计算机视觉系统期终汇报

第12组

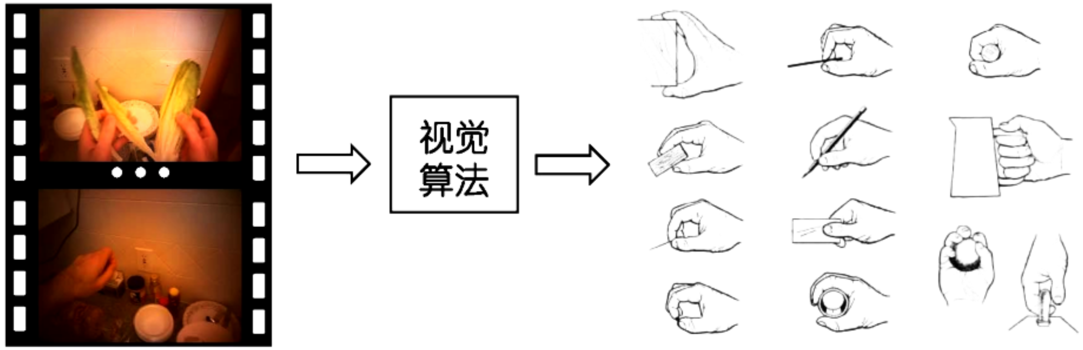
E202008000106邹雪丹
201708010207 杨佳政
201708010214 唐协成
201508010628 李鑫

目录

CONTENTS

1.问题定义

编程任务-需求



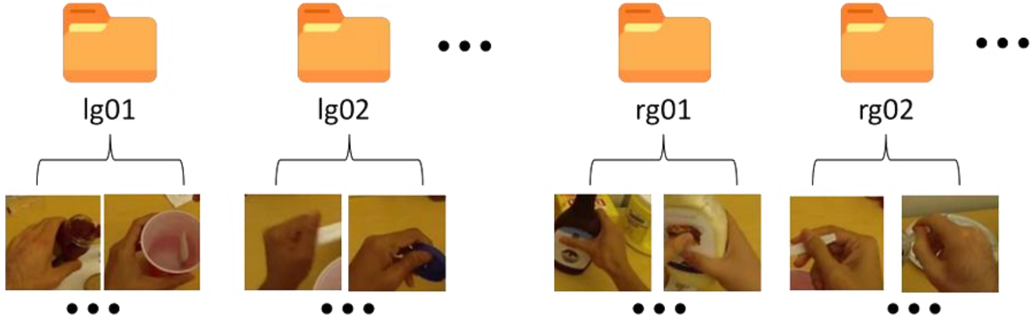
手部动作图像集　　　　抓取手势类型集

需求描述：
从大量的视频数据中自动发掘常见的抓取手势类型集，在机器人学和医疗康复等领域有重要的应用价值。此次编程任务希望通过设计**计算机视觉算法**，对包括手部动作的图像集合进行处理和分析，从中**自动发现不同的抓取手势类型**。

编程任务-具体要求

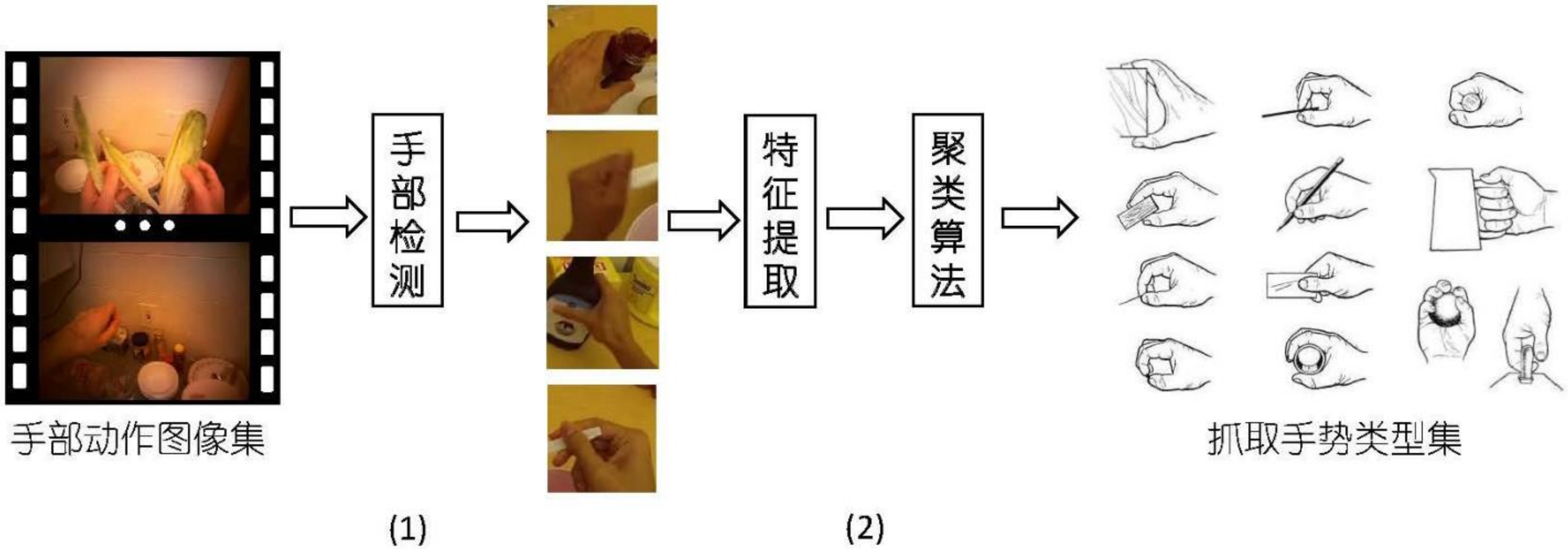
输入：包含手部动作的图像集

输出：对应不同抓取手势的手部图像集合(建议10-15个类型)



注意：输出图像保留原始图像的文件名

1.问题定义



手部识别

↓

图像分割

↓

特征提取

↓

图像聚类

我们认为这个问题可以分解为识别手部并分割对应图片，对分割到的手部进行特征提取，最终聚类手部这几个子问题。对于每一个子问题的解决应该是相对独立的，我们可以单独地提升其中的每个步骤的效果。

2. 手部识别和分割

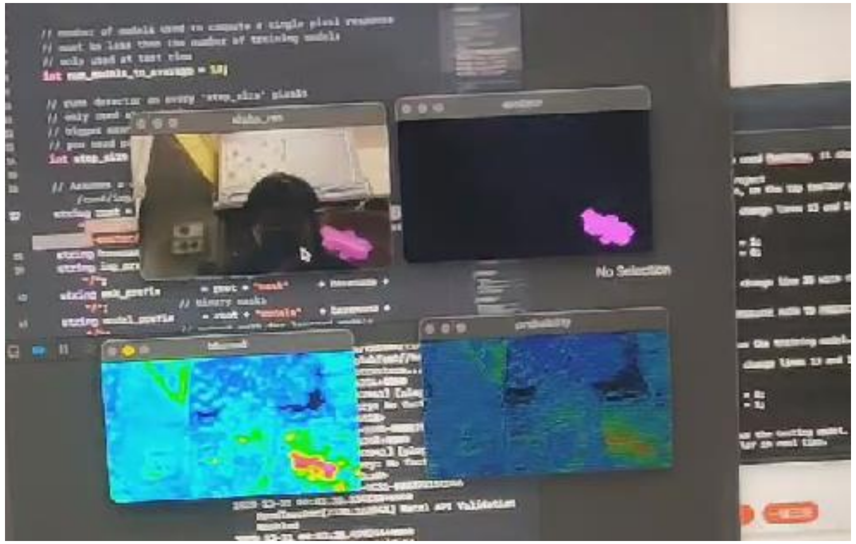
传统的肤色检测方法显然不能很好地工作于我们复杂的数据集上。

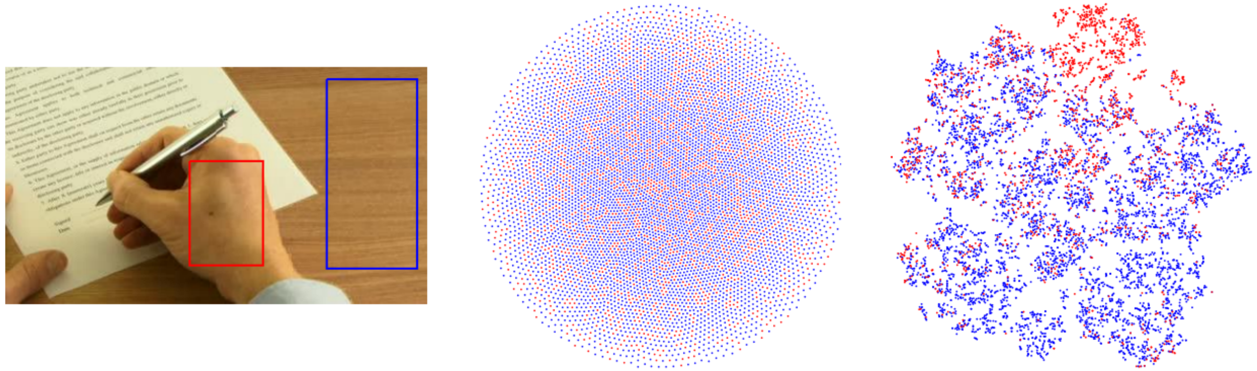
2. 手部识别和分割

Method A

我们尝试了Li, Cheng等人在2013年基于第一人称视频手部像素研究肤色和背景颜色区别并以此提出的一个利用了推荐系统方法的基于局部手部特征和全局特征相结合的分类器手部识别分割方法。

由于他们原始的代码是基于视频输入的，我们这里将源代码稍作修改并重新规定了输出图片的尺寸来提取我们数据集——第一人称视频中的手部图像。





(a) Image regions (b) Color features (c) Color + texture features

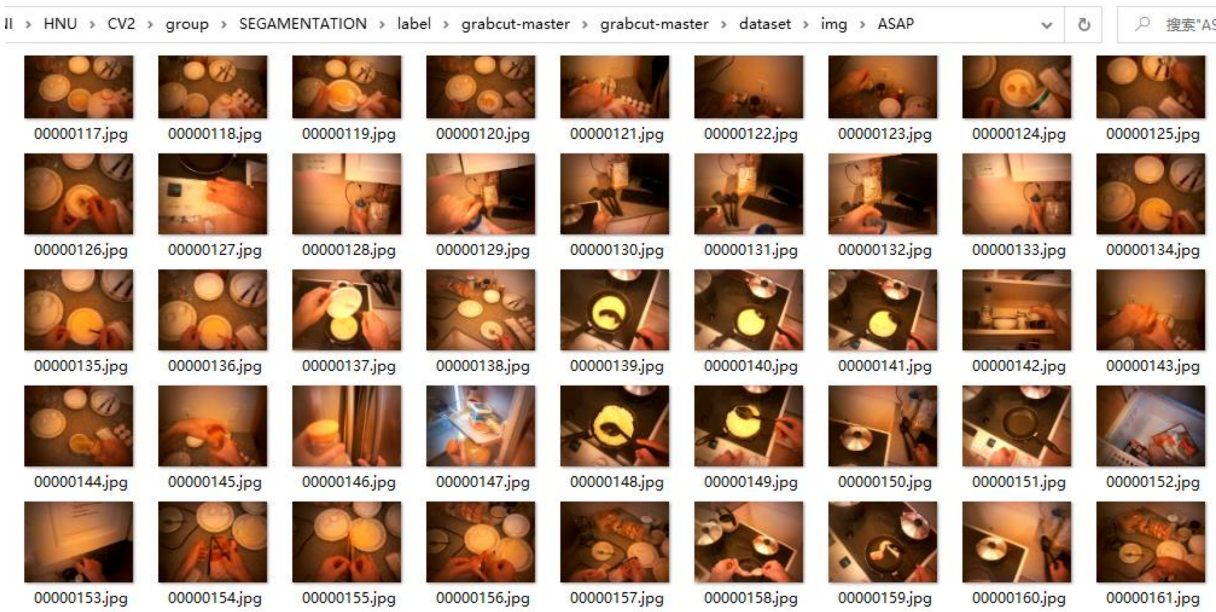
Figure 2. Visualization of feature spaces with t-SNE [27]. Skin features in **red** and the desk features in **blue**. Texture features allows for better separation.

Li, Cheng, and Kris M. Kitani. "Pixel-level hand detection in ego-centric videos." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

Method A

由于他们的模型需要预先标记出手部的训练图片来训练分类器，我们使用了他们提供的基于grabcut方法编写的打标工具来给我们的数据集打标。 大部分情况下，在3次迭代以内，通过标记出前景区和背景区可以较好地标记出图片中手部区域。

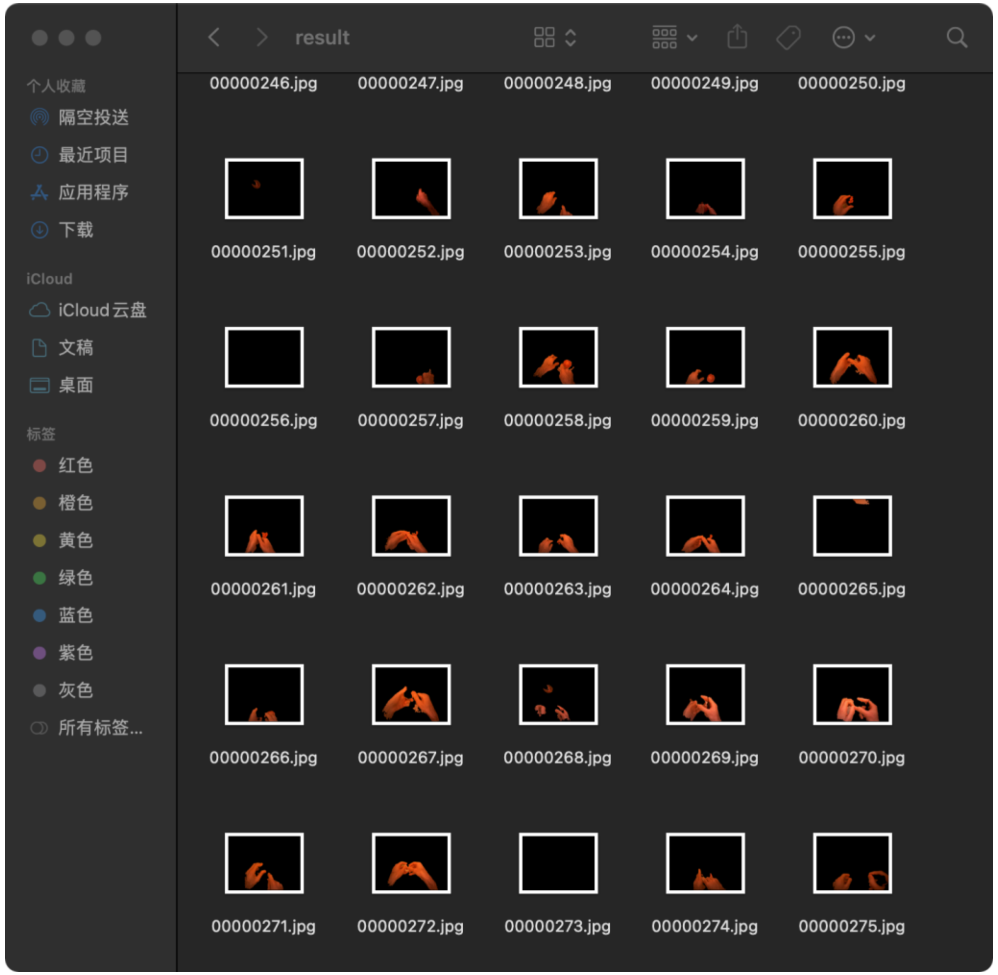
Method A

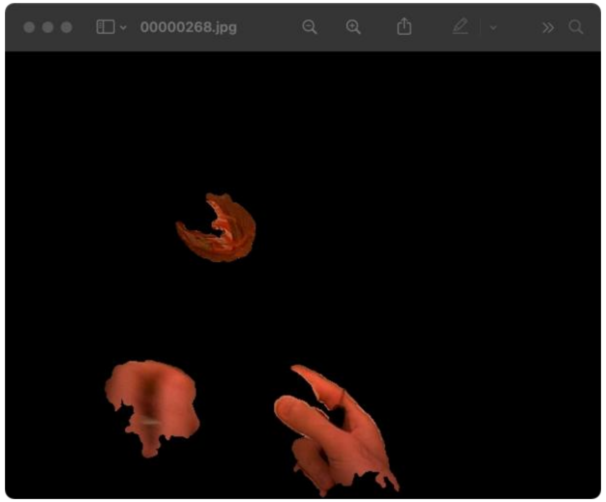


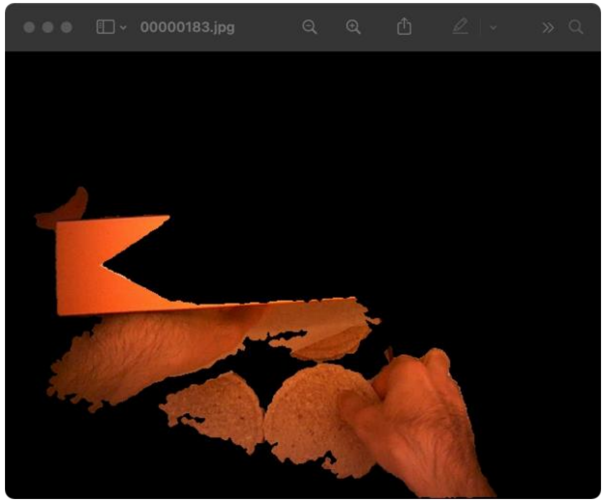
我们在OSX下，搭建OpenCV2环境测试了这个方法。

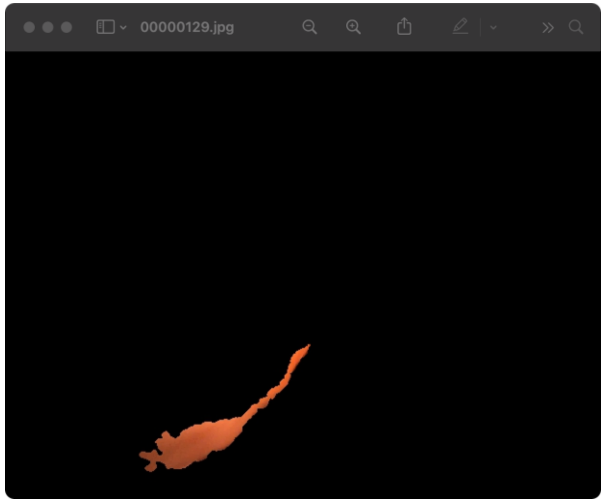
考虑到我们的数据集有大概1500余张，我们按照10:1法则选取了其中的150张按照上述的打标工具打上了手部标签，它们作为训练集来训练这个分类器进行手部分类。

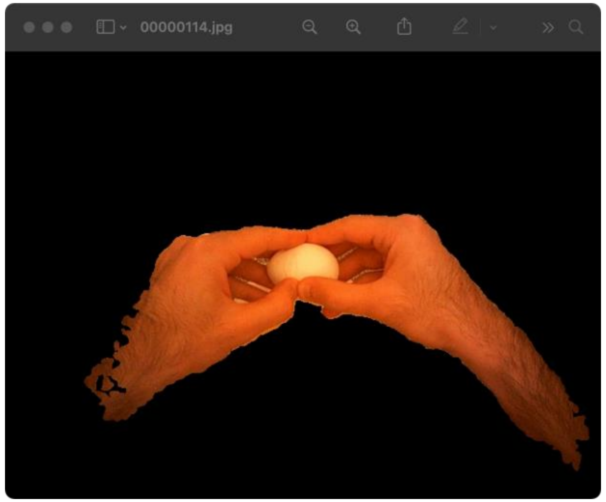
Method A











使用这种方法得到的分割效果只能说勉强，有一些分割结果完全错误。当有两只手时也不容易很好地分割出来。

2. 手部识别和分割

Method B

另一种我们尝试的方法是Cai, Minjie等人于2020年提出的一种贝叶斯CNN神经网络方法。他们的方法只需要输入无标签的第一人称视频图像就能够很好地提取其中的手部图像。

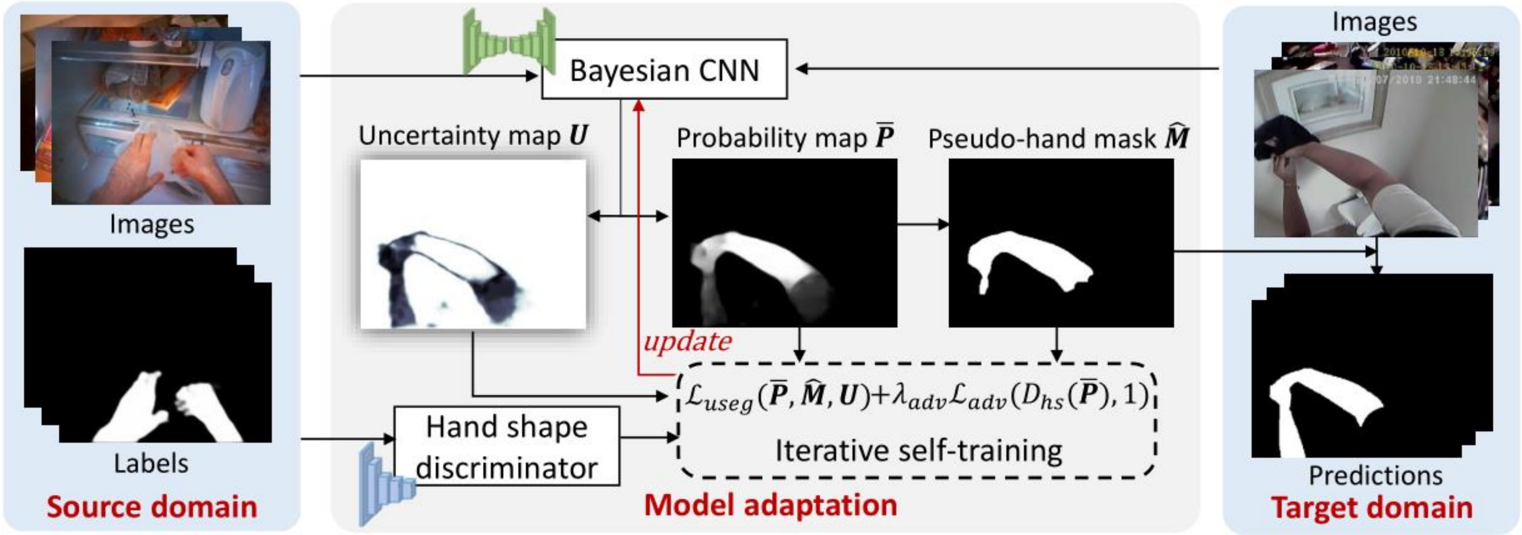


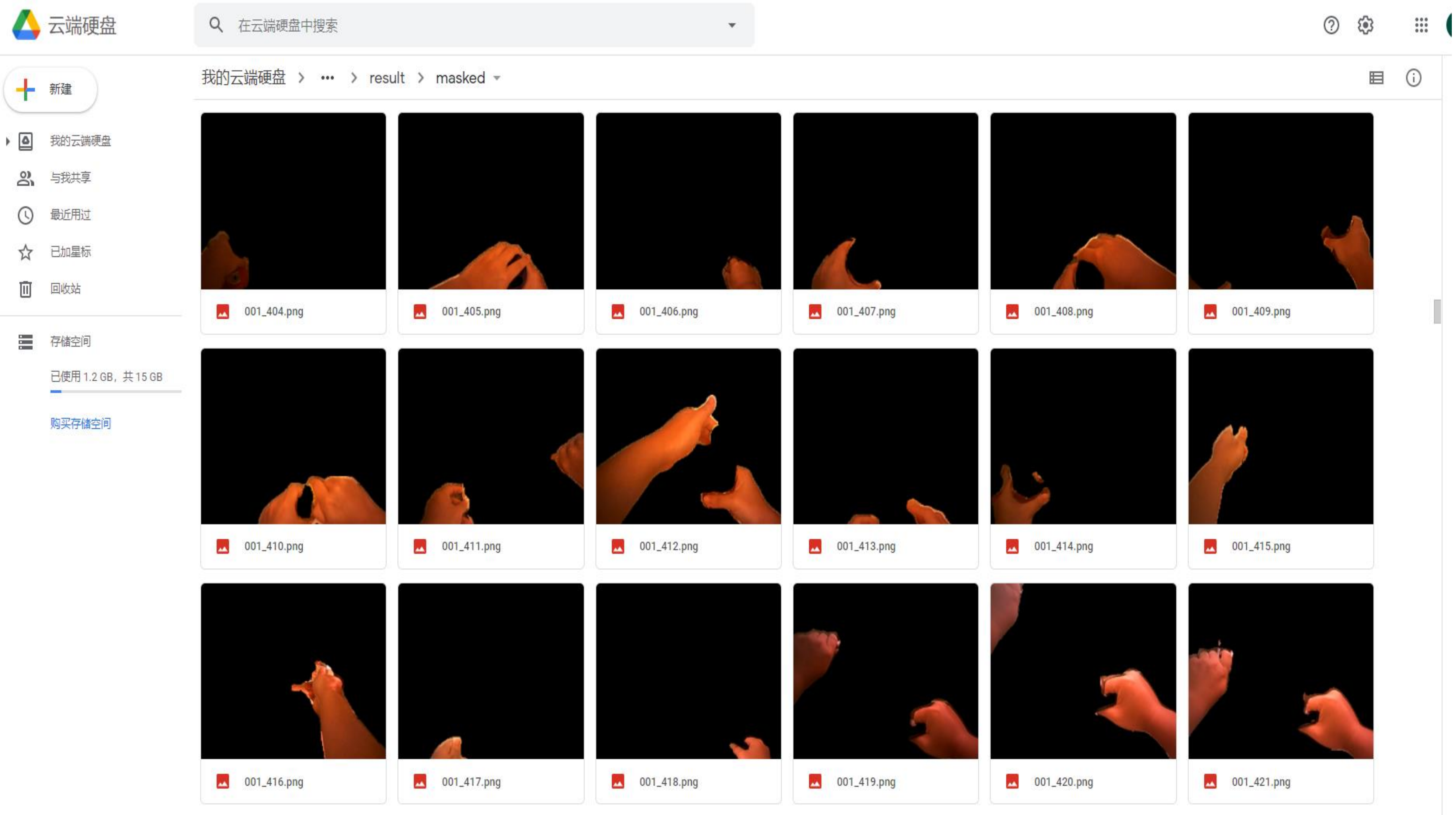
Figure 3. Overview of the proposed uncertainty-guided model adaptation.

Cai, Minjie, Feng Lu, and Yoichi Sato. "Generalizing Hand Segmentation in Egocentric Videos With Uncertainty-Guided Model Adaptation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

.

Method B

我们在Google Colab上，
修改了源代码中的训练部分，
注意到其预训练模型正是基于EGTEA Gaze+数据集，我们直接使用了预训练模型，修改后得到了测试代码来讲我们的数据集丢入网络得到分割结果。
可以看到，这个方法的识别风格结果比Li,Cheng等人的方法结果好上不少。



Method B

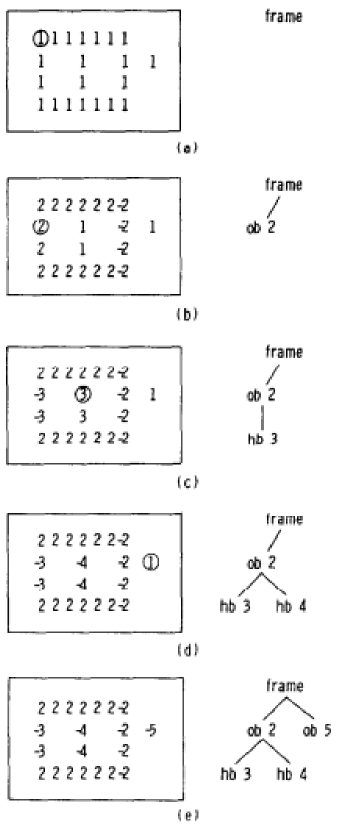
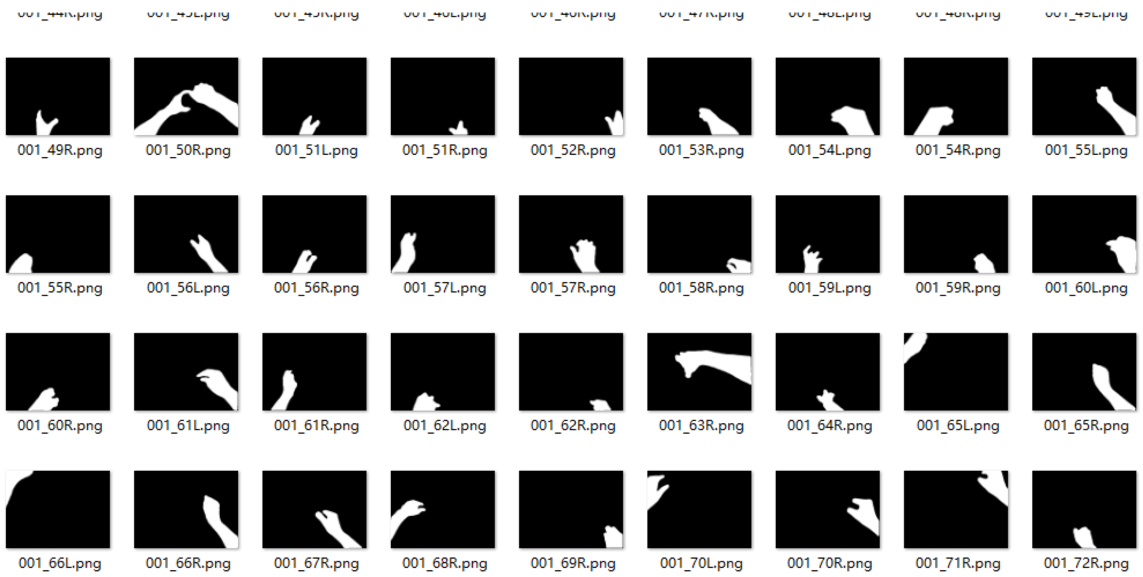


FIG. 3. An illustration of the process of Algorithm 1. The left-hand figures show the pixel values and the right-hand the extracted structures among borders (ob: outer border, hb: hole border). The circled pixels are the starting points of border following.

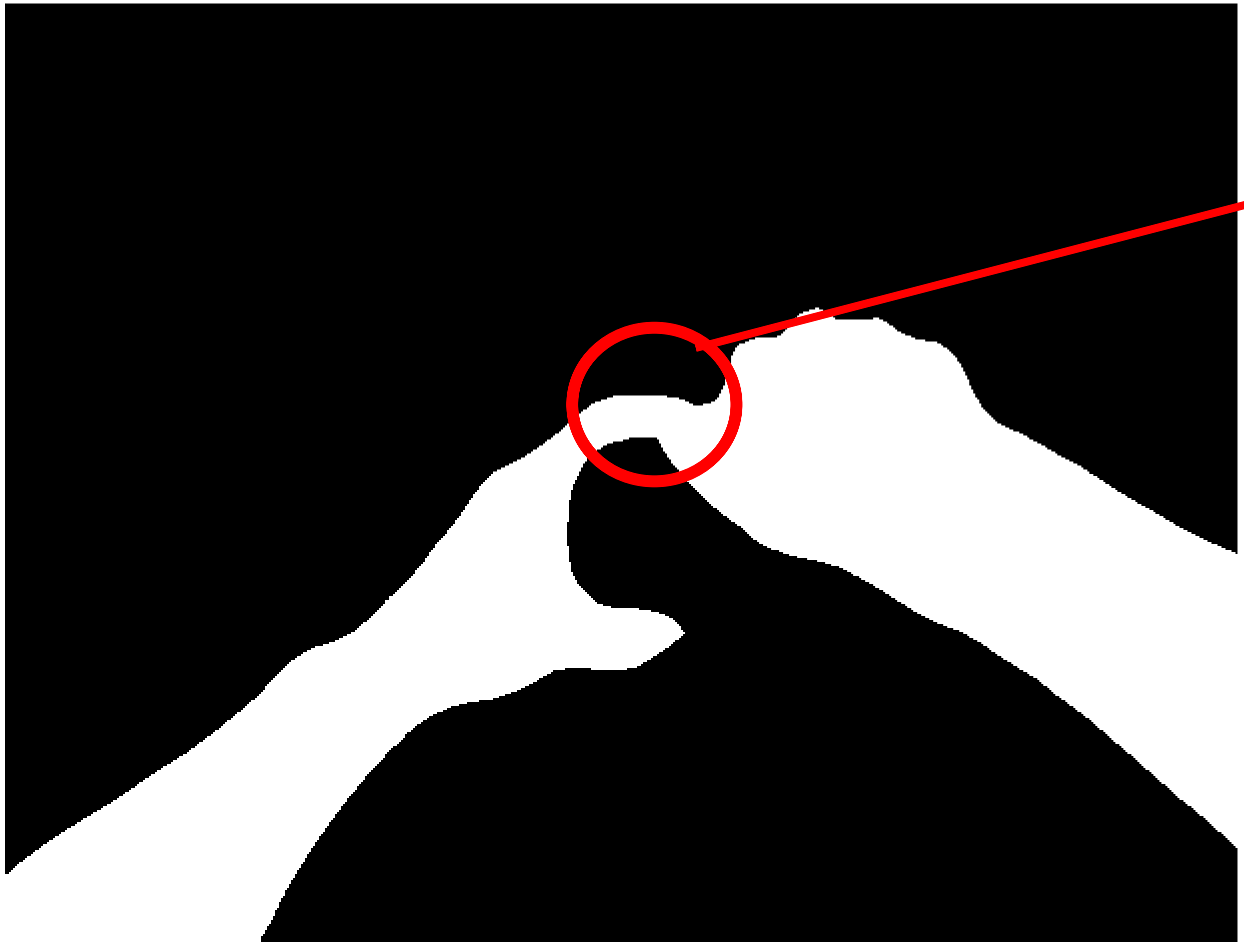


接下来我们需要对分割的结果进行必要的处理，以尽可能地**去除掉错误划分出来的像素块**，并**分离左右手**。

我们使用opencv自带的基于Satoshi等人提出的方法实现的**findContours()函数**来找到所有的轮廓，并且设置面积阈值以去除掉太小的轮廓（我们认为是错误的像素块），保留剩下的最多2个最大的轮廓，得到所有手部单独的划分。

Satoshi Suzuki and others. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1):32–46, 1985.

Method B

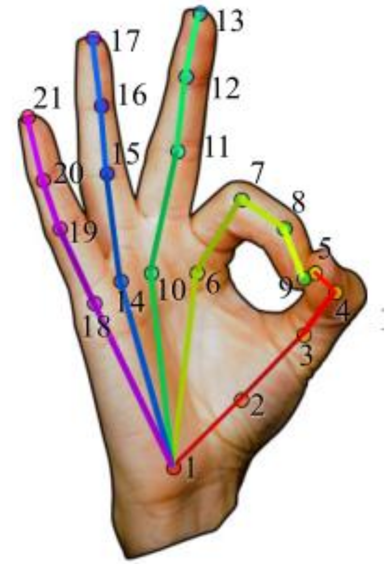


这种方法的一个问题是当两只手贴得太近的时候只能得到一个轮廓。可以考虑使用腐蚀膨胀技术对原图做处理让左右手的轮廓大致分开后再提取。

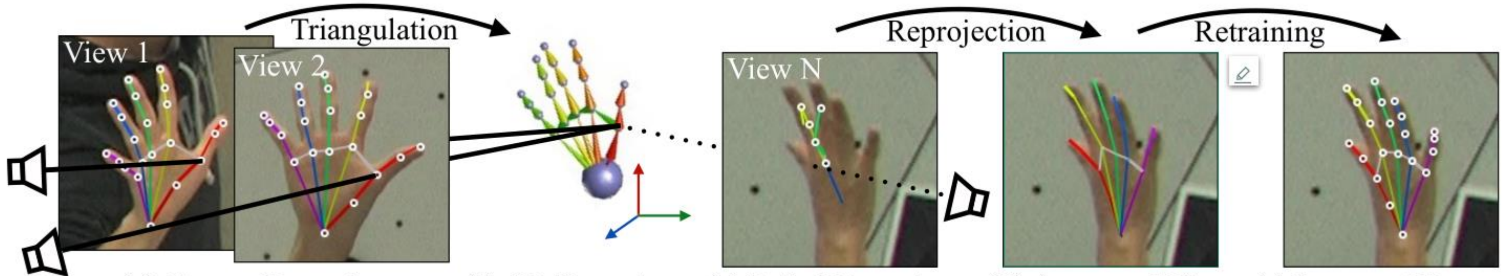
但是考虑到有的情况连通区域太大，对所有的图像都大做腐蚀操作可能导致部分图像的手指部分数据严重丢失，我们最后放弃了处理这种潜在的改进方法。

3.手部聚类和分类

Method 1



(a) Detections



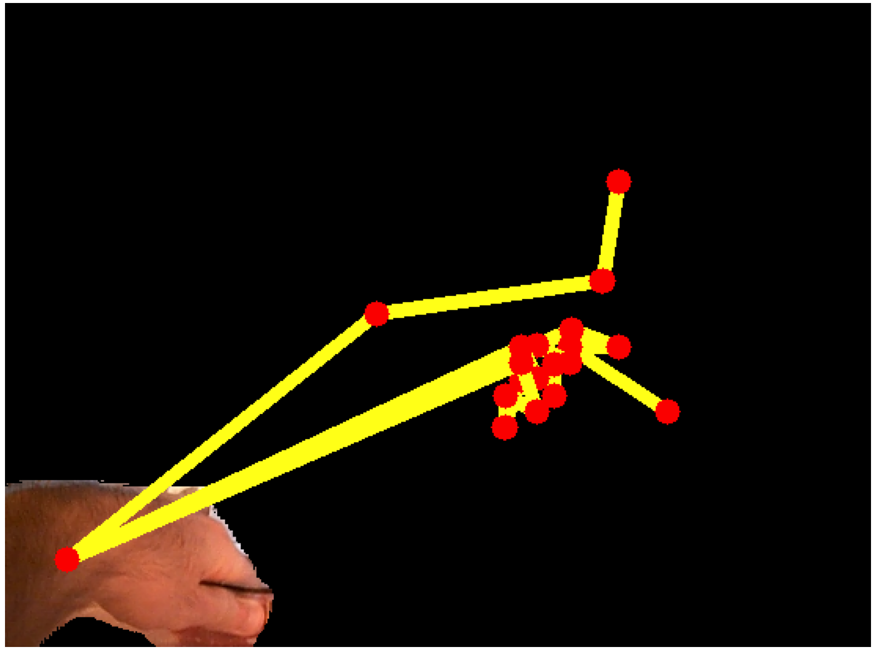
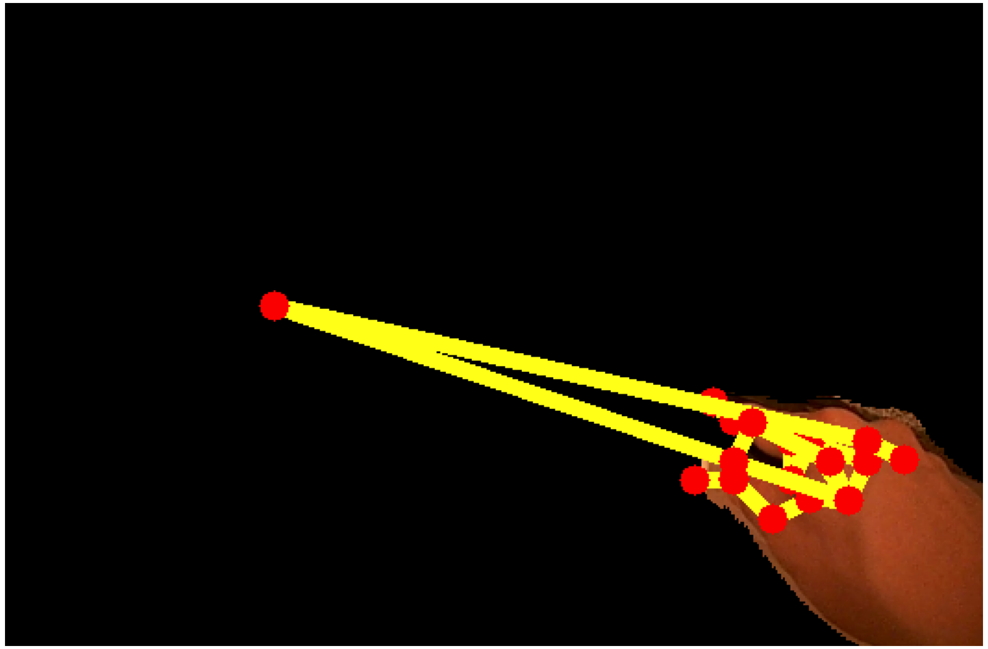
(a) Correct Detections (b) 3D Keypoints (c) Failed Detection (d) Annotated View (e) Improved Detector

Figure 3: Multiview Bootstrapping. (a) A multiview system provides views of the hand where keypoint detection is easy, which are used to triangulate (b) the 3D position of the keypoints. Difficult views with (c) failed detections can be (d) annotated using the reprojected 3D keypoints, and used to retrain (e) an improved detector that now works on difficult views.

近年T. Simon等人通过标记少量多视角下的手部图像为数据集，训练CPM网络，使得可以对单张二维手部图像标记出对应的关键点。我们采用了这个方法来提取我们得到二维图像的手部关键点。

T. Simon, H. Joo, I. Matthews and Y. Sheikh, "Hand Keypoint Detection in Single Images Using Multiview Bootstrapping," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

Method 1

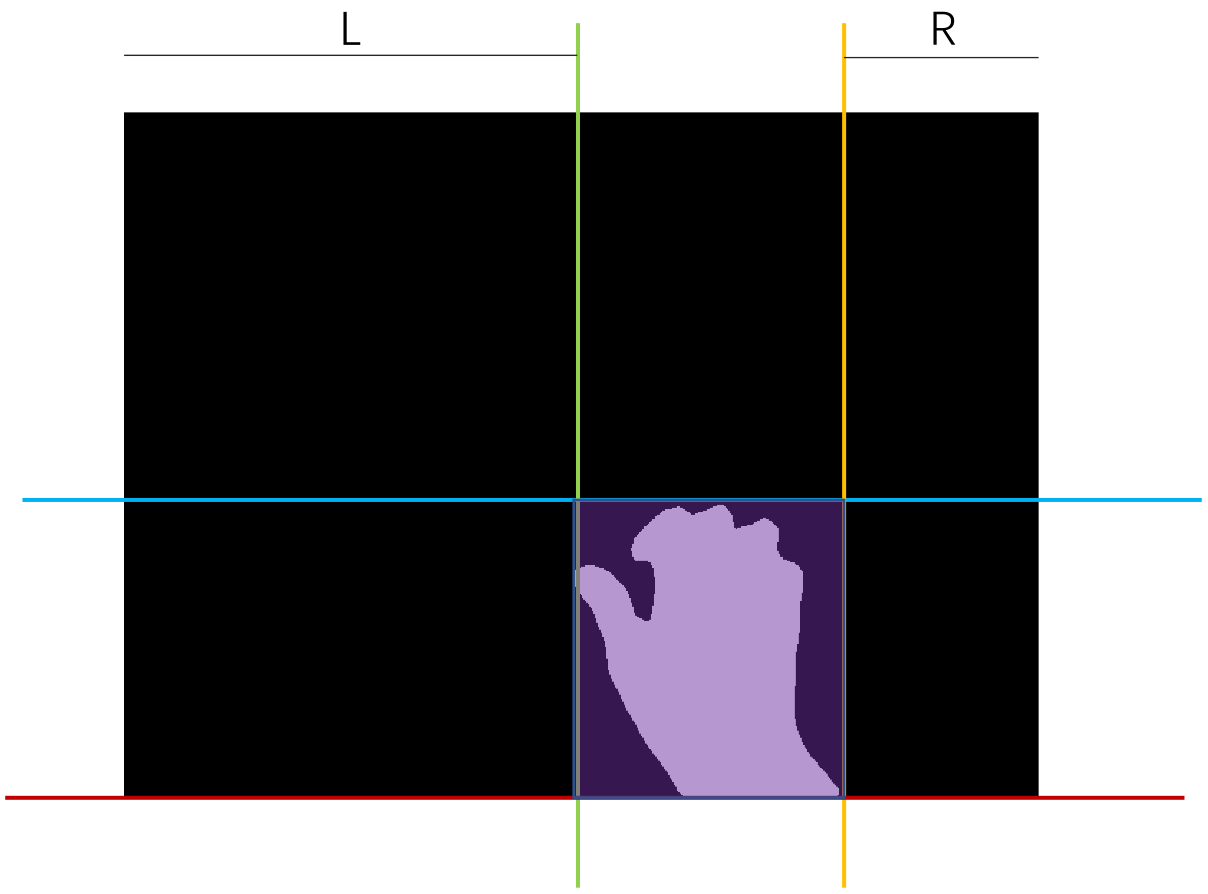
不过我们很快便发现，这种方法对没有单独提取手部区域，留有大量黑色区域的图像效果是很差的。
为了提升方法的正确率，我们有必要对分割结果的手部区域进行提取。

Method 1

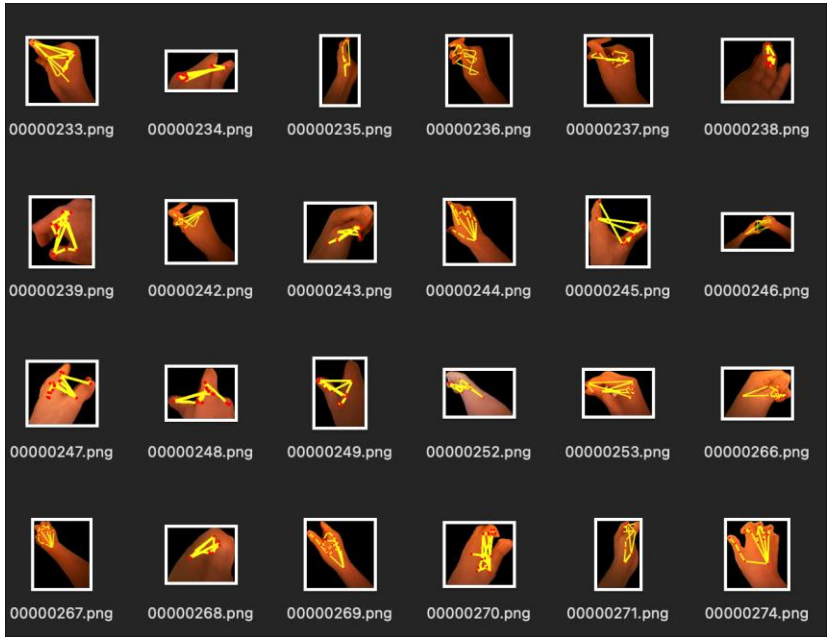
提取最小手部区域的矩形很简单。只需要遍历整张图，找到非黑色的mask区域的最左坐标，最右坐标，最上坐标和最下坐标。

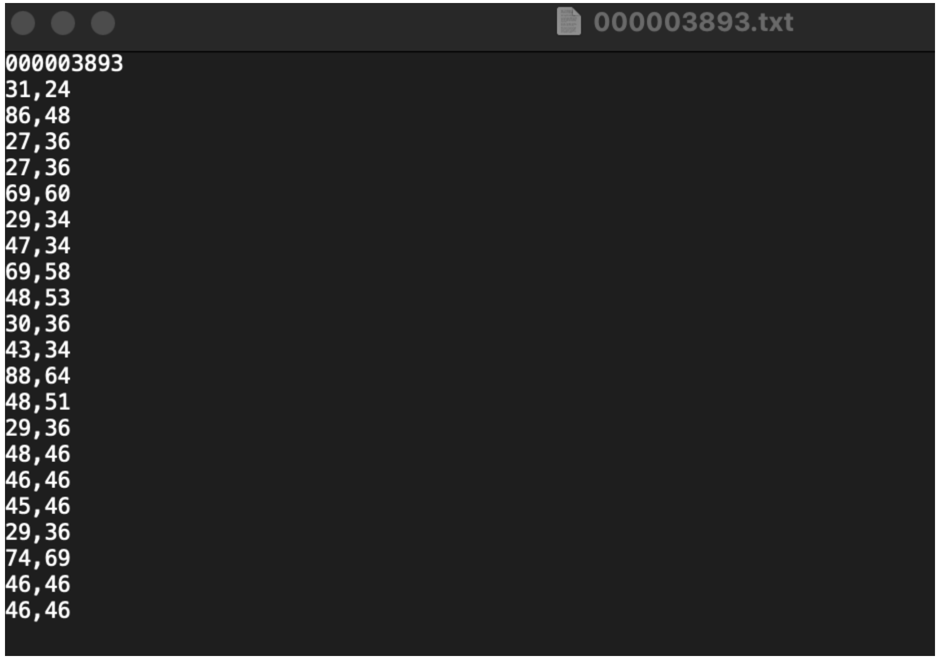
由最左上，最右下坐标组合得到的rect矩形区域就是最小手部矩形区域。

类似地，注意到左手大部分在图像左侧，右手的大部分在图像右侧，用rect矩形的左边缘和右边缘坐标还可以大致判断是左手还是右手。当然实际上，由于数据集的复杂性，这种简单的特征分辨左右手正确率不算特别高。

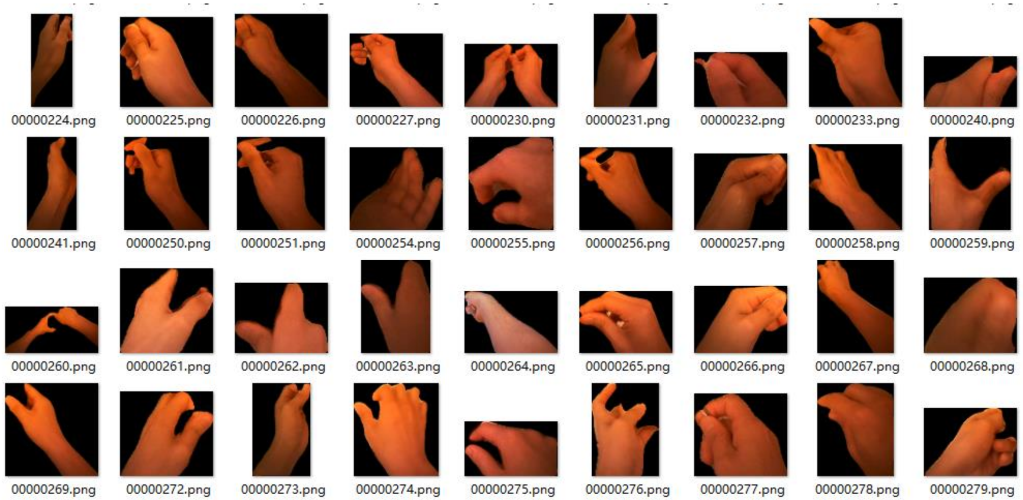


Method 1




000003893.txt

000003893
31,24
86,48
27,36
27,36
69,60
29,34
47,34
69,58
48,53
30,36
43,34
88,64
48,51
29,36
48,46
46,46
45,46
29,36
74,69
46,46
46,46



之后我们再用T. Simon等人的方法作用于我们提取后的手部局部图片上，成功得到每张图的手部特征点坐标，为了方便后续聚类操作不再花费额外的时间重新提取每张图的特征点，我们让每张图生成的手部特征点坐标同时保存在对应的txt文本上，并记录对应的图片编号。

Method 1

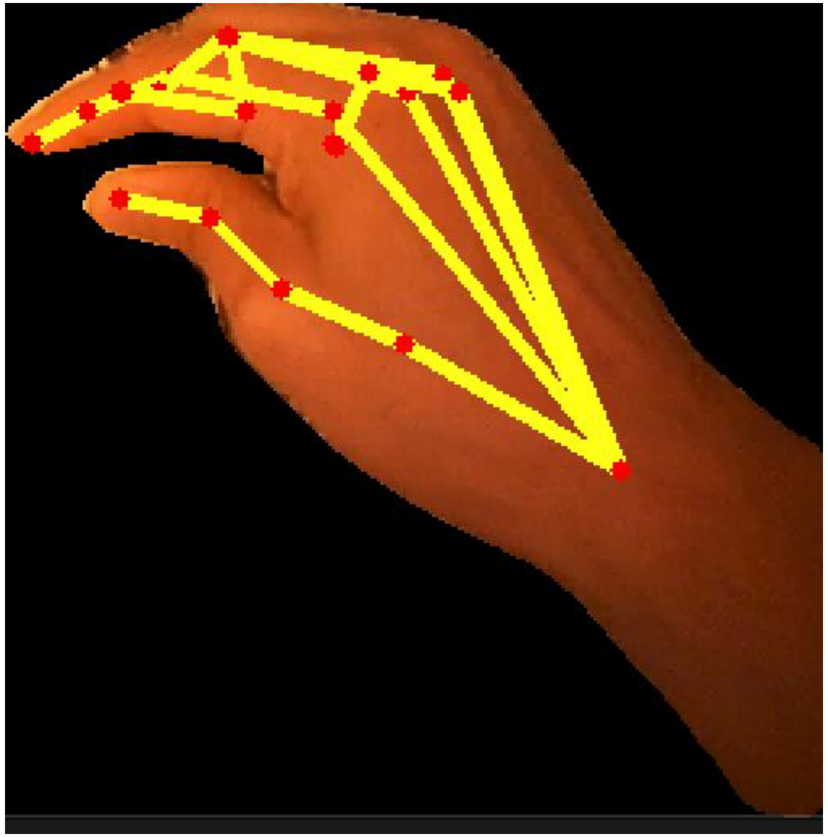
如何对得到的手部特征点聚类?

考虑使用K-means聚类算法，对得到的关键点坐标进行聚类。

可将得到的21个关键点的坐标**展开成42维的向量**，那么就是**对近3000余个42维的向量做K-means聚类**（1500余张任务图片做单手分割后得到的图片总量将近3000张）。

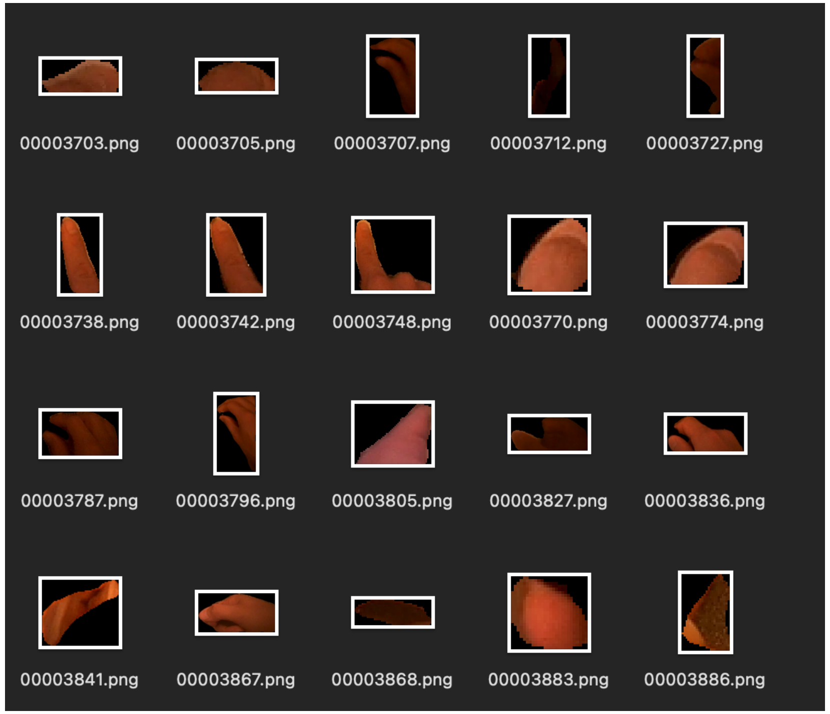
或者**直接将所有点的横纵坐标求均值**，得到中心坐标，变成**对3000余个2维向量做K-means聚类**。

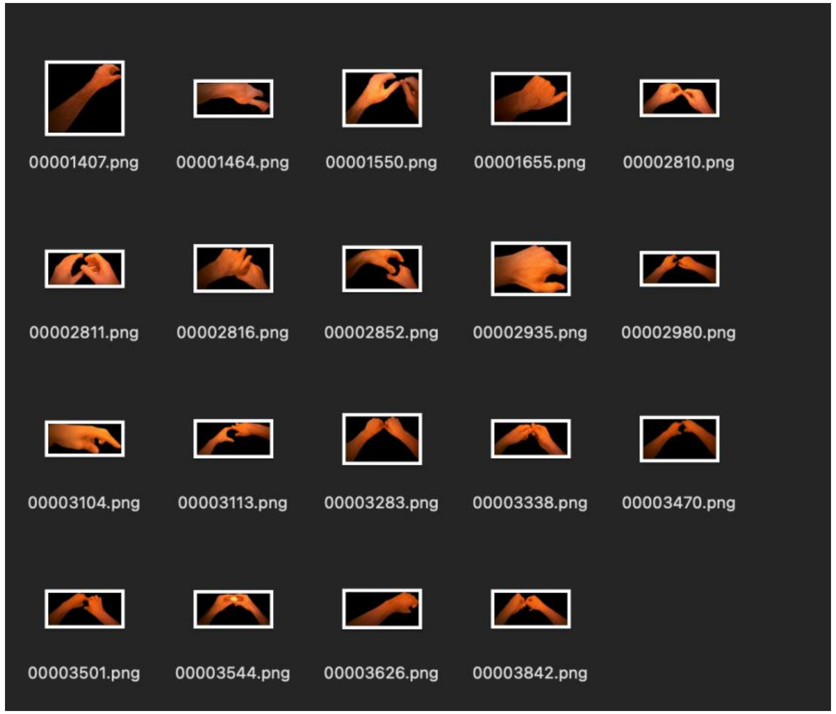
由于我们的算力有限，我们简单地采用了后者方法。

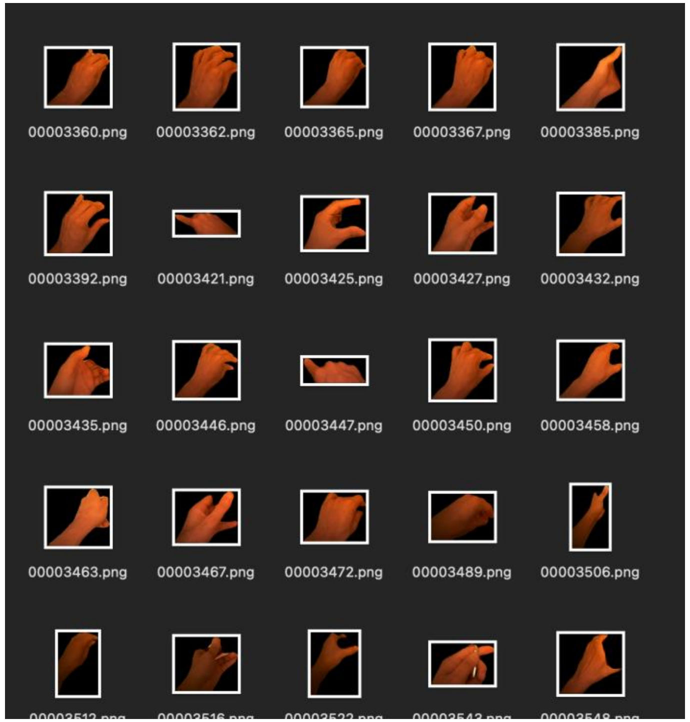


几何中心: $\left(\frac{\Sigma x_i}{n}, \frac{\Sigma y_i}{n}\right)$

Method 1

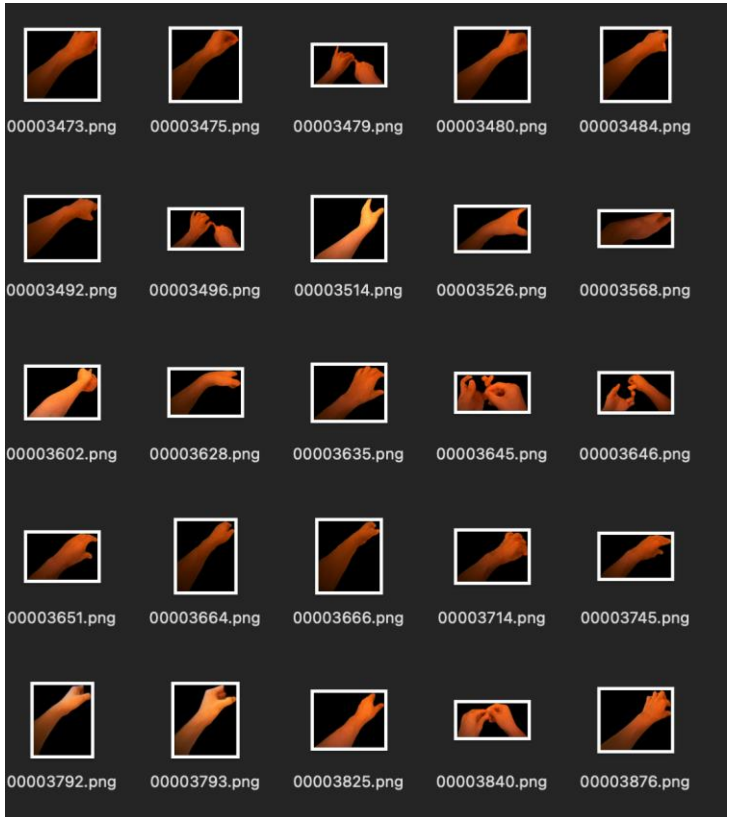


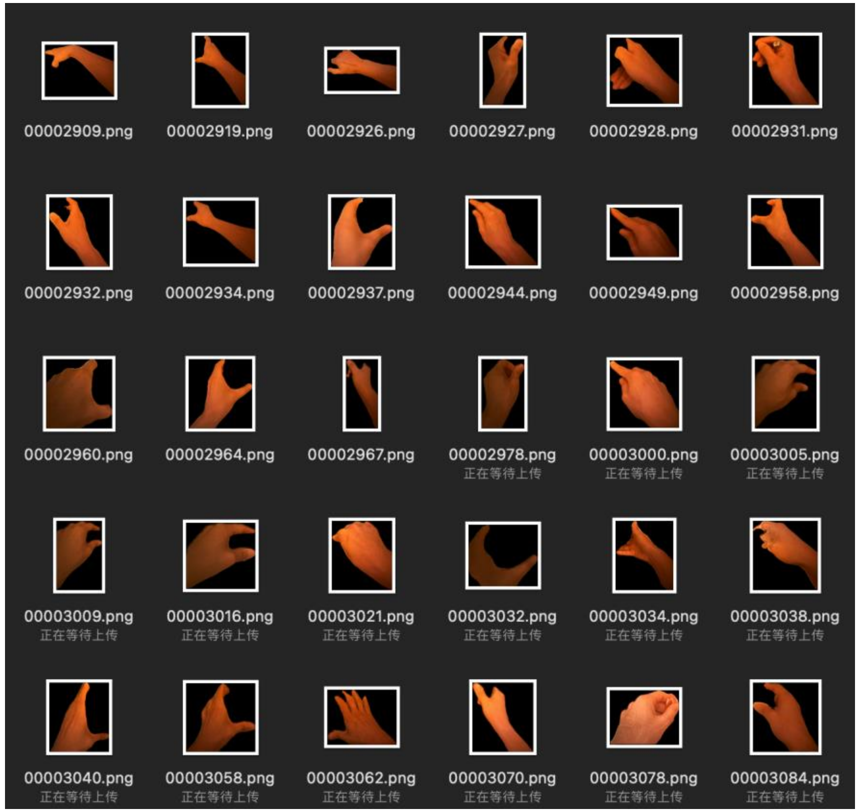


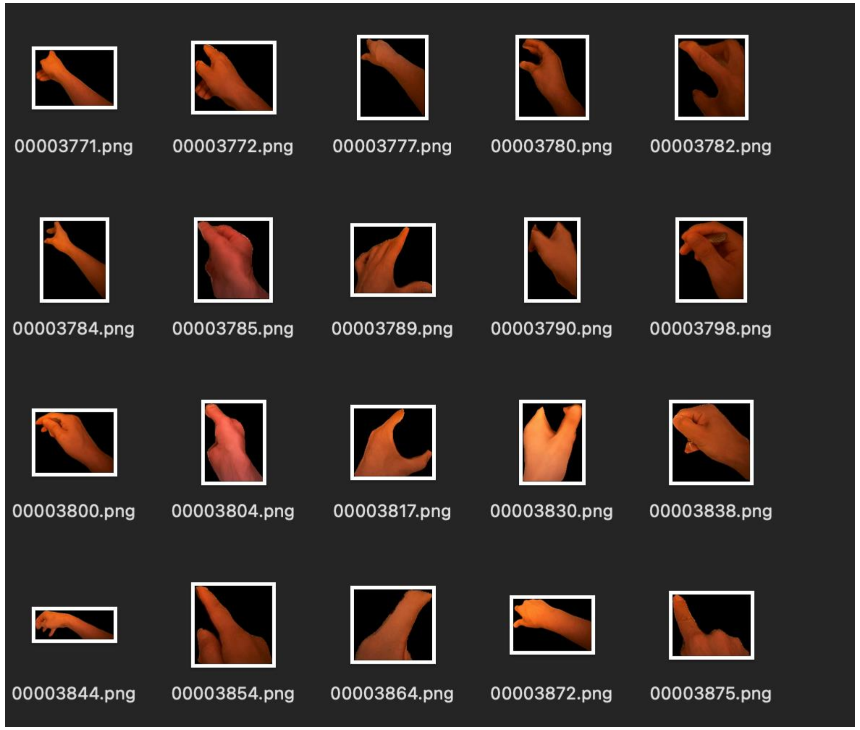


我们设置了10个类别来对所有的中心坐标点做聚类。由于K-means的聚类中心是随机生成的，很多时候会出现多个类别最终没有任何图片的情况，我们反复运行了K-means方法多次得到了一个各类别都有分类结果的较好结果。

Method 1



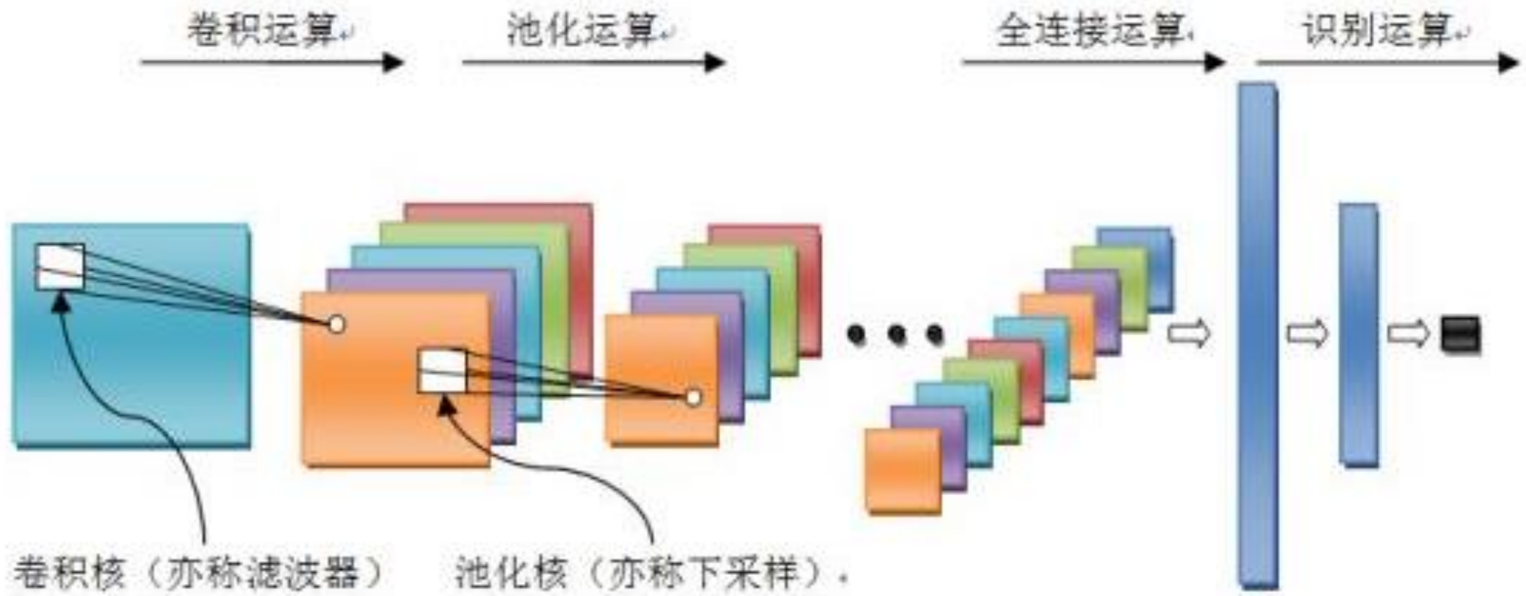




可以看到，尽管有一定的错误率，这个聚类方法还是可以对大致类型的图片起到一定的划分作用。
由于我们简单粗暴地用关键点的几何中心坐标来做聚类，损失了每个关键点的大部分信息，聚类结果必然不算好。同时，由于是聚类算法，我们无法真正做到对手部姿势的分类。

3.手部聚类和分类

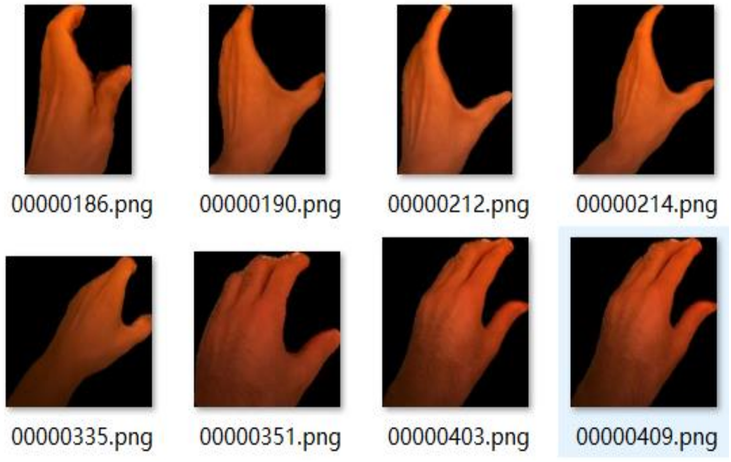
Method 2



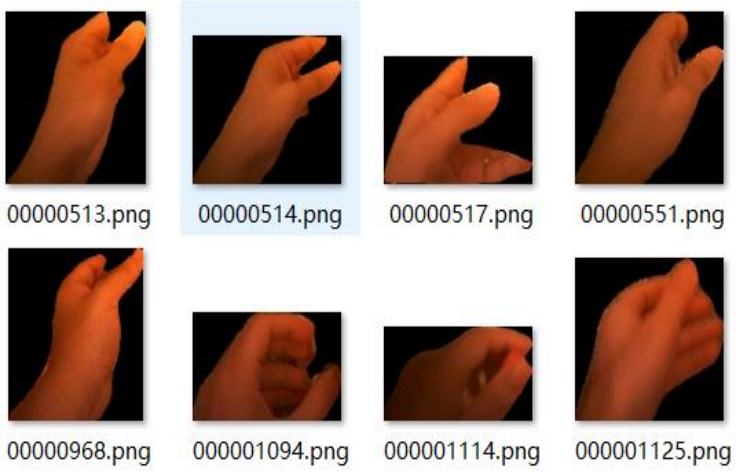
具有自然空间顺序的图像分类非常适合于CNN，卷积神经网络长期以来是图像识别领域的核心算法之一，并在学习数据充足时有稳定的表现。

数据集中的手部动作有较强的一致性特征，考虑采用手动打标签构造训练集+训练CNN模型+整体数据预测分类的方式来实现手势分类。

Method 2



左手大拇指夹持



左手C字形握持

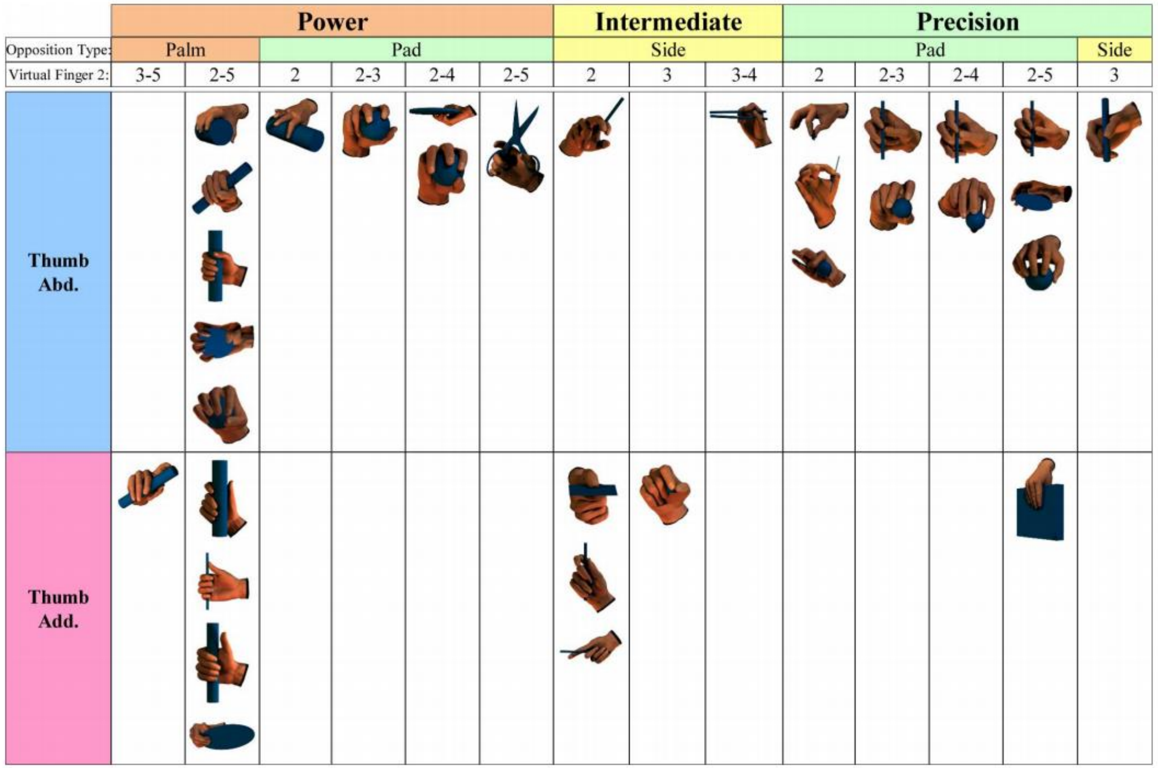


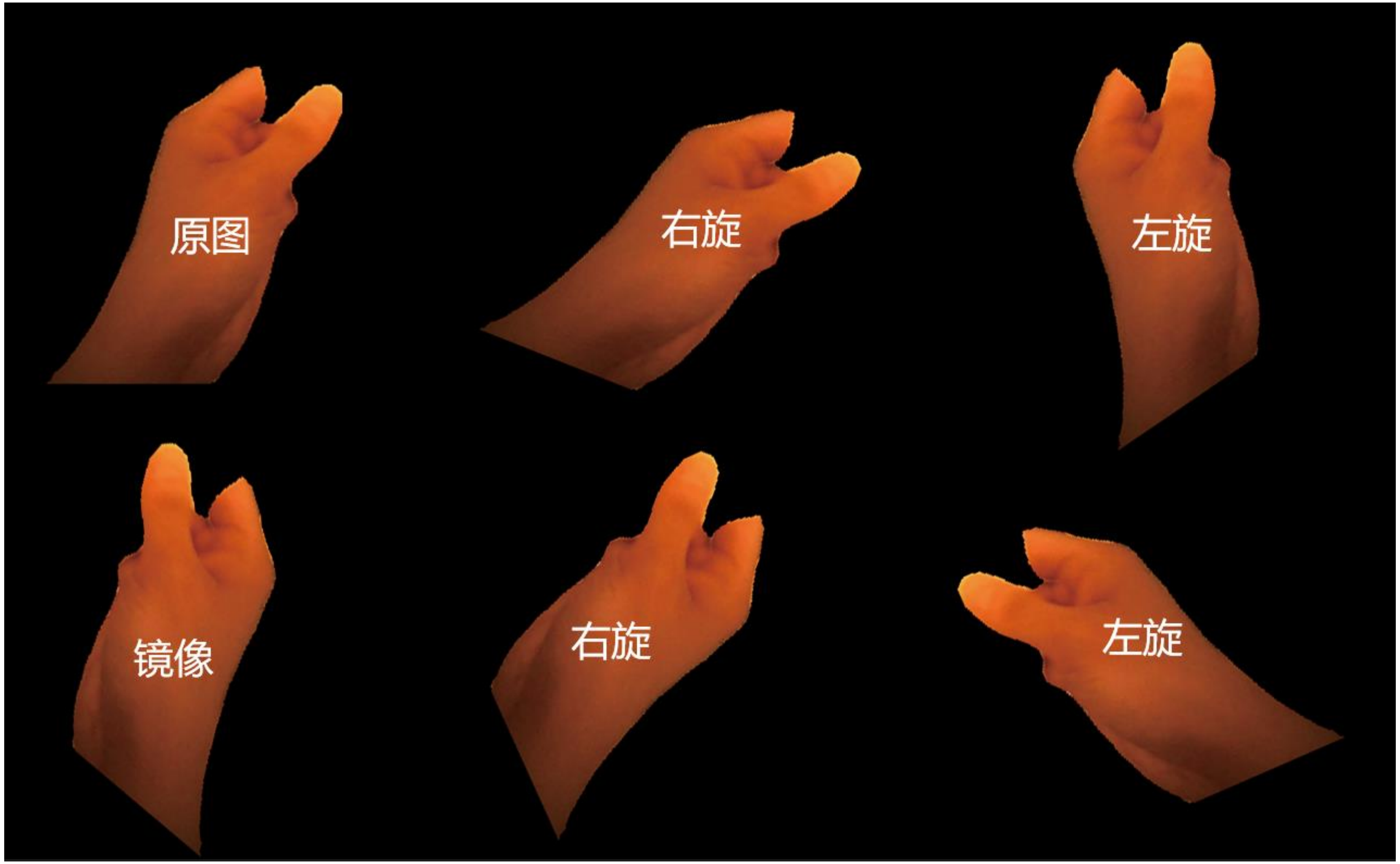
Fig. 2. Comprehensive Grasp Taxonomy which includes 33 grasp types.

参考T. Feix于2009年提出的一种手抓取姿势分类方法并做了简化，我们将我们的姿势分为左右各9种，并额外规定了双手图片以及其他图片2种类别累计20种分类。

T. Feix, R. Pawlik, H.-B. Schmiedmayer, J. Romero, and D. Kragic, “A comprehensive grasp taxonomy,” in Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation, 2009, pp. 2–3.

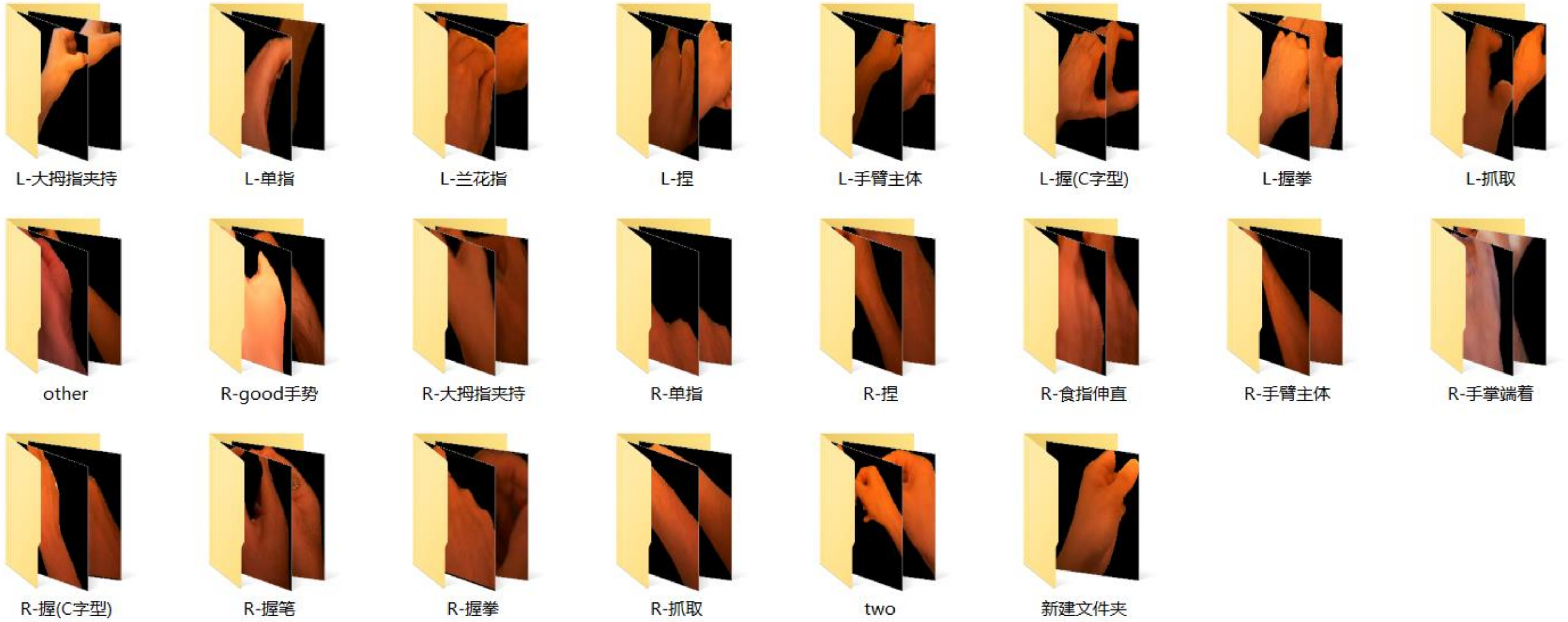
Method 2

我们按照10:1原则对近300张图片进行了人工打标分类，为了让CNN对我们有限的训练集得到的结果更有泛化能力，我们对训练集做了简单的**图像增强**。



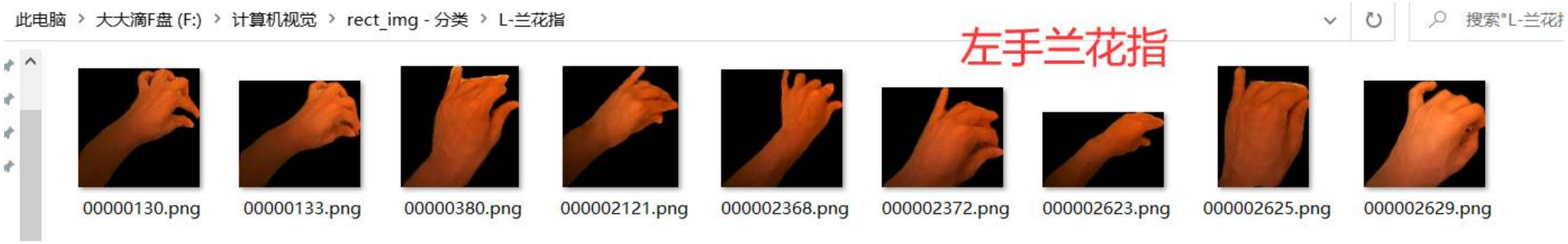
Method 2

按照文件夹排列的最终分类结果



Method 2

按照文件夹排列的最终分类结果例



Method 2

按照文件夹排列的最终分类结果例

Question?

第12组

E202008000106邹雪丹
201708010207 杨佳政
201708010214 唐协成
201508010628 李鑫