E-Commerce Letters 电子商务评论, 2026, 15(5), 84-91
Published Online May 2026 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2026.155491

Hans汉斯

多尺度特征融合CNN的高效电商商品分类研究

张泰恒

贵州大学大数据与信息工程学院，贵州 贵阳

收稿日期：2026年3月6日；录用日期：2026年3月18日；发布日期：2026年5月14日

摘 要

针对电商图像商品尺度跨度大、背景干扰多及分类实时性要求高等挑战，本文提出一种多尺度特征融合的高效卷积神经网络模型MSFF-EComNet。首先，设计多尺度并行卷积架构(MSFF)，利用多分支感受野同步捕获商品的局部细节与全局语义特征；其次，引入改进的注意力协同机制(IASM)，通过空间与通道维度的联合增强，引导模型精准聚焦商品主体并有效抑制背景噪声。实验结果表明，在DeepFashion数据集上，该模型的精确率达到89.56%。相较于经典ResNet-50，本文模型准确率提升4.44%，参数量压缩54%。该研究为大规模电商场景下的图像识别提供了一种兼顾精度与效率的平衡方案。

关键词

电子商务，卷积神经网络，多尺度融合，注意力机制，高效分类

Research on Efficient E-Commerce Commodity Classification Using Multi-Scale Feature Fusion CNN

Taiheng Zhang

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: March 6, 2026; accepted: March 18, 2026; published: May 14, 2026

Abstract

To address the challenges of large-scale variations in e-commerce product images, numerous background interferences, and high real-time classification requirements, this paper proposes an efficient convolutional neural network model, MSFF-EComNet, featuring multi-scale feature fusion. First, a multi-scale parallel convolution architecture (MSFF) is designed to simultaneously capture local details and global semantic features through multi-branch receptive fields. Second, an improved

文章引用: 张泰恒. 多尺度特征融合 CNN 的高效电商商品分类研究[J]. 电子商务评论, 2026, 15(5): 84-91.
DOI: 10.12677/ecl.2026.155491

attention synergy mechanism (IASM) is introduced to enhance spatial and channel dimensions, guiding the model to precisely focus on the main object while effectively suppressing background noise. Experimental results demonstrate that the model achieves an accuracy of 89.56% on the DeepFashion dataset. Compared to the classic ResNet-50, our model achieves a 4.44% accuracy improvement while reducing parameters by 54%. This research provides a balanced solution that prioritizes both accuracy and efficiency for image recognition in large-scale e-commerce scenarios.

Keywords

E-Commerce, Convolutional Neural Networks, Multi-Scale Fusion, Attention Mechanism, Efficient Classification

1. 引言

1.1. 研究背景与意义

在数字经济浪潮的推动下，电子商务已跃升为全球零售业的核心引擎。根据行业报告，全球电商平台的视觉数据(图像与视频)正以指数级速度增长。在这一背景下，商品图像不仅是消费者获取信息的第一媒介，更是平台进行自动化管理的基础数据。如何从海量、杂乱的图像中精准、高效地识别并分类商品，直接关系到搜索引擎的召回率、推荐系统的精准度以及仓储物流的自动化水平[1]。

早期的商品分类主要依赖于人工标注的文本标签，但这种方式在处理具有复杂视觉属性(如面料纹理、精细剪裁)的商品时，显得力不从心。随着计算机视觉技术的演进，基于卷积神经网络(CNN)的图像识别技术展现出了卓越的特征提取能力[2]。然而，电商场景下的图像具有其独特性：拍摄光照不均、背景高度复杂、商品尺度跨度剧烈以及类间极高的视觉相似性。因此，研发一种能够兼顾局部细节与全局语义的多尺度特征融合模型，对于提升电商平台的智能化水平具有重要的科学价值。

1.2. 国内外研究现状

卷积神经网络自 AlexNet 夺冠以来，已成为图像分类领域的主流工具[3]。随后，VGGNet 加深了网络深度，ResNet 通过残差连接解决了深层网络的退化问题，为复杂图像处理奠定了基础[4]。在电商领域，国内外学者针对商品识别开展了大量研究。例如，利用 DeepFashion 等大规模数据集，研究者们通过改进的分类网络实现了对服装属性的细粒度识别[5]。

近年来，多尺度特征融合(Multi-scale Feature Fusion)成为提升分类鲁棒性的关键。现有研究多采用特征金字塔(FPN) [6]或不同步长的卷积堆叠来获取不同感受野的特征。同时，注意力机制(Attention Mechanism)的引入使得模型能够自适应地关注图像的关键区域，如 SE-Net 通过通道注意力提升了特征表达能力[7]。然而，在平衡识别精度与计算成本方面，针对移动端电商应用的高效多尺度融合模型依然是当前计算机视觉与电子商务交叉领域的研究热点[8]。

1.3. 科学问题与挑战

尽管深度学习在通用图像分类上取得了巨大成功，但在面对电商实际应用时，仍面临以下科学挑战：1) 特征表达的尺度不一致性：电商商品在图像中的占比不一，且关键特征(如纽扣、标识)往往隐藏在极

小的像素区域，传统的单一尺度卷积核难以同时兼顾全局结构与微观细节[9]。2) 复杂背景的干扰：电商图像常包含直播间装饰、模特背景或室外实拍场景，杂乱的背景噪声极易导致特征提取偏移。3) 细粒度特征的区分难题：不同品牌或型号的商品可能在视觉上仅有微小差异，如何强化模型对关键判别区域的感知能力是本文的核心任务。

1.4. 本文研究内容与创新点

针对上述挑战，本文提出了一种基于多尺度特征融合的高效 CNN 商品分类模型(MSFF-EComNet)。本文的主要研究工作包括：

1) 提出多尺度并行卷积架构：设计了一种多分支特征提取模块，通过不同大小的感受野并行处理图像，实现对商品多层级特征的同步捕获。2) 引入改进的注意力协同机制：结合空间与通道注意力，构建特征增强模块，引导网络在复杂的电商背景中自动聚焦于商品主体。在公开的大规模电商数据集上进行对比实验，证明了本文算法在分类准确率与运行效率上的优越性。

2. 相关工作

2.1. 卷积神经网络的发展历程

卷积神经网络的性能提升很大程度上源于网络拓扑结构的创新。在基础骨干网络方面，VGGNet 通过减小卷积核尺寸并增加网络深度，验证了小尺寸卷积核堆叠对特征提取的有效性[10]。随后，为了在移动端和嵌入式电商设备上实现高效推理，轻量化架构成为研究重点。MobileNet 系列引入了深度可分离卷积(Depthwise Separable Convolution)，极大地降低了模型的参数量与计算复杂度[11]；而 ShuffleNet 则通过通道洗牌(Channel Shuffle)操作解决了组卷积带来的信息流通不畅问题[12]。这些轻量化技术为本文设计高效分类模型提供了结构基础。

2.2. 多尺度特征提取技术

在复杂的图像识别场景中，目标的大小和位置具有极大的随机性。为了提升模型对不同尺度目标的鲁棒性，研究者们探索了多种空间特征集成方案。Inception 结构是多尺度特征提取的典型代表，它通过在同一层并联不同感受野的卷积分支，实现了特征在空间维度上的多尺度整合[13]。此外，空洞空间卷积池化金字塔(ASPP)通过引入不同采样率的空洞卷积(Atrous Convolution)，在不增加参数的前提下捕获了丰富的上下文信息，有效提升了模型对大尺寸目标的感知能力[14]。本文的模型设计将借鉴这种并行分支的思想，构建多分支特征提取模块。

2.3. 图像识别中的注意力机制

为了使模型能够在杂乱的背景中精准定位目标主体，注意力机制被广泛应用。卷积块注意力模块(CBAM)通过串联空间和通道注意力，进一步强化了模型对物体位置和特征重要程度的感知能力[15]。针对计算效率问题，ECA-Net 提出了一种不降维的局部跨通道交互策略，利用一维卷积代替全连接层，以极小的计算成本显著提升了分类精度[16]。本文将在多尺度融合过程中，引入类似的轻量化注意力单元，以实现对关键特征的自适应增强。

3. 本文提出的方法

3.1. 总体架构设计

针对电商场景下商品尺度不同以及背景干扰严重的挑战，本文提出了一种集成多尺度特征捕捉与注意力协同增强的高效神经网络模型：MSFF-EComNet。该模型的整体架构逻辑如图 1 所示。模型首先通

过轻量化的基础卷积层对输入图像进行特征预提取；随后，核心特征处理层由两个关键模块组成：多尺度并行卷积架构(MSP-Conv)用于同步捕获不同粒度的商品特征，改进的注意力协同机制(IASM)用于对提取的特征进行显著性增强与噪声抑制。最后，特征经由全局平均池化层输入分类器完成识别任务。

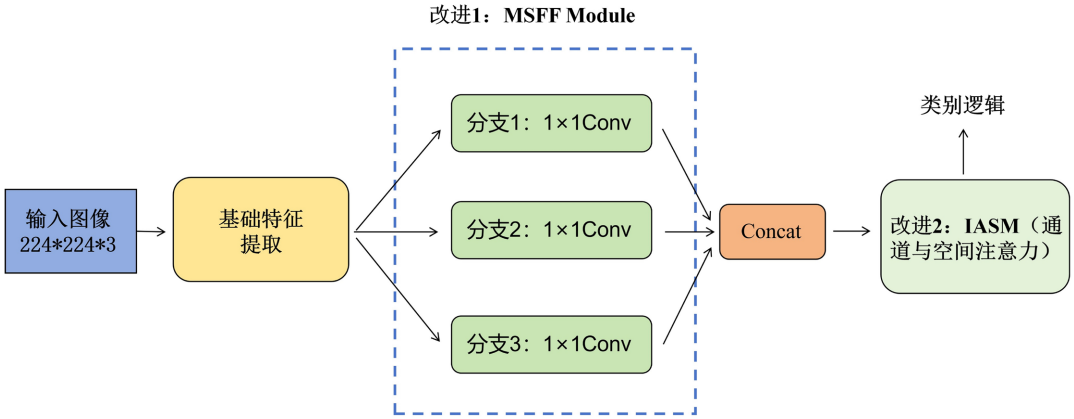

Figure 1. Overall framework of MSFF-EComNet
图 1. MSFF-EComNet 整体框架

3.2. 多尺度并行卷积架构设计

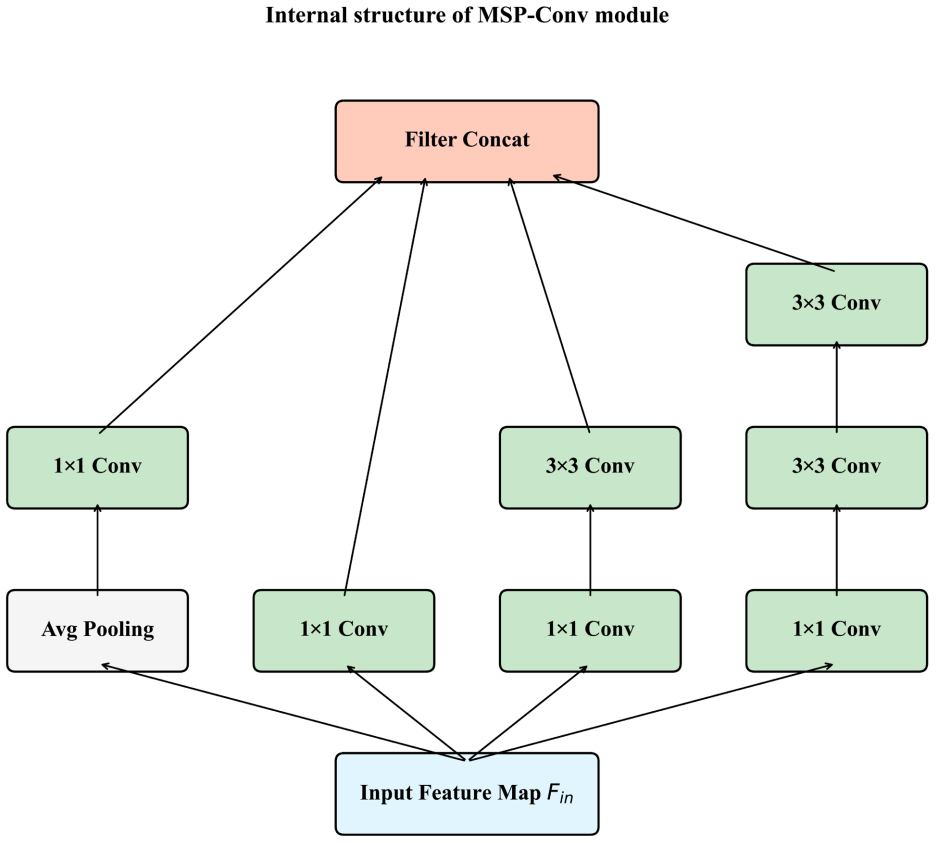

Figure 2. Schematic diagram of internal structure of multi-scale parallel convolution module
图 2. 多尺度并行卷积模块内部构造示意图

为了解决单一尺寸卷积核难以兼顾商品全局轮廓与细节部分(如面料纹理、品牌标识)的问题，本文设计了一种多分支并行特征提取模块。

该模块将输入的特征图 F_{in} 同时送入四个具有不同感受野的并行分支，如图 2 所示。MSP-Conv 模块由四个具有不同感受野的特征提取分支组成：细粒度特征分支：包含 1×1 卷积层，主要用于实现跨通道的信息交互与维度校正，捕捉商品图像中微小的像素级细节。标准局部特征分支：采用经典的 3×3 卷积层，作为提取商品主体几何结构与局部纹理的核心组件。宏观语义特征分支：为了在不显著增加参数量的情况下扩大感受野，本文参考 Inception 架构中的分解卷积策略[17]，利用两个级联的 3×3 卷积层来模拟 5×5 卷积的效果。该设计不仅能有效捕获商品的整体廓形信息，还能提高特征提取的非线性表达能力。特征聚合与融合：各分支提取的特征图最终通过滤波器拼接(Filter Concat)技术在通道维度进行融合。这种设计确保了模型能够在同一层级内实现对多尺度语义信息的同步感知，显著提升了模型在应对电商图像中商品尺度跨度剧烈时的分类稳健性。

提取的特征图在通道维度上进行拼接，融合后的特征 F_{ms} 可表示为：

$$F_{ms} = \text{Concat}\left[f^{1\times1}(F_{in}), f^{3\times3}(F_{in}), f^{5\times5}(F_{in})\right] \tag{1}$$

其中，$f^{k\times k}$ 表示卷积核尺寸为 k 的变换操作。这种并行结构确保了网络能够在同一层级内实现对多尺度特征的同步感知。

传统的 Inception 结构旨在通过增加网络宽度和深度来提升通用图像识别性能，其设计较为繁琐。相比之下，本文的 MSFF 模块专门针对电商图像中细节与整体并存的特点，精简了分支结构。

3.3. 改进的注意力协同机制

电商图像常包含直播间装饰、模特背景或复杂的室外实拍场景。为了引导网络忽略非目标干扰，本文引入了改进的注意力协同机制，从“通道”和“空间”两个维度对特征进行重构[15]。1) 通道注意力权重分配：通过全局池化压缩空间维度，学习不同特征通道的重要程度。对于电商图像而言，该模块能自动增强与商品属性相关的通道(如颜色通道、边缘通道)，而抑制与背景相关的冗余通道；2) 空间注意力特征聚焦：在通道增强的基础上，通过空间注意力掩码(Spatial Mask)对特征图进行加权，使得网络能够自动聚焦于图像中商品主体的所在区域，有效抑制边缘噪声。

标准 CBAM 采用简单的通道与空间注意力串联，而 IASM 采用了残差注意力协同连接。不同于传统的串行结构，IASM 经过注意力处理后的特征图与原始特征图进行像素级相加，确保了在增强关键特征的同时，不会丢失原始的多尺度语义信息。其最终输出特征 F_{final} 的计算公式为：

$$F_{final} = F_{ms} \oplus \left(M_s\left(M_c\left(F_{ms}\right)\right) \otimes F_{ms}\right) \tag{2}$$

其中，M_c 和 M_s 分别表示通道和空间注意力映射函数，$\otimes$ 表示元素级乘法。

4. 实验结果与分析

4.1. 实验环境与数据集

为了保证实验结果的可重复性与公平性，本文所有的训练与测试过程均在统一的硬件平台与软件环境下完成。实验环境配置如表 1 所示。本文选用大规模电商服装数据集 DeepFashion 进行实验。为确保实验的严谨性，从该数据集中抽取了 15 类常见商品(短裙、毛衣、西装、牛仔裤等)，总计 50,000 张图像。数据按 8:1:1 的比例划分为训练集、验证集和测试集。

Table 1. Experimental environment configuration table
表 1. 实验环境配置表

类别	相关配置
CPU	Intel(R) Core(TM) i9-13900K @ 3.00 GHz
GPU	NVIDIA GeForce RTX 4060TI 16 GB
软件工具	Cuda 11.6、Pytorch 1.13、Python 3.9
操作系统	Windows 11 64 位

4.2. 评价指标

为全面评估 MSFF-EComNet 的性能，本文采用以下学术通用指标：

精确率(Precision)，表示模型正确预测的所有样本(包括正例和负例)占总样本量的比例。其计算公式为：

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

召回率(Recall)，衡量的是在所有实际为“正例”的样本中，被模型正确预测出来的比例。其计算公式为：

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$F1$-Score，是两者的调和平均数，综合考虑了漏报和误报的影响。其计算公式为：

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

其中，真正例(True Positive, TP)表示模型预测该图像为 A 类商品，其实际类别也确实为 A 类。真负例(True Negative, TN)表示模型预测该图像不属于 A 类，其实际类别也确实不属于 A 类。假正例(False Positive, FP)表示模型预测该图像为 A 类商品，但其实际类别并非 A 类。假负例(False Negative, FN)表示模型预测该图像不属于 A 类，但其实际类别为 A 类。

Parameters 反映了模型对内存资源的占用情况。

4.3. 消融实验

为了深入探讨两个核心改进模块对模型分类性能的贡献，本文进行了一系列对比消融实验。通过在基准模型上逐步引入 MSP-Conv、IASM 模块，分析了各模块对模型精度、计算复杂度和推理速度的影响。

Table 2. Ablation experimental data
表 2. 消融实验数据

Model	MSP-Conv	IASM	Precision	Recall	$F1$
基准模型	-	-	82.35	72.9	81.7
1	√	-	86.42	74.7	86
2	-	√	85.18	75.1	80.8
MSFF-EComNet	√	√	**89.56**	**81.8**	**89.14**

当同时融合 MSFF 与 IASM 两个模块时，本文提出的 MSFF-EComNet 达到了最优性能。如表 2 所

示，最终模型的精确率升至 89.56%，*F*1 分数达到 89.14%。相比基准模型，总精确率提升了 7.21%。在精度上的巨大飞跃证明了这种策略的高效性。

4.4. 主流模型对比实验

为了验证 MSFF-EComNet 的先进性，本文将其与当前主流模型进行了性能对标：ResNet-50 [4]、MobileNet-V3 [18]、EfficientNet-B0 [19]。

Table 3. Comparative test of different models
表 3. 不同模型对比试验

Model	Precision/%	Recall/%	*F*1-score/%	Params (M)
ResNet-50	85.12	80.5	84.1	25.56
MobileNet-V3	79.74	77.8	78.9	7.3
EfficientNet-B0	87.35	84.4	86.4	10.1
MSFF-EComNet	89.56	88.1	88.3	11.79

由表 3 可知，ResNet-50 的精确率为 85.12%，*F*1 分数为 84.1%。本文模型在准确率上高出其 4.44%。MobileNet-V3 追求极端的计算效率，其准确率为 79.74%。本文模型在准确率上领先其 10%左右。这证明了在电商这类对细粒度特征要求极高的场景中，过度的轻量化会导致严重的漏报。EfficientNet-B0 以其优异的结构搜索闻名，准确率为 87.35%。本文模型仍保持了 2.21%的领先优势，这得益于 IASM 注意力机制对电商特有噪声的针对性抑制。

5. 总结

本文针对电子商务环境下商品图像尺度多变、拍摄背景复杂以及分类效率要求高等挑战，提出了一种基于多尺度特征融合的高效卷积神经网络模型 MSFF-EComNet。通过在 DeepFashion 公开数据集上的广泛实验与对比分析，得出以下结论：1) 本文设计的多尺度并行卷积架构(MSFF)通过并行的不同尺寸感受野分支，成功解决了单一卷积核难以兼顾商品局部细粒度特征与整体结构的问题。消融实验显示，该模块使模型的分类准确率提升了 4.07%。2) 本文提出的改进注意力协同机制(IASM)通过空间与通道维度的联合赋权，显著增强了模型在复杂背景中的特征提纯能力。在仅增加极小计算开销的前提下，将 *F*1 分数提升至 89%。3) 精度与效率的平衡：对比实验表明，MSFF-EComNet 的 Top-1 准确率达到了 89.56%。在性能上优于经典模型 ResNet-50 和 EfficientNet-B0 的同时，参数量仅为 11.79 M。这证明了本文模型能够很好地适配电商平台对高精度与高实时性的双重需求。

尽管本文提出的 MSFF-EComNet 模型在 DeepFashion 数据集上实现了分类精度与推理效率的平衡，但研究仍存在一定的局限性。因为实验样本主要集中于服装领域，模型对于数码产品、美妆工具等具有差异化几何特征的商品类目，其泛化能力仍需进一步验证；还可以引入模型量化与剪枝等轻量化技术，以提升模型在低算力环境下对小样本类目商品的识别准确性。

参考文献

[1] 季长清, 高志勇, 秦静, 等. 基于卷积神经网络的图像分类算法综述[J]. 计算机应用, 2022, 42(4): 1044-1049.

[2] 夏文生, 夏晓婧, 邹金宝. 基于卷积神经网络的遥感图像分类综述[J]. 机电工程技术, 2025, 54(21): 1-8.

[3] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. https://doi.org/10.1145/3065386

[4] 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

[5] Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X. (2016) DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 1096-1104. https://doi.org/10.1109/CVPR.2016.124

[6] Lin, T.Y., *et al*. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2117-2125.

[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] 朱子飘. 基于多尺度特征融合的医学图像分类研究[D]: [硕士学位论文]. 南宁: 广西大学, 2025.

[9] 陈超, 齐峰. 卷积神经网络的发展及其在计算机视觉领域中的应用综述[J]. 计算机科学, 2019, 46(3): 63-73.

[10] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556.

[11] Howard, A.G., *et al*. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv: 1704.04861.

[12] Zhang, X., Zhou, X.Y., Lin, M.X. and Sun, J. (2018) ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6848-6856. https://doi.org/10.1109/CVPR.2018.00716

[13] Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., *et al*. (2015) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[14] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. AND Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, 801-818. https://doi.org/10.1007/978-3-030-01234-2_49

[15] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[16] Wang, Q., *et al*. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 11531-11539.

[17] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2017). Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**, 4278-4284. https://doi.org/10.1609/aaai.v31i1.11231

[18] Howard, A., Sandler, M., Chu, G., *et al*. (2019) Searching for MobileNetV3. 2019 P*roceedings of the IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 1314-1324. https://doi.org/10.1109/ICCV.2019.00140

[19] Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International Conference on Machine Learning (ICML). arXiv: 1905.11946.