主要思想：特征匹配很重要，是视觉SLAM的瓶颈。改进了ASD，提出了二进制描述子，既提高了精度，又保证了效率，有实际意义。精度从FPR95，mAP说明

主要就是将神经网络二进制描述符用于视觉SLAM，效果更好。

关键点提取，描述符计算，描述子匹配，后端优化

基于深度学习二进制描述符的SLAM系统研究

Abstract：基于特征匹配的SLAM受描述符的影响比较大，CNN对于image patch信息的提取的效果比较好，适用于生成描述符。所以本文设计了一种基于CNN的描述符，效果比较好，在数据集HPatches和Brown dataset上的鲁棒性都比较好，改进的SLAM系统在数据集kitti和tartanair上的结果比传统的ORB-SLAM2要好

介绍二进制描述符的生成，主要是几个损失函数介绍一下，然后就是改进的ORB-SLAM2，进行对比实验，说明我的二进制描述符的FPR95比较好，HPatches数据集的几种模式我的这个表现好，来间接说明我的二进制描述符比一般的更鲁棒，然后说我的这个效率比浮点型的高，回环检测时间比浮点型的短。

系统框架：描述符的生成（几个损失函数介绍），SLAM系统的搭建（ORB-SLAM2）.

实验：Brown Dataset的FPR95；HPatches的图片检索等等（来说明我的描述符比较鲁棒）；Tartanair数据集的SLAM的对比；与浮点型描述子的回环检测时间的对比；

BASD-SLAM: A Deep-Learning Visual SLAM System Based On Binary Adaptive-Scale descriptor

Abstract:

The feature quality plays an important role in visual SLAM (Visual Simultaneous Localization and Mapping) based on feature matching, and becomes the bottleneck of positioning accuracy improvement. Now lots of hand-crafted descriptors like BRIEF and ORB don't work very well in complex scenarios. The Convolutional Neural Network is proved to have tremendous advantages on image feature extraction. In this paper, we design a CNN model to extract binary visual feature descriptor from image patches. Based on this deep feature descriptor, we design a monocular SLAM system, named BASD-SLAM, by replacing ORB descriptor in ORB-SLAM2. We also train visual Bag of Words to detect loop closure. Experiments show that our BASD achieves better results on the HPatches dataset and UBC benchmark. In the meantime, the BASD-SLAM system outperforms other current popular SLAM system on KITTI odometry dataset and Tartanair dataset.

描述子的质量在基于关键点的描述子匹配的视觉SLAM中很重要，成为了提高定位精度的瓶颈。现在大量的手工打造的描述子，如ORB，SIFT在复杂场景效果不是很好。而CNN被证明在图片编码方面有很大的优势。所以本文设计了一种基于CNN的二进制视觉描述子，并利用ORB-SLAM2的框架设计了SLAM系统。深度学习二进制描述子可以在提高轨迹估计精度的同时，也能提高特征匹配和后端优化效率。大量在brown dataset，HPatches和tartanair数据集上的实验证明了描述子的鲁棒性和高精度性。

1. INTRODUCTION

Introduction：

Visual SLAM has got prosperous development in recent years. The result of feature matching in the keypoint-based vSLAM system depends on the descriptor quality. The traditional descriptors rely on the pixel-level

视觉SLAM发展迅速。对于基于特征点的视觉SLAM来说，特征匹配的结果取决于描述子的质量。传统的描述子进行像素级的匹配，匹配误差会累积，最终影响pose estimation result。

Related work

Local feature descriptor

讲讲传统的描述子，浮点型的，二进制型的。讲讲CNN描述子，浮点型的，二进制型的。这些描述子都只是用于image verification,等等，没有在SLAM和SFM上进一步测试，

Proposed method

Deep learning enhanced SLAM

3 System Overview（SLAM框架）

In our BASD-SLAM system, we still adopt traditional visual SLAM pipeline. ORB-SLAM2 is a classical visual SLAM system. Unlike other end-to-end SLAM system, we just replace the traditional hand-crafted descriptor ORB with our learned descriptor and evaluate the efficiency and effectiveness of our descriptor. This also enables our descriptor suitable to other SLAM system like SFM.

3.1 Local feature design

Compared with hand-crafted methods, learned descriptors has tremendous advantages, such as compact structure, evenly distribution, robust to noise and so on. Moreover, learning-based descriptors are data-adaptive. In order to make our descriptor more effective, we just adopt shallow convolutional neural network to generate our descriptor, and the shallow network has also been proved to be suitable to extract low-level image information [14]. float descriptors sacrifice the effectiveness of feature matching, loop closure. Instead, our shallow network will obtain binary local feature descriptor, and also maintain the high precision.

[15] reveals that triplet network has greater advantages in metric learning than Siamese network, so we also adopt the former to train our descriptor. There are eight convolutional layer, which is followed by a ReLU non-linearity and Batch Norm operations. And the output of network is normalized to unit-length. In order to reduce the possibility of overfitting, we add a dropout layer in the last of our network. After lots of tuning step and training process, we set the dropout rate to 0.3.

Loss function plays an important role in descriptor generation.

1. Adaptive-Scale Triplet Loss

Triplet loss has been proved to have great advantages in descriptor generation. So we also adopt this loss function. [ASD-SLAM] proposed the scale uncertainty influence in triplet loss, and modified the prime triplet loss function to reflect the changes of scale by adding a scale reminder factor. Given three image patches, Pa, Pb and Pc, which represent the anchor, positive and negative image patches. After the reasoning of network, we get descriptors xa, xb and xc respectively. And the adaptive-scale triplet loss function is defined as





Where d- and d+ is the L2 distance of anchor descriptor with negative descriptor, anchor descriptor with positive descriptor, respectively.

Because we set batch size to 1024, so the choice of d- and d+ matters. We also adopt the adaptive-scale sampling strategy to obtain suitable d- and d+.

The native training strategy is too complex and performs not well, so we turn to the hard negative mining strategy proposed in [hardnet], which is proved to be effective and easy to converge in training.

1. Even-Distribution Loss

The distribution of binary bits reflects the encoding quality of neural network. In large dataset, same bit of every descriptor generated by all image patches should have same numbers of -1 and +1. However, the sign function is not differentiable, so we cannot reduce even-distribution loss by optimizing the numbers of -1 and +1. We just constraint the means of every float descriptor dimension in one batch size descriptors to 0. Even-distribution loss is defined as :



1. Quantization Loss

In quantization step, we use sign function to obtain binarization result of float descriptor. However, the difference between real-value and ±1 can bring a great drop in accuracy. So we minimize the quantization loss to get a better binary descriptor. Quantization loss is defined as:



1. Correlation Loss

In order to make the descriptor contain more information, the bits of every descriptor should have less correlation[L2-net]. So we introduce the correlation loss penalty to get more differentiable descriptors. We use the descriptor generated from anchor image patch, where  is row vector of one image descriptor.

The correlation matrix  is defined as:



Where  is mean of ith row of . Obviously, the off-diagonal elements of R should be 0. So the correlation loss is:



3.2 SLAM system

4.Experiments

4.1 descriptor evaluation

4.1.1 UBC benchmark

4.1.2 Matching result in hard scenarios

4.2 SLAM system evaluation

4.2.1 Evaluation of KITTI Odometry dataset

4.2.2 Evaluation of Tartanair dataset

四个损失函数的讲解

Slam系统的搭建，词袋的训练，

Experiments

Brown dataset的fpr95

图片的匹配

Tartanair的实验

Kitti数据集

[14] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, and T. Brox, “Discriminative unsupervised feature learning with exemplar convolutional neural networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 9, pp. 1734–1747, Sep. 2016.