**Abstract：**

直接从图像数据中学习通用的相似性函数用来比较图像块。选择了一个基于CNN的模型，可应对图像外观的广泛变化。研究了多种神经网络架构。这种方法优于最先进的技术。

**Introduction**

比较图像间的块是一类最基本的任务，扮演重要角色。

决定块是否对应很有挑着性因为有很多因素影响图像外观。比较块的需求促进了很多手工设计的特征描述符的发展，如SIFT。

手工设计的描述符不能顾及所有影响因素。

当下，可容易访问包含图像块间对应关系的大型数据库。

本论文的目标即利用数据集来学习图像块地相似函数：（也收到神经网络和深度架构最近的进展）选择用深度卷积神经网络来表示此函数。并且要解决什么样的网络架构最适用。因此探索和提出了不同的网络类型。训练时，使用包含原始图像块对的（匹配和不匹配）数据集最为单独输入。

本文贡献：1）直接从图像数据学针对块的通用的相似性函数，它考虑了各种类型的变换和影响。2）探索和提出了不同的神经网络模型 3）将方法应用于不同的问题和基准数据集，显著优于最先进的方法；它产生的特征描述符比起手动设计的有更好的表现，并在密集方式下也能有效计算

**Related work**

比较块的传统方法是描述符和平法欧式距离。

大多描述符是手工制作的，如SIFT或者DAISY。

最近，学习描述符的方法已被提出。【19：凸优化】

----------------------------------------------------

我们的方法受最近成功的卷积神经网络的启发。

Fisher[2014]分析了AlexNet网络卷积描述符的表现（在Imagenet数据集训练），在大多数情况下比SIFT好；他们也提出了一个非监督方法获取描述符，比SIFT和Imagenet训练的都好。

MC-CNN-1中提出了一种基于CNN来比较块计算小基线立体问题中的代价并且在KITTI取得了最好表现。然而，它仅仅是比较有非常小的块组成的对，比如在窄基线立体中。相反，我们针对的相似函数，是能够应付更广的外观改变、更广泛、更具挑战性的应用中，如宽基线立体，特征匹配，图像检索。

**Architectures**

为了和其他方法对比，仅选择灰度级的块训练。除了SPP，块的大小固定64×64（原始块需重新调整大小）。

研究三种基本神经网络架构，2-channel，Siamese，Pseudo-siamese。本质上，架构的不同源于他们解决下面问题的不同方式：是先计算每个块的的描述符然后在此基础上创建相似性还是直接进行相似性估计？

**3.1Basic models**

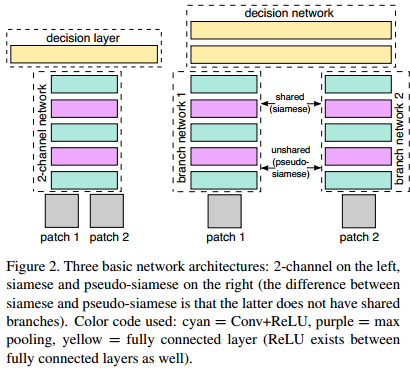
**Siamese**：效仿了有描述符的思想。

网络中的两个分支共享相同的结构和相同的权重集。

分支输出被连接并给到一个顶网络。顶网路被ReLU激活曾分隔开。

分支能够被看作描述符计算模块，顶网络被看作相似函数。

**Pseudo-siames:**复杂性在siamese和2-channel之间。权重不共享，增加了参数，比Siamese灵活，但不如2-channel，保持了测试时siamese网络的效率。

**2-channel**：将输入的两块看作一个双通道的图像。顶层仅由一个带着一个输出的全连接线性决定层组成。

灵活性最高（因为从共同处理两个补丁开始），训练快，测试时更昂贵（expensive）（需要所有块的组合相互之间测试以一种粗暴的方式）

**3.1Additional models**

**Deep network**

将更大的卷积打破成为3×3的核，由ReLU激活分开，它提高了网络内部的非线性性并使决定函数更加有判断力。

**Central-surround two-stream network**

包含两个分开的流，中心和周围，这使得能够在两个不同分辨率的空间区域进行处理。

中心高分辨率流接收两个32×32的块作为输入（通过在64×64的块的中心剪裁得到）。

环绕的地分辨的流接收两个32×32的块作为输入（通过对原始图像块的一般进行下采样得到）

得到的两个流能够被基础架构使用。

多分辨率信息在提高图像匹配表现方面是重要的。

通过两次考虑一个块的中心部分，更多的注意力靠近块中心的像素上，更少地关注图像的外围，有助于提高匹配的精确性。（本质上，因为池化被用于下采样图像，外围的像素在匹配的过程中被允许有更大的变化））

总的输入维度减少了两倍，训练进行得更快

**SPP network for comparing patches**

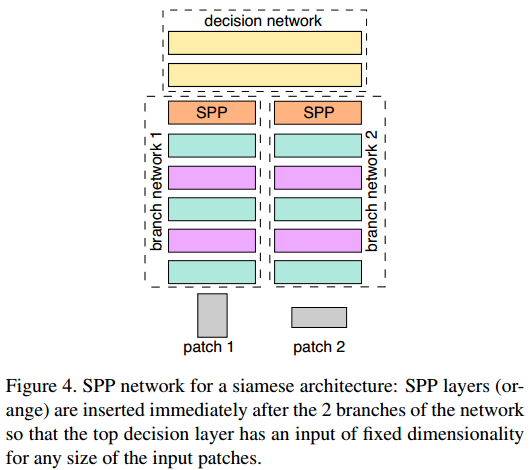
本质上相当于网络的卷积层和全连接层之间插入一个空间金字塔池化层。这样的一层聚集了最后一个卷积层的调整。

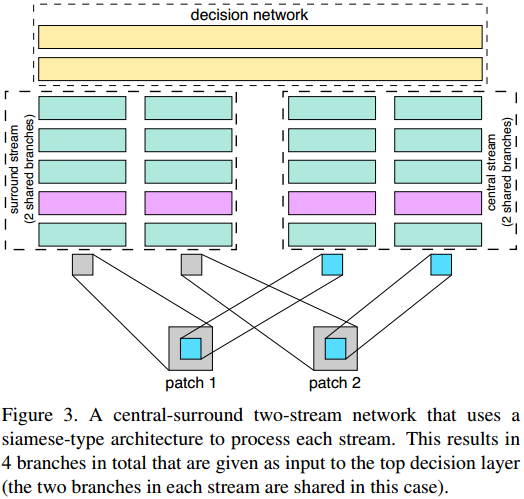
**4.Learning**

**Optimization.**

使用一个基于hinge损失和平方L2范式正则化，得到下面的学习目标函数

,





ASGD，学习率为1，动量为0.9，权重衰减λ=0.0005。128的小批量。权重被随机初始化

**Data Augmentation and preprocessing**

扩充了训练数据通过水平和垂直翻转两个补丁并旋转到90度，180度，270度。

在Torch中使用Titan GPU，卷积例行程序（routines）取自Nvidia cuDNN库

siamese描述符在GPU上只比在CPU上计算SIFT描述符慢两倍并在GPU上的Imagenet描述符快两倍。

**5.Experient**

应用我们的模型到各种各样的模型和数据集。

5.1. **Local image patches benchmark**

对于模型的第一个分析，使用了标准基准数据集[3:2010], 它由三个子集组成，Yosemite，Notre Dame，和Liberty，它们中的每一个包含超过450000个图像对（64×64）在高斯差分特征点周围取样。块被尺度和方向标准化。每个子集被生成通过使用实际的3D对应被获取通过多视图立体深度图。这些图被用来产生500000个真值特征点对为每个数据集，拥有相等数量的正（正确的）和负（错误的）匹配。

评价模型，通过给描述符空间的特征对之间设置阈值生成ROC曲线，FPR95。mean(1，4)为[1][3]中使用的四种组合的均值。

表1公布了几个模型的表现，也详述了他们的架构（我们也实验了更小的核，更少的池化层，和添加正则化，并没有注意到在表现上的显著提高）。

基于2-channel的架构展示了最好的表现，这表明从网络的第一层直接联合使用两个块的信息是重要的。

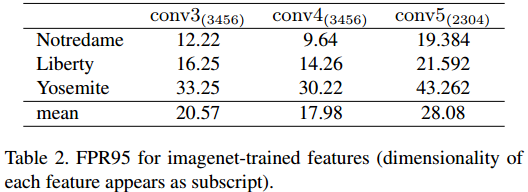
2ch-2stram网络是表现最好的网络在这个数据集上，2ch-deep紧随其后（这证明了在匹配过程中多分辨率信息的重要性和提高网络深度也有帮助？？）

基于siamese的架构，这些也能实现更好的表现比起已有的最好的系统。这是相当有意思的因为，比如这些siamese网络没有尝试学习池化区域的形状、大小和位置（像，例如[19,3]做的那样），而是仅仅使用标准的最大池化层。

在这些siamese网络中，二流网络（siam-2stream）有最好的表现，再次证实了在比较图像块时多分辨率信息的重要性。

psudo-siamese网络（pseudo-siam）比对应的siamese（siam）更好。

也测试了siamese模型的表现当它们的顶层决定层被由网络的两个分支产生的两个卷积描述符的l2欧式距离所替换。双流网络（siam-2stream-l2）计算了更好的距离比起samese网络（siam-l2）。siam-2stream-l2网络甚至比之前最先进的描述符[19: co]表现得更好，这相当得惊人考虑到这些siamese模型从没有用L2距离训练过。

我们在表2中展示了imagenet训练的CNN特征（这些被l2标准化来提高效果）。 在这些之中，conv4给出了最好的FPR95分数，等于17.98。这比SIFT要好，但仍比我们的模型差很多。

