基于双分支注意力机制的图像自动标注研究

张国有,崔永强
(太原科技大学 计算机科学与技术学院,山西 太原 030024)

摘 要:图像自动标注技术能够将图像低层视觉特征转化为人类理解的高层语义信息,增强图像的可理解性和可搜索性,在图像检索和图像分类领域具有重要的应用价值。目前,基于卷积神经网络模型的图像自动标注技术,仍存在浅层网络无法捕捉足够的特征信息、容易忽视标签之间的相互关系以及标注时难以确定标签数量的问题。该文提出的基于双分支注意力机制的图像自动标注模型,首先使用双分支注意力网络,增强图像特征和标签的相关性以及学习标签之间的相关性;其次在空间注意力分支增加多尺度特征提取模块,以提取图像的多尺度特征,解决浅层网络特征提取不充分的问题;再次通过融合模块,融合两个分支的输出,将图像特征进一步增强;最后通过标签数量预测模块,预测待标注图像的标签数量,进一步提高标注的准确性。该模型分别在三个基准数据集 Corel 5K、ESP Game 和 IAPR-TC-12 上进行实验分析,实验结果表明该模型可以有效解决上述问题,提高标注的有效性与准确性。

关键词:图像自动标注;卷积神经网络;多尺度特征;注意力机制;特征融合
中图分类号:TP311.51 **文献标识码**:A **文章编号**:1673-629X(2024)09-0167-07
doi:10.20165/j.cnki.ISSN1673-629X.2024.0172

Research on Automatic Image Annotation Based on Dual-branch Attention Mechanism

ZHANG Guo-you, CUI Yong-qiang
(School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: Automatic image annotation technology can transform low-level visual features of images into high-level semantic information understood by humans, enhancing the comprehensibility and searchability of images, and has important application value in the fields of image retrieval and classification. At present, automatic image annotation technology based on convolutional neural network models still faces problems such as shallow networks being unable to capture sufficient feature information, easily ignoring the interrelationships between labels, and difficulty in determining the number of labels during annotation. The proposed automatic image annotation method based on dual-branch attention mechanism first uses a dual-branch attention network to enhance the correlation between image features and labels, as well as learn the correlation between labels. Secondly, a multi scale feature extraction module is added to the spatial attention branch to extract multi scale features of the image, solving the problem of insufficient feature extraction in shallow networks. By fusing the outputs of the two branches again through the fusion module, the image features are further enhanced. Finally, the label quantity prediction module is used to predict the number of labels in the image to be annotated, further improving the accuracy of annotation. The proposed model was experimentally analyzed on three benchmark datasets, Corel 5K, ESP Game, and IAPR-TC-12. The experimental results showed that the proposed method can effectively solve the above problems and improve the effectiveness and accuracy of labeling.

Key words: automatic image annotation; convolutional neural network; multi scale feature; attention mechanism; feature fusion

0 引 言

随着智能手机及社交媒体的盛行,人们拍摄并分享图像日益增多。图像数据已成为文化资源数字化的最主要的方式。但对图像数据的应用能力未能同步提升。问题根源在于计算机难以从图像低层特征中提取人类理解的高层语义信息,即存在“语义鸿沟”。为弥合此鸿沟、提升图像应用能力,图像自动标注成为紧迫而关键的任务。图像自动标注是计算机程序自动为图像分配一个或多个标签以描述图像内容的过程。例如,一张猫玩耍的图像,程序通过分析内容,自动加上

收稿日期:2024-01-23 **修回日期**:2024-05-24
基金项目:山西省自然科学基金项目(202203021221145);国家自然科学基金项目(62072325);太原科技大学科技创新基金项目(20212039)
作者简介:张国有(1972-),男,副教授,硕导,博士,CCF 会员(80252M),研究方向为三维模型数字水印技术、深度学习;通讯作者:崔永强(1998-),男,硕士研究生,研究方向为深度学习与图像语义标注。

“猫”“玩耍”“宠物”等标签来表达图像的语义概念。近年来,卷积神经网络(CNN)构建的模型为图像自动标注提供了多种创新途径。常见方法是利用预训练模型进行图像特征提取,然后对提取的特征进行深入分析[1]。其次是改进模型本身,以提高模型的特征提取能力[2]。这些方法在提高图像自动标注的性能方面提供了有力的支持,拓宽了标注的应用领域。图像自动标注技术,可以减轻人工标注的负担,提高数据处理效率,降低标注成本,并为图像增加语义信息,使内容更易搜索和分类,进而推动图像搜索引擎、智能图像分类系统等领域发展,为用户提供更好的图像检索和分类体验。

1 相关研究

图像自动标注任务使用机器学习或深度学习模型对图像进行分析,并预测多个概念或标签,如物体类别、场景描述和图像特征等。这一任务面临着处理多个标签的语义关系和相关性的挑战,因此可以将其视为多标签分类问题。在图像自动标注领域,CNN 自 2012 年以来成为核心算法模型[3-4]。然而,浅层 CNN 网络提取的特征过于单一,在处理多个目标的情况下,可能会出现混淆或忽略其他目标的问题,而且对于小尺寸的目标,其信息可能分布在感受野的边缘,而浅层网络的感受野较小,容易导致小目标的特征被部分遗漏,这使得浅层网络在处理小尺寸目标时性能受限。这仍然是图像自动标注任务的一大难题。为此,Zhang 通过使用基于稀疏编码的空间金字塔匹配和深度卷积神经网络来建模图像特征[5],并进一步使用度量学习技术来结合这些特征,以实现特征图更有效地表示图像。同时使用标签转移机制,通过使用图像的类别信息,自动将那些有希望的标签推荐给每幅图像。Palekar 提出了一种基于混合深度学习(DL)的 AIA 优化图像标注模型[6],利用 Slantlet 变换、YCrCb 色彩空间和局部二值模式(LBP)提取图像的形状、颜色和纹理等特征,并通过特征选择来降低特征维数。最后,利用深度优化的卷积残差图像标注模型获得自适应图像标注。为了提高网络的特征表达能力和特征提取能力,毛静怡等人对 ResNet 网络进行改进[7],添加了多尺度模块增加感受野,深度模块降低背景噪声并关注病灶区域,增强模块捕捉不同尺度信息,提升分类精度。Adnan 提出一种改进的基于卷积神经网络和 Slantlet 变换的自动图像注释模型[8]。该模型结合基于 CNN 特征和邻居图像的多个特征(SLT, YCbCr, LBP),通过选择具有 Slantlet 变换的 CNN 来实现精确度和召回率之间的平衡。实现了灵活的注释和提高的准确性。为了提高医学放射图像标注的准确性,Li 采用 CNN 对医学辐射图像进行特征提取[9]。模型构建了图像的梯度信息分布模型,采用块模板匹配法将图像分割成大小不同的影像特征块,提取医学放射图像的多分辨率特征,使得模型能够提取更具价值的特征。Adnan 提出的增强型自动图像标注系统将多个标签映射到单个图像中,提供对视觉内容含义的深入理解[10]。将多种特征类型的信息合并到图像标注系统的一个新的特征向量表示中。同时通过具有高斯-拉普拉斯金字塔(DL-MCNN-GLP)的深度学习多重卷积神经网络,提高图像编码和注释的有效性。

此外,图像的自动标注涉及到选择适当的标签,而标签的数量可能是不确定的。标签的选取直接影响着标注的准确性和完整性。如果标签数量太少,可能无法充分描述图像的复杂内容;而如果标签数量太多,可能导致过拟合或模型难以训练。因此,解决这一问题需要考虑如何动态地确定标签数量,使其能够适应不同图像的特点,提高自动标注的鲁棒性和准确性。为此,赵爱迪使用 CNN-THOP[11] 模型进行图像标注。首先,利用 CNN 模型预测各类标签的概率,并同时进行最优阈值匹配,根据最优阈值对标签进行标注。Wei 提出一种结合卷积神经网络与 KNN 标签语义拓展模型[12]。通过 CNN 获得图像特征,然后使用基于 KNN 的标签预测获得预测标签;同时提出一种基于图像特征相似度的标签数量预测模型来预测标签的数量。王琳提出一种 CNN 和加权贝叶斯相结合的最近邻图像标注模型[13],使用贝叶斯后验概率公式计算候选标签与待标注图像视觉特征之间的概率值,得出候选标签的标注与待标注图像的概率。根据自定义权重结合贝叶斯优化概率值进行排序,获得新的标签集,以此来优化标注结果。文献[14]通过使用 AlexNet 网络提取的 CNN 特征收集近邻图像,再使用自定义的贝叶斯方法来预测或扩展语义标签。文献[15]提出了一个加权 KNN 的模型,结合了多标签线性判别方法来计算权重。同时,利用基于 KNN 的模型来获取测试图像在每个标签类别中的 k 近邻,并根据其近邻的贡献来获得图像的预测。

综上所述,基于 CNN 的图像自动标注仍存在的问题主要有:

(1)CNN 模型中,浅层网络的感受野较小,意味着每个神经元只能捕捉到图像中较小的局部特征。对于尺度较小的物体,无法捕捉到足够的特征信息,导致自动标注的准确性有所限制。

(2)现有模型通常过于聚焦于提取图像低层视觉特征与单一标签的关联,缺乏对标签之间相互关系的充分考虑。这导致在进行图像自动标注时,模型可能无法全面理解图像的语义内容。

(3)图像的自动标注,意味着标签数量的不确定,标签数量的选取直接影响着标注的准确性。

为了解决这些问题,该文提出了基于双分支注意力机制的图像自动标注模型。在深度学习模型优化中,空间注意力机制关注图像中不同位置,有助于捕获对象的空间结构,从而更好地处理图像中的物体遮挡和变形。通道注意力机制动态调整每个通道的重要性,提高模型对不同类别或属性的区分度,从而提高分类任务性能。通过并行使用这两种注意力机制,能够充分利用空间和通道信息,更好地理解数据的结构和特征。最终,通过预测标签数量,使得网络能够更准确地完成图像的自动标注。

2　基于双分支注意力机制的图像自动标模型

该文提出一种基于双分支注意力机制的图像自动标注模型(Convolutional Neural Network Based on Double Branch Attention Mechanism, CNN-DBAM)。该模型采用双分支注意力机制,不仅关注特征细节,还强化了标签之间的关系。同时,通过预测图像结构相似度来估计标签数量,从而提升标注性能。模型的提出旨在细致特征提取的同时,加强标签关联,提高自动标注的精确性。

CNN-DBAM 模型由以下几部分组成:

(1)双分支注意力网络,结合了空间和通道注意力机制,以关注图像特征与标签之间以及标签与标签之间的关联。空间注意力分支强调图像特征与标签的相关性,而通道注意力分支则着眼于标签之间的关联。通过并行运用这两种注意力机制,模型更好地学习图像的空间结构和通道关联性,从而获得更有效的特征表示;(2)多尺度特征提取模块,加强了模型的上下文感知和目标定位能力,有助于获得更丰富的图像特征。这使得模型能够在不同尺度上更全面地理解图像内容;(3)特征融合模块,将两个分支的特征图融合,筛选出更有价值的特征图,从而实现特征的增强。这有助于提升模型对图像信息的表达能力;(4)标签数量预测模块用于预测标签数量,提高标签预测准确度,减少遗漏和误标情况的发生。这有助于改善标注的精确性。完整的标注模型如图 1 所示。

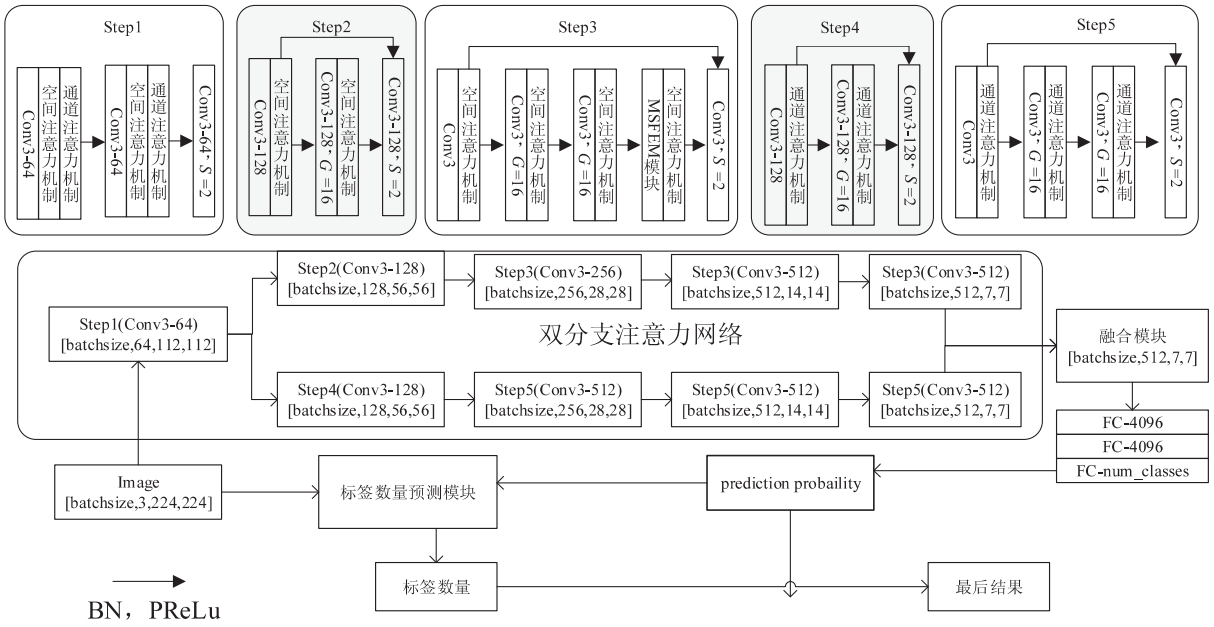


图 1　模型整体结构

2.1　双分支注意力网络

2.1.1　双分支注意力网络结构

双分支注意力网络采取不同的注意力机制,提取各自相关的特征信息。通过空间注意力机制增强图像与标签的相关性;通过通道注意力机制建模通道之间的相关性,进而学习标签之间的关联。在空间注意力分支,采用多尺度特征提取模块提取图像的多尺度特征,解决图像中物体尺度不一致的问题,同时使用残差结构减少特征重复。最后通过特征融合模块来选取两个分支中最适合的特征进行融合,提高图像特征和标签之间以及标签和标签之间的相关性。模型的整体网络结构如图 1 中双分支注意力网络部分所示。

2.1.2　空间注意力机制

空间注意力机制的整体结构如图 2 所示。输入的特征图 F 经过最大池化和平均池化得到两个 $1\times H\times W$ 的特征图,使用 Concat 操作进行特征拼接,再经过一个 7×7 的卷积核得到 1 通道的特征图。随后,通过 Sigmoid 激活函数得到空间注意力特征 M_s。最终,

将 M_s 与原特征图相乘,使其恢复为 $C \times H \times W$ 的大小,得到最终的输出 F_s。计算过程如公式 1 所示。

$$F\_s = M\_s(F) \times F = \delta(f7 \times 7([\text{AvgPool}(F); \text{MaxPool}(F)])) \times F \quad (1)$$

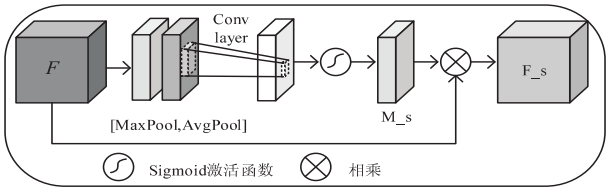


图 2 空间注意力结构

2.1.3 通道注意力机制

通道注意力机制的整体结构如图 3 所示。首先,将输入的特征图 F 经过全局最大池化和全局平均池化,将特征图的尺寸从 $C \times H \times W$ 压缩为 $C \times 1 \times 1$。再将两个特征图输入一个双层神经网络(Multilayer Perceptron,MLP),第一层将通道数压缩至原通道数的 1/4,第二层将通道数再扩展回原通道数。经过 ReLU 激活函数,得到两个激活后的特征图。将两个激活后的特征逐元素相加,并通过 Sigmoid 激活函数,得到通道注意力特征 M_c。最终,将 M_c 与原始特征图相乘,将其恢复为 $C \times H \times W$ 大小,得到最终的输出 F_c。计算过程如公式 2 所示。

$$F\_c = M\_c(F) \times F = \delta(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \times F \quad (2)$$

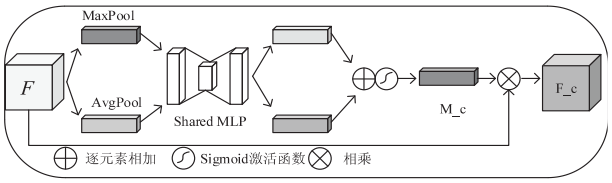


图 3 通道注意力结构

2.2 多尺度特征提取模块

多尺度特征的提取由多尺度特征提取模块(Multi Scale Feature Extraction Module,MSFEM)完成。整体结构如图 4 所示。

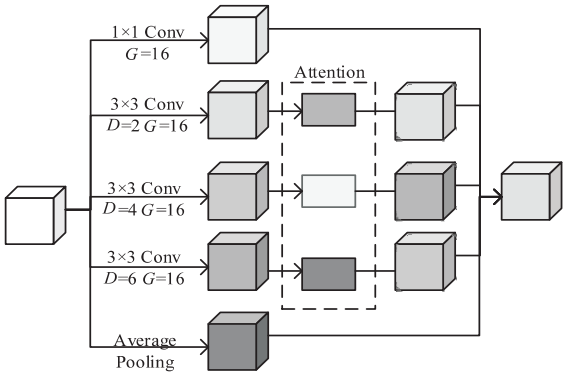


图 4 MSFEM 模块结构

该模型包含了一系列卷积层、注意力机制和特征融合操作。卷积层包含一个 1×1 的分组卷积,三个不同扩张率和填充方式的 3×3 分组卷积,以及一个自适应平均池化层;注意力机制层将 3×3 卷积的输出在通道维度上缩放为原来的 1/4,以学习不同尺度信息的重要特征;特征融合则是将 4 个卷积层与池化层的输出在通道维度上进行拼接。该模块通过不同大小的空洞卷积核,生成不同尺度的特征图,然后通过对每个尺度的特征图计算自注意力权重,从而提取多尺度特征并进行融合,得到更丰富的表示。

2.3 特征融合模块

两个分支提取后的特征图分别关注不同的特征信息,为了能够更好地利用这些信息,需要将不同分支的特征输入到改进融合模块中进行融合。融合模块如图 5 所示。

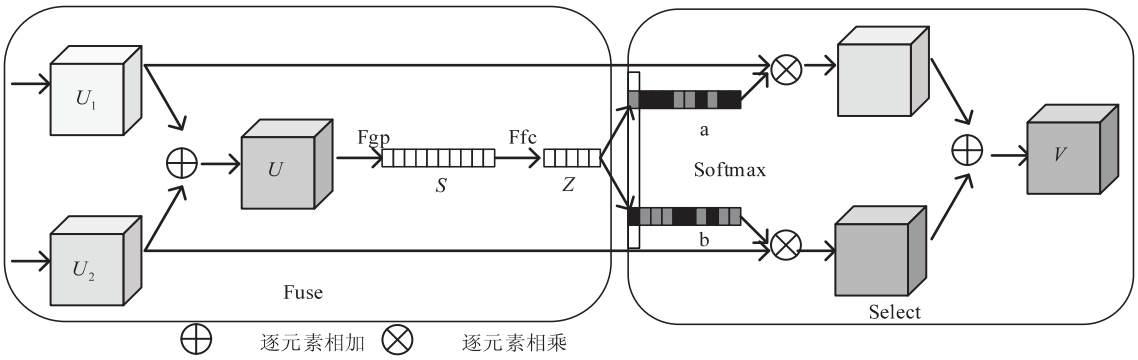


图 5 融合模块结构

融合模块在通道维度上增强关键特征、减少冗余特征,并针对不同层次的卷积特征的重要性进行特征融合,从而使不同分支提取的特征能够相互补充,而这一过程则由网络自动学习完成。该操作包含以下步骤:

(1)将两个分支的特征图 U_1、U_2 进行对应元素相加后,得到融合特征 U。

(2)对 U 在通道维度上做全局平均池化得到长度为 C 的全局信息特征 S,计算过程如公式 3 所示。

$$S = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c(i,j) \quad (3)$$

(3)特征 S 经过两个全连接层处理,首先,第一层

将特征 S 降维得到特征 Z;接着,第二层将特征 Z 升维,并通过 Softmax 函数激活,得到各自分支的注意力权重 a 和 b。具体计算如公式 4 和公式 5 所示。

$$Z = F_{fc}(s) = \mathrm{Relu}(B(WS)) \tag{4}$$

$$a = \frac{e^{W_a Z}}{e^{W_a Z} + e^{W_b Z}}, \quad b = \frac{e^{W_b Z}}{e^{W_a Z} + e^{W_b Z}} \tag{5}$$

其中, $W \in R^{d\times c}$ 表示第一层全连接层的参数; $W_a, W_b \in R^{d\times c}$ 表示第二层全连接层的参数; $a + b = 1$; B 为批正则化处理。

(4)最终的输出特征 V,是由不同分支的特征 U_1、U_2 与各自注意力权重 a 和 b 乘积的加和。具体计算如公式 6 所示。

$$V = (U_1 \times a) + (U_2 \times b) \tag{6}$$

2.4　标签数量预测模块

在图像自动标注过程中,标签数量的选取一直是一个重要问题。需要合理地选取标签数量。为此,提出了标签数量预测模块,将图像结构的相似度近似看为标签数量的相似度,通过多张图片与待预测图片的平均相似计算,近似求出待预测图片的标签数量。该操作包含以下步骤:

(1)通过双分支网络提取图像特征;

(2)统计每个图像,预测概率最高的 k 个标签;

(3)每个标签随机抽取 m 个图像计算图像相似度;

(4)根据图像结构相似度最终求得待标注图像的标签数量。

计算公式如公式 7 和公式 8 所示。

相似度计算:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7}$$

其中, μ_x , μ_y 分别表示图像 x,y 的均值; σ_x^2 , σ_y^2 分别表示图像 x,y 的方差; σ_{xy} 表示图像在 x 与 y 的协方差; $c_1 = (k_1 L)$, $c_2 = (k_2 L)$ 是用来维持稳定的常数, L 是像素值的动态范围, $k_1 = 0.01$, $k_2 = 0.03$。

数量预测:

$$\mathrm{number}(j) = \frac{\sum_{i=1}^{k\times m} \alpha^i \times \mathrm{SSIM}\ (j,i)^i}{\sum_{i=1}^{k\times m} \mathrm{SSIM}\ (j,i)^i} \tag{8}$$

其中, j 表示待预测图像; k 表示预测概率最高的 k 个标签; m 表示每个标签随机抽取 m 个图像; α^i 表示第 i 个图像的标签数;SSIM(j,i)表示图像 j 与图像 i 的结构相似度,且 SSIM 的值大于 0.1。

3　实验分析

为了验证 CNN-DBAM 模型的效性,将在三个基准数据集 Corel 5K、IAPR-TC-12、ESP Game[16] 上进行对比实验。

3.1　数据集与评价指标

该文使用了三个基准图像数据集:Corel 5K、IAPR-TC-12 和 ESP Game,按数据集自身划分的训练集与测试集进行模型训练,这些数据集通过网络进行下载获取。数据集的数据如表 1 所示。

表 1　数据集详细信息

数据集	Corel 5K	IAPR-TC-12	ESP Game
图像数	5 000	19 627	20 770
标签数	260	291	268
训练图像数	4 500	17 665	18 689
测试图像数	500	1 962	2 081
平均标签数	3.4	5.7	4.7
标签中值数	4	5	5
最大标签数	5	23	15
标签出现平均次数	58.6	347.7	326.7

实验使用评价指标:精确率(precision, P)、召回率(recall, R)以及 F1 值(F1),对实验结果进行评测。计算过程如公式 9 ~ 11 所示。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

$$\mathrm{F1} = \frac{2 \times P \times R}{P + R} \tag{11}$$

其中,TP(True Positive)表示被预测准确的正例,即预测为正类并且标签也为正类;FN(False Negative)表示被预测错误的反例,即预测为假类但标签为正类;FP(False Positive)表示被预测错误的正例,即标签为假类但预测为正类。P 计算了被正确预测的正类实例数量占所有被预测为正类实例数量的比例; R 表示被正确预测的正类实例数量占所有实际正类实例数量的比例;F1 表示 P 与 R 的调和平均数,其范围在 0 到 1 之间,F1 的数值越大,意味着模型效果越好。

3.2　参数分析

文中预测模块的参数 k 与 m 分别表示预测概率最高的 k 个标签以及每个标签随机抽取 m 个图像,在本节中,将对两个参数进行分析,以使得模型达到最好的效果。

为探索 k 值的变化对评价指标的影响,首先将 m 设定为 30。从图 6 可以看出,随着 k 值的增加,评价指标在 Corel 5K 数据集上呈现先上升后下降的趋势;在 ESP Game、IAPR-TC-12 数据集上先上升后趋于平稳。由图 6 可知,当 $k = 4$ 时,评价指标在三个数据集上整体最优。因此 k 的值选为 4。从图 7 可以看出,当

k =4 时,随着 m 值的增加,评价指标在 Corel 5K 数据集上呈现先上升后下降的趋势;在 ESP Game、IAPR-TC-12 数据集上先上升后趋于平稳。当 m =45 时,评价指标在 Corel 5K 数据集上达到最优;在 ESP Game、IAPR-TC-12 数据集上的变化逐渐稳定。因此 m 的值选为 45。

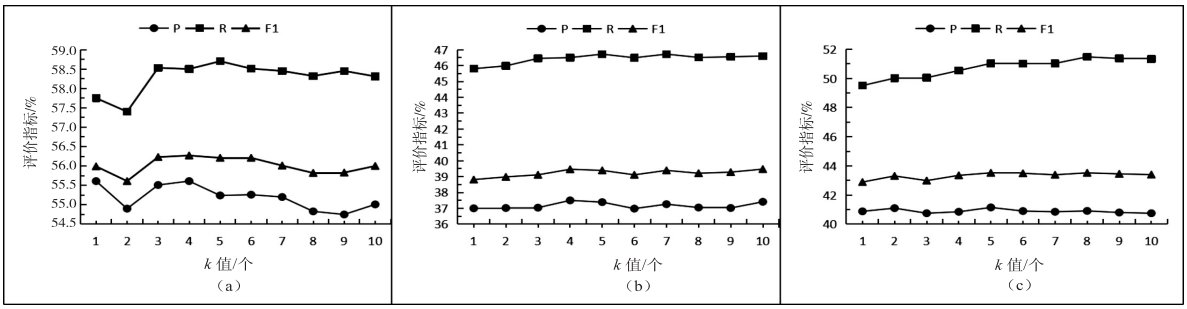


图 6 在 Corel 5K(a)、ESP Game(b)、IAPR-TC-12(c)数据集上评价指标随 k 值的变化曲线

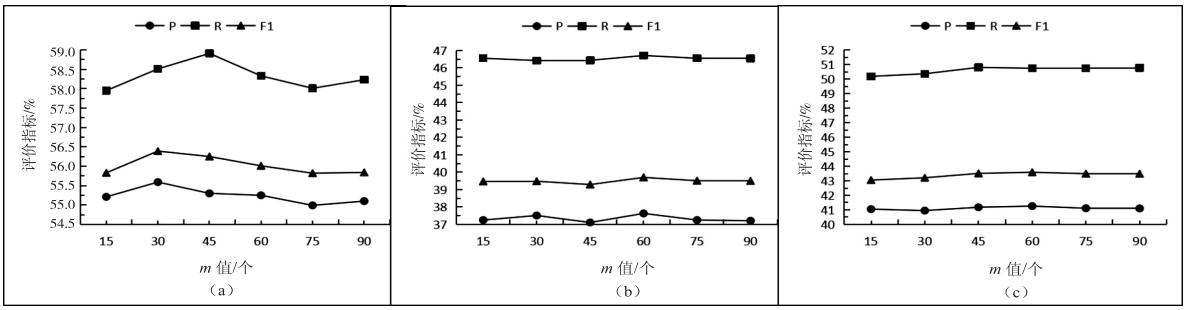


图 7 在 Corel 5K(a)、ESP Game(b)、IAPR-TC-12(c)数据集上评价指标随 m 值的变化曲线

3.3 对比实验

由于 CNN-DBAM 模型采用卷积神经网络来提取图像特征,将与目前较好的基于 CNN 的标注模型进行比较。

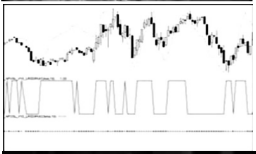
在 Corel 5K、ESP Game 和 IAPR-TC-12 三个数据集上,CNN-DBAM 模型与近几年的图像自动标注模型的比较结果如表 2 所示。

表 2 文中模型与其他模型在三个数据集上的对比结果

数据集	Corel 5K			ESP Game			IAPR-TC-12		
	P	*R*	F1	*P*	*R*	F1	*P*	*R*	F1
CNN+WB[13]	0.45	**0.62**	0.52	0.39	0.43	0.39	**0.46**	0.44	0.42
TSEL+LQP[12]	0.45	0.40	0.43	0.44	0.43	0.43	0.42	0.42	0.41
Weight+KNN[15]	0.22	0.15	0.18	**0.46**	0.22	0.30	0.42	0.17	0.24
SEM[14]	0.37	0.52	0.43	0.38	0.42	0.40	0.41	0.39	0.40
CNN-ST[8]	0.40	0.55	0.42	0.38	0.46	**0.50**	0.42	0.41	0.39
CNN-ECC[5]	–	–	–	–	–	–	0.46	0.35	0.40
GAN[17]	0.46	0.47	0.46	–	–	–	0.56	0.38	0.43
PGCF[18]	0.28	0.27	0.27	0.34	0.18	0.24	0.28	0.27	0.27
CNN-DBAM	**0.58**	0.61	**0.59**	0.40	**0.49**	0.42	0.43	**0.53**	**0.46**

由表 2 可知,CNN-DBAM 模型与整体效果较好的 CNN-WB 模型相比,在 Corel 5K 数据集上, P 值提高了 13 百分点, R 值降低了 1 百分点,F1 值提高了 5 百分点。在 ESP Game 数据集上,CNN-DBAM 模型与精确率最高的 Weight+KNN 模型相比, P 值降低了 6 百分点, R 值提升了 29 百分点,F1 值提升了 12 百分点,与整体效果较好的 CNN-WB 相比, P 值提升了 1 百分点, R 值提高了 6 百分点,F1 值提高了 3 百分点。在 IAPR-TC-12 数据集上,CNN-DBAM 模型与精确率最高的 CNN-ECC 模型相比, P 值降低了 3 百分点, R 值提升了 18 百分点,F1 值提升了 6 百分点,与整体效果较好的 CNN-WB 相比, P 值降低了 3 百分点, R 值提高了 9 百分点,F1 值提高了 4 百分点。这是由于在提取多尺度特征的同时,还借助通道注意力机制捕捉标签之间的相互关系,获取更加完善的特征信息的同时,初步考虑标签的相互关系,使得模型整体效果提升。表 3 表示在 Corel 5K、ESP Game 和 IAPR-TC-12 三个数据集上抽样的预测结果,深色为预测正确标签。

表 3　三个数据集上随机抽样的预测效果

图片	原始标签	预测标签
	forest cat tiger bengal	forest cat tiger bengal
	black chart graph line white	black chart colors graph line
	door frame glass round table tee-shirt woman	man plate table wall woman

由上述分析及表 3 的预测结果可以看出,提出的 CNN-DBAM 模型在实践中表现出色。在 Corel 5K 数据集上,CNN-DBAM 模型整体上表现更优,虽然召回率略有降低,但准确率和 F1 值均超过对比模型;在 ESP Game 和 IAPR-TC-12 数据集上,CNN-DBAM 模型同样整体优于对比模型,尽管查准率有所下降,但召回率和 F1 值均高于对比模型。

4　结束语

该文提出了一种基于双分支注意力机制的图像自动标注模型,通过双分支注意力网络提取图像多尺度特征来获取更具表达力的图像特征,同时依靠注意力机制关注特征与标签以及标签与标签之间的相互关系,从而得到更准确的预测标签。最后通过预测模块,预测标签数量来提高标签标注的准确性。该模型分别在 Corel 5K、ESP Game 和 IAPR-TC-12 三个数据集上与近几年的先进模型进行实验对比,结果表明该模型可以有效提高标注的有效性及准确性。

参考文献:

[1] 顾广华,曹宇尧,崔　冬,等. 基于形式概念分析和语义关联规则的目标图像标注[J]. 自动化学报,2020,46(4):767-781.

[2] HELMY T,DJATMIKO F. Framework for automatic semantic annotation of images based on image's low-level features and surrounding text[J]. Arabian Journal for Science and Engineering,2023,48(2):1991-2007.

[3] BENSACI R, KHALDI B, AIADI O, et al. Deep convolutional neural network with KNN regression for automatic image annotation[J]. Applied Sciences,2021,11(21):10176.

[4] PARALIC M, ZELENAK K, KAMENCAY P, et al. Automatic approach for brain aneurysm detection using convolutional neural networks[J]. Applied Sciences,2023,13(24):13313.

[5] ZHANG Weifeng, HU Hua, HU Haiyang, et al. Automatic image annotation via category labels[J]. Multimedia Tools and Applications,2020,79(17-18):11421-11435.

[6] PALEKAR V, KUMAR L S. Adaptive optimized residual convolutional image annotation model with bionic feature selection model[J]. Computer Standards & Interfaces, 2024, 87:103780.

[7] 毛静怡,宋余庆,刘　哲. 多尺度深度特征提取的肝脏肿瘤 CT 图像分类[J]. 中国图象图形学报,2021,26(7):1704-1715.

[8] ADNAN M M, RAHIM M S M, KHAN A R, et al. An improved automatic image annotation approach using convolutional neural network-Slantlet transform[J]. IEEE Access, 2022,10:7520-7532.

[9] LI Xiaofeng, WANG Yanwei, CAI Yinjie. Automatic annotation algorithm of medical radiological images using convolutional neural network[J]. Pattern Recognition Letters,2021, 152(9):158-165.

[10] ADNAN M M, KURDI W H M, ALOTAIBI S, et al. Image annotation with YCbCr color features based on multiple deep CNN-GLP[J]. IEEE Access,2024,12:11340-11353.

[11] 曹建芳,赵爱迪,张自邦. 融合阈值寻优的卷积神经网络在图像标注中的应用[J]. 计算机应用,2020,40(6):1587-1592.

[12] WEI Wei, WU Qiong, CHEN Deng, et al. Automatic image annotation based on an improved nearest neighbor technique with tag semantic extension model[J]. Procedia Computer Science,2021,183:616-623.

[13] 王　琳,张素兰,杨海峰. 基于 CNN 和加权贝叶斯的最近邻图像标注方法[J]. 计算机技术与发展,2021,31(10):63-69.

[14] MA Yanchun, LIU Yongjiang, XIE Qing, et al. CNN-feature based automatic image annotation method[J]. Multimedia Tools & Applications,2019,78(3):3767-3780.

[15] MA Yanchun, XIE Qing, LIU Yongjie, et al. A weighted KNN-based automatic image annotation method[J]. Neural Computing and Applications,2020,32(11):6559-6570.

[16] SALAR A, AHMADI A. Improving loss function for deep convolutional neural network applied in automatic image annotation[J]. The Visual Computer, 2023, 40(3):1617-1629.

[17] LIU Jian, WU Weisheng. Automatic image annotation using improved wasserstein generative adversarial networks[J]. IAENG International Journal of Computer Science,2021,48(3 Pt.1):507-513.

[18] WANG Mengke, LIU Yan, LIU Weifeng, et al. Feature fusion based parallel graph convolutional neural network for image annotation[J/OL]. Neural Processing Letters: 2023(9). https://doi.org/10.1007/s11063-022-11131-x.