# 计算机视觉实践-改进实验报告

组员：张潇，李喆，王浩骅

[计算机视觉实践-改进实验报告 1](#_Toc20438)

[一、 实验目的 1](#_Toc7193)

[二、 实验原理 1](#_Toc6010)

[三、 实验步骤和代码 11](#_Toc28873)

[四、 实验结果 19](#_Toc32440)

[五、 实验分析与总结 24](#_Toc21192)

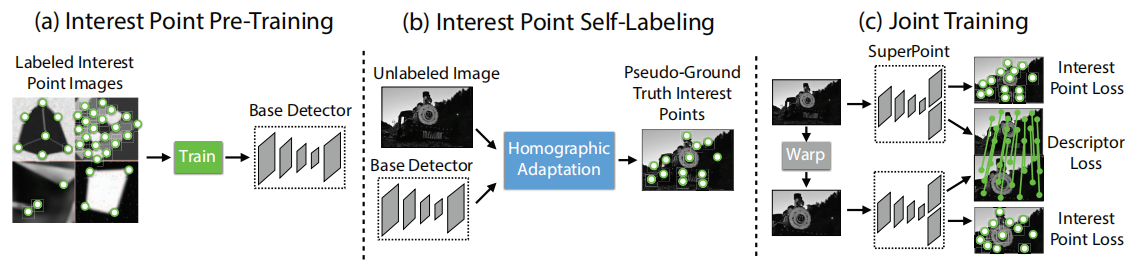
1. 实验目的

* 改进练习1，从三个方面优化图像拼接效果：特征点提取方面，特征匹配方面，图像变形方面。

1. 实验原理

**Superpoint：**

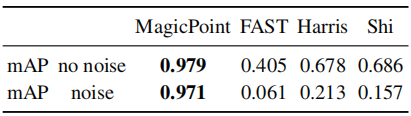
使用Superpoint技术替换原有的SIFT技术，提取图像中的角点，角点往往代表图像中局部结构的突变，大量的角点可以反映图像中物体的结构信息。Superpoint是Magic Leap提出的一种自监督的特征提取网络，可以提取图像上的关键点，特别是角点。其在MSCOCO数据集上取得了优秀的效果，且在GPU上的运行速度超过了一些传统的角点检测算法。其主要构建流程如下图：



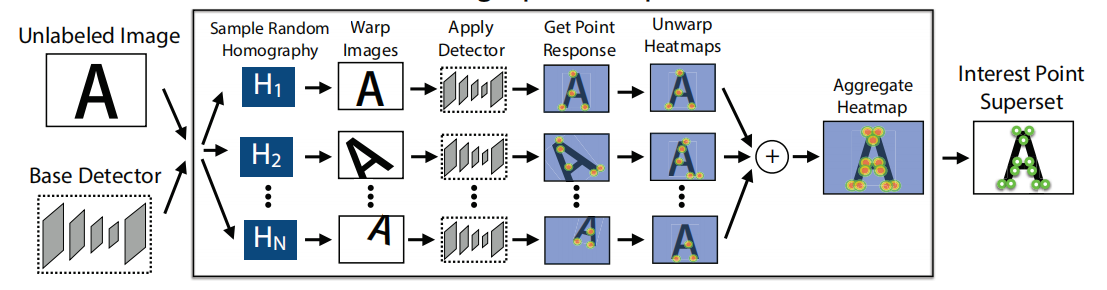
为避免进行大量的人工标注，如上图，该方法提出了一个合成数据集，该数据集构建了许多虚拟的三维物体，这些物体上的角点的位置在建模阶段容易得出，之后在不同视角观察这些合成物体并渲染图像，得出最终的合成数据集。随后使用这些数据训练一个特征点提取器作为特征点标注器，即MagicPoint(如图中的Base Detector)。MagicPoint是一种基于CNN的网络，可以根据输入的图像得出一张热力图，之后根据阈值可以得出一系列关键点。

这样做的好处是特征点标签的位置是绝对精准的，因为特征点是建模得来，其位置是已知的。但是，这样的特征点还不足以反映现实世界的特征，因为合成的形状只是现实图像的一个子集。

通过与传统算法对比，Magicpoint已经具备很好的效果，如下图：



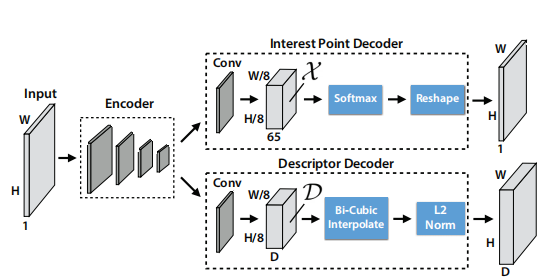
通过对比1000幅图像上的全类平均正确率(mAP)，可以发现Magicpoint的效果已经超过了许多传统方法。



上图展示了特征点的自标注过程。对于一张未标注的图像，对其进行若干次单应性变换后，使用MagicPoint对这些变形后的图片提取热力图，并将所有热力图叠加，得到一个合并的热力图，此过程被称作Homographic Adaptation。其优势在于，增强了检测器对不同角度、不同位置的角点的鲁棒性。

随后使用阈值可以得到在真实图像上标注的若干关键点。使用此自标注数据集进行训练的特征点提取器就是Superpoint。

即对原图进行单应性变换后，将原图和变换后的图片一起输入Superpoint网络并得到输出。随后使用这些输出计算损失并反向传播。Superpoint的网络结构如下：



其中的Encoder结构类似于VGG，主要用于对图像进行降维并提取特征。特征点解码器以H/8×W/8×65的张量作为输入，其中64个通道为8×8个互不重叠的局部图像块构成，还有一个通道作为dustbin。经过Softmax操作后，dustbin会被移除，重新将图像块转换成H×W的输出图像，其中的数据表示每一个像素点是关键点的概率。

这种网络结构的优点在于，可以对特征点解码器和描述符解码器同时训练，提升了训练效率；其次，其网络结构简单，运行效率高；解码器并没有使用上采样，这样可以避免网格效应，同时也减少了计算复杂度。

对描述符解码器进行训练时，不存在已经标注的真值。优化时的思想在于，对原始图像进行单应性变换后，其中的特征点的对应关系是可以知道的，因此需要让相对应的点的描述符的相似性尽量大，不对应的点的描述符的相似性尽量小。

特征点解码器的损失函数如下：

其中，特征点解码器的损失为：

容易看出，这是一个交叉熵损失的形式，代码中就使用了交叉熵来表达。

其中的X，X'分别代表原图和对原图进行随机变换的图像，形成一个图像对，其中。

下面的公式计算出图像块之间的匹配关系：

其中代表第(h,w)个图像块的中心坐标，H代表单应性矩阵。

描述符解码器的损失为：

其中

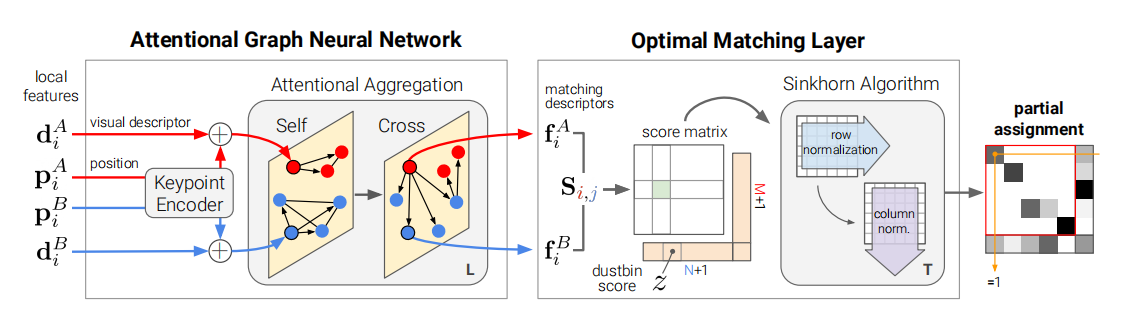
为描述符，。

为了使最后输出的特征点分布更均匀，并尽可能减少冗余信息，采用一种非极大值抑制方法。首先对输出的热力图进行一次最大池化，若该点得分在自身一定范围内最大则该点为选取的最终点。当进行一次选取后，已被选取的点及其周围的一定范围不再参与选取，而是在剩余的区域进行进一步选取，不停循环此过程。上述过程中的一定范围指radius×radius的正方形区域，参数radius可以手动指定，循环次数通过人为指定，一般取2次即可。

**Superglue：**

Superglue提出了一种能够同时进行特征匹配以及滤除外点的网络。其中特征匹配是通过求解可微分最优化转移问题（optimal transport problem）来解决。

整个框架由两个主要模块组成：注意力GNN以及最优匹配层。其中注意力GNN将特征点以及描述子编码成为一个向量（该向量可以理解为特征匹配向量），随后利用自我注意力以及交叉注意力来回增强（重复L次）这个向量f的特征匹配性能；随后进入最优匹配层，通过计算特征匹配向量的内积得到匹配度得分矩阵，然后通过Sinkhorn算法（迭代T次）解算出最优特征分配矩阵。



给定两张图片，每张图片上都有特征点位置以及对应的描述子，所以我们经常用来表示图像特征。第个特征可以表示为，其中表示特征点提取置信度，表示特征坐标；描述子可以表示为，其中表示特征维度，这里的特征可以是CNN特征，如SuperPoint，或者是传统特征SIFT。假设图像分别有个特征，可以表示为及。使用软分配矩阵表示点与点之间的匹配概率，。

对每个特征点进行特征映射，。令为图像A第i个元素在第l层的中间形式，是聚合了所有特征点之后的结果，其中,分别表示自注意力和交叉注意力。图像A中所有特征传递的残差信息是：

其中[|]表示串联操作，图像B上的特征进行类似的更新。经过L次自注意力和交叉注意力后，就可以得到GNN的输出，对图像A：

其中的理解为匹配描述子，用于特征匹配。通过内积计算得分：

使用Sinkhorn算法求解总体最大化问题：

其中为待求的分配矩阵。

对于已有的匹配标签，最小化负对数似然：

**高斯混合模型GMM:**

对于已有的点集，该方法提出的高斯混合模型：

其中，θ={T，σ2，γ}，T代表空间位置变换，T(xn)则是对点xn进行了空间位置变换后的点，γ∈[0,1]代表异常值的比率，这三个参数都是未知参数。其中的1/a为异常值的均匀分布。中的值可以由借助Kmn计算得出，预先设定参数τ，0≦τ≦1，以此代表匹配的可信度，若Kmn=1，则πmn=τ，且令中的值均为(1 − τ )/(N − 1)，若Kmn=0，则中的值均为1/N，以此保证。

给出该模型的对数似然：

其中，，zm为隐变量，zm=n代表将点ym分配给T(xn)。若能找到一组参数θ={T，σ2，γ}使得该对数似然达到最大，就认为得出了最优的结果。

为了求出最优参数，该方法提出了一种半监督的EM算法。

EM算法(Expectation-Maximization Algorithm)在两个阶段交替进行，分别是期待步(Expectation step，E-step)和极大步(Maximization step，M-step)。E-step用于使用已有的参数θold估计隐变量的后验分布，即。M-step尝试找出最大化似然函数的参数。两步不停交替迭代，直到参数不发生变化。

E-step中， 若Kmn=1，则令pmn=πmn，反之则使用如下公式更新pmn的值：

M-step中，取Q(θ)对γ和σ2的导数：

当EM算法收敛，即参数稳定后，使用一个预先定义的参数η，就可以得到最终的匹配结果集合：

(m,n)表示PY中的第m个点和PX中的第n个点是正确匹配关系。

下面求取EM算法中使用的空间变换T，该方法提出的T的建模：

其中为PX中随机选取的L个点，每个点都是一个二维向量，包含了点的二维平面坐标。其中，为对角高斯核函数。参数可由下式解出：

其中，为待求解的L×2的参数矩阵。符号d代表将一个列向量的值分布在对角线上。为N×L的矩阵。X为PX中的点，转化为N×2的矩阵，同理Y为M×2的矩阵，包含点的二维平面坐标，并对坐标值进行标准化。P代表一个M×N的矩阵，且。，W为一个N×N的权重矩阵，首先找到X中每一个点的K个最近邻，若对于X中的每一个点，若xj不属于xi的K个最近邻，令Wij=0，随后，解以下最小二乘问题，最小化代价函数：

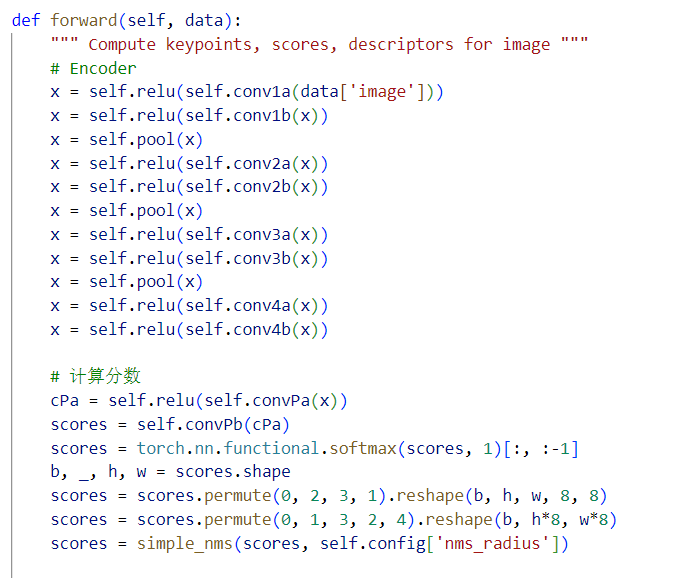
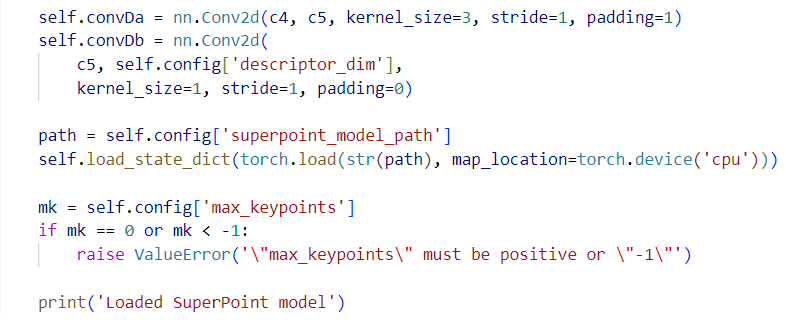
其中需要满足约束，计算出使代价最小的W。

以上式子中，λ、β、L、K均为预设参数，在本文的环境中，设置λ=1000，β=0.1，L=K=(N+M)/20，并向上取整。

1. 实验步骤和代码

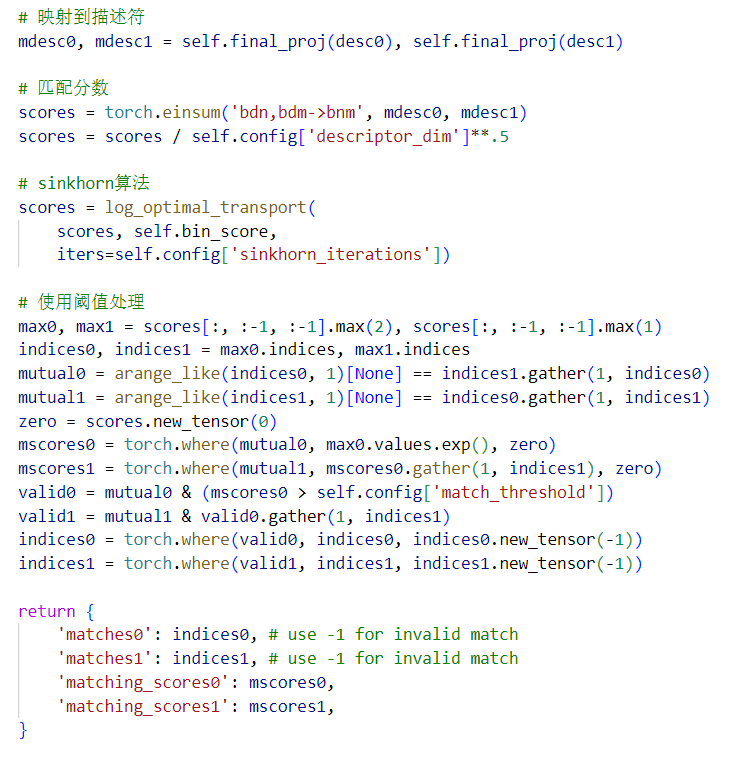
**下面仅展示主要代码：**

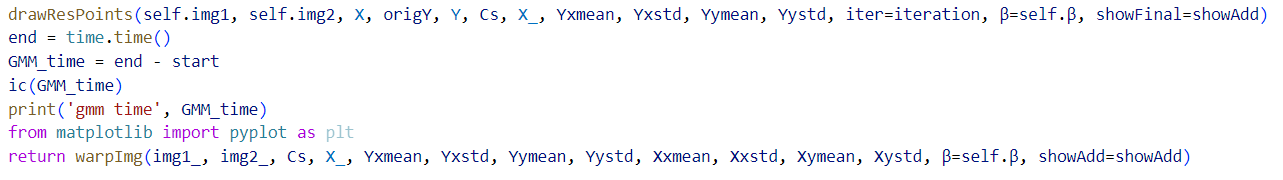
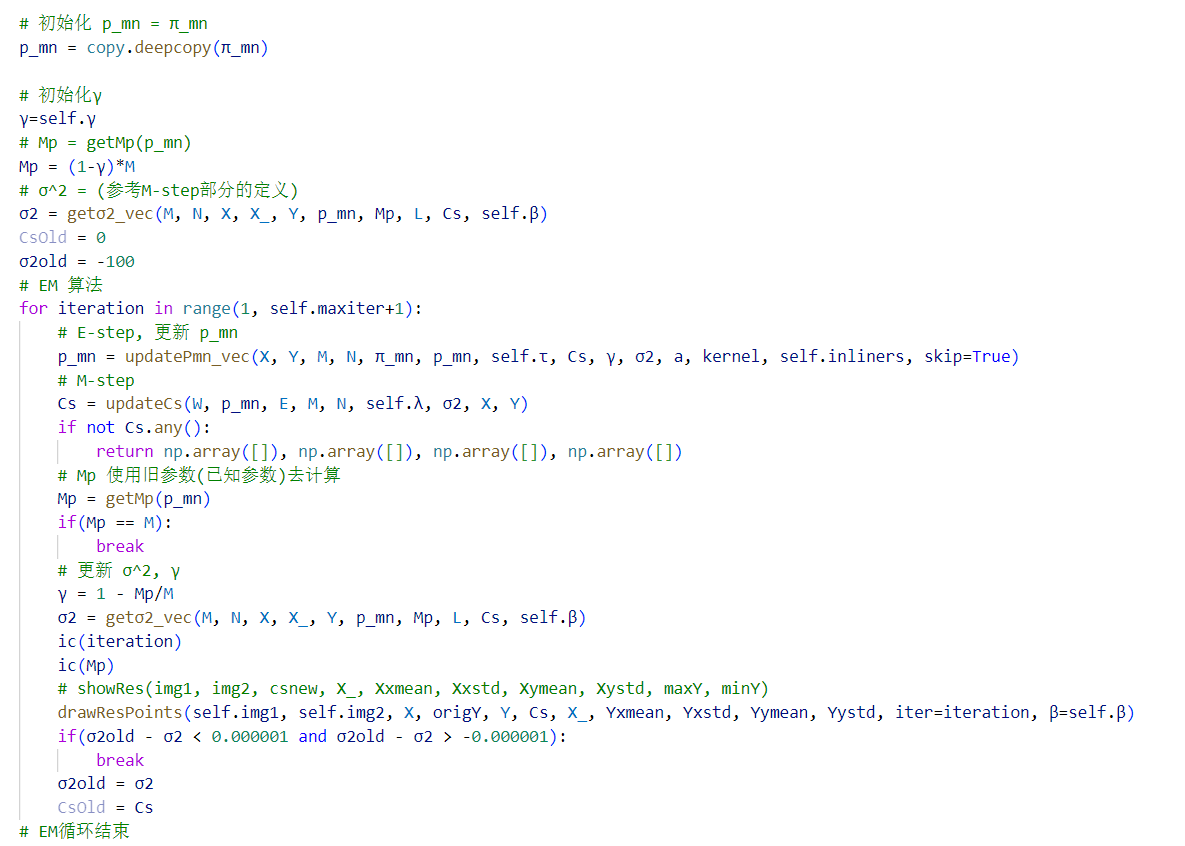
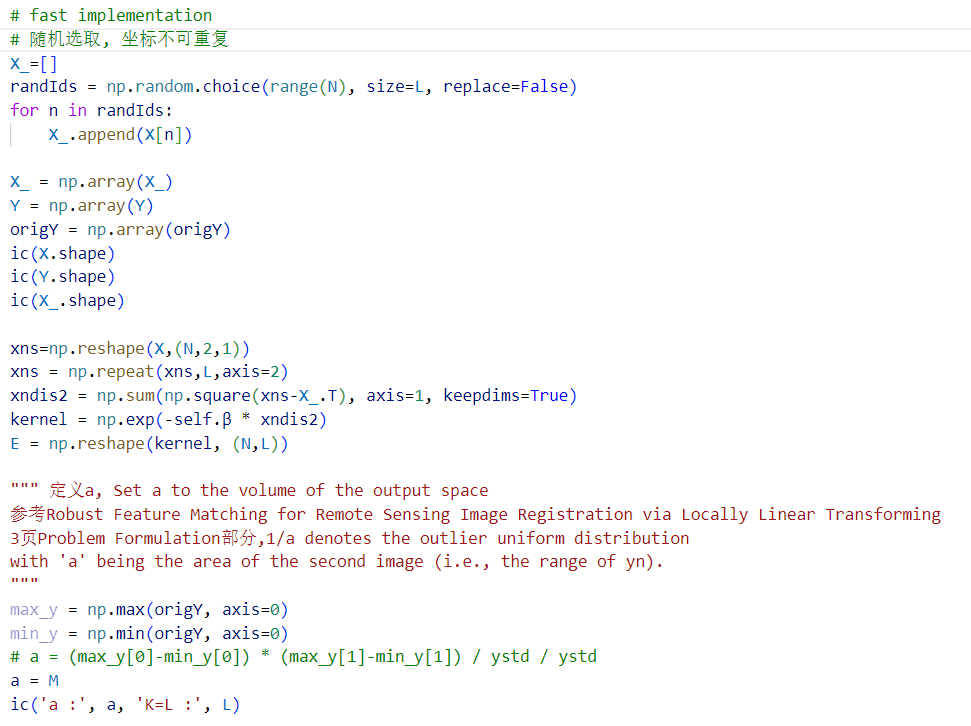
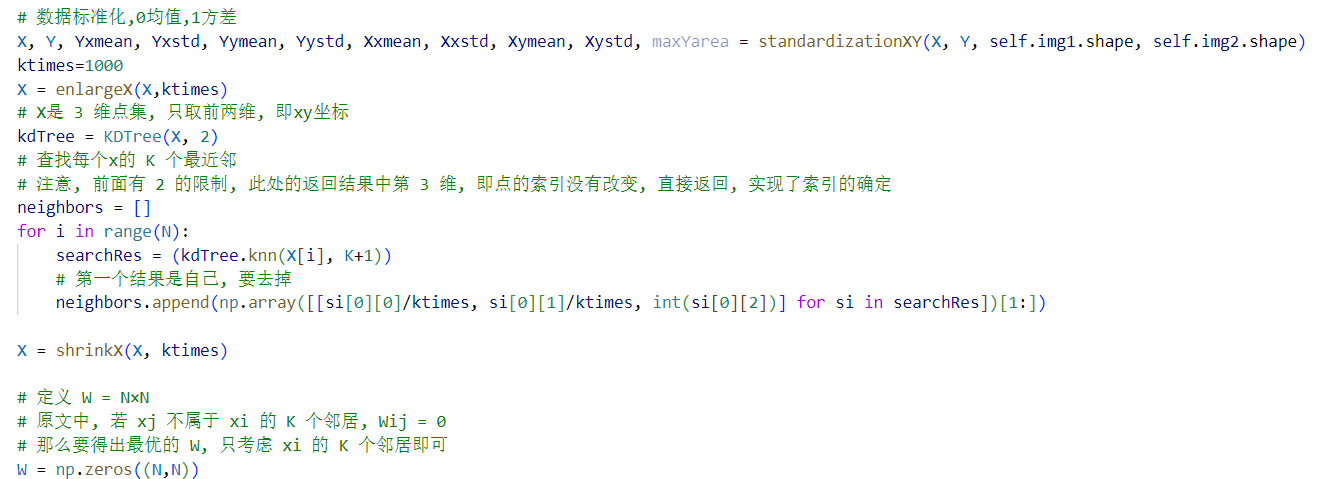
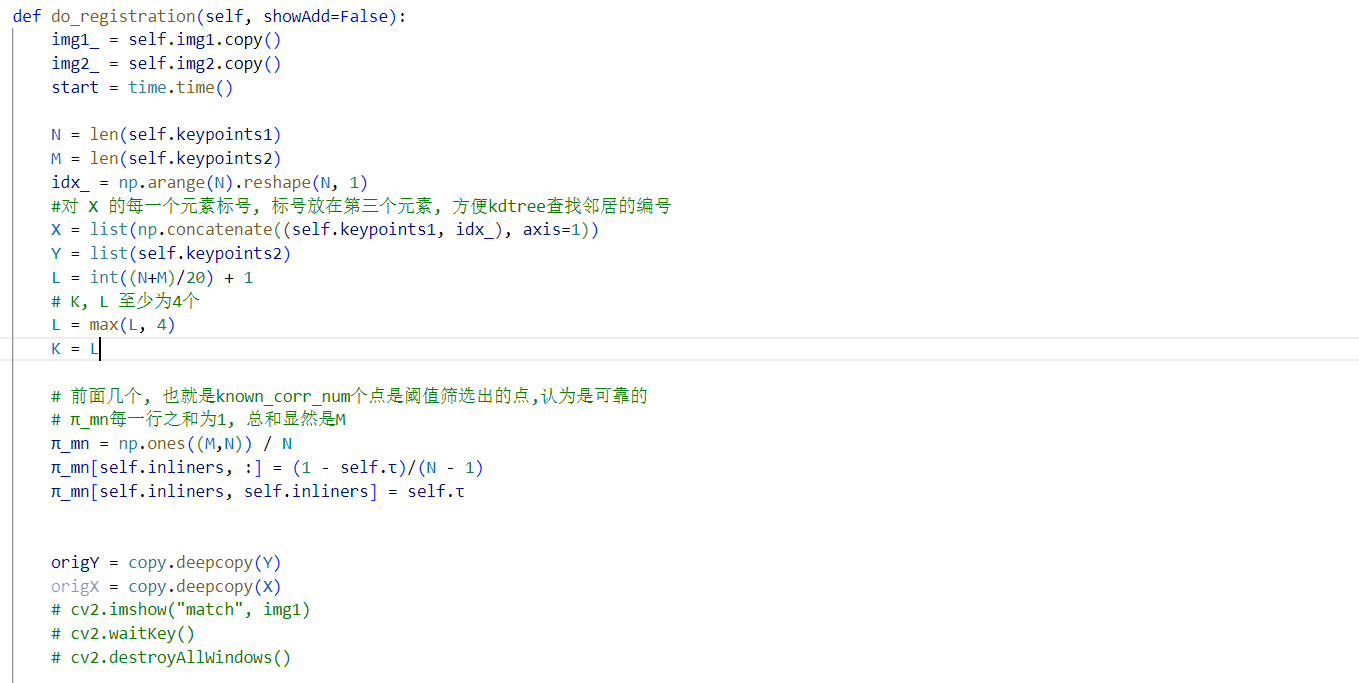
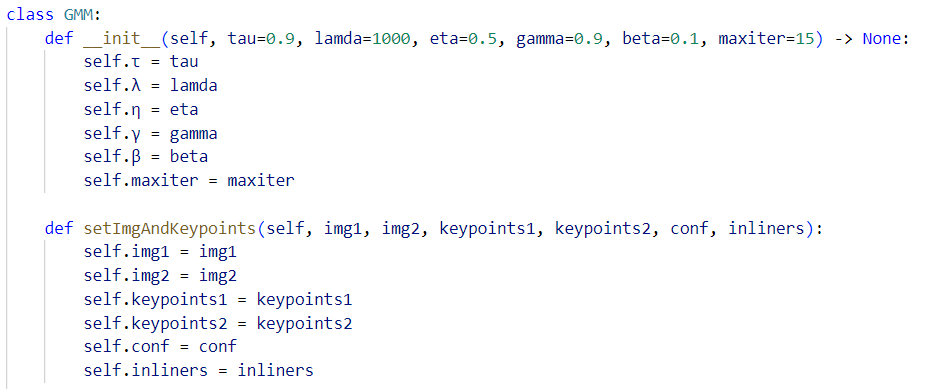
1. 检测图像中的特征点  
   Superpoint的网络结构：



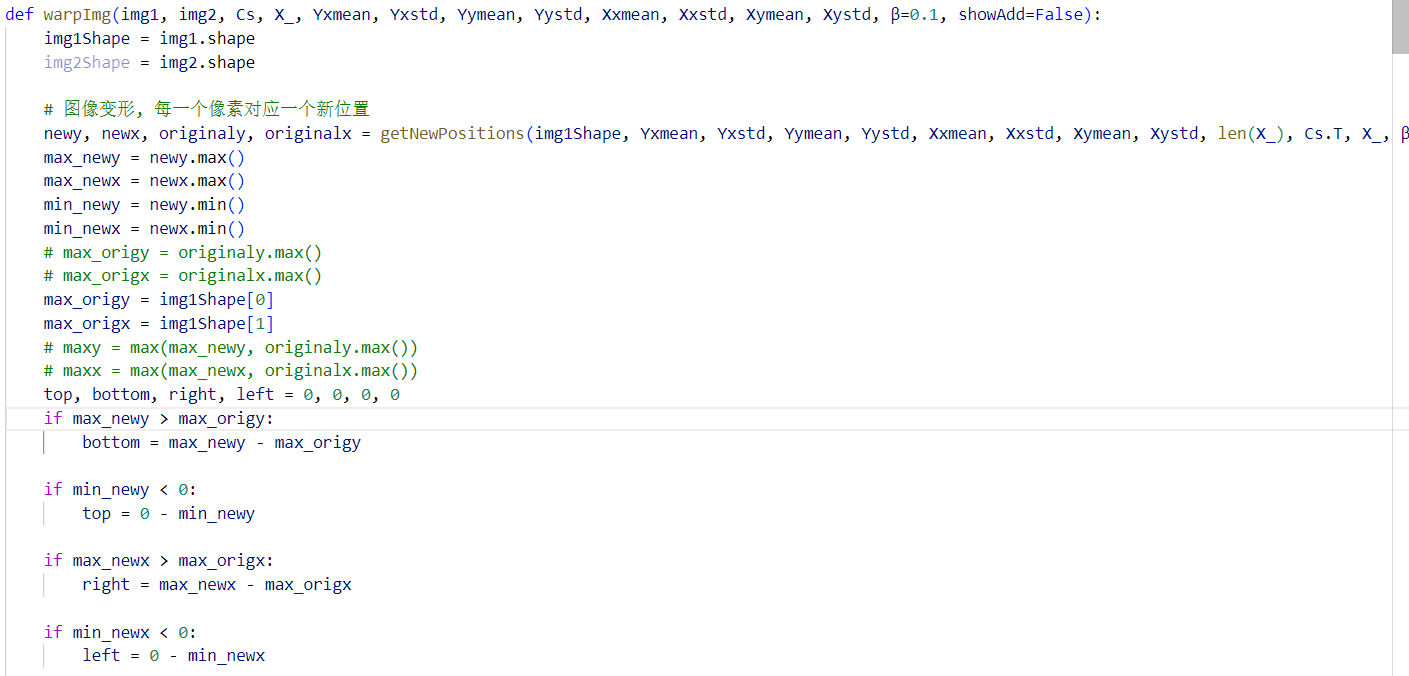
1. Superglue特征匹配  
   使用获得的特征点和描述符进一步输入Superglue网络进行匹配，得到匹配以及对应的分数。

网络结构：



1. GMM图像变换  
   

图像变形：





1. 实验结果

**输入图像：**







**对比：**

**使用SIFT提取特征点，暴力匹配特征，刚性变形：**







**使用本流程：**







1. 实验分析与总结

通过上述实验，从三个方面设计并实现了图像拼接的优化流程。其中，特征点提取和特征匹配的重要性是第一位的，是否能提取大量的匹配点对决定了后续拼接效果的上限。其次，非刚性变形的作用体现在，对于图像中不同的区域，形变程度往往是不同的。而刚性形变会引起整幅图像的改变，从而引起一些局部区域被过度拉伸（使用SIFT流程的第二幅图中右上角被拉伸）。

目前依然存在的问题是，缺少一种鲁棒的融合手段来处理光照不同引起的“接缝”问题（使用本流程测试的最后一幅图），一个可能的思路是在拼接处附近使用直方图匹配，由于时间有限，没有展开详细设计。此外，当匹配点较少时，非刚性变形容易受到一些外点的影响。当匹配点较多时，非刚性变形速度降低，可用性变差。