

# 基于注意力机制和软匹配的多标签遥感图像检索方法

张永梅 徐敏 李小冬

(北方工业大学信息学院 北京 100144)

**摘要** 针对卷积神经网络对于多标签遥感图像特征提取能力弱、不能准确反映遥感图像多标签复杂性的问题,提出基于注意力机制和软匹配的多标签遥感图像检索方法。在特征提取阶段,以密集卷积神经网络模型为基础,在每个密集块(Dense Block)后添加CBAM(Convolutional Block Attention Module)层,实现对多标签图像区域特征提取。在模型训练时,利用区分硬匹配与软匹配的联合损失函数,学习图像的哈希编码表示。通过评估遥感图像哈希编码间的汉明距离,实现相似图像的检索。实验结果表明,所提方法在数据集NUS-WIDE和多标签遥感图像数据集DLRSD上与其他基于全局特征的深度哈希方法相比,明显提升了检索准确率。

**关键词** 遥感图像检索 密集卷积神经网络 深度哈希 多标签 软匹配

中图分类号 TP751 TP3 文献标志码 A DOI:10.3969/j.issn.1000-386x.2024.06.027

## A MULTI-LABEL REMOTE SENSING IMAGE RETRIEVAL METHOD BASED ON ATTENTION MECHANISM AND SOFT MATCHING

Zhang Yongmei Xu Min Li Xiaodong

(School of Information Science and Technology, North China University of Technology, Beijing 100144, China)

**Abstract** To address the problems that convolutional neural networks are weak in extracting features of multi-label remote sensing images and the reflection of complexity multiple labels in remote sensing images, a multi-label remote sensing image retrieval method based on attention mechanism and soft matching is proposed. In the feature extraction stage, the method was based on the densely connected convolutional neural networks, and a CBAM (Convolutional Block Attention Module) layer was added after each dense block to achieve feature extraction of multi-label image regions. During model training, the joint loss function that distinguished hard matching and soft matching was used to learn the Hash code representation of the images. The retrieval results were obtained by evaluating the Hamming distance between the image Hash code and the retrieved image Hash code. The experimental results show the proposed method has a significant improvement in retrieval accuracy compared with other deep hashing methods based on global features on the universal dataset NUS-WIDE and multi-label remote sensing image dataset DLRSD.

**Keywords** Remote sensing image retrieval Densely connected convolutional neural networks Deep hashing Multi-label Soft matching

## 0 引言

近年来,卫星技术的发展使得高分辨率遥感图像数量激增的同时,也给遥感图像检索带来了巨大考验。如何对大规模遥感图像进行快速检索,学者们进行了大量研究。基于内容的图像检索是目前最为主流的图

像检索技术,包括图像的特征表示和相似性度量两部分<sup>[1]</sup>。

早期的基于内容的图像检索技术主要通过尺度不变特征变换、局部二值模式、方向梯度直方图等方法提取图像的颜色、纹理等低层特征。在相似性判定时,利用近似最近邻算法来进行相似度比对。由于这些方法提取到的低层特征往往难以表达图像的高层语义信

息,对图像的特征能力较弱,容易产生“语义鸿沟”问题,且提取的特征维度大、检索速度慢。

为了解决大规模图像检索时,早期基于内容的图像检索技术导致的特征维度高、检索慢的问题。哈希方法逐渐被用于图像检索领域,通过将高维特征转化为低维的二进制编码,使得编码在保留图像高层语义信息的同时,相似图像哈希编码的汉明距离尽可能小,这样可以大大减少存储开销,并缩短检索时间。传统的哈希检索方法主要通过改进哈希函数、损失函数以及相似性度量方法等来提高检索精度,但由于对特征的提取能力较弱,难以表达复杂的语义信息,使得检索效果并不理想。随着以卷积神经网络为主要代表的深度学习模型在图像领域取得的一系列突破,深度哈希检索方法也逐渐发展起来。

## 1 相关研究工作

深度哈希检索方法包括无监督哈希方法和有监督哈希方法。常见的无监督哈希方法包括位置敏感哈希(Local Sensitive Hashing, LSH)<sup>[2]</sup>、谱哈希(Spectral Hashing, SH)<sup>[3]</sup>、迭代量化哈希(Iterative Quantization, ITQ)<sup>[4]</sup>等。无监督哈希方法无须使用标签数据,利用多个随机函数构成函数簇,将数据的特征映射为与哈希编码维度相同的特征向量,最后通过阈值化获得对应的哈希编码。

有监督哈希方法通过分析训练数据与对应的标签信息来构造投影函数。特别是基于深度学习的哈希检索方法的提出,使得深度哈希方法成为目前最主要的图像检索方法。深度哈希方法的大致思路是利用卷积神经网络(Convolutional Neural Networks, CNN)、AlexNet、ResNet(Residual Network)等深度学习模型提取图像的高层语义特征,对图像的特征进行相似度度量,得到最终的检索结果。卷积神经网络哈希(Convolutional Neural Network Hashing, CNNH)方法<sup>[5]</sup>最先将深度学习运用于哈希,基于迁移学习,将训练好的网络模型迁移到高分辨率遥感图像上,利用迁移后的模型提取图像特征,阈值化后得到哈希编码。随后出现的DSDH(Deep Supervised Discrete Hashing)<sup>[6]</sup>、DPSH(Deep Pairwise-Supervised Hashing)<sup>[7]</sup>等通过端到端的方式,使得模型可以直接对输入图像进行哈希编码的学习,大大加快了图像检索模型的训练速度。

虽然上述方法在图像检索方面具有一定的效果,但检索目标对象均为单标签图像,对于多标签图像的检索,相关的研究很少。Liu<sup>[8]</sup>等提出的直接二进制嵌入(Direct binary embedding, BDE)方法,基于ResNet模

型利用分类任务训练生成哈希编码,完成对多标签图像的检索。深度多标签哈希(Deep Multi-Label Hashing, DMLH)方法<sup>[9]</sup>基于AlexNet模型通过使用样本权重交叉熵损失函数来实现端到端的多标签图像检索。

为充分利用多标签遥感图像的区域特征与不同组图像间的相似关系,本文提出了一种基于注意力机制和软匹配的多标签遥感图像检索方法。特征提取阶段,以密集卷积神经网络为基础,结合注意力机制提取图像深层语义特征。图像相似性判定时,摒弃原本粗略的图像相似性定义,采用基于余弦距离的软匹配方式来定义图像间的相似度,并通过联合损失函数学习图像的哈希编码。在多标签数据集NUS-WIDE和遥感图像数据集DLRSD上分别进行实验,实验结果表明结合注意力机制与软匹配的检索方法在检索准确率上与其他方法相比具有一定的提升。

## 2 问题定义

在监督哈希方法中,对于输入数据 $X = \{x_i\}_{i=1}^N$ ,标签数据 $Y = \{y_i\}_{i=1}^N \in \mathbf{R}^{c \times N}$ , $x_i$ 对应的标签信息为 $y_i \in \{0, 1\}^c$ , $c$ 表示类别的数量,一个样本可能同时对应几个标签。深度哈希的目的就是学习输入数据 $X$ 的哈希编码 $B \in \{-1, 1\}^{K \times N}$ ,其中第 $i$ 列向量 $b_i \in \{-1, 1\}^K$ 表示第 $i$ 个样本 $x_i$ 的哈希码, $K$ 表示哈希编码的位数。

对于给定的语义标签信息,可以得到成对样本 $x_i$ 、 $x_j$ 的相似关系 $S = \{S_{ij}\}$ 。在以往的多标签检索中,使用的是较为粗略的相似性判定标准,也就是说,对于样本 $x_i$ 、 $x_j$ ,只要对应的 $y_i$ 、 $y_j$ 存在相同位置的值为1,就代表 $x_i$ 、 $x_j$ 相似,将 $S_{ij}$ 的值设为1,否则为0,我们称这种方式称为硬匹配相似。Zhang等<sup>[10]</sup>利用成对标签的余弦距离作为 $S_{ij}$ 的值,可以在一定程度上对样本之间的相似性做更细致的区分。

## 3 多标签遥感图像检索方法

本文所提出的检索方法如图1所示,可分为特征提取、哈希编码以及相似度度量三部分。本文方法通过基于DenseNet的网络模型结合CBAM提取遥感图像的深层特征,利用两层全连接层对图像进行二进制编码,并通过构造损失函数来量化误差,完成对哈希码的学习。检索时,将被检索图像的二进制编码与数据库中的编码进行相似度比对,选取汉明距离最小的前 $N$ 项作为检索结果。

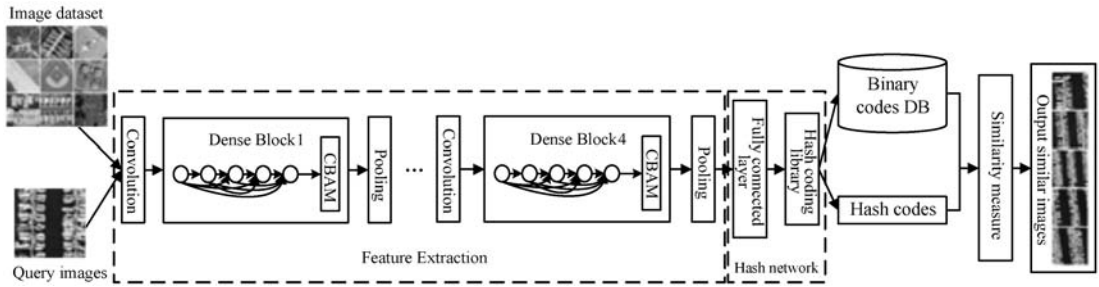


图 1 本文方法实现框架

### 3.1 特征提取方法

#### 3.1.1 DenseNet 模型

DenseNet<sup>[11]</sup>是基于 CNN 的具有密集连接结构的神经网络模型,通过特征重用的方式来提升模型性能。该模型主要由卷积层、池化层以及密集块 (Dense Block) 组成。在密集块中,每层将其所有前部卷积层的输出作为输入,使得任意两层可以直接联系,即第  $N$  层将会接收所有来自该密集块的前部层。DenseNet 结构可以很好地缓解梯度消失与梯度爆炸的问题,而且省参数、省计算、具有较好的抗过拟合效果。

#### 3.1.2 引入 CBAM 的 DenseNet 模型

卷积块的注意力机制 (Convolutional Block Attention Module, CBAM)<sup>[12]</sup>能同时结合空间与通道的注意力模块,解决在通道和空间维度上关注的问题。

通道注意力模块将输入的特征  $F$  同时进行最大池化与平均池化,对池化后得到的特征分别经过多层感知机 (Multilayer Perceptron, MLP),将 MLP 输出的两组特征进行加和,经过 sigmoid 后得到通道注意力的权

重系数。

$$M_c(F) = \sigma(MLP(AvgPool(F))) + MLP(MaxPool(F)) \quad (1)$$

将原始特征  $F$  与权重  $M_c(F)$  相乘,得到新的特征  $F' = M_c(F) \otimes F$ 。

空间注意力模块主要关注目标的位置信息。空间注意力模块首先在通道维度上对输入特征  $F'$  采用最大池化与平均池化,将两个通道拼接起来,利用卷积操作得到空间注意力的权重系数。

$$M_s(F') = \sigma(f^{7 \times 7}[AvgPool(F'); MaxPool(F')]) \quad (2)$$

式中:  $\sigma$  表示 sigmoid 激活函数;  $f^{7 \times 7}$  表示卷积操作,并且过滤器的大小为  $7 \times 7$ 。最后将输入特征  $F'$  与权重系数  $M_s(F')$  相乘,得到新的特征  $F'' = M_s(F') \otimes F'$ 。

本文提出结合 CBAM 的 DenseNet 网络模型,在原始 DenseNet 的基础上,在每个 Dense Block 的后面添加 CBAM 层,实现利用 DenseNet 网络模型的同时结合通道与空间注意力机制。改进后的 Dense Block 结构如图 2 所示。

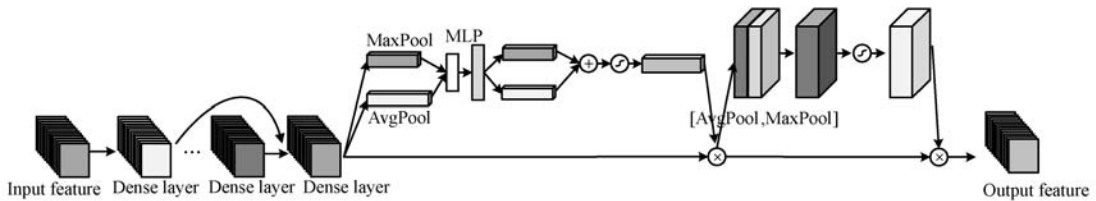


图 2 改进的 Dense Block 结构

### 3.2 哈希编码

哈希编码由两个全连接层组成,最后一层全连接层包含有  $k$  个神经元,  $k$  表示哈希编码的维度。对于经过基于 DenseNet 模型生成的特征向量  $x$ ,深度哈希网络可以表示为:

$$h(x) = \tan(W_2 \cdot \tan(W_1 x + bias)) \quad (3)$$

$$b = \text{sgn}(h(x)) = \begin{cases} 1 & h(x) > 0 \\ -1 & h(x) \leq 0 \end{cases} \quad (4)$$

式中:  $W_1$ 、 $W_2$  表示全连接层的权值;  $\text{sgn}$  表示符号函数。

### 3.3 结合软匹配相似性的损失函数

对于给定的语义标签信息  $Y = \{y_i\}_{i=1}^N \in \mathbf{R}^{c \times N}$ ,可

以得到成对的标签信息的相似关系  $S = \{S_{ij}\}$ ,  $S_{ij} \in [0, 1]$ 。在现有方法中,如果  $y_i$  与  $y_j$  存在共同标签,则  $S_{ij}$  为 1,表示  $y_i$  与  $y_j$  相似;否则,  $S_{ij}$  为 0,表示  $y_i$  与  $y_j$  不相似。本文方法将图像间相似关系  $S_{ij}$  表示为标签  $y_i$  与  $y_j$  的余弦距离。

$$S_{ij} = \frac{\langle y_i, y_j \rangle}{\|y_i\| \|y_j\|} \quad (5)$$

式中:  $\langle y_i, y_j \rangle$  表示内积操作,  $\|\cdot\|$  表示取模操作。当  $S_{ij}$  的值为 0 时,表明样本标签  $y_i$  与  $y_j$  之间不存在相似性,即  $y_i$  与  $y_j$  之间不存在任何共享的类别标签。  $S_{ij}$  的值为 1 时,表明样本标签  $y_i$  与  $y_j$  完全相同。当  $S_{ij}$  的值介于 0 到 1 之间时,值越大,标签  $y_i$  与  $y_j$  之间相似度越

大。当  $S_{ij}$  的值为 0 或者 1 时, 本文认为样本为硬匹配相似, 否则, 称为软匹配相似。

对硬匹配相似的样本, 采用 Gui 等<sup>[6]</sup>提出的 FSDH 方法的损失函数, 基于相似关系  $S = \{S_{ij}\}$ , 哈希编码的最大似然估计可以表示为:

$$p(B|S) \propto p(S|B)p(B) = \prod_{s_{ij} \in S} p(s_{ij}|B)p(B) \quad (6)$$

式中:  $p(B|S)$  表示似然函数;  $p(B)$  为先验分布;  $p(s_{ij}|B)$  表示的是给定哈希码  $B$  时,  $s_{ij}$  的条件概率, 其定义如下:

$$p(s_{ij}|B) = \begin{cases} \sigma(\Phi_{ij}) & s_{ij} = 1 \\ 1 - \sigma(\Phi_{ij}) & s_{ij} = 0 \end{cases} \quad (7)$$

式中:  $\sigma(x) = \frac{1}{1 + e^{-x}}$  为激活函数;  $\Phi_{ij} = \frac{1}{2} \langle b_i, b_j \rangle =$

$\frac{1}{2} b_i^T b_j$ 。损失函数为:

$$F = - \sum_{s_{ij} \in S} (s_{ij} \Psi_{ij} - \log(1 + e^{\Psi_{ij}})) \quad (8)$$

式中:  $\Psi_{ij} = \frac{1}{2} h_i^T h_j$ ,  $h_i$  ( $i = 1, 2, \dots, N$ ) 是模型最后一层的输出。

对于软匹配相似性的样本对, 损失函数为:

$$F_s = \sum_{s_{ij} \in S} \left( \frac{\langle h_i, h_j \rangle + q}{2} - s_{ij} \cdot q \right)^2 \quad (9)$$

式中:  $q$  是编码的位数。最终, 联合损失函数为:

$$L = \sum_{s_{ij} \in S} (M_{ij} (\log(1 + e^{\Psi_{ij}}) - s_{ij} \Psi_{ij}) + \gamma (1 - M_{ij}) \left( \frac{\langle h_i, h_j \rangle + q}{2} - s_{ij} \cdot q \right)^2) \quad (10)$$

式中:  $M_{ij} = 1$  表示硬匹配相似;  $M_{ij} = 0$  表示为软匹配相似;  $\gamma$  为超参数。

## 4 实验

### 4.1 实验数据

为验证本文方法的有效性, 将本文方法分别在多标签数据集 NUS-WIDE 和多标签遥感图像数据集 DLRSD 上与其他方法进行了对比实验。

NUS-WIDE 数据集包含 269 648 幅多标签图像, 共 5 018 个标签, 实验时采用包含最高频的 21 类标签的子数据集, 共 19 万余幅图像, 其中每个标签至少关联 5 000 幅图像。DLRSD 数据集是在单标签遥感图像数据集 UC-Merced 的基础上由 Chaudhuri<sup>[13]</sup>等制作。该数据集包含 2 100 幅高分辨率遥感图像, 包含飞机、裸土、建筑物、汽车等 17 类标签。

### 4.2 实验环境和参数设置

实验过程中使用的深度学习开发环境为 Pytorch

1.4.0, 操作系统为 Ubuntu 18.04, GPU 加速显卡为 NVIDIA GeForce RTX 2080Ti。为加快训练速度, 实验使用的 DenseNet 模型的主体结构基于 DenseNet121, 训练时学习率设置为 0.001, 权重衰减设置为  $1e-5$ , batch\_size 设置为 64, 训练时 epoch 值设置为 150, 超参数  $\gamma$  设为 0.001。

对 NUS-WIDE 数据集进行训练时, 从 21 个最常见的标签中随机选取 2 100 幅图像作为查询图像 (每类 100 幅), 其余的图像作为训练样本。对 DLRSD 数据集进行训练时, 随机选取 30% 作为测试集, 其余 70% 作为训练集。实验过程中, 对图像的哈希编码长度分别设置为 12、24、32 和 48。

### 4.3 实验对比结果及分析

将本文方法与多种传统方法和深度哈希方法进行了比较, 对 NUS-WIDE 数据集检索时选取基于传统方法的无监督哈希方法 SH、ITQ 和有监督学习方法 SPLH (Sequential Projection Learning for Hashing)<sup>[14]</sup>、KSH (Supervised Hashing with Kernels)<sup>[15]</sup> 和 SDH (Supervised Discrete Hashing)<sup>[16]</sup> 作为对比, 基于深度的哈希方法选取 CNNH、DPSH、DSDH、IDHN (Improved Deep Hashing) 作为对比。实验采用在信息检索中最为常用的 MAP (Mean Average Precision) 值作为评价指标, 本文方法与其他方法的实验结果如表 1 所示。

表 1 NUS-WIDE 数据集 MAP 指标

| 方法   | 编码长度    |         |         |         |
|------|---------|---------|---------|---------|
|      | 12      | 24      | 32      | 48      |
| SH   | 0.454 2 | 0.406 4 | 0.405 5 | 0.423 5 |
| ITQ  | 0.452 0 | 0.468 0 | 0.472 6 | 0.477 3 |
| SPLH | 0.568 3 | 0.589 2 | 0.597 0 | 0.601 0 |
| KSH  | 0.556 6 | 0.572 0 | 0.581 0 | 0.588 0 |
| SDH  | 0.568 0 | 0.600 0 | 0.608 0 | 0.637 0 |
| CNNH | 0.611 0 | 0.618 0 | 0.625 0 | 0.608 0 |
| DPSH | 0.752 0 | 0.790 0 | 0.794 0 | 0.812 0 |
| IDHN | 0.729 2 | 0.758 5 | 0.763 9 | 0.769 2 |
| DSDH | 0.776 0 | 0.808 0 | 0.820 0 | 0.829 0 |
| Our  | 0.838 0 | 0.840 7 | 0.852 9 | 0.852 9 |

从实验中可以明显看到, 与无监督学习方法相比, 有监督的哈希检索方法在检索效果上具有明显优势。与基于全局特征学习的深度哈希方法相比, 本文方法比 CNNH、DPSH、DSDH 等基于 CNN 的图像检索方法在检索性能上也有一定的提升, 表明本文基于 DenseNet 与注意力机制的方法, 对于多标签图像的特征提取能力与基于 CNN 模型的哈希检索方法相比, 具有明显优势。

在多标签遥感图像数据集 DLRSD 上与基于深度学习的哈希方法 DCH(Deep Cauchy Hashing)<sup>[17]</sup>、DHN(Deep Hashing Network)<sup>[18]</sup>、DTQ(Deep Triplet Quantization)<sup>[19]</sup>、DQN(Deep Quantization Network)<sup>[20]</sup>、MLSIR(Multi-label Remote Sensing Image Retrieval)<sup>[21]</sup>进行对比,实验结果如图3-图4所示,分别表示检索结果的 Top 10 与 Top 100 所对应的 MAP 值。

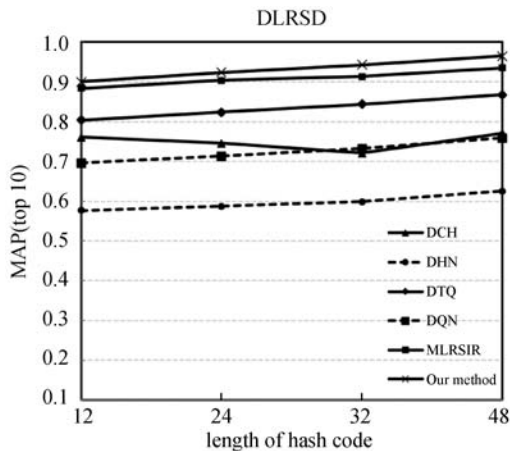


图3 检索样本为10时MAP指标

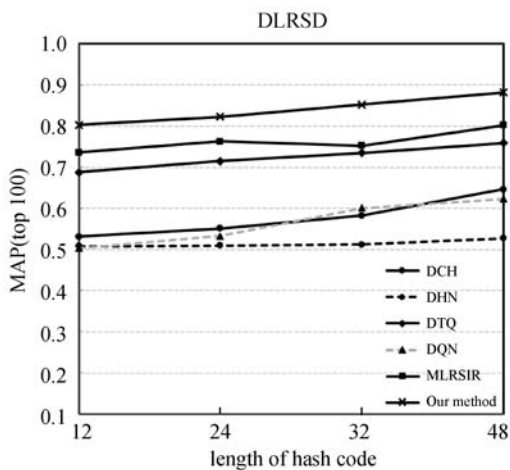


图4 检索样本为100时MAP指标

可以看到,针对多标签遥感图像数据集 DLRSD,在大多数情况下,MAP 准确率会随着哈希编码维度的上升而增加。对于不同的哈希编码长度,DCH 与 DTQ 两种基于批次的哈希方法相比 DHN 具有相对较好的检索性能。本文方法与对比方法相比较,无论在检索结果为10 或者100 的情况下,MAP 性能指标均有明显提升,表明本文方法无论在通用数据集 NUS-WIDE 和多标签遥感图像数据集 DLRSD 上,均能基于 DenseNet 的网络模型有效提取区域特征,提高检索性能。

## 5 结语

本文改进了 DenseNet 模型的原始结构,为充分挖掘多标签遥感图像的高层语义信息,在每个 Dense

Block 后面添加 CBAM 模块用于获取输入图像的通道与空间注意力特征,同时利用区分软匹配与硬匹配数据信息的联合损失函数,能够更好地生成图像的哈希编码。在多标签通用数据集 NUS-WIDE 与遥感图像数据集 DLRSD 上进行实验,实验结果验证了本文方法的可行性,明显提升了检索准确率。由于使用的多标签遥感数据集规模较小,未来将扩充训练数据规模,继续优化模型性能。

## 参考文献

- [1] 王志明,张航.融合多层卷积神经网络特征的快速图像检索方法[J].计算机辅助设计与图形学学报,2019,31(8):1410-1416.
- [2] Chafik S, Daoudi I, Ouardi H E, et al. Locality sensitive hashing for content based image retrieval: A comparative experimental study [C]//International Conference on Next Generation Networks and Services,2014;38-43.
- [3] Weiss Y, Torralba A, Fergus R. Spectral hashing [C]//21st International Conference on Neural Information Processing Systems,2018;1753-1760.
- [4] Gong Y C, Lazebnik S, Gordo A, et al. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013,35(12):2916-2929.
- [5] Xia R K, Pan Y, Lai H J, et al. Supervised hashing for image retrieval via image representation learning [C]//28th AAAI Conference on Artificial Intelligence,2014;2156-2162.
- [6] Gui J, Liu T, Sun Z, et al. Deep supervised discrete hashing[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,27(12):5996-6009.
- [7] Li W J, Wang S, Kang W C. Feature learning based deep supervised hashing with pairwise labels [C]//25th International Joint Conference on Artificial Intelligence,2016;1711-1717.
- [8] Liu L, Rahimpour A, Taalimi A, et al. End-to-end binary representation learning via direct binary embedding [C]//IEEE International Conference on Image Processing,2017:1257-1261.
- [9] Zhong C L, Yu Y, Tang S H, et al. Deep multi-label hashing for large-scale visual search based on semantic graph [J]. Lecture Notes in Computer Science,2017,10366(14):169-184.
- [10] 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.

- 24th ACM International Conference on Multimedia,2016:640 – 644.
- [ 5 ] Sam D B, Babu R V. Top-down feedback for crowd counting convolutional neural network[EB]. arXiv:1807.08881,2018.
- [ 6 ] Liu J, Gao C Q, Meng D Y, et al. DecideNet: Counting varying density crowds through attention guided detection and density estimation[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:5197 – 5206.
- [ 7 ] Valloli V K, Mehta K. W-Net: Reinforced u-net for density map estimation[EB]. arXiv:1903.11249,2019.
- [ 8 ] Li Y H, Zhang X F, Chen D M. CsrNet: Dilated convolutional neural networks for understanding the highly congested scenes[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:1091 – 1100.
- [ 9 ] Cao X K, Wang Z P, Zhao Y, et al. Scale aggregation network for accurate and efficient crowd counting[C]//European Conference on Computer Vision,2018:734 – 750.
- [10] Liu N, Long Y C, Zou C Q, et al. AdcrowdNet: An attention-injective deformable convolutional network for crowd understanding[C]//IEEE Conference on Computer Vision and Pattern Recognition,2019:3225 – 3234.
- [11] Hossain M A, Hosseinzadeh M, Chanda O, et al. Crowd counting using scale-aware attention networks [ C ]//IEEE Winter Conference on Applications of Computer Vision, 2019:1280 – 1288.
- [12] Ma Z H, Wei X, Hong X P, et al. Bayesian loss for crowd count estimation with point supervision[C]//IEEE International Conference on Computer Vision,2019:6142 – 6151.
- [13] Wang B Y, Liu H D, Samaras D, et al. Distribution matching for crowd counting[C]//34th International Conference on Neural Information Processing Systems, 2020: 1595 – 1607.
- [14] Viola P, Jones M J, Snow D. Detecting pedestrians using patterns of motion and appearance[J]. International Journal of Computer Vision,2005,63(2):153 – 161.
- [15] Chan A B, Liang Z S, Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition,2008:1 – 7.
- [16] Cao Z, Hidalgo G, Simon T, et al. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,43(1):172 – 186.
- [17] Zhu L, Zhao Z J, Lu C, et al. Dual path multi-scale fusion networks with attention for crowd counting [EB]. arXiv: 1902.01115,2019.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB]. arXiv:1409.1556,2014.
- [19] Kingma D P, Ba J. Adam: A method for stochastic optimization[EB]. arXiv:1412.6980,2014.
- [20] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition,2015:1 – 9.
- [21] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [ C ]//International Conference on Medical Image Computing and Computer-Assisted Intervention,2015:234 – 241.
- 
- (上接第 185 页)
- [11] Huang G, Liu Z, Laurens V D M, et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition,2017:2261 – 2269.
- [12] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module [ C ]//European Conference on Computer Vision,2018:3 – 19.
- [13] Chaudhuri B, Demir B, Chaudhuri S, et al. Multilabel remote sensing image retrieval using a semi-supervised graph-theoretic method[J]. IEEE Transactions on Geoscience and Remote Sensing,2018,56(2):1144 – 1158.
- [14] Wang J, Kumar S, Chang S F. Sequential projection learning for hashing with compact codes[C]//27th International Conference on International Conference on Machine Learning,2010:1127 – 1134.
- [15] Liu W, Wang J, Ji R, et al. Supervised hashing with kernels[C]//IEEE Conference on Computer Vision and Pattern Recognition,2012:2074 – 2081.
- [16] Shen F M, Shen C H, Liu W, et al. Supervised discrete hashing[C]//IEEE Conference on Computer Vision and Pattern Recognition,2015:37 – 45.
- [17] Cao Y, Long M S, Liu B, et al. Deep Cauchy hashing for hamming space retrieval[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:1229 – 1237.
- [18] Zhu H, Long M S, Wang J M, et al. Deep hashing network for efficient similarity retrieval [ C ]//30th AAAI Conference on Artificial Intelligence,2016:2415 – 2421.
- [19] Liu B, Cao Y, Long M S, et al. Deep triplet quantization [ C ]//ACM Multimedia Conference on Multimedia, 2018: 755 – 763.
- [20] Cao Y, Long M S, Wang J M, et al. Deep quantization network for efficient image retrieval [ C ]//30th AAAI Conference on Artificial Intelligence,2016:3457 – 3463.
- [21] Moustafa M, Ahmed S, Hamed A. Learning to hash with convolutional network for multi-label remote sensing image retrieval[J]. International Journal of Intelligent Engineering and Systems,2020,13(5):539 – 547.