

浮游动物特征总结

王如晨 朱亚菲

2015 年 8 月

目录

1	13 类浮游动物的特征	5
2	PkID 中用到的特征	9
2.1	位置特征	9
2.2	尺寸特征	10
2.3	灰度值特征	10
2.4	形状特征	11
2.5	生物统计特征	12
2.6	还没有查找到的特征	12
2.7	其他特征	13
3	采用 PkID 中的特征进行实验	14
3.1	采用 PkID 中的 67 个特征进行实验	14
3.1.1	PkID 中 67 个特征 +SVM Linear	14
3.1.2	PkID 中 67 个特征 +RandomForest	15
3.2	将 67 个特征分类进行实验	16
3.2.1	形状特征 +SVM Linear	16
3.2.2	形状特征 +RandomForest	16
3.2.3	灰度特征 +SVM Linear	17
3.2.4	灰度特征 +RandomForest	17
3.2.5	生物统计特征 +SVM Linear	17
3.2.6	生物统计特征 +RandomForest	18
3.2.7	不包括位置特征 +SVM Linear	18
3.2.8	不包括位置特征 +RandomForest	18

3.2.9	不包括位置特征和尺度特征 +SVM Linear	19
3.2.10	不包括位置特征和尺度特征 +RandomForest	19
3.2.11	不包括位置特征、尺度特征和灰度特征 +SVM Linear	19
3.2.12	不包括位置特征、尺度特征和灰度特征 +RandomForest	21
3.2.13	不包括位置特征、尺度特征、灰度特征和形状特征 +SVM Linear	21
3.2.14	不包括位置特征、尺度特征、灰度特征和形状特征 +RandomForest	22
3.3	从 67 个特征中选取部分特征进行实验	22
3.3.1	选取 16 个特征 +SVM Linear	22
3.3.2	选取 16 个特征 +RandomForest	23
3.3.3	选取 18 个特征 +SVM Linear	23
3.3.4	选取 18 个特征 +RandomForest	23
3.3.5	选取 19 个特征 +SVM Linear	23
3.3.6	选取 19 个特征 +RandomForest	24
3.3.7	选取 21 个特征 +SVM Linear	25
3.3.8	选取 21 个特征 +RandomForest	25
3.3.9	选取 25 个特征 +SVM Linear	26
3.3.10	选取 25 个特征 +RandomForest	27
3.3.11	选取 27 个特征 +SVM Linear	27
3.3.12	选取 27 个特征 +RandomForest	28
3.3.13	选取 29 个特征 +RandomForest	29
3.4	实验分析	29
4	计算机视觉特征提取	32
4.1	几何参数	32
4.1.1	边界的周长	32
4.1.2	边界的曲率	32
4.1.3	面积	32
4.1.4	宽度和高度	32
4.1.5	矩形度	32
4.1.6	体态比	32
4.1.7	圆形性	32
4.1.8	偏心率	33

4.1.9	凸率	33
4.1.10	密集度	33
4.1.11	球状性	33
4.1.12	伸长度	34
4.1.13	叶状性	34
4.2	几种典型的特征描述方法	34
4.2.1	边界描述子	34
4.2.2	区域描述子	34
4.3	特征融合	35
4.3.1	贝叶斯融合 (Bayes Fusion)	35
5	选取特征进行实验	38
5.1	参数特征选取实验	38
5.1.1	实验一	38
5.1.2	实验二	38
5.1.3	实验三	39
5.1.4	实验四	40
5.1.5	实验五	41
5.1.6	实验六	42
5.2	特征融合方法实验	43
5.2.1	实验一 (特征融合方法一)	43
5.2.2	实验二 (特征融合方法二)	43
5.3	融合不同特征实验	44
5.3.1	实验一	45
5.3.2	实验二	45
5.3.3	实验三	45
5.3.4	实验四	46
5.3.5	实验五	46
5.4	ELM 作为分类器的实验	46
5.4.1	实验一	47
6	实验总结	48

7 实验尝试	50
7.1 AdaBoost	50
7.1.1 尝试一 (MATLAB 中的 AdaBoost)	50
7.1.2 尝试二 (特征融合后采用 AdaBoost)	50
7.1.3 尝试三 (将 SVM 作为弱分类器进行级联)	50
7.1.4 尝试四 (将二分类改为多分类问题)	51
7.1.5 AdaBoost 应用	51
7.2 Multi-view Learning (多视角学习)	52
7.3 Bayes Fusion	52
7.4 Fuzzy Neural network	52

1. 13 类浮游动物的特征

Appendicularia (尾海鞘纲) 属于脊索动物门，体型像蝌蚪，身体分为躯干和尾两部分。躯干为椭圆形；尾部扁平，比躯干要长。¹ 大小：小于 5mm。

观察采集的图像发现：

- 形状像蝌蚪，分为躯干和尾部。
- 躯干较大且灰度较深，并不是呈现规则的椭圆（还有部分突出的东西还不知道是什么）。
- 尾部大致呈现两种形状：一种细长弯曲；另一种较粗（粗细甚至于头部差不多），呈现柳叶状。尾部的灰度相比于躯干较浅，轮廓不太清晰。

Bubble(气泡) 非生物。

观察采集的图像发现：

- 圆形。
- 气泡四周灰度深，中间灰度很浅，呈亮白色。

Chaetognatha (毛颚动物门) 全体分为头、躯干和尾 3 部分。头部略膨大，躯干与头连接处稍微缢缩为颈部，尾部末端尖细。头部两侧各着生一系列 4 ~ 13 根镰刀状、几丁质的颚毛。² 大小：最大成年成年个体长 105mm，该类成年个体的长度一般都大于 5mm。

观察采集的图像发现：

- 身体修长，可以明显看出身体分为头、躯干和尾三部分。
- 头部小且圆滑，在头与躯干连接的地方略窄。
- 躯干较粗，轮廓清晰。
- 尾部慢慢变窄，末尾尖。

Cladocera Penilia (Penilia avirostris, 鸟喙尖头蚤) 属于节肢动物门，鳃足纲，枝角目，俗称水跳蚤。大小：大约为 1mm 左右。

观察采集的图像发现：

¹<https://zh.wikipedia.org/wiki/%e5%b0%be%e6%b5%b7%e9%9e%98%e7%ba%b2>

²<https://zh.wikipedia.org/wiki/%e6%af%9b%e9%a2%9a%e5%8a%a8%e7%89%a9%e9%97%a8>

- 身体短小，有两条长长的触角（但并不是每一幅图像中都可以看到。有时触角是向前的，可以看的很清楚；有时触角是向后的，和身体重合在了一起）。
- 该类浮游动物身体中轴线的地方灰度较深（感觉类似人体的脊柱），这个颜色较深的中轴线上还有一条纹理线连向边缘（就像人体脊柱上连着的骨骼）。由于运动，扫描得到的图像中浮游动物的中轴线并不是都在其身体中间。

Copepoda(桡脚类) 属于节肢动物门，颚足纲，桡足类属于其下的一个亚纲。体形像泪珠，有大的触角。分为前体部和后体部，前体部较为宽大，后体部较为短小。³前体部前体部由头和胸部组成，头部有两对触角，胸部有颚足、五对胸足。后体部无附肢，由 3—5 节组成。最末的腹节称尾节，末端具 1 对尾叉，尾叉的末端有 5 根不等长的刚毛，常呈羽状。⁴

观察采集的图像发现：

- 该类动物身体呈长椭圆形，尾部长在椭圆形一段（由于尾叉末端有几根不等长的刚毛，因此尾部呈一簇），触角长在椭圆的另一端（一共有两对触角，但最多只能看到一两个，有的图像甚至看不到）。
- 从该类动物正上方扫描得到的目标关于其自身的中轴对称。从该类动物侧面扫描得到的目标不对称，其身体一侧长着几对胸足。
- 该类中有八十几张图片中有多个目标，应该分在 Multiple 类中。

Decapoda(十足目) 属于节肢动物门，软甲纲。分为两类：Lucifer hansenii 和 Crab larvae。体躯延长呈虾形（腹部发达）或缩短扁圆呈蟹形（腹部化）。⁵

观察采集的图像发现：

- 该类浮游动物形状特征并不是很统一，大致分为两类：虾形和蟹形。
- 一些图像中可以看到目标有一条尾巴（像虾的尾巴）。
- 一些图像中可以看到目标有一对灰度较深的复眼。

Doliolida (海樽目) 属于脊索动物门，樽海鞘纲⁶。体型一般呈桶状，体壁最外是被囊层，其内层是外套膜。被囊层下有 8～9 条肌带环绕着体躯。

观察采集的图像发现：

³<https://zh.wikipedia.org/wiki/%e6%a9%88%e8%85%b3%e9%a1%9e>

⁴<http://baike.baidu.com/view/665478.htm>

⁵http://baike.baidu.com/link?url=LWmrgD_DVUcwOupg_ziOLTlJWj6quxa_juRrS3zUt91A-FjPM6VQwYfZ5fFZckzIyEGCaXypikXUGg2JsYMXUX-uFEkmkLqC5lfxvXvApK3WRBcWQkfbDhMlfTdgrWvh-728gSoUylWZG2UstFK

⁶<https://zh.wikipedia.org/wiki/%e6%a8%bd%e6%b5%b7%e9%9e%98%e7%ba%b2>

- 由于该类浮游动物比较透明，因此在图像中灰度较浅，并且其桶状轮廓也不完整了，但最明显的是能看到大概 7、8 条环状的肌肉带，有的图像中还能看到内部器官。

Egg 很多种类生物的卵。

观察采集的图像发现：

- 形状大致都呈圆形。
- 有的卵整体灰度都很深；有的卵中间有一块灰度较深的区域，四周灰度较浅（结构像细胞）。
- 由于是不同动物的卵，因此其灰度特征差异较大。

Fiber (纤维) 非生物。

- 弯曲的线状，有的纤维有分叉和交叉。
- 该类图像中噪声较多，纤维的边缘也不是很规则。

Gelatinous (明胶) 该类包括很多不同种类的胶状生物，它们体内都含有很高的水分。包括 Aglaura (属于刺胞动物门，这一个没有搜的中文名字，但也属于水母类)、Medusa (水母，属于刺胞动物门，水螅纲)、Siphonophora (管水母，属于刺胞动物门，水螅纲，管水母目)、Radiolaria (放射虫，属于原声动物门，辐足纲) 和 Salps (樽海鞘，属于脊索动物门 Chordata，樽海鞘纲 Thaliacea，纽鳃樽目 Salpida)。其中水母大部分都有三个主要部位：圆伞状或是钟状（寺院里面敲得那种钟）的身体，触器和口腕。

观察采集的图像发现：

- 由于该类呈胶状，因此该类物体灰度整体较浅，边缘也不是十分清晰。大部分是水母，有小部分的樽海鞘（与海樽目形态很相似），小部分的放射虫。
- 其中水母也包括很多类，形态大致呈现以下几种：
 - 一些水母身体呈现类似钟状（这里呈现钟状有长有短，有粗有细，还有的会发生一点弯曲），灰度较浅，内部有一块颜色较深的椭圆形区域。
 - 一些水母也呈钟状，但内部没有颜色较深的椭圆形区域，整个身体灰度均匀。
 - 还有的个头稍微偏小，形状有的类似圆形、像半个胶囊（应该是由拍摄原因，有的拍到顶部，有的拍到侧面），体内有颜色较深的一个大点和几个小点。（可能是灯塔水母）
 - 还有四张看不出形状的，不知道是什么。

- 放射虫：形状近似圆形（但由于整体灰度较浅，形状保存是完整），中间有一块灰度较深的区域，四周灰度较浅，可以看到淡淡的细纹从中心连接到边界。

Multiple (多个生物) 由于浮游动物的重叠，导致分割过程中多个浮游动物被分割到一张图像上。

Nonbio 非生物的集合。（不符合以上集中浮游动物的形态特征）

Pteropoda (翼足目) 属于软体动物门，腹足纲。

观察采集的图像发现：

- 该类浮游动物灰度较深，形状总体都呈现一头宽一头窄。
- 形状总体呈现三类：有的呈现象牙状，有点弯曲；有的较粗短，像一顶尖的小帽子；有的呈现细长的三角形状。

参考网站：<http://www.imas.utas.edu.au/zooplankton/home>。

2. PkID 中用到的特征

PkID 中用到的特征一共有 67 个：Area, Mean, StdDev, Mode, Min, Max, X, Y, XM, YM, Perim., BX, BY, Width, Height, Major, Minor, Angle, Circ., Feret, IntDen, Median, Skew, Kurt, %Area, XStart, YStart, Area_exc, Fractal, Skelarea, Slope, Histcum1, Histcum2, Histcum3, XMg5, YMg5, Compentropy, Compmean, Compslope, CompM1, CompM2, CompM3, Symetrieih, Symetriev, Tag, ESD, Elongation, Range, MeanPos, CentroidsD, CV, SR, PerimAreaexc, FeretAreaexc, PerimFeret, PerimMaj, Circexc, CDexc, [Nb1](#), [Nb2](#), [Nb3](#), [Symetrieihc](#), [Symetrieivc](#), [Convperim](#), [Convarea](#), [Fcons](#), [ThickR](#) (这几个特征没有找到具体的含义)

从训练集的 PID 文件文件中看到, Compentropy, Compmean, Compslope, CompM1, CompM2, CompM3 这 6 个特征在所有图像上的值都为 0, Tag 这个特征在所有图像上的值都为 1, 在训练分类器时是不起作用的, 同时这 7 个特征的具体含义也没有找到。

2.1 位置特征

BX 能够包围物体, 且平行于图像两条边的最小外界矩形的左上角顶点的 X 坐标

BY 能够包围物体, 且平行于图像两条边的最小外界矩形的左上角顶点的 Y 坐标

Height 能够包围物体, 且平行于图像两条边的最小外界矩形的高

Width 能够包围物体, 且平行于图像两条边的最小外界矩形的宽

XStart 图像最左上角像素点的 X 坐标

YStart 图像最左上角像素点的 Y 坐标

XM 物体灰度重心的 X 坐标

YM 物体灰度重心的 Y 坐标

XMg5 gamma 值为 51 时的物体灰度重心的 X 坐标 (gamma 值表示图像输出值与输入值关系的斜率)

YMg5 gamma 值为 51 时的物体灰度重心的 Y 坐标

X 物体重心点的 X 坐标

Y 物体重心点的 Y 坐标

Angle 浮游动物主轴与图片 x 轴形成的夹角，在图片切割后旋转图片测量相关参数使用

这类特征反映的是浮游动物在图像中的位置信息，浮游动物特征与位置信息无关，因此它们不适合作为特征直接用于分类（会降低分类的准确率），而是用来计算其他特征（尺寸特征、灰度特征和形状特征）。

2.2 尺寸特征

Area 物体的表面积，方形像素的个数

Perim 周长，物体最外层边缘的长度

Major 物体的最佳拟合椭圆的长轴

Minor 物体的最佳拟合椭圆的短轴

Feret Maximum feret diameter（最大费雷特征），沿物体边缘任意两个点的最长距离

Area_exc 去掉物体空洞后的表面积，空洞是指灰度值与背景相同的部分

%area 物体表面积中空洞所占的百分比，即背景所占的比例

这类特征表示了图像中目标的大小尺寸。它的根据是同类浮游动物的表面积、周长等尺寸特征应该是大致相同的。但是这些特征还存在着问题：1、同类浮游动物在不同时期（如幼年和成年）的个体大小尺寸是不同的。2、拍摄照片的方位不同（比如正面和侧面）得到的尺寸特征也是不同的。

2.3 灰度值特征

Min 物体内部所有像素点的最小灰度值 (0 = black)

Max 物体内部所有像素点的最大灰度值 (255 = white)

IntDen (Integrated density) 总密度，物体内部像素点的灰度值的总和 ($IntDen = Area * Mean$)

Slope 归一化的灰度累计直方图的斜率

Histcum1 灰度累计直方图的值为 25% 时所对应的灰度值

Histcum2 灰度累计直方图的值为 50% 时所对应的灰度值

Histcum3 灰度累计直方图的值为 75% 时所对应的灰度值

CentroidsD $\sqrt{(XM - X)^2 + (YM - Y)^2}$ 目标物体重心和灰度重心之间的距离。

根据是同类浮游动物的灰度特征（灰度的范围和整体灰度变换趋势）应该是相似的，但观察图像发现并不是所有同类浮游动物的灰度都是相似的，例如 Gelatinous 类中有的个体灰度跨度较小，整体灰度都较浅，而有的个体灰度跨度较大；同时由于拍摄时光线的原因，会造成同类浮游动物中个体灰度的深浅不一。

2.4 形状特征

Fractal 物体边界的分形维数 (Berube and Jebrak, 1999)，表明物体边界的不规则程度

Skelarea 骨架像素的表面积（在二值图像中，不断地从物体边缘处减去像素点直到仅剩一个像素的宽度，最后所得图形的像素点数）

Symetrieih 关于水平轴的对称性

Symetriev 关于竖直轴的对称性

Circ $Circularity = (4 * Pi * Area) / Perim^2$ 圆形度，表征物体接近圆的程度，值等于 1 时，说明物体为正圆形，值越接近 0，物体体形越长。

ESD $2 \times \sqrt{\frac{Area}{\pi}}$ 相应球形直径（也称为等效球直径），是指一不规则外形物体，其体积相同球体的直径。

Elongation $\frac{Major}{Minor}$ 延伸率，最佳拟合椭圆的长轴和短轴之比。

Circexc $\frac{4 \times \pi Area_exc}{Perim^2}$ 去掉目标内部空洞的圆形度。

这类特征描述的是浮游动物的灰度特征，根据的是不同种类浮游动物的形状不同。存在的问题是有不同种类的浮游动物形状相似，例如 Appendicularia 和 Chaetognatha, Bubble 和 Egg；也有同种浮游动物形状不同，例如 Decapoda、Gelatinous。

2.5 生物统计特征

Mean 物体内的平均灰度值；物体中所有像素点的灰度值的总和除以总的像素个数

Range $Max - Min$ 极差，灰度的范围。

CV $100 \times \frac{StdDev}{Mean}$ 变异系数（也称离散系数或相对偏差），是灰度标准偏差与平均值之比，用百分数表示。

SR $100 \times \frac{StdDev}{Max - Min}$ 灰度标准差比上极差。

Skew 灰度直方图的偏度，衡量灰度分布的不对称性。偏度为负就意味着在概率密度函数左侧的尾部比右侧的长，绝大多数的值位于平均值的右侧。偏度为正就意味着在概率密度函数右侧的尾部比左侧的长，绝大多数的值位于平均值的左侧。偏度为零就表示数值相对均匀地分布在平均值的两侧，但不一定意味着其为对称分布。

Kurt 峰度，描述灰度直方图的陡缓程度。

Mean_exc 物体内部去掉空洞后的平均灰度值 ($Mean_exc = IntDen / Area_exc$)

Median 物体内部像素的灰度值的中值

StdDev 物体内部像素的灰度值的标准差

Mode Modal grey value within the object（可能表示灰度的众数）

2.6 还没有查找到的特征

MeanPos $\frac{Mean - Max}{Max - Min}$

PerimAreaexc $\frac{Perim}{\sqrt{Area_exc}}$

FeretAreaexc $\frac{Feret}{\sqrt{Area_exc}}$

PerimFeret $\frac{Perim}{Feret}$

$$\text{PerimMaj} \frac{Perim}{Major}$$

$$\text{CDexc} \frac{\sqrt{(XM - X)^2 + (YM - Y)^2}}{\sqrt{Area_exc}}$$

Nb1 在图像在用阈值 Histcum1 二值化后剩余对象的数量

Nb2 在图像在用阈值 Histcum2 二值化后剩余对象的数量

Nb3 在图像在用阈值 Histcum3 二值化后剩余对象的数量

Symetriehc 在图像在用阈值 Histcum1 二值化后物体的水平对称性

Symetrievc 在图像在用阈值 Histcum1 二值化后物体的垂直对称性

Convperim 包围物体凸包的周长

Convarea 包围物体凸包的面积

Fcons 灰度对比度

ThickR 物体最大厚度和平均厚度（不包括最大厚度）的比值

2.7 其他特征

这些特征并没有在 PkID 中使用，而是在作者的一个幻灯片中提到的新特征。

Neighborhood gray-tone difference Matrix (NGTDM) 邻域灰度差别矩阵

Cumulation Histogram 累积直方图

Convex Area

Symmetry

Thickness Ratio

3. 采用 PkID 中的特征进行实验

3.1 采用 PkID 中的 67 个特征进行实验

3.1.1 PkID 中 67 个特征 +SVM Linear

在 PkID 软件中，选用 67 个特征和 SVM Linear 分类器进行训练和分类得到的结果如图 1，其分类准确率为 70.6%。

Error rate			0.2593													
Values prediction			Confusion matrix													
Value	Recall	1-Precision	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8400	0.1960	2289	0	92	0	6	11	0	1	45	6	123	151	1	2725
Bubble	0.7959	0.2640	0	577	0	0	20	0	0	52	0	0	3	73	0	725
Chaetognatha	0.8883	0.0941	133	0	1550	0	0	10	0	0	14	1	20	17	0	1745
CladoceraPenilia	0.7894	0.1993	0	0	0	3651	18	0	4	0	0	29	5	918	0	4625
Copepoda	0.7984	0.2842	0	0	0	17	5884	187	0	0	2	1	180	1011	88	7370
Decapoda	0.6747	0.2538	2	0	0	0	470	1879	0	0	0	0	88	339	7	2785
Doliolida	0.7502	0.1790	0	0	0	8	0	0	1069	0	0	131	2	215	0	1425
Egg	0.6640	0.1766	0	129	0	25	27	2	0	1142	0	65	0	320	10	1720
Fiber	0.7342	0.2146	44	0	28	0	0	0	0	9	1116	0	39	284	0	1520
Gelatinous	0.6844	0.2238	6	0	1	119	21	10	142	36	0	2060	73	540	2	3010
Multiple	0.3002	0.4646	264	0	21	101	589	174	15	8	50	87	953	897	16	3175
Nonbio	0.8041	0.2890	105	78	16	639	946	185	72	123	194	274	288	12375	95	15390
Pteropoda	0.4525	0.3085	4	0	3	0	239	60	0	16	0	0	6	266	491	1085
Sum			2847	784	1711	4560	8220	2518	1302	1387	1421	2654	1780	17406	710	47300

图 1: PkID 软件中 SVM Linear 分类器分类结果

PkID 系统中 SVM Linear 的参数如图 2。根据图中的参数，自己设计的 SVM Linear 分类器得到的混淆矩阵如图 3。在采用 SVM 进行训练之前，要先对特征进行归一化。

Parameters	
Kernel type	LINEAR
Degree (poly)	1.00
Gamma in kernel function (poly/rbf/sigmoid)	0
Coef0 in kernel function(poly/sigmoid)	0
Tolerance of termination criteria (eps)	0.0001
C (Complexity Cost)	1
Compute probability estimates	0
Use shrinking heuristics	1
Data normalization	1

图 2: PkID 系统中 SVM Linear 的参数

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2325	0	95	3	9	15	0	0	51	5	106	116	0	2725	0.853211	0.202948
Bubble	0	591	0	1	17	0	0	60	0	0	9	47	0	725	0.815172	0.316763
Chaetognatha	144	0	1556	0	0	6	0	0	9	0	14	16	0	1745	0.891691	0.099016
CladoceraPenilia	0	0	0	3797	19	0	8	1	0	30	5	765	0	4625	0.820973	0.196402
Copepoda	0	0	0	14	6119	197	0	0	2	3	167	802	66	7370	0.830258	0.296586
Decapoda	1	0	0	0	565	1894	0	0	0	0	68	247	10	2785	0.680072	0.261022
Doliolida	0	0	0	16	0	0	1148	0	0	94	0	167	0	1425	0.805614	0.157122
Egg	0	172	0	28	90	3	1	1117	0	61	0	240	8	1720	0.649419	0.176254
Fiber	54	0	31	0	0	0	0	9	1108	0	43	275	0	1520	0.728947	0.216407
Gelatinous	8	0	2	131	22	13	127	44	2	2103	60	498	0	3010	0.698671	0.213538
Multiple	260	1	23	114	599	189	15	12	56	107	948	840	11	3175	0.298583	0.443662
Nonbio	120	101	17	621	1003	184	63	99	186	270	279	12346	101	15390	0.802209	0.256265
Pteropoda	5	0	3	0	256	62	0	14	0	1	5	241	498	1085	0.458986	0.282421
Total	2917	865	1727	4725	8699	2563	1362	1356	1414	2674	1704	16600	694	47300	0.717985	0.239877

图 3: PkID-SVM 交叉验证, folds 取 2, repetitions 取 5

3.1.2 PkID 中 67 个特征 + RandomForest

在 PkID 软件中, 选用 67 个特征和随机森林分类器进行训练和分类得到的结果如图 4, 其分类准确率为 75.4%。

Error rate			0.2155														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8051	0.2039	Appendicularia	2194	0	111	8	1	19	0	0	57	1	157	177	0	2725
Bubble	0.8359	0.1049	Bubble	0	606	0	0	0	0	0	29	0	0	3	87	0	725
Chaetognatha	0.8934	0.0983	Chaetognatha	137	0	1559	0	5	2	0	0	7	1	18	14	2	1745
CladoceraPenilia	0.8830	0.1044	CladoceraPenilia	0	0	0	4084	22	0	7	0	0	7	19	486	0	4625
Copepoda	0.8419	0.2108	Copepoda	3	0	0	9	6205	145	0	0	0	1	234	745	28	7370
Decapoda	0.7989	0.2157	Decapoda	4	0	0	0	339	2225	0	0	0	0	87	126	4	2785
Doliolida	0.7474	0.1418	Doliolida	0	0	0	2	1	0	1065	0	0	156	7	194	0	1425
Egg	0.7424	0.1401	Egg	0	66	0	13	2	1	1	1277	0	28	3	320	9	1720
Fiber	0.7414	0.2069	Fiber	59	0	18	0	0	0	0	0	1127	0	62	254	0	1520
Gelatinous	0.6914	0.2042	Gelatinous	6	0	0	72	10	5	93	39	0	2081	79	624	1	3010
Multiple	0.3238	0.5093	Multiple	267	0	23	51	491	202	9	23	45	91	1028	941	4	3175
Nonbio	0.8406	0.2452	Nonbio	86	5	17	321	716	189	66	117	185	246	390	12937	115	15390
Pteropoda	0.6627	0.1848	Pteropoda	0	0	1	0	70	49	0	0	0	3	8	235	719	1085
			Sum	2756	677	1729	4560	7862	2837	1241	1485	1421	2615	2095	17140	882	47300

图 4: PkID 软件中随机森林分类器分类结果

自己设计随机森林分类器, 对 67 个特征进行训练和分类得到的混淆矩阵如图 5。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2220	0	110	13	1	21	0	0	61	2	137	160	0	2725	0.8147	0.195069
Bubble	0	629	0	0	3	0	0	28	0	0	4	61	0	725	0.8676	0.096264
Chaetognatha	147	0	1550	0	2	6	0	0	4	1	23	12	0	1745	0.8883	0.093567
CladoceraPenilia	0	0	0	4156	18	0	7	0	0	1	11	432	0	4625	0.8986	0.079717
Copepoda	4	0	0	7	6309	161	0	0	0	0	193	679	17	7370	0.856	0.190323
Decapoda	1	0	0	0	316	2270	0	0	0	0	76	121	1	2785	0.8151	0.221269
Doliolida	0	0	0	1	0	0	1104	0	0	160	7	153	0	1425	0.7747	0.110395
Egg	0	57	0	15	3	1	0	1361	0	30	1	246	6	1720	0.7913	0.098675
Fiber	46	0	19	0	0	0	5	0	1129	0	67	254	0	1520	0.7428	0.204369
Gelatinous	6	0	0	49	12	6	67	19	0	2127	88	635	1	3010	0.7066	0.196145
Multiple	266	2	16	38	452	219	13	17	43	93	1054	955	7	3175	0.332	0.481299
Nonbio	68	8	13	237	626	187	45	85	182	227	366	13240	106	15390	0.8603	0.230635
Pteropoda	0	0	2	0	50	44	0	0	0	5	5	261	718	1085	0.6618	0.161215
Total	2758	696	1710	4516	7792	2915	1241	1510	1419	2646	2032	17209	856	47300	0.77	0.181457

图 5: PkID-RF 交叉验证, folds 取 2, repetitions 取 5

3.2 将 67 个特征分类进行实验

3.2.1 形状特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中选取形状特征 (8 个), 并采用 SVM Linear 进行训练和分类得到的结果如图 6, 其分类准确率为 37%。

Error rate			0.5353														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.6095	0.4093	Appendicularia	1661	0	102	0	9	54	0	0	153	5	41	700	0	2725
Bubble	0.0000	1.0000	Bubble	0	0	0	0	1	0	0	415	0	0	0	309	0	725
Chaetognatha	0.8080	0.1334	Chaetognatha	227	0	1410	0	0	18	0	0	61	2	19	8	0	1745
CladoceraPenilia	0.0465	0.7018	CladoceraPenilia	0	0	0	215	52	0	0	1	0	39	0	4318	0	4625
Copepoda	0.0837	0.6425	Copepoda	14	0	0	17	617	63	0	0	0	16	70	6572	1	7370
Decapoda	0.2743	0.4264	Decapoda	71	0	1	8	147	764	0	0	0	25	121	1648	0	2785
Doliolida	0.3354	0.5047	Doliolida	0	0	0	1	3	0	478	0	0	495	1	447	0	1425
Egg	0.5860	0.4886	Egg	0	0	0	18	0	0	1	1008	0	72	0	621	0	1720
Fiber	0.6039	0.3035	Fiber	167	0	53	0	0	0	0	6	918	0	28	348	0	1520
Gelatinous	0.4924	0.3789	Gelatinous	7	0	0	23	73	49	201	303	0	1482	21	840	11	3010
Multiple	0.0967	0.5518	Multiple	295	0	25	41	126	191	46	17	34	70	307	2023	0	3175
Nonbio	0.8500	0.5874	Nonbio	309	0	31	397	619	120	238	204	152	166	73	13081	0	15390
Pteropoda	0.0359	0.2353	Pteropoda	61	0	5	1	79	73	1	17	0	14	4	791	39	1085
			Sum	2812	0	1627	721	1726	1332	965	1971	1318	2386	685	31706	51	47300

图 6: 采用形状特征 SVM Linear 进行分类的结果

3.2.2 形状特征 +RandomForest

在 PkID 软件中, 从 67 个特征中选取形状特征 (8 个), 并采用随机森林进行训练和分类得到的结果如图 7, 其分类准确率为 63.1%。

Error rate			0.3285														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7006	0.3417	Appendicularia	1909	0	129	3	8	54	5	0	110	1	188	289	29	2725
Bubble	0.5959	0.3143	Bubble	0	432	0	0	3	0	0	149	0	0	0	141	0	725
Chaetognatha	0.8521	0.1053	Chaetognatha	194	0	1487	0	0	8	0	0	22	0	25	8	1	1745
CladoceraPenilia	0.7529	0.3051	CladoceraPenilia	0	0	0	3482	141	27	2	4	0	15	45	900	9	4625
Copepoda	0.7331	0.3082	Copepoda	12	0	0	197	5403	205	8	0	1	35	280	1214	15	7370
Decapoda	0.6668	0.3284	Decapoda	57	0	5	71	284	1857	0	0	0	15	174	311	11	2785
Doliolida	0.5979	0.3804	Doliolida	0	0	0	17	41	0	852	2	0	306	7	198	2	1425
Egg	0.5593	0.3109	Egg	0	175	0	81	0	0	4	962	2	71	0	421	4	1720
Fiber	0.7158	0.3008	Fiber	88	0	15	0	4	0	0	5	1088	1	59	258	2	1520
Gelatinous	0.5638	0.3279	Gelatinous	7	3	0	84	121	13	260	127	0	1697	46	593	59	3010
Multiple	0.2243	0.6229	Multiple	312	0	12	220	436	242	54	11	86	93	712	986	11	3175
Nonbio	0.7364	0.3324	Nonbio	258	20	6	845	1336	307	188	136	247	256	344	11333	114	15390
Pteropoda	0.5060	0.3189	Pteropoda	63	0	8	11	33	52	2	0	0	35	8	324	549	1085
			Sum	2900	630	1662	5011	7810	2765	1375	1396	1556	2525	1888	16976	806	47300

图 7: 采用形状特征随机森林进行分类的结果

3.2.3 灰度特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中选取灰度特征 (18 个, 包括生物统计特征), 并采用 SVM Linear 进行训练和分类得到的结果如图 8, 其分类准确率为 37.6%。

Error rate			0.4549														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.6914	0.3144	Appendicularia	1884	0	178	102	36	15	0	0	0	49	100	361	0	2725
Bubble	0.0566	0.6168	Bubble	0	41	0	5	94	0	0	46	0	0	1	538	0	725
Chaetognatha	0.5433	0.2582	Chaetognatha	329	0	948	5	31	5	5	0	0	48	76	298	0	1745
CladoceraPenilia	0.4415	0.4855	CladoceraPenilia	0	0	0	2042	12	0	7	0	0	2	1	2561	0	4625
Copepoda	0.6836	0.4252	Copepoda	0	6	0	28	5038	301	0	30	0	0	81	1860	26	7370
Decapoda	0.5451	0.3286	Decapoda	0	0	0	3	737	1518	0	13	0	0	66	433	15	2785
Doliolida	0.3705	0.4477	Doliolida	1	0	0	10	0	0	528	0	0	63	6	817	0	1425
Egg	0.1994	0.6163	Egg	0	3	0	67	417	35	4	343	0	14	9	776	52	1720
Fiber	0.0020	0.0000	Fiber	49	0	14	40	26	0	19	0	3	6	22	1341	0	1520
Gelatinous	0.2066	0.4842	Gelatinous	78	1	52	476	26	18	217	3	0	622	144	1373	0	3010
Multiple	0.1644	0.5639	Multiple	276	2	61	331	683	162	35	12	0	153	522	930	8	3175
Nonbio	0.7843	0.4873	Nonbio	126	49	25	860	1309	164	141	185	0	249	163	12071	48	15390
Pteropoda	0.2037	0.4027	Pteropoda	5	5	0	0	356	43	0	262	0	0	6	187	221	1085
			Sum	2748	107	1278	3969	8765	2261	956	894	3	1206	1197	23546	370	47300

图 8: 采用灰度特征 SVM Linear 进行分类的结果

3.2.4 灰度特征 +RandomForest

在 PkID 软件中, 从 67 个特征中选取灰度特征 (18 个, 包括生物统计特征), 并采用随机森林进行训练和分类得到的结果如图 9, 其分类准确率为 61.3%。

Error rate			0.3204														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7160	0.2929	Appendicularia	1951	0	235	54	24	25	4	0	36	41	150	205	0	2725
Bubble	0.8166	0.1834	Bubble	0	592	0	0	3	0	0	8	0	0	5	117	0	725
Chaetognatha	0.7278	0.2853	Chaetognatha	229	0	1270	0	11	3	5	0	5	46	73	99	4	1745
CladoceraPenilia	0.8666	0.1918	CladoceraPenilia	8	0	0	4008	30	0	19	2	0	49	39	470	0	4625
Copepoda	0.7659	0.3438	Copepoda	10	0	7	13	5645	408	0	9	0	11	196	987	84	7370
Decapoda	0.7260	0.3090	Decapoda	6	0	0	1	527	2022	0	5	0	2	86	114	22	2785
Doliolida	0.7200	0.2914	Doliolida	1	0	0	23	0	0	1026	0	0	187	24	164	0	1425
Egg	0.4645	0.2455	Egg	0	41	0	12	188	45	0	799	2	23	49	491	70	1720
Fiber	0.2605	0.3333	Fiber	75	0	9	37	28	0	12	0	396	6	34	923	0	1520
Gelatinous	0.4837	0.4245	Gelatinous	56	0	78	191	16	11	200	7	2	1456	203	789	1	3010
Multiple	0.2466	0.6139	Multiple	297	6	100	90	627	172	38	34	8	220	783	777	23	3175
Nonbio	0.7637	0.3136	Nonbio	126	86	78	526	1194	187	144	163	145	486	378	11753	124	15390
Pteropoda	0.4074	0.4260	Pteropoda	0	0	0	4	310	53	0	32	0	3	8	233	442	1085
			Sum	2759	725	1777	4959	8603	2926	1448	1059	594	2530	2028	17122	770	47300

图 9: 采用灰度特征随机森林进行分类的结果

3.2.5 生物统计特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中选取生物统计特征 (16 个, 包括没有查找到的特征), 并采用 SVM Linear 进行训练和分类得到的结果如图 10, 其分类准确率为 52.4%。

Error rate			0.3764														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7831	0.2149	Appendicularia	2134	0	118	25	11	4	0	0	31	16	191	195	0	2725
Bubble	0.2345	0.3950	Bubble	0	170	0	4	31	0	0	76	0	0	10	434	0	725
Chaetognatha	0.8350	0.1275	Chaetognatha	171	0	1457	4	8	6	0	0	32	0	31	36	0	1745
CladoceraPenilia	0.5952	0.3541	CladoceraPenilia	0	0	0	2753	9	0	7	0	0	40	8	1808	0	4625
Copepoda	0.6782	0.3999	Copepoda	0	0	0	14	4998	337	0	0	6	0	191	1715	109	7370
Decapoda	0.5110	0.3544	Decapoda	3	0	0	0	911	1423	0	3	0	0	75	351	19	2785
Doliolida	0.3586	0.3122	Doliolida	0	0	0	9	0	0	511	0	0	218	12	675	0	1425
Egg	0.3855	0.3094	Egg	0	42	0	70	53	7	4	663	0	47	3	810	21	1720
Fiber	0.6217	0.2248	Fiber	38	0	61	3	17	0	0	0	945	0	35	421	0	1520
Gelatinous	0.4286	0.3505	Gelatinous	10	2	0	384	31	8	95	9	0	1290	81	1100	0	3010
Multiple	0.2539	0.5322	Multiple	270	0	19	208	627	161	35	15	50	62	806	914	8	3175
Nonbio	0.7778	0.4178	Nonbio	89	67	15	788	1189	192	91	135	155	313	275	11971	110	15390
Pteropoda	0.3465	0.4152	Pteropoda	3	0	0	0	444	66	0	59	0	0	5	132	376	1085
			Sum	2718	281	1670	4262	8329	2204	743	960	1219	1986	1723	20562	643	47300

图 10: 采用生物统计特征 SVM Linear 进行分类的结果

3.2.6 生物统计特征 +RandomForest

在 PkID 软件中, 从 67 个特征中选取生物统计特征 (16 个, 包括没有查找到的特征), 并采用随机森林进行训练和分类得到的结果如图 11, 其分类准确率为 69.9%。

Error rate			0.2699														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7828	0.2412	Appendicularia	2133	0	137	0	6	22	3	0	55	15	204	150	0	2725
Bubble	0.7986	0.1201	Bubble	0	579	0	0	15	1	0	29	0	6	14	81	0	725
Chaetognatha	0.8223	0.1345	Chaetognatha	210	0	1435	0	11	8	0	0	35	0	16	29	1	1745
CladoceraPenilia	0.8322	0.1980	CladoceraPenilia	0	0	0	3849	21	0	53	0	0	18	54	630	0	4625
Copepoda	0.8262	0.2703	Copepoda	2	0	0	12	6089	243	0	0	0	0	214	764	46	7370
Decapoda	0.7056	0.2660	Decapoda	2	0	1	0	513	1965	0	0	0	0	101	192	11	2785
Doliolida	0.7404	0.2596	Doliolida	0	0	0	19	0	0	1055	0	0	94	12	245	0	1425
Egg	0.7157	0.1475	Egg	0	57	0	5	8	6	5	1231	0	82	19	303	4	1720
Fiber	0.6553	0.2735	Fiber	58	0	42	1	3	0	0	0	996	1	87	326	6	1520
Gelatinous	0.5721	0.2963	Gelatinous	13	3	0	127	9	1	94	32	0	1722	109	898	2	3010
Multiple	0.2825	0.5849	Multiple	270	2	24	139	571	200	30	35	70	94	897	830	13	3175
Nonbio	0.7767	0.2817	Nonbio	123	17	19	647	917	208	185	117	215	411	427	11954	150	15390
Pteropoda	0.5797	0.2703	Pteropoda	0	0	0	0	181	23	0	0	0	4	7	241	629	1085
			Sum	2811	658	1658	4799	8344	2677	1425	1444	1371	2447	2161	16643	862	47300

图 11: 采用生物统计特征随机森林进行分类的结果

3.2.7 不包括位置特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中去掉 13 个位置特征, 并采用 SVM Linear 进行训练和分类得到的结果如图 12, 其分类准确率为 70.8%。

3.2.8 不包括位置特征 +RandomForest

在 PkID 软件中, 从 67 个特征中去掉 13 个位置特征, 并采用随机森林进行训练和分类得到的结果如图 13, 其分类准确率为 76.1%。

Error rate			0.2592														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8426	0.1901	Appendicularia	2296	0	84	1	10	11	0	1	39	4	122	157	0	2725
Bubble	0.8400	0.2784	Bubble	0	609	0	1	19	0	0	23	0	0	2	71	0	725
Chaetognatha	0.8968	0.0826	Chaetognatha	124	0	1565	0	0	8	0	0	15	0	18	15	0	1745
CladoceraPenilia	0.7948	0.2026	CladoceraPenilia	0	0	0	3676	19	0	5	0	0	31	6	888	0	4625
Copepoda	0.7986	0.2908	Copepoda	0	0	0	16	5886	185	0	0	1	0	190	1011	81	7370
Decapoda	0.6661	0.2604	Decapoda	1	0	0	0	515	1855	0	0	0	0	92	313	9	2785
Doliolida	0.7481	0.1704	Doliolida	0	0	0	7	0	0	1066	0	0	132	3	217	0	1425
Egg	0.6448	0.1515	Egg	0	154	0	23	29	1	0	1109	0	65	0	328	11	1720
Fiber	0.7342	0.2196	Fiber	54	0	21	0	0	0	0	8	1116	0	42	279	0	1520
Gelatinous	0.6973	0.2220	Gelatinous	5	0	1	114	19	10	132	31	3	2099	62	531	3	3010
Multiple	0.3002	0.4591	Multiple	241	0	19	116	592	178	12	10	58	84	953	901	11	3175
Nonbio	0.8016	0.2880	Nonbio	109	81	15	656	965	197	70	115	198	283	266	12336	99	15390
Pteropoda	0.4387	0.3101	Pteropoda	5	0	1	0	246	63	0	10	0	0	6	278	476	1085
			Sum	2835	844	1706	4610	8300	2508	1285	1307	1430	2698	1762	17325	690	47300

图 12: 不采用位置特征 SVM Linear 进行分类的结果

Error rate			0.2092														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8048	0.2109	Appendicularia	2193	0	117	8	1	19	0	0	47	4	177	159	0	2725
Bubble	0.8262	0.1152	Bubble	0	599	0	0	1	0	0	44	0	0	0	81	0	725
Chaetognatha	0.8871	0.0979	Chaetognatha	147	0	1548	0	2	4	0	0	4	1	27	12	0	1745
CladoceraPenilia	0.8962	0.0944	CladoceraPenilia	0	0	0	4145	19	0	4	0	0	6	15	436	0	4625
Copepoda	0.8468	0.2051	Copepoda	4	0	0	5	6241	138	0	0	2	5	220	723	32	7370
Decapoda	0.7953	0.2173	Decapoda	7	0	0	0	321	2215	0	0	0	0	101	135	6	2785
Doliolida	0.7740	0.1239	Doliolida	0	0	0	2	0	0	1103	0	0	141	9	170	0	1425
Egg	0.7669	0.1052	Egg	0	68	0	12	1	2	0	1319	0	26	2	276	14	1720
Fiber	0.7586	0.2172	Fiber	37	0	12	0	0	0	0	0	1153	0	65	251	2	1520
Gelatinous	0.7213	0.2021	Gelatinous	7	0	2	58	9	4	69	27	0	2171	91	571	1	3010
Multiple	0.3203	0.5176	Multiple	291	0	19	52	486	213	5	14	57	95	1017	920	6	3175
Nonbio	0.8441	0.2344	Nonbio	93	10	12	295	711	185	78	70	209	267	375	12991	94	15390
Pteropoda	0.6562	0.1788	Pteropoda	0	0	6	0	59	50	0	0	1	5	9	243	712	1085
			Sum	2779	677	1716	4577	7851	2830	1259	1474	1473	2721	2108	16968	867	47300

图 13: 不采用位置特征随机森林进行分类的结果

3.2.9 不包括位置特征和尺度特征 +SVM Linear

在 PkID 软件中，从 67 个特征中去掉位置特征和尺度特征，并采用 SVM Linear 进行训练和分类得到的结果如图 14，其分类准确率为 70.3%。

3.2.10 不包括位置特征和尺度特征 +RandomForest

在 PkID 软件中，从 67 个特征中去掉位置特征和尺度特征，并采用随机森林进行训练和分类得到的结果如图 15，其分类准确率为 76.4%。

3.2.11 不包括位置特征、尺度特征和灰度特征 +SVM Linear

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征和灰度特征，并采用 SVM Linear 进行训练和分类得到的结果如图 16，其分类准确率为 69.2%。

Error rate			0.2618														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8378	0.1904	Appendicularia	2283	0	89	4	10	10	0	0	43	4	122	160	0	2725
Bubble	0.8262	0.2826	Bubble	0	599	0	1	18	0	0	32	0	0	3	72	0	725
Chaetognatha	0.8968	0.0885	Chaetognatha	125	0	1565	0	0	5	0	0	16	1	16	16	1	1745
CladoceraPenilia	0.7825	0.2098	CladoceraPenilia	0	0	0	3619	23	0	5	0	0	33	7	938	0	4625
Copepoda	0.7996	0.2926	Copepoda	0	0	0	18	5893	182	0	0	1	1	184	1010	81	7370
Decapoda	0.6614	0.2570	Decapoda	0	0	0	0	531	1842	0	0	0	0	91	311	10	2785
Doliolida	0.7396	0.1778	Doliolida	0	0	0	7	0	0	1054	0	0	140	3	221	0	1425
Egg	0.6256	0.1639	Egg	0	155	0	31	42	1	0	1076	0	69	0	337	9	1720
Fiber	0.7342	0.2201	Fiber	55	0	27	2	0	0	0	2	1116	0	39	279	0	1520
Gelatinous	0.6924	0.2259	Gelatinous	5	0	2	112	19	11	142	38	3	2084	62	530	2	3010
Multiple	0.2989	0.4555	Multiple	236	0	19	129	587	181	11	9	60	87	949	895	12	3175
Nonbio	0.8029	0.2902	Nonbio	111	81	14	657	964	190	70	120	192	273	262	12357	99	15390
Pteropoda	0.4424	0.3084	Pteropoda	5	0	1	0	244	57	0	10	0	0	5	283	480	1085
			Sum	2820	835	1717	4580	8331	2479	1282	1287	1431	2692	1743	17409	694	47300

图 14: 不采用位置特征和尺度特征 SVM Linear 进行分类的结果

Error rate			0.2085														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8015	0.2075	Appendicularia	2184	0	106	7	0	29	0	0	54	4	182	159	0	2725
Bubble	0.8414	0.1003	Bubble	0	610	0	0	2	0	0	34	0	0	0	79	0	725
Chaetognatha	0.8951	0.0924	Chaetognatha	133	0	1562	0	1	3	0	0	13	1	19	13	0	1745
CladoceraPenilia	0.8893	0.0848	CladoceraPenilia	0	0	0	4113	21	0	5	0	0	7	17	462	0	4625
Copepoda	0.8488	0.2019	Copepoda	2	0	0	5	6256	137	0	0	2	3	228	712	25	7370
Decapoda	0.8025	0.2213	Decapoda	1	0	0	0	300	2235	0	0	0	0	96	145	8	2785
Doliolida	0.7860	0.1257	Doliolida	0	0	0	2	0	0	1120	0	0	134	10	159	0	1425
Egg	0.7669	0.1009	Egg	0	61	0	7	4	1	1	1319	0	25	7	282	13	1720
Fiber	0.7493	0.2335	Fiber	49	0	18	0	0	0	1	0	1139	0	66	247	0	1520
Gelatinous	0.7226	0.1986	Gelatinous	7	0	0	54	8	5	69	24	0	2175	84	584	0	3010
Multiple	0.3282	0.5133	Multiple	277	0	20	39	479	219	7	15	52	100	1042	921	4	3175
Nonbio	0.8427	0.2359	Nonbio	102	7	12	267	712	187	78	75	226	261	379	12969	115	15390
Pteropoda	0.6599	0.1873	Pteropoda	1	0	3	0	56	54	0	0	0	4	11	240	716	1085
			Sum	2756	678	1721	4494	7839	2870	1281	1467	1486	2714	2141	16972	881	47300

图 15: 不采用位置特征和尺度特征随机森林进行分类的结果

Error rate			0.2691														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8374	0.1945	Appendicularia	2282	0	93	3	11	10	0	0	42	5	126	153	0	2725
Bubble	0.8055	0.2930	Bubble	0	584	0	3	19	0	0	45	0	0	1	73	0	725
Chaetognatha	0.8980	0.0884	Chaetognatha	134	0	1567	0	0	7	0	0	10	0	14	13	0	1745
CladoceraPenilia	0.7619	0.2172	CladoceraPenilia	0	0	0	3524	23	0	4	0	0	46	6	1022	0	4625
Copepoda	0.7965	0.3007	Copepoda	0	0	0	17	5870	171	0	0	1	4	189	1021	97	7370
Decapoda	0.6309	0.2667	Decapoda	2	0	0	0	592	1757	0	0	0	0	97	327	10	2785
Doliolida	0.6975	0.1990	Doliolida	0	0	0	8	0	0	994	0	0	175	2	246	0	1425
Egg	0.6105	0.1758	Egg	0	168	0	36	43	3	0	1050	0	75	0	337	8	1720
Fiber	0.7349	0.2139	Fiber	54	0	25	1	0	0	0	2	1117	0	43	278	0	1520
Gelatinous	0.6857	0.2431	Gelatinous	5	0	1	108	20	10	149	44	4	2064	60	542	3	3010
Multiple	0.2961	0.4588	Multiple	241	0	18	130	593	189	10	13	57	88	940	883	13	3175
Nonbio	0.8026	0.2955	Nonbio	111	74	14	672	978	183	84	111	190	270	254	12352	97	15390
Pteropoda	0.4323	0.3271	Pteropoda	4	0	1	0	245	66	0	9	0	0	5	286	469	1085
			Sum	2833	826	1719	4502	8394	2396	1241	1274	1421	2727	1737	17533	697	47300

图 16: 不采用位置特征、尺度特征和灰度特征 SVM Linear 进行分类的结果

3.2.12 不包括位置特征、尺度特征和灰度特征 +RandomForest

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征和灰度特征，并采用随机森林进行训练和分类得到的结果如图 17，其分类准确率为 77.2%。

Error rate			0.2056														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8154	0.2076	Appendicularia	2222	0	110	5	0	17	0	0	41	5	166	159	0	2725
Bubble	0.8855	0.0802	Bubble	0	642	0	0	3	0	0	28	0	0	0	52	0	725
Chaetognatha	0.8957	0.0849	Chaetognatha	140	0	1563	0	0	4	0	0	4	0	22	12	0	1745
CladoceraPenilia	0.8878	0.0936	CladoceraPenilia	0	0	0	4106	22	0	5	0	0	6	16	470	0	4625
Copepoda	0.8514	0.2060	Copepoda	2	0	0	6	6275	132	0	0	2	3	223	709	18	7370
Decapoda	0.7932	0.2080	Decapoda	8	0	0	0	307	2209	0	0	0	0	103	149	9	2785
Doliolida	0.7832	0.1240	Doliolida	0	0	0	0	0	0	1116	0	0	157	10	142	0	1425
Egg	0.7878	0.1032	Egg	0	51	0	12	0	0	0	1355	0	21	2	277	2	1720
Fiber	0.7461	0.2141	Fiber	51	0	11	0	0	0	3	0	1134	0	59	262	0	1520
Gelatinous	0.7279	0.1945	Gelatinous	5	0	0	54	12	1	73	21	1	2191	91	561	0	3010
Multiple	0.3288	0.5153	Multiple	270	0	12	47	499	213	8	15	63	101	1044	901	2	3175
Nonbio	0.8428	0.2318	Nonbio	105	5	12	300	739	163	69	92	196	234	403	12970	102	15390
Pteropoda	0.6903	0.1508	Pteropoda	1	0	0	0	46	50	0	0	2	2	15	220	749	1085
			Sum	2804	698	1708	4530	7903	2789	1274	1511	1443	2720	2154	16884	882	47300

图 17: 不采用位置特征、尺度特征和灰度特征随机森林进行分类的结果

3.2.13 不包括位置特征、尺度特征、灰度特征和形状特征 +SVM Linear

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征、灰度特征和形状特征，并采用 SVM Linear 进行训练和分类得到的结果如图 18，其分类准确率为 63.3%。

Error rate			0.3126														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8198	0.2053	Appendicularia	2234	0	102	15	16	10	0	0	48	1	135	164	0	2725
Bubble	0.7545	0.3058	Bubble	0	547	0	9	25	0	0	83	0	0	0	61	0	725
Chaetognatha	0.8877	0.1041	Chaetognatha	129	0	1549	0	3	3	0	0	19	2	24	16	0	1745
CladoceraPenilia	0.6904	0.3092	CladoceraPenilia	0	0	0	3193	27	0	0	0	0	6	5	1394	0	4625
Copepoda	0.7501	0.3325	Copepoda	0	0	0	32	5528	159	0	0	3	0	166	1383	99	7370
Decapoda	0.5774	0.2674	Decapoda	2	0	0	0	618	1608	0	0	0	0	100	448	9	2785
Doliolida	0.5432	0.2110	Doliolida	0	0	0	4	0	0	774	0	0	242	3	402	0	1425
Egg	0.5453	0.2514	Egg	0	152	0	98	34	0	1	938	0	39	0	447	11	1720
Fiber	0.6947	0.2201	Fiber	62	0	35	2	3	0	0	0	1056	0	40	322	0	1520
Gelatinous	0.5571	0.2583	Gelatinous	16	0	0	247	36	17	108	7	2	1677	66	834	0	3010
Multiple	0.2743	0.4772	Multiple	267	2	28	176	567	162	19	4	54	88	871	929	8	3175
Nonbio	0.7907	0.3513	Nonbio	96	86	15	846	1064	159	79	145	172	206	251	12169	102	15390
Pteropoda	0.3419	0.3817	Pteropoda	5	1	0	0	361	77	0	76	0	0	5	189	371	1085
			Sum	2811	788	1729	4622	8282	2195	981	1253	1354	2261	1666	18758	600	47300

图 18: 不采用位置特征、尺度特征、灰度特征和形状特征 SVM Linear 进行分类的结果

3.2.14 不包括位置特征、尺度特征、灰度特征和形状特征 +RandomForest

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征、灰度特征和形状特征，并采用随机森林进行训练和分类得到的结果如图 19，其分类准确率为 75.7%。

Error rate			0.2160														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8187	0.2066	Appendicularia	2231	0	104	3	2	16	1	0	55	9	158	146	0	2725
Bubble	0.8910	0.0785	Bubble	0	646	0	0	2	0	0	30	0	0	1	46	0	725
Chaetognatha	0.8848	0.0891	Chaetognatha	129	0	1544	0	5	4	0	0	19	1	36	7	0	1745
CladoceraPenilia	0.8778	0.1304	CladoceraPenilia	0	0	0	4060	27	0	6	0	0	7	23	502	0	4625
Copepoda	0.8472	0.2132	Copepoda	1	0	0	22	6244	183	0	0	0	4	219	667	30	7370
Decapoda	0.7842	0.2175	Decapoda	7	0	0	0	326	2184	0	0	0	1	109	155	3	2785
Doliolida	0.7825	0.1449	Doliolida	0	0	0	0	0	0	1115	0	0	127	8	175	0	1425
Egg	0.7814	0.0937	Egg	0	52	0	6	1	0	0	1344	0	22	1	285	9	1720
Fiber	0.6980	0.2106	Fiber	65	0	25	0	0	0	3	0	1061	0	75	291	0	1520
Gelatinous	0.6688	0.2034	Gelatinous	13	0	0	88	18	7	74	12	6	2013	127	652	0	3010
Multiple	0.3291	0.5276	Multiple	281	2	12	53	496	198	15	16	40	90	1045	917	10	3175
Nonbio	0.8378	0.2407	Nonbio	85	1	10	437	720	168	90	79	163	253	397	12893	94	15390
Pteropoda	0.6461	0.1724	Pteropoda	0	0	0	0	95	31	0	2	0	0	13	243	701	1085
			Sum	2812	701	1695	4669	7936	2791	1304	1483	1344	2527	2212	16979	847	47300

图 19: 不采用位置特征、尺度特征、灰度特征和形状特征随机森林进行分类的结果

3.3 从 67 个特征中选取部分特征进行实验

3.3.1 选取 16 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 16 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc），并采用 SVM Linear 进行训练和分类得到的结果如图 20。

Error rate			0.3391														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8048	0.2218	Appendicularia	2193	0	101	20	3	8	0	0	28	13	129	227	3	2725
Bubble	0.7807	0.3213	Bubble	0	566	0	2	23	1	0	10	0	0	5	118	0	725
Chaetognatha	0.8315	0.1004	Chaetognatha	193	0	1451	4	2	1	0	0	8	5	38	43	0	1745
CladoceraPenilia	0.5963	0.2995	CladoceraPenilia	0	0	0	2758	18	0	2	0	0	56	0	1791	0	4625
Copepoda	0.7004	0.3956	Copepoda	0	0	0	5	5162	234	0	0	5	10	155	1779	20	7370
Decapoda	0.5059	0.2905	Decapoda	4	0	0	0	986	1409	0	1	0	0	78	306	1	2785
Doliolida	0.5726	0.3131	Doliolida	0	0	0	4	0	0	816	0	0	257	0	348	0	1425
Egg	0.5721	0.2096	Egg	0	157	0	27	42	7	0	984	0	70	1	419	13	1720
Fiber	0.6428	0.2184	Fiber	53	0	26	9	26	0	0	3	977	1	39	386	0	1520
Gelatinous	0.6063	0.3074	Gelatinous	6	0	0	133	40	8	204	95	0	1825	86	612	1	3010
Multiple	0.2485	0.4933	Multiple	251	0	21	194	659	156	19	18	57	119	789	887	5	3175
Nonbio	0.7822	0.3768	Nonbio	114	111	14	781	1205	118	147	131	175	279	232	12038	45	15390
Pteropoda	0.2682	0.2322	Pteropoda	4	0	0	0	375	44	0	3	0	0	5	363	291	1085
			Sum	2818	834	1613	3937	8541	1986	1188	1245	1250	2635	1557	19317	379	47300

图 20: 选取 16 个特征采用 SVM Linear 进行分类的结果

3.3.2 选取 16 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 16 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc），并采用随机森林进行训练和分类得到的结果如图 21，其分类准确率为 72.76%。

Error rate			0.2404														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8018	0.2295	Appendicularia	2185	0	111	14	1	12	0	0	42	4	188	167	1	2725
Bubble	0.8097	0.1186	Bubble	0	587	0	0	2	0	0	37	0	0	4	95	0	725
Chaetognatha	0.8562	0.0890	Chaetognatha	161	0	1494	0	5	3	0	0	23	0	30	29	0	1745
CladoceraPenilia	0.8603	0.1373	CladoceraPenilia	0	0	0	3979	7	0	16	0	0	20	22	581	0	4625
Copepoda	0.8303	0.2521	Copepoda	4	0	0	7	6119	186	0	1	0	5	242	760	46	7370
Decapoda	0.7382	0.2135	Decapoda	6	0	0	1	462	2056	0	1	0	0	102	149	8	2785
Doliolida	0.7495	0.1772	Doliolida	0	0	0	7	0	0	1068	0	0	169	11	170	0	1425
Egg	0.7349	0.1684	Egg	0	60	0	8	2	5	1	1264	0	45	16	308	11	1720
Fiber	0.6868	0.2457	Fiber	50	0	18	0	1	0	0	0	1044	0	75	330	2	1520
Gelatinous	0.6741	0.2329	Gelatinous	6	0	0	78	15	3	122	52	1	2029	102	601	1	3010
Multiple	0.3030	0.5603	Multiple	291	0	6	110	525	181	14	25	49	103	962	898	11	3175
Nonbio	0.8116	0.2572	Nonbio	132	19	11	408	884	145	77	140	225	267	426	12490	166	15390
Pteropoda	0.6028	0.2733	Pteropoda	1	0	0	0	159	23	0	0	0	3	8	237	654	1085
			Sum	2836	666	1640	4612	8182	2614	1298	1520	1384	2645	2188	16815	900	47300

图 21: 选取 16 个特征采用随机森林进行分类的结果

3.3.3 选取 18 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 18 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev），并采用 SVM Linear 进行训练和分类得到的结果如图 22。

3.3.4 选取 18 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 18 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev），并采用随机森林进行训练和分类得到的结果如图 23，其分类准确率为 74.59%。

3.3.5 选取 19 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 19 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、

Error rate			0.3111														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8213	0.2030	Appendicularia	2238	0	96	4	6	8	0	0	43	5	127	194	4	2725
Bubble	0.8041	0.3244	Bubble	0	583	0	0	26	3	0	11	0	0	5	97	0	725
Chaetognatha	0.8877	0.0899	Chaetognatha	139	0	1549	0	3	0	0	0	16	0	13	25	0	1745
CladoceraPenilia	0.6986	0.2502	CladoceraPenilia	0	0	0	3231	33	0	5	0	0	65	0	1291	0	4625
Copepoda	0.7228	0.3449	Copepoda	0	0	0	15	5327	168	0	0	0	2	175	1661	22	7370
Decapoda	0.5436	0.2885	Decapoda	5	0	0	0	707	1514	0	0	0	0	91	466	2	2785
Doliolida	0.5811	0.3106	Doliolida	0	0	0	8	0	0	828	0	0	271	0	318	0	1425
Egg	0.5773	0.1894	Egg	0	168	0	27	48	1	0	993	0	70	0	403	10	1720
Fiber	0.6967	0.2392	Fiber	54	0	29	2	5	1	0	2	1059	3	41	324	0	1520
Gelatinous	0.6355	0.2875	Gelatinous	7	0	0	158	50	6	202	84	2	1913	60	526	2	3010
Multiple	0.2583	0.4699	Multiple	248	0	12	129	585	185	14	18	66	90	820	1003	5	3175
Nonbio	0.7938	0.3537	Nonbio	113	112	16	735	1048	162	152	110	206	266	205	12216	49	15390
Pteropoda	0.2894	0.2304	Pteropoda	4	0	0	0	294	80	0	7	0	0	10	376	314	1085
			Sum	2808	863	1702	4309	8132	2128	1201	1225	1392	2685	1547	18900	408	47300

图 22: 选取 18 个特征采用 SVM Linear 进行分类的结果

Error rate			0.2270														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8143	0.2189	Appendicularia	2219	0	110	3	4	11	0	0	46	5	186	141	0	2725
Bubble	0.8138	0.1246	Bubble	0	590	0	0	4	0	0	36	0	0	3	92	0	725
Chaetognatha	0.8699	0.0894	Chaetognatha	172	0	1518	0	5	0	0	0	7	1	24	18	0	1745
CladoceraPenilia	0.8720	0.1189	CladoceraPenilia	0	0	0	4033	11	0	11	0	0	24	16	530	0	4625
Copepoda	0.8347	0.2228	Copepoda	2	0	0	9	6152	171	0	0	0	3	229	779	25	7370
Decapoda	0.7709	0.2136	Decapoda	2	0	0	2	356	2147	0	1	0	1	107	162	7	2785
Doliolida	0.7347	0.1704	Doliolida	0	0	0	4	0	0	1047	0	0	179	4	191	0	1425
Egg	0.7297	0.1572	Egg	0	58	0	12	3	3	2	1255	0	43	14	317	13	1720
Fiber	0.7342	0.2398	Fiber	50	0	10	0	5	0	1	0	1116	1	60	277	0	1520
Gelatinous	0.6867	0.2296	Gelatinous	10	0	0	81	20	3	105	54	0	2067	74	594	2	3010
Multiple	0.3187	0.5188	Multiple	257	1	16	62	495	194	11	20	60	82	1012	953	12	3175
Nonbio	0.8221	0.2528	Nonbio	129	25	11	371	806	167	85	122	238	273	366	12652	145	15390
Pteropoda	0.6949	0.2129	Pteropoda	0	0	2	0	55	34	0	1	1	4	8	226	754	1085
			Sum	2841	674	1667	4577	7916	2730	1262	1489	1468	2683	2103	16932	958	47300

图 23: 选取 18 个特征采用随机森林进行分类的结果

PerimFeret、PerimMaj、CDexc、Mean、stdDev、Mode)，并采用 SVM Linear 进行训练和分类得到的结果如图 24。

3.3.6 选取 19 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 19 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Mean、stdDev、Mode），并采用随机森林进行训练和分类得到的结果如图 25，其分类准确率为 73.04%。

Error rate			0.3290															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.8070	0.2210	Appendicularia	2199	0	107	27	1	8	0	0	26	6	140	211	0	2725	
Bubble	0.7448	0.3366	Bubble	0	540	0	6	20	1	0	52	0	0	4	102	0	725	
Chaetognatha	0.8332	0.0975	Chaetognatha	193	0	1454	3	6	0	0	0	11	5	37	36	0	1745	
CladoceraPenilia	0.6067	0.2967	CladoceraPenilia	0	0	0	2806	16	0	3	0	0	57	2	1741	0	4625	
Copepoda	0.7288	0.3900	Copepoda	0	0	0	7	5371	262	0	0	3	5	177	1452	93	7370	
Decapoda	0.5436	0.2932	Decapoda	4	0	0	0	983	1514	0	0	0	0	73	205	6	2785	
Doliolida	0.5881	0.2922	Doliolida	0	0	0	4	0	0	838	0	0	253	1	329	0	1425	
Egg	0.5698	0.2379	Egg	0	172	0	31	49	3	1	980	0	65	1	409	9	1720	
Fiber	0.6520	0.2122	Fiber	56	0	23	9	18	0	0	2	991	2	40	379	0	1520	
Gelatinous	0.6116	0.3008	Gelatinous	6	0	0	121	37	12	195	103	0	1841	81	611	3	3010	
Multiple	0.2683	0.4867	Multiple	250	0	14	186	710	164	18	19	51	116	852	789	6	3175	
Nonbio	0.7760	0.3539	Nonbio	112	102	13	790	1246	140	129	125	176	283	247	11943	84	15390	
Pteropoda	0.3770	0.3295	Pteropoda	3	0	0	0	348	38	0	5	0	0	5	277	409	1085	
			Sum	2823	814	1611	3990	8805	2142	1184	1286	1258	2633	1660	18484	610	47300	

图 24: 选取 19 个特征采用 SVM Linear 进行分类的结果

Error rate			0.2384															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.7908	0.2315	Appendicularia	2155	0	125	15	4	12	0	0	54	8	186	166	0	2725	
Bubble	0.8234	0.0982	Bubble	0	597	0	0	1	0	0	27	0	0	3	97	0	725	
Chaetognatha	0.8619	0.0967	Chaetognatha	163	0	1504	0	5	1	0	0	19	0	32	20	1	1745	
CladoceraPenilia	0.8605	0.1329	CladoceraPenilia	0	0	0	3980	16	0	22	0	0	19	23	565	0	4625	
Copepoda	0.8259	0.2459	Copepoda	1	0	0	5	6087	197	0	1	0	7	263	765	44	7370	
Decapoda	0.7594	0.2334	Decapoda	2	0	0	0	424	2115	0	1	0	0	89	146	8	2785	
Doliolida	0.7544	0.1831	Doliolida	0	0	0	4	0	0	1075	0	0	154	15	177	0	1425	
Egg	0.7320	0.1476	Egg	0	44	0	6	3	10	1	1259	0	47	17	324	9	1720	
Fiber	0.6921	0.2448	Fiber	44	0	17	1	4	0	1	0	1052	0	67	333	1	1520	
Gelatinous	0.6824	0.2313	Gelatinous	11	0	0	74	22	2	108	42	0	2054	106	591	0	3010	
Multiple	0.3017	0.5606	Multiple	295	1	9	93	521	205	18	22	48	104	958	893	8	3175	
Nonbio	0.8149	0.2559	Nonbio	133	20	9	412	823	189	91	125	219	276	413	12541	139	15390	
Pteropoda	0.5954	0.2453	Pteropoda	0	0	1	0	162	28	0	0	1	3	8	236	646	1085	
			Sum	2804	662	1665	4590	8072	2759	1316	1477	1393	2672	2180	16854	856	47300	

图 25: 选取 19 个特征采用随机森林进行分类的结果

3.3.7 选取 21 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 21 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode），并采用 SVM Linear 进行训练和分类得到的结果如图 26。

3.3.8 选取 21 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 21 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、

Error rate			0.3008															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.8253	0.2028	Appendicularia	2249	0	101	7	8	11	0	0	48	4	139	156	2	2725	
Bubble	0.7766	0.3249	Bubble	0	563	0	3	17	2	0	48	0	0	6	86	0	725	
Chaetognatha	0.8865	0.0916	Chaetognatha	141	0	1547	0	8	0	0	0	15	0	17	17	0	1745	
CladoceraPenilia	0.7053	0.2503	CladoceraPenilia	0	0	0	3262	21	0	7	0	0	69	1	1265	0	4625	
Copepoda	0.7419	0.3313	Copepoda	0	0	0	10	5468	180	0	0	0	2	195	1418	97	7370	
Decapoda	0.5662	0.2871	Decapoda	4	0	0	0	711	1577	0	0	0	0	100	385	8	2785	
Doliolida	0.5951	0.2886	Doliolida	0	0	0	7	0	0	848	0	0	258	1	311	0	1425	
Egg	0.5791	0.2157	Egg	0	170	0	36	47	2	0	996	0	79	1	378	11	1720	
Fiber	0.7079	0.2380	Fiber	57	0	25	1	4	0	0	2	1076	1	43	311	0	1520	
Gelatinous	0.6565	0.2809	Gelatinous	6	0	0	150	27	5	187	86	1	1976	60	511	1	3010	
Multiple	0.2781	0.4744	Multiple	247	0	13	135	596	185	16	20	62	89	883	922	7	3175	
Nonbio	0.7928	0.3316	Nonbio	114	101	16	740	996	181	134	110	210	270	223	12201	94	15390	
Pteropoda	0.3926	0.3406	Pteropoda	3	0	1	0	274	69	0	8	0	0	11	293	426	1085	
			Sum	2821	834	1703	4351	8177	2212	1192	1270	1412	2748	1680	18254	646	47300	

图 26: 选取 21 个特征采用 SVM Linear 进行分类的结果

PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode), 并采用随机森林进行训练和分类得到的结果如图 27, 其分类准确率为 74.91%。

Error rate			0.2240														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7949	0.2256	Appendicularia	2166	0	148	3	0	10	0	0	50	2	191	155	0	2725
Bubble	0.8124	0.0952	Bubble	0	589	0	0	6	0	0	31	0	0	1	98	0	725
Chaetognatha	0.8665	0.1153	Chaetognatha	172	0	1512	0	5	1	0	0	9	0	22	24	0	1745
CladoceraPenilia	0.8765	0.1225	CladoceraPenilia	0	0	0	4054	10	0	15	0	0	19	22	505	0	4625
Copepoda	0.8452	0.2216	Copepoda	4	0	0	10	6229	167	0	0	0	4	194	737	25	7370
Decapoda	0.7724	0.2089	Decapoda	5	0	0	0	349	2151	0	0	0	0	92	180	8	2785
Doliolida	0.7600	0.1752	Doliolida	0	0	0	5	0	0	1083	0	0	153	11	173	0	1425
Egg	0.7448	0.1262	Egg	0	46	0	8	5	1	2	1281	0	44	15	307	11	1720
Fiber	0.7401	0.2388	Fiber	50	0	23	0	1	0	2	1	1125	0	46	272	0	1520
Gelatinous	0.7013	0.2170	Gelatinous	11	0	0	75	20	0	104	25	0	2111	83	577	4	3010
Multiple	0.3068	0.5233	Multiple	264	0	14	72	526	180	15	20	69	88	974	948	5	3175
Nonbio	0.8237	0.2489	Nonbio	125	16	10	393	792	173	92	108	225	271	385	12676	124	15390
Pteropoda	0.6940	0.1903	Pteropoda	0	0	2	0	59	36	0	0	0	4	7	224	753	1085
			Sum	2797	651	1709	4620	8002	2719	1313	1466	1478	2696	2043	16876	930	47300

图 27: 选取 21 个特征采用随机森林进行分类的结果

3.3.9 选取 25 个特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中根据自己的判断选取部分特征进行 (选取 25 个特征: Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR), 并采用 SVM Linear 进行训练和分类得到的结果如图 28。

Error rate			0.2834														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8283	0.1957	Appendicularia	2257	0	104	3	11	10	0	0	40	5	121	174	0	2725
Bubble	0.8041	0.3125	Bubble	0	583	0	0	24	0	0	46	0	0	4	68	0	725
Chaetognatha	0.8860	0.0959	Chaetognatha	148	0	1546	0	0	4	0	0	12	0	18	17	0	1745
CladoceraPenilia	0.7332	0.2300	CladoceraPenilia	0	0	0	3391	27	0	9	0	0	61	1	1136	0	4625
Copepoda	0.7859	0.3068	Copepoda	0	0	0	17	5792	162	0	0	2	3	171	1123	100	7370
Decapoda	0.6083	0.2720	Decapoda	2	0	0	0	624	1694	0	0	0	0	94	361	10	2785
Doliolida	0.6232	0.2839	Doliolida	0	0	0	7	0	0	888	0	0	258	1	271	0	1425
Egg	0.6006	0.1749	Egg	0	183	0	33	41	1	0	1033	0	68	0	351	10	1720
Fiber	0.7329	0.2215	Fiber	50	0	34	1	0	0	0	2	1114	0	38	281	0	1520
Gelatinous	0.6591	0.2799	Gelatinous	5	0	0	142	44	6	186	46	4	1984	63	529	1	3010
Multiple	0.2693	0.4571	Multiple	239	0	11	120	583	183	12	13	65	93	855	992	9	3175
Nonbio	0.7991	0.3131	Nonbio	101	82	14	690	977	197	145	103	194	283	204	12298	102	15390
Pteropoda	0.4258	0.3343	Pteropoda	4	0	1	0	232	70	0	9	0	0	5	302	462	1085
			Sum	2806	848	1710	4404	8355	2327	1240	1252	1431	2755	1575	17903	694	47300

图 28: 选取 25 个特征采用 SVM Linear 进行分类的结果

3.3.10 选取 25 个特征 +RandomForest

在 PkID 软件中, 从 67 个特征中根据自己的判断选取部分特征进行 (选取 25 个特征: Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR), 并采用随机森林进行训练和分类得到的结果如图 29, 其分类准确率为 76.5%。

Error rate			0.2133														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8015	0.2138	Appendicularia	2184	0	126	3	1	13	0	0	57	6	189	144	2	2725
Bubble	0.9090	0.0948	Bubble	0	659	0	0	1	0	0	19	0	0	2	44	0	725
Chaetognatha	0.8716	0.0957	Chaetognatha	169	0	1521	0	0	2	0	0	17	0	19	17	0	1745
CladoceraPenilia	0.8804	0.1125	CladoceraPenilia	0	0	0	4072	20	0	19	0	0	20	23	471	0	4625
Copepoda	0.8495	0.2066	Copepoda	1	0	0	8	6261	131	0	0	2	2	212	738	15	7370
Decapoda	0.7835	0.2068	Decapoda	6	0	0	0	327	2182	0	0	0	0	96	170	4	2785
Doliolida	0.7593	0.1467	Doliolida	0	0	0	5	0	0	1082	0	0	156	13	169	0	1425
Egg	0.7878	0.1038	Egg	0	62	0	10	0	0	3	1355	0	40	1	248	1	1720
Fiber	0.7533	0.2437	Fiber	51	0	17	0	1	0	0	0	1145	0	59	247	0	1520
Gelatinous	0.7080	0.2254	Gelatinous	5	0	0	72	25	1	76	39	2	2131	99	557	3	3010
Multiple	0.3156	0.5278	Multiple	257	3	3	70	489	214	15	16	73	101	1002	923	9	3175
Nonbio	0.8361	0.2354	Nonbio	105	4	10	348	722	170	73	83	217	292	395	12868	103	15390
Pteropoda	0.6894	0.1548	Pteropoda	0	0	5	0	44	38	0	0	1	3	12	234	748	1085
			Sum	2778	728	1682	4588	7891	2751	1268	1512	1514	2751	2122	16830	885	47300

图 29: 选取 25 个特征采用随机森林进行分类的结果

3.3.11 选取 27 个特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中根据自己的判断选取部分特征进行 (选取 27 个特征: Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、

PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR、Convperim、Convarea)，并采用 SVM Linear 进行训练和分类得到的结果如图 30。

Error rate			0.2819														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8341	0.1937	Appendicularia	2273	0	94	3	11	9	0	0	37	5	122	171	0	2725
Bubble	0.8041	0.3084	Bubble	0	583	0	0	24	0	0	46	0	0	3	69	0	725
Chaetognatha	0.8888	0.0909	Chaetognatha	143	0	1551	0	0	8	0	0	15	0	16	12	0	1745
CladoceraPenilia	0.7325	0.2305	CladoceraPenilia	0	0	0	3388	29	0	10	0	0	60	1	1137	0	4625
Copepoda	0.7855	0.3078	Copepoda	0	0	0	17	5789	155	0	0	3	2	171	1136	97	7370
Decapoda	0.6144	0.2676	Decapoda	2	0	0	0	611	1711	0	0	0	0	94	357	10	2785
Doliolida	0.6344	0.2745	Doliolida	0	0	0	7	0	0	904	0	0	242	0	272	0	1425
Egg	0.6023	0.1738	Egg	0	180	0	34	41	1	0	1036	0	67	0	352	9	1720
Fiber	0.7322	0.2184	Fiber	60	0	26	1	0	0	0	2	1113	0	40	278	0	1520
Gelatinous	0.6625	0.2728	Gelatinous	5	0	1	139	37	7	183	45	4	1994	58	532	5	3010
Multiple	0.2731	0.4530	Multiple	233	0	17	122	584	189	11	13	61	90	867	978	10	3175
Nonbio	0.7992	0.3126	Nonbio	98	80	16	692	998	186	138	103	191	282	208	12299	99	15390
Pteropoda	0.4221	0.3343	Pteropoda	5	0	1	0	239	70	0	9	0	0	5	298	458	1085
			Sum	2819	843	1706	4403	8363	2336	1246	1254	1424	2742	1585	17891	688	47300

图 30: 选取 27 个特征采用 SVM Linear 进行分类的结果

3.3.12 选取 27 个特征 +RandomForest

在 PKID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 27 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR、Convperim、Convarea），并采用随机森林进行训练和分类得到的结果如图 31，其分类准确率为 76.3%。

Error rate			0.2123														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8040	0.2200	Appendicularia	2191	0	122	0	0	19	0	0	67	0	178	147	1	2725
Bubble	0.8690	0.0735	Bubble	0	630	0	0	5	0	0	30	0	0	3	57	0	725
Chaetognatha	0.8831	0.1015	Chaetognatha	148	0	1541	0	2	4	0	0	10	1	27	12	0	1745
CladoceraPenilia	0.8819	0.1140	CladoceraPenilia	0	0	0	4079	12	0	22	0	0	16	20	476	0	4625
Copepoda	0.8613	0.2072	Copepoda	3	0	0	12	6348	124	0	0	0	5	211	647	20	7370
Decapoda	0.7903	0.2048	Decapoda	5	0	0	0	323	2201	0	0	0	0	99	154	3	2785
Doliolida	0.7551	0.1567	Doliolida	0	0	0	5	0	0	1076	0	0	154	9	181	0	1425
Egg	0.7942	0.1025	Egg	0	47	0	9	2	0	1	1366	0	42	2	247	4	1720
Fiber	0.7329	0.2359	Fiber	58	0	24	0	0	0	1	0	1114	0	57	266	0	1520
Gelatinous	0.7050	0.2230	Gelatinous	10	0	0	80	24	1	92	23	0	2122	88	567	3	3010
Multiple	0.3165	0.5230	Multiple	287	1	15	67	499	216	9	10	60	104	1005	896	6	3175
Nonbio	0.8340	0.2320	Nonbio	107	2	12	352	738	164	75	93	207	283	400	12836	121	15390
Pteropoda	0.6922	0.1738	Pteropoda	0	0	1	0	54	39	0	0	0	4	8	228	751	1085
			Sum	2809	680	1715	4604	8007	2768	1276	1522	1458	2731	2107	16714	909	47300

图 31: 选取 27 个特征采用随机森林进行分类的结果

3.3.13 选取 29 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 29 个特征：Mean, StdDev, Fractal, Skelarea, Circ, Kurt, Skew, Symetrieih, Symetriev, Symetrieihc, Symetrieivc, Convperim, Convarea, Elongation, MeanPos, CV, SR, PermAreaexc, FeretAreaexc, PerimFeret, Circexc, CDexc, PerimMaj, ESD, ThickR, Fcons, Nb1, Nb2, Nb3），并采用随机森林进行训练和分类得到的结果如图 32，其分类准确率为 76.3%。

Error rate			0.2065														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8158	0.2027	Appendicularia	2223	0	118	9	0	16	0	0	39	2	174	144	0	2725
Bubble	0.8993	0.0856	Bubble	0	652	0	0	2	0	0	24	0	0	1	46	0	725
Chaetognatha	0.8848	0.0960	Chaetognatha	142	0	1544	0	0	5	0	0	20	0	18	16	0	1745
CladoceraPenilia	0.8902	0.0920	CladoceraPenilia	0	0	0	4117	22	0	3	0	0	11	14	458	0	4625
Copepoda	0.8516	0.2045	Copepoda	3	0	0	8	6276	133	0	0	3	3	223	704	17	7370
Decapoda	0.7849	0.2097	Decapoda	4	0	0	0	325	2186	0	0	0	0	109	156	5	2785
Doliolida	0.7832	0.1129	Doliolida	0	0	0	0	0	0	1116	0	0	159	5	145	0	1425
Egg	0.7791	0.1037	Egg	0	56	0	9	0	0	0	1340	0	25	1	286	3	1720
Fiber	0.7441	0.2290	Fiber	59	0	22	0	0	0	0	0	1131	1	44	263	0	1520
Gelatinous	0.7256	0.2093	Gelatinous	8	0	2	49	18	1	71	26	0	2184	103	548	0	3010
Multiple	0.3169	0.5182	Multiple	255	4	6	54	484	219	9	17	64	104	1006	948	5	3175
Nonbio	0.8464	0.2327	Nonbio	92	1	11	288	723	155	59	88	210	270	375	13026	92	15390
Pteropoda	0.6756	0.1427	Pteropoda	2	0	5	0	39	51	0	0	0	3	15	237	733	1085
			Sum	2788	713	1708	4534	7889	2766	1258	1495	1467	2762	2088	16977	855	47300

图 32: 选取 29 个特征采用随机森林进行分类的结果

3.4 实验分析

- 位置特征、尺度特征和灰度特征对分类结果没有太大影响。
- 使用 PkID 中的特征能达到的最高分类识别率大概在 76% 左右。
- 我们自己的 SVM 和随机森林分类结果要比 PkID 中的分类结果高 1% 左右。
- 随机森林的分类准确率要比 SVM 要高。

根据以上实验，判断可以从 PkID 的 67 个特征中选取一下特征：Mean, StdDev, Fractal, Skelarea, Circ, Kurt, Skew, Symetrieih, Symetriev, Elongation, MeanPos, CV, SR, PermAreaexc, FeretAreaexc, PerimFeret, Circexc, CDexc, [Symetrieihc](#), [Symetrieivc](#), [Convperim](#), [Convarea](#), (PerimMaj, ESD, [ThickR](#), [Fcons](#), Nb1, Nb2, Nb3)。

前 22 个特征，在 PkID 软件采用 SVM Linear 进行训练和分类得到的结果如图 33，其分类准确率为 66.07%。

Error rate			0.2947														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8316	0.1973	Appendicularia	2266	0	99	0	5	13	0	0	47	7	103	181	4	2725
Bubble	0.8483	0.2832	Bubble	0	615	0	2	25	0	0	7	0	0	5	71	0	725
Chaetognatha	0.8928	0.0968	Chaetognatha	133	0	1558	0	1	5	0	0	18	0	9	21	0	1745
CladoceraPenilia	0.6737	0.2847	CladoceraPenilia	0	0	0	3116	36	0	10	0	0	72	0	1391	0	4625
Copepoda	0.7902	0.3163	Copepoda	0	0	0	16	5824	139	0	0	1	2	172	1200	16	7370
Decapoda	0.5921	0.2671	Decapoda	3	0	0	0	656	1649	0	0	0	0	84	391	2	2785
Doliolida	0.6140	0.2792	Doliolida	0	0	0	13	0	0	875	0	0	255	1	281	0	1425
Egg	0.6110	0.1665	Egg	0	168	0	31	38	2	0	1051	0	69	0	351	10	1720
Fiber	0.7039	0.2229	Fiber	61	0	31	1	0	1	0	3	1070	1	33	319	0	1520
Gelatinous	0.6385	0.2871	Gelatinous	5	0	0	201	40	5	180	41	5	1922	56	554	1	3010
Multiple	0.2699	0.4351	Multiple	241	0	21	139	601	187	12	9	52	90	857	961	5	3175
Nonbio	0.7926	0.3319	Nonbio	109	75	15	837	1009	179	137	126	184	278	192	12198	51	15390
Pteropoda	0.3309	0.1987	Pteropoda	5	0	1	0	283	70	0	24	0	0	5	338	359	1085
			Sum	2823	858	1725	4356	8518	2250	1214	1261	1377	2696	1517	18257	448	47300

图 33: 选取 22 个特征采用 SVM Linear 进行分类的结果

前 22 个特征，用自己设计的 SVM Linear 进行训练和分类得到的结果如图 34。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2309	0	88	4	6	14	0	0	49	6	89	157	3	2725	0.847339	0.199653
Bubble	0	633	0	3	24	1	0	14	0	0	6	44	0	725	0.873103	0.330867
Chaetognatha	134	0	1565	0	0	7	0	0	10	0	10	19	0	1745	0.896848	0.096942
CladoceraPenilia	0	0	0	3528	37	0	10	1	0	67	0	982	0	4625	0.762811	0.271074
Copepoda	0	0	0	22	6062	131	0	0	3	0	153	985	14	7370	0.822524	0.328236
Decapoda	3	0	0	0	706	1677	0	0	0	0	63	336	0	2785	0.602154	0.268005
Doliolida	0	0	0	28	0	0	966	0	0	209	0	222	0	1425	0.677895	0.253478
Egg	0	203	0	40	134	0	1	1013	0	58	0	260	11	1720	0.588953	0.171031
Fiber	53	0	38	2	0	1	0	5	1078	0	40	303	0	1520	0.709211	0.218274
Gelatinous	5	0	0	209	39	11	154	45	5	1951	56	535	0	3010	0.648173	0.269288
Multiple	273	0	25	164	603	184	15	14	49	100	826	917	5	3175	0.260157	0.42439
Nonbio	103	110	17	840	1044	174	148	108	185	279	187	12143	52	15390	0.789019	0.291334
Pteropoda	5	0	0	0	369	91	0	22	0	0	5	232	361	1085	0.332719	0.190583
Total	2885	946	1733	4840	9024	2291	1294	1222	1379	2670	1435	17135	446	47300	0.677762	0.254858

图 34: 选取 22 个特征采用 SVM Linear 进行分类的结果

前 22 个特征，在 PkID 软件采用随机森林进行训练和分类得到的结果如图 35，其分类准确率为 76.66%。

前 22 个特征，用自己设计的随机森林进行训练和分类得到的结果如图 35。

Error rate			0.2116														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8136	0.2042	Appendicularia	2217	0	119	2	0	16	0	0	46	2	179	143	1	2725
Bubble	0.8952	0.0846	Bubble	0	649	0	0	3	0	0	29	0	0	1	43	0	725
Chaetognatha	0.8831	0.0957	Chaetognatha	137	0	1541	0	0	5	0	0	24	0	19	19	0	1745
CladoceraPenilia	0.8709	0.1188	CladoceraPenilia	0	0	0	4028	17	0	19	0	0	22	23	516	0	4625
Copepoda	0.8498	0.2076	Copepoda	2	0	0	10	6263	130	0	0	3	5	234	712	11	7370
Decapoda	0.7752	0.2057	Decapoda	8	0	0	0	332	2159	0	0	0	0	110	172	4	2785
Doliolida	0.7656	0.1490	Doliolida	0	0	0	7	0	0	1091	0	0	148	8	171	0	1425
Egg	0.7965	0.1087	Egg	0	54	0	10	0	0	4	1370	0	32	6	242	2	1720
Fiber	0.7500	0.2375	Fiber	63	0	14	0	1	0	0	0	1140	0	54	248	0	1520
Gelatinous	0.6977	0.2111	Gelatinous	8	0	0	84	25	4	91	30	2	2100	101	562	3	3010
Multiple	0.3294	0.5263	Multiple	253	1	12	79	504	210	6	21	72	99	1046	865	7	3175
Nonbio	0.8400	0.2321	Nonbio	96	5	12	351	712	153	71	87	207	253	413	12927	103	15390
Pteropoda	0.6986	0.1474	Pteropoda	2	0	6	0	47	41	0	0	1	1	14	215	758	1085
			Sum	2786	709	1704	4571	7904	2718	1282	1537	1495	2662	2208	16835	889	47300

图 35: 选取 22 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2276	0	97	4	0	15	0	0	53	2	134	144	0	2725	0.835229	0.198592
Bubble	0	632	0	0	3	0	0	30	0	0	1	59	0	725	0.871724	0.063704
Chaetognatha	143	0	1549	0	1	5	0	0	12	0	22	12	1	1745	0.887679	0.080712
CladoceraPenilia	0	0	0	4051	15	0	26	0	0	12	17	504	0	4625	0.875892	0.106922
Copepoda	0	0	0	10	6334	137	0	0	0	4	208	660	17	7370	0.85943	0.189404
Decapoda	2	0	0	0	309	2216	0	0	0	0	92	163	3	2785	0.795691	0.208571
Doliolida	0	0	0	7	0	0	1092	0	0	149	3	174	0	1425	0.766316	0.137441
Egg	0	41	0	8	1	0	0	1414	0	33	1	218	4	1720	0.822093	0.097639
Fiber	53	0	12	0	0	0	1	0	1130	0	57	267	0	1520	0.743421	0.228669
Gelatinous	5	0	0	81	28	3	77	22	3	2064	86	638	3	3010	0.685714	0.208589
Multiple	274	1	12	72	458	222	12	11	61	94	1022	930	6	3175	0.32189	0.499265
Nonbio	87	1	11	303	629	163	58	90	206	249	387	13093	113	15390	0.850747	0.233745
Pteropoda	0	0	4	0	36	39	0	0	0	1	11	225	769	1085	0.708756	0.16048
Total	2840	675	1685	4536	7814	2800	1266	1567	1465	2608	2041	17087	916	47300	0.771122	0.185672

图 36: 选取 22 个特征采用随机森林进行分类的结果

4. 计算机视觉特征提取

4.1 几何参数

4.1.1 边界的周长

轮廓边界的周长。对轮廓边缘上的像素点的统计。

4.1.2 边界的曲率

4.1.3 面积

描述区域大小的特征。对区域内总像素点的统计。

4.1.4 宽度和高度

最小外接矩形的宽度和高度

4.1.5 矩形度

反映被检测目标的最小外接矩形的充满程度，当目标的形状越接近矩形时，矩形度的值越接近 1。

$$R = \frac{A}{WH}$$

A 为目标的面积，W、H 分别为最小外接矩形的宽度和高度。

4.1.6 体态比

为目标最小外接矩形的长与宽的比值。

$$C = \frac{W}{H}$$

4.1.7 圆形性

用目标区域的所有边界点定义的特征向量。

$$C_I = \frac{\mu_R}{\sigma_R}$$

μ_R 为区域重心到边界点的平均距离， σ_R 为从区域重心到边界点的距离的平均方差。

4.1.8 偏心率

在一定程度上反映了区域的紧凑程度。定义为目标区域长短主轴的平方根的比值。

$$E = \frac{p}{q}$$

设目标区域在 XY 平面上，区域像素点绕 X 轴的转动惯量为 A，绕 Y 轴的转动惯量为 B，惯性积为 C。目标区域的长度分别是 p 和 q。

$$p = \sqrt{\frac{2}{(A+B) + \sqrt{(A-B)^2 + 4C^2}}}$$

$$q = \sqrt{\frac{2}{(A+B) - \sqrt{(A-B)^2 + 4C^2}}}$$

4.1.9 凸率

为目标区域面积与目标区域凸包面积之比，该特征包含着描述边界不规则特性的信息。

$$C_R = \frac{A}{\sum_{x=1}^M \sum_{y=1}^N k(x, y)}$$

分母为凸包区域的面积。

4.1.10 密集度

描述目标密集度的量化特征，提供了目标形状的重要信息。在周长确定后，密集度越高，所围成的面积越大。

$$C_2 = \frac{L^2}{4\pi A}$$

L 为周长。

4.1.11 球状性

内切圆的直径与外接圆的直径之比。

$$S = \frac{r_i}{r_c}$$

4.1.12 伸长度

周长与目标区域最小外接矩形面积之比。

$$P = \frac{L}{WH}$$

4.1.13 叶状性

叶状反映了边界的幅度特征，为区域重心到边界的最短距离与目标区域的最大宽度之比。

$$B = \frac{R_1}{W_{max}}$$

4.2 几种典型的特征描述方法

4.2.1 边界描述子

- 链码
- 多边形近似
- 骨架
- 形状数
- 统计矩：边界线段的形状可以通过简单的统计矩进行定量的描述，如均值、方差和高阶矩。
- 傅里叶描述子
- 曲率尺度空间
- 形状上下文 (KNN)

4.2.2 区域描述子

- 拓扑描述：欧拉数
- 不变矩
- 角半径变换 (Angular Radial Transformation, ART)：通过使用一组半径变换系数，描述单个连通区域或者不连通区域，对旋转和噪声具有鲁棒性。

- 纹理
 - 统计方法：灰度共生矩阵
 - 模型法：马尔科夫随机场
 - 频谱方法：Gabor 滤波、小波变换

4.3 特征融合

特征融合分为三个层次：数据级融合、特征级融合和决策级融合。数据级融合是结合未加工的信息来得到更加丰富的信息。特征级融合是选择并结合特征来去除多余和无关的特征。决策级融合是用多个相同或不同的分类器，相同或不同的分类器。

图像融合方法：

像素级：PCA（主成分分析）、HIS 变换、Brovey 变换、线性加权法、SFIM、IHS 变换、高通滤波法、小波变换融合算法。

特征级：聚类分析法、贝叶斯估计法、信息熵法、神经网络法、带权平均法、Dempster-shafer 推理法、表决法及神经网络法。

决策级：神经网络法、贝叶斯融合、模糊聚类法、模糊集理论、可靠性理论以及逻辑模板法。

4.3.1 贝叶斯融合 (Bayes Fusion)

当在进行图像分类过程中，可能需要用到不止一种特征。贝叶斯融合可以通过结合不同分类器的结果实现特征融合。

贝叶斯融合规则：

如果要将一幅图像分类到 n 个可能的种类中 $(\omega_1, \dots, \omega_n)$, x_i 表示第 i 个分类器产生的待识的属性，它属于 n 个模式类之一。记 $P(\omega_k)$ 为先验概率， $P(x_i|\omega_k)$ 为每类的概率密度函数， $P(x_1, \dots, x_R|\omega_k)$ 联合概率分布函数， R 为用来分类的分类器数目。

根据贝叶斯最小错误率理论，如果

$$P(\omega_j|x_1, \dots, x_R) = \max_k P(\omega_k|x_1, \dots, x_R) \quad (1)$$

则 $Z \in \omega_j$ 。

且有：

$$P(\omega_j|x_1, \dots, x_R) = \frac{P(x_1, \dots, x_R|\omega_k)P(\omega_k)}{P(x_1, \dots, x_R)} \quad (2)$$

其中

$$P(x_1, \dots, x_R) = \sum_{j=1}^n P(x_1, \dots, x_R | \omega_j) P(\omega_j) \quad (3)$$

假定分类器度量之间是相互独立的，有：

$$P(x_1, \dots, x_R | \omega_k) = \prod_{i=1}^R P(x_i | \omega_k) \quad (4)$$

将式 (3)(4) 带入式 (2) 有：

$$P(\omega_j | x_1, \dots, x_R) = \frac{P(\omega_j) \prod_{i=1}^R P(x_i | \omega_j)}{\sum_{j=1}^n P(\omega_j) \prod_{i=1}^R P(x_i | \omega_j)} \quad (5)$$

将式 (5) 带入式 (1)，得到：

1. 融合规则 1：乘法规则

$$P(\omega_j) \prod_{i=1}^R P(x_i | \omega_j) = \max_{k=1}^n P(\omega_k) \prod_{i=1}^R P(x_i | \omega_k) \quad (6)$$

转化为后验概率，得到：

$$P^{-(R-1)}(\omega_j) \prod_{i=1}^R P(\omega_j | x_i) = \max_{k=1}^n P^{-(R-1)}(\omega_k) \prod_{i=1}^R P(\omega_k | x_i) \quad (7)$$

这样就可以将该图片归类为 ω_j 。

2. 融合规则 2：加法规则

在乘法规则中，如果假定由分类器输出的后验概率与相应的先验概率之间只有微小的偏差：

$$P(\omega_k | x_i) = P(\omega_k)(1 + \delta_{ki}), \delta_{ki} \ll 1 \quad (8)$$

由此得到：

$$P^{-(R-1)}(\omega_k) \prod_{i=1}^R P(\omega_k | x_i) = P(\omega_k) \prod_{i=1}^R (1 + \delta_{ki}) \quad (9)$$

如果将上式右边的乘积展开忽略二次以上的项，即：

$$P(\omega_k) \prod_{i=1}^R (1 + \delta_{ki}) = P(\omega_k) + P(\omega_k) \sum_{i=1}^R \delta_{ki} \quad (10)$$

再将式 (8)(10) 带入 (7) 可以得到加法规则如下：

$$(1 - R)P(\omega_j) + \sum_{i=1}^n P(\omega_j|x_i) = \max_{k=1}^n [(1 - R)P(\omega_k) + \sum_{i=1}^R P(\omega_k|x_i)] \quad (11)$$

通常我们假定各类的先验概率是相等的，则上式等价于：

$$\sum_{i=1}^R P(\omega_j|x_i) = \max_{k=1}^n \sum_{i=1}^R P(\omega_k|x_i) \quad (12)$$

由于加法规则对估计误差不敏感，因而具有比其他规则更好的性能。

5. 选取特征进行实验

该部分进行的实验：用 Matlab 和 C（以 MATLAB 为主）实现对浮游动物特征的提取（特征包括 PkID 中部分特征以及计算视觉中的一些特征提取方法），并进行分类。在该实验中使用的去噪方法是去掉连通区域小于 50 的噪声。

5.1 参数特征选取实验

5.1.1 实验一

选用特征： Mean、StdDev、CV、SR、MeanPos、Elongation、Circ、Feret、PerimAreaexc、CDexc、Skelarea、FeretAreaexc、PerimFeret。（这些特征是从 PkID 的 67 个特征中选取的）

分类器： 随机森林、SVM

- MATLAB：采用随机森林进行训练和分类得到的结果如图 37，其分类准确率为 61.6%。采用 SVM Linear 进行训练和分类得到的结果如图 38，其分类准确率为 39.9%
- C：采用随机森林进行训练和分类得到的结果如图 39，其分类准确率为 59.7%。采用 SVM Linear 进行训练和分类得到的结果如图 40，其分类准确率为 33.4%

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	1857	0	220	49	20	12	5	0	64	64	124	310	0	2725	0.681468	0.355208
Bubble	0	527	0	4	11	0	0	94	0	2	5	82	0	725	0.726897	0.155449
Chaetognatha	234	0	1279	0	3	0	2	0	134	7	36	50	0	1745	0.732951	0.244982
CladoceraPenilia	15	0	0	3587	16	2	8	5	1	80	46	865	0	4625	0.775568	0.270045
Copepoda	5	0	0	21	5715	379	5	0	2	5	189	984	65	7370	0.775441	0.310865
Decapoda	6	0	0	1	866	1421	0	2	3	4	128	332	22	2785	0.510233	0.394288
Doliolida	0	0	0	7	1	0	971	0	0	204	9	233	0	1425	0.681404	0.354388
Egg	0	71	0	39	0	18	0	1210	0	23	41	314	4	1720	0.703488	0.212752
Fiber	173	0	127	5	13	19	7	0	779	17	98	282	0	1520	0.5125	0.386614
Gelatinous	42	1	2	237	18	0	284	20	5	1122	137	1139	3	3010	0.372757	0.509615
Multiple	227	0	30	168	463	227	56	43	79	238	765	866	13	3175	0.240945	0.608896
Nonbio	321	25	36	796	987	229	166	162	203	522	365	11396	182	15390	0.740481	0.333801
Pteropoda	0	0	0	0	180	39	0	1	0	0	13	253	599	1085	0.552074	0.32545
Total	2880	624	1694	4914	8293	2346	1504	1537	1270	2288	1956	17106	888	47300	0.615862	0.343258

图 37: Matlab-13 个特征采用随机森林进行分类的结果

5.1.2 实验二

选用特征： Mean、StdDev、CV、SR、MeanPos、Elongation、Circ、Feret、PerimAreaexc、CDexc、Skelarea、FeretAreaexc、PerimFeret、矩形度、体态比、凸率、伸长度、灰度共生矩阵（对比度）、对称性（左右），共 19 个特征。（前 13 个特征为 5.1.1 实验一所使用的特征）

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	1891	0	145	104	36	3	7	0	61	51	62	365	0	2725	0.693945	0.362441
Bubble	0	254	0	19	33	0	0	260	0	0	2	157	0	725	0.350345	0.498024
Chaetognatha	273	0	1227	0	11	2	2	0	113	7	10	100	0	1745	0.703152	0.245387
CladoceraPenilia	1	0	0	3312	29	0	1	0	0	8	17	1257	0	4625	0.716108	0.407725
Copepoda	3	0	0	35	5465	10	0	1	2	0	88	1762	4	7370	0.74152	0.52116
Decapoda	7	0	0	0	2341	155	0	0	0	0	42	240	0	2785	0.055655	0.448399
Doliolida	2	0	0	50	0	0	359	0	0	447	0	567	0	1425	0.25193	0.544416
Egg	0	157	0	128	217	0	3	406	0	2	18	788	1	1720	0.236047	0.533869
Fiber	189	0	197	10	43	10	3	0	558	14	36	460	0	1520	0.367105	0.403209
Gelatinous	20	0	0	663	55	0	204	3	0	487	177	1411	0	3010	0.161794	0.674683
Multiple	274	0	17	340	929	21	35	0	48	208	373	929	1	3175	0.11748	0.605291
Nonbio	306	95	40	941	1722	21	174	199	153	273	108	11332	26	15390	0.736322	0.427503
Pteropoda	0	0	0	0	532	59	0	2	0	0	12	426	54	1085	0.04977	0.372093
Total	2966	506	1626	5592	11413	281	788	871	935	1497	945	19794	86	47300	0.398552	0.464938

图 38: Matlab-13 个特征采用 SVM Linear 进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	1800	0	263	72	26	11	20	0	68	20	171	274	0	2725	0.66055	0.3998
Bubble	0	532	0	2	12	0	0	82	0	4	15	78	0	725	0.733793	0.173913
Chaetognatha	321	0	1129	0	4	1	22	0	127	2	36	103	0	1745	0.646991	0.307787
CladoceraPenilia	40	0	1	3449	12	1	28	1	1	65	65	962	0	4625	0.74573	0.289452
Copepoda	26	1	0	27	5480	495	5	0	4	1	175	1099	57	7370	0.743555	0.324041
Decapoda	6	0	0	0	806	1424	0	13	6	0	85	421	24	2785	0.511311	0.424414
Doliolida	22	0	4	10	3	0	840	0	28	234	26	258	0	1425	0.589474	0.389535
Egg	0	77	0	13	3	16	0	1173	0	60	22	347	9	1720	0.681977	0.226253
Fiber	171	0	141	2	21	19	9	3	646	7	85	416	0	1520	0.425	0.439722
Gelatinous	59	0	0	167	22	2	206	48	3	1529	97	874	3	3010	0.507973	0.408053
Multiple	262	4	42	220	456	199	65	34	45	112	717	1006	13	3175	0.225827	0.627145
Nonbio	292	30	51	892	1127	261	181	157	225	549	414	11069	142	15390	0.719233	0.355629
Pteropoda	0	0	0	0	135	45	0	5	0	0	15	271	614	1085	0.565899	0.287703
Total	2999	644	1631	4854	8107	2474	1376	1516	1153	2583	1923	17178	862	47300	0.596716	0.357957

图 39: C-13 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	1438	0	141	170	65	0	60	0	2	15	29	805	0	2725	0.527706	0.579163
Bubble	0	58	0	11	55	0	0	580	0	0	0	21	0	725	0.08	0.546875
Chaetognatha	474	0	668	2	21	0	22	0	3	0	7	548	0	1745	0.382808	0.37277
CladoceraPenilia	52	0	0	2900	17	0	6	0	0	74	6	1570	0	4625	0.627027	0.399462
Copepoda	9	0	0	24	5271	0	2	9	0	0	13	2035	7	7370	0.715197	0.55635
Decapoda	8	0	0	0	2414	0	0	0	0	0	6	357	0	2785	0	0
Doliolida	89	0	1	55	0	0	311	0	0	242	3	724	0	1425	0.218246	0.591327
Egg	4	43	0	73	240	0	0	1054	0	62	2	240	2	1720	0.612791	0.486355
Fiber	156	0	159	6	113	0	18	2	24	3	6	1033	0	1520	0.015789	0.538462
Gelatinous	220	6	1	493	53	0	141	59	0	1022	14	1001	0	3010	0.339535	0.433481
Multiple	545	0	10	329	997	0	52	17	0	27	81	1113	4	3175	0.025512	0.588832
Nonbio	422	21	85	766	1965	0	149	312	23	359	29	11225	34	15390	0.72937	0.465145
Pteropoda	0	0	0	0	670	0	0	19	0	0	1	315	80	1085	0.073733	0.370079
Total	3417	128	1065	4829	11881	0	761	2052	52	1804	197	20987	127	47300	0.334439	0.456023

图 40: C-13 个特征采用 SVM Linear 进行分类的结果

分类器： 随机森林、SVM

MATLAB：采用随机森林进行训练和分类得到的结果如图 41，其分类准确率为 72.9%。采用 SVM Linear 进行训练和分类得到的结果如图 42，其分类准确率为 58.9%

5.1.3 实验三

选用特征： 在5.1.2实验二所用特征的基础上增加的了不变矩特征，共 20 个特征。

分类器： 随机森林、SVM

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2106	0	166	41	10	16	0	0	57	0	152	177	0	2725	0.772844	0.241628
Bubble	0	624	0	0	5	0	0	14	0	0	0	82	0	725	0.860669	0.023474
Chaetognatha	173	0	1491	0	2	11	0	0	7	1	41	19	0	1745	0.854441	0.126026
CladoceraPenilia	4	0	0	3856	22	1	13	0	0	42	37	650	0	4625	0.83373	0.175716
Copepoda	2	0	1	15	6076	208	0	0	3	0	193	851	21	7370	0.824423	0.234182
Decapoda	6	0	0	0	453	2027	0	3	0	4	86	203	3	2785	0.727828	0.236822
Doliolida	0	0	0	1	0	0	1061	0	0	215	0	148	0	1425	0.744561	0.226112
Egg	0	10	0	13	2	7	0	1322	0	23	12	328	3	1720	0.768605	0.127393
Fiber	75	0	11	4	2	2	0	0	1081	1	55	286	3	1520	0.711184	0.229508
Gelatinous	5	0	0	148	10	6	194	34	0	1810	89	714	0	3010	0.601329	0.29736
Multiple	253	1	27	125	463	209	14	27	62	129	914	949	2	3175	0.287874	0.534149
Nonbio	153	4	10	475	770	144	89	115	193	348	367	12646	76	15390	0.821702	0.267026
Pteropoda	0	0	0	0	119	25	0	0	0	3	16	200	722	1085	0.665438	0.13012
Total	2777	639	1706	4678	7934	2656	1371	1515	1403	2576	1962	17253	830	47300	0.728819	0.219194

图 41: Matlab-19 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2154	0	150	46	18	7	0	0	50	0	147	153	0	2725	0.790459	0.330432
Bubble	0	636	0	15	16	7	0	4	0	0	0	47	0	725	0.877241	0.105485
Chaetognatha	314	0	1371	0	1	11	0	0	6	1	33	8	0	1745	0.785673	0.164534
CladoceraPenilia	1	0	0	3321	34	0	8	0	0	24	13	1224	0	4625	0.718054	0.362082
Copepoda	1	0	0	32	5762	108	0	0	2	0	164	1283	18	7370	0.781818	0.417568
Decapoda	7	0	0	0	1209	1251	0	0	0	0	84	234	0	2785	0.449192	0.256684
Doliolida	3	0	0	5	0	0	606	0	0	547	2	262	0	1425	0.425263	0.437848
Egg	0	44	0	98	171	0	10	1053	0	6	5	328	5	1720	0.612209	0.2171
Fiber	224	0	52	7	30	10	1	0	823	1	39	333	0	1520	0.541447	0.22723
Gelatinous	7	0	0	445	76	2	260	14	0	1326	80	800	0	3010	0.440532	0.434059
Multiple	320	0	45	278	699	161	30	12	28	149	699	749	5	3175	0.220157	0.524813
Nonbio	186	31	23	959	1392	96	163	262	156	289	193	11614	26	15390	0.754646	0.32902
Pteropoda	0	0	0	0	485	30	0	0	0	0	12	274	284	1085	0.261751	0.159763
Total	3217	711	1641	5206	9893	1683	1078	1345	1065	2343	1471	17309	338	47300	0.589111	0.305125

图 42: Matlab-19 个特征采用 SVM Linear 进行分类的结果

MATLAB：采用随机森林进行训练和分类得到的结果如图 43，其分类准确率为 73.7%。采用 SVM Linear 进行训练和分类得到的结果如图 49，其分类准确率为 61.0%

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2209	0	120	34	11	10	0	0	36	0	163	142	0	2725	0.810642	0.207391
Bubble	0	627	0	0	5	0	0	14	0	0	1	78	0	725	0.864828	0.021841
Chaetognatha	131	0	1537	0	2	5	0	0	15	0	39	16	0	1745	0.880802	0.108469
CladoceraPenilia	2	0	0	3902	23	3	6	0	0	19	18	652	0	4625	0.843676	0.161762
Copepoda	1	0	0	15	6112	195	0	0	0	0	191	833	23	7370	0.829308	0.227405
Decapoda	5	0	0	0	440	2033	0	4	0	4	78	215	6	2785	0.729982	0.227877
Doliolida	0	0	0	3	0	0	1073	0	0	205	0	144	0	1425	0.752982	0.207533
Egg	0	13	0	9	6	9	0	1358	0	20	5	300	0	1720	0.789535	0.100066
Fiber	64	0	13	2	1	1	0	0	1121	0	59	259	0	1520	0.7375	0.222068
Gelatinous	5	0	0	129	19	7	182	23	0	1800	88	757	0	3010	0.598007	0.291896
Multiple	247	0	33	112	461	197	13	18	62	147	875	1007	3	3175	0.275591	0.526771
Nonbio	123	1	21	449	719	150	80	92	207	346	317	12823	62	15390	0.833203	0.266083
Pteropoda	0	0	0	0	112	23	0	0	0	1	15	246	688	1085	0.634101	0.120205
Total	2787	641	1724	4655	7911	2633	1354	1509	1441	2542	1849	17472	782	47300	0.736935	0.206874

图 43: Matlab-20 个特征采用随机森林进行分类的结果

5.1.4 实验四

选用特征： 采用内距离形状上下文提取图像特征：

1. 先从图像中挑选 39 张图像作为模板（每种浮游动物中选取 3 张）。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2170	0	154	48	12	9	0	0	40	1	140	151	0	2725	0.79633	0.332513
Bubble	0	630	0	15	20	2	0	3	0	0	2	53	0	725	0.868966	0.111425
Chaetognatha	308	0	1379	0	0	11	0	0	8	1	28	10	0	1745	0.790258	0.163228
CladoceraPenilia	3	0	0	3356	18	0	8	0	0	16	15	1209	0	4625	0.725622	0.349108
Copepoda	2	0	0	34	5881	101	0	0	3	1	175	1142	31	7370	0.797965	0.372225
Decapoda	11	0	0	0	877	1488	0	5	0	0	83	321	0	2785	0.534291	0.237314
Doliolida	1	0	0	1	0	0	554	0	0	592	1	276	0	1425	0.388772	0.438134
Egg	0	43	0	75	52	9	10	1196	0	2	4	329	0	1720	0.695349	0.10412
Fiber	237	0	51	7	41	11	2	0	816	0	39	316	0	1520	0.536842	0.214629
Gelatinous	6	0	0	408	72	4	243	3	0	1373	70	831	0	3010	0.456146	0.438905
Multiple	318	0	42	280	674	163	19	3	25	152	718	774	7	3175	0.226142	0.51552
Nonbio	195	36	22	932	1274	125	150	125	147	309	193	11837	45	15390	0.769136	0.322322
Pteropoda	0	0	0	0	447	28	0	0	0	0	14	218	378	1085	0.348387	0.180043
Total	3251	709	1648	5156	9368	1951	996	1335	1039	2447	1482	17467	461	47300	0.610323	0.29073

图 44: Matlab-20 个特征采用 SVM Linear 进行分类的结果

2. 采用内距离形状上下文分别计算训练集和测试集中所有图像和上一步中 39 张图像间的距离。
3. 将上面计算得到的距离矩阵作为训练集和测试集的特征，输入到 ELM 中进行学习和分类。

分类器： ELM（经过实验发现 ELM 的分类效果要比 SVM 和随机森林的好）

MATLAB：采用 ELM 进行训练和分类得到的结果如图 45，其分类准确率为 63.4%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2092	0	90	2	17	43	3	1	105	25	102	234	11	2725	0.767706	0.314772
Bubble	0	258	1	5	5	1	0	310	0	4	1	140	0	725	0.358862	0.5
Chaetognatha	164	0	1413	2	5	5	2	0	104	19	14	17	0	1745	0.809742	0.172232
CladoceraPenilia	0	0	0	3525	97	21	9	4	1	19	15	932	2	4625	0.762162	0.264245
Copepoda	12	0	3	98	5742	124	6	0	8	49	145	1179	4	7370	0.779104	0.266198
Decapoda	23	0	1	120	303	1639	5	16	2	46	53	572	5	2785	0.58851	0.270258
Doliolida	1	0	0	36	19	3	645	0	1	83	30	607	0	1425	0.452632	0.388046
Egg	0	216	0	65	10	6	4	948	1	33	11	425	1	1720	0.551163	0.377544
Fiber	168	0	147	0	11	5	1	4	887	17	87	191	2	1520	0.583553	0.377544
Gelatinous	13	15	1	81	136	10	57	110	7	1546	112	894	28	3010	0.513621	0.337618
Multiple	289	1	10	157	377	114	82	14	118	146	506	1361	0	3175	0.15937	0.649827
Nonbio	247	26	25	679	1065	253	239	114	189	337	356	11807	53	15390	0.767186	0.364737
Pteropoda	44	0	16	21	38	22	1	2	2	10	13	227	689	1085	0.635023	0.133333
Total	3053	516	1707	4791	7825	2246	1054	1523	1425	2334	1445	18586	795	47300	0.59428	0.33972

图 45: Matlab-IDSC 特征采用 ELM 进行分类的结果（39 张图像作为模板）

5.1.5 实验五

选用特征： 采用内距离形状上下文提取图像特征。该实验在5.1.4实验四的基础上，将 39 张图像模板增加到 65 张。

分类器： ELM

MATLAB：采用 ELM 进行训练和分类得到的结果如图 46，其分类准确率为 64.8%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2268	0	40	0	4	34	0	0	73	6	79	218	3	2725	0.832294	0.291693
Bubble	0	202	0	7	6	0	0	364	5	1	1	139	0	725	0.278621	0.422857
Chaetognatha	145	0	1514	2	1	6	0	1	54	1	7	14	0	1745	0.867622	0.149916
CladoceraPenilia	0	0	0	3829	44	31	4	3	0	12	9	692	1	4625	0.827892	0.213111
Copepoda	7	0	0	56	5972	78	5	0	7	27	161	1053	4	7370	0.810312	0.21771
Decapoda	16	0	0	106	115	1865	1	3	2	22	47	606	2	2785	0.669659	0.231878
Doliolida	0	0	0	10	6	0	1011	4	0	48	23	323	0	1425	0.709474	0.256071
Egg	1	109	0	61	5	13	9	1136	11	24	11	340	0	1720	0.660465	0.332158
Fiber	140	0	158	0	9	3	1	5	902	3	101	198	0	1520	0.593421	0.330861
Gelatinous	13	7	0	89	132	11	60	82	7	1652	101	847	9	3010	0.548837	0.233766
Multiple	281	0	3	123	367	118	62	10	99	100	537	1473	2	3175	0.169134	0.632695
Nonbio	280	32	56	581	953	261	206	93	186	246	379	12085	32	15390	0.78525	0.335989
Pteropoda	51	0	10	2	20	8	0	0	2	14	6	212	760	1085	0.700461	0.065191
Total	3202	350	1781	4866	7634	2428	1359	1701	1348	2156	1462	18200	813	47300	0.650265	0.285684

图 46: Matlab-IDSC 特征采用 ELM 进行分类的结果 (65 张图像作为模板)

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2170	0	154	48	12	9	0	0	40	1	140	151	0	2725	0.79633	0.332513
Bubble	0	630	0	15	20	2	0	3	0	0	2	53	0	725	0.868966	0.111425
Chaetognatha	308	0	1379	0	0	11	0	0	8	1	28	10	0	1745	0.790258	0.163228
CladoceraPenilia	3	0	0	3356	18	0	8	0	0	16	15	1209	0	4625	0.725622	0.349108
Copepoda	2	0	0	34	5881	101	0	0	3	1	175	1142	31	7370	0.797965	0.372225
Decapoda	11	0	0	0	877	1488	0	5	0	0	83	321	0	2785	0.534291	0.237314
Doliolida	1	0	0	1	0	0	554	0	0	592	1	276	0	1425	0.388772	0.438134
Egg	0	43	0	75	52	9	10	1196	0	2	4	329	0	1720	0.695349	0.10412
Fiber	237	0	51	7	41	11	2	0	816	0	39	316	0	1520	0.536842	0.214629
Gelatinous	6	0	0	408	72	4	243	3	0	1373	70	831	0	3010	0.456146	0.438905
Multiple	318	0	42	280	674	163	19	3	25	152	718	774	7	3175	0.226142	0.51552
Nonbio	195	36	22	932	1274	125	150	125	147	309	193	11837	45	15390	0.769136	0.322322
Pteropoda	0	0	0	0	447	28	0	0	0	0	14	218	378	1085	0.348387	0.180043
Total	3251	709	1648	5156	9368	1951	986	1335	1039	2447	1482	17467	461	47300	0.610323	0.29073

图 47: Matlab-20 个特征采用 SVM Linear 进行分类的结果

5.1.6 实验六

选用特征： 采用内距离形状上下文提取图像特征。该实验在5.1.4实验四的基础上，将 39 张图像模板增加到 104 张。

分类器： ELM

MATLAB：采用 ELM 进行训练和分类得到的结果如图 48，其分类准确率为 65.9%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2279	0	42	0	5	35	0	1	50	7	78	228	0	2725	0.83633	0.275358
Bubble	0	164	0	5	1	0	0	382	5	1	2	165	0	725	0.226207	0.422535
Chaetognatha	138	0	1530	0	3	1	0	0	52	0	8	13	0	1745	0.876791	0.114071
CladoceraPenilia	2	0	0	3844	49	33	0	3	0	12	8	674	0	4625	0.831135	0.201993
Copepoda	2	0	0	59	5987	66	6	1	5	31	156	1053	4	7370	0.812347	0.209428
Decapoda	11	0	0	90	109	1972	11	1	2	20	43	526	0	2785	0.708079	0.214968
Doliolida	0	0	0	13	7	2	1018	5	2	43	14	320	1	1425	0.714386	0.249816
Egg	1	82	1	66	6	11	11	1185	10	22	12	313	0	1720	0.689953	0.33013
Fiber	141	0	99	0	10	1	3	5	939	10	95	217	0	1520	0.617763	0.319565
Gelatinous	7	5	0	78	138	18	59	81	3	1643	97	873	8	3010	0.545847	0.23439
Multiple	263	0	8	125	344	110	55	9	97	106	516	1538	4	3175	0.16252	0.629044
Nonbio	260	33	37	536	886	253	194	96	206	241	352	12259	37	15390	0.796556	0.332916
Pteropoda	41	0	10	1	28	10	0	0	9	10	10	198	768	1085	0.707834	0.065693
Total	3145	284	1727	4817	7573	2512	1357	1769	1380	2146	1391	18377	822	47300	0.65575	0.276916

图 48: Matlab-IDSC 特征采用 ELM 进行分类的结果 (104 张图像作为模板)

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2170	0	154	48	12	9	0	0	40	1	140	151	0	2725	0.79633	0.332513
Bubble	0	630	0	15	20	2	0	3	0	0	2	53	0	725	0.868966	0.111425
Chaetognatha	308	0	1379	0	0	11	0	0	8	1	28	10	0	1745	0.790258	0.163228
CladoceraPenilia	3	0	0	3356	18	0	8	0	0	16	15	1209	0	4625	0.725622	0.349108
Copepoda	2	0	0	34	5881	101	0	0	3	1	175	1142	31	7370	0.797965	0.372225
Decapoda	11	0	0	0	877	1488	0	5	0	0	83	321	0	2785	0.534291	0.237314
Doliolida	1	0	0	1	0	0	554	0	0	592	1	276	0	1425	0.388772	0.438134
Egg	0	43	0	75	52	9	10	1196	0	2	4	329	0	1720	0.695349	0.10412
Fiber	237	0	51	7	41	11	2	0	816	0	39	316	0	1520	0.536842	0.214629
Gelatinous	6	0	0	408	72	4	243	3	0	1373	70	831	0	3010	0.456146	0.438905
Multiple	318	0	42	280	674	163	19	3	25	152	718	774	7	3175	0.226142	0.51552
Nonbio	195	36	22	932	1274	125	150	125	147	309	193	11837	45	15390	0.769136	0.322322
Pteropoda	0	0	0	0	447	28	0	0	0	0	14	218	378	1085	0.348387	0.180043
Total	3251	709	1648	5156	9368	1951	986	1335	1039	2447	1482	17467	461	47300	0.610323	0.29073

图 49: Matlab-20 个特征采用 SVM Linear 进行分类的结果

5.2 特征融合方法实验

该部分进行的是特征融合实验，确定哪种特征融合方式可以得到更高的分类准确率。

5.2.1 实验一（特征融合方法一）

该实验进行的是特征融合。由于在实验一——三中使用的特征都是特征值，而实验四中使用的特征以及 HOG、LBP 和其他的一些特征都是特征向量的形式。如果要将这些特征一起使用就需要进行特征融合。在该实验中，将5.1.3中的 20 个特征和 LBP 特征融合，具体的融合方法：

1. 用训练集不同种的特征（这里的将特征分为两种：实验三中 20 个特征作为一种，LBP 特征作为一种）分别进行训练得到分类器（20 个特征采用随机森林进行训练，LBP 采用 SVM 进行训练）。然后将训练集对应的这两种特征分别输入到其对应的分类器中进行预测，这两种特征会分别得到训练集中每个样本属于每个类别的分类概率（ $m \times n$ 维， m 为训练集样本数， n 为类别数）。
2. 将每种特征得到的概率进行拼接（ $m \times 2n$ 维），再输入到分类器（这里的分类器使用的是 SVM）进行训练。
3. 然后将测试集的分类概率（用和步骤 1 中同样的方法可以得到测试集的分类概率）输入到步骤 2 中训练好的分类器，得到最终的分类结果。

在该实验得到的分类结果如图50，其分类准确率为 76.1%。

5.2.2 实验二（特征融合方法二）

该实验也是将5.1.3中的 20 个特征和 LBP 特征融合，采用的融合方法：

1. 用训练集不同种的特征（这里的将特征分为两种：实验三中的 20 个特征作为一种，LBP 特征作为一种）分别进行训练得到分类器（20 个特征采用随机森林进行训练，LBP 采用 SVM 进行训

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2188	0	123	22	6	9	0	0	44	0	182	151	0	2725	0.802936	0.197359
Bubble	0	661	0	0	5	0	0	1	0	0	0	58	0	725	0.911724	0.001511
Chaetognatha	137	0	1543	0	3	3	0	0	8	0	36	15	0	1745	0.884241	0.113218
CladoceraPenilia	0	0	0	3962	24	4	4	0	0	18	29	584	0	4625	0.856649	0.1536
Copepoda	1	0	0	15	6129	198	0	0	0	1	242	756	28	7370	0.831615	0.218937
Decapoda	4	0	0	0	390	2079	0	0	0	2	95	210	5	2785	0.746499	0.224254
Doliolida	0	0	0	3	0	0	1150	0	0	148	0	124	0	1425	0.807018	0.14307
Egg	0	1	0	13	5	8	0	1440	0	20	7	226	0	1720	0.837209	0.08803
Fiber	58	0	19	1	2	0	0	0	1135	0	64	241	0	1520	0.746711	0.220467
Gelatinous	5	0	0	113	11	7	108	19	0	2004	81	662	0	3010	0.665781	0.265934
Multiple	230	0	32	119	446	186	9	21	61	135	996	940	0	3175	0.313701	0.534579
Nonbio	103	0	22	433	725	156	71	98	208	401	386	12717	70	15390	0.826316	0.247114
Pteropoda	0	0	1	0	101	30	0	0	0	1	22	207	723	1085	0.666359	0.124697
Total	2726	662	1740	4681	7847	2680	1342	1579	1456	2730	2140	16891	826	47300	0.761289	0.194828

图 50: Matlab-20 个特征和 LBP 特征融合方法一

练)。

2. 计算每个分类器的权重：对于训练样本集中的每一个样本，分别将其每种特征输入到对应的特征分类器中进行识别，如果能够识别正确，则其对应的特征分类器的权重加一，最终得到每种特征的权重。
3. 预测概率：每幅的不同种类特征通过分类器可以得到其属于每个类别的分类概率（与特征融合方法一中步骤 1 相同）。根据权重和分类概率，计算出最终属于各个类别的概率。

在该实验得到的分类结果如图51，其分类准确率为 73.7%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2202	0	109	27	5	11	0	0	57	1	160	153	0	2725	0.808073	0.208198
Bubble	0	624	0	0	5	0	0	13	0	0	0	83	0	725	0.86069	0.023474
Chaetognatha	141	0	1544	0	4	2	0	0	11	0	33	10	0	1745	0.884814	0.101281
CladoceraPenilia	1	0	0	3922	23	3	7	0	0	22	19	628	0	4625	0.848	0.161787
Copepoda	1	0	0	19	6122	191	0	0	0	1	191	822	23	7370	0.830665	0.228384
Decapoda	5	0	0	0	452	2013	0	4	0	5	83	221	2	2785	0.722801	0.226662
Doliolida	0	0	0	5	0	0	1068	0	0	187	0	165	0	1425	0.749474	0.208302
Egg	0	9	0	16	6	9	0	1366	0	17	5	291	1	1720	0.794186	0.102497
Fiber	70	0	12	1	0	1	0	0	1131	0	54	249	2	1520	0.744079	0.235294
Gelatinous	3	0	0	118	18	5	178	23	0	1818	75	772	0	3010	0.603987	0.277998
Multiple	238	1	36	122	462	193	16	14	70	132	882	1007	2	3175	0.277795	0.515651
Nonbio	120	5	17	449	720	142	80	102	210	334	305	12833	73	15390	0.833853	0.265679
Pteropoda	0	0	0	0	117	33	0	0	0	1	14	242	678	1085	0.624885	0.131882
Total	2781	639	1718	4679	7934	2603	1349	1522	1479	2518	1821	17476	781	47300	0.737177	0.206699

图 51: Matlab-20 个特征和 LBP 特征融合方法二

5.3 融合不同特征实验

该部分实验采用特征融合方法一。在实验过程中，采用的针对不同特征采用的分类器：

- 5.1.3实验三中的 20 个特征：随机森林
- LBP 特征、Gabor 特征、Fourier 描述子、SIFT 特征：SVM
- IDSC：ELM

5.3.1 实验一

选用特征： 在5.2.1特征融合方法一的基础上，将5.1.3中的 20 个特征、LBP 特征融和 Gabor 特征融合。

MATLAB：在该实验得到的分类结果如图52，其分类准确率为 73.6%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2178	0	75	13	1	7	0	0	31	3	259	158	0	2725	0.799266	0.190335
Bubble	1	663	0	0	5	2	0	10	0	0	2	42	0	725	0.914483	0.005997
Chaetognatha	158	0	1526	0	0	4	0	0	13	0	32	11	1	1745	0.874499	0.093286
CladoceraPenilia	0	0	0	3753	35	3	1	0	0	16	33	784	0	4625	0.811459	0.138825
Copepoda	3	0	1	25	5866	186	0	1	0	3	276	996	13	7370	0.795929	0.195226
Decapoda	4	0	0	11	300	2022	0	0	0	1	124	319	4	2785	0.726032	0.239278
Doliolida	0	0	0	6	6	0	1007	0	0	169	5	232	0	1425	0.706667	0.071033
Egg	0	3	0	9	1	14	1	1437	0	33	9	213	0	1720	0.835465	0.072305
Fiber	62	0	33	0	2	2	3	2	1064	0	89	263	0	1520	0.7	0.2
Gelatinous	5	0	1	76	17	4	17	7	3	1982	127	771	0	3010	0.658472	0.236518
Multiple	194	1	16	91	315	188	10	7	51	90	1127	1081	4	3175	0.354961	0.561137
Nonbio	81	0	27	374	654	158	45	85	168	299	461	13001	37	15390	0.844769	0.284991
Pteropoda	4	0	4	0	87	68	0	0	0	0	24	312	586	1085	0.540092	0.091473
Total	2690	667	1683	4358	7289	2658	1084	1549	1330	2596	2568	18183	645	47300	0.735546	0.183108

图 52: Matlab-20 个特征、LBP 和 Gabor 特征融合方法一

5.3.2 实验二

选用特征： 在5.2.1特征融合方法一的基础上，将5.1.3中的 20 个特征、LBP 特征融 Fourier 描述子融合。

MATLAB：在该实验得到的分类结果如图53，其分类准确率为 76.2%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2204	0	116	20	9	10	0	0	42	0	176	148	0	2725	0.808807	0.195914
Bubble	0	659	0	0	5	0	0	4	0	0	0	57	0	725	0.908966	0.003026
Chaetognatha	133	0	1549	0	3	2	0	0	9	0	33	16	0	1745	0.887679	0.10669
CladoceraPenilia	1	0	0	3944	22	5	4	0	0	16	34	599	0	4625	0.852757	0.152557
Copepoda	0	0	0	16	6135	181	0	0	0	0	222	792	24	7370	0.832429	0.223516
Decapoda	5	0	0	0	416	2061	0	0	0	3	87	207	6	2785	0.740036	0.214857
Doliolida	0	0	0	3	0	0	1145	0	0	154	0	123	0	1425	0.803509	0.143605
Egg	0	1	0	13	8	7	0	1450	0	20	5	216	0	1720	0.843023	0.091479
Fiber	58	0	17	1	1	0	0	0	1150	0	64	229	0	1520	0.756579	0.217155
Gelatinous	4	0	0	110	10	6	108	20	0	1998	78	676	0	3010	0.663787	0.267058
Multiple	226	0	28	119	452	177	9	21	63	130	981	969	0	3175	0.308976	0.524939
Nonbio	109	1	23	428	738	145	71	101	205	405	365	12729	70	15390	0.827096	0.249735
Pteropoda	1	0	1	0	102	31	0	0	0	0	20	205	725	1085	0.668203	0.121212
Total	2741	661	1734	4654	7901	2625	1337	1596	1469	2726	2065	16966	825	47300	0.76168	0.193211

图 53: Matlab-20 个特征、LBP 和 Fourier 描述子融合方法一

5.3.3 实验三

选用特征： 在5.2.1特征融合方法一的基础上，将5.1.3中的 20 个特征、LBP 特征和 SIFT 特征融合。

MATLAB：在该实验得到的分类结果如图54，其分类准确率为 76.1%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2193	0	122	21	8	9	0	0	48	0	172	152	0	2725	0.804771	0.199927
Bubble	0	661	0	0	5	0	0	1	0	0	0	58	0	725	0.911724	0.001511
Chaetognatha	138	0	1544	0	3	3	0	0	8	0	33	16	0	1745	0.884814	0.111623
CladoceraPenilia	0	0	0	3964	25	4	4	0	0	19	27	582	0	4625	0.857081	0.154076
Copepoda	2	0	0	15	6140	194	0	0	0	1	239	751	28	7370	0.833107	0.220416
Decapoda	4	0	0	0	392	2073	0	0	0	2	97	212	5	2785	0.744345	0.221262
Doliolida	0	0	0	4	0	0	1146	0	0	149	0	126	0	1425	0.804211	0.142857
Egg	0	1	0	12	4	8	0	1438	0	21	9	227	0	1720	0.836047	0.088142
Fiber	60	0	17	1	1	0	0	0	1130	0	64	247	0	1520	0.743421	0.220152
Gelatinous	5	0	0	116	12	7	108	19	0	2001	81	661	0	3010	0.664784	0.268105
Multiple	231	0	32	118	447	188	9	21	60	137	992	940	0	3175	0.312441	0.531191
Nonbio	108	0	22	435	734	148	70	98	203	403	381	12720	68	15390	0.826511	0.247248
Pteropoda	0	0	1	0	105	28	0	0	0	1	21	206	723	1085	0.666359	0.122573
Total	2741	662	1738	4686	7876	2662	1337	1577	1449	2734	2116	16898	824	47300	0.76074	0.194545

图 54: Matlab-20 个特征、LBP 和 SIFT 特征融合方法一

5.3.4 实验四

选用特征： 在5.2.1特征融合方法一的基础上，将5.1.3中的 20 个特征和5.1.4中的 IDSC 特征融合。

MATLAB：在该实验得到的分类结果如图55，其分类准确率为 76.3%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2222	0	106	15	8	9	0	0	44	1	185	135	0	2725	0.815413	0.195219
Bubble	0	628	0	0	5	0	0	18	0	0	4	70	0	725	0.866207	0.017214
Chaetognatha	128	0	1560	0	4	1	0	0	9	0	32	11	0	1745	0.893983	0.097744
CladoceraPenilia	0	0	0	4023	32	4	7	0	0	23	31	505	0	4625	0.869838	0.151803
Copepoda	1	0	0	20	6155	181	0	0	0	0	215	786	12	7370	0.835142	0.216123
Decapoda	4	0	0	0	373	2119	0	4	0	0	93	190	2	2785	0.760862	0.202484
Doliolida	0	0	0	5	0	0	1116	0	0	146	1	157	0	1425	0.783158	0.180617
Egg	0	10	1	10	5	10	0	1403	0	17	7	256	1	1720	0.815698	0.119272
Fiber	61	0	17	0	5	0	0	0	1147	1	57	232	0	1520	0.754605	0.230201
Gelatinous	3	0	0	112	14	3	151	27	0	1975	88	637	0	3010	0.656146	0.262509
Multiple	225	0	25	116	454	181	13	21	64	141	1003	932	0	3175	0.315906	0.533922
Nonbio	117	1	17	442	725	126	75	120	226	370	423	12690	58	15390	0.824561	0.243833
Pteropoda	0	0	3	0	72	23	0	0	0	4	13	181	789	1085	0.727189	0.084687
Total	2761	639	1729	4743	7852	2657	1362	1593	1490	2678	2152	16782	862	47300	0.762977	0.195048

图 55: Matlab-20 个特征和 IDSC 特征融合方法一

5.3.5 实验五

选用特征： 在5.2.1特征融合方法一的基础上，将5.1.3中的 20 个特征、LBP 特征和5.1.4中的 IDSC 特征融合。（采用特征融合方法一）

MATLAB：在该实验中，IDSC 采用 39 个模板得到的分类结果如图56，其分类准确率为 77.1%。

MATLAB：在该实验中，IDSC 采用 65 个模板得到的分类结果如图57，其分类准确率为 77.7%。

MATLAB：在该实验中，IDSC 采用 104 个模板得到的分类结果如图58，其分类准确率为 77.7%。

5.4 ELM 作为分类器的实验

该部分使用极限学习机（ELM）作为分类器。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2250	0	116	14	7	5	0	0	34	0	154	145	0	2725	0.825688	0.196715
Bubble	0	651	0	1	5	0	0	0	2	0	0	1	65	0	725	0.897931
Chaetognatha	151	0	1544	0	1	0	0	0	8	0	31	10	0	1745	0.884814	0.105446
CladoceraPenilia	0	0	0	4020	30	2	5	0	0	14	29	525	0	4625	0.869189	0.136041
Copepoda	1	0	0	16	6190	158	0	0	2	0	230	763	10	7370	0.839891	0.216852
Decapoda	5	0	0	1	369	2120	0	0	0	1	96	193	0	2785	0.761221	0.191457
Doliolida	0	0	0	7	0	0	1156	0	0	114	0	148	0	1425	0.811228	0.122914
Egg	0	0	0	6	5	6	0	1462	0	15	4	222	0	1720	0.85	0.075269
Fiber	66	0	11	1	4	0	0	0	1134	0	65	239	0	1520	0.746053	0.222755
Gelatinous	3	0	0	99	10	3	86	17	0	2040	76	676	0	3010	0.677741	0.23994
Multiple	223	0	30	93	481	175	9	13	67	122	997	965	0	3175	0.314016	0.528605
Nonbio	100	0	18	395	730	128	62	87	214	376	419	12803	58	15390	0.831904	0.24417
Pteropoda	2	0	7	0	72	25	0	0	0	2	13	185	779	1085	0.717972	0.080283
Total	2801	651	1726	4653	7904	2622	1318	1581	1459	2684	2115	16939	847	47300	0.771358	0.181573

图 56: Matlab-20 个特征、LBP 和 IDSC 特征融合方法一

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2253	0	106	14	4	5	0	0	35	0	165	143	0	2725	0.826789	0.192473
Bubble	0	658	0	0	5	0	0	0	0	0	1	61	0	725	0.907586	0.004539
Chaetognatha	119	0	1582	0	0	3	0	0	4	0	21	16	0	1745	0.90659	0.094966
CladoceraPenilia	0	0	0	4063	29	3	3	0	0	16	17	494	0	4625	0.878486	0.135164
Copepoda	1	0	0	14	6216	145	0	0	0	1	201	784	8	7370	0.843419	0.217128
Decapoda	8	0	1	1	367	2122	0	0	0	2	85	198	1	2785	0.761939	0.186038
Doliolida	0	0	0	6	0	0	1181	0	0	108	1	129	0	1425	0.828772	0.116018
Egg	0	1	0	15	6	10	1	1438	0	19	7	223	0	1720	0.836047	0.075835
Fiber	51	0	12	1	3	1	0	0	1137	0	66	249	0	1520	0.748026	0.214779
Gelatinous	4	0	2	94	12	2	75	12	0	2066	86	657	0	3010	0.686379	0.23792
Multiple	230	0	23	99	473	166	14	18	74	115	994	969	0	3175	0.313071	0.50962
Nonbio	122	2	18	391	762	125	62	88	197	380	370	12813	60	15390	0.832554	0.242596
Pteropoda	2	0	4	0	63	25	0	0	1	4	13	181	792	1085	0.729954	0.080139
Total	2790	661	1748	4698	7940	2607	1336	1556	1448	2711	2027	16917	861	47300	0.776893	0.177478

图 57: Matlab-20 个特征、LBP 和 IDSC 特征融合方法一

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2250	0	116	13	7	6	0	0	46	0	148	139	0	2725	0.825688	0.189773
Bubble	0	660	0	0	5	0	0	0	0	0	1	59	0	725	0.910345	0.001513
Chaetognatha	120	0	1574	0	1	1	0	0	6	0	30	13	0	1745	0.902006	0.105173
CladoceraPenilia	0	0	0	4076	25	1	5	0	0	15	18	485	0	4625	0.881297	0.142075
Copepoda	0	0	0	15	6206	139	0	0	0	1	222	775	12	7370	0.842062	0.214729
Decapoda	7	0	0	0	381	2127	0	0	0	1	79	187	3	2785	0.763734	0.174942
Doliolida	0	0	0	8	0	0	1154	0	0	108	0	155	0	1425	0.809825	0.110254
Egg	0	0	0	11	5	11	0	1444	0	24	7	217	1	1720	0.839535	0.079082
Fiber	47	0	15	1	4	0	0	0	1150	0	67	236	0	1520	0.756579	0.21875
Gelatinous	3	0	0	106	17	2	68	18	0	2075	80	641	0	3010	0.689369	0.237132
Multiple	229	0	25	106	470	147	12	21	74	122	999	969	1	3175	0.314646	0.521781
Nonbio	120	1	22	415	725	120	58	85	196	372	424	12785	67	15390	0.830734	0.241156
Pteropoda	1	0	7	0	57	24	0	0	0	2	14	187	793	1085	0.730876	0.095781
Total	2777	661	1759	4751	7903	2578	1297	1568	1472	2720	2089	16848	877	47300	0.776669	0.179396

图 58: Matlab-20 个特征、LBP 和 IDSC 特征融合方法一

5.4.1 实验一

选用特征： 采用5.1.3中的 20 个特征进行实验。

MATLAB：在该实验中 ELM 隐藏神经元设置为 650 个，实验得到的分类结果如图59，其分类准确率为 72.4%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	1138	0	40	22	3	9	3	1	25	3	71	44	0	1359	0.83738	0.28652
Bubble	0	323	0	0	1	0	0	1	0	0	5	36	0	366	0.882514	0.027108
Chaetognatha	133	0	667	6	8	3	1	1	6	4	22	23	1	875	0.762286	0.087551
CladoceraPenilia	0	0	0	2041	8	4	1	0	0	3	6	273	0	2336	0.873716	0.262911
Copepoda	5	0	1	18	3104	63	1	0	1	4	73	372	8	3650	0.850411	0.258303
Decapoda	5	0	2	12	209	1070	0	4	2	3	24	93	1	1425	0.750877	0.222948
Doliolida	4	0	1	10	2	0	589	0	2	59	3	60	0	730	0.806849	0.232073
Egg	2	3	0	22	11	9	13	629	0	5	6	147	0	847	0.742621	0.103989
Fiber	48	1	3	6	14	5	3	0	518	1	34	131	1	765	0.677124	0.226866
Gelatinous	11	0	0	107	8	4	74	8	0	988	34	287	0	1521	0.649573	0.234109
Multiple	148	2	8	106	304	75	12	5	29	72	387	407	1	1556	0.248715	0.538186
Nonbio	99	3	8	408	430	104	68	50	87	145	155	6067	31	7655	0.792554	0.246429
Pteropoda	2	0	1	11	83	31	2	3	0	3	18	111	300	565	0.530973	0.125364
Total	1595	332	731	2769	4185	1377	767	702	670	1290	838	8051	343	23650	0.723507	0.219412

图 59: Matlab-20 个特征采用 ELM 进行分类

6. 实验总结

使用的特征：

- 20 个统计特征：Mean、StdDev、CV、SR、MeanPos、Elongation、Circ、Feret、PerimAreaexc、CDexc、Skelarea、FeretAreaexc、PerimFeret、矩形度、体态比、凸率、伸长度、灰度共生矩阵（对比度）、对称性（左右）、不变矩。
- LBP 特征
- 内距离形状上下文

分类器：SVM，随机森林，ELM

（不同特征适合不同的分类器：20 个统计特征适合采用随机森林，LBP 适合采用 SVM，内距离形状上下文采用的是 ELM）。

实验结果：

单类特征实验结果		
采用的特征	分类器	分类准确率 (%)
20 个统计特征	随机森林	73.7
LBP 特征	SVM	67.4
内距离形状上下文 (39)	ELM	63.4
内距离形状上下文 (65)	ELM	64.8
内距离形状上下文 (104)	ELM	65.9
多类特征融合实验结果		
采用的特征	分类器	分类准确率 (%)
20 个统计特征+LBP 特征	随机森林+SVM	76.1
20 个统计特征+内距离形状上下文 (39)	随机森林+ELM+SVM	76.3
20 个统计特征+LBP+内距离形状上下文 (39)	随机森林+ELM+SVM	77.1
20 个统计特征+LBP+内距离形状上下文 (65)	随机森林+ELM+SVM	77.7
20 个统计特征+LBP+内距离形状上下文 (104)	随机森林+ELM+SVM	77.7

图 60: 采用不同特征与分类的实验结果统计

实验分析：

- 实验发现, 5.1.3 中 20 个特征、LBP 特征和 IDSC 特征对分类准确率提高有较大作用, 其中最好的是 5.1.3 中 20 个特征。
- 采用 IDSC 提取图像特征时, 先选用了 39 张图像作为模板, 后来又选用 65 张图像作为模板。模板图像数量越多, 分类的准确率越高, 但同时需要的时间越多。在将 IDSC 与另外两类特征融合后, 随着模板数量的增长最终分类准确率的提高并不明显。
- IDSC 通常采用 K-NN 进行分类, 但是计算时间较长, 因此在实验中没有采用 K-NN 进行分类。而是通过选取模板, 计算所有图像和模板之间的距离, 作为图像的特征。

接下来的工作:

- 目前采用的融合方法在融合过程中, 这三种特征所占的权重是一样的。下一步想更改特征融合方法, 根据这三种特征对提高分类准确率的贡献, 在融合过程中设置不同的权重。但是目前对融合方法还是不太了解, 我想接下来先大致学习一下特征融合的方法。
- 由于目标物体轮廓不平滑, 骨架提取的结果比较差。解决方法可以分为两种: 1. 进行骨架剪枝。
2. 平滑物体轮廓边缘。
- 不同的特征适用的分类器不同, 分类器也会影响分类准确率的提高。目前使用的分类器有 SVM、随机森林和 ELM, 想要寻找更加合适的分类器。
- 寻找特征描述方法。

7. 实验尝试

7.1 AdaBoost

AdaBoost 常用的弱分类器有：CART (classification and regression tree), decision stump, SVM, logistic regress 等。AdaBoost 最初设计用于解决二分类，对二分类问题，弱分类器条件要求对任意样本分布，分类器的分类准确率只需略高于 0.5，如果分类器过强容易产生过拟合。实际问题以多分类居多，AdaBoost 在处理多分类问题时，对分类数目为 K 的多分类，要求弱分类器比随机猜测准确率 $\frac{1}{K}$ 略高。

7.1.1 尝试一 (MATLAB 中的 AdaBoost)

20 个统计特征用 AdaBoost 算法进行分类，实验采用的 MATLAB 自带函数 fitensemble，最大迭代次数设为 50 次，实现 AdaBoost 算法。实验结果如图 61，分类准确率为 63.1%，低于随机森林得到的分类准确率 73.7%。

	Appendicularia	Bubble	Chaetognatha	Cladocera	Penilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	1814	0	255	48	5	5	0	0	0	222	3	264	109	0	2725	0.665688	0.33186
Bubble	0	640	0	14	19	0	0	19	0	0	0	5	28	0	725	0.882759	0.215686
Chaetognatha	282	0	1378	0	0	9	0	0	0	39	1	26	10	0	1745	0.789685	0.240771
Cladocera	6	0	0	2648	7	0	6	0	1	82	158	1717	0	0	4625	0.572541	0.348104
Copepoda	5	0	12	9	5371	161	0	0	16	2	242	1365	187	0	7370	0.728765	0.34396
Decapoda	6	0	13	0	791	1495	0	6	5	0	151	316	2	0	2785	0.536804	0.239573
Doliolida	0	0	0	22	0	0	867	0	1	393	0	142	0	0	1425	0.608421	0.472628
Egg	0	39	0	58	11	24	11	1373	2	11	2	189	0	0	1720	0.798256	0.293361
Fiber	115	0	66	6	31	8	2	0	1037	0	28	227	0	0	1520	0.682237	0.379042
Gelatinous	8	19	0	347	69	4	491	16	1	1271	83	701	0	0	3010	0.422259	0.441809
Multiple	291	0	56	199	573	130	32	23	53	171	879	744	24	0	3175	0.27685	0.606183
Nonbio	187	118	27	711	975	119	235	506	292	343	381	11246	250	0	15390	0.730734	0.336949
Pteropoda	1	0	8	0	335	11	0	0	1	0	13	167	549	0	1085	0.505991	0.45751
Total	2715	816	1815	4062	8187	1966	1644	1943	1670	2277	2232	16961	1012	0	47300	0.630845	0.362111

图 61: 20 个特征采用 MATLAB 中的 AdaBoost 算法得到的分类结果

7.1.2 尝试二 (特征融合后采用 AdaBoost)

在特征融合方法一的基础上，将最终的分类方法 SVM 改为 AdaBoost。实验结果如图 62，分类准确率为 70.9%。

7.1.3 尝试三 (将 SVM 作为弱分类器进行级联)

该实验没有完成，在实现过程中发现两个问题：

- SVM 作为弱分类器，在迭代的过程中不知道怎样输入样本权重。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2107	0	89	13	15	10	0	6	20	0	129	336	0	2725	0.773211	0.180793
Bubble	0	647	0	0	5	0	0	4	0	0	0	69	0	725	0.892414	0.007669
Chaetognatha	147	0	1485	0	4	0	0	2	4	0	34	69	0	1745	0.851003	0.072455
CladoceraPenilia	0	0	0	3771	23	1	2	7	0	8	14	799	0	4625	0.815351	0.122411
Copepoda	0	0	0	6	6046	112	0	5	0	0	131	1061	9	7370	0.820353	0.235168
Decapoda	4	0	0	0	479	1883	0	3	0	2	62	352	0	2785	0.676122	0.149503
Doliolida	0	0	0	0	4	0	1010	0	0	115	0	296	0	1425	0.708772	0.095792
Egg	0	1	0	7	10	5	0	1348	0	11	3	335	0	1720	0.783721	0.14899
Fiber	48	0	4	0	6	0	0	2	1041	0	47	372	0	1520	0.684868	0.187354
Gelatinous	2	0	0	81	31	1	68	17	0	1720	60	1030	0	3010	0.571429	0.223827
Multiple	185	0	13	82	477	119	6	48	41	66	688	1449	1	3175	0.216693	0.509272
Nonbio	79	4	10	337	692	74	31	141	175	294	225	13287	41	15390	0.863353	0.328838
Pteropoda	0	0	0	0	113	9	0	1	0	0	9	342	611	1085	0.563134	0.077039
Total	2572	652	1601	4297	7905	2214	1117	1584	1281	2216	1402	19797	662	47300	0.709263	0.179932

图 62: 采用 Adaboost+SVM 进行特征融合

- 采用 SVM 根据已有的特征进行分类得到的分类结果已经超过 50%，这应该不是弱分类器，对它采用 AdaBoost 进行级联是不是会产生过拟合。

7.1.4 尝试四（将二分类改为多分类问题）

将解决二分类问题的算法改为解决多分类问题，采用拆分的方法。将多分类问题改成多个二分类问题，二分类为一对其余。实验结果如图63，分类准确率为 76.6%。实验中遇到问题：

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2257	0	104	18	4	7	0	0	33	0	161	141	0	2725	0.828257	0.192198
Bubble	0	639	0	0	5	0	0	6	0	0	1	74	0	725	0.881379	0.025915
Chaetognatha	123	0	1577	0	0	2	0	0	3	0	24	16	0	1745	0.903725	0.096275
CladoceraPenilia	0	0	0	4034	29	4	3	0	0	15	18	522	0	4625	0.872216	0.142979
Copepoda	1	0	0	15	6217	150	0	0	0	1	204	772	10	7370	0.843555	0.219852
Decapoda	8	0	2	1	388	2100	0	0	0	1	85	199	1	2785	0.754039	0.190127
Doliolida	0	0	0	7	0	0	1135	0	0	156	0	127	0	1425	0.796491	0.167278
Egg	0	14	0	13	7	10	1	1405	0	19	7	244	0	1720	0.81686	0.094716
Fiber	51	0	12	1	3	1	0	0	1143	0	67	242	0	1520	0.751974	0.217659
Gelatinous	4	0	1	105	13	2	150	15	0	1983	86	651	0	3010	0.658804	0.259522
Multiple	227	0	25	100	478	167	14	17	75	116	989	967	0	3175	0.311496	0.514482
Nonbio	121	3	20	413	755	126	60	109	206	383	381	12739	74	15390	0.827745	0.245945
Pteropoda	2	0	4	0	70	24	0	0	1	4	14	200	766	1085	0.705991	0.099882
Total	2794	656	1745	4707	7969	2593	1363	1552	1461	2678	2037	16894	851	47300	0.765579	0.189756

图 63: 采用 Adaboost+SVM 进行特征融合

- 实验中部分样本不能被识别（识别的结果 13 类都不是）。在实验中，我将这些不能识别的样本再用 SVM 进行分类。

7.1.5 AdaBoost 应用

AdaBoost 可以用于：

1. 用于检测、识别和分类问题：AdaBoost 应用最多的就是分类，通过级联弱分类器而得到强分类器，提高分类准确率。

2. 用于特征选择：在原始特征集合中，挑选出一些最具有代表性、可分性最好的特征子集。基于 AdaBoost 的特征选择过程：1、初始化训练样本权重 2、设计每个特征的分类器 3、根据加权训练样本最小错误率准则，选择分类器，也就是选择了特征 4、调整样本权重 5、通过循环，最后得到分类器的线性组合

近年来提出了对集成学习方法进行优化、剪枝。

7.2 Multi-view Learning (多视角学习)

1. co-training (协同训练)：在未标记数据的两个不同视角下，轮流的训练，使相互一致性最大化。标准协同训练算法的步骤为：
 - (a) 输入：标记数据集 L ，未标记数据集 U 。
 - (b) 用 L_1 训练视图 X_1 上的分类器 f_1 ，用 L_2 训练视图 X_2 上的分类器 f_2 ；
 - (c) 用 f_1 和 f_2 分别对未标记数据 U 进行分类；
 - (d) 把 f_1 对 U 的分类结果中，前 k 个最置信的数据（正例 p 个反例 n 个）及其分类结果加入 L_2 ；把 f_2 对 U 的分类结果中，前 k 个最置信的数据及其分类结果加入 L_1 ；把这 2 ($p+n$) 个数据从 U 中移除；
 - (e) 重复上述过程，直到 U 为空集。
2. multiple kernel learning (多核学习)：核指的是核函数。我们学过的 SVM 都是单核的，在使用的时候，需要根据经验或试验来选择用哪种核函数、怎样指定它的参数。然而实际对图像分类的实验中会用到纹理、形状等不同特征，这些特征最适合的核函数不一定相同，此时可以使用多核学习。多核学习：给定一些基本核函数，用它们的线性组合作为最终的核函数。通过训练得到线性组合中每一个核函数的权重。⁷
3. subspace learning (子空间学习)：是指通过投影，实现高维特征向低维空间的映射，是一种经典的降维思想。

7.3 Bayes Fusion

7.4 Fuzzy Neural network

模糊神经网络是模糊理论和神经网络结合的产物，是具有模糊权系数或者输入信号是模糊量的神经网络。在特征融合方法一的基础上，将最终的分类方法 SVM 改为 FNN，实验结果如图64，分类准

⁷<http://www.jianshu.com/p/ad010080a200>

确率为 75.8%。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2189	0	134	10	8	19	0	0	49	3	184	128	1	2725	0.803303	0.221827
Bubble	0	660	0	2	6	0	0	6	1	1	1	48	0	725	0.910345	0.001513
Chaetognatha	130	0	1552	0	0	6	0	0	15	2	18	20	2	1745	0.889398	0.153762
CladoceraPenilia	3	0	0	3927	33	11	6	0	0	44	82	518	1	4625	0.849081	0.145375
Copepoda	10	0	8	33	6017	152	2	1	2	4	361	742	38	7370	0.816418	0.232428
Decapoda	11	0	5	6	390	2035	0	3	0	7	143	178	7	2785	0.7307	0.216102
Doliolida	0	0	0	8	2	1	1120	0	2	121	19	152	0	1425	0.785965	0.116719
Egg	1	1	0	7	8	12	1	1439	1	20	10	220	0	1720	0.836628	0.086929
Fiber	68	0	41	1	5	2	0	0	1095	1	68	236	3	1520	0.720395	0.265594
Gelatinous	10	0	3	95	25	2	63	14	2	2066	103	624	3	3010	0.686379	0.259764
Multiple	240	0	41	114	458	173	11	10	73	112	1012	924	7	3175	0.31874	0.617391
Nonbio	143	0	41	392	822	160	65	103	248	408	620	12290	98	15390	0.798571	0.244297
Pteropoda	8	0	9	0	65	23	0	0	3	2	24	183	768	1085	0.707834	0.172414
Total	2813	661	1834	4595	7839	2596	1268	1576	1491	2791	2645	16263	928	47300	0.757981	0.210316

图 64: 采用 FNN+SVM 进行特征融合