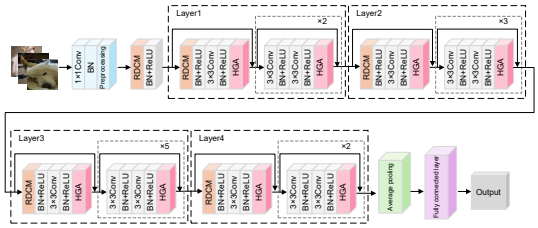
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 4 月, 第 53 卷, 第 4 期

旋转可变形卷积的图像分类网络



姜文涛[1], 董金良[1*], 张晟翀[2]

[1]辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105;
[2]光电信息控制和安全技术重点实验室, 天津 300308

摘要: 针对传统卷积操作在面对旋转、缩放等几何变换时特征提取能力有限, 以及难以有效区分关键特征与噪声的问题, 本文提出了旋转可变形卷积的图像分类网络 (image classification network with rotational deformable convolution, RoDeNet)。该网络基于 ResNet-34 进行改进, 首先引入旋转可变形卷积模块 (rotational deformable convolution module, RDCM), 通过结合预定义旋转偏移量与动态学习偏移量, 增强模型对方向敏感特征的捕捉能力。RDCM 采用双路径结构, 分别提取基础特征和旋转形变特征, 并通过特征保留模块 (feature preserve module, FPM) 平衡特征选择与信息保留。此外, 提出层次门控注意力机制 (hierarchical gated attention, HGA), 通过多粒度特征选择机制和动态权重优化, 提升模型对特征与噪声的辨别能力, 同时保持特征表达的丰富性。RoDeNet 在 CIFAR-10、CIFAR-100、Imagenette 和 Imagewoof 数据集上的准确率达到了 96.87%、81.81%、94.72% 和 89.69%, 对比 ResNet-34 准确率分别提高了 0.58%、3.01%、4.74% 和 5.09%。该网络模型通过旋转可变形卷积与层次门控注意力的协同作用, 有效解决了分类网络在几何变换下的特征提取不足和噪声敏感问题, 提升了对几何变换的鲁棒性和特征选择能力。

关键词: 图像分类; 残差网络; 旋转可变形卷积; 特征保留; 层次门控注意力

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250307 | CSTR: 32245.14.oee.2026.250307 | 中图分类号: TP391 | 文献标识码: A

姜文涛, 董金良, 张晟翀. 旋转可变形卷积的图像分类网络 [J]. 光电工程, 2026, **53**(4): 250307

Jiang W T, Dong J L, Zhang S C. Image classification network with rotating deformable convolution[J]. *Opto-Electron Eng*, 2026, **53**(4): 250307

1 引 言

图像分类是计算机视觉领域的基础性任务之一, 其目的是使计算机能够自动识别图像中所包含的对象或场景, 并将其划分到一个具体的类别中[1]。

随着深度学习技术的持续发展, 基于深度卷积神经网络 (CNN) 的图像分类方法已成为最主流且最成熟的网络架构之一[2]。文献 [3-8] 提出的经典残差深度网络及其变体通过加深、加宽和多分辨率建模不断提升特征表达能力。文献 [9-10] 则通过密集连接与复合缩放策略提升了特征重用效率。文献 [11-12] 提出的可变形卷积突破了固定卷积核结构的限制, 使网络能够自适应感受野以建模几何形变。近年来, 文献 [13-14] 通过改进归一化机制与大规模可变形卷积训练进一步提升了 CNN 的稳定性与扩展性。此外, TEC-Net[15] 在分支中引入动态可变形卷积, 并与 Transformer 分支的自适应注意力协同, 使网络在处理多尺度与几何变换任务时具备更强适应性。

在视觉分类任务中, 传统 CNN 固定感受野难以捕捉长距离依赖, 且不同区域的重要性被等同处理。为克服这些局限, 注意力机制被引入视觉模型。文献 [16-18] 提出的注意力与自注意力机制通过动态权重分配增强模

收稿日期: 2025-10-13; 修回日期: 2026-01-13; 录用日期: 2026-01-14; 网络出版日期: 2026-04-25
基金项目: 国家自然科学基金资助项目 (61601213); 辽宁省自然科学基金 (20170540426); 辽宁省教育厅重点基金 (LJYL049)
*通信作者: 董金良, 1606592672@qq.com

型的长程依赖建模能力。SENet[18] 的提出进一步提升了全局空间关系与通道依赖建模能力。随着注意力机制演进，文献 [19-20] 通过结构化稀疏策略和特征对齐方面进行了系统扩展。此外，文献 [21-22] 采用可变形建模与大感受野注意力的融合策略，从而提升模型在复杂纹理场景下的适应性。

尽管深度卷积神经网络和注意力机制在图像分类任务中取得了显著进展，但现有方法仍面临多方面的挑战。一方面，传统的卷积神经网络采用固定的卷积核几何结构与采样网格，面对具有旋转、缩放等几何变换的图像时，固定结构的卷积核无法自适应调整其感受野，导致特征提取错位，使得同一类别的不同形态实例在特征空间中产生差异，削弱了模型的泛化能力和鲁棒性。此外，网络对图像中背景、噪声或无关部分的特征与对目标物体的特征同样进行处理，导致信息冗余。难以有效区分重要与不重要的信息。

基于上述问题与研究现状，本文提出了旋转可变形卷积图像分类网络 (image classification network with rotational deformable convolution, RoDeNet)，该网络基于 ResNet34 进行改进，将网络输入层及各层级首个残差块的初始卷积替换为旋转可变形卷积模块 (rotational deformable convolutionmodule, RDCM)，将显式的几何先验知识与数据驱动学习相结合，使模型能够更好地适应各种几何变换，防止偏移发散；并在 RDCM 中加入特征保留模块 (feature preserve module, FPM) 减少信息损失，增强关键特征的稳定性，减少冗余内存访问；最后在残差结构中引入层次门控注意力 (hierarchical gated attention, HGA)，强化关键特征，同时抑制无关背景噪声的干扰。RoDeNet 通过几何自适应特征提取和注意力机制，提升了模型在复杂变换条件下的分类性能。

2 旋转可变形卷积的图像分类网络

2.1 RoDeNet 网络结构

RoDeNet 的主干部分包括以下四个阶段：

1) 输入预处理阶段：首先网络使用 1×1 卷积将原始图像进行通道维度扩展，再通过随机水平翻转、填充后随机裁剪以及对旋转、剪切等几何变换的幅度进行 0.5 倍缩放的 TrivialAugmentWide 图像增强操作，与 RDCM 方向角度设置形成互补，增强模型对微小方向变化的鲁棒性，配合批归一化处理实现数据标准化。

2) 初始特征提取阶段：经过预处理后的特征图经过 RDCM 特征提取卷积层，通过基础卷积路径和旋转可变形卷积路径提取低级特征，并使得特征提取过程能够自动适应输入数据的几何变换特性。

3) 深层特征提取阶段：特征图通过四个 layer 层，在每个 layer 层的第一个残差块使用 RDCM 提取深层特征，并在每个残差块后面插入 HGA 增强特征多样性，进行多层次的深特征提取，并降低背景噪声。

4) 分类输出阶段：网络首先通过全局平均池化将空间特征压缩，经展平操作后送入全连接层完成类别投影，得到最终分类结果。RoDeNet 的结构图如图 1 所示。

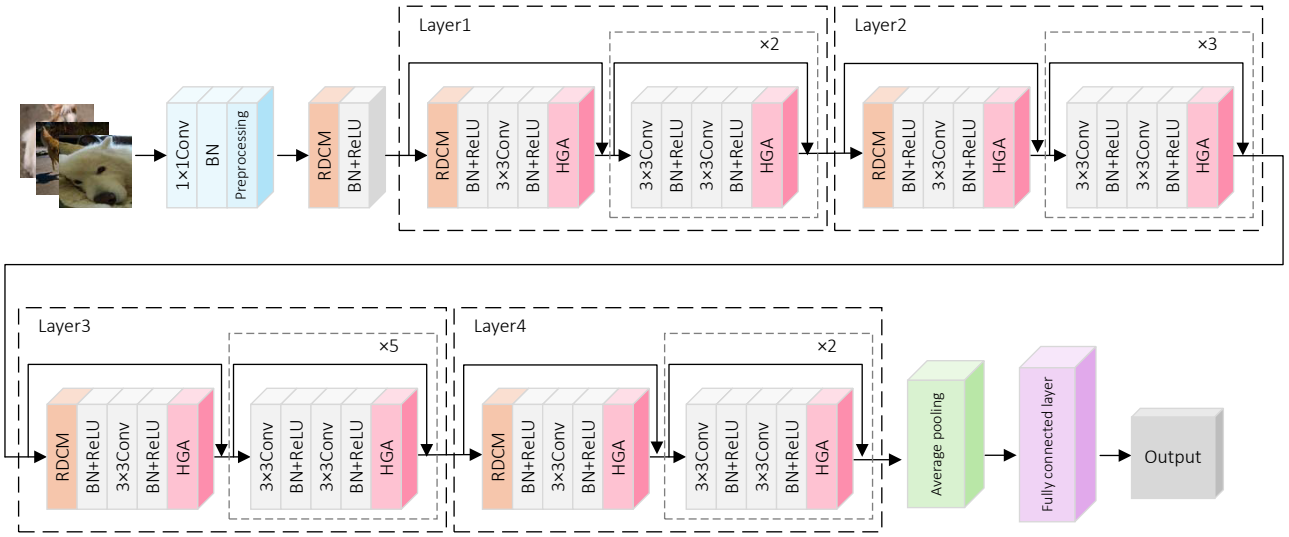


图 1 RoDeNet 结构图

Fig. 1 Architecture of RoDeNet

2.2 旋转可变形卷积模块

2.2.1 RDCM 整体结构

旋转可变形卷积模块 (RDCM) 是一种结合了基础卷积与具备动态旋转偏移的可变形卷积 (RDConv) 路径的新型卷积策略，再引入特征保留模块 (FPM)，实现更鲁棒的特征表达能力。与当前可变形卷积 (DCNv1/v2/v3) 相比，RDCM 在偏移生成方式、方向建模能力以及采样稳定性方面均具有本质差异。现有 DCN 系列完全依赖网络从数据中自由学习偏移量，缺乏显式几何约束，因此在深层网络中容易出现偏移发散、采样点越界以及特征错位等问题，且其偏移学习未引入方向先验，对旋转、倾斜等方向性强的结构敏感性不足。

为解决这一问题，RDCM 在偏移设计中显式引入四个基础旋转角度的预定义偏移，并与动态偏移进行加性合成，使卷积核的自适应偏移受限于稳定的几何基底，从而显著提升采样过程的方向一致性与几何稳定性。此外，RDCM 通过方向分组机制将通道划分为多个独立子空间，使不同方向的形变特征在各自子空间中建模，有效避免了传统 DCN 中多方向特征竞争造成的干扰，提高了方向结构的判别能力。RDCM 结构图如图 2 所示。

由图 2 可知，将特征图 Input 分组，分别进入基础卷积路径和旋转可变形卷积路径，基础卷积路径使用标准的卷积操作，保持稳定的初始特征提取；旋转可变形卷积路径通过初始定义四个角度 (0°, 45°, 90°, 135°)，将特征图分为四个方向组，各组分别通过偏移生成网络得到旋转偏移量，接着每个基础角度对应的卷积核位置都经过旋转变换，指导可变形卷积在特定方向上自适应调整感受野，再将每个方向组的计算结果进行加权求和。两条路径的输出在通道维度拼接后输入到特征保留模块，进行特征优化，增强重要特征并抑制噪声，并保持原始特征的基本结构和信息。最终输出融合了方向感知与空间自适应能力的特征图 Output。

RDCM 不仅能够有效捕捉图像中的方向性结构，还能在分类任务中缓解旋转导致的特征表达不一致、深层偏移不稳定以及方向纹理难以建模的问题。

2.2.2 动态旋转偏移

动态旋转偏移是在标准卷积核固定采样网格的基础上，为每个采样点添加可学习的位移向量，使标准卷积核具备旋转感知能力。

首先，为了在卷积采样中引入稳定的旋转先验，对卷积核的 $K\times K$ 网格坐标进行中心化处理得到 (x,y)，使采样点以卷积核几何中心作为旋转轴。在此基础上，基于预设的旋转角度集合应用旋转矩阵计算得到旋转后的坐标 (x',y')，旋转前后坐标的差值构成预定义旋转偏移量Δx和Δy：并由超参数 init_scale 控制偏移强度，以避免旋转过大导致采样点跳出局部邻域。最后，将每个点的偏移量按交替存储得到预定义偏移量$\boldsymbol{O}_{\text{predef}}$。中心化旋转与非中心化旋转对比如图 3 所示。

在引入旋转先验的同时，通过动态学习偏移让卷积核的每个采样点根据看到的图像内容自动调整位置。卷积核通过一个小型偏移生成网络 offset_head，从数据中学习相对于标准网格的调整量。设输入特征图为 $\boldsymbol{X}\in\mathbb{R}^{B\times C\times H\times W}$，其中 B 为批处理维度，C 为输入通道数，H 和 W 分别为高度和宽度。

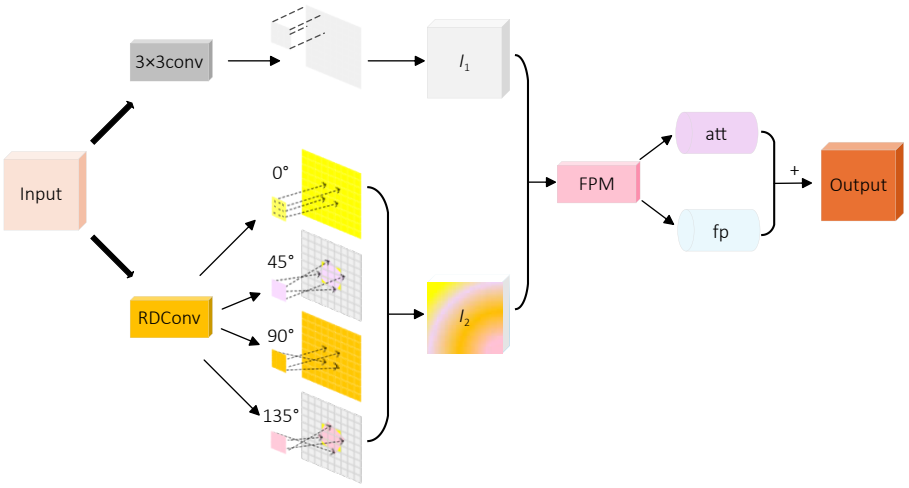


图 2 RDCM 结构图

Fig. 2 Architecture of RDCM

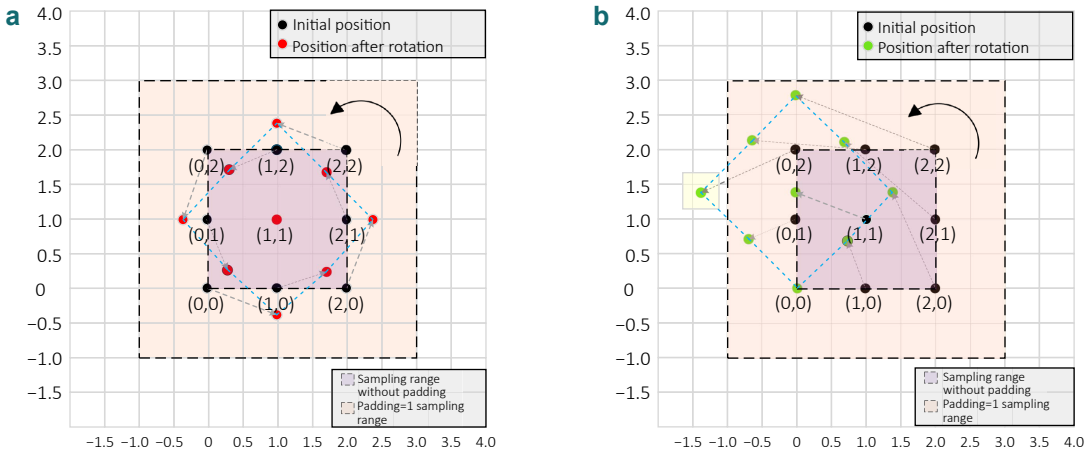


图 3 中心化与非中心化旋转对比图。(a) 中心化；(b) 非中心化

Fig. 3 Comparison of centered and non-centered rotation. (a) Centered; (b) Non-centered

输入特征图 $\boldsymbol{X}$ 先通过一个卷积层，生成初始偏移量预测值 $\boldsymbol{O}_{\text{raw}}$，如式 (1) 所示。

$$\boldsymbol{O}_{\text{raw}} = \boldsymbol{W}_{\text{conv}} \cdot \boldsymbol{X} + \boldsymbol{b}_{\text{conv}}\,, \tag{1}$$

式中：$\boldsymbol{W}_{\text{conv}} \cdot \boldsymbol{X}$ 为卷积操作；$\boldsymbol{b}_{\text{conv}}$ 是用于生成动态偏移量的偏置项初始化值，为不同方向组的偏移量提供初始值。

再对 $\boldsymbol{O}_{\text{raw}}$ 进行 GroupNorm 分组归一化，防止不同方向组的偏移量数值差异过大，得到归一化的偏置量，并通过 GELU 函数激活，实现平滑的非线性变换，得到最终预测的动态偏移量 $\boldsymbol{O}_{\text{learn}}$，如式 (2) 所示。

$$\boldsymbol{O}_{\text{learn}} = GELU(GroupNorm(\boldsymbol{O}_{\text{raw}}))\,. \tag{2}$$

GELU 函数表示为

$$\begin{cases} \boldsymbol{GELU}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\phi}(\boldsymbol{x}) \\ \boldsymbol{\phi}(\boldsymbol{x}) = \dfrac{1}{2}\left(1 + \text{erf}(\boldsymbol{x}/\sqrt{2})\right) \end{cases}, \tag{3}$$

式中：$\boldsymbol{\phi}(\boldsymbol{x})$ 为标准高斯 CDF，能够平滑调整偏移量，使偏移量学习更符合数据分布，增强模型对形变的适应能力。

最后，为同时利用显式旋转先验与数据驱动的自适应能力，将预定义旋转偏移 $\boldsymbol{O}_{\text{predef}}$ 与动态学习偏移 $\boldsymbol{O}_{\text{learn}}$ 逐元素相加，得到最终偏移量 $\boldsymbol{O}_{\text{final}}$，如式 (4) 所示。

$$\boldsymbol{O}_{\text{final}} = \boldsymbol{O}_{\text{predef}} + \boldsymbol{O}_{\text{learn}}\,. \tag{4}$$

2.2.3 旋转可变形卷积

旋转可变形卷积 (RDConv) 通过把特征图分为 D 个方向组，并且使用偏移量 (offset) 调整标准网格坐标，动态调整采样位置，每组使用独立的通道映射和卷积核权重 (weight)，对 K^2 个采样点和 D 个方向组的计算结果进行加权求和，生成输出特征图 $\boldsymbol{Y}$。旋转可变形卷积结构图如图 4 所示。

首先，将输入特征图 $\boldsymbol{X} \in \mathbb{R}^{B\times C\times H\times W}$ 按照 D (旋转角度数量) 把通道分为 D 个子集，每个子集独立处理：先把输出特征图空间位置 (h,w) 的每个采样点 P 的标准网格坐标根据最终偏移量张量 $\boldsymbol{O}_{\text{final}}$ 提取的偏移量进行调整，如果坐标 $P(h,w)$ 为整数，直接索引，如果坐标为非整数，采用双线性插值从输入特征图中获取对应值，以实现连续空间下的平滑采样，得到最终采样坐标 $P_*(h,w)$，如式 (5) 所示。

$$P_*(h,w) = P(h,w) + \boldsymbol{O}_{\text{final}}(h,w)\,, \tag{5}$$

接着对得到的最终采样坐标 $P_*(h,w)$ 进行可变形卷积操作并融合所有分组特征输出特征图 $\boldsymbol{Y}$，如式 (6) 所示。

$$\boldsymbol{Y} = \sum\nolimits_{d,k} \boldsymbol{X}(P_*(h,w)) \cdot \boldsymbol{W}_{d,k}\,, \tag{6}$$

式中：d 表示遍历所有方向组；k 表示遍历卷积核采样点，最终累加所有方向组的贡献；$\boldsymbol{W}_{d,k}$ 表示方向 d 下第 k 个采样点的卷积权重。

2.2.4 特征保留模块

特征保留模块 (FPM) 通过结合注意力驱动的注意力分支与特征保留分支，能够在增强特征判别力的同时避免信息损失。设输入 FPM 模块的特征图为 $\boldsymbol{F} \in \mathbb{R}^{B\times C\times H\times W}$，同时进入注意力分支 (att) 和特征保留分支 (fp)，FPM 结构如图 5 所示。

在注意力分支中，特征图 $\boldsymbol{F}$ 先经过一个 3×3 的降维卷积层，将输出通道维数压缩为 $C/4$，减少计算量，接着进行分组归一化 (GroupNorm, GN)，对特征图沿通道维度分为 4 组，每组独立归一化，再经过GELU激活函

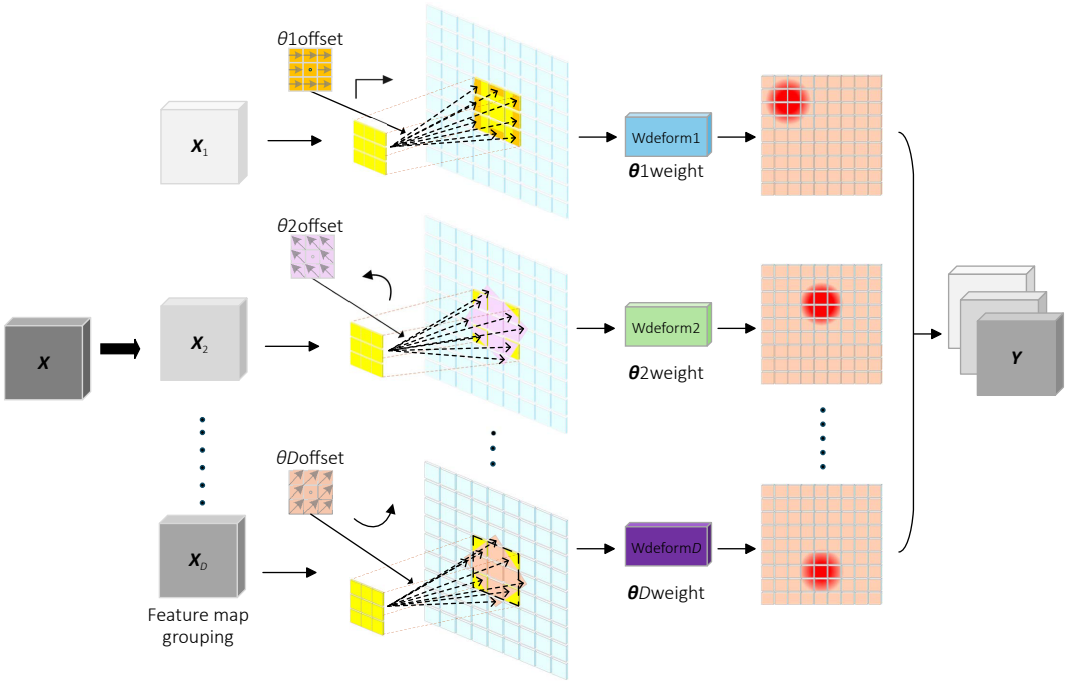


图 4 旋转可变形卷积结构图

Fig. 4 Architecture of rotational deformable convolution

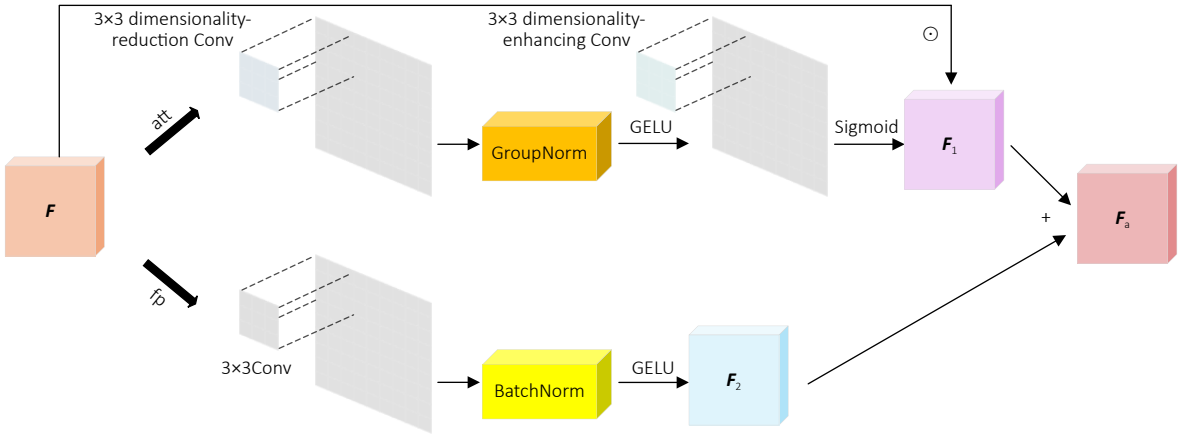


图 5 特征保留模块结构图

Fig. 5 Architecture of the feature preserve module

数，增强表达能力，得到特征图$\boldsymbol{F}_{\mathrm{g}}$，如式 (7) 所示。

$$\boldsymbol{F}_{\mathrm{g}}=GELU(GroupNorm(\boldsymbol{W}_1\cdot\boldsymbol{F}+\boldsymbol{b}_1)), \tag{7}$$

式中：$\boldsymbol{W}_1$为降维卷积核权重；$\boldsymbol{b}_1$为降维偏置项。

然后进入一个 3×3 升维卷积层，与降维卷积形成“瓶颈”结构，增强非线性建模能力，再通过Sigmoid激活函数，生成空间注意力权重，得到特征图$\boldsymbol{F}_1$，如式 (8) 所示。

$$\boldsymbol{F}_1=Sigmoid(\boldsymbol{W}_2\cdot\boldsymbol{F}_{\mathrm{g}}+\boldsymbol{b}_2), \tag{8}$$

式中：$\boldsymbol{W}_2$为升维卷积核权重；$\boldsymbol{b}_2$为升维偏置项。Sigmoid函数计算公式如式 (9) 所示。

$$Sigmoid(x)=\frac{1}{1+\mathrm{e}^{-x}}, \tag{9}$$

在特征保留分支中，特征图$\boldsymbol{F}$先通过一个 3×3 卷积层，保留原始特征结构，再通过批量归一化(BatchNorm, BN)，稳定特征分布，最后通过GELU函数激活，引入非线性，增强特征多样性，得到特征图$\boldsymbol{F}_2$，如式 (10) 所示。

$$\boldsymbol{F}_2 = GELU(BatchNorm(\boldsymbol{W}_3 \cdot \boldsymbol{F}) + \boldsymbol{b}_3)\,, \tag{10}$$

式中：$\boldsymbol{W}_3$为卷积核权重；$\boldsymbol{b}_3$为偏置项。

最后进行特征融合操作，对输入特征图 $\boldsymbol{F}$ 的每个位置进行逐元素乘法，用注意力权重调整特征强度，再与特征保留分支中的输出$\boldsymbol{F}_2$按相同维度逐元素相加，得到最终输出$\boldsymbol{F}_a$，如式 (11) 所示。

$$\boldsymbol{F}_a = \boldsymbol{F} \odot \boldsymbol{F}_1 + \boldsymbol{F}_2\,. \tag{11}$$

2.3 RDCM 特征提取卷积层

在特征表示层面，传统首层卷积缺乏对旋转特征的解耦表达能力，其所有通道共享相同的空间采样模式，无法区分不同旋转方向的特征。因此，本文提出了 RDCM 模块替换首层卷积，并加入输入预处理层，得到 RDCM 特征提取卷积层。

在 RDCM 特征提取卷积层中，确保网络在训练初期就具备基本的旋转感知能力，并使网络能够自主决定在什么情况下依赖哪种特征。此外，其输入预处理层通过 1×1 卷积在通道维度进行线性组合，将原始的 RGB 颜色空间特征转换为更高维的抽象特征表示，能够确保每个旋转方向都能分配到均等的特征通道进行独立处理，每个扩展通道可以专注于特定方向的模式提取。RDCM 特征提取卷积层如图 6 所示。

2.4 层次门控注意力

层次门控注意力 (HGA) 通过分组计算与双分支协同机制，有效缓解了多尺度特征融合的语义冲突问题。设输入特征图为$\boldsymbol{Z} \in \mathbb{R}^{B\times C\times H\times W}$，首先进行 Reshape 操作将特征图按通道维度分为 N 组，分组后的特征图，同时分别输入空间池化分支和深度可分离卷积分支。HGA 结构如图 7 所示。

在空间池化分支中，首先经过一个全局平均池化层，得到水平与垂直方向的注意力描述向量，如式 (12) 所示。

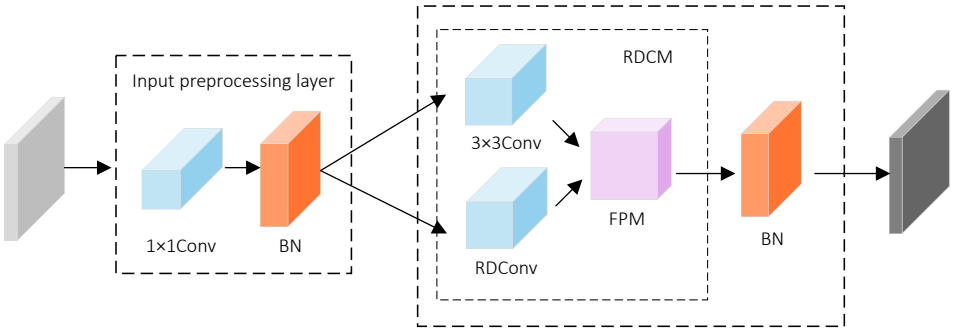


图 6 RDCM 特征提取卷积层结构图

Fig. 6 Architecture of the feature extraction convolutional layer in RDCM

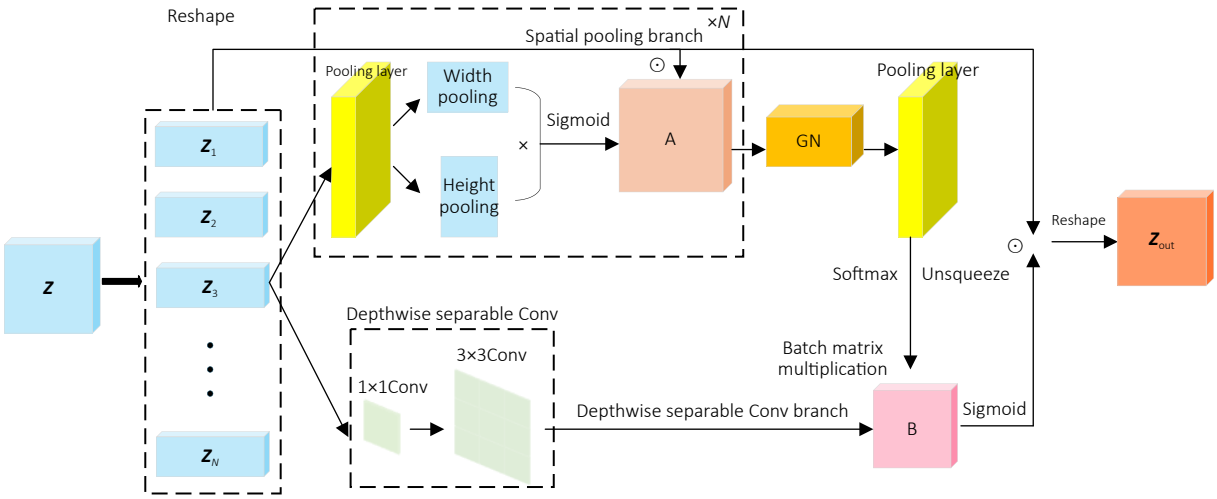


图 7 层次门控注意力结构图

Fig. 7 Architecture of hierarchical gated attention

$$\begin{cases} \boldsymbol{S}_{\mathrm{w}} = AvgPool_{\mathrm{W}}(\boldsymbol{Z}_N) \\ \boldsymbol{S}_{\mathrm{h}} = AvgPool_{\mathrm{H}}(\boldsymbol{Z}_N)^{\mathrm{T}} \end{cases}, \tag{12}$$

式中：$AvgPool_{\mathrm{W}}$和$AvgPool_{\mathrm{H}}$分别表示在宽度与高度维度上进行全局平均池化。

接着将宽度和高度方向注意力进行合并，通过矩阵乘法生成空间注意力矩阵，再经过Sigmoid激活函数，生成空间注意力图，最后对输入特征与空间注意力图进行逐元素相乘得到特征图S_{att}，如式 (13) 所示。

$$\boldsymbol{S}_{\mathrm{att}} = \boldsymbol{Z}_N \odot Sigmoid(\boldsymbol{S}_{\mathrm{w}} \cdot \boldsymbol{S}_{\mathrm{h}}) . \tag{13}$$

最后通过组归一化 (GN)，稳定数值分布，再对得到的特征进行全局池化，提取通道全局信息，并通过Softmax函数归一化得到特征图$\boldsymbol{S}_{\mathrm{f}}$，如式 (14) 所示。

$$\boldsymbol{S}_{\mathrm{f}} = Softmax(AvgPool(GroupNorm(\boldsymbol{S}_{\mathrm{att}}))) , \tag{14}$$

$Softmax$函数表达式为

$$Softmax(\boldsymbol{x}) = \frac{\mathrm{e}^{\boldsymbol{x}_i}}{\sum_{j=1}^{K} \mathrm{e}^{\boldsymbol{x}_j}} , \tag{15}$$

式中：$\boldsymbol{x}_i$表示输入向量的第 i 个元素；K为向量的总长度；j是求和过程的索引变量。

在深度可分离卷积分支中，将分组后的特征图先通过一个 1×1 点卷积，用于混合通道信息，提升特征表达能力，再通过一个 3×3 深度卷积，进行空间特征提取，保持通道独立性，最后将空间维度展平，得到特征图$\boldsymbol{D}_{\mathrm{a}}$，如式 (16) 所示。

$$\boldsymbol{D}_{\mathrm{a}} = Vec(Conv_{3\times3}(Conv_{1\times1}(\boldsymbol{Z}_N))) , \tag{16}$$

式中：Vec表示将张量的指定维度展平为一维向量，使通道权重能同时作用于所有空间位置。

再将两个分支的权重融合，先将$\boldsymbol{S}_{\mathrm{f}}$进行unsqueeze操作，确保适配批量矩阵乘法的输入要求，接着进行批量矩阵乘法，进行加权求和，并把特征图的形状重塑，保持与输入特征图空间对齐，得到特征图$\boldsymbol{T}$，如式 (17) 所示。

$$\boldsymbol{T} = Vec(unsqueeze(\boldsymbol{S}_{\mathrm{f}}) \cdot \boldsymbol{D}_{\mathrm{a}}) . \tag{17}$$

最后，特征图$\boldsymbol{T}$经过Sigmoid激活函数，并且进行特征重加权操作，通过动态生成的注意力权重对输入特征进行像素级调整，最后合并分组，将分组维度合并回通道维度，得到最终特征图，如式 (18) 所示。

$$\boldsymbol{Z}_{\mathrm{out}} = Reshape(\boldsymbol{Z}_N \odot Sigmoid(\boldsymbol{T})) , \tag{18}$$

式中：Reshape 操作表示将分组处理后的特征恢复为标准的 4D 张量格式，确保输出与输入维度完全一致。

HGA 通过对输入特征进行分组处理，并引入深度可分离卷积，进一步提高了特征提取的效率。同时，该模块通过共享池化操作分别提取水平和垂直方向的特征信息，并在此基础上生成空间注意力权重，提升了网络对目标的感知能力和整体特征质量。

2.5 融合 RDCM 与 HGA 的残差块

传统 ResNet 中的残差块 (记为 Block)，如图 8(a) 所示，采用简单的两个 3×3 卷积堆叠结构，配合恒等映射的残差连接，保证了梯度的有效传播。然而，当输入特征发生旋转或形变后，经过第一个 3×3 卷积时，卷积核的采样点无法自适应调整，固定网格采样模式会强制将各种几何变换的特征压缩到同一表示空间，导致特征错位，并且传统卷积对所有通道特征均等处理，无法区分关键特征与噪声。其残差映射关系如式 (19) 所示。

$$y = F(x) + x , \tag{19}$$

式中：$F(x)$表示主路径处理后的特征；x 表示输入特征。

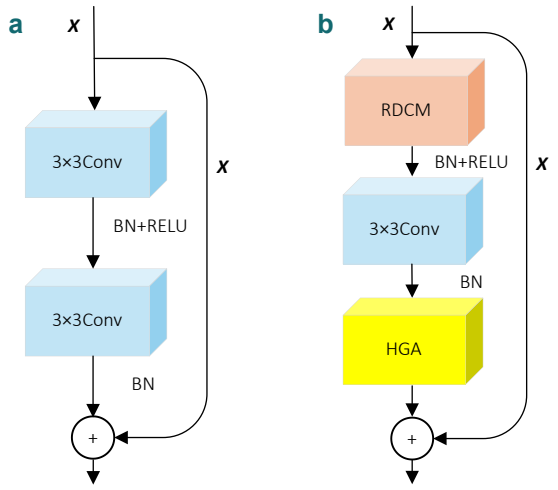


图 8 残差结构对比图。(a) Block；(b) RH-Block

Fig. 8 Residual structure comparison chart. (a) Block; (b) RH-Block

因此，本文将 Block 中的首层卷积替换为 RDCM，得到 RH-Block，如图 8(b) 所示。通过 RDCM 自动调整采样位置，保持特征表达的一致性，阻断了首层特征偏差的级联放大，为后续层提供了稳定的特征基础；当特征通过首层 RDCM 完成几何自适应校正后，特征逐渐从几何敏感的低级特征向语义相关的高级特征转变，因此第二层卷积保持标准 3×3 结构，用于提取与几何变换无关的高层语义特征；在经过两层卷积后，引入 HGA 模块，作用于经过几何校正的低级特征和初步抽象的高级语义特征，注意力机制可以专注于语义层面的特征选择。RH-Block 的残差映射关系如式 (20) 所示：

$$y = H(R(x)) + x, \tag{20}$$

式中：$H(R(x))$表示 RDCM 处理后再经过 HGA 处理后的特征。

3 实验与结果分析

3.1 实验环境及实验数据集

本文实验采用的操作系统：Ubuntu 22.04；GPU：NVIDIA RTX 3090 显卡，显存容量：24 GB；Pytorch：2.1.0；CUDA：12.1.0；CUDNN：8.6.0。为了评估 RoDeNet 网络性能，本文实验采用 CIFAR-10、CIFAR-100、Flowers-102、STL-10、Imagenette 和 Imagewoof 数据集，数据集信息如表 1 所示。

实验评价指标采用图像分类的分类准确率 (accuracy)。在模型训练中，迭代次数 (epochs) 为 200，CIFAR-10、CIFAR-100、Flowers-102 和 STL-10 数据集的批次大小 (batch_size) 为 128，Imagenette 和 Imagewoof 数据集的批次大小为 32，学习率 (lr) 设置为 0.1，动量 (momentum) 设置为 0.9，权重衰减 (weight_decay) 设置为 5e-4，优化器为随机梯度下降 (SGD)。

3.2 base_angles 参数实验

为了研究不同旋转角度集合对网络性能的影响，本节在 CIFAR-100 数据集上进行 base_angles 参数实验，其中方案 1 角度集合为 [0°，90°]，方案 2 角度集合为 [0°，45°，90°，135°]，方案 3 角度集合为 [0°，60°，120°，180°]，方案 4 角度集合为 [0°，22.5°，45°，67.5°，90°，112.5°，135°，157.5°]。由于模块通过引入动态可学习偏移量机制，模型可自动补偿离散角度间隔的不足，因此无需穷举所有可能的角度组合。不同方案的准确率 (accuracy)、参数量 (parameters) 和模型计算量 (FLOPs) 如表 2 所示。

可知，方案 1 虽然具有最低的参数量和计算量，但准确率也达到 79.88% 的最低准确率，而方案 2 和方案 3 在参数量和计算量完全相同的条件下，准确率却分别达到 80.73% 和 80.48%，显然方案 2 更有效，而方案 4 在参数量和计算量最高的情况下，准确率不如方案 2。因此，基于旋转对称群的完备性理论[23] 和实验结果，选择方案 2 作为预定义旋转角度集合，该组合满足二维平面上 SO (2) 群的最小完备表示要求[24]。

3.3 init_scale 参数实验

init_scale 在 RDCM 中用于控制预定义旋转形变的初始强度，在先验知识与数据驱动学习之间建立动态平衡。它决定了模型在初始化时，对输入特征施加的几何变换的幅度，直接影响模型的旋转适应能力和训练稳定性。因此本节实验针对 init_scale 的不同取值对网络性能的影响进行实验分析。init_scale 在实验中取值范围为 0~1，在CIFAR-10、CIFAR-100、Imagenette 和Imagewoof 数据集上的实验结果如表 3 所示。

表 1 实验数据集信息

Table 1 Experimental dataset information

Dataset	Pixels	Categories	Testset	Trainset
CIFAR-10	32×32	10	50000	10000
CIFAR-100	32×32	100	50000	10000
Flowers-102	96×96	102	14740	15153
STL-10	96×96	10	5000	8000
Imagenette	224×224	10	9469	3925
Imagewoof	224×224	10	9025	3929

表 2 base_angles 参数实验信息

Table 2 base_angles parameter experiment information

Method	Accuracy/%	FLOPs/G	Parameters/M
1	79.88	1.432	25.462
2	80.73	1.463	25.630
3	80.48	1.463	25.630
4	80.69	1.534	25.971

表 3 init_scale 参数实验结果

Table 3 init_scale parameter experiment results

Init_scale	CIFAR-10/%	CIFAR-100/%	Imagenette/%	Imagewoof/%
0.0	96.49	80.73	94.34	88.75
0.1	96.72	80.84	94.42	88.97
0.2	96.66	80.12	94.26	88.75
0.3	96.66	79.86	94.34	88.69
0.4	96.6	80.17	94.39	88.62
0.5	96.46	80.58	94.36	88.87
0.6	95.53	80.79	94.26	88.67
0.7	96.55	80.5	94.34	88.34
0.8	96.52	80.43	94.31	88.69
0.9	96.49	80.43	93.88	88.26
1.0	96.56	80.39	94.29	87.96

从实验数据来看，init_scale=0.1 在四个数据集上均取得了最优结果，这一现象表明适度的初始旋转偏移能够有效平衡几何先验与数据驱动的自适应能力。当 init_scale 取值在 0.1~0.4 区间时，准确率波动范围普遍小于 1%，这说明 RoDeNet 对这一参数范围内的设置具有较好的鲁棒性。当参数值超过 0.6 时，模型性能开始出现明显分化，这种差异可能源于不同数据集对几何变换的敏感度不同，特别是在 init_scale=1.0 时，表明过强的几何先验会严重制约模型对自然形变的适应能力。当 init_scale=0.0 时，所有数据集的性能都未能达到最优水平，这验证了纯数据驱动学习在旋转建模中的局限性。实验表明，init_scale=0.1 时能够取得最佳平衡，说明预定义偏移强度并非越大越好。

3.4 HGA 模块插入位置实验

为了验证 HGA 模块在残差块的不同位置对于网络性能的影响，本节实验在 CIFAR-10、CIFAR-100、Flowers-102 和 STL-10 数据集上进行验证，其中位置 A 为 HGA 模块嵌入到残差块第一个卷积后，位置 B 为 HGA 模块嵌入到残差块第二个卷积后，位置 C 为 HGA 模块嵌入到残差连接中，位置 D 为 HGA 模块嵌入到残差连接之后。嵌入位置如图 9 所示。实验结果如表 4 所示。

综合四个数据集的实验结果，HGA 模块插入在位置

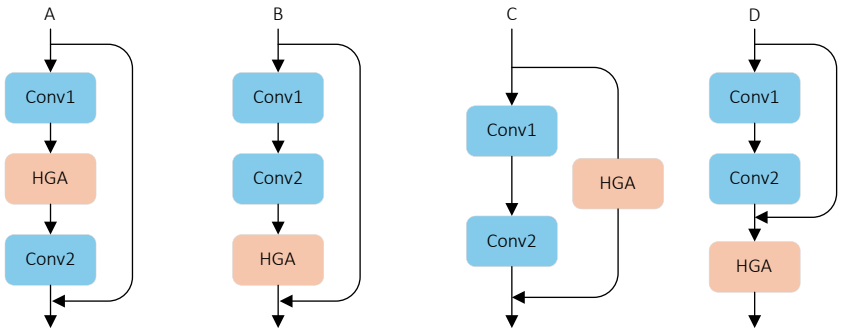


图 9 HGA 嵌入位置图

Fig. 9 HGA embedding location diagram

表 4 HGA 插入位置实验

Table 4 HGA insertion position experiment

Position	CIFAR-100/%	Flowers-102/%	STL-10/%	Imagenette/%
A	79.74	90.35	80.37	92.2
B	80.37	91.2	82.86	92.86
C	79.88	86.44	75.62	91.03
D	80.09	88.76	79.78	91.54

B 普遍表现最优，在 CIFAR-100、Flowers-102、STL-10 和 Imagenette 数据集上分别达到 80.37%、91.2%、82.86% 和 92.86% 的准确率，显著优于其他位置。相比之下，插入在 C 位置的效果最差，尤其在 Flowers-102 和 STL-10 数据集上表现明显下降。A 位置的嵌入因特征过于底层，注意力机制难以有效聚焦于有判别性的区域。而 D 位置的嵌入会破坏残差结构的核心设计，导致梯度传播受阻。B 位置嵌入注意力模块的优势在于既保留了足够的空间信息，又具备一定的语义抽象能力，适合通过注意力机制增强重要特征的权重。因此选择将 HGA 模块嵌入在 B 位置。

3.5 网络模型对比实验

3.5.1 参数量、计算复杂度和运行速度对比实验

为全面评估模型的综合性能与效率，将 RoDeNet 与其他网络模型在参数量 (parameters)、计算复杂度 (FLOPs) 和运行速度 (speed) 上进行对比。对比结果如表 5 所示。

其中，RoDeNet 参数量为 25.65 M，与 CAPR-DenseNet 相近，但明显低于 HO-ResNet 和 Multi-ResNet 大规模增强型网络，相比 Couplformer、QKFormer 这类 Transformer 模型，RoDeNet 也避免了注意力机制带来的大量投影矩阵开销，具有良好的参数利用效率。RoDeNet 的 FLOPs 为 5.73 G，虽高于 ResNet-34，但远低于 DenseNet 和 Transformer 架构的高计算量，保证了在性能提升的同时保持较高计算效率。在推理速度方面，RoDeNet 的 Speed 为 1.78 f/s，虽不及 ResNet-34 的 3.02 f/s，但略快于 Multi-ResNet 的 1.62 f/s，且相较于参数量和 FLOPs 更高的多路径网络和 Transformer 模型，它能够获得更显著的分类性能。总体来看，RoDeNet 在不显著增加参数量与计算量的前提下实现了性能提升，用合理的速度牺牲换来了更强的特征提取能力与分类性能。

3.5.2 分类准确率对比实验

为了验证 RoDeNet 的有效性，本节实验选择 CIFAR-10、CIFAR-100、Imagenette 和 Imagewoof 数据集进行网络模型对比实验。对比网络模型有：ResNet-34、FAVOR+[25]、HO-ResNet[26]、WideResnet-28-10、Multi-ResNet、ABNet-2G-R3[27]、SSCNet[28]、MSMDFormer[29]、ATONet[30]、Couplformer[31]、MMA-CCT-7/3×2[32]、QKFormer[33]、ViT-Lite-7/4[34]、FMDConv[35] 和 CIARD[36] 共 16 个网络模型。其中，未开源的代码优先采用对比网络所对应论文提供的实验结果；开源的代码通过论文提供的开源代码复现。实验结果如表 6 所示。

由表 6 可知，RoDeNet 在各个数据集的准确率均展现出显著优势，在 CIFAR-10 和 CIFAR-100 数据集上准确率分别达到了 96.87% 和 81.81%，在 Imagenette 和 Imagewoof 这两个具有复杂几何变换的数据集中，准确率分别达到了 94.72% 和 89.69%，领先于其他模型。其中 HO-ResNet 通过高阶重构模块增强浅层特征表达能力，适合捕获纹理边缘等局部信息，对类别结构复杂的图像难以提供稳定的表征；ATONet 的多路径设计增强了类型区分能力，但多路径之间缺乏有效的依赖建模，对细节密集或局部结构变化丰富的图像处理能力受限；FMDConv 优势主要体现在频域增强，但其忽略空间结构内部的关联，对结构变化明显或纹理干扰较强的图像适应性不高。

对于基于 Transformer 结构的模型，MMA-CCT-7/3×2 通过多流混合注意力机制在全局建模方面表现突出，但其局部特征提取能力弱于传统卷积架构；MSMDFormer 结合多尺度特征，但其缺乏强局部先验，在数据量有限时难以稳定捕获细节结构；ViT-Lite-7/4 作为轻量 Transformer 对数据规模和分辨率依赖较强，

表 5 参数量、计算复杂度和运行速度对比实验结果

Table 5 Comparative results for parameter count, computational complexity, and inference speed

Network	Parameters/M	FLOPs/G	Speed/(f/s)
ResNet-34	21.33	3.61	3.02
CAPR-DenseNet	25.51	17.73	2.86
Couplformer	27.63	14.29	2.73
WideResnet-28-10	36.5	5.96	–
HO-ResNet	50.26	11.05	1.93
Multi-ResNet	51.23	11.73	1.62
QKFormer	35.62	26.39	–
RoDeNet	25.65	5.73	1.78

表 6 网络模型对比实验结果

Table 6 Comparative results for different network models

Network	CIFAR-10/%	CIFAR-100/%	Imagenette/%	Imagewoof/%
ResNet-34	96.29	78.8	89.98	84.6
FAVOR+	91.42	72.56	88.16	77.57
HO-ResNet	96.32	77.12	87.23	79.74
WideResnet-28-10	95.87	80.50	88.34	78.71
Multi-ResNet	94.56	78.69	87.71	81.26
ABNet-2G-R3	96.08	80.83	–	–
DAMSNet	96.51	80.50	–	–
SSCNet	96.69	80.63	88.75	82.09
MSMDFormer	95.65	77.46	–	–
ATONet	94.51	78.56	86.68	80.21
Couplformer	93.54	73.92	87.91	77.89
MMA-CCT-7/3×2	94.74	77.51	–	–
QKFormer	96.18	80.27	88.42	81.67
ViT-Lite-7/4	93.57	73.94	–	–
FMDConv	94.21	74.99	–	–
CIARD	91.8	73.4	–	–
RoDeNet	96.87	81.81	94.72	89.69

但缺乏稳健的卷积先验，导致特征表达偏弱。

3.5.3 精确率、召回率和 F1 分数对比实验

精确率 (precision) 是在所有被模型预测为正类的样本中，真正是正类的比例，召回率 (recall) 是在所有实际是正类的样本中。F1 分数 (F1-Score) 是精确率和召回率的调和平均数，F1-Score 高表明模型在精确率和召回率两方面都表现良好，性能均衡且整体优秀。本节实验对比了 RoDeNet 与 ResNet-34、WideResnet-28-10 在三项指标上的表现，实验结果如表 7 所示。

实验结果显示，RoDeNet 在三项指标上均取得了最佳表现，其精确率为 0.8191，召回率为 0.8181 和 F1 分数为 0.8180，均显著高于另外两个对比模型。其中 ResNet-34 通过标准卷积层堆叠与残差连接，其相对简单和固定的结构在复杂数据分布上可能受限；WideResnet-28-10 单纯增加宽度虽然提升了模型容量，但也引入了更多的参数，其性能提升受限。实验证明了 RoDeNet 在分类任务中有更高的可靠性与泛化性，它不仅能够更准确地将正样本识别出来，而且在做出正类预测时出错的概率更低，最终实现了较好的综合性能。

3.6 消融实验

为了对比 RDCM 和 HGA 模块在网络中的有效性，本节在 CIFAR-10、CIFAR-100、Imagenette 和 Imagewoof 数据集上进行 RoDeNet 的消融实验。图 10 为各个模块在四个数据集上的准确率的趋势变化，实验结果如表 8 所示，其中“-----”为基线网络。

实验结果表明，分别引入 RDCM 和 HGA 模块都为模型性能带来了显著提升。RDCM 提供的方向敏感特征为注意力机制提供了丰富的底层特征，HGA 则通过权重分配突出 RDCM 提取的判别性方向特征，完整模型的性能提升超过各模块单独性能提升效果，二者协同作用在保持旋转鲁棒性的同时实现判别性特征的自适应增

表 7 精确率、召回率和 F1 分数对比实验结果

Table 7 Comparative results for precision, recall, and F1 score

Network	Precision/%	Recall/%	F1-Score/%
ResNet-34	78.31	78.28	78.20
WideResnet-28-10	80.51	80.42	80.39
RoDeNet	81.91	81.81	81.80

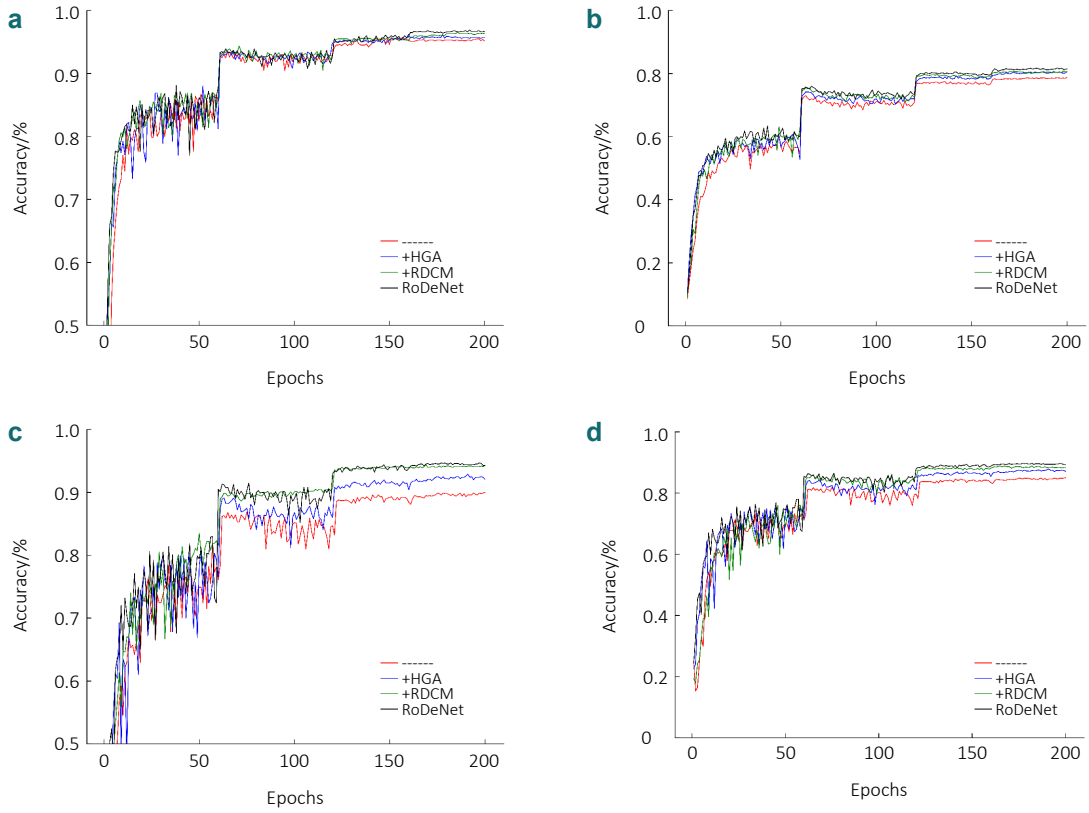


图 10 准确率折线图。(a) CIFAR-10；(b) CIFAR-100；(c) Imagenette；(d) Imagewoof

Fig. 10 Accuracy line chart. (a) CIFAR-10; (b) CIFAR-100; (c) Imagenette; (d) Imagewoof

表 8 消融实验结果

Table 8 Ablation study results

Network	CIFAR-10/%	CIFAR-100/%	Imagenette/%	Imagewoof/%
-----	96.29	78.8	89.98	84.6
+HGA	96.86	80.37	92.86	85.26
+RDCM	96.71	80.84	94.42	88.97
RoDeNet	96.87	81.81	94.72	89.69

强，使模型既具备方向敏感性又保持特征选择灵活性。

3.7 可视化分析

为了进一步观察 RoDeNet 对关键特征识别的效果，本节实验选取 DenseNet、ResNet-34 和 WideResnet-28-10 三个网络模型与 RoDeNet 进行可视化分析对比。实验数据图为 Imagenette 数据集上随机选取的五张不同类别图片，如图 11(a) 所示，各网络模型可视化对比结果如图 11 所示。

从可视化分析结果中的四组对比图可看出，DenseNet 和 ResNet-34 的热力分布存在明显的注意力分散现象，特别是在处理复杂场景时，热力图呈现多区域泛化激活模式，表明这些网络难以准确定位关键判别特征。WideResnet-28-10 虽然展现出相对集中的激活区域，但仍存在背景噪声干扰和主体边缘模糊的问题。相比之下，RoDeNet 能自适应调整不同层级特征的贡献权重的热力图具体精准定位特性，在动物样本中精确聚焦于头部关键特征，在建筑场景中完整捕捉结构轮廓，在复杂景物中能够有效抑制背景干扰。

4 结 论

针对固定几何结构的卷积核在应对旋转、缩放等几何变换时特征提取能力不足以及传统网络难以有效区分重要特征与背景噪声的问题，本文提出的旋转可变形卷积的图像分类网络，通过引入旋转可变形卷积模块结合

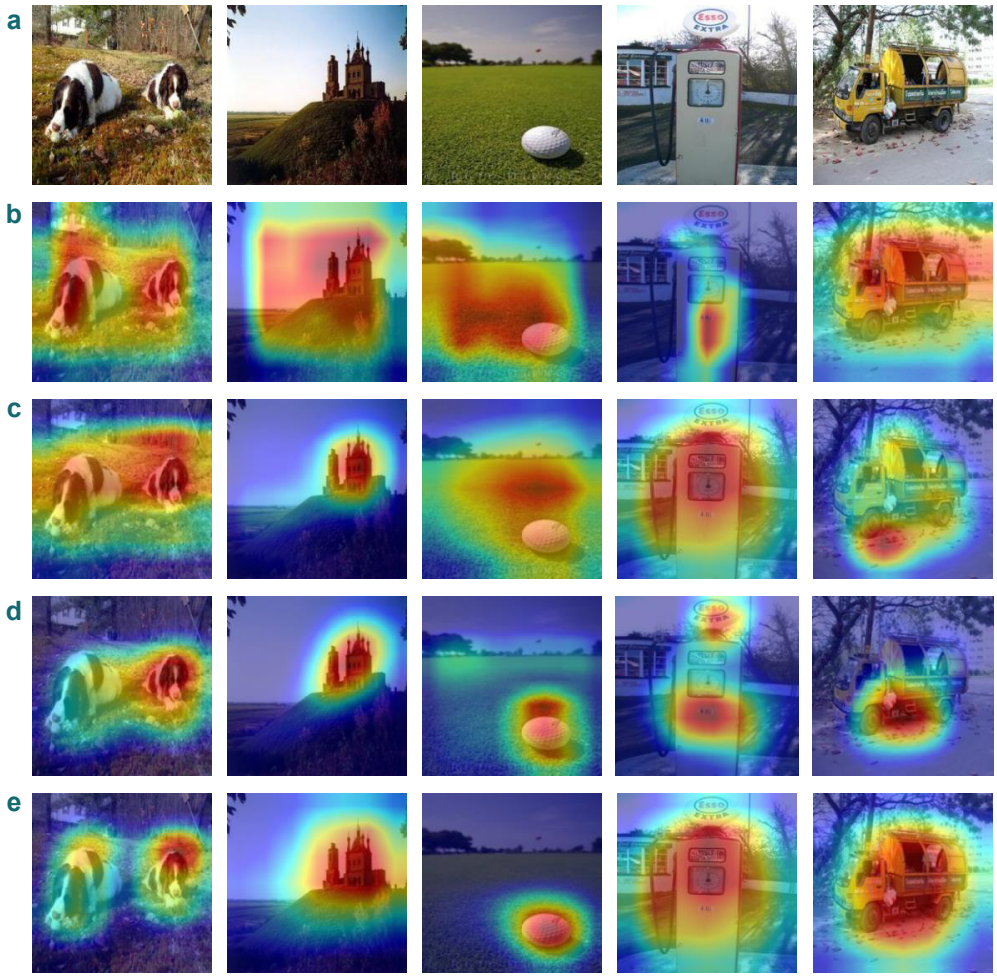


图 11 各网络模型可视化对比图。(a) 原始图像；(b) DenseNet；(c) ResNet-34；(d) WideResnet-28-10；(e) RoDeNet

Fig. 11 Comparison of visualizations of various network models. (a) Original image; (b) DenseNet; (c) ResNet-34; (d) WideResnet-28-10; (e) RoDeNet

特征保留模块，使卷积核具备方向敏感的特征提取能力，并且在增强特征判别力的同时避免了信息损失。并且提出 HGA 模块，通过分组计算和多尺度特征融合，实现了对重要特征的增强和对噪声的抑制，提升了特征选择的精准度。实验验证了 RoDeNet 的有效性和优越性，相比于基线网络模型 ResNet-34，在 CIFAR-10、CIFAR-100、Imagenette 和 Imagewoof 数据集上的准确率分别提高了 0.58%、3.01%、4.74% 和 5.09%，结果充分证明了 RoDeNet 在几何变换条件下的鲁棒性和特征提取能力。

本文提出的 RoDeNet 虽然在多个基准数据集上取得了显著的性能提升，但仍存在需要改进的问题。首先，在计算效率方面，旋转可变形卷积的偏移量计算和层次门控注意力的多分支结构引入了额外的计算开销，其次，当前采用的四个固定旋转角度 (0°、45°、90°、135°) 虽然能够覆盖主要的几何变换，但对于连续空间的所有可能变换和非刚性形变的适应能力有限。未来工作将会在计算效率优化方面和几何建模能力提升方面展开，有效解决当前网络存在的局限性，在保持几何自适应优势的同时，显著提升计算效率和泛化能力。

作者贡献

姜文涛、董金良：研究资料的收集、分析，实验的设计与完成，论文写作，指导论文选题、论文修改；张晟翀：指导论文选题、论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Liu Y, Pang Y L, Zhang W D, et al. Active learning based image classification technology: status and future[J]. *Acta Electron Sin*, 2023, **51**(10): 2960−2984.

刘颖, 庞羽良, 张伟东, 等. 基于主动学习的图像分类技术: 现状与未来[J]. 电子学报, 2023, **51**(10): 2960−2984.

2. Zhu L Z, Wang J Z, Lee W, et al. Incorporating simulated spatial context information improves the effectiveness of contrastive learning models[J]. *Patterns*, 2024, **5**(5): 100964.
3. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012: 1097–1105.
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]//*Proceedings of the 3rd International Conference on Learning Representations*, 2015.
5. He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778. https://doi.org/10.1109/CVPR.2016.90.
6. Zagoruyko S, Komodakis N. Wide residual networks[C]//*Proceedings of British Machine Vision Conference (BMVC)*, 2016.
7. Xie S N, Girshick R, Dollar P, et al. Aggregated residual transformations for deep neural networks[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 5987-5995. https://doi.org/10.1109/CVPR.2017.634.
8. Abdi M, Nahavandi S. Multi-residual networks: improving the speed and accuracy of residual networks[Z]. arXiv: 1609.05672, 2016. https://doi.org/10.48550/arXiv.1609.05672.
9. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2261–2269. https://doi.org/10.1109/CVPR.2017.243.
10. Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//*Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019: 6105–6114.
11. Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//*Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2017: 764–773. https://doi.org/10.1109/ICCV.2017.89.
12. Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 9300–9308. https://doi.org/10.1109/CVPR.2019.00953.
13. Wang W H, Dai J F, Chen Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 14408-14419. https://doi.org/10.1109/CVPR52729.2023.01385.
14. Woo S, Debnath S, Hu R H, et al. ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 16133–16142. https://doi.org/10.1109/CVPR52729.2023.01548.
15. Sun R, Tao L, Zhang W C, et al. TEC-Net: Vision Transformer embrace convolutional neural Networks for medical image segmentation[EB/OL]. arXiv:2306.04086, (2023-11-20) [2025-04-25]. https://doi.org/10.48550/arXiv.2306.04086.
16. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[C]//*Proceedings of the 3rd International Conference on Learning Representations*, 2015.
17. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010.
18. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.
19. Li W X, Guo Y C, Zheng J L, et al. SparseFormer: detecting objects in HRW shots via sparse vision Transformer. [C]// Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), New York, USA, 2024, 4851–4860. https://doi.org/10.1145/3664647.3681043.
20. Yuan J Y, Gao H Z, Dai D M, et al. Native sparse attention: hardware-aligned and natively trainable sparse attention[C]//*Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, 2025: 23078–23097. https://doi.org/10.18653/v1/2025.acl-long.1126.
21. Reza A, Leon N, Michael H, et al. Beyond self-attention: deformable large kernel attention for medical image segmentation[EB/OL]. arXiv: 2309.00121, (2023-08-31) [2025-04-25]. https://doi.org/10.48550/arXiv.2309.00121.
22. Ashtekar A. Black hole horizons and their mechanics[Z]. arXiv: 2308.08729, 2023. https://doi.org/10.48550/arXiv.2308.08729.
23. Worrall D E, Garbin S J, Turmukhambetov D, et al. Harmonic networks: deep translation and rotation equivariance[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 7168–7177. https://doi.org/10.1109/CVPR.2017.758.
24. Serre J P. *Linear Representations of Finite Groups*[M]. New York: Springer, 1977. https://doi.org/10.1007/978-1-4684-9458-7.
25. Choromanski K M, Likhosherstov V, Dohan D, et al. Rethinking attention with performers[C]//*Proceedings of the 9th International Conference on Learning Representations*, 2021.
26. Luo Z B, Sun Z T, Zhou W L, et al. Rethinking ResNets: improved stacking strategies with high order schemes[Z]. arXiv: 2103.15244, 2021. https://doi.org/10.48550/arXiv.2103.15244.
27. Daliparthi V S S A. ANDHRA Bandersnatch: training neural networks to predict parallel realities[Z]. arXiv: 2411.19213, 2023. https://doi.org/10.48550/arXiv.2411.19213.
28. Jiang W T, Chen C, Zhang S C. Sparse feature image classification network with spatial position correction[J]. *Opto-Electron Eng*, 2024, **51**(5): 240050.
姜文涛, 陈晨, 张晟翀. 空间位置矫正的稀疏特征图像分类网络[J]. 光电工程, 2024, **51**(5): 240050.
29. Yang Y T, Li L L, Liu X, et al. Multiscale and multidirectional transformer-based image recognition[J]. *Chin J Comput*, 2025, **48**(2): 249−265.
杨育婷, 李玲玲, 刘旭, 等. 基于多尺度-多方向 Transformer 的图像识别[J]. 计算机学报, 2025, **48**(2): 249−265.
30. Wu X D, Gao S Q, Zhang Z Y, et al. Auto- train-once: controller network guided automatic network pruning from scratch[C]//*Proceedings of IEEE/CVF International Conference on Computer Vision and Pattern Recognition*, 2024: 16163–16173. https://doi.org/10.1109/CVPR52733.2024.01530.
31. Lan H, Wang X H, Shen H, et al. Couplformer: rethinking vision transformer with coupling attention[C]//*Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023: 6464–6473. https://doi.org/10.1109/WACV56688.2023.00641.
32. Konstantinidis D, Papastratis I, Dimitropoulos K, et al. Multi-manifold attention for vision transformers[J]. *IEEE Access*, 2023, **11**: 123433−123444.

33. Zhou C L, Zhang H, Zhou Z K, et al. QKFormer: hierarchical spiking transformer using Q-K attention[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2024.
34. Tan J H. Pre-training of lightweight vision transformers on small datasets with minimally scaled images[Z]. arXiv: 2402.03752, 2024. https://doi.org/10.48550/arXiv.2402.03752.
35. Zhang T Y, Wan F, Duan H R, et al. FMDConv: fast multi-attention dynamic convolution via speed–accuracy trade-off[J]. *Knowl-Based Syst*, 2025, **317**: 113393.
36. Lu L M, Pang S C, Zheng X, et al. CIARD: cyclic iterative adversarial robustness distillation[C]//*Proceedings of 2025 IEEE/CVF International Conference on Computer Vision (ICCV 2025)*, 2025: 350–359.

作者简介

姜文涛 (1986-)，男，博士，副教授，硕士生导师，主要研究方向为图像与视觉信息计算。主持国防预研基金项目、辽宁省教育厅科学技术项目和辽宁省自然科学基金面上项目，发表中文学术论文 30 余篇，英文学术论文 5 篇。

E-mail：lntuwulue@163.com

【通信作者】董金良 (2000-), 男，硕士研究生，主要研究方向为深度学习与图像处理、模式识别与人工智能。

E-mail：1606592672@qq.com

Image classification network with rotating deformable convolution

Jiang Wentao[1], Dong Jinliang[1*], Zhang Shengchong[2]

[1] College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;
[2] Key Laboratory of Optoelectronic Information Control and Security Technology, Tianjin 300308, China

Abstract

Objective Image classification serves as a fundamental component in computer vision and supports a wide range of applications including object recognition, scene understanding, and intelligent perception systems. Deep convolutional neural networks have made substantial progress through hierarchical feature learning and residual architectures. Despite these advances, conventional convolution operations rely on fixed sampling grids and rigid kernel geometry, which limits their ability to adapt to geometric transformations such as rotation, scale variation, and orientation changes. When object orientation varies, features extracted by fixed convolution kernels become spatially misaligned, leading to inconsistent representations within the same category and reduced classification robustness. Deformable convolution partially alleviates this limitation by learning adaptive sampling offsets, yet fully data-driven offsets lack explicit geometric constraints and often result in offset divergence, unstable sampling patterns, and weak sensitivity to directional structures in deeper layers. In addition, existing convolution-based networks tend to assign similar importance to foreground and background features, which amplifies the influence of noise and reduces discriminative capability. The objective of this work was to construct an image classification network that simultaneously enhances rotation robustness and improves feature selection stability under complex geometric variations.

Methods A rotational deformable convolution-based image classification network, named RoDeNet, was developed on the basis of the ResNet-34 backbone. The core component of RoDeNet was the rotational deformable convolution module (RDCM), which integrated explicit geometric priors with adaptive offset learning. Instead of relying solely on unconstrained learned offsets, RDCM introduced predefined rotational offsets corresponding to four base orientations, namely 0°, 45°, 90°, and 135°. These offsets provided stable directional references for convolutional sampling. A compact offset prediction network was employed to generate dynamic offsets from feature responses. The predefined offsets and learned offsets were combined through element-wise addition, which constrained the deformation process within a geometrically meaningful range while preserving data-driven adaptability. RDCM adopted a dual-path structure. One path employed standard convolution to preserve stable base features and ensure consistent low-level representations. The other path applied rotational deformable convolution to capture orientation-adaptive features through direction-aware sampling. Feature

Foundation item: National Natural Science Foundation of China Fund (61601213), Natural Science Foundation of Liaoning Province Fund (20170540426), and Key Project of Education Department of Liaoning Province Fund (LJYL04)

*Correspondence: Dong J L, E-mail: 1606592672@qq.com

outputs from both paths were concatenated along the channel dimension to form enriched representations. To prevent information loss during feature enhancement, a feature preserve module (FPM) was integrated after feature fusion. FPM consisted of an attention-guided enhancement branch and a feature retention branch. The attention branch amplified discriminative spatial responses, while the retention branch maintained the original structural information, ensuring balanced feature optimization. To further improve feature discrimination and noise suppression, a hierarchical gated attention (HGA) mechanism was incorporated into residual blocks. HGA divided feature channels into multiple groups and processed them through two cooperative branches. The spatial pooling branch extracted horizontal and vertical contextual information to model long-range spatial dependencies. The depthwise separable convolution branch focused on local spatial patterns with reduced computational cost. The outputs of the two branches were fused to generate adaptive attention weights, which were used to reweight feature responses across spatial and channel dimensions. In the overall architecture, RDCM replaced the initial convolution layer and the first residual block of each stage, while HGA was inserted after the second convolution within residual blocks to avoid interference with early geometric correction.

Results and Discussions Extensive experiments were conducted on CIFAR-10, CIFAR-100, Imagenette, and Imagewoof datasets to evaluate classification accuracy, robustness, and efficiency. RoDeNet achieved classification accuracies of 96.87% on CIFAR-10, 81.81% on CIFAR-100, 94.72% on Imagenette, and 89.69% on Imagewoof. Compared with the baseline ResNet-34, accuracy improvements of 0.58%, 3.01%, 4.74%, and 5.09% were achieved, respectively. The performance gains were more pronounced on Imagenette and Imagewoof, which contain significant orientation variation and background complexity, indicating improved geometric robustness. Parameter sensitivity experiments demonstrated that moderate predefined rotation offset strength produced the most favorable performance balance. Excessively strong geometric priors constrained adaptive learning, while purely data-driven offsets failed to provide stable rotational modeling. Ablation studies confirmed that RDCM and HGA each contributed independently to performance improvement. RDCM enhanced direction-sensitive feature extraction, while HGA improved discriminative feature selection and suppressed background noise. Their combination produced greater performance gains than either module alone, indicating complementary functionality. Computational analysis showed that RoDeNet introduced a moderate increase in parameter count and computational complexity compared with ResNet-34, while remaining significantly more efficient than many Transformer-based classification models. Visualization analysis further revealed that RoDeNet focused more accurately on object regions and structural contours, whereas baseline models exhibited dispersed or background-dominated activation patterns. These observations confirmed the effectiveness of the proposed geometric adaptation and hierarchical attention mechanisms.

Conclusions A rotational deformable convolution-based image classification network was presented to address the limitations of fixed convolutional structures under geometric transformations. By integrating predefined rotational priors with dynamically learned offsets, RDCM enabled stable and direction-aware adaptive sampling. The feature preserve module balanced feature enhancement and information retention. The hierarchical gated attention mechanism improved multi-level feature discrimination through grouped attention modeling. Experimental results demonstrate that the proposed network achieves superior classification accuracy and robustness under rotation and orientation variations while maintaining acceptable computational efficiency. The method provides an effective solution for rotation-sensitive image classification tasks. Future work will focus on reducing computational overhead and extending the framework to continuous rotation modeling and to scenarios involving non-rigid geometric deformations.

Keywords: image classification; residual networks; rotational deformable convolution; feature preservation; hierarchical gated attention

Jiang W T, Dong J L, Zhang S C. Image classification network with rotating deformable convolution[J]. *Opto-Electron Eng*, 2026, **53**(4): 250307; DOI: 10.12086/oee.2026.250307