

RCSHIP-1.0 数据集使用说明

数据主编：冷祥光、谭向东、计科峰、匡纲要（国防科技大学电子科学学院，长沙 410073）

下载说明：SAR 距离压缩域舰船目标检测数据集 1.0 (RCSHIP-1.0) 所有权归国防科技大学电子科学学院所有，《雷达学报》编辑部具有编辑出版权等。读者可免费使用该数据进行教学、科研等，但需在论文、报告等成果中引用或致谢。该数据禁止私自用于商业目的，如有商业需求，请与《雷达学报》编辑部联系。首次数据下载，请关注微信公众号后注册，并通过邮箱验证；以后数据下载，在开始时微信扫码即可。英文网站数据下载和注册，可直接通过邮箱验证进行。如有更多需求，欢迎联系冷祥光 (luckight@163.com)、谭向东 (13549207302@163.com)。

本数据集参考文献与引用格式：

[1] X. Tan, X. Leng, K. Ji and G. Kuang. RCSHIP: A Dataset Dedicated to Ship Detection in Range-Compressed SAR Data[J]. *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1-5, 2024, Art no. 4004805, doi: 10.1109/LGRS.2024.3366749.

[2] X. Tan, X. Leng, R. Luo, Z. Sun, K. Ji and G. Kuang. YOLO-RC: SAR Ship Detection Guided by Characteristics of Range-Compressed Domain[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 18834-18851, 2024, doi: 10.1109/JSTARS.2024.3478390.

[3] 冷祥光, 谭向东, 计科峰, 匡纲要. SAR 距离压缩域舰船目标检测数据集 (RCSHIP-1.0) [OL]. *雷达学报*, 2025. <https://radars.ac.cn/web/data/getData?newsColumnId=3ba34ef9-045e-4453-b918-2650dfa3cfd1>.

Xiangguang Leng, Xiangdong Tan, Kefeng Ji, Gangyao Kuang. RCSHIP: A Dataset Dedicated to Ship Detection in Range-Compressed SAR Data (RCSHIP-1.0) [OL]. *Journal of Radars*, 2025.

<https://radars.ac.cn/web/data/getData?dataType=SARMulti-domainShipDetectionDatasetEN&pageType=en>.

致谢：感谢国防科技大学电子科学学院谭向东、冷祥光*、计科峰、匡纲要等人对 RCSHIP-1.0 数据集构建与发布做出的贡献。

数据集简介：RCSHIP 数据集旨在构建一套基于 Sentinel-1 卫星数据的距离压缩域舰船检测数据集，推动合成孔径雷达 (Synthetic Aperture Radar, SAR) 在成像前舰船目标检测算法的研究与发展。该数据集由 1580 幅大场景 SAR 距离压缩域数据组成，这些图像进一步被划分为 18322 张子图像，以方便后续的大场景距离压缩域 SAR 图像检测结果展示与分析。数据集覆盖了多种海洋场景，包含大量舰船目标实例，数据总量达 500GB，是目前针对 SAR 距离压缩域舰船检测**规模最大、标注最规范、最早公开可用**的数据集。

RCShip 数据集的基本数据基于欧洲空间局网站发布的 Sentinel-1 卫星 IW (Interferometric Wide swath) 模式 SAR 图像。这些图像均属于不含舰船目标的纯海洋场景。然后, 选取若干舰船切片图像作为检测目标, 将其嵌入到背景图像中, 生成包含舰船目标的大场景合成图像, 并利用嵌入位置信息进行标注, 生成标注文件。接着, 采用逆线性调频变标 (Inverse Chirp Scaling, ICS) 算法和距离向脉冲压缩, 从大场景图像中获得距离压缩数据。最后, 进行数据增强, 并按照预定比例划分训练集和验证集, 其中训练集与验证集的比例为 9:1, 训练集共 13689 张图像, 验证集共 1521 张图像。测试集则通过对筛选出的 580 张包含若干数量舰船的海域大场景 SAR 图像进行细致标注后, 应用 ICS 算法和距离压缩技术生成, 共得到 3112 张距离压缩数据作为 RCShip-1.0 的测试集。

RCShip 数据集的组织形式便于学者使用, 包含七个不同文件目录, 分别为 Tools、JPEGImages、Annotations、JPEGSubImages、SubAnnotations、ImageSets 和 ImagesTiff, 涵盖了图像文件、标注文件、数据集划分文件以及 SAR 距离压缩域复值数据文件等。

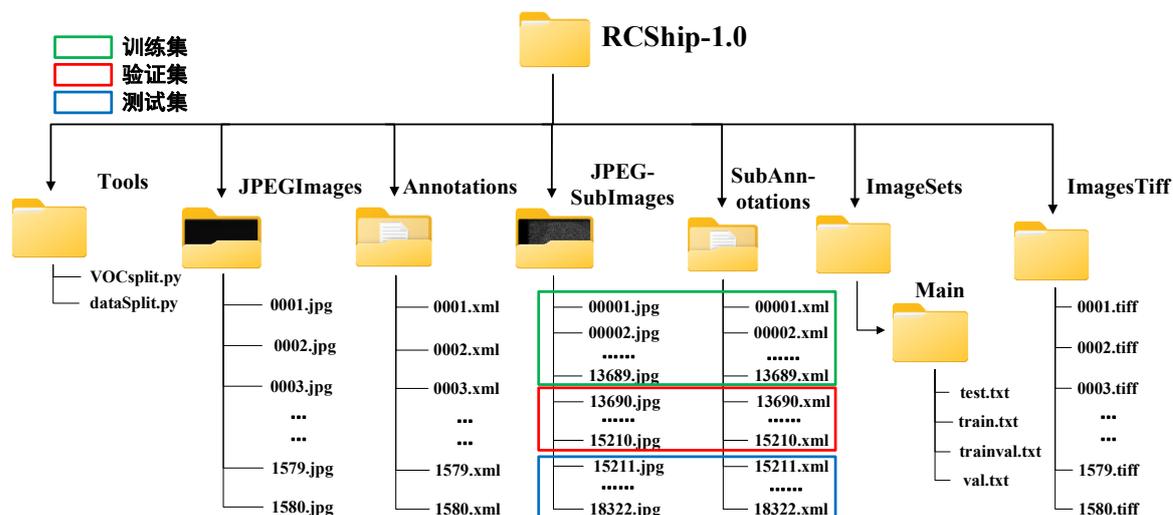


图 1 RCShip-1.0 文件组织结构图

各文件夹内文件具体内容与样本示例如下:

1) Tools 工具文件: 包含两个代码文件, 分别是 ImagesSplit.py 和 VOCsplit.py。ImagesSplit.py 用于将大图像数据分割成指定大小的图像, 而 VOCsplit.py 则用于将图像数据分割成符合 VOC2007 格式的数据集。这些工具文件为数据集的处理和组织提供了便利。

```

1 import sys
2
3 # 文件路径定义
4 xmlfilepath = './sunVOT.txt' # Annotation文件所在位置
5 jpgfilepath = './JPEGSubImages' # 图片文件夹所在位置
6 txtsavepath = './ImageSets/' # ImageSets文件夹下的Main文件夹所在位置
7 output_train = './ImageSets/train'
8 output_val = './ImageSets/val'
9 output_test = './ImageSets/test'
10
11 # 创建输出文件夹
12 os.makedirs(output_train, exist_ok=True)
13 os.makedirs(output_val, exist_ok=True)
14 os.makedirs(output_test, exist_ok=True)
15
16 total_jpg = os.listdir(jpgfilepath)
17
18 # 按文件名排序
19 total_jpg = [os.path.splitext(jpg_file)[0] for jpg_file in total_jpg]
20 total_jpg.sort(key=lambda x: int(x.split('-')[0]))
21
22 # 计算训练集、验证集和测试集的数量
23 total_num = len(total_jpg)
24 trainval_num = 15322
25 train_num = 15210
26
27 # Main文件夹下所有对应的txt文件路径
28 ftrainval = open(os.path.join(txtsavepath, 'trainval.txt'), 'w')
29 ftest = open(os.path.join(txtsavepath, 'test.txt'), 'w')
30 ftrain = open(os.path.join(txtsavepath, 'train.txt'), 'w')
31 fval = open(os.path.join(txtsavepath, 'val.txt'), 'w')
32
33 # 遍历所有图片文件，并读取并写入相应的txt文件中，复制文件路径和文件名
34 for i, jpg_file in enumerate(total_jpg):
35     jpg_name = jpg_file + '.jpg'
36     xml_name = jpg_file + '.xml'
37     print(jpg_name)
38     jpg_path = os.path.join(jpgfilepath, jpg_name)

```

```

1 # coding:utf-8
2 import sys
3 parser = argparse.ArgumentParser()
4 # xml文件路径，根据自己数据集路径修改 xml-放在ImageAnnotations+
5 parser.add_argument('--xml_path', default='xml', type=str, help='input xml_label_path')
6 # 图片文件夹，根据自己数据集路径修改 ImageSets/Main
7 parser.add_argument('--txt_path', default='dataSet', type=str, help='output txt_label_path')
8 opt = parser.parse_args()
9 trainval_percent = 1.0
10 train_percent = 0.9
11 xmlfilepath = opt.xml_path
12 txtsavepath = opt.txt_path
13 totalxml = os.listdir(xmlfilepath)
14 if not os.path.exists(txtsavepath):
15     os.makedirs(txtsavepath)
16
17 num = len(total_xml)
18 list_index = range(num)
19 tv = int(num * trainval_percent)
20 tr = int(tv * train_percent)
21 trainval = random.sample(list_index, tv)
22 train = random.sample(trainval, tr)
23 file_trainval = open(txtsavepath + '/trainval.txt', 'w')
24 file_test = open(txtsavepath + '/test.txt', 'w')
25 file_train = open(txtsavepath + '/train.txt', 'w')
26 file_val = open(txtsavepath + '/val.txt', 'w')
27
28 for i in list_index:
29     name = total_xml[i][:-4] + '\n'
30     if i in trainval:
31         file_trainval.write(name)
32         if i in train:
33             file_train.write(name)
34         else:
35             file_val.write(name)
36     else:
37         file_test.write(name)
38
39 file_trainval.close()
40 file_train.close()
41 file_val.close()
42 file_test.close()

```

图 2 代码文件示例

2) JPEGImages: 包含 1580 个张大场景 SAR 距离压缩域图像。这些图像文件从 0001.jpg 到 1580.jpg 依次编号，每张图像的大小为 24208 × 1512 像素。这些图像构成了 RCSHP 数据集的核心内容，为船舶检测提供了丰富的数据基础。

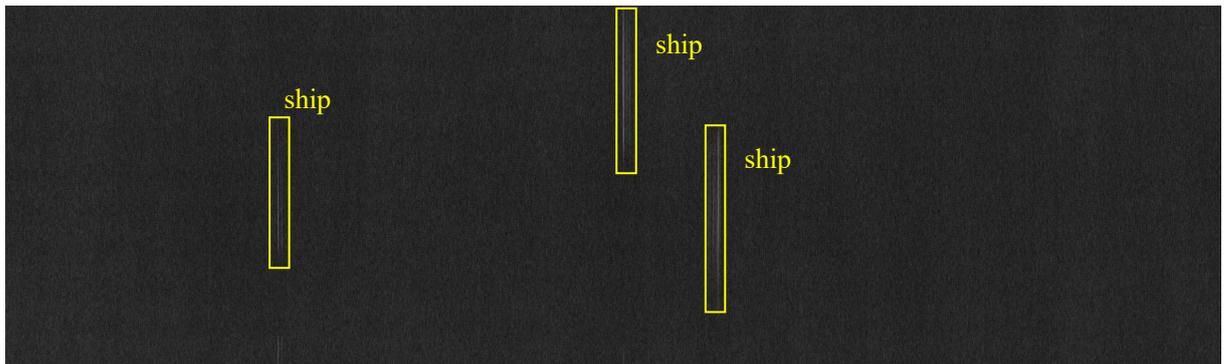


图 3 大场景图像舰船目标样例

3) Annotations 数据标注: 包含 1580 个标注文件。这些标注遵循 PASCAL VOC 标准，为每张图像中的舰船目标提供了精确的位置信息，是模型训练和评估的重要依据

```

▼<annotation>
  <folder/>
  <filename>1000_S1B_IW3_SLC__1SDV_20190906T084108_20190906T084
  ▼<size>
    <width>24208</width>
    <height>1512</height>
    <depth>2</depth>
  </size>
  ▼<object>
    <name>ship</name>
    <pose/>
    <truncated>0</truncated>
    <difficult>0</difficult>
    ▼<bndbox>
      <xmin>23061</xmin>
      <ymin>575</ymin>
      <xmax>23107</xmax>
      <ymin>1027</ymin>
    </bndbox>
  </object>
  ▼<object>
    <name>ship</name>
    <pose/>
    <truncated>0</truncated>
    <difficult>0</difficult>
    ▼<bndbox>
      <xmin>13289</xmin>
      <ymin>90</ymin>
      <xmax>13330</xmax>
      <ymin>558</ymin>
    </bndbox>
  </object>
</annotation>

```

图 4 标注文件样例

4) JPEGSubImages: 包含 18322 张 1024×1024 像素的子图像。这些子图像是通过将 1580 张大规模 SAR 图像进行裁剪得到的，其中 1 到 13689 的序号代表训练集，13690 到 15210 的序号对应验证集，15211 到 18322 的序号则属于测试集。

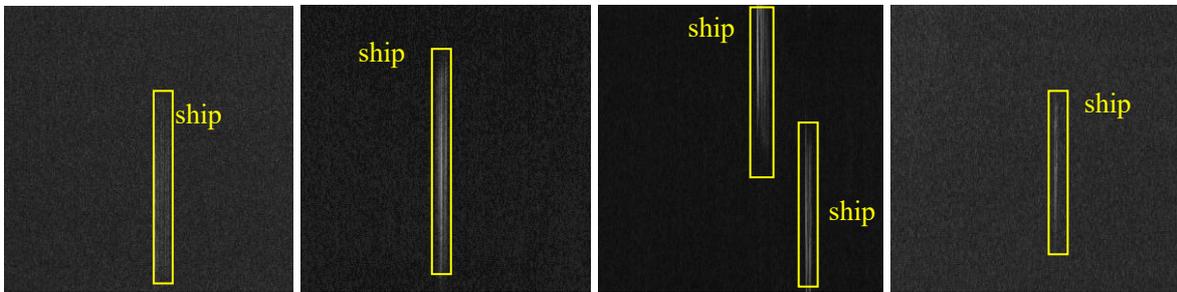


图 5 子图像舰船目标样例

5) SubAnnotations: 包含 18322 个标注文件，从 1.xml 到 18322.xml 依次编号，每个文件对应一张与之同号的子图像。这些标注文件为子图像中的船舶目标提供了相应的标注信息，确保了模型在子图像级别上的训练和评估的准确性

```

▼<annotation verified="yes">
  <folder>Desktop</folder>
  <filename>103_S1B_IW3_SLC__1SDV_20190906T084108_20190906T084136_0
  <path>D:\TXD\openSAR_process\RCSHIP\split\images103_S1B_IW3_SLC__
  ▼<source>
    <database>Unknown</database>
  </source>
  ▼<size>
    <width>1024</width>
    <height>1024</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  ▼<object>
    <name>ship</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    ▼<bndbox>
      <xmin>43</xmin>
      <ymin>1</ymin>
      <xmax>84</xmax>
      <ymax>151</ymax>
    </bndbox>
  </object>
</annotation>

```

图 6 子图像标注样例

6) ImageSets: 组织划分训练集、验证集和测试集的文件。通过这个目录可以方便地管理不同数据集的划分情况，便于实验的重复和比较

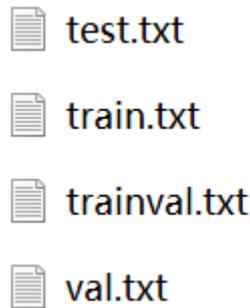


图 7 数据集划分示意

7) ImagesTiff: 包含从 0001.tiff 到 1580.tiff 编号的 1580 张大规模 SAR 距离压缩图像。这些图像文件将用于未来的复值 SAR 研究。



图 8 复数据示意