【大规模图像检索的利器】Deep哈希算法介绍

原创 2016-06-20 刘昊淼 深度学习大讲堂

点击上方“深度学习大讲堂”可订阅哦！

深度学习大讲堂致力于推送人工智能，深度学习方面的最新技术，产品以及活动。

前言

在最近邻搜索（nearest neighbor search）问题中，给定一个查询（query），目标是要找到空间中离它最近的点。这里所说的空间可以是任意的空间，比如特征空间，或者语义空间。具体来说，在图像检索这个问题中，每张图像对应空间中的一个点，而所谓的“近”既可以是外观上的近（看着像），也可以是语义上的近（同类）。以下图为例，当我把左侧的图像扔给两个不同的搜索引擎后，得到的返回结果就对应上面的两种情况。

事实上，如果不对效果做什么要求的话，想要实现上面这样的功能其实并不难。最直接的方法就是用一种特征来表示每张图像（比如AlexNet中fc7层的输出），然后通过计算查询图像（上图左）和数

据库中图像在特征空间中的欧式距离，并按照距离从小到大的顺序，返回数据库中的图像。

上面这种方法虽然看上去简单粗暴，但是却不失为一种有效的做法。但是，随着互联网上的图像越来越多，这种方法的短板也体现得淋漓尽致——存储空间消耗大，检索速度慢。具体来说，如果使用上面提到的AlexNet中fc7层的输出来表示每张图像，那么表示一百万张图像就需要大约15GB的存储空间（单精度浮点数），而计算查询图像和数据库中每张图像的距离，则需要8192次加法操作和4096次乘法操作，遍历完所有的一百万张图像再返回结果的话，恐怕用户早就等得不耐烦了。考虑到现在互联网上的数据规模动辄就是上亿的级别，这种方法就显得更不靠谱了。

为了解决上面方法对存储空间和检索时间的不切实际的要求，近年来近似最近邻搜索（approximate nearest neighbor search）技术发展迅猛，因为其对空间和时间的需求大幅降低，而且能够得到不错的检索结果，因此成为了一种实用的替代方案。在这其中，哈希（hashing）作为一种代表性方法，近年来受到了广泛的关注。本文首先对哈希算法的发展历程进行简单的介绍，然后按照相关性，对近年来的一些主要的深度哈希算法进行介绍，最后对现有深度哈希方法进行简单的总结。

发展历程

在哈希算法中，通常的目标是将样本表示成一串固定长度的二值编码（通常使用0/1或-1/+1表示其中的每个bit），使得相似的样本具有相似的二值码（使用Hamming距离度量二值码之间的相似性）。
在最初的工作中，作者提出在特征空间中随机选择一些超平面对空间进行划分，根据样本点落在超平面的哪一侧来决定每个bit的取值。这类方法虽然有严格的理论证明保证其效果，但是在实际操作中通常需要比较多的bit才能得到令人满意的检索效果。

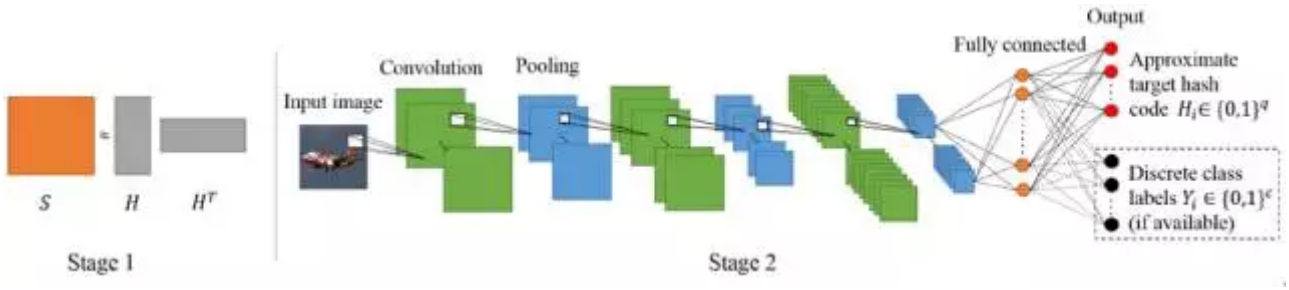
在之后的工作中，为了得到编码长度更短、检索效果更好的二值码，人们进行了很多尝试，包括构建不同的目标函数、采用不同的优化方法、利用图像的标签信息、使用非线性模型等。随着研究的深入，利用二值编码进行检索的性能也逐步提升。

与常见的机器学习算法不同，哈希方法的目标是得到二值编码，所以优化过程中经常会遇到离散取值的约束，因此通常来说无法使用基于梯度的方法对目标函数进行优化。为了简化问题，通常的做法是改用一个更宽松的约束，比如不再要求“二值码”是二值的，而是只要在一个规定的范围中即可。优化结束后，再对松弛过的“二值码”进行量化，得到最终的真二值码，深度哈希算法通常也采用这种做法。

深度哈希算法

最早的基于深度学习的哈希算法应该是2009年由Hinton研究组提出的Semantic Hashing方法[1]。对于这个方法来说，深度模型只是提供了一定的非线性表示能力，而网络的输入仍是手工设计的特征，和现在通常意义上的深度学习算法还是有一定的区别，因此本文中不作具体介绍。在这之后，基于Semantic Hashing出现了一些改进，但是也都没有掀起什么大的风浪，在此一笔带过。

时间来到2014年，受到CNN强大学习能力的鼓舞，中山大学的潘炎老师研究组和颜水成老师合作，在美国人工智能协会年会（AAAI 2014）上发表的论文提出了一种名为CNNH（Convolutional Neural Network Hashing）的方法[2]，把基于CNN的深度哈希算法推到了前台。CNNH的做法如下图所示，首先通过对相似度矩阵（矩阵中的每个元素指示对应的两个样本是否相似）进行分解，得到样本的二值编码；然后，利用CNN对得到的二值编码进行拟合。拟合的过程相当于一个多标签预测问题，作者使用了交叉熵损失来达到这个目的，这一步对应图中最右侧红色节点。此外，作者还提出加入分类的损失函数来进一步提升性能（softmax，对应图中最右侧黑色节点）。



尽管实验中CNNH相比传统的基于手工设计特征的方法取得了显著的性能提升，但是这个方法仍然不是端到端的方法，学到的图像表示不能反作用于二值编码的更新，因此并不能完全发挥出深度学习的能力。为了更好地挖掘深度模型的潜力，在这之后，出现了不少改进方法。

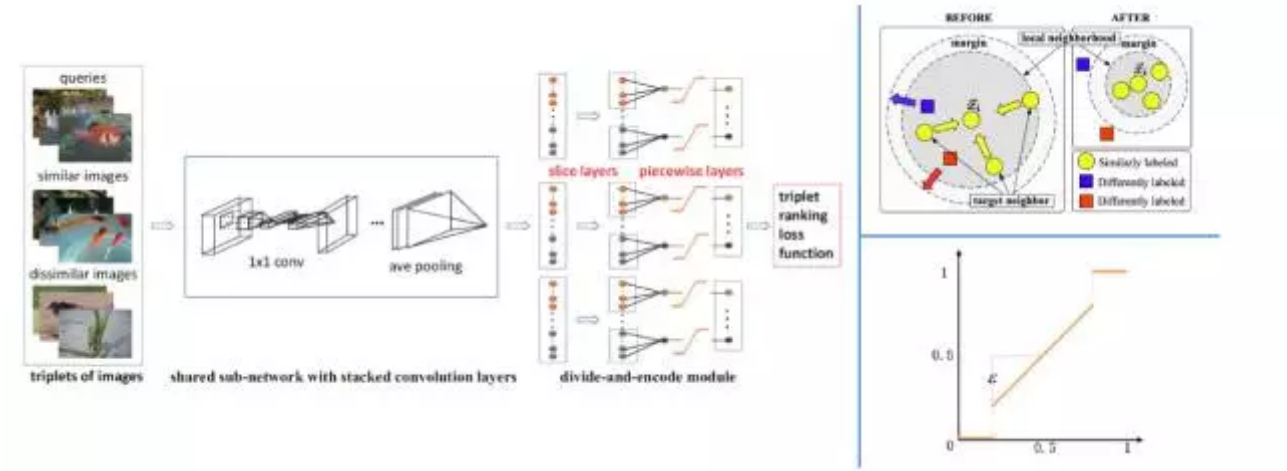
2015年计算机视觉与模式识别会议（CVPR 2015）中，一下子出现了四篇基于深度学习的哈希算法。在这四篇文章之中，其中一篇文章（在此不对这篇文章做详细介绍，有兴趣的同学可以参考[3]）使用手工设计特征作为输入之外，其余的三篇均为完全的端到端模型。下面具体介绍一下这三篇文章。

第一篇文章[4]与上面介绍的CNNH一样，同样是来自中山大学的潘炎老师研究组和颜水成老师。因为这篇文章中使用了一个比CNNH中的网络深得多的Network in Network的网络结构，因此被简称为NINH（NIN Hashing）或DNNH（Deep Neural Network Hashing）。这篇文章的做法如下图所示。网络使用三张图像构成的三元组进行训练。在三元组中，其中的第一张图像和第二张图像是相似的，而第一张图像和第三张图像则是不相似的。基于三元组的损失函数的目标是：在得到的Hamming空间中，相似样本间的距离小于不相似样本间的距离（下图右上）。值得一提的是，这项工作为了适配哈希学习这个任务，在网络结构上做了一些有针对性的设计，包括：

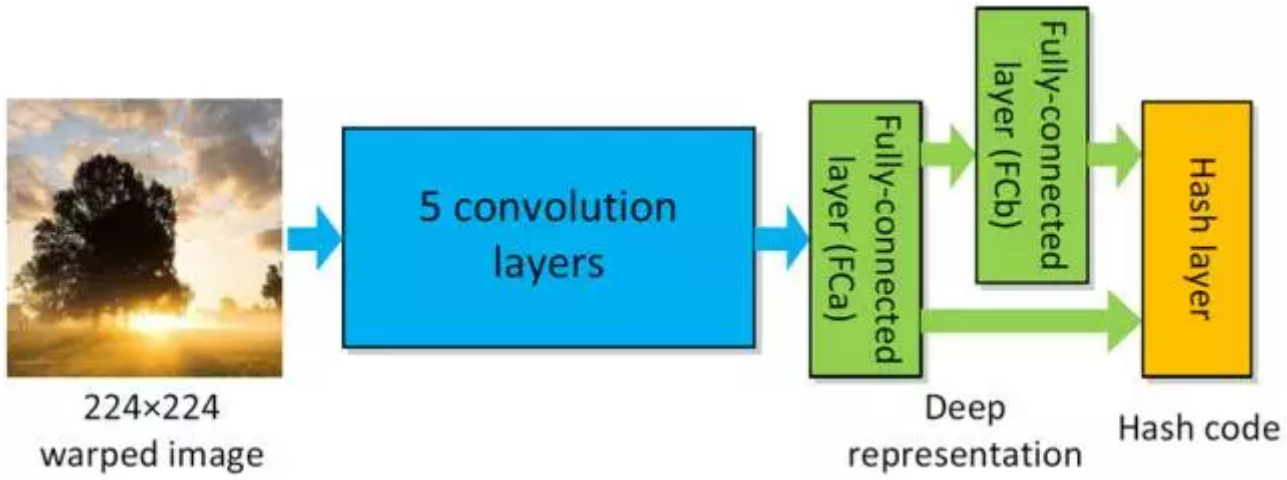
（1）为了减小二值码不同bit之间的冗余性，作者提出使用部分连接层代替全连接层，每个部分负

责学习一个bit，各部分之间无连接（下图左slice layers）；

（2）为了避免二值码学习中的离散取值约束，像大多数哈希方法一样，作者使用sigmoid激活函数将离散约束松弛为范围约束（{0,1}→(0,1)），同时为了保持学到的特征空间和Hamming空间相似，引入了分段量化函数（下图右下）。以上这两部分合在一起，构成了图中的divide-and-encode模块。该方法可以端到端的训练，学到的图像表示可以反作用于二值码，因此相比于CNNH，性能有所提升。

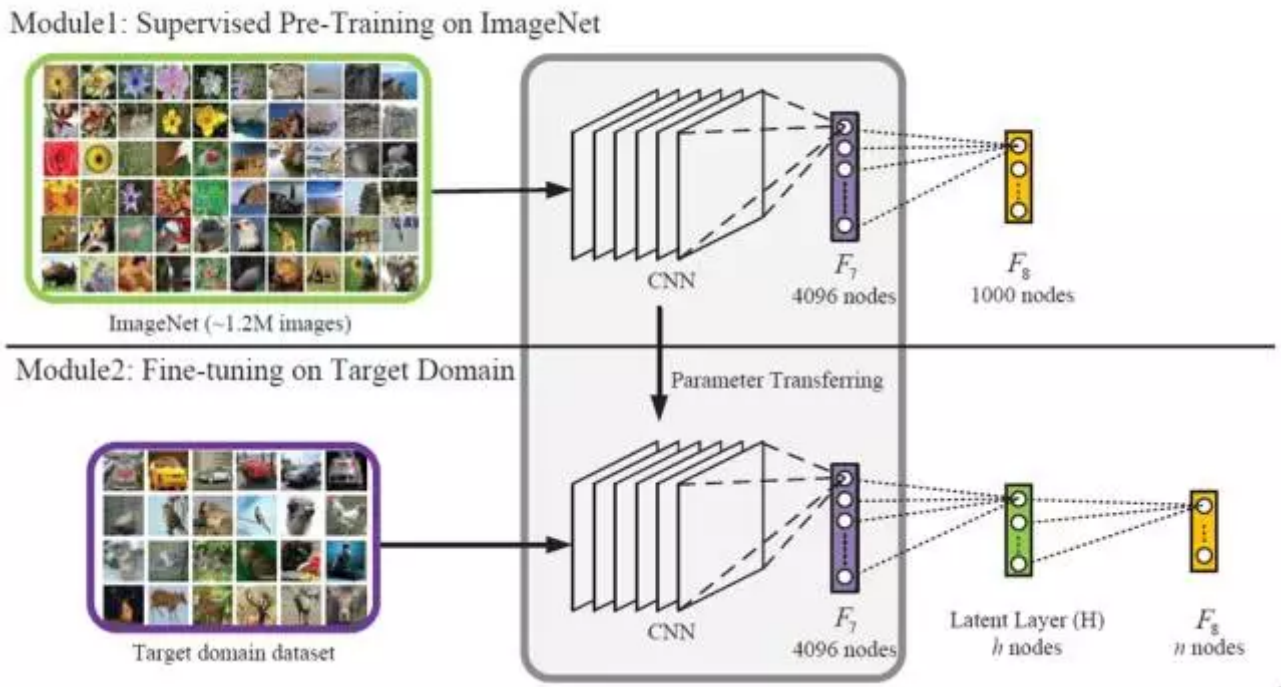


第二篇文章[5]来自中科院自动化研究所的谭铁牛老师的研究组，和DNNH相比没有在网络结构上下太大功夫，而是更多地关注了损失函数这一块。文章中使用了类似于DeepID2的网络结构，如下图所示。回头来看图像检索这个任务，任务的目的无非就是把数据库中的图像，按照和查询图像的相关性由大到小的顺序，依次返回。基于这个思想，这篇文章中提出直接让网络学习这个排序，因此该方法称为DSRH（Deep Semantic Ranking Hashing）。事实上，这种做法相当于直接对最终的评测指标进行优化，相当于开启了上帝模式。但是实际中上帝模式并不是那么容易开的，直接优化排序并不容易，因此作者使用了一个凸上界作为替代，进行优化。

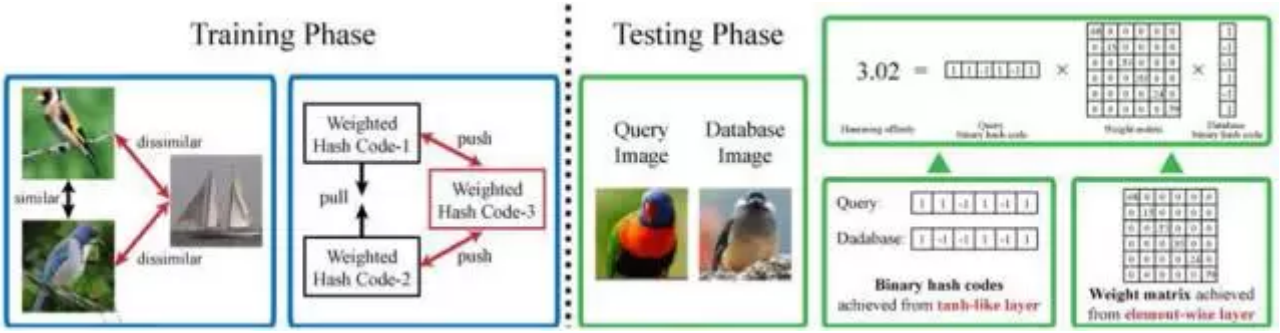


第三篇文章[6]来自台北中央研究院的陈祝嵩研究组，出现在当年的CVPR workshop中，其中使用

了一种比较直接的方法来学习二值编码，该方法名为DLBHC（Deep Learning of Binary Hash Codes），方法流程如下图所示。其核心点为：在预训练好的网络倒数第二层和最终的任务层中间，插入一个新的全连接层，这个层使用sigmoid激活函数来提供范围约束，节点数即为目标二值码的码长。通过端到端的finetune，可以将语义信息嵌入到这个新加入的全连接层输出之中。虽然这么做得到的二值码中包含语义信息，但是由于在训练过程中没有显式地考虑样本点之间的相对位置关系，并不能保证Hamming距离近的点在语义上也相近，因此和最终的检索任务还是有些偏离。

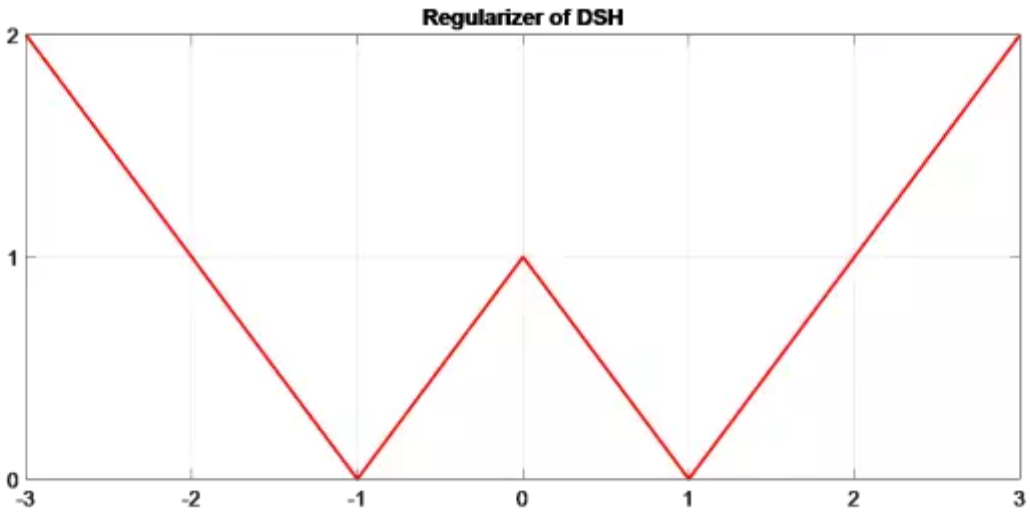


同年，由中山大学林倞老师、哈尔滨工业大学左旺孟老师和香港理工大学张磊老师等人合作的文章发表在当年的Transactions on Image Processing (TIP 2015)中[7]，作者提出了一种使用加权Hamming距离代替标准Hamming距离的哈希方法DRSCH（Deep Regularized Similarity Comparison Hashing），如下图所示。该方法同样使用基于三元组的损失，同时使用图像对（image pair）之间的相似性作为正则项，希望得到的网络能够同时保持三元组确定的关系和图像对确定的关系（实验表明这两者虽然看上去是在描述一样的东西，但是这种做法确实能取得一定的效果提升）。在网络学习的过程中，加权Hamming距离的权值也作为参数进行更新，从而得到与网络匹配的权值。虽然这篇文章中使用的加权Hamming距离的时间复杂度要略大于标准Hamming距离，但是也带来了额外的好处：一方面，可以以很高的效率计算更加精确的距离；另一方面，可以通过权值来选择bit，从而得到不同码长的二值码，而不必像之前的方法一样每换一个码长就重新训练一个模型。此外，和DNNH相似，这篇文章的作者也设计了一种操作来保证学到的空间接近于Hamming空间，其形式类似于双曲正切函数（tanh）。



以上四篇文章中的框架，可以代表大多数深度哈希文章的做法，可以总结为：深度模型学习图像表示 + sigmoid/tanh函数限制输出范围 + 不同的损失函数 + （可选）有针对性的网络结构。这四个部件合在一起，组合出了很多种不同的方法，在此就不再详细介绍这些衍生方法了。

上述框架中，问题比较大的一个地方在于sigmoid/tanh的使用。由于这类激活函数具有饱和的性质，越是当输出接近期望的值的时候（0/1或-1/+1），梯度就越小，网络训练也就越困难。因此，最近的一些工作开始关注sigmoid/tanh的替代品。例如，我们发表在CVPR 2016的工作DSH（Deep Supervised Hashing）[8]中，使用了如下图所示的一个正则项，来对网络的输出进行约束，使之接近二值编码。当网络的输出和期望得到的值偏差越大的时候，损失也越大，但是同时，梯度的值保持在-1或+1，来保证训练过程的稳定性。此外，类似的正则思想在清华大学的Haiyi Zhu博士等人发表在AAAI 2016和李武军老师研究团队发表在2016年国际人工智能联合会议（IJCAI 2016）的两篇工作中也有体现[9,10]。和侧重于设计损失函数的方法相比，这类方法的关注点在于量化部分，而这在传统哈希方法中也是一个重要的研究方向。



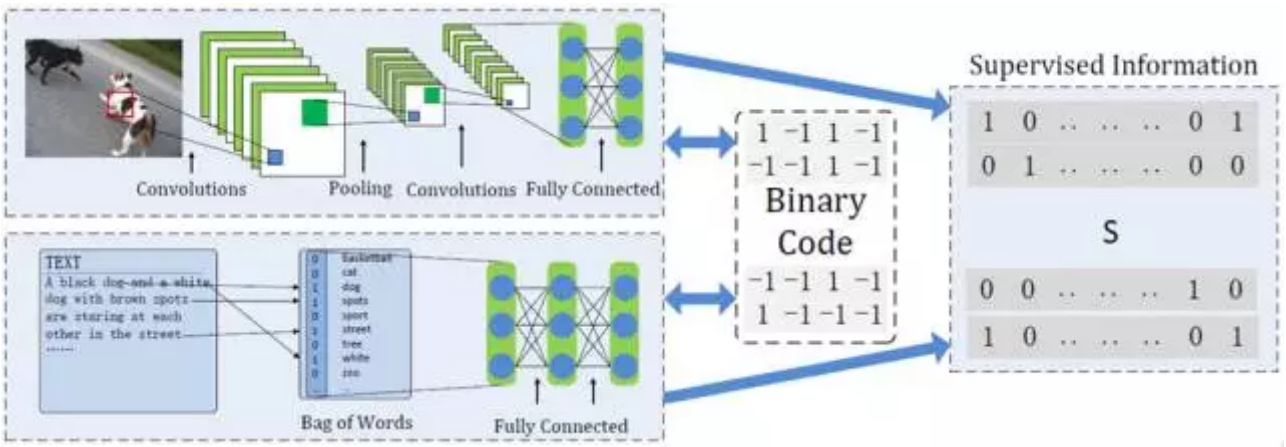
以上介绍的深度哈希方法在生成二值码的时候，只需要将图像送入训练好的网络，并将网络输出进行量化。值得一提的是，由于传统哈希方法需要同时使用多种特征才能达到和深度哈希方法可比的性能，提取特征的时间严重拖慢了传统方法在实际应用中的编码速度，因此深度哈希方法在编码速度上甚至有时会优于传统方法。

需要指出，目前很多深度哈希算法在对比同类方法的时候，用的都是对比方法原文中的网络结构，而自己却用更深、更复杂的结构。在我看来，这种对比并不能很好地反映方法本身的好坏，更合适地对比方法应该是大家使用基本一致的结构进行对比。

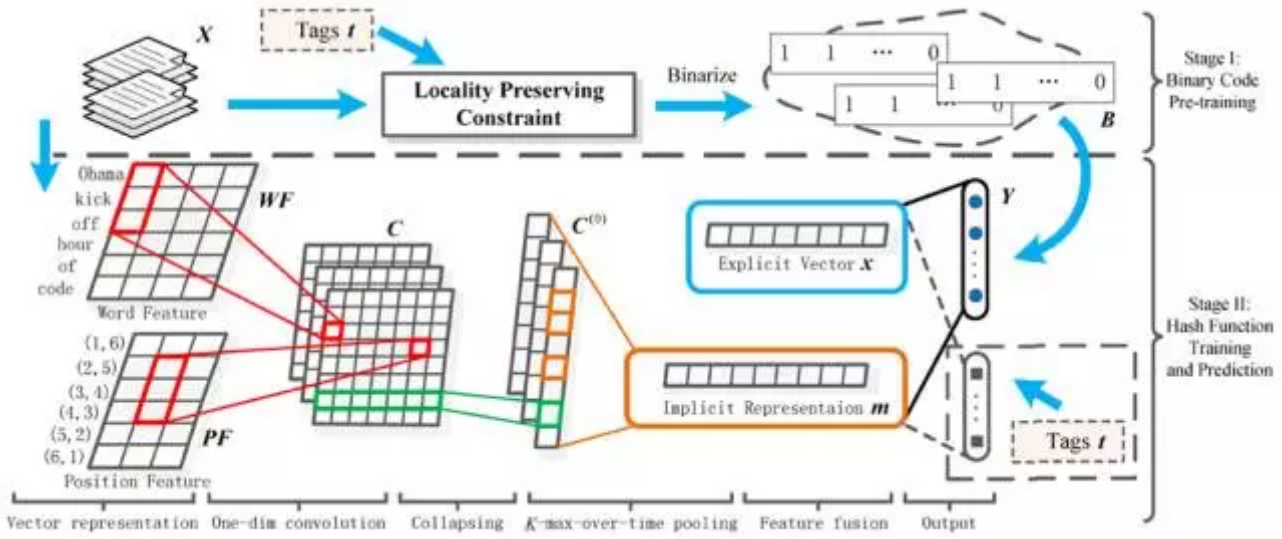
其他应用

上面介绍的方法解决的都是以图搜图的问题，在其他的一些应用方面，深度哈希算法也有用武之地。下面我举两个例子进行说明。

第一个应用是跨模态检索。一个最常见的例子是：在搜索引擎中输入一些关键词，找相关的图像。通常来说，关键词（文本）和图像并不在同一个空间中，因此无法直接比较。在2016年年初，李武军老师带领的研究团队在arXiv上发布了一篇文章，其中介绍了一种跨模态深度哈希算法DCMH（Deep Cross-Modal Hashing）[11]。这篇文章中，作者利用一个两路的深度模型将两种不同模态的数据（文章中是文本和图像）变换到一个公共空间，并要求相似的样本在这个公共空间中相互靠近，如下图所示。通过同时对图像和图像、图像和文本、文本和文本这几种不同类型的样本对施加这个约束，可以保证两种模态样本的对齐。如此一来，即可实现在公共空间中的跨模态检索。



另一种应用是以文本搜文本，即给定一段查询文本，找到和这段文本相似的其他文本。作为一个经典问题，在IJCAI 2015上，来自中科院自动化研究所的许家铭博士等人提出了一种基于卷积网络的解法[12]，如下图所示。该方法首先将文本中的每个单词表示成一个词向量，将文本转化为一个宽度等于句子长度，高度等于1，通道数等于词向量维度的张量。之后通过对文本进行卷积、pooling等一系列操作，得到一组中间表示，并对这组中间表示进行一系列操作得到二值码。这个方法中词向量的提取并不能和最终的任务连在一起，因此不能算是完全的端到端模型。作为利用CNN进行文本哈希算法的初步探索工作，这篇文章为深度哈希算法的更广阔应用开启了一扇新的大门。



以上两个应用作为深度哈希算法在不同领域上的具体实现，都根据手头的问题对模型进行了相应的修改，以适应对应的任务，而这种灵活性，也正是深度学习的一大优势。

结语

基于深度学习的哈希算法，凭借其强大的特征学习能力，一出现就迅速超越了基于手工设计特征的传统哈希方法。但是，目前的研究还远没有到尽头，更适合这一任务的网络结构、优化算法等都还有待进一步探索。目前来看，传统方法非但没有过时，反而可以在新的深度哈希算法研究中提供一些指导，进一步提升深度哈希算法的能力。

参考文献

[1] Ruslan Salakhutdinov, Geoffrey Hinton. Semantic Hashing. IJAR 2009.
[2] Rongkai Xia, Yan Pan, Hanjiang Lai, Cong Liu, Shuicheng Yan. Supervised Hashing for Image Retrieval via Image Representation Learning. AAAI 2014.
[3] Venice Erin Liong, Jiwen Lu, Gang Wang, Pierre Moulin, Jie Zhou. Deep Hashing for Compact Binary Codes Learning. CVPR 2015.
[4] Hanjiang Lai, Yan Pan, Ye Liu, Shuicheng Yan. Simultaneous Feature Learning and Hash Coding with Deep Neural Networks. CVPR 2015.
[5] Fang Zhao, Yongzhen Huang, Liang Wang, Tieniu Tan. Deep Semantic Ranking Based Hashing for Multi-Label Image Retrieval. CVPR 2015.
[6] Kevin Lin, Huei-Fang Yang, Jen-Hao Hsiao, Chu-Song Chen. Deep Learning of Binary Hash Codes for Fast Image Retrieval. CVPR 2015 workshop.
[7] Ruimao Zhang, Liang Lin, Rui Zhang, Wangmeng Zuo, Lei Zhang. Bit-Scalable Deep Hashing with Regularized Similarity Learning for Image Retrieval and Person Re-identification. TIP 2015.

[8] Haomiao Liu, Ruiping Wang, Shiguang Shan, Xilin Chen. Deep Supervised Hashing for Fast Image Retrieval. CVPR 2016.
[9] Han Zhu, Mingsheng Long, JianminWang,Yue Cao. Deep Hashing Network for Efficient Similarity Retrieval. AAAI 2016.
[10] Wu-Jun Li, Sheng Wang, Wang-Cheng Kang. Feature Learning based Deep Supervised Hashing with Pairwise Labels. IJCAI 2016.
[11] Qing-Yuan Jiang, Wu-Jun Li. Deep Cross-Modal Hashing. arXiv:1602.02255.
[12] Jiaming Xu, Peng Wang, Guanhua Tian, Bo Xu, Jun Zhao, Fangyuan Wang, Hongwei Hao. Convolutional Neural Networks for Text Hashing. IJCAI 2015.

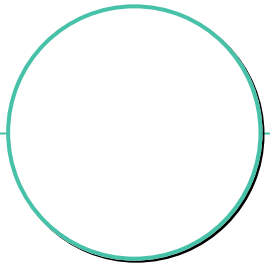
作者简介

刘昊淼

中国科学院计算技术研究所VIPL课题组博士生，专注于深度学习在大规模图像检索、分类中的应用研究。相关研究成果发表在计算机视觉国际顶级学术会议ICCV、CVPR。

往期精彩回顾

Deep learning for arts
深度学习在智能电网图像识别与故障检测中的应用
【脑洞】Hinton剑桥演讲：大脑神经元的误差反向传播机制
基于深度学习的目标检测研究进展
【阿里集团卜居深度解析】卷积神经网络的硬件加速
长文干货！走近人脸检测：从 VJ 到深度学习（上）
长文干货！走近人脸检测：从VJ到深度学习（下）
面部特征点定位概述及最近研究进展



欢迎关注我们！

深度学习大讲堂致力于推送人工智能，深度学习的最新技术，产品和活动！

深度学习大讲堂