二进制描述子需要达到的优点：

尺度不变性、旋转不变性、噪声鲁棒

对于光照等变化不敏感、高重复场景鲁棒、视角变化大时仍能较好匹配

量化损失小

看论文时要关注如何进行二进制编码的

Learning Deep Binary Descriptor with Multi-Quantization

BRIEF，BRISK，ORB，FREAK等二进制描述符表现并不完美，因为对于原始像素强度的比较易受尺度，变换和噪声的影响。（对于不同光照时的重定位有影响？？），所以如果可以用一种更稳定的图片描述子替换orb描述符来进行视觉里程计的计算，则SLAM更准确。

目标：提出一种新的描述子代替ORB，然后用orbslam2的框架编写代码，验证slam算法的稳定性和准确度。

CNN特征有较强差异性，如果训练一个检测器检测稳定的描述子，并用二进制表示，则计算开销小。因此提出了很多二进制描述符。

Contribution：

提出一种无监督特征学习方法叫deep binary descriptor with multi-quantization（DBD-MQ），提出deep multi-quantization network来学习基于无监督的和数据的二值化方法。并且设计了K-Autoencoders (KAEs) network在深度学习框架来同时学习特征提取器的参数和二值化方程。

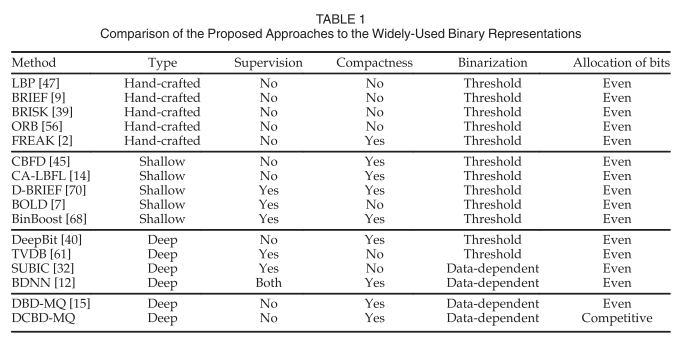
但是DBD-MQ不管信息量的不同，而给每个实数特征维度分配了同样数量的bits。但是不同维度需要不同bits，所以设计了deep competitive binary descriptor with multi-quantization (DCBD-MQ)的学习方法。

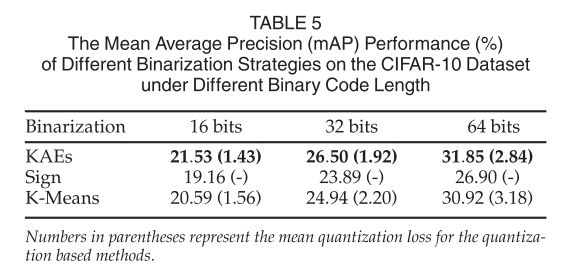
重视实数特征维度的分布和bits的分布

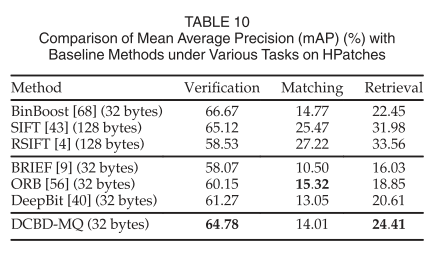
Drawback：

喷的同行的描述子缺点：compact binary face descriptor (CBFD) and DeepBit仅仅利用rigid sign function（sgn(x)）来进行二值化，而不考虑数据分布，导致严重的量化损失。

一些实验结果：







现存的许多二进制描述子只是简单利用了rigid sign function，但是许多分布中，手设的零作为二进制的阈值并不合理，会导致比较严重的量化损失。

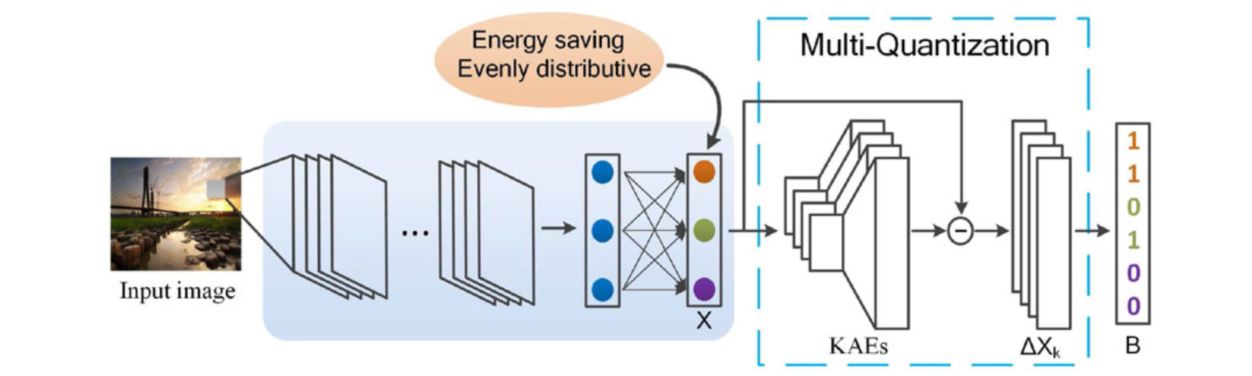
本文提出了一种deep multi-quantization network用无监督方式学习基于数据的二进制方法。与认为设置阈值相比，这种方法可以探索每个特征维度的数据分布，获取细腻的二进制结果。

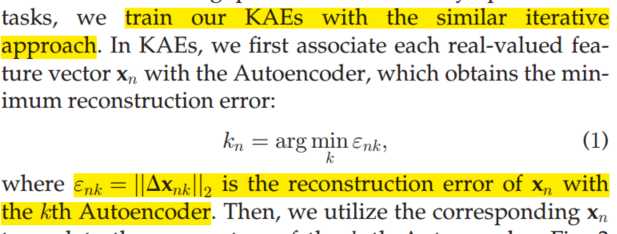
我们提出了一种K-Autoencoders（KAEs）网络来进行基于数据的二值化，并且利用multi-quantization（DBD-MQ），提出了一个深度二进制描述子。

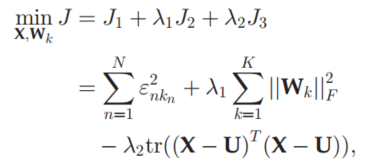
主要优点：

1）多量化深度二进制算法，利用固定的总的bit长度，让不同的实数特征维度对应多个bits来获取更多信息；

2）通过设计一个基于相似度量的Autoencoders的EMD，设计了一种similarity-aware binary encoding strategy，让相似的量化器有更小的汉明距离。

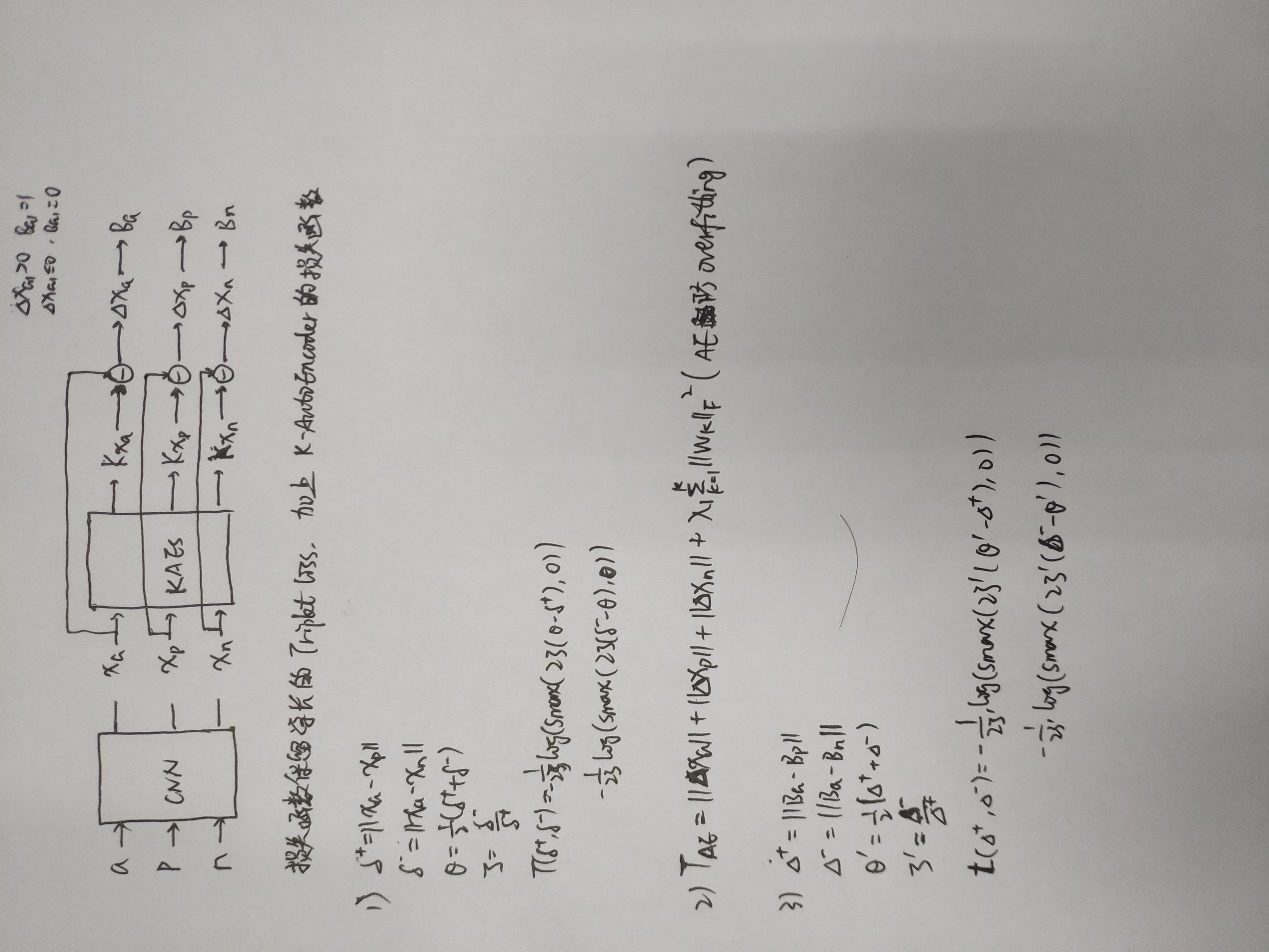






本文是利用EMD方法，设计两个AE，来让相似的样本得到的二进制编码的距离最小。（这里可以仔细阅读一下）。

想法：而在ASDNet里面，主要是利用的triplet network来实现浮点描述子的相似图片距离更小，不同图片距离更大来实现的。所以在我们的网络里面，可以利用triplet优势，让生成的二进制码的anchor图片和positive的汉明距离更小，anchor和negative的hamming距离更大。



模型量化误差：<https://zhuanlan.zhihu.com/p/132561405>

<https://mooc.yanxishe.com/course/473/learn?lessonid=2590#lesson/2590>

Deep Binaries: Encoding Semantic-Rich Cues for Efficient Textual-Visual Cross Retrieval

arXiv:1708.02531v1

contribution：提出了deep binary encoding method termed as Textual-Visual Deep Binaries (TVDB)，能够编码信息丰富的图片和描述性句子。两个modal-specific binary encoding networks使用LSTMhetext-CNN搭建的，在单词之间提升了图片区块信息和语义信息来获取高质量二进制表示。并且提出了随机批量编码学习方法来进行训练

drawback：没有特别说用于slam的描述子

Learning multifunctional binary codes for both category and attribute oriented retrieval tasks

现存的大多数hash算法只能用于防止单一的相似性，不能同时解决不同人物。我们设计一个新的hash算法，叫Dual Purpose Hashing (DPH)，能用CNN去分层获取类别和属性之间的联系来同时保存类别和属性。我们用的是互联网上大量部分标记的图片进行训练

![F:\Users\gxf\AppData\Roaming\Tencent\Users\740297955\QQ\WinTemp\RichOle\R[4RZ6G`9I{2{IHBNMMR](9.png](data:image/png;base64,)

Learning Rotation-Invariant Local Binary Descriptor（跟Learning Compact Binary Descriptors with Unsupervised Deep Neural Networks的损失函数设计很像）

1. BRIEF，BRISK，ORB，FREAK等二进制描述子对于局部变化具有很强的鲁棒性，并且通过汉明距离代替二进制特征表现出很高的计算效率。然而，大多数局部二进制特征是手工制作的，这需要很强的先验知识和启发式

最近，一些基于学习的局部二进制描述器被提出来直接学习散列过滤器将图像块投影到二进制码[2]，[20]来解决这个局限性。与手工编码方法相比，学习后的二进制编码具有结构紧凑、节能、分布均匀等优点，具有更强的鉴别能力。此外，基于学习的局部二进制描述符更具有数据自适应性。然而，这些基于学习的方法对旋转非常敏感，不适用于具有大旋转变化的数据库，如纹理分类、不对齐的人脸识别和场景分类，也不适用于一些测试中旋转未知的实际应用。针对这一缺陷，我们提出了一种基于局部区域学习方向和特征投影散列函数的旋转不变局部二进制描述子(RI-LBD)。

3. 如何实现旋转不变性的：O-PDV进行特定角度的旋转，计算权重和，然后极小化。

为O-PDV设计了一个benchmark，使其RBP的能量最小，则所有旋转变化的O-PDV都有相同的benchmark O-PDV。

一个显然的想法就是将每个O-PDV旋转到benchmark O-PDV，然而benchmark不好获取，对于一个patch来说benchmark orientation对于好的二进制编码来说不是最好的选择。我们设计一个benchmark O-PDVs作为初始值，然后设计一个联合学习的方法，通过把RBPs的能量添加到objective function里面来获取更多数据适应性好的orientation。

4. 如何实现二进制的：O-PDV：周围的pixel减去中心pixel的intensity，然后从左上角开始顺时针排列成vector，θr=2π/2r，如果旋转角度不是θr的倍数，则应该用插值。因为图像块很小，所以可以视为圆形。

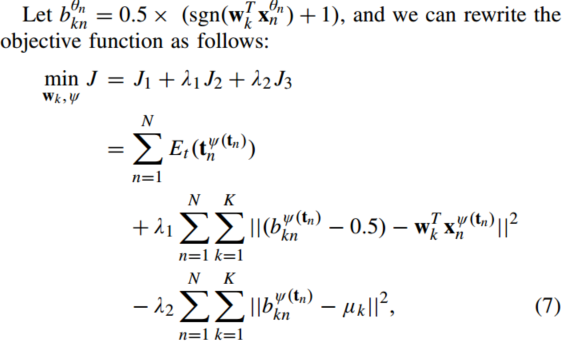
为了减小噪音的影响，利用sgn(x)函数，将O-PDV二值化（0.5\*(sgn(x)+1)≥0的位置则为1，否则为0）

相关文献：21-25提出了一系列复杂的二进制描述符

3,26-29 feature extraction 2,20基于无监督的学习方法也获得了较好地效果，但是对旋转敏感

Co-occurrence feature：有关联的feature。与一般的co-occurence features手动设置不同，我们直接从原始像素中学习了triple rotation-invariant co-occurrence local binary descriptors

二进制的实现跟下面的论文很像



where **N** is the number of patches extracted from originalimages, **K** is the length of each binary feature, **μk** is themean of the **k**th bit of all **N** input vectors（所有输入向量的第k位的均值）, and **λ**1 and **λ**2are parameters to balance the weight of different terms.

J1 is to minimize the energy of each rotated O-PDV in order to obtain rotation-invariance, J2 is to reduce the loss of the quantization between the original O-PDV and the learned binary codes, and J3 is to maximize the variance of the learned binary codes to make them more independent。

We relax the non-linear sgn(**·**) function as its signed magnitude [36], [37] as it makes (9) an NP-hard problem

Learning Compact Binary Descriptors with Unsupervised Deep Neural Networks

Abstract：

没有使用random projections 或者linear hash functions，用无监督方式来设计二进制描述子。

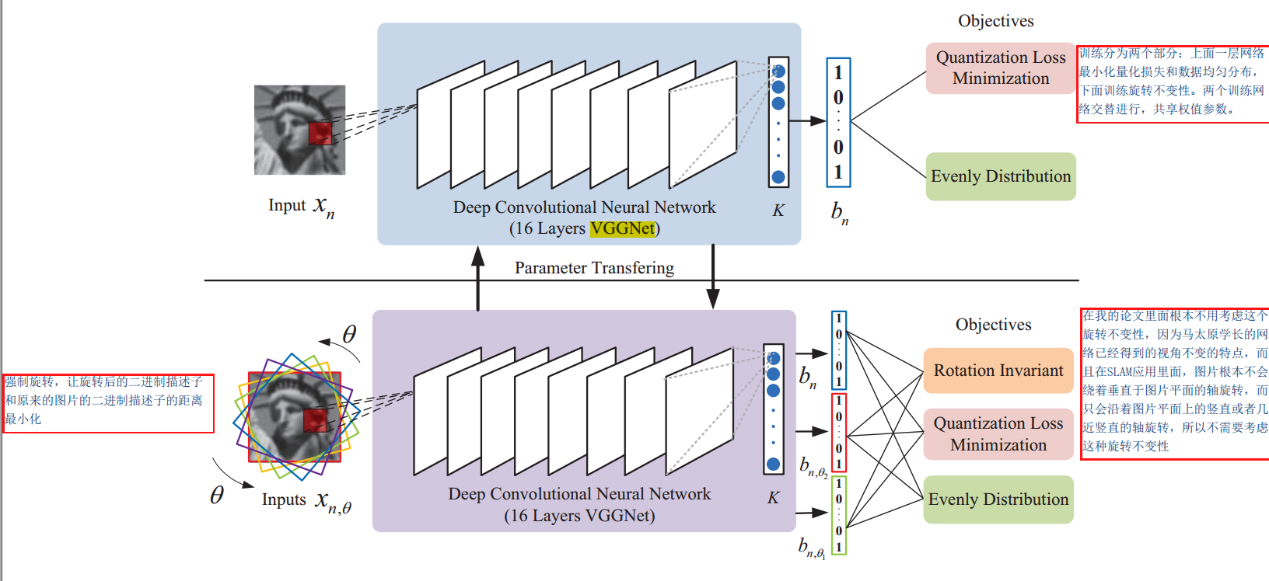
网络三个优点：

1）最小的量化损失

2）平均分布的编码

3）相互无关联的bits

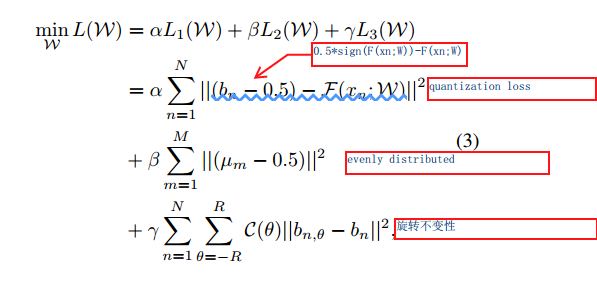
网络结构：



训练分为两个部分：上面一层网络最小化量化损失和数据均匀分布，下面训练旋转不变性。两个训练网络交替进行，共享权值参数。

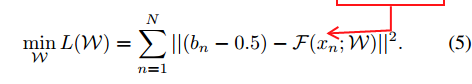
F:\Users\gxf\AppData\Local\Temp\1595899762(1).png

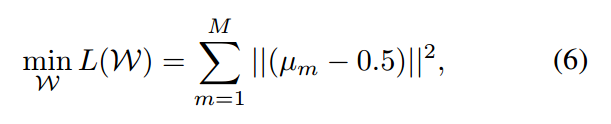
输入x（实数型的128列描述子，其实这里只是一个image patch），然后经过一系列的网络，里面有很多不同的权重W，最终得到F(x;W),这个也应该是128列，然后经过公式(1),即可得到128列的二进制串

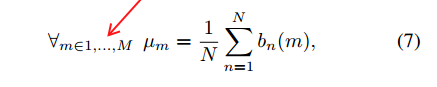


F:\Users\gxf\AppData\Local\Temp\1595900063(1).png

直接学习更新W，而不是迭代更新W和b

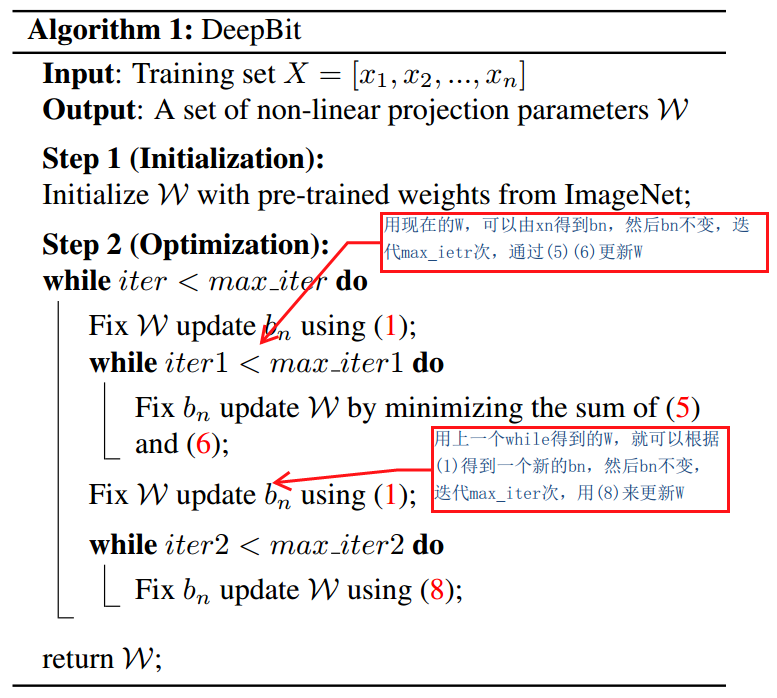


M表示bit length



所有的训练数据生成的二进制描述子的每个bit的值进行求和取平均，最后让他尽量靠近0.5，则0和1的总数量就相等了，则bit分布均匀了。

α=1.0，β=1.0



注意：在我的论文里面根本不用考虑这个旋转不变性，因为马太原学长的网络已经得到的视角不变的特点，而且在SLAM应用里面，图片根本不会绕着垂直于图片平面的轴旋转，而只会沿着图片平面上的竖直或者几近竖直的轴旋转，所以不需要考虑这种旋转不变性。SLAM的图片描述子和用于检索的图片描述子还是有区别的。所以在SLAM里面只需要考虑perspective change（视角改变），而不需要考虑rotation invariant（旋转不变性）

|  |  |
| --- | --- |
| F:\Users\gxf\AppData\Local\Temp\1595856822(1).png | F:\Users\gxf\AppData\Local\Temp\1595856858(1).png |
| SLAM | 图片，目标识别，图片匹配，无人机SLAM |

式（5）的优化不需要sign函数的反向传播，不过

Binarized Neural Networks: Training Neural Networks with Weights and Activations Constrained to +1 or -1

D-Brief：Efficient Discriminative Projectionsfor Compact Binary Descriptors

与SURF，BRISK，ORB作了对比，也是用的FAST，然后得到D-Brief描述子

二进制描述符的bits之间的compactness（关联）必须要尽可能小，则二进制描述符表达的内容才会更多。

Receptive fields selection for binary feature description

**总结：**

**二进制的方法就是最小化量化损失，以及让数据平均分布**