

基于视觉注意力和 FCA 的古建筑图像语义完备

牛少刚,张素兰,张继福

(太原科技大学 计算机科学与技术学院,山西 太原 030024)

摘要:准确完备的古建筑图像语义不仅可提高古建筑图像检索效率,且能有效反映古建筑的历史文化信息。针对不同古建筑图像轮廓特征明显不同且建筑语义互相关联,为有效丰富古建筑图像语义,提出一种基于视觉注意力机制和形式概念分析(Formal Concept Analysis, FCA)的古建筑图像语义完备方法。首先使用注意力算子网络和 VGG16 网络模型生成待标注古建筑图像注意力图,并通过 softmax 分类器进行分类,获取图像初始标签集;其次构造基于待标注图像初始标签及其近邻标签的概念格;然后,利用概念格上下文分析语义的特点,通过概念节点之间的相似度度量,获取待标注图像潜在的语义标签。最后,在古建筑图像数据集上进行实验,结果验证了该方法能够有效地提高古建筑图像标注精度,丰富古建筑图像语义。

关键词:古建筑图像;标签完备;卷积神经网络;视觉注意力机制;形式概念分析

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)09-0214-07

doi:10.3969/j.issn.1673-629X.2022.09.033

Semantic Completion Annotation of Ancient Architecture Based on Visual Attention Mechanism and FCA

NIU Shao-gang, ZHANG Su-lan, ZHANG Ji-fu

(School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: Accurate and complete ancient architectural image semantics can improve the efficiency of image retrieval and effectively reflect the historical and cultural information. Aiming at the problem that the contour features of different ancient architectural images are different and the architectural semantics are interrelated, we propose a semantic completion method of ancient architectural images based on visual attention mechanism and FCA (Formal Concept Analysis), which effectively enriches image semantics. Firstly, the attention operator network and VGG16 network model are used to generate the attention map of ancient architecture, and the initial label set of the image is obtained by softmax classifier. Secondly, we construct a concept lattice based on the initial label of the image to be labeled and its neighbor labels. Then we use the context of the concept lattice to analyze the semantic characteristics and obtain the latent semantic labels of the image to be labeled by measuring the similarity between the concept nodes. Finally, experiments on the ancient architecture image datasets show that the proposed method can effectively complete the semantics of the ancient architecture image and improve the accuracy of image annotation.

Key words: ancient architecture images; tag completion; convolutional neural networks; visual attention mechanism; formal concept analysis

0 引言

中国古建筑作为中国的艺术文化瑰宝,蕴含着强烈的人文气息和文化气息^[1-3]。随着互联网古建筑图像资源的快速增长,用户可以通过网络便捷地查找古建筑信息。但在检索古建筑图像时,由于古建筑图像缺乏相应语义标签或存在错误标签,检索到的图像往往难以满足用户需求。因此,对古建筑图像进行准确

完备的标注,不仅能提高古建筑图像的检索效率,而且可提高人们对古建筑鉴赏的文化修养。

图像特征的有效提取是提高完备标注精度的关键,而传统的基于颜色、形状、纹理等图像特征提取^[4-5]方法无法提取图像的一些深层特征,且不能有效地对古建筑轮廓边界等重要特征进行语义分析。近年来,深度学习中基于卷积神经网络(Convolutional

收稿日期:2021-10-13

修回日期:2022-02-23

基金项目:山西省平台基地专项(201805D131007)

作者简介:牛少刚(1996-),男,硕士研究生,研究方向为数据挖掘与图像处理;通讯作者:张素兰(1971-),女,博士,教授,研究方向为概念格、数据挖掘与图像处理;张继福(1963-),男,博士,教授,研究方向为机器学习、数据挖掘与高性能计算。

Neural Networks, CNN) 提取图像特征的方法,因容易建立图像显著特征与语义之间的映射关系,得到了研究者的关注^[6-8]。但由于古建筑图像结构复杂、特征分布不均匀,影响古建筑图像 CNN 模型分类效果,而视觉注意力机制是深度学习中的一种资源分配方式,可将学习到的权重分配到提取的特征上面,权重的大小表示了特征的显著性程度。该方法因能在提取重要特征时,抑制一些非重要的信息,在各种视觉识别任务中得到很好的应用^[9-15]。其中,林泓等人^[13]在无监督图像翻译中融合注意力机制,加强图像生成过程中像素间远近距离的关联关系,解决了图像无关区域导致图像翻译细节模糊以及真实性低等问题。Leng 等人^[14]根据视觉注意力的思想,提出了上下文感知的注意力网络用于图像识别,通过注意力模块进行注意转移,定位注意区域来完成精确的图像识别。汤文兵等人^[15]通过视觉注意力机制实现用户多重交互,再结合卷积网络获取更高级别的特征交互,有效提升了物品识别效率。因此,为有效提取古建筑图像轮廓边界等重要特征,该文通过 CNN 结合视觉注意力机制,建立古建筑图像轮廓特征与语义信息之间的映射关系并获取古建筑图像初始语义。

古建筑图像的完备标注,主要通过计算图像视觉特征与语义标签,以及语义标签之间的关联度,来选择适合的标签作为标注结果,而完备标注效果的好坏关键在于古建筑图像视觉内容与语义标签,语义标签之间关联关系的有效建立。概念格又称形式概念分析 (Formal Concept Analysis, FCA)^[16],是一种数据分析和知识提取的有力工具,因其能有效地描述对象与属性之间的关系,反映概念之间泛化与特化的关系,已在

知识表示、数据挖掘、图像标注等众多领域得到了广泛的应用^[17-20]。其中,Hao 等人^[17]构造了能够动态处理上下文的三支概念格,提升了构造效率并缩短了构造时间,能够更有效地进行知识规则提取。Zou 等人^[19]提出一种模糊分层概念格的规则提取方法,通过设置不同的置信度构造概念算子来获取规则之间的关联性。顾广华等人^[20]通过构造基于图像簇的概念格,去除噪声图像块并获取目标语义簇,大大提升了图像标注的准确率。由于古建筑结构复杂、形式多变,其结构语义之间又存在一定联系,为进一步完备古建筑图像语义,该文利用 FCA 具有上下文语义分析的特点,建立了一种基于 FCA 的古建筑图像标签完备方法。该方法利用概念格分析古建筑构件之间隐含的关联关系,通过概念相关度计算,有效提取出古建筑图像的潜在语义,完备了古建筑图像语义。

1 文中方法

在利用 CNN 提取图像特征时,卷积层深度的增加使 CNN 所提取的深层特征更具有语义判别性。但在用 CNN 对古建筑图像进行特征提取时,图像特征与语义标签之间的映射较为单一,无法得到全面的图像语义信息,而概念格有层次化分析语义的特点,可以有效对古建筑图像进行语义补充。为尽可能地描述古建筑的屋顶风格、屋顶类型等深层语义,该文利用视觉注意力机制和卷积神经网络结合的方法得到初始的代表性标签,然后通过概念格节点相关度计算,获取更深层的古建筑语义标签(例如屋顶风格重檐、建筑结构木结构等),进而丰富古建筑图像的语义信息,该方法框架如图 1 所示。

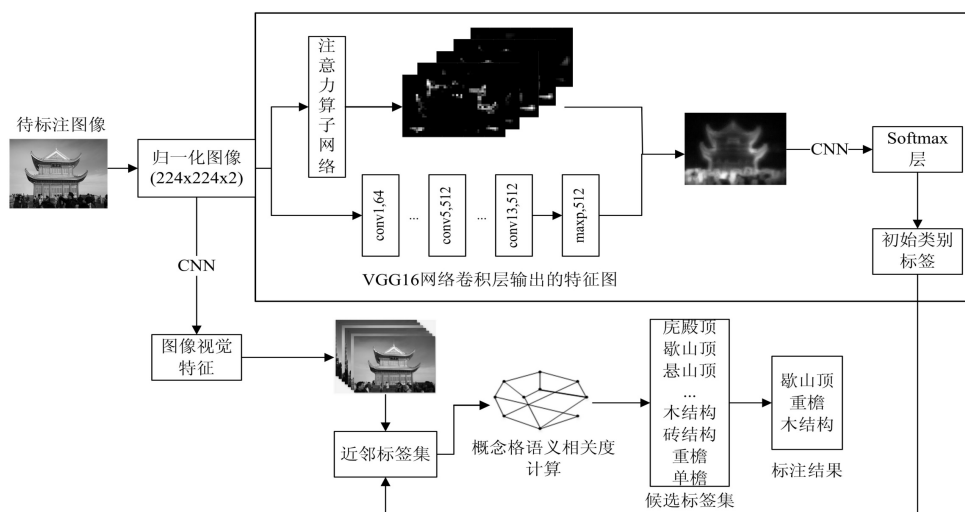


图1 基于注意力机制和概念格语义扩展的古建筑图像语义完备标注模型

1.1 基于 CNN 的古建筑注意力图的获取及初始类别标签生成

通过注意力算子网络和 VGG16 网络模型卷积层

提取待标注古建筑图像的特征,得到尺寸和通道数不变的权重图和特征图,然后将权重图与特征图进行融合获取能够表述古建筑图像特征的注意力图,最后,通

过 CNN 结构中的 softmax 层获取古建筑类别标签。其中注意力机制可以表示为:

$$f^{\text{att}} = N(I) \quad (1)$$

$$F_s^{\text{att}} = f^{\text{att}} \otimes F \quad (2)$$

其中, $N(\cdot)$ 表示注意力算子网络^[9], I 表示输入图像矩阵, f^{att} 表示经 $N(\cdot)$ 运算后输出的权重图矩阵, F 是输入图像 I 的特征矩阵, F_s^{att} 则表示权重图和特征图对位相乘得到的注意力显著图, \otimes 表示数组元素对位相乘。为能更有效地获得古建筑图像的初始标签, 通过 VGG-Net^[21] 模型对古建筑图像数据进行训练, 采用 VGG-Net 中性能更好的 VGG16 模型来提取古建筑图像的特征。VGG16 总共有 16 层隐藏层, 包含 13 个卷积层和 3 个全连接层, 包含 1 个输入层和 1 个输出层, 均使用相同大小的卷积核 (3×3) 和最大池化 (2×2), 输入图像均为 224×224 。为增加古建筑图像数据量, 对古建筑数据进行了旋转、尺度缩放处理和均值化处理, 并且对输入古建筑图像进行尺度 (224×224) 和灰度归一化处理, 以提高类内标注精度。

1.2 近邻图像的生成及近邻标签生成

为得到待标注图像 I_0 的初始近邻集及其近邻标签, 首先利用 1.1 节中已训练好的 VGG16 模型提取图像 4 096 维特征向量, 计算待标注图像 I_0 与训练图像的视觉相似性得分并选取相似性得分高的图像集作为近邻集。该文使用欧氏距离计算图像的视觉相似性, 距离越小说明二者越相似。计算公式如下:

$$S_i(I_i, I_j) = \exp\left(\frac{-\text{dist}(I_i, I_j)}{\delta}\right) \quad (3)$$

其中, $\text{dist}(I_i, I_j)$ 为图像 I_i 和 I_j 的视觉特征向量的欧氏距离, δ 是一个标量, 根据经验, 将图像距离矩阵的中位数设置为其值。通过图像相似性得分的降序排序, 获取了待标注图像的图像近邻集标签, 并将其与 1.1 节中的初始类别标签进行融合作为最终的近邻标签。

1.3 基于概念格的语义扩展

1.3.1 形式背景的构造

概念格是一种有效的语义层次分析工具。为利用概念格进行图像标签语义相关性分析, 该文将待标注图像与其近邻图像集作为对象集 U , 将待标注图像的初始标签和近邻标签合并形成新的标签集作为属性集 M , 并将近邻图像与标签映射关系进行归一化处理, 即存在映射关系记为 1, 反之记为 0, 从而构造出基于待标注图像及近邻图像集标签相关的形式背景。

假设待标注图像为 I_0 , 其经过 softmax 分类器获得的典型的初始预测标签为 t_1 , 由 1.2 节中搜索近邻图像得到 m 幅 (假设 $m=3$) 与 I_0 最相似的近邻图像 $I_1 - I_3$, 获取图像 I_0 近邻图像的所有标签并入集合 T 中, 则 $I_m = \{I_1, I_2, I_3\}$, $T = \{\text{“WuDianDing”}, \text{“XieShanDing”},$

“XuanShanDing”, “YingShanDing”, “CuanJianDing”, “DoubleEaves”, “SingleEaves”, “WoodStructure”, “BrickStructure” $\}$ 。参考文献[18], 根据近邻图像与标签之间映射关系构造形式背景 $K = (U, M, R)$, 如表 1 所示。

表 1 形式背景 K

近邻图像	语义标签								
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
I_0	1	0	0	0	0	1	0	1	1
I_1	0	0	1	0	0	0	1	1	0
I_2	0	1	0	0	0	1	0	1	0
I_3	0	0	0	1	0	0	1	0	1

根据形式背景 K 生成概念格结构 Hasse 图, 如图 2 所示。

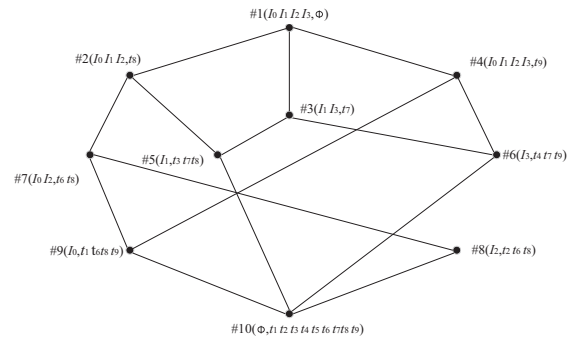


图 2 形式背景 K 的 Hasse 图

1.3.2 基于概念格的古建筑语义扩展

古建筑图像往往包含多种语义信息, 各语义信息之间又具有一定关联。为利用概念格进行图像标签语义相关性分析, 该文采用粗糙概念格 (Rough Concept Lattice, RCL)^[22-23] 相似性模型来度量两个概念节点之间的相似性。与其他相似性算法相比, RCL 相似性算法结构性更加清晰, 能有效地挖掘出标签之间的相似性, 可以更方便地度量形式背景中上下文概念的相似性^[23]。因此, 利用 RCL 来处理古建筑语义之间的关联问题。概念格中任意两个节点 a 和 b 的相似性度量方法如下:

$$\text{sim}(a, b) =$$

$$\frac{|(a \vee b)_{\text{LA}}|}{|(a \vee b)_{\text{LA}}| + \alpha |a_{\text{LA}} - b_{\text{LA}}| + (1 - \alpha) |b_{\text{LA}} - a_{\text{LA}}|} \quad (4)$$

其中, $\text{sim}(a, b)$ 的值介于 0 和 1 之间, 1 表示完全映射, 0 表示错误映射; $a \vee b$ 表示 a 和 b 的交集; $||$ 表示集合的基数; a_{LA} 为下近似属性集, 即包含概念 a 最大可定义集; $a_{\text{LA}} - b_{\text{LA}}$ 表示属性集属于 a_{LA} , 不属于 b_{LA} ; α 为相似性度量阈值, 其值设置为 0.5, 当相似性度量值大于等于 0.5 时, 可以认为两个概念节点之间存在映射关系, 当 α 小于阈值 0.5 时, 则节点之间不存在关

联。在图2中,对节点#5、#6进行相似性度量,节点#3是其连接节点且与#5、#6存在上下位关系。

$$\begin{aligned} 5_{LA} &= \{\{1,2\}, \{1,3\}\} & 6_{LA} &= \{\{1,3\}, \{1,4\}\} \\ 3_{LA} &= \{\{1,3\}\} & 5_{LA} - 6_{LA} &= \{\{1,2\}\} \\ 6_{LA} - 5_{LA} &= \{\{1,4\}\} \\ \text{sim}(5,6) &= \frac{1}{1 + 0.5 \times 1 + (1 - 0.5) \times 1} = 0.5 \end{aligned}$$

通过计算概念之间的相似性度量可以得到,节点5和6之间的相似性度量阈值为0.5,说明#5和#6可以相互映射。由式(3)计算得到待标注图像与近邻图像的相似度,将其作为近邻图像 I_m 对待标注图像的支持度指标之一,根据概念格获得的概念之间的相似度,计算候选标签集中每个标签对待标注图像的支持度:

$$\text{sup}(t_i, I_i) = \beta S_v(I_i, I_m) + (1 - \beta) \sum_{m=1}^m \text{sim}(I_i, I_m) \gamma(I_m, t_i) \quad (5)$$

其中, β 为相似性权重系数, $\gamma(I_m, t_i)$ 为近邻图像 I_m 与标签 t_i 之间的所属关系,即若 I_m 存在标签 t_i , $\gamma(I_m, t_i)$ 的值为1,反之为0。为减少不相关的语义标签,在计算得到的概念之间的支持度后,选择支持度高于0.1的标签词,作为待标注图像的最终标签词。

例如,待标注图像 I_0 的初始标签为 t_1 ,结合1.2节中计算图像之间的相似度,可得 I_0 与近邻图像之间的相似性得分,选取得分较高的前三张图像 $I_1 - I_3$,将其标签 $\{t_2, t_3, t_4, t_6, t_7, t_8, t_9\}$ 赋予待标注图像。然后通过式(5)得到各个标签对待标注古建筑图像 I_0 的支持度,去掉了近邻标签中与待标注图像无关的标签 $\{t_2, t_3, t_4, t_7\}$ 后,得到待标注图像的最终标签 $\{t_1, t_6, t_8, t_9\}$ 。与传统的完备标注方法相比,基于概念格的语义扩展通过挖掘图像上下文潜在的语义关系,不仅扩展了图像标签,而且也去掉了近邻图像中出现次数多的冗余标签,使标注效果得到显著提升。

1.4 算法描述

文中方法的主要步骤如下:

输入:待标注的古建筑图像 I_0 ,待完备标签集 $L = \{L_1, L_2, \dots, L_n\}$ 。

输出:待标注图像完善后的标签集 L' 。

Step1:预处理。对古建筑数据进行归一化处理。

Setp2:图像特征提取。利用注意力机制和卷积神经网络获取古建筑数据集的抽象特征图。

Setp3:初始标签获取。将上一步得到的古建筑注意力特征图做softmax线性分类,得到古建筑图像的初始标签集 L' 。

Setp4:获取近邻图像及近邻标签。通过式(3)得到的待标注图像 I_0 的近邻图像和近邻标签,并入并更新标签集 L' 。

Step5:概念格的构造以及标签完善。利用近邻图像和标签集抽象形式背景并构造概念格,利用概念格中节点之间的映射关系,由式(5)得到标签对图像的支持度 $\text{sup}(a, b)$,选取对待标注图像支持度高的标签进行标注并更新标签集 L' 。

Step6:标签输出。输出待标注图像的最终标签集 L' 。

语义完备算法流程主要包括图像归一化、提取图像特征、获取初始标签、获取近邻图像集近邻标签以及概念格构造和概念格语义相似度计算等。流程如图3所示。

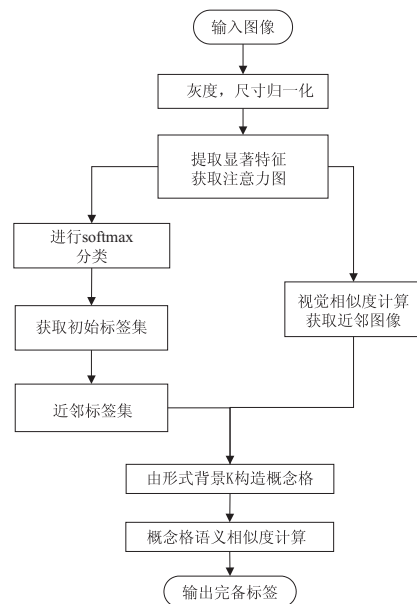


图3 语义完备算法流程框图

2 实验结果及分析

2.1 古建筑数据集及语义标签设计

通过网络收集了唐宋时期古建筑图像,构建了唐宋时期古建筑图像数据集。为验证文中方法的有效性,在唐宋时期古建筑数据集和中国科学院自动化研究所模式识别国家重点实验室提供的传统古建筑数据集(<http://vision.ia.ac.cn/data>)进行了实验。唐宋时期数据集和传统古建筑数据集的属性如表2所示。

表2 古建筑图像数据集属性

数据集	图像数	标签数	训练图像数	测试图像数
唐宋古建筑数据	500	9	450	50
传统古建筑数据	2 400	9	2 100	260

综合中国古建筑的特点,从建筑屋顶风格、建筑构件以及屋顶类型等方面设置语义标签。古建筑的屋顶主要有五种风格,分别为:庑殿顶(WuDianDing)、歇山顶(XieShanDing)、悬山顶(XuanShanDing)、硬山顶(YingShanDing)、攒尖顶(CuanJianDing);建筑构件主

要有木结构 (WoodStructure) 和砖结构 (BrickStructure); 屋檐形式分为单檐 (SingleEaves) 和重檐 (DoubleEaves) 两种形式。

2.2 评价标准

采用的评价标准^[20]有: 准确率 P 、召回率 R 以及 F 值, 分别定义为:

$$P = \frac{1}{M} \sum_{j=1}^M \frac{\text{Correct}(w_j)}{\text{Predicted}(w_j)} \times 100\% \quad (6)$$

$$R = \frac{1}{M} \sum_{j=1}^M \frac{\text{Correct}(w_j)}{\text{Truth}(w_j)} \times 100\% \quad (7)$$

$$F = \frac{2P \cdot R}{P + R} \times 100\% \quad (8)$$

其中, M 为待标注图像的总数目, w_j 为第 j 张待标注图像的生成的语义标签, $\text{Correct}(w_j)$ 为生成语义标签 w_j 中正确标签的数目, $\text{Predicted}(w_j)$ 为生成语义标签 w_j 的总标签数, $\text{Truth}(w_j)$ 为第 j 张待标注图像真值标签的数目。

2.3 注意力图对标注效果的影响及训练模型损失对比

为了确定生成的注意力特征图对古建筑图像的特征表示能力, 将经过注意力机制处理的古建筑数据集与原古建筑数据集进行了标注准确率实验对比。在训练网络时, 将训练集图像大小进行归一化, 输入尺寸为 224×224 , 使用动量为 0.9 的异步随机梯度下降, 网络初始的学习率设为 0.01, 每 10 步下降到当前学习率的十分之一, epoch 为 200, 损失函数为交叉熵损失函数, 为克服过拟合, 对全连接层中前两层丢失率为 50% 的进行删除操作。采用二分类模型常用指标 ROC-AUC 曲线图来评价所训练的网络模型。ROC 为受试者操作特征曲线, AUC 指曲线下方的面积, 其值大小范围在 0.5 ~ 1 之间, AUC 值越大, 说明训练模型性能越好。图 4(a) 和图 4(b) 分别给出了训练模型 VGG16 在使用了注意力机制和未使用注意力机制的唐宋古建筑数据集和传统古建筑数据集上的 ROC-AUC 曲线图。

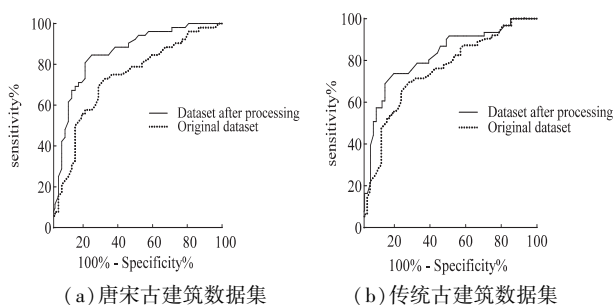


图 4 不同数据集 ROC-AUC 比较

由图 4 可以看到, VGG16 在传统古建筑数据集上模型的训练精度不如唐宋古建筑数据集, 这是因为在唐宋古建筑数据集构建过程中, 人工过滤掉了一些噪

声严重的数据。而传统古建筑数据集中存在与建筑无关的物体以及建筑物被遮挡等情况, 影响了模型的训练精度。从图 4 中可以观察到, 融合视觉注意力机制之后的古建筑数据在 VGG16 网络模型上有着更高的识别精度, 且在训练模型的 ROC 曲线上的 AUC 值分别达到了 85.12% 和 80.18%。说明该文的古建筑注意力特征图能够有效地增强古建筑的边界、轮廓等特征, 进而表述了古建筑的局部特征以及整体特征。图 5 为 VGG16 网络模型在两类数据集上的损失收敛曲线, 可以看到两类数据集的损失率都在 epoch 为 160 之后达到了最低。

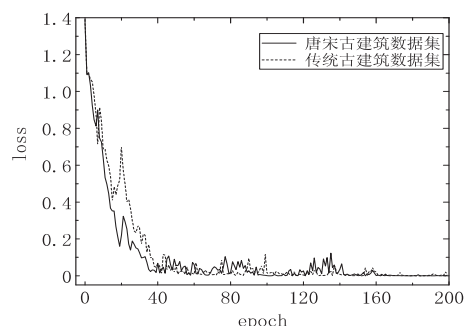


图 5 VGG16 训练网络在两类数据集上的损失曲线

2.4 参数分析

本节在唐宋古建筑数据集和传统古建筑数据集上通过实验进行了参数分析。参数 β 是图像相似度权重系数, β 值越小, 图像视觉相似度对标注的结果影响越小。通过寻找其最佳值, 获取图像标注的最佳结果, 用 F 值作为评价该参数的指标。图 6 是对参数 β 进行分析的结果, 从图中可以看出, 随着 β 的变化, 评估指标曲线波动程度较小, 总体呈先升后降的趋势, 并在 $\beta = 0.2$ 时达到最佳, 说明概念格语义相关度对标注的结果影响较大, 因此设置参数 β 的值为 0.2。

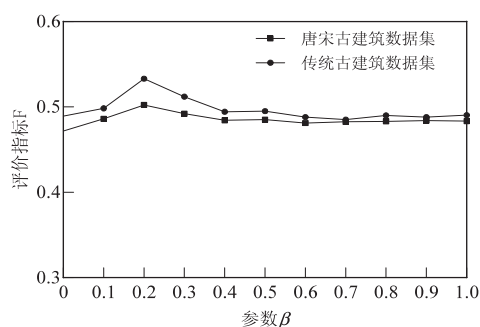


图 6 评估指标 F 随参数 β 的变化

2.5 标签完备

本节验证概念格语义丰富的有效性, 在文中两类数据集上分别随机选取一定数量的图像进行实验, N 为概念格节点数, 选取准确率 P 为评价指标, 计算方法见公式(6), 准确率 P 的大小与标注图像获取的标签数目以及正确标注图像的标签数目有关。实验中, 通过计算每幅图像所产生语义标签的数目, 将图像所生

成的正确标签在候选标签集中进行标记,来获取每幅图像语义扩展之后的准确率,实验三次,取平均值,实验结果如表3所示。

表3 概念格语义扩展实验

数据集	图像数目	平均每幅 图像初始 标签数目	平均每幅 图像补充 标签数目	<i>N</i>	<i>P</i> /%
唐宋古建筑数据集	50	1.08	1.98	31	64.5
传统古建筑数据集	200	1.035	2.025	74	66.3

由表3可以看出,在两类数据集上,图像初始标签数目基本为1,完善后的标签数目为3,这是因为古建筑语义标签设置为三大类(屋顶形式、屋檐以及建筑构件),文中方法通过概念格上下文语义相关度计算,有效地避免了类间语义的错误传播,比如屋顶形式庑殿顶和歇山顶两个标签之间并没有关联,概念格通过计算语义类间的关联,补充了与待标注图像初始标签关联度高的标签,对古建筑图像进行了有效标注。

2.6 实验对比

为了验证文中方法的有效性,在两类数据集上进行了标注实验对比,实验三次,对 *P*、*R*、*F* 三个指标取平均值,实验结果如表4所示。

在表4中,VAM表示文中所用到的视觉注意力机制。从表4中可以看出,文中方法在两类数据集上均取得了较高的准确率和召回率,特别是通过概念格完备语义标签之后,标签的召回率得到了较大提升。这

表4 文中方法对比



Method	唐宋古建筑数据集			传统古建筑数据集		
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
VGG16	0.52	0.38	0.44	0.50	0.37	0.43
VGG16+VAM	0.56	0.39	0.46	0.54	0.38	0.45
VGG16+VAM+FCA	0.54	0.51	0.53	0.51	0.49	0.50

是因为在特征提取阶段利用了视觉注意力机制加强了古建筑图像的边界、轮廓等重要特征,提高了图像的初始标注效果,然后计算了图像之间的相似度,根据相似度系数越高得到图像特异性越强的特点,得到具有高准确度的候选图像,提高了标注的准确率 *P*。在语义学习阶段利用概念格挖掘出了标签之间的相关性,扩展了图像标签,提高了标注的召回率 *R*,与当下多数研究者使用 VGG16 网络相比,文中方法的准确率和召回率都有所提升,在标签召回率上提升较大,证明了文中方法能够有效扩展图像标签,从而丰富了古建筑图像的语义信息。在文中方法中,加入概念格之后图像标注准确率比视觉注意力机制方法稍有降低,这是由于在概念格进行语义扩展时,候选标签集中标签的数量会不断增多,错误地标记了部分噪声标签,导致准确率降低。

2.7 标注实例

表5给出了部分图像的标注实例,其中第一列是文中使用的数据集,第二列为待标注的图像,第三列为图像的初始标签,即图像的待完备标签词,第四列为利用文中方法得到的标注结果。

表5 标注实例

数据集	待标注图像	初始标签	标注结果
唐宋古建筑数据		XieShanDing(歇山顶)	XieShanDing(歇山顶) SingleEaves(单檐) WoodStructure(木结构)
		WuDianDing(庑殿顶)	WuDianDing(庑殿顶) DoubleEaves(重檐) BrickStructure(砖结构)
传统古建筑数据		XieShanDing(歇山顶)	XieShanDing(歇山顶) SingleEaves(单檐) BrickStructure(砖结构) WoodStructure(木结构)
		XieShanDing(歇山顶)	XieShanDing(歇山顶) DoubleEaves(重檐) BrickStructure(砖结构)

其中加粗字体为正确标注的标注词,非加粗字体为图像中不应该存在但被文中方法标注的标签词。在初始标签预测阶段,文中方法可以利用卷积神经网络和视觉注意力识别古建筑图像中的具体信息,并产生能够反映古建筑图像内容的标签,但仅使用深度模型

进行古建筑图像标签完备,会带来许多噪声标签。由于古建筑图像本身含有多层语义,各语义之间又存在关联,而深度学习并不能很好地学习关联语义的特征,造成语义信息的缺失。比如在第二张图中,标注结果标注了图像的深层语义标签“砖结构”,该标签词无法

从图像的视觉特征得到,但是经过概念格的提取标签的相关性,挖掘了古建筑图像更深层的语义内容。由此可以证明概念格可以有效地改善标注结果,丰富古建筑图像的语义信息。但在第三行第一张图的标注结果中,产生了错误的标签“单檐”,这是因为“单檐”和“歇山顶”相关性较高,而图像中的初始标签为“歇山顶”,所以导致了错误的标注。这也说明标签之间的相似性会影响最终的标注结果,若能进一步细化古建筑图像与标签的相关性,去除噪声标签,将会极大地提升古建筑图像标注性能。

3 结束语

提出一种基于视觉注意力和 FCA 的古建筑图像标注方法。该方法利用视觉注意力提取了古建筑图像的边界轮廓等重要特征,通过卷积神经网络得到了视觉特征和语义标签之间的映射关系,在此基础上利用概念格对语义标签进行扩展,通过在古建筑数据集上进行标注实验,证明了该方法可以有效地提升古建筑图像标注性能。下一步工作是进一步探索古建筑语义标签之间的关联性以及细化古建筑语义标签,以提高古建筑图像标注精度,丰富古建筑图像语义。

参考文献:

- [1] 陈馨玉. 中国建筑的中国标志[J]. 中国美学, 2018(1): 72-92.
- [2] LIU A, ZHOU X. The reflection of Chinese ancient religious philosophy thought in the artistic conception of Chinese traditional architecture [C]//Proceedings of 2019 international conference on religion, culture and art (ICRCA). Canada: Clausius Scientific Press, 2019.
- [3] 朱善华, 徐雪宁. 中国古建筑文化与当代建筑文化的结合[J]. 建筑与文化, 2018(6): 42-43.
- [4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//Proceedings of the 2005 IEEE conference on computer vision and pattern recognition (CVPR). San Diego: IEEE, 2005: 886-893.
- [5] YANG X, LATECKI L J. Affinity learning on a tensor product graph with applications to shape and image retrieval [C]//IEEE conference on computer vision and pattern recognition (CVPR). Providence: IEEE, 2011: 2369-2376.
- [6] RAJKOMAR A, LINGAM S, TAYLOR A G, et al. High-throughput classification of radiographs using deep convolutional neural networks [J]. Journal of Digital Imaging, 2017, 30(1): 95-101.
- [7] ZHENG Y, LI Z, ZHANG C. A hybrid architecture based on CNN for cross-modal semantic instance annotation [J]. Multimedia Tools and Applications, 2018, 77(7): 8695-8710.
- [8] LI Z, LIU F, YANG W, et al. A survey of convolutional neural networks: analysis, applications, and prospects [J]. arXiv: 2004. 02806, 2020.
- [9] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(11): 1254-1259.
- [10] LIU T, YUAN Z, SUN J, et al. Learning to detect a salient object [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(2): 353-367.
- [11] WANG F, JIANG M, CHEN Q, et al. Residual attention network for image classification [C]//2017 IEEE conference on computer vision and pattern recognition (CVPR). Honolulu: IEEE, 2017: 6450-6458.
- [12] ZHANG Q, SHI Y, ZHANG X. Attention and boundary guided salient object detection [J]. Pattern Recognition, 2020, 107(7): 107484.
- [13] 林泓, 任硕, 杨益, 等. 融合自注意力机制和相对鉴别的无监督图像翻译 [J]. 自动化学报, 2021, 47(9): 2226-2237.
- [14] LENG J, LIU Y, CHEN S. Context-aware attention network for image recognition [J]. Neural Computing and Applications, 2019, 31(12): 9295-9305.
- [15] 汤文兵, 任正云, 韩芳. 基于注意力机制的协同卷积动态推荐网络 [J]. 自动化学报, 2021, 47(10): 2438-2448.
- [16] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts [C]//International conference on formal concept analysis. Berlin: Springer, 2009.
- [17] HAO Fei, YANG Yixuan, MIN Geyong, et al. Incremental construction of three-way concept lattice for knowledge discovery in social networks [J]. Information Sciences, 2021, 578: 257-280.
- [18] ZHANG Sulan, ZHANG Jifu, GUO Ping, et al. A FWCL-based method for visual vocabulary formation [J]. Multimedia Tools and Applications, 2016, 75(1): 647-665.
- [19] ZOU Li, LIN Hongmei, SONG Xiaoying, et al. Rule extraction based on linguistic-valued intuitionistic fuzzy layered concept lattice [J]. International Journal of Approximate Reasoning, 2021, 133: 1-16.
- [20] 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.
- [21] BAI C, HUANG L, PAN X, et al. Optimization of deep convolutional neural network for large scale image retrieval [J]. Neurocomputing, 2018, 303: 60-67.
- [22] ZHAO Yi, HALANG W. Rough concept lattice based ontology similarity measure [C]//Proceedings of the first international conference on scalable information systems. Hong Kong: ACM, 2006: 15-es.
- [23] WANG L, LIU X. A new model of evaluating concept similarity [J]. Knowledge-Based Systems, 2008, 21(8): 842-846.