第三章 可视化

第一节 草图特征分析

首先我们手绘了199张草图，分辨率为512×512，可以分成11类：1添加头发、2添加新属性、3改变脸型、4改变眉毛、5改变眼睛形状、6改变眼睛大小、7无参照手绘、8改变嘴巴、9改变鼻子、10改变嘴巴（但眼睛与9相同）、11改变鼻子（但眼睛与8相同），除了第7类无参照手绘外，其余10个类别每一类中的草图都只改变相应的位置或属性而其他部分保持不变，无参照手绘中的草图相互之间没有关联。除第8类与第11类、第9类与第10类之间眼睛部位相同以外，其余各类间眼睛都不相同。手绘草图作为模型的输入，测试模型的生成效果。

Pix2pixHD的网络结构如下图所示：

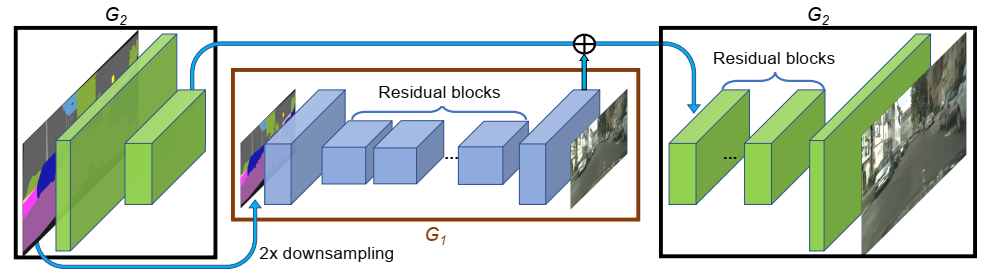


图2.1

我们以指代global generator，指代localenhancer。由于外层的只是单纯为了提高生成图像的分辨率，为了可以更直接地探究问题的本质，我们截取了作为模型的生成器。由图2.1可知，由5个卷积层、9个残差模块和5个逆卷积层组成，我们分别称前5个卷积层为global layer 0至global layer 4。由第二章可知，我们去掉了前两个卷积层的实例标准化操作。

把草图输入，得到global layer 0到global layer 4的特征图，然后对应输入草图的左眼眼角位置坐标(170, 250)，分别在5张特征图上抽取一个点的全通道特征向量，我们称之为vector 0到vector 4。5张特征图上对应点的坐标分别为(170, 250)、(85, 125)、(43, 63)、(22, 32)和(11, 16)，它们在输入草图上的感知野大小分别为7、9、13、21和37。感知野的示意图如下图所示：

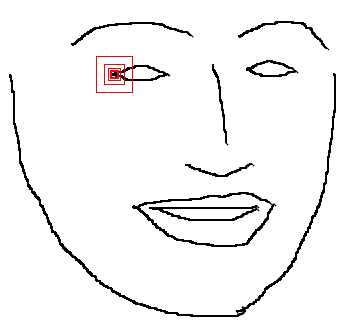


图2.2

将199张输入草图的vector 0和vector 1分别放在一起比较，发现没有改变眼睛的组，比如改变嘴巴和鼻子，这两个向量在类内是一致的；而改变了眼睛的组，如改变眼睛形状、改变眼睛大小、无参照手绘三组，这两个向量在它们各自的类内是不相等的。

第二节 PCA可视化结果

接下来用PCA对vector 0和vector 1进行降维后的可视化分析：

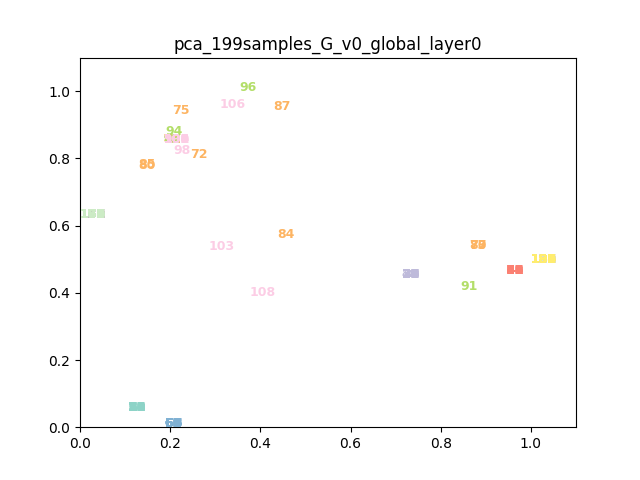


图2.3 vector 0的PCA可视化结果

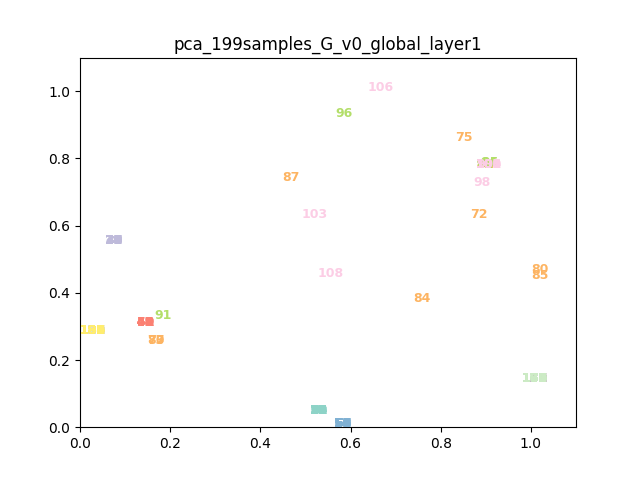


图2.4 vector 1的PCA可视化结果

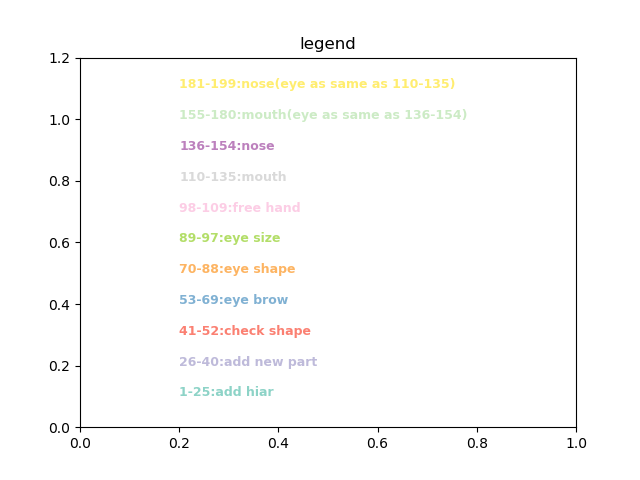


图2.5 图例

图2.3和图2.4分别是对vector 0和vector 1用PCA降维后可视化的结果，从图中可以看出第1、2、3、4类数据各自分散地分布于同一点，第8、11类数据分布于同一点，第9、10类数据分布于同一点，而第5、6、7类数据分布的非常分散。这说明去掉前两层的实例标准化操作之后，左眼角角点位置的特征只受感知野内的草图内容的影响，感知野内草图的改变会使提取到的对应点的特征发生变化，而改变草图其他部分不会对该点的特征产生作用。

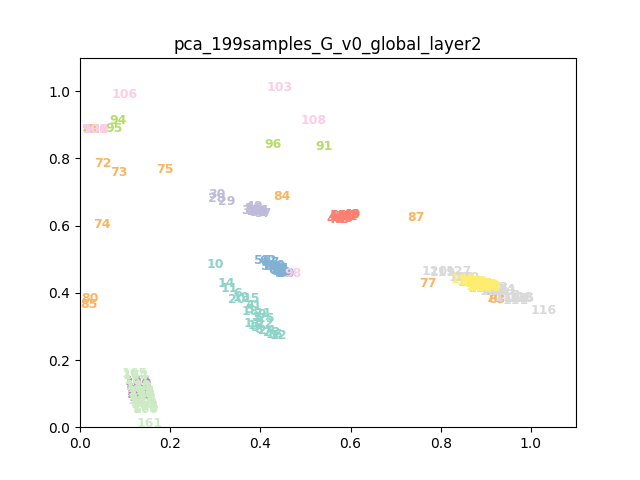


图2.6 vector 2的PCA可视化结果

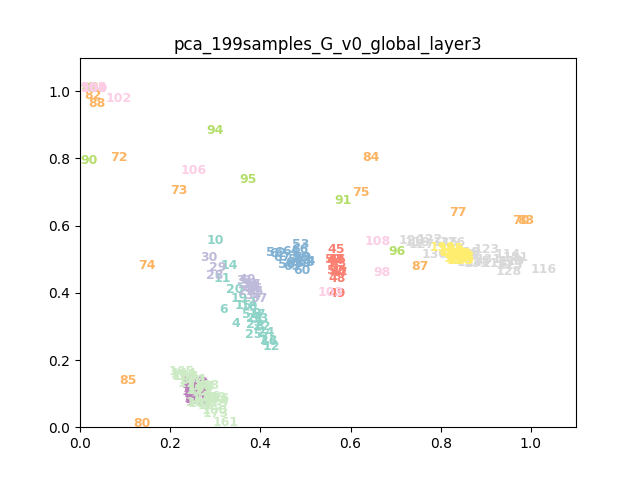


图2.7 vector 3的PCA可视化结果

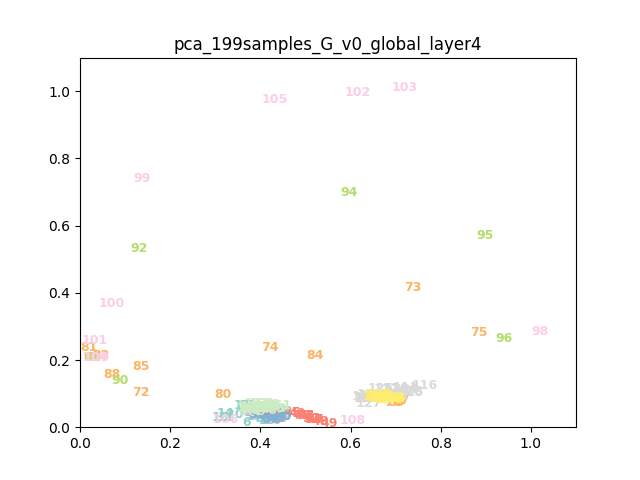


图2.8 vector 4的PCA可视化结果

观察vector 2的PCA可视化结果可以发现，没有改变眼睛的类别，类内的特征分布开始不同，说明实例标准化操作将草图上其他位置的变化传达到了眼睛位置，使眼睛位置的特征发生改变，而不单单只影响改动位置的特征。

对比vector 2和vector 3的PCA可视化结果，可以发现，没有改动眼睛的草图类别，如第1、2、3、4、8、9、10、11类，类内距离逐渐增大，即同一个类别内特征分布得更为分散。这说明随着逐层增加实例标准化操作，在输入草图上改动其他部位对眼睛位置的特征影响越来越明显，导致最终生成图像的眼睛也发生改变。

对比从vector 2到vector 4的PCA可视化结果，还可以发现，随着卷积层数的加深，第1、2、3、4、5、8、9、10、11类特征的类间距离逐渐减小，类间界限逐渐模糊，即不同类别的特征分布得更为集中。这说明随着逐层增加实例标准化操作，由于实例标准化的平均化作用，即使在输入草图上改变眼睛，对特征图上眼睛位置的影响作用也越来越弱。这种现象最终导致生成图像的纹理细节变得模糊，不够清晰。

接下来做的是模型的对比实验。所用的模型结构是pix2pixHD的，按前述同样的方法得到前5个卷积层的特征图，对应输入草图的左眼眼角位置坐标(170, 250)，分别在5张特征图上抽取一个点的全通道特征向量，维数分别为48、96、192、384和768，与前文对应，我们称之为vector0’ 到 vector4’。用PCA 对其进行降维后的可视化结果如下：

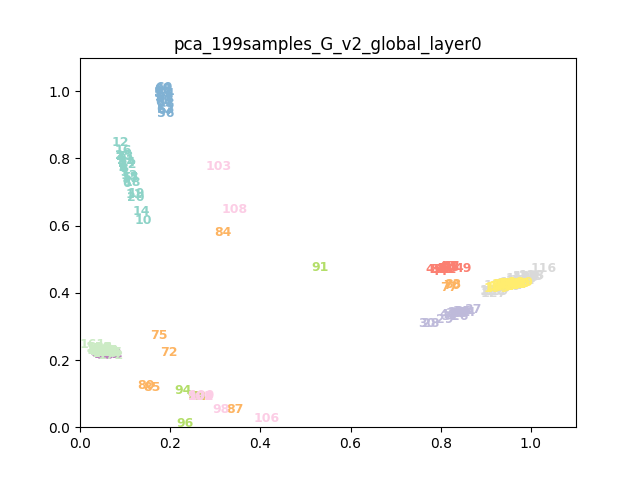


图2.9 vector0’的PCA可视化结果

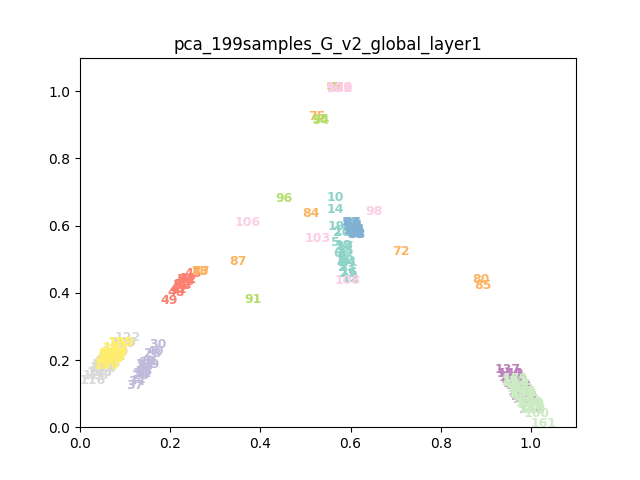


图2.10 vector1’的PCA可视化结果

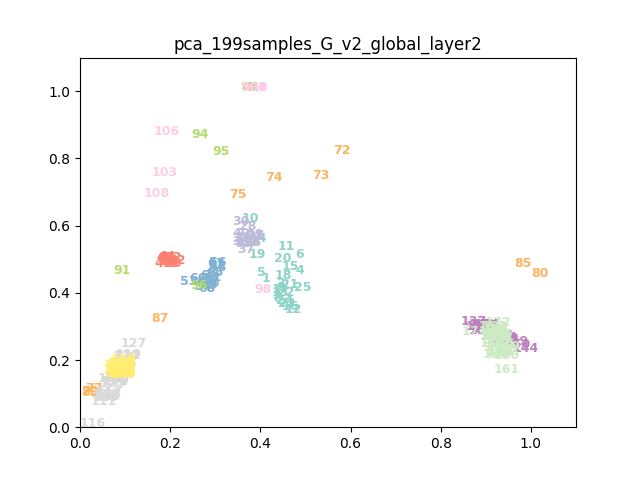


图2.11 vector2’的PCA可视化结果

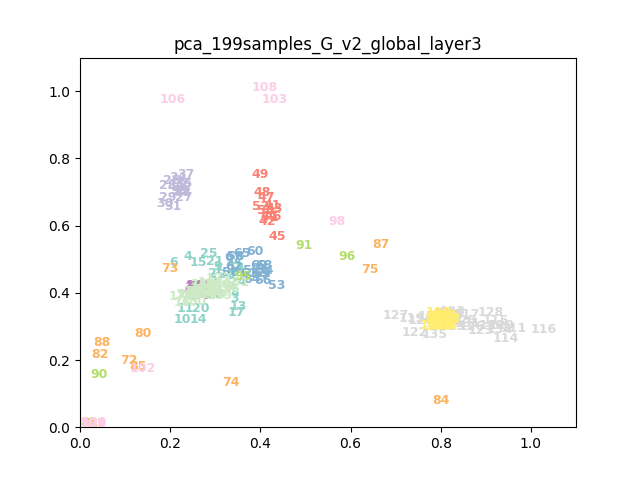


图2.12 vector3’的PCA可视化结果

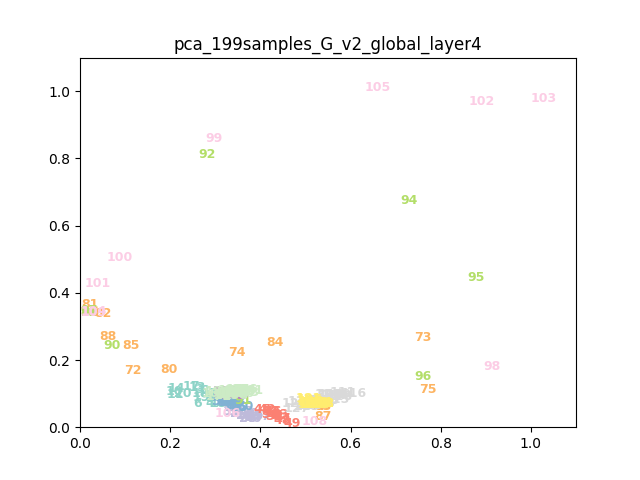


图2.13 vector4’的PCA可视化结果

将vector0’与vector1’的可视化结果分别与vector0和vector1的进行对比，因为前两个卷积层有实例标准化的缘故，所以第1、2、3、4、8、9、10、11类的特征各自分散地分布，而没有像vector0与vector1一样重合于一点。

分别将vector2，vector3，vector4与vector2’，vector3’，vector4’的可视化结果进行横向对比，发现vector2’，vector3’，vector4’中没有改变眼睛的类别，如第1、2、3、4、8、9、10、11类，其类内距离更大，类别内部分布得更为分散，而vector2，vector3，vector4则有效减少了类内距离，是类内分布得更为集中。这恰恰说明了我们将模型前两个卷积层的实例标准化操作去掉，可以有效降低在草图上改变某一部位而对生成图像其他部位产生的影响，也丰富了生成图像的纹理细节。

第三节 t-SNE可视化结果

从图2.14和2.15可以发现，第1、2、3、4、8、9、10、11类的vector0聚类在一起，各类集中地分布，而vector0’则在类内分布得比较分散。这也说明在第一个卷积层去掉实例标准化操作后，在草图上改变其他部位对特征图的的眼睛位置没有改变。

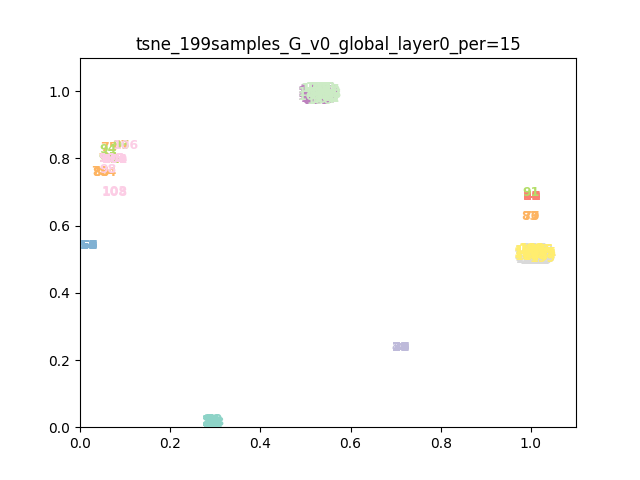


图2.14 vector0的t-SNE可视化结果

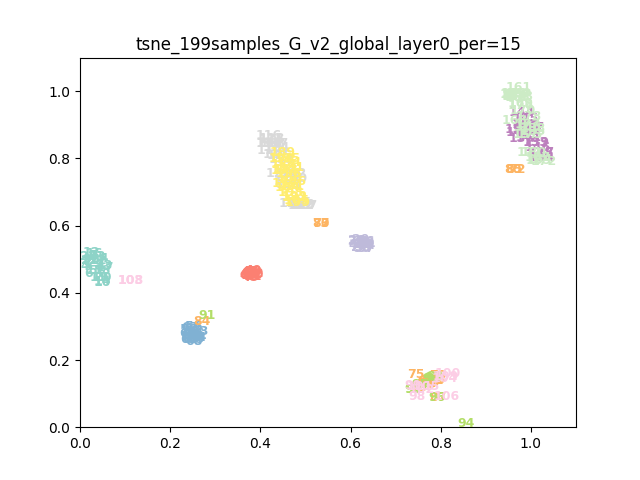


图2.15 vector0’ 的t-SNE可视化结果

由图2.16可以发现，第8类与第11类、第9类与第10类的vector4分布在一起，并没有互相分开。这说明，对眼睛部位的特征来说，在草图上修改眼睛依然是最重要的改变，在草图眼睛相同的情况下，无论改变嘴巴还是鼻子都不能对眼睛特征产生足以区分类别显著影响。

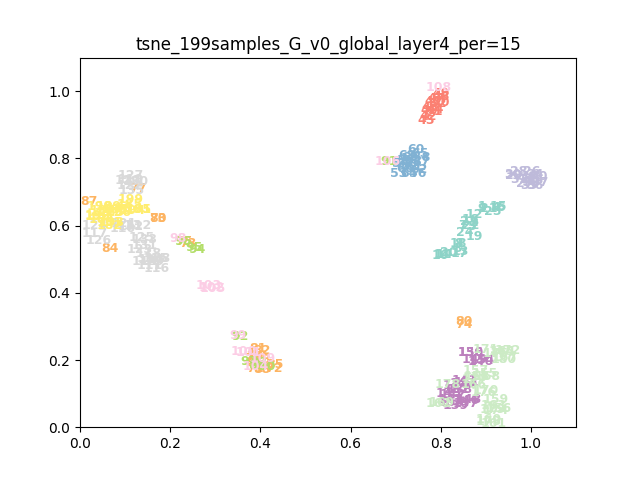


图2.16 vector4 的t-SNE可视化结果

第四节 t-SNE原理

高维数据的可视化在许多不同的领域一直是一个非常重要的问题，近几十年来提出了许多可视化的方法，数据降维是其中非常重要的一种。降维，顾名思义就是将高维空间中的数据降低维度。通常变成二维或三维的形式，方便以散点图的形式展现，帮助人们更好地分析数据。降维后的数据能否保留原始数据之间的关系以及降维后的可视化效果成为衡量一种降维方法的两大标准。

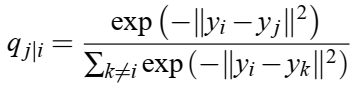
降维方法一般分为线性降维和非线性降维，线性降维包括主成分分析（PCA）、多维尺度变换（MDS）等，非线性降维包括等度量映射（Isomap）、局部线性嵌入（LLE）和随机近邻嵌入（SNE）等。线性降维通常会将高维空间中不同的数据点在低维表示中远远分开，对于高维空间中相同或相近的数据点其表示能力则存在不足。非线性降维可以很好地弥补这一点。t-SNE便是一种非线性降维方法，它在SNE的基础上做了几点改进。

一、SNE

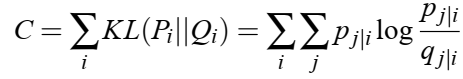
随机近邻嵌入首先将高维空间数据点之间的欧氏距离转换为表示相似性的条件概率分布。考虑高维空间中两个数据点和，以表示选择作为其紧邻点的条件概率。与的欧氏距离越小，则越大。如果用高斯分布描述这种条件概率，则其数学表达形式为：



其中表示以为中心的高斯分布的方差。设和映射到低维空间的点分别为和，同样地，用条件概率表示低维空间点的相似性，其数学表达形式为:

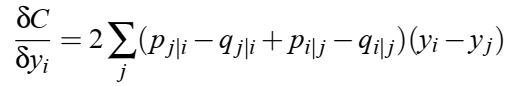


为了使和可以真实地反映和之间的关系，理论上应该让条件概率与完全相等。用表示给定与其余所有点之间的条件概率分布，表示与其余个点之间的条件概率分布，应该使与完全相等，所以用与的KL距离作为代价函数：

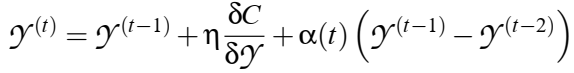


由于KL散度是非对称的，因此在低维空间中，不同类型的误差在两两距离上的权重并不相等。特别地，如果高维空间中相距很近的点被映射到低维空间后距离很远，将得到很大的代价；而如果在高维空间中相距很远的点被映射成距离很近的点，得到的代价不会很大。这说明，SNE的一大缺陷就是更关注局部结构而忽视了全局。

代价函数对求梯度后有一个简单的形式：



利用带动量的随机梯度下降算法优化低维空间点的分布：



二、t-SNE

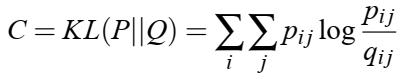
t-SNE在SNE的基础上做了两点改进：

1. 将原始的SNE转变成对称SNE

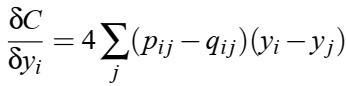
在原来的SNE中，，且，是非对称的，则在高维空间和低维空间中分别构造出对称的联合概率分布和，使得对任意的，。构造的有如下形式：



其中，n为数据点的数量，这种定义既满足了对称性，也保证了对于离群点代价惩罚不致太小。代价函数重写为：



梯度函数变为：

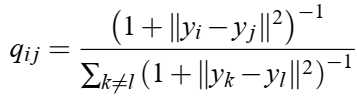


可见比之前的形式更为简洁，计算效率更高。

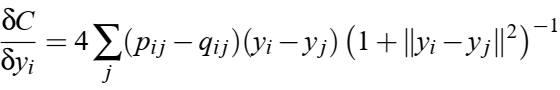
2. 将低维空间中的数据表示为t分布

t分布是长尾分布，相比于正态分布可以更好地将高维数据映射到低维空间。举例来说，在高维空间中距离相近的点，为了满足，映射到低维空间后距离会更近；而在高维空间中距离较远的点，映射到低维空间后距离会拉大。这使得同类别的点或者说相似的点在低维空间分布得更为紧密，而不同类别的点或者说相似性较低的点分布得更为稀疏，从而使可视化的效果更好。

用t分布重新定义低维空间的联合概率分布：



代价函数与之前相同，梯度的形式变为：



之后用带动量的随机梯度下降算法低维空间数据表示Y即可。t-SNE的算法框图如下：

