 为 0.9876，表明存在极少数正常样本被误分类为病变类别。对应的 F1-Score 同样为 0.9938，表现出较高的分类稳定性。

此外，宏平均指标与加权平均指标之间差异极小，说明模型在不同类别上的表现较为均衡，没有明显的类别偏置问题。

Table 2. Classification performance evaluation results on the OCT2017 test set
表 2. OCT2017 测试集分类性能评估结果

	Precision	Recall	F1-Score	样本量
DME	98.78%	100%	99.38%	242
NORMAL	100%	98.76%	99.38%	242
accuracy			99.38%	484
macro avg	99.39%	99.38%	99.38%	484
weighted avg	99.39%	99.38%	99.38%	484

4.2. 模型性能对比分析

为验证本文所提方法的有效性，选取 ResNet18、DenseNet121 和 Vision Transformer (ViT)作为对比模型，在相同实验设置下对 OCT 病变分类任务进行测试，实验结果如表 3 和图 4 所示。

从表 3 中可以看出，本文方法在 Accuracy、Precision、Recall 和 F1-Score 四项指标上均取得了最优结果。其中，本文方法的 Accuracy 达到 99.38%，相比 ResNet18、DenseNet121 和 ViT 分别提升了 0.62、1.86 和 4.34 个百分点；Precision、Recall 和 F1-Score 也分别达到 99.39%、99.38%和 99.38%。这说明本文方法在整体分类性能、预测精度以及类别识别均衡性方面均优于对比模型，表现出更强的综合判别能力。

结合图 4 进一步分析可以发现，不同模型在 DME 与 NORMAL 两类样本上的误分类情况存在差异。ResNet18 的混淆矩阵中，238 个 DME 被正确识别，4 个 DME 被误分类为 NORMAL；240 个 NORMAL

被正确识别，2 个 NORMAL 被误分类为 DME。DenseNet121 的误分类数量进一步增加，其中 9 个 DME 被误判为 NORMAL，3 个 NORMAL 被误判为 DME。ViT 的误分类最为明显，22 个 DME 被误分类为 NORMAL，说明其在病变样本识别上存在较明显的漏检问题。相比之下，本文方法的混淆矩阵结果最优，242 个 DME 样本全部被正确识别，仅有 3 个 NORMAL 样本被误分类为 DME，表明本文方法在病变识别方面实现了零漏检，同时保持了较低的误检率。

Table 3. Comparison of classification performance between the proposed method and baseline models
表 3. 本文方法与对比模型的分类性能比较

方法	Accuracy	Precision	Recall	F1-Score
Resnet18 [8]	98.76%	98.76%	98.76%	98.76%
Densnet121 [9]	97.52%	97.55%	97.52%	97.52%
Vit [10]	95.04%	95.35%	95.04%	95.03
Ours	99.38%	99.39%	99.38%	99.38%

综合表 3 与图 4 可以说明，本文方法不仅在整体性能上优于对比模型，而且在病变样本识别、误分类控制以及类别均衡性方面均表现出更好的稳定性。

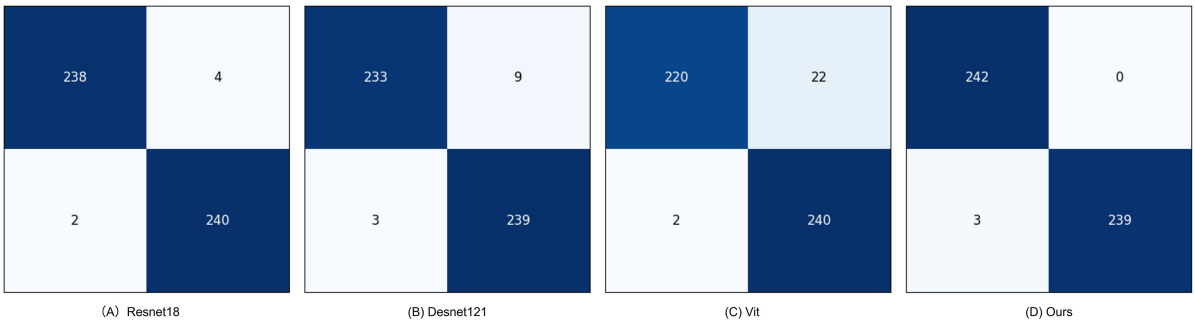

Figure 4. Confusion matrices of the proposed method and baseline models
图 4. 本文方法与对比模型的混淆矩阵

5. 总结与展望

本文围绕 OCT 图像病变分类任务，针对单流模型在多尺度信息建模与特征融合方面的不足，提出了一种基于双流双向注意力与门控融合的分类方法。该方法通过构建双流输入，从原始图像与辅助表征两个角度提取互补特征；引入双向注意力机制增强跨流特征交互能力；并通过门控融合策略实现特征的自适应选择与加权融合。在 OCT2017 数据集上的实验结果表明，本文方法在 Accuracy、Precision、Recall 和 F1-Score 等指标上均优于对比模型，同时在混淆矩阵和分类报告中表现出更低的误分类率和更好的类别均衡性，验证了所提方法的有效性与稳定性。

尽管本文方法取得了较好的实验效果，但仍存在改进空间：首先，当前实验主要基于单一公开数据集，后续可在更多数据集或实际临床数据上验证模型的泛化能力；其次，本文方法主要针对二维图像进行建模，未来可扩展至三维 OCT 数据或时序数据分析，以进一步挖掘结构与动态信息；此外，还可以结合病灶定位或分割任务，提高模型的可解释性与临床应用价值。

参考文献

[1] Kermany, D.S., Goldbaum, M., Cai, W.J., *et al*. (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-

Based Deep Learning. *Cell*, **172**, 1122-1131.e9.

[2] Bouma, B.E., de Boer, J.F., Huang, D., *et al*. (2022) Optical Coherence Tomography. *Nature Reviews Methods Primers*, **2**, 79.

[3] Yanagihara, R.T., Lee, C.S., Ting, D.S.W. and Lee, A.Y. (2020) Methodological Challenges of Deep Learning in Optical Coherence Tomography for Retinal Diseases: A Review. *Translational Vision Science & Technology*, **9**, 11. https://doi.org/10.1167/tvst.9.2.11

[4] Pang, S., Zou, B., Xiao, X., Peng, Q., Yan, J., Zhang, W., *et al*. (2024) A Novel Approach for Automatic Classification of Macular Degeneration OCT Images. *Scientific Reports*, **14**, Article No. 19285. https://doi.org/10.1038/s41598-024-70175-2

[5] Pan, H., Miao, J., Yu, J., Dong, J., Zhang, M., Wang, X., *et al*. (2025) A Lightweight Model for the Retinal Disease Classification Using Optical Coherence Tomography. *Biomedical Signal Processing and Control*, **101**, Article 107146. https://doi.org/10.1016/j.bspc.2024.107146

[6] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: *Lecture Notes in Computer Science*, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[7] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/cvpr.2018.00745

[8] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[9] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2261-2269. https://doi.org/10.1109/cvpr.2017.243

[10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., *et al*. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929.