

基于注意力机制的医学影像深度哈希检索算法

朱承璋, 黄嘉儿, 肖亚龙, 王晗, 邹北骥

引用本文

朱承璋, 黄嘉儿, 肖亚龙, 王晗, 邹北骥. [基于注意力机制的医学影像深度哈希检索算法](#)[J]. 计算机科学, 2022, 49(8): 113-119.

ZHU Cheng-zhang, HUANG Jia-er, XIAO Ya-long, WANG Han, ZOU Bei-ji. [Deep Hash Retrieval Algorithm for Medical Images Based on Attention Mechanism](#)[J]. Computer Science, 2022, 49(8): 113-119.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多时间尺度时空图网络的交通流量预测模型](#)

Multi-time Scale Spatial-Temporal Graph Neural Network for Traffic Flow Prediction

计算机科学, 2022, 49(8): 40-48. <https://doi.org/10.11896/jsjcx.220100188>

[基于卷积神经网络的 APP 用户行为分析方法](#)

Analysis Method of APP User Behavior Based on Convolutional Neural Network

计算机科学, 2022, 49(8): 78-85. <https://doi.org/10.11896/jsjcx.210700121>

[基于非局部注意力生成对抗网络的视频异常事件检测方法](#)

Non-local Attention Based Generative Adversarial Network for Video Abnormal Event Detection

计算机科学, 2022, 49(8): 172-177. <https://doi.org/10.11896/jsjcx.210600061>

[基于图卷积神经网络的文本分类方法研究综述](#)

Review of Text Classification Methods Based on Graph Convolutional Network

计算机科学, 2022, 49(8): 205-216. <https://doi.org/10.11896/jsjcx.210800064>

[基于双图神经网络信息融合的文本分类方法](#)

Text Classification Method Based on Information Fusion of Dual-graph Neural Network

计算机科学, 2022, 49(8): 230-236. <https://doi.org/10.11896/jsjcx.210600042>

基于注意力机制的医学影像深度哈希检索算法

朱承璋^{1,2,3,4} 黄嘉儿^{1,3,4} 肖亚龙^{1,2} 王 晗^{1,3,4} 邹北骥^{1,3,4}

1 中南大学计算机学院 长沙 410083

2 中南大学文学与新闻传播学院 长沙 410083

3 “移动医疗”教育部-中国移动联合实验室 长沙 410083

4 湖南省机器视觉与智慧医疗工程技术研究中心 长沙 410083

(anandawork@126.com)

摘要 针对现阶段医学影像检索中检索性能差、精度低、缺乏可解释性等一系列问题,提出了一种结合了注意力机制的医学影像检索算法。以深度卷积神经网络为基础,以贝叶斯模型为框架,所提算法引入了由语义特征引导的注意力机制模块,通过分类网络的引导,生成包含语义信息的局部特征描述子,同时使用全局特征与富含语义信息的局部特征作为哈希网络的输入,引导哈希网络从全局和局部的角度关注重要特征区域,增强了哈希编码的特征表达能力,并引入加权似然估计函数解决了正负样本对数量不均衡的问题。采用 MAP 和 NDCG 作为评价指标,选择 ChestX-ray14 数据集进行实验,将所提算法与目前常用的深度哈希方法进行对比。实验结果表明,本文算法在哈希编码不同码位下的 MAP 值和 NDCG 值都远优于现有的深度哈希方法,证明了其有效性。

关键词: 医学影像检索;注意力机制;深度哈希;贝叶斯框架;卷积神经网络

中图分类号 TP391

Deep Hash Retrieval Algorithm for Medical Images Based on Attention Mechanism

ZHU Cheng-zhang^{1,2,3,4}, HUANG Jia-er^{1,3,4}, XIAO Ya-long^{1,2}, WANG Han^{1,3,4} and ZOU Bei-ji^{1,3,4}

1 School of Computer Science and Engineering, Central South University, Changsha 410083, China

2 School of Literature and Journalism, Central South University, Changsha 410083, China

3 Mobile Health Ministry of Education-China Mobile Joint Laboratory, Changsha 410083, China

4 Hunan Engineering Research Center of Machine Vision and Intelligent Medicine, Changsha 410083, China

Abstract A medical image retrieval method combining attention mechanism is proposed for a series of problems such as poor retrieval performance, low accuracy and lack of interpretability in current medical image retrieval. Based on deep convolutional neural networks and taking Bayesian models as the framework, the proposed algorithm introduces an attention mechanism module guided by semantic features. Local feature descriptors containing semantic information are generated under the guidance of the classification network. Both global features and local features rich in semantic information are used as inputs to the hash network, which enhances the feature representation capability of hash coding by guiding the hash network to pay attention to important feature regions from both global and local perspectives. And the weighted likelihood estimation function is introduced to solve the problem of the unbalanced number of positive and negative sample pairs. MAP and NDCG are used as evaluation metrics, and the ChestX-ray14 dataset is selected for experiments. The proposed algorithm is compared with the current commonly used deep hashing methods. Experiment results show that the MAP and NDCG values are much better than the existing deep hashing methods at different code levels of hash coding, which proves the effectiveness of the proposed algorithm.

Keywords Medical image retrieval, Attention mechanism, Deep hashing, Bayesian framework, Convolutional neural networks

到稿日期:2021-07-14 返修日期:2021-10-23

基金项目:科技创新 2030—“新一代人工智能”重大项目(2018AAA0102100);湖南省高新技术产业科技创新引领计划(2020GK2021)

This work was supported by the National Key R&D Program of China(2018AAA0102100) and Hunan Province High-tech Industry Science and Technology Innovation Leading Program(2020GK2021).

通信作者:肖亚龙(ylxiao@csu.edu.cn)

1 引言

随着医学成像技术和计算机视觉技术的飞速发展,计算机辅助检测和计算机辅助诊断(Computer Aided Diagnosis, CAD)为医生的工作带来了极大的便利,医生可以对海量的医学影像进行检索,在数据库中迅速找到相似的病例,从而加速诊断流程。传统的图像检索系统通过图像的颜色、形状、纹理等浅层特征对其建立特征索引库,并通过计算特征间的距离来判断图像的相似性。而医学影像往往具有高同一性,且普遍存在类间差异小、类内差异大的特点,仅采用浅层特征往往无法很好地表达影像中影响诊断的语义性特征,检索任务驱动的特征提取算法也缺少一定的泛化能力。同时,随着医学大数据技术的发展,传统的图像检索系统在面对海量医学影像数据时无法同时满足高精度和高效率这两大需求。

近年来,深度卷积神经网络(Convolutional Neural Network, CNN)在图像分类^[1]、分割^[2]、生成^[3]任务中都取得了极其优异的成果。深度卷积神经网络可从海量的医学影像中学习得到其特征,其对抽象的语义信息更敏感,抽取的深层特征能更好地表达图像本身的语义信息。基于深度学习强大的特征表达能力及其端到端的架构,深度哈希学习应运而生。

CNN 提取到的图像高级特征中包含图像高级的语义信息。深度哈希学习^[4]将语义特征二值化为哈希编码,建立了图像数据到哈希编码的二值映射。相比传统图像检索系统对特征进行距离计算,深度哈希方法直接将图像编码为哈希编码,使用汉明距离对相似性进行度量,大大提高了检索的效率。哈希编码的有效性很大程度取决于高级语义特征的有效性。而对于医学影像检索任务而言,图像本身包含着一些无用的背景信息,如器官的整体结构信息、非病灶区域信息。在码长有限的情况下,这些无用信息给哈希编码带来了一定的冗余。为了兼顾存储和检索性能,同时平衡哈希编码的码长与码字的特征表达效率,则需要尽量排除冗余信息的干扰。

注意力机制^[5]模拟了人类的注意力思维方式,被广泛应用于自然语言处理、图像分类及图像分割等不同类型的深度学习任务中。医生在诊断病情时会先关注整体,再关注病灶区域,自动屏蔽不重要的背景信息。

受医生诊断过程的启发,本文提出了一个基于注意力机制的医学影像深度哈希框架。在 ResNet50^[6]模型的基础上,引入了由全卷积网络(Fully Convolutional Network, FCN)^[7]引导的注意力机制,用于捕捉与分类任务强相关的特征描述子,将其编码为局部信息。令深度哈希网络同时学习全局信息和局部信息,增强了哈希编码的语义特征表达能力和鲁棒性。本文方法在保证检索效率的同时,提升了对医学影像的检索精度。

2 相关工作

2.1 深度哈希学习

CNN 在各类计算机视觉任务上都有着优异的表现,深度哈希学习也引起了广泛的关注。

Xia 等^[8]基于深度卷积网络构造了基于相似度矩阵变换的哈希学习方法(Convolutional Neural Network Hashing,

CNNH)。该方法将通过相似性矩阵 S 进行变换,将矩阵 S 分解为 H 和 H^T 。其中, H 将作为训练集样本的初始哈希编码。CNNH 将哈希编码生成问题转换为哈希编码 H 与相似性矩阵 S 之间的关系问题,通过优化哈希目标函数,学习得到图像训练集的哈希编码。然而,此方法无法通过网络更新二值编码,无法实现端到端的哈希编码生成。Pan 等^[9]提出的 DNNH(Deep Neural Network Hashing)对 CNNH 提出了改进,并设计了一种端到端的网络结构学习哈希编码。其选用了基于三元组的损失函数,减小相似样本对之间的距离,增大不相似样本对之间的距离,通过动态组合样本对,解决了相似样本对与不相似样本对之间数量不均衡的问题,并使用部分连接层代替全连接层,使码位间相互独立,有效地减少了哈希编码间的冗余。Zhu 等^[10]提出的 DHN(Deep Hashing Network)方法则选用了贝叶斯框架对哈希编码生成问题进行建模,将哈希编码生成问题转换为样本对标签的二分类问题,在 NUS-WIDE, CIFAR-10 等数据集上都获得了远优于 DNNH 的性能。然而,上述方法没有对哈希编码的离散问题及量化误差进行进一步的阐述。Cao 等^[11]提出的 HashNet 使用双曲正切函数,在训练过程中动态地调整参数,不断逼近阶跃函数,有效解决了哈希编码的离散问题。在贝叶斯框架下,常使用 sigmoid 作为激活函数。Cao 等^[12]提出的 DCH(Deep Cauchy Hashing)指出了 sigmoid 函数在汉明距离过小时模型收敛过慢的问题,提出了一种基于柯西分布的概率函数,使哈希编码在小码长的情况下检索性能更稳定。Yuan 等^[13]引入了中心相似度的机制,将相似性的度量从单个样本粒度上升到集群的粒度,构建了汉明空间中的集群中心,使得相似性的度量更精准。Wang 等^[14]则针对哈希二值码的离散优化问题,提出了 Kurtosis Loss,在学习连续值的过程中通过 KT 损失函数对其进行阈值约束,解决了连续值离散化过程中的信息丢失问题。

2.2 医学图像哈希检索

基于医学图像的哈希检索系统可以自动在数据库中寻找图像语义相似的医学影像,以辅助医生的初步诊断。有别于自然场景图像,医学影像具有结构固定、成像方式复杂、多模态、类间差异小和类内差异大等特点。因此,无法直接将自然场景中的哈希检索系统直接迁移到医学影像场景中。

Cai 等^[15]提出了一种基于孪生网络的 CT 器官哈希检索框架,其使用 ImageNet 上预训练过的 AlexNet 模型构建孪生网络,将相似样本对和不相似样本对输入到网络中,对不同类别的 CT 图像进行检索。Peng 等^[16]则提出了一种基于三元组的组织病理学多任务模型,其将检索任务与分类任务相结合,同时优化检索任务与分类任务,提高了分类任务的可解释性。Chen 等^[17]则根据 DSRH 的思想,设计了基于多标签的 CT 图像哈希检索系统。相比 DSRH,此方法根据标签交集信息设计了一个加权的损失函数,共有的标签越多,排序结果就越靠前,实现了排序敏感的 CT 图像哈希检索系统。Zheng 等^[18]则充分挖掘病理图像切片在空间维度上的信息,引入了图神经网络,将切片图像转换为图进行深度哈希的学习,更好地保留了图像的结构信息,避免了病理切片图像类间差异小的问题。Sliva 等^[19]引入了可解释性的机制,将具有可解释性

的分类热力图与原图叠加输入网络进行训练,提高了系统的可解释性。Fang 等^[20]则同时利用了空间注意力机制和多任务学习,引入了三元组网络,解决了医学深度哈希检索中的小样本丢失问题。

2.3 注意力机制

注意力机制模仿了人类观察物品时的行为过程,可快速提取稀疏数据中的重要特征。在对医学影像进行诊断时,医生除了从整体把握图像的结构外,还会更加关注局部的病灶区域信息。深度哈希在提升检索效率的同时,也因为码长或码字的限制,或多或少地丢失了特征中的语义信息。引入注意力机制可以使深度哈希网络关注图像中重要区域的信息,使得哈希编码尽可能地保存更重要的语义信息,减少冗余。Hu 等^[21]提出的 SENet 在图像通道维进行压缩,将全局的空间信息压缩到一个通道描述符中,令通道描述符与网络计算得到的权重相乘,自适应地调整通道维的特征响应。相比只关注通道域的 SENet, Woo 等^[22]提出的 CBAM(Convolutional Block Attention Module)同时考虑了空间域和通道域的信息,对空间域和通道域依次采用了全局平均池化和全局最大池化,得到对应域的两个特征描述符,使提取得到的高层次特征更加丰富。将描述符输入到多层感知器(Multilayer Perceptron, MLP)中学习权重系数,再将权重系数与特征相乘,对特征进行自适应的调整。注意力机制在图像分类、自然语言处理等领域都有广泛的应用,但在深度哈希任务上的潜能还有待挖掘。

3 本文算法

在医学图像检索系统中,给定一个包含了 N 幅图像的数据集 $I = \{x_i\}_{i=1}^N$, 其中 x_i 代表数据集中索引为 i 的一幅图像。 $L = \{y_i\}_{i=1}^N$ 是数据集对应的标签向量,其中, $y_i \in \{0, 1\}^C$, C 代表了数据集的标签类别总数。样本对之间的关系通过相似性

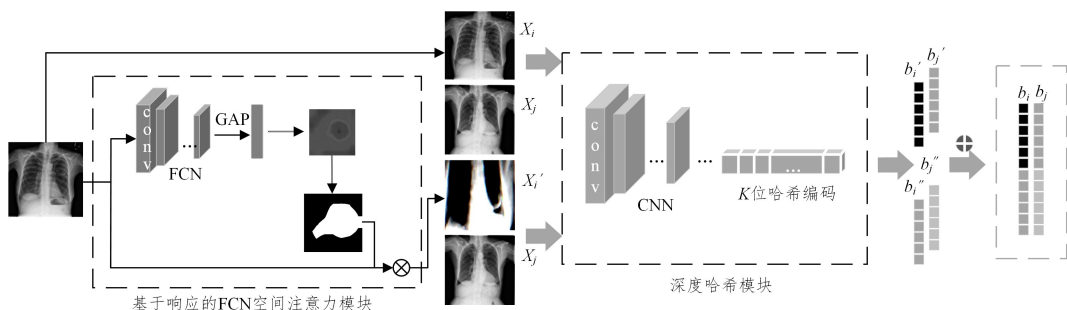


图1 基于注意力机制的深度哈希框架

Fig. 1 Attention-based deep hashing framework

3.1 基于响应的 FCN 注意力模块

当医生诊断案例时,除了关注整体图像外,通常更加关注病灶区域。因此,本文引入了基于响应的 FCN 注意力模块,将包含丰富语义信息的热力图与原图叠加,以获得与分类结果强相关的局部图像块,引导深度哈希网络学习更具有区分度的哈希编码。

Grad-CAM 采用梯度的全局平均来获取每个特征图的权重,其数学原理如式(1)所示:

矩阵 $S = \{s_{ij}\} \{0, 1\}^{N \times N}$ 连接。对于有监督深度哈希而言,相似性矩阵 S 可以通过数据标签或用户反馈来构建。当样本对之间至少有一个标签相同时, $s_{ij} = 1$; 当样本对所有标签都不相同时, $s_{ij} = 0$ 。假设图像 x_i 的哈希编码为 $\{b_i\} \in \{-1, 1\}^k$, k 为哈希编码的位数。深度哈希的目标是通过卷积神经网络建立一个从图像空间到汉明空间的一个非线性映射关系 $f: X \rightarrow B$, X 代表图像数据集, B 代表存储了所有数据哈希编码的矩阵。此映射关系 f 应尽量保证汉明空间中的数据分布靠近原始图像空间中的数据分布。即对于相似的样本对 (x_i, x_j) , 其对应的哈希编码对 (b_i, b_j) 之间的汉明距离应尽量小; 反之, 不相似样本对之间的汉明距离应尽量大。

本文提出了一个基于注意力机制的医学影像深度哈希框架。图1为本文算法的框架图,其主要由4个模块构成。

(1) 基于响应的 FCN 空间注意力模块。借鉴多任务学习的思想,先对数据集进行分类任务的训练,在得到可靠的分类模型后,通过梯度加权类激活映射方法^[23] (Gradient-weighted Class Activation Mapping, Grad-CAM)对最后一层的特征图计算权重,并对其对应类别的权重进行加权求和,通过 ReLU 激活函数得到分类任务的热力图。将热力图通过 Otsu's 二值化算法^[24], 自动选择阈值,变换得到掩膜,将掩膜与原图叠加,获得与分类结果强相关的图像块 x_i' 。

(2) 基于 ResNet50 的二分支的哈希网络。将图 x_i 和图 x_j 分别输入到双分支深度哈希网络中,训练得到 b_i', b_j', b_i'' 和 b_j'' 。其中, b_j' 和 b_j'' 作为 b_i' 和 b_i'' 的辅助训练编码。

(3) 损失函数。损失函数包括分类损失函数 L_{cls} 和基于贝叶斯框架的加权相似性损失 $L_{weightedPair}$ 。

(4) 量化约束。由于离散问题不可微,无法对其反向传播,因此本文算法将离散问题转化为趋近于极限的连续问题。通过调整参数,利用双曲正切函数逼近阶跃函数,生成离散的哈希编码。

$$w_k^c = \frac{1}{R} \sum_i \sum_j \frac{\partial g^c}{\partial P_{ij}^k} \quad (1)$$

其中, k 表示网络最后一层特征图的个数, w_k^c 代表第 k 个特征图对应标签 c 的权重, g^c 代表第 c 类得分的梯度, P_{ij}^k 代表第 k 个特征图中位置 (i, j) 的特征值, R 代表特征图的像素总数。

获得每个特征图的权重后,根据对应类别,对相应的权重进行加权求和,通过 ReLU 激活函数得到热力图,如式(2)所示:

$$Heatmap_{Grad-CAM}^c = \text{ReLU}(\sum_k w_k^c P^k) \quad (2)$$

本文将图像 x_i 输入到已经在分类任务上预训练好的 DenseNet121 网络中, 提取最后一层卷积层特征 $f \in \mathbb{R}^{H \times W \times C}$, H, W, C 分别表示特征图的高、宽和通道数。将 f 输入到 GAP 层后, 得到空间特征描述子, 对空间描述子求其权重。通过式(2)计算得到标签 c 所对应的热力图 $Heatmap_{G_{\text{rand-CAM}}^c}$ 。对 $Heatmap_{G_{\text{rand-CAM}}^c}$ 使用 Otsu's 二值化算法自动选择阈值, 计算得到热力图对应的掩膜 $Mask_c$ 。

$$Mask_c = f^{\circ} (Heatmap_{G_{\text{rand-CAM}}^c}) \quad (3)$$

将 $Mask_c$ 与原图 x_i 相乘。

$$x_i' = Mask_c \odot x_i \quad (4)$$

得到和分类结果强相关的图像块 x_i' 。

为了同时保存全局信息和局部信息, 本文将包含全局信息的哈希编码 b_i' 和包含局部信息的哈希编码 b_i'' 拼接得到的哈希编码 b_i 作为图像最终的哈希编码, 同时保留粗粒度和细粒度的信息, 以获得更具有鲁棒性的哈希编码。

3.2 损失函数

3.2.1 分类损失

本文采用 sigmoid 作为输出层的激活函数, 将二分类交叉熵损失函数 (Binary Cross Entropy Loss, BCE Loss) 作为损失函数:

$$L_{\text{cls}} = -\frac{1}{C} \sum_{i=1}^C y_i \log f_i(x) + (1 - y_i) \log(1 - f_i(x)) \quad (5)$$

其中, x 表示输入样本, C 表示多标签分类的样本总数, y_i 表示第 i 个类别对应的真实标签, $f_i(x)$ 表示激活后的模型的输出值。 L_{cls} 对样本的每个标签进行了二分类交叉熵损失的计算, 再将所有类别的损失进行求和并取平均值, 将平均值作为最终的多分类损失。

3.2.2 基于贝叶斯框架的相似性损失

样本对哈希编码间的相似性由汉明距离的大小表示, 汉明距离指两个等长哈希编码中同一码位不同码字的个数。由于无法直接对哈希编码生成这一离散问题进行优化, 深度哈希将此问题通过汉明距离与哈希编码之间内积的关系, 将其转换为一种连续的表达。汉明距离与哈希编码的内积之间有如下关系:

$$dist_{\text{H}}(b_i, b_j) = \frac{1}{2(K - \langle b_i, b_j \rangle)} \quad (6)$$

其中, b_i, b_j 表示两个哈希编码, $dist_{\text{H}}(b_i, b_j)$ 表示两个哈希编码间的汉明距离, K 表示哈希编码的码长, $\langle b_i, b_j \rangle$ 表示 b_i, b_j 的内积。通过式(6)可将离散的哈希编码生成问题转换为哈希编码间内积的优化问题。

贝叶斯公式可以将先验概率转换为后验概率。在已知相似性标签 s_{ij} 的情况下, 可以对哈希编码进行学习。此框架的表达式为:

$$\log P(S|B) = \sum_{s_{ij} \in S} \log P(s_{ij} | b_i, b_j) \quad (7)$$

其中, S 为所有数据的相似度标签, B 为所有数据的哈希编码, s_{ij} 为对应样本对的相似度标签, b_i, b_j 为样本对之间的哈希编码。概率函数的表达式为:

$$P(s_{ij} | b_i, b_j) = \begin{cases} \sigma(b_i, b_j), & s_{ij} = 1 \\ 1 - \sigma(b_i, b_j), & s_{ij} = 0 \end{cases}, \sigma = \frac{1}{1 + e^{-\alpha x}} \quad (8)$$

其中, $P(s_{ij} | b_i, b_j)$ 表示当给定哈希编码对 b_i, b_j 时, 其对应的相似性标签 s_{ij} 的条件概率; α 为调节反向传播效率的参数,

通常将 α 设置为 $\alpha < 1$ 。

在数据集中, 常常会出现相似性样本对分布不平衡的问题, 不相似的样本对数量往往远大于相似的样本对数量。因此, 需要引入加权参数, 动态调整损失函数。加权参数的定义如下:

$$w_{ij} = \begin{cases} \frac{|S|}{|S_1|}, & s_{ij} = 1 \\ \frac{|S|}{|S_0|}, & s_{ij} = 0 \end{cases} \quad (9)$$

其中, $|S|$ 表示样本对总数, $|S_1|$ 表示相似样本对总数, $|S_0|$ 表示不相似样本对总数。

将式(8)代入到式(7), 即可得到最终的损失函数:

$$L_{\text{WeightedPair}} = -\sum_{s_{ij} \in S} w_{ij} (\log(1 + \exp(\alpha \langle b_i, b_j \rangle)) - \alpha s_{ij} \langle b_i, b_j \rangle) \quad (10)$$

3.3 量化约束

哈希编码由离散值构成, 由于离散问题不可微, 深度神经网络无法通过反向传播直接学习到图像对应的哈希编码。因此, 大部分的深度哈希方法将此问题拆解为两步。首先, 将哈希层的 k 维特征松弛到实数域中, 生成哈希编码的近似连续编码 Z , 再通过阶跃函数 $\text{sgn}(Z)$ 对近似值进行阈值约束, 从而获得离散的哈希编码。再引入量化损失 $|B - Z|$, 控制量化误差。

$$B = \text{sgn}(Z) = \begin{cases} +1, & Z \geq 0 \\ -1, & Z < 0 \end{cases} \quad (11)$$

本文则采用了 HashNet 中的设置, 利用双曲正切函数, 在训练过程中逐渐减小平滑量, 将一个离散化问题转换为极限情况下的连续性问题。双曲正切函数的性质如下:

$$\lim_{\beta \rightarrow \infty} \tanh(\beta z) = \text{sgn}(z) \quad (12)$$

在训练过程中将 β 逐渐增大, 逐渐将双曲正切函数逼近到阶跃函数, 将离散问题转换为可微的连续问题。

4 实验与分析

4.1 实验环境

本次实验在一台 GPU 服务器上进行, 基于 CUDA 和 cudnn 进行 GPU 加速, 实际实验中使用了 4 块 GPU, 机器的配置如表 1 所列。

表 1 实验机器配置信息

Table 1 Configuration information of experimental machine

OS	Ubuntu 16.04.3 LTS
CPU	Intel(R) Xeon(R) Silver 4114 CPU @ 2.20 GHz
Memory	128 GB
GPU	4 × GTX1080ti
CUDAversion	9.0
Cudnnversion	7.6.5

本次实验采用了 Anaconda 4.9.2 进行包管理, 使用 Pytorch 1.1.0 实现了所提模型。

4.2 数据集及评价指标

本文算法采用 ChestX-ray14 数据集进行实验评估。ChestX-ray14^[25] 数据集总共包含 112120 张正面 X 射线图片, 这些图像来自 30805 位患者。该数据集属于多标签数据集, 包含了 14 类常见的胸腔病变图像和 1 类无病变图像。其中,

51 708 幅图片包含多个标签。本文将以 1:3:3 的比例将数据集中的数据随机划分为检索数据集、训练集和检索数据库。

本文采用了多标签哈希检索常用的两个评价指标:平均精度均值(Mean Average Precision, MAP)及归一化折损累计增益(Normalized Discounted Cumulative Gain)。

$$MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q} \quad (13)$$

其中, Q 代表查询数据的总数, $AveP(q)$ 代表数据在不同召回率下相加的平均精度。

$$nDCG_p = \frac{DCG_p}{IDCG_p} = \frac{\sum_{i=1}^P \frac{2^{rel_i} - 1}{\log_2(i+1)}}{\sum_{i=1}^{|REL|} \frac{2^{rel_i} - 1}{\log_2(i+1)}} \quad (14)$$

其中, P 代表前 P 个位置, rel_i 表示第 i 个位置样本的相关度等级, 可以将多标签数据集中的标签交集长度用于定义 rel_i 指标。|REL| 表示按照相关度从大到小的顺序排序, 取前 P 个样本组成的集合, 即排序最优集合。

为了更好地从局部和整体的角度对检索性能进行评估, 本文选用 $MAP@10$ 和 $NDCG@ALL$ 作为评价指标。

4.3 对比方法

为了验证本文算法的有效性, 本文选择了 8 种先进的深度哈希算法进行对比, 包括 DSH (Deep Supervised Hashing)^[26], DCH, DFH (Deep Fisher Hashing)^[27], DPN (Deep Polarized Network)^[28], DPSH (Deep Supervised Hashing with Pairwise Labels)^[29], ISDH (Improved Soft pairwise similarity Deep Hashing)^[30], DTSH (Deep Triplet Supervised Hashing)^[31] 和 LCDSH (Locality-Constrained Deep Supervised Hashing)^[32]。

4.4 实验设置

为了充分利用 FCN 的性能, 本文首先对 FCN 进行预训练, 避免生成与分类任务无关的热力图。待 FCN 的训练收敛后, 再对深度哈希网络进行训练。

为了充分挖掘分类任务与检索任务间的关系, 首先训练一个成熟的分类网络, 再将分类网络中的参数作为 FCN 的初始值。本文对 FCN 网络采用自适应矩估计 (Adam) 优化器, 初始学习率设置为 10^{-4} , 一阶矩估计的指数衰减率 β_1 设置为 0.9, 二阶矩估计的指数衰减率 β_2 设置为 0.999, ϵ 设置为 10^{-8} 。

本文对深度哈希网络采用随机梯度下降 SGD 优化器, 初始学习率的选择范围为 $10^{-5} \sim 10^{-3}$, 更新乘法步长设为 5。对多个学习率进行验证并选择最优结果的学习率。本文的初始学习率设置为 0.000 1, 哈希生成全连接层的学习率是特征层的 10 倍, $momentum$ 设置为 0.9, $batchsize$ 为 32, 训练迭代总次数为 10 000, 权重衰减参数设置为 0.000 5。

为了实现公平的比较, 本文方法和所有对比方法在数据集上都采用相同的数据集划分策略和相同的预处理策略。在网络架构上, 所有对比方法都采用了与本文一致的 ResNet50 网络结构作为骨架网络。在训练策略上, 所有对比方法的训练迭代总次数均为 10 000。在训练参数上, 均采用原始参考文献中建议的最优参数作为训练。

4.5 结果分析

4.5.1 对比实验

表 2 列出了不同算法在 ChestX-ray14 数据集上不同码位数下的 $MAP@10$ 的对比结果。本文算法明显优于所有对比算法, 可以证明其取得了良好的检索效果。从结果中可以看出, 本文算法在 ChestX-ray14 数据集上的 16 bit, 32 bit, 48 bit, 64 bit 哈希编码的 $MAP@10$ 分别比对比算法中最好的结果高 4.3%, 3.4%, 11.7%, 11.6%。不同于对比算法, 本文算法随着码长的增加, $MAP@10$ 值也随之上升, 体现了码长与其包含的信息量的一致性。

表 2 ChestX-ray14 数据集上本文方法与其他方法在检索任务上的 $mAP@10$ 比较

Table 2 Comparison of $mAP@10$ between the proposed method and other methods in retrieval task on ChestX-ray14 dataset

Algorithm	Number of hash coding bits/bits			
	16	32	48	64
DSH	0.401	0.313	0.434	0.485
DCH	0.473	0.481	0.482	0.455
DFH	0.531	0.543	0.504	0.517
DPN	0.486	0.405	0.464	0.473
DPSH	0.428	0.534	0.469	0.515
ISDH	0.536	0.491	0.472	0.473
DTSH	0.424	0.487	0.375	0.509
LCDSH	0.437	0.417	0.421	0.405
本文方法	0.579	0.597	0.621	0.633

在对比算法中, DSH 算法和 LCDSH 算法没有对量化损失进行约束, 明显弱于在学习过程中考虑了量化损失的其他对比算法, 说明了对量化损失进行约束的必要性。DTSH 算法、DPSH 算法和 DFH 算法则着重于挖掘样本对之间的关系, 利用二元/三元组损失生成哈希编码, 使得网络能学习到样本对间的差异, 明显优于没有充分挖掘样本差异的对比算法。而 ISDH 则着重于多标签信息的挖掘, 在 16 bit 码长下获得了对比方法中最好的性能。

表 3 列出了不同算法在 ChestX-ray14 数据集上 48 bits 的 $nDCG@ALL$ 的对比结果。从结果中可以看出, 本文算法明显优于所有对比算法, 可以证明本文算法在全局检索范围内取得了良好的检索效果。

在对比算法中, 引入了多标签损失的 ISDH 算法的检索效果明显优于其他没有考虑多标签损失的算法。

表 3 ChestX-ray14 数据集上本文方法与其他方法在检索任务上的 $nDCG@ALL$ 比较

Table 3 Comparison of $nDCG@ALL$ between the proposed method and other methods on ChestX-ray14 dataset at 48 bits

Algorithm	$nDCG@ALL$
DSH	0.2779
DCH	0.3018
DFH	0.3025
DPN	0.2752
DPSH	0.3064
ISDH	0.3072
DTSH	0.2927
LCDSH	0.2953
本文方法	0.3147

4.5.2 消融实验

为了验证本文提出的基于响应的 FCN 空间注意力模块的有效性,本文采用了 3 种对比方法进行消融实验。表 4 列出了消融实验的对比结果。本文主要设置了 3 个对比项: 1) 本文-A, 去除注意力模块; 2) 本文-A+SE, 引入基于通道域的注意力机制 SEBlock; 3) 本文-A+CBAM, 引入基于空间通道域的注意力机制 CBAMBlock。

通过对比本文方法和对比项 1), 本文提出的基于 FCN 响应的注意力机制对 MAP 值有显著提升。通过对比对比项 1) 一对比项 3), 引入空间域和通道域的注意力机制后, MAP 值有了显著提升, 这验证了注意力机制能够关注重要的特征并抑制不重要特征, 提高了哈希编码的特征表达能力。通过对比本文方法、对比项 2) 以及对比项 3), 基于响应的 FCN 空间注意力模块在 MAP 值的提升上明显优于基于特征图的注意力模块, 这验证了在同样采用注意力机制的情况下, 本文算法在注意力机制上的优越性。

表 4 ChestX-ray14 数据集上本文方法与消融实验在检索任务上的 $mAP@10$ 比较

Table 4 Comparison of $mAP@10$ between the proposed method and ablation experiments on ChestX-ray14 dataset

Algorithm	Number of hash coding bits/bits			
	16	32	48	64
本文方法	0.579	0.597	0.621	0.633
本文-A	0.375	0.385	0.405	0.377
本文-A+SE	0.388	0.405	0.576	0.416
本文-A+CBAM	0.403	0.388	0.458	0.449

结束语 本文提出了一种结合了注意力机制的医学影像深度哈希检索方法。本文选用贝叶斯框架, 对深度哈希问题进行建模, 通过分类网络的引导, 利用加权梯度类激活映射生成包含语义信息的局部特征描述子, 同时使用全局特征子和局部特征子作为哈希网络的输入, 引导哈希网络从全局和局部的角度出发, 关注重要特征区域, 抑制非重要区域, 增强了哈希编码的特征表达能力, 减少了哈希编码的冗余。本文为基于注意力机制的医学影像深度哈希方法提供了一套新的解决框架, 从分类任务驱动的角度出发, 有效地提升了检索任务的性能, 对于大规模可解释的医学图像检索具有一定的实际应用意义。

参考文献

- [1] 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. Red Hook, USA: Curran Associates Inc., 2012; 1097-1105.
- [2] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C] // 2015 Medical Image Computing and Computer Assisted Intervention 2015. Cham, Switzerland: Springer, 2015; 234-241.
- [3] LI Y, JIANG H N, WANG H B, et al. Generating GF-3 orthophoto image based on indirect rectification[J]. Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition), 2020, 32(3): 441-451.
- [4] WANG J D, ZHANG T K, SONG J K, et al. A survey on learning to hash[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 769-790.
- [5] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention[C] // Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2014; 2204-2212.
- [6] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE Press, 2016; 770-778.
- [7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // 2015 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE Press, 2015; 3431-3440.
- [8] XIA R K, PAN Y, LAI H J, et al. Supervised hashing for image retrieval via image representation learning[C] // Proceedings of the 28th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2014; 2156-2162.
- [9] LAI H J, PAN Y, LIU Y, et al. Simultaneous feature learning and hash coding with deep neural networks[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE Press, 2015; 3270-3278.
- [10] ZHU H, LONG M S, WANG J M, et al. Deep hashing network for efficient similarity retrieval[C] // Proceedings of the 30th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2016; 2415-2421.
- [11] CAO Z J, LONG M S, WANG J M, et al. HashNet: Deep learning to hash by continuation[C] // 2017 IEEE International Conference on Computer Vision. Piscataway, USA: IEEE Press, 2017; 5609-5618.
- [12] CAO Y, LONG M S, LIU B, et al. Deep cauchy hashing for hamming space retrieval[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE Press, 2018; 1229-1237.
- [13] YUAN L, WANG T, ZHANG X P, et al. Central similarity quantization for efficient image and video retrieval[C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE Press, 2020; 3080-3089.
- [14] WANG R K, WANG R P, QIAO S S, et al. Deep position-aware hashing for semantic continuous image retrieval[C] // 2020 IEEE Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE Press, 2020; 2482-2491.
- [15] CAI Y H, LI Y Y, QIU C Y, et al. Medical image retrieval based on convolutional neural network and supervised hashing[J]. IEEE Access, 2019, 7: 51877-51885. <https://ieeexplore.ieee.org/document/8692349>.
- [16] PENG T Y, BOXBERG M, WEICHERT W, et al. Multi-task learning of a deep k-nearest neighbour network for histopathological image classification and retrieval[C] // 2019 Medical Image Computing and Computer Assisted Intervention. Cham, Switzerland: Springer, 2019; 676-684.
- [17] CHEN Z X, CAI R J, LU J W, et al. Order-sensitive deep hashing for multimorbidity medical image retrieval[C] // 2018 Me-

- dical Image Computing and Computer Assisted Intervention. Cham, Switzerland; Springer, 2018; 620-628.
- [18] ZHENG Y S, JIANG B N, SHI J, et al. Encoding histopathological WSIs using GNN for scalable diagnostically relevant regions retrieval[C] // 2019 Medical Image Computing and Computer Assisted Intervention. Cham, Switzerland; Springer, 2019; 550-558.
- [19] SILVA W, POELLINGER A, CARDOSO J S, et al. Interpretability-guided content-based medical image retrieval[C] // 23th Medical Image Computing and Computer Assisted Intervention. Cham, Switzerland; Springer, 2020; 305-314.
- [20] FANG J S, FU H Z, LIU J. Deep triplet hashing network for case-based medical image retrieval[J]. Medical Image Analysis, 2021, 69; 101981. <https://www.sciencedirect.com/science/article/pii/S136184152100027X>.
- [21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA; IEEE Press, 2018; 7132-7141.
- [22] WOO S, PARK J, LEE J, et al. CBAM; Convolutional block attention module[C] // Proceedings of the 15th European Conference on Computer Vision. Piscataway, USA; Springer, 2018; 3-19.
- [23] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C] // 2017 IEEE International Conference on Computer Vision. Piscataway, USA; IEEE Press, 2017; 618-626.
- [24] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1); 62-66.
- [25] RAJPURKAR P, IRVIN J, ZHU K L, et al. CheXNet: Radiologist-level pneumonia detection on chest X-Rays with deep learning[EB/OL]. (2019-06-12) [2021-08-01]. <http://arxiv.org/abs/1711.05225>.
- [26] LIU H M, WANG R P, SHAN S G, et al. Deep supervised hashing for fast image retrieval[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA; IEEE Press, 2016; 2064-2072.
- [27] LI Y Q, PEI W J, ZHA Y F, et al. Push for quantization: Deep fisher hashing[C] // 30th British Machine Vision Conference. Cardiff, United Kingdom; BMVC, 2019; 1-12.
- [28] FAN L X, NG K W, JU C, et al. Deep polarized network for supervised learning of accurate binary hashing codes[C] // Proceedings of the 29th International Joint Conference on Artificial Intelligence. Palo Alto, USA; AAAI Press, 2020; 825-831.
- [29] LI W J, WANG S, KANG W C. Feature learning based deep supervised hashing with pairwise labels[C] // Proceedings of the 25th International Joint Conference on Artificial Intelligence. Palo Alto, USA; AAAI Press, 2016; 1711-1717.
- [30] ZHANG Z, ZOU Q, LIN Y W, et al. Improved deep hashing with soft pairwise similarity for multi-label image retrieval[J]. IEEE Transactions on Multimedia, 2020, 22(2); 540-553.
- [31] WANG X F, SHI Y, KITANI K M. Deep supervised hashing with triplet labels[C] // 13th Asian Conference on Computer Vision. Cham, Switzerland; Springer, 2017; 70-84.
- [32] ZHU H, GAO S H. Locality constrained deep supervised hashing for image retrieval[C] // Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto, USA; AAAI Press, 2017; 3567-3573.



ZHU Cheng-zhang, born in 1978, Ph.D., associate professor, master supervisor. Her main research interests include pattern recognition, computer vision, and image processing.



XIAO Ya-long, born in 1985, Ph.D., associate professor, is a member of China Computer Federation. His main research interests include wireless sensing and computational communication.

(责任编辑:喻黎)