Journal of Sensor Technology and Application 传感器技术与应用, 2023, 11(6), 505-509
Published Online November 2023 in Hans. https://www.hanspub.org/journal/jsta
https://doi.org/10.12677/jsta.2023.116057

Hans汉斯

基于全局注意力机制的图像检索算法研究

汤海冰，蒋新发，杨　影

湖南文理学院计算机与电气工程学院，湖南 常德

收稿日期：2023年9月26日；录用日期：2023年10月29日；发布日期：2023年11月7日

摘　要

针对图像检索中由于图像尺度变化大、目标相似性等影响检索精度的问题，本文提出了一种基于多特征融合的图像检索算法，采用残差网络(ResNet50)提取图像特征，加入全局注意力机制(Global Attention Mechanism)，将网络提取的原始特征与GAM注意力机制提取的特征融合，使图像中的关键部分得到更多的关注，实验证明了所提出的算法具有较高的检索准确率和鲁棒性。

关键词

残差网络，注意力机制，图像检索，特征融合

Research on Image Retrieval Algorithms Based on Global Attention Mechanism

Haibing Tang, Xinfa Jiang, Ying Yang

School of Computer and Electrical Engineering, Hunan University of Arts and Sciences, Changde Hunan

Received: Sep. 26th, 2023; accepted: Oct. 29th, 2023; published: Nov. 7th, 2023

Abstract

This paper proposes an image retrieval algorithm based on multi feature fusion to address the issues of significant changes in image scale and target similarity that affect retrieval accuracy in image retrieval. The algorithm uses a residual network (ResNet50) to extract image features, adds a Global Attention Mechanism, and fuses the original features extracted by the network with the features extracted by the GAM attention mechanism, so that key parts of the image receive more attention, The experiment has proven that the proposed algorithm has high retrieval accuracy and robustness.

文章引用：汤海冰，蒋新发，杨影．基于全局注意力机制的图像检索算法研究[J]．传感器技术与应用，2023，11(6)：505-509. DOI: 10.12677/jsta.2023.116057

Keywords

Residual Network, Attention Mechanism, Image Retrieval, Feature Fusion

1. 引言

在图像库中快速有效的检索到用户想要的图像，是很有价值的商业应用。已有许多图像检索算法被提出，而基于深度学习的图像检索技术取得了巨大进展。深度学习通过学习图像的特征表达，提取高级语义特征，相比传统的基于统计特征的图像检索，取得了较高的检索精度。Zhou 等人[1]提出一种基于深度神经网络的图像检索方法，该方法采用卷积神经网络提取图像特征，并使用倒排索引进行检索。国外的 Lowe 等[2]学者则提出了基于局部不变特征的图像检索方法，该方法能够在旋转、平移等变换下依然能够提取出具有区分性的特征，实现了更加鲁棒的图像检索效果。2018 年，中国科学院计算技术研究所的 Liang 和他的团队[3]提出了一种基于多特征融合的图像检索算法。该算法将多种特征融合在一起，包括局部特征、全局特征、颜色特征和纹理特征等。该算法利用深度学习方法进行特征提取，并采用对抗网络对图像进行重构，使得提取的特征更加具有代表性。在多个数据集上的实验结果表明，该算法能够取得较好的检索效果。董华、王涛等[4]学者则在卷积神经网络的不同层次提取不同的特征进行融合，实现了更加准确的图像检索结果。与单特征算法相比，多特征融合算法具有更强的鲁棒性和抗干扰能力，可以有效地解决特征冗余和过拟合等问题，同时也可以克服图像尺度变化和目标相似性等问题，从而提高检索结果的准确性和鲁棒性，渐成图像检索算法主流。Wu 等人[5]研究利用多层特征融合的方法，包括全局特征、局部特征和颜色特征等。与其他方法相比，该算法在特征提取时使用了多个卷积神经网络，而且使用了反卷积神经网络进行图像重构，提高了特征的鲁棒性和鉴别性。在多个公共数据集上的实验表明，该算法在精度和效率方面都优于其他方法。除此之外，还有许多其他的多特征融合算法应用于图像检索中，比如基于视觉词袋模型的多特征融合算法[6]、基于卷积神经网络的多特征融合算法[7]等等。

但前述多特征融合的图像检索方法没有很好地区分图像中不同区域和内容的重要性。这会导致计算资源分配不合理，并且检索结果的准确率会受到影响。为了解决上述问题，本文提出了一个融合注意力特征的图像检索算法。该算法使用 ResNet [8]网络提取图像的特征，使用 GAM 注意力提取图像的重要信息，GAM 注意力机制[9]可以学习到全局特征信息，从而增强网络对输入图像的泛化能力，同时通过 GAM 注意力机制，网络可以更加聚焦于重要的特征通道，避免了过多地关注无用特征通道。算法将原始特征与通过 GAM 模块获取的特征融合，使图像中的关键部分得到更多的关注。

2. 模型设计

2.1. GAM 注意力机制

GAM 注意力机制是一种用于图像分类和目标检测任务的注意力机制，旨在提高神经网络对图像中重要区域的关注度。GAM 机制通常用于在特征图上进行全局加权操作，以获取图像中最具有代表性的特征。

与其他注意力机制不同，GAM 机制是一种全局操作，它通过考虑整个特征图而不是单个通道来计算每个通道的权重，GAM 由通道注意力和空间注意力模块组成。具体来说，对于给定的特征图，GAM 会

首先将其压缩为一维向量，然后使用一个全连接层来学习每个通道的权重。这些权重随后被应用于特征图上，以加权对应通道的每个特征。最终，加权的特征被级联在一起形成全局特征表示，该表示被馈送到分类器中以进行分类。GAM 注意力结构图如图 1 所示。

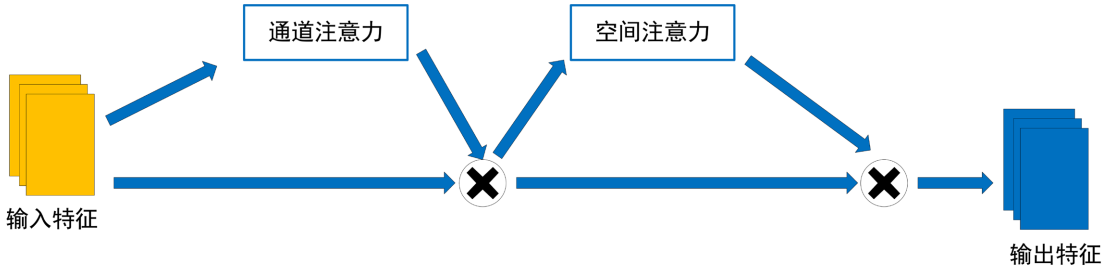

Figure 1. GAM attention structure diagram
图 1. GAM 注意力结构图

通道注意力结模块采用三维数组来保留三个维度上的信息，从而实现通道注意力。一个包含两个层的多层感知器(MLP)被用于放大通道与空间之间的依赖性。(MLP 是编码－解码结构，类似于 BAM，且其压缩比为 r) 该通道注意子模块的结构如图 2 所示。

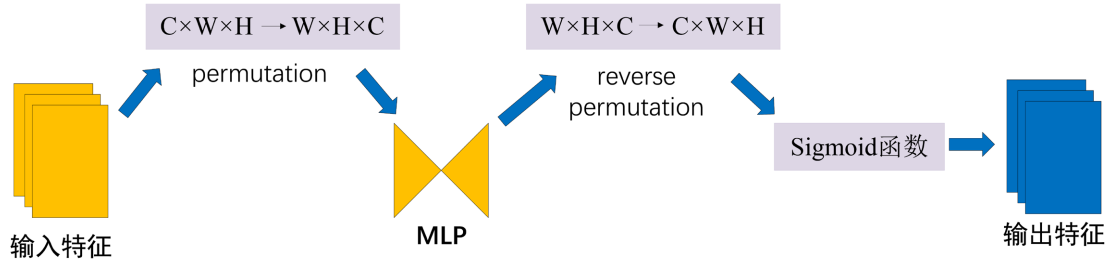

Figure 2. Channel attention structure diagram
图 2. 通道注意力结构图

为了关注空间信息，空间注意力子模块使用两个卷积层进行空间信息融合，并从通道注意力子模块中使用了与 BAM 相同的缩减比 r。由于最大池化操作减少了信息的使用，产生了消极的影响，因此该模块删除了池化操作，以进一步保留特性映射。空间注意力子模块示意如图 3。

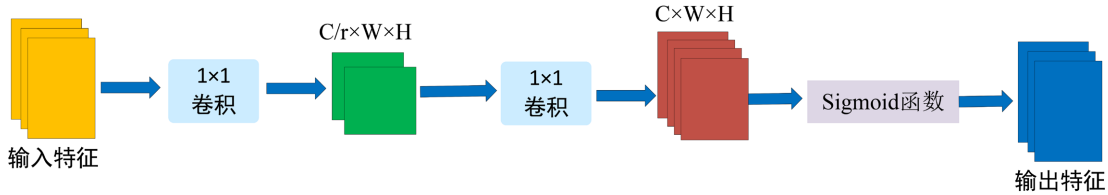

Figure 3. Spatial attention structure diagram
图 3. 空间注意力结构图

GAM 注意力机制的主要优点在于，它可以将注意力集中在最具代表性的特征上，从而提高模型的分类准确率。此外，由于 GAM 机制是一种全局操作，因此它可以自适应地计算每个通道的权重，而不受任何先验假设的约束，从而提高了模型的灵活性和鲁棒性。

2.2. 系统模型

系统的骨干网络采用 Resnet50，采用 Resnet 作为骨干网络是因为其实分类网络的标准架构，系统模型结构即为 Resnet50 + GAM。

3. 实验

3.1. 数据集与实验配置

采用 COREL-10K 数据集[10]。该数据集包含 10,000 张图像，涵盖了 100 个类别，每个类别有 100 张图像。这些图像涵盖了许多不同的主题，如动物、自然风景、建筑、运动、人物等，被用于许多不同的任务，如图像分类、图像检索、目标检测、目标跟踪等。

实验环境设置。操作系统为 Ubuntu 16.04.7 LTS，Pytorch 深度学习框架，版本为 1.12.1 + cu116，GPU 为 NVIDIA Tesla T4 16G，运行系统为文理学院高性能服务器的一部分。

实验参数在 Resnet50 上微调，GAM 参数的学习率为 Resnet50 参数的 10 倍，训练轮数为 200 epochs。

3.2. 评价标准

采用工程上广泛应用的重要评价指标：Top-K 分类精度(precision)。计算公式为：

$$precision@K = \sum_{i=1}^{k} R(i)/K \quad \text{(公式 1)}$$

检索出的前 K 个图像中，只要有一幅图像与检索图像标签相同，那么 $R(i)=1$。本实验检索标签为图像的类别。

3.3. 实验比较

为验证 GAM 注意力机制的有效性，我们研究了不使用注意力机制和 GAM 注意力机制与 SENet [11]、CBAM [12]这两种常见注意力机制应用在图像检索上检索精度的对比。即 Resnet50 + GAM 与 Resnet50、Resnet50 + SENet 注意力机制和 Resnet50 + CBAM 注意力机制的对比。检索结果如下表 1：

Table 1. Comparison of retrieval accuracy
表 1. 检索精度对比

检索架构	参数大小	Top-1 精度	Top-10 精度
Resnet50	25.6 M	77.25	90.62
Resnet50 + SENet	28.1 M	78.08	91.51
Resnet50 + CBAM	28.2 M	78.57	91.84
Resnet50 + GAM	151.4 M	79.18	91.93

可以看出 Resnet50 + GAM 同其他三种方法比较，Top-1 和 Top-10 上精度都是最高的，只是在 Top-10 上精度提升没那么明显，但在参数大小上 Resnet50 + GAM 几乎是其他三种方法的六倍，这也是本文模型精度更高的原因，也同当前大模型(如 GPT)趋势一致。

4. 结束语

本文提出了 Resnet + GAM 网络模型，该模型把从 ResNet50 提取的特征通过 GAM 注意力机制进行特征融合，让模型学习到图像重要部分的关注，从而提高图像检索的性能。实验结果表明，与 Resnet 和 SENet、CBAM 注意力机制相比，本文模型有更高的检索精度。

基金项目

湖南省教育厅科学研究项目：2019 年湖南省教育厅科学研究项目“基于三维建模的旅游图像处理技术研究”(19C1276)；2017 年湖南文理学院博士科研启动项目“旅游图像处理技术研究”(E07017005)。

参考文献

[1] Zhou, B., Lapedriza, A., Xiao, J., *et al*. (2014) Learning Deep Features for Image Retrieval Using Convolutional Neural Networks. *Advances in Neural Information Processing Systems* 27 (*NIPS* 2014), Montreal, 8-13 December 2014, 157-172.

[2] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[3] Liang, L., Yang, H., Zhang, J., *et al*. (2018) Multi-Feature Fusion Based on Deep Learning for Image Retrieval. *Neurocomputing*, **275**, 2357-2364.

[4] 董华, 王涛. 基于深度神经网络的图像检索研究[J]. 计算机工程与设计, 2018, 39(7): 1662-1666.

[5] Wu, X., Zha, Z.J., Yang, Y., *et al*. (2019) A Multilayer Feature Fusion Framework for Image Retrieval. *IEEE Transactions on Image Processing*, **28**, 147-162.

[6] Liu, Y., Wang, M., Cao, L., *et al*. (2017) A Multiple Feature Fusion Model for Image Retrieval Based on Bag of Words. 2017 *IEEE* 2*nd Advanced Information Technology*, *Electronic and Automation Control Conference* (*IAEAC*), Chongqing, 25-26 March 2017, 2644-2649.

[7] Wang, Y., Yao, H., Shi, H., *et al*. (2018) Deep Multiple Feature Fusion for Image Retrieval. *Multimedia Tools and Applications*, **77**, 19529-19547.

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

[9] Zhang, H., Wang, L., Liu, Y. and Shen, H.T. (2018) Graph Attention Networks. *Proceedings of the* 32*nd Conference on Neural Information Processing Systems*, Montreal, 3-8 December 2018, 9110-9119.

[10] Lee, S.W., Kim, S.H., Lee, S.U., and Kim, S.J. (2006) Content-Based Image Retrieval Using Color and Texture Combined Features. *Pattern Recognition Letters*, **27**, 1805-1811.

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

[12] 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, Cham, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1