通过CNN实现的语义保留的哈希监督学习算法论文的研读

姓名：汤振杰 学号：201832102002 专业：计算机技术

1. 摘要

这篇论文提出了一个简单高效的监督深度哈希算法。但是用该方法生成的哈希编码对大的数据集检索很高效。我们先假设语义标签是被几个潜藏的属性开或者关来决定的。分类就依靠着这些属性。基于这个假设，我们把哈希函数当成深度神经网络的隐含层，通过最小化一个目标函数和限制哈希编码的属性来学习这些二进制编码。构造出的算法我们叫做SSDH。通过这样的设计，SSDH在一个简单的学习模型中把分类和检索结合起来了。除此之外，SSDH把图像的表示，构建哈希编码，和分类以逐点的方式结合了起来。SSDH是简单的因为它可以通过稍微修改一下已经存在的分类深度神经网络来构建。然而SSDH同时又是有效的在一些基准数据集和大数据集上比其他哈希算法更出色。与最先进的方法做对比，SSDH实现了更高的检索准确率。可喜的是分类精度也没有牺牲。

1. 课题介绍：

语义检索在基于内容的图片检索中是非常重要的。在基于内容的检索中哈希方法构造的相似保留的二进制编码很受关注。构造哈希方法的关键原理是投影相似图像到相似的二进制代码，也就是把高维的看的见得数据投影到低维的哈希编码空间。投影后，我们就可以通过计算两个哈希码之间的哈西距离，来确定与检索图片最近的图片，该过程是非常快的。

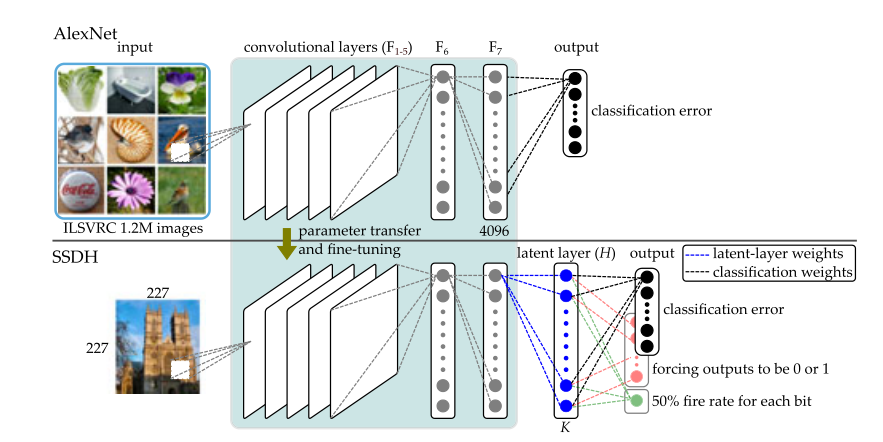
最近，基于学习的哈希算法由于在构建哈希编码时可以充分利用训练样本而变得很流行。学习到的二进制编码比通过随机投影以大的可能性投影相似的图片到相同的类别里的局部敏感哈希算法（没有使用任何训练数据因此需要更长的编码来获取更好的性能）更有效。在众多的监督学习算法中。获取监督信息的哈希算法在构建哈希函数的过程中可以捕获数据的语义结构。虽然监督哈希方法产生了有希望的性能，但是由于最近好多算法使用成对的训练样本来训练因此需要长时间的训练和大的存储空间。他们适合小型的数据集一旦数据集变大他们也就没有了用武之地了。

最近的研究表明深度卷积神经网络有能力获取图片丰富的中层表达，这些表达对分类目标检测，语义分割。在多类别的大数据集上训练好的卷积神经网络能被用在其他的领域提出特征。(如：检索)这比手工（GIST [16] and HOG）提取的特征效果好多了。而且这些预训练好的网络能获得更好的语义特征。

CNN在分类和目标检测中的成功很有影响力的。它揭示了在大的多类别的数据集上训练好的卷积神经网络有领域适应和转移学习。在图像检索中，一个值得研究的问题很自然的提了出来：出来分类，这个微调与预训练的网络是不是有能力学习哈希编码来进行图片检索？？除此之外，如果可以，那么应该如何去修改这个网络的结构？？

在这篇论文中，为了回答这个问题和能有效的去训练大的数据集，我们利用深度学习的优势，然后提出了监督的语义保留的哈希算法（SSDH）来从带标签的图片中学习二进制编码。在该算法中我们假设图像标签是由于一系列潜在的相互独立的属性来决定的并且分类就是靠这些潜在的属性。这个假设既朴实有具有创新。基于这个想法，我们构造了一个哈希方法来作为卷积神经网络的隐含层。这个二进制编码通过最小化一个损失函数来实现。这个设计虽然简单但是在学习过程中把分类和检索很好的结合起来。保证语义相似的图片有相似的二进制编码。

另外，为了使每一个隐含层的编码尽可能的输出在0和1之间和最后得到的哈希编码更加分散。我们在目标函数上加了一些限制来使二进制编码的每一个二进制位携带更多的信息和更有辨别性。在网络的学习中，我们使用训练好的卷积神经网络来修改结构，然后再目标数据集上来进行微调。网络的结构如下图1：



图一

我们的网络能使用已经训练好表现好的深度卷积神经网络。并且提供了一个简单的方法去加强它。仅仅需要微调一下权重就可以实现分类和检索。并且当我们训练完成时，检索准确率提高的同时分类的准确率也没有下降。这篇论文主要的贡献在以下几个方面：

1．把检索与分类结合了起来。SSDH是一个监督的哈希算法，所以它有深度学习的优势。在一个简单的学习模型中把分类检索结合起来。也把学习图片表达，学习哈希方程，和分类结合了起来。

2. 可扩展的深度哈希。SSDH是逐点的方法，因此它不需要成对的训练输入。因为这个特点使它可以在大的数据集上进行学习和检索。

3．轻量级深度哈希。SSDH是在训练好的有效的深度网络上建立的，所以它可以获得监督迁移学习的优点。可以通过修改一下现存的网络来简单的实现。

该论文在一些基准数据和一些超过百万的数据集上进行充分的实验。实验结果表明：该方法简单但是效率高，效果好。比最先进的算法能得到更好的结果。

1. 研究现状：
2. 基于学习的哈希算法

通过训练数据生成哈希码，希望克服不依赖数据的方法的限制。例如：LSH系列算法。基于学习的算法根据使用监督信息的多少分为三类：一类：无监督的学习算法，二类：半监督的学习算法，三类：监督的学习算法。

无监督的学习算法：它使用没有标签的数据来生成哈希编码，然后尽量在原空间中保留图片的相似性。又代表性的算法有：SH和ITQ。

半监督的算法即使用有标签的数据又使用无标签的数据来生成哈希编码。例如，ssh在成对的数据上最小化经验损失函数最大化这些哈希码之间的距离。半监督标签散列（SSTH）在类标签和哈希码之间建立模型然后在无监督的模式下去保留两张图片之间的相似性。

监督的算法：它的目的是充分使用数据集的监督信息来生成高效的二进制表示，因此可以获得比非监督和半监督的算法更高的检索准确率。利用样本数据集中的成对关系，BRE最小化原始的欧几里得距离和二进制编码的哈希距离，这些相同的不同的标签的信息在训练的过程中能被收集到。MLH最小化经验损失来生成哈希编码。基于排序的方法可以结合这些；来自成对样本的排队信息来训练。那些依靠成对训练的算法训练时需要大的存储空间而且不适合使用在大的数据集上。而逐点的方法使用提供的标签数据来生成哈希编码将更适合可扩展的哈希。使用典型相关分析（CCAITQ）[1]的迭代量化将CCA与标签信息一起应用于降维，然后通过最小化量化误差来执行二值化。受监督的离散散列（SDH）[37]制定了哈希码的学习分类条款，以便学习二进制代码最优用于分类。虽然SDH与我们提出的方法在学习哈希码和分类上有相似的原理，但是SDH把该问题分成了小问题。而且需要很小心的选择分类时的损失你函数才能使整体最优和最有效。我们基于深度网络的方法简化了这个最优化过程，所以很自然的就可以使用到大的数据集上。

基于深度网络的监督哈希算法，形成了一个系列，所以我们在这分开讨论它们。最早使用深度网络在哈希算法上是语义哈希。它使用一族受限的玻尔兹曼机来在没有标签的数据集上进行生成哈希码。这个学习到的二进制编码作为内存中分地址来使用。因此相似的项可以通过哈希距离来查找出来。可以自动学习数据压缩表示的自动编码机能被用来投影图像到二进制编码。深度自编码器使用去训练好的玻尔兹曼机来初始化权重，这个二进制编码层使用逻辑单元它的输出能被现在在0或者1.

深度网络也用于深度散列（DH）和监督DH（SDH）[41]用于学习紧凑二进制码通过寻找多个非线性投影来映射图片写成二进制代码，深度多视图散列（DMVH）构建具有视图特定和共享隐藏单元的网络，以处理多视图数据。然而，这些方法依赖于手工制作的特征，这些特征在事先设计之前需要强大，并且不会随着代码学习而发展。相比之下，我们的SSHD将功能学习和代码构建融合在一个模型中。在语义监督下，它们都演变成一个特征空间，其中语义相似的内容倾向于共享相似的代码。在语义监督下，它们都演变成一个特征空间，其中语义相似的内容倾向于共享相似的代码。最近，基于CNN的哈希算法已经被提出来了， CNNH和CNNH + [43]采用两阶段学习方法，首先基于数据标签将成对相似性矩阵分解为近似哈希码，然后训练CNN以学习哈希函数。这些方法和基于深度语义排名的哈希算法采用源自标签的三元组排名损失用于代码构建。跟这些算法一样，我们的算法也在编码学习中探索标签信息。然而我们的方法在以下几个方面有别于他们。首先我们的SSDH对掩藏层添加了额外的限制去让获得的二进制编码更加分散，而其他算法没有这个限制。第二，我们的网路能在现有的分类网络上轻微的改进来实现而其他算法的网络必须用有意义的改动来实现更复杂的网络。最后我们的网络是逐神经元的但是其他的方法，这些方法中的一些需要在散列函数学习之前执行矩阵分解（例如，CNNH和CNNH + [43]），并且一些需要以图像对的形式（例如，SDH [41]）或图像三元组（例如， ，[40]和DSRH [44]），当数据量很大时，它们不那么有利。

1. 监督深度转移学习

在深度学习中，网络可以预先训练基于能量概率的无监督方式RBM模型和深层信念网络[46]，或通过自我在自动编码器中复制[39]，然后，随后进行监督训练（即，微调），可以针对特定任务优化网络。最近预先训练已经推进了监督学习的步伐。监督的预训练然后微调网络已经在CNN上被采纳了也取得了令人瞩目的成绩。它遵循归纳转移学习原则[47]，它采用的理念是，人们无法学习如何在爬行前行走，或者如何在行走前跑步。因此，从神经网络的一个或多个任务训练的连接强度可以用作初始条件并且进一步适应于适合其他域中的新的和/或更高级别的任务。在DeCAF调查的监督预训练[12]表明，在ImageNet数据集[48]上进行预训练的深度CNN可用作特征提取器。获得的深度卷积特征对于其他视觉任务是有效的，例如场景分类，域适应和细粒度识别。在[13]中研究了深度表示的能力，其中转移了预训练的CNN的中级表示，并且将两个适应层添加到深度特征的顶部以用于学习新任务。这项工作表明，只需要有限的训练数据就可以实现转学。不同于[13]，其中微调仅在附加层中进行分类，基于区域的卷积网络（R-CNN）[8]，[21]对整个网络进行微调，以针对特定领域的对象任务 检测和分割。

此外，这些深层特征最近也在图像检索中引起了广泛关注。如Krizhevsky等人所示。 [5]，在大数据上学习的CNN的特征可用于检索。从那以后，深度特征在图像搜索中被广泛采用。例如，[15]中的工作已经广泛地评估了深度特征作为全局描述符的性能。龚等人。 [49]建议使用局部聚合描述符向量（VLAD）在多个尺度上汇集局部补丁的深度特征。Babenko和Lempitsky [50]提出了一种汇总聚合方法，用于从局部深度特征生成紧凑的全局描述符，[14]中的工作研究空间搜索策略以提高检索性能。

如何利用监督深度转移学习的优势来完成哈希函数的构建尚未进行探索，在本文中，我们不仅仅为了任务域转换的目的而进行归纳传递学习，而是进一步研究了功能级别的适应问题。所提出的方法将权重微调到新域以进行分类，并且还实现了功能级调整以生成语义感知二进制代码。我们的方法依赖于现有分类体系结构的增强，并且我们表明分类性能不会在实验上降级。因此，它提供了一种对检索和分类都有效的多用途体系结构。

1. 通过深度网络学习哈希编码

让I={In}Nn=1表示N幅图片。是它们相对应的标签向量。M表示类标签的总数。如果图片属于该类那么为1否则为0。我们的目标是学习一个映射这个映射把图片映射到kbit的二进制编码。M\*N并且保留图片之间的语义相似性。特别地，我们的目标是设计一个监督哈希算法。该算法使语义标签去创造具有下列属性的二进制编码：

第一：这个编码准从两个图片的语义相似性。有相同种类的图片被映射到相同或者相似的二进制代码。

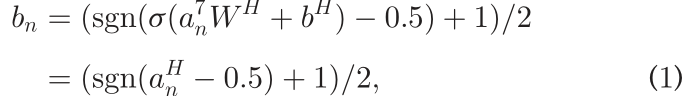
第二：编码中的位均匀分布且具有区别性。

1. 深度哈希方程

我们利用深度学习的最新进展，在CNN上构建哈希函数，能够从图像中学习语义表示。我们的方法是基于已经存在的深度模型，如：Alxnet 和 VGG。它也能跟其他深度模型结合。在不失一般性的情况下，我们在下面介绍基于AlexNet的方法。

AlexNet的架构如图1的上半部分所示。它有5个卷积层（F 1~5），最大池操作后跟2个完全连接的层（F 6~7）和一个输出层。在卷积层中，单元被组织成特征图并且本地连接到前一层的输出（即，特征图）中的块。当任务是识别图像时，可以将完全连接的层视为分类器。卷积和前两个完全连接的层（F 6~7）由整流线性单元（ReLU）组成，因为ReLU导致更快的训练。AlexNet特别针对多类分类问题而设计，因此其输出层是一个分类层，具有相同数量的类标签的单元。输出单元具有softmax功能，并且训练网络以最大化多项逻辑回归目标函数以进行多类别分类。为了将深度表示结合到散列函数学习中，我们将具有K个单位的潜在层H添加到层F 7的顶部（即，恰好在输出层之前的层），如图1的下半部分所示。该潜在层完全连接到F 7并使用S形单元，使得激活在0和1之间。

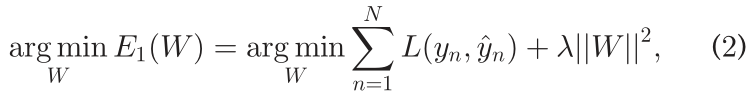
让表示F7与添加的潜在层之间的权重对于给定的图片有特征向量在F7层的。这H层的激活单元能被计算是k维的特征向量。是偏移量。是逻辑sigmoid函数，定义为：z为真是的值。这个二进制编码方法为：



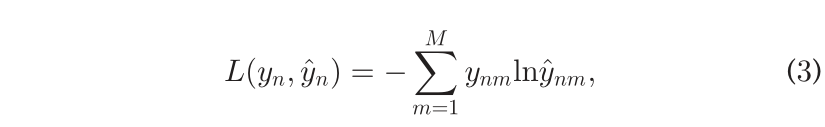
如果v>0其他情况v=-1。对矩阵或向量执行元素操作。

1. 标记一致的二进制代码

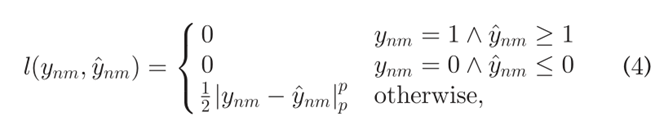
图像标签不仅提供分类图像的知识，而且是用于学习散列函数的有用的监督信息，我们建议对标签和二进制代码之间的关系进行建模，以构造保留语义的二进制代码。我们假设语义标签可以从一组K潜在概念（或隐藏属性）中导出，每个属性都打开或关闭。当输入图像与二进制值输出相关联（在0或者1中）时，分类取决于隐藏的属性。这意味着通过优化在分类错误上定义的损失函数，我们可以确保将语义相似的图像映射到类似的二进制代码。考虑一个矩阵，它执行二进制隐藏属性到类标签的线性映射。将这样的矩阵结合到我们的网络中相当于在潜层的顶部添加分类层（参见图1，其中黑色虚线表示）。令表示图像的网络预测（图1中的黑色节点）。在分类公式方面，为了解决，可以选择优化以下目标函数：



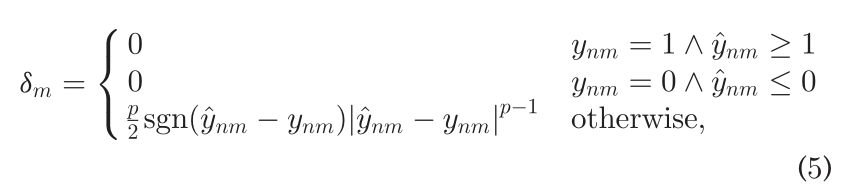
L是一种损失函数，可以最大限度地减少分类错误，详情如下，W表示网络的权重，并且管理正规化术语的相对重要性。损失函数的选择取决于问题本身。对于多类分类，我们只需遵循AlexNet中使用softmax输出的设置并最小化交叉熵误差函数：



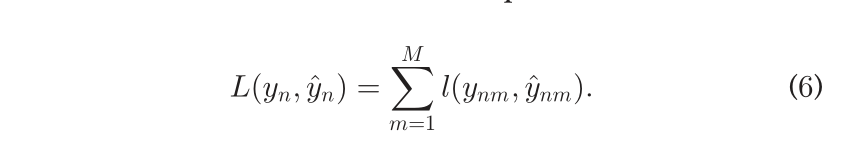
其中和是第m单元的期望的输出和预测输出。

我们引入了最大边际损失函数来实现多标签分类的目标，因为AlexNet中的损失函数仅用于单标签目的。遵循相同的观念，让表示与M类标签的N个图像相关联的标签矢量。在多标签分类中，图像与多个类相关联，因此的多个条目可以是1，我们网络中的输出是二元分类器，给定第n张照片标签为，我们希望网络的第m个输出节点对所需标签（即正样本）具有正响应，对（即负样本）具有负响应。具体来说，要扩大分类边界的边界对于特定标签的样本，我们把的节点设置为1. 的节点设置为0。每个输出节点的损失函数定义如下：  


当p = 1（或2）时，这种损失函数实际上实现了线性L1范数（或L2范数）支持向量机（SVM）[51]，其阈值为0.5。网络将AlexNet架构，二进制潜在层和SVM分类器组合在一起，用于多标签分类。注意，为了训练大规模线性SVM，最先进的方法[51]，[52]采用SVM双域（DCD）中的坐标下降优化，证明它等同于执行随机原始域中的梯度下降（SGD）[51]。由于SGD是训练神经网络的标准程序，当我们的网络仅针对SVM层进行训练并且其他层的参数是固定的时，它相当于使用原始域SGD方法求解SVM的凸二次规划问题。 在[51]，[52]中（SGD的学习率对应于某些SVM的模型参数C）。在训练整个网络时，参数随后演变为更有利的特征表示（在AlexNet架构中），潜在二进制表示（在隐藏层中）和二进制分类器（在SVM层中）。具有激活输出单元m的梯度表达式如下：



P为1或者2。由于损失函数几乎可以在任何地方区分，因此它适用于基于梯度的优化方法。最后损失函数为所以损失单元的和表达式如下：



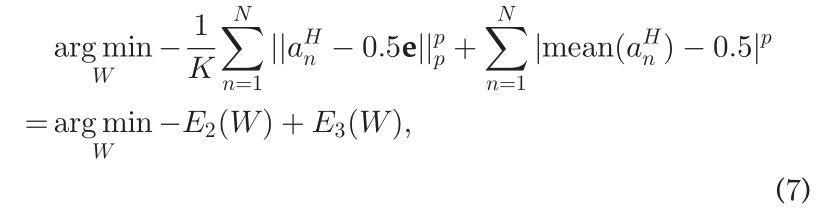
1. 高效的二进制代码

除了语义相似的图像具有相似的二进制代码之外，我们还鼓励每个潜在节点的激活接近0或者1。让是隐藏层向量的第k个元素。因为已经被sigmod函数激活了，所以它的值在0或者1之间。为了进一步使代码接近0或1，可以通过添加最大化潜在层激活与0.5之间的平方误差之和的约束来实现，既，其中e是具有所有元素1的K维向量。通过此约束，网络生成的代码可以更恰当地满足二进制值要求。

除了使代码二值化之外，我们还考虑了平衡属性。这可以通过让训练样本中的50％的值为0并且将每个位k的另外50％为1来实现，如[27]中所建议的那样。但是，由于所有训练数据共同参与以实现此约束，因此当SGD应用于优化时，很难以小批量实施。

在本文中，我们希望将约束可分解到每个样本上，以便它们可以通过SGD以点对点的方式实现。为了使二进制代码平衡，我们考虑使用小批量实现不同的约束。给一个图片，让在{0,1}上形成离散概率分布。我们希望隐藏值不是0或1。也就是说，每个比特的开启或关闭的发生概率是相同的，或者离散分布的熵最大化。为此，我们希望每一位通过最小化来触发50％的次，其中计算向量中元素的平均值。它还扩大了最小间隙并使代码更加分离，因为具有相同0和1的两个二进制字符串之间的最小汉明距离是2（但不是1）。

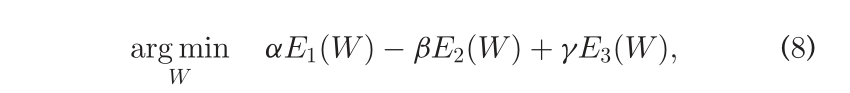
总之，组合这两个约束使得接近长度为K的二进制字符串，每个位为0或1的概率为50％，我们的目标是优化以下目标以获得二进制代码：



第一项鼓励H中单位的激活接近0或1，第二项进一步确保每个节点的输出有近50％的概率为0或1。请注意，方程式中设计的目标函数任然是损失和的形式。它保留了每个损失项仅由单个训练样本贡献的属性，并且损失函数中不涉及交叉样本项。因此，目标仍然是有意义的，并且可以通过将训练样本（但不是它们的对或三元组）分成批次来有效地通过SGD最小化。因此，我们的网络依赖于潜在概念驱动的分类目标的最小化，在潜在代码上具有一些充分条件来学习语义感知二进制表示，这可以在实验中显示出相当有效的数据集。在网络设计中，我们添加一个单元（图1下半部分的绿色节点），在潜层的节点上执行平均池操作（绿色虚线），以获得的平均激活在 Eq（7）中。与该单元的连接相关联的权重固定为1 / K。方程式中的项直接对潜层中的单位施加约束。无需修改网络。然而，为了清晰呈现，我们在图1中绘制了额外的红色节点以指示该约束。

1. 总体目标和实施

总的目标函数是为了建造语义保留和二进制编码有属性的函数，如下：



和是每一项的权重。执行多标签分类时，在我们的实现中，输出层将替换为最大边距损失层。由于我们的网络改编自AlexNet [5]，该网络已经在ImageNet的120万个ILSVRC子集上进行了1000级识别任务的训练，因此我们网络中F 1到7层的初始权重被设置为预先训练的网络 其余的权重随机初始化。我们将SGD与反向传播一起应用于小批量网络培训，以最大限度地减少方程式中的总体目标。我们还采用了dropout，其中中间单位的激活被设置为零，在训练期间概率为0.5，以避免过度拟合。首先在数据集上评估参数和，然后在我们的实验中将所有参数设置为1。我们的模型是对现有网络的轻量级修改，因此易于实现，这些代码是公开的。



与“AlexNet功能+ LSH”的关系。 我们的方法和AlexNet功能+ LSH之间的关系值得一提。因为随机高斯权重用于初始化F 7和潜在层之间的权重，所以我们的网络可以被视为用LSH（即，随机权重）初始化以将在ImageNet中学习的深度特征（AlexNet特征）映射到二进制代码。通过SGD学习，预训练，潜在和分类层的权重演变出更适合新域的多层函数。与AlexNet功能和LSH的简单组合相比，我们的方法可以获得更有利的结果，如第4节中的实验所示。

1. 检索二进制代码

图2示出了用于提取二进制代码并检索用于查询的类似图像的方案。首先，将图像馈送到网络，并提取潜在层的激活。然后经由等式1量化所提取的激活来获得二进制码。通过计算查询的二进制代码与数据库图像之间的汉明距离并在数据库中选择具有小汉明距离的图像作为检索结果，找到与新查询类似的图像。

1. 实验

我们在几个基准测试中进行实验，以比较我们的方法和最先进的方法。我们还将我们的方法应用于包含超过100万个图像的大型数据集，以显示其可扩展性。数据集中的图像有各种各样的图像类型，包括CIFAR-10的微小物体，NUS-WIDE的网络图像，MNIST的手写数字，UT-ZAP50K的目录图像，以及SUN397，Oxford的场景图像， 和巴黎。大型数据集Yahoo-1M和ILSVRC分别包含具有异构类型的产品和对象图像。评估协议和数据集总结如下。

1. 评价标准

我们使用文献中广泛采用的三种评估指标进行性能比较。 它们从不同方面衡量哈希算法的性能。

1. 平均精度（mAP）: 我们根据汉明距离对查询排序所有图像并计算mAP。mAP计算召回精度曲线下的面积，并且是哈希函数的所有性能的指标。
2. k样本的精度：它被计算为前k个检索图像中真实邻居的百分比;
3. 汉明半径r内的精度：我们计算在查询图像的汉明半径r内的桶中的图像的精度，其中r = 2被选择为先前的工作。

遵循评估哈希方法性能的常用设置，我们使用类标签作为基础事实，并通过检查返回的图像和查询是否共享公共类标签来计算所有上述三个度量。对于缺少类标签的数据集，通过为其测试集中的查询提供的地面实况检索列表来评估性能。

1. CIFAR-10上做的实验

CIFAR-10介绍：一个由60,000个32\*32图片组成的数据集。彩色图像分为10类。类标签是互斥的，因此每个类有6000个图像。整个数据集被划分为两个非重叠集：具有50,000个图像的训练集和具有10,000个图像的测试集。按照[40]，[43]中的设置，我们从测试集中随机抽样1,000个图像，每个类100个图像，以形成用于性能评估的查询集。 CIFAR-10是用于评估基于散列的图像检索方法的最常用数据集之一。

结果：

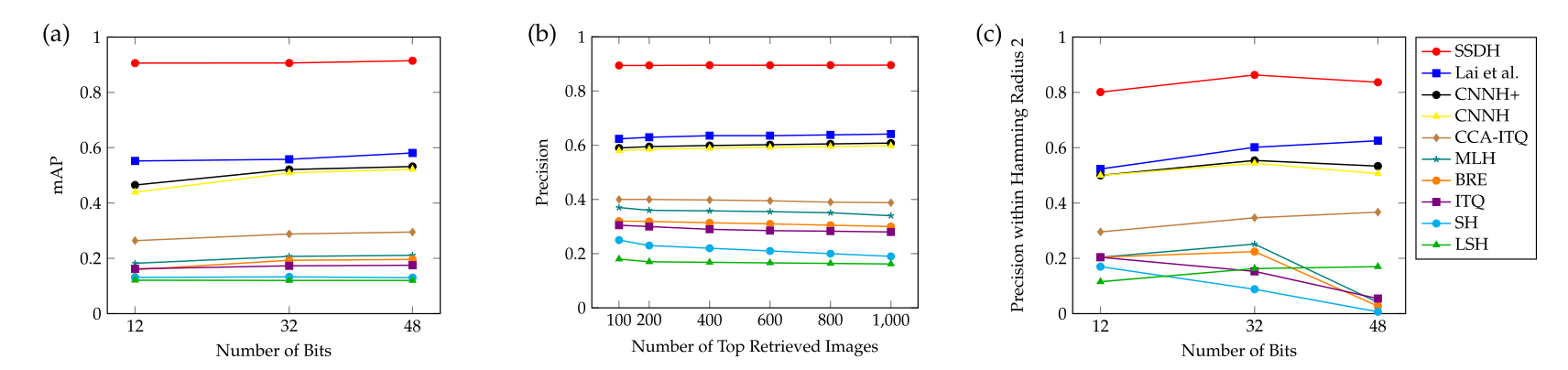


图4. CIFAR-10数据集上不同散列算法的比较评估。（a）关于不同哈希比特数的mAP曲线。（b）当在评估中使用48位哈希码时，关于不同数量的顶部检索样本的精确度曲线（c）相对于不同数量的散列位，在汉明半径2曲线内的精度。

参考结果，SSDH为不同的代码长度提供稳定和最有利的性能，并且与竞争方法相比，mAP提高了约34％。结果表明，在单一学习模型中统一检索和分类，其中哈希码学习由语义标签管理，可以更好地捕获图像中的语义信息，从而产生更有利的性能。

图4b示出了当在评估中使用48位散列码时k个样本的精度，其中k的范围从100到1,000。这些曲线传达了与mAP度量中观察到的类似的消息。SSDH具有优于其他散列方法的一致优势，并且在学习散列函数中利用标签信息的方法（我们的，Lai等人，CNNH +，CNNH和CCA-ITQ）比不具有这种方法的那些更好。汉明半径2内的精度评估如图4c所示。我们的方法在此指标上也更有利于其他方法。由于不清楚r对于不同任务和代码长度的适当值是什么，我们认为前面两个评估指标（k个样本的mAP和精度）将比一般的度量更好地反映检索性能。在这里，我们仅使用r = 2来遵循性能比较的惯例。

随着我们的网络从分类网络中得到增强，值得注意的是分类性能是否仍然保持不变。为了验证这一点并进行公平的比较，我们在CIFAR-10数据集上微调原始的AlexNet（即没有添加潜在层的模型），使用ImageNet上训练的功能进行初始化。AlexNet +微调实现了89.28％的分类精度，我们的SSDH架构（具有潜在层）的代码长度分别为12,32和48，分别达到89.74,89.87和89.89％的精度。它表明，使用我们的架构仍然可以实现稳定的分类性能。

1. 结论

我们提出了一种监督的深度哈希模型SSDH，它保留了图像之间的标签语义。SSDH将散列函数构造为网络中提取特征层和分类层之间的潜在层。通过优化在分类错误和二进制代码的期望标准上定义的目标函数，SSDH共同学习二进制代码，特征和分类。这种网络设计有几个优点：（1）SSDH在单一模型中统一检索和分类。（2）SSDH简单，易于通过对现有深度网络进行略微修改来实现分类（3）SSDH自然可扩展到大规模搜索，我们已经进行了大量的实验，并提供了SSDH的比较评估，其中包括多种基准测试，具有多种图像类型。结果表明，SSDH实现了卓越的检索性能，并提供了有前景的分类结果。

1. 自我总结

以上是我对论文的理解。这是第三次整篇读这篇论文，因为要写1万字的论文阅读报告，所以我把论文又自己翻译了一遍。下面就把我看这篇论文从开始到现在的过程说一下。

首先刚拿到这篇论文时，处于一种懵逼的状态。不知道这是要做什么。虽然暑假的时候也学习过一段时间的卷积神经网络。理解里面每一层的具体操作和每一层的作用。也写过一些简单的网络，比如DNN和简单的CNN之类的。但是没有实际使用过。所以看到这篇论文时基本是一无所知的状态。然后就仔细的去读第一遍，读到后面就把前面的忘记了，没一个整体的认识。当读完地一边时，能明白它里面大致讲什么东西。但是每一步还是不太清楚。第一遍大概花了我不到一周的时间。

当看第二遍的时候，就仔细看每一个需要用到的式子，看每一个式子的含义，这大概花了我三天时间。读完第二遍基本就明白了它里面的每一步具体怎么做还有这么做的意义是什么。回国头来看这篇文章其实就觉得特别简单了。我说的是原理特别简单。然后我自己就有想试一试的冲动。所以下一步我就开始着手开始实现了。

因为有一些tensorflow的基础所以我就选择了tensorflow的框架。作者用的是CAFFE框架。所以得自己写代码了。如果自己从头开始写这个网络的话明显就有点复杂。所以我现在github上寻找微调的AlexNet，然后下载了下来。下载下来代码后就开始看代码，代码很容易看懂。但是我需要修改代码里面的输入函数，因为我使用的是cifar-10的数据集，网络的输入不是太合适。修改输入的过程花了我大概3天的时间。最后在我电脑上跑起来的时候吃内存太多，所以我只能在服务器上跑。最后实验到70%的准确率时上升的就特别慢了。修改超参数后，速度又上升了。这是做实验的整个过程。因为在华为云服务器上跑一个小时要近8块钱。所以晚上走的时候就关掉了。这就是实验的整个过程。

通过读和实验本篇论文，首先我了解了论文的整体架构，其次对CNN的理解更加深入了，然后调试网络时也积累了一定的经验。为以后自己看论文和写论文打下来基础。同时也让自己对深度学习有了进一步的认识。

最后放上自己修改的微调AlexNet的代码：https://github.com/tangzhenjie/fineTune\_AlexNet\_For\_Hash