Convolutional neural network architecture for geometric matching

阅读报告

1. 论文主要内容

估计不同图像中物体之间的几何关系是算计视觉的基本问题之一。

传统方法的主要过程是：检测和匹配特征点（如SIFT和HOG），用几何约束来筛选匹配，最后用鲁棒的方法来估计几何变换，如RANSAC和Hough变换。在大多数情况下，传统方法都有比较好的效果。但在物体的外貌发生变换（类内替换）时传统方法很难成功。而在场景发生较大的变换，或发生复杂的非刚性变换时，传统方法的效果也并不理想。

这篇文章在传统方法的基础上设计了一个卷积神经网络，来模拟一般的匹配方法。他们用卷积神经网络来代替特征提取，并且用可训练的匹配层和参数估计层来提高对噪声和错误匹配的鲁棒性。最后的结果可以用来处理外表有较大变化的物体，因此在实例匹配和类别匹配上都可以应用。

主要贡献有三方面：

1. 提出了一种卷积神经网络来进行几何关系匹配，这个神经网络主要由三个部分组成：特征提取、匹配和参数估计。
2. 证明了该神经网络可以用合成的图片进行训练，而不需要人工标注。并且对从未见过的图像有较好的泛化能力。
3. 该模型可以同时应用在实例匹配和类别匹配。
4. 相关工作
5. 网络架构

传统架构：

1. 采用局部描述子（如SIFT）从输入图像中提取数据
2. 通过描述子之间的匹配得到两张输入图像的初步匹配
3. 用得到的初步匹配和一些鲁棒的方法（如RANSAC和Hough变换）来估计几何模型的参数

论文中模仿了传统方法的架构，其结构如下：

1. 用卷积层来来提取输入图像A和B的特征图fa，fb
2. 匹配特征图fa和fb，得到初步的特征对应图fab
3. 最后用回归网络来输出几何模型的参数（一个6维向量）

结构如图：

* 1. 特征提取

第一部分是一个用CNN实现的特征提取层。它的输入是一张图像，输出是一个h\*w\*d的特征图，其中d是特征维度。一些物体检测中也有用到类似的CNN结构，因此，文中使用了VGG-16网络的前几层（到pool4层，在ReLU之前），并对每个特征进行了L2归一化。文中使用了预先训练好的模型，它最初是用来做ImageNet上的图像分类任务的。如图X中所示，两张输入图像分别经过两个网络，这两个网络是共享参数的。

* 1. 匹配网络

通过特征提取网络得到的特征通过匹配层结合成一个单一张量输入到回归网络中。文章模拟了传统方法设计了该匹配层。

传统方法中的匹配过程主要是计算两个特征之间的相似度，并对其进行阈值化、保留最近邻等操作，去掉模糊匹配。整个过程不关心特征点本身。

在本文的匹配层中，模仿了传统方法，只关心特征点之间的相似度和空间位置，而不关心特征点本身。具体结构是一个匹配层加上正则化。匹配层计算所有特征之间的相似度，然后通过一些加工和正则化来减少模糊匹配的权重。

其中匹配层的输入是特征提取层的输出，即L2正则化后的特征图fa和fb，输出是一个相关性图cab如公式1（ij）和。。。。k=。。。。

如图

然后用一个正则化操作来减小模糊匹配的权重来生成初步的特征对应图fab。正则化的过程是一个ReLU加上一个L2正则化的操作。ReLU会去除负相关。而当fb中的某个特征只与fa中的一个特征有较大的相似度时，正则化会增加匹配的相关度，就像最近邻匹配一样。当fb中的某个特征与fa中的多个点匹配时，匹配的相关性就会减少，就像二近邻检验。

其中求相关性和正则化的操作都是可微的，这有利于反向传播。

* 1. 回归网络

回归网络用来估计两张图像间几何变换的参数。

传统方法中，通常用局部几何约束来对初步的特征匹配图进行约束，最后用RANSAC或Hough投票。

本文的用网络模仿了传统方法。用了两组卷积层接正则化层和ReLU层，并在最后加上一个全连接层来回归变换参数，如图X。

这种结构是以一种自底向上的方式进行的，有点像Hough投票，前面的卷积层进行投票，之后的网络来统计投票。其中，第一个卷积层能增强局部一致性。

全连接层的输入是特征数的平方，这样多的参数很难学习，而且会因为占用了太多内存而降低速度。另外，虽然文中用的是卷积网络，但它任然能学到比较大的变换，因为在相似图的局部空间中，包含了图像B的特征点对图A中所有特征点的相关性。

3.4层次架构

估计从图像之间的变换通常先估计一个简单地变换，然后逐步增加模型的复杂度。文章遵循同样的做法，先估计了一个仿射变换，这是一个6个自由度的线性变换，能进行平移、旋转、非线性同相的缩放和剪切。然后用估计出的变换对B进行操作得到B‘，再估计B‘和A之间的薄板样条变换（18个参数）。最终得到的几何变换是把两个变换结合起来的结果，也是一个薄板样条变换。如图X

1. 参数训练
   1. 损失函数

训练数据包括两张图像和给出的标准变换（），损失函数l是用来衡量估计出的变换（）和标准变换（）之间的距离。更重要的是用损失函数来计算梯度，并用随机梯度下降来最小化损失函数。

这个损失函数应该不针对特定的模型，因此它可以用来估计仿射变换和薄板样条变换。它应该独立于变换参数，因此，它不能直接对参数本身进行操作。所以，我把对变换后的图像进行比较，并以此定义损失函数。如下

其中。。。

4.2 用合成数据进行训练

这篇文章用已知几何变换的图像来进行训练，并用人工合成的方法来生成数据。对于任意一个二维的几何变换，和原始图像IA，可以生成变换后的图像Ib，并组成一组训练数据。如图

1. 实验结果

在Proposal Flow数据集上进行了实验。数据集中包括很多图片的类，每一类图片中有很多种图片，如汽车和鸭子的不同角度的图片，并有不同的背景。用文章的方法估计从图B变换到图A的几何变换。结果如图X。

文章用关键点的平均正确概率（PCK）来评判算法估计结果的好坏。当预测位置与关键点的距离在α。。。之内的话，就认为关键点预测正确。

用Tokyo Time Machine 数据集通过仿射变换和薄板长条变换生成了两个训练数据集。。。和。。。每个数据集包括4万张图片，其中2万张做训练，2万张做确认。