

FSAR-Cap: 大规模细粒度 SAR 图像描述数据集使用说明

数据主编：张腊梅（电子与信息工程学院，哈尔滨工业大学，哈尔滨，150001）

下载说明：FSAR-Cap 大规模细粒度 SAR 图像描述数据集所有权归哈尔滨工业大学电子与信息工程学院所有，《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，欢迎联系张腊梅 (lmzhang@hit.edu.cn) 或张金琪 (24b905016@stu.hit.edu.cn)。

本数据集参考文献与引用格式：

[1] 张金琪，庄迪，张腊梅，等. DGS-CapNet: 基于空间-频率感知的 SAR 图像描述模型[J]. 雷达学报（中英文），待出版. doi: 10.12000/R25250

ZHANG Jinqi, ZHUANG Di, ZHANG Lamei, et al. DGS-CapNet: A Spatial-Frequency-Aware Model for SAR Image Captioning[J]. Journal of Radars, in press. doi: 10.12000/R25250

数据集简介：FSAR-Cap 数据集旨在为 SAR 图像语义理解与跨模态建模构建一

套具备精细语义描述能力的图像-文本对照语料,推动 SAR 图像自动解译、图像描述生成(Captioning)以及遥感多模态模型的发展。该数据集基于 FAIR-CSAR 的目标检测数据集构建,共包含 14,480 张 SAR 图像和 72,400 条配套描述文本。FSAR-Cap 采用两阶段标注方式:先利用检测结果和空间位置信息,通过多种模板自动生成基础描述;随后再结合人工校验与语言大模型润色。最终,每幅图像都会生成 5 条风格不同、信息互补的描述语句,文本内容涵盖目标类型、数量、位置关系、外形特征等,以下为数据集的统计分析结果:



作为目前首个面向 SAR 图像、具备细粒度描述层级的大规模语义标注数据集，FSAR-Cap 不仅提升了 SAR 图像语义表达质量，也为 SAR 领域的图像 captioning、遥感视觉语言模型训练、多模态推理、SAR - 自然语言对齐研究提供

了统一且高质量的数据基准，为 SAR 自动化解译与智能语义理解技术体系的进一步发展奠定了基础。

数据集包含两部分，第一部分 all_images 中存放了所有的原始图像，第二部分为 captions.json 中存放着对图片的标注内容，create_input_files.py 文件为数据的预处理程序，示例如下：



文本标注格式如下：

```
{"filename":  
"GF3_SAY_SL_017051_E118.7_N32.1_20191105_L1A_DH_L10000000001_01024_09945.png",  
"imgid": 14108,  
"split": "train",  
"sentences": [  
  {"tokens":["a","sar","image","reveals","an","open","water","body","with","five","large","elongated","  
cargo","ships","located","in","the","top","right","center","and","bottom","left","regions"],  
"raw": "A SAR image reveals an open water body with five large, elongated cargo ships located in the  
top-right, center, and bottom-left regions.",  
"imgid": 14108, "sentid": 70540},  
  {"tokens":["the","image","displays","a","continuous","water","surface","with","ships","crossing","thro  
ugh","the","top","right","center","and","bottom","left","areas"],  
"raw": "The image displays a continuous water surface, with ships crossing through the top-right, center,  
and bottom-left areas.",  
"imgid": 14108, "sentid": 70541},  
  {"tokens":["the","scene","includes","a","body","of","water","spanning","across","the","image","with",  
"ships","positioned","in","the","top","right","center","and","bottom","left","sections"],  
"raw": "The scene includes a body of water spanning across the image, with ships positioned in the top-
```

```
right, center, and bottom-left sections.",
"imgid": 14108, "sentid": 70542},
{"tokens":
["five", "large", "cargo", "ships", "are", "visible", "in", "the", "top", "right", "center", "and", "bottom", "left", "re-
gions", "of", "the", "open", "water", "body"],
"raw": "Five large cargo ships are visible in the top-right, center, and bottom-left regions of the open
water body.",
"imgid": 14108, "sentid": 70543},
{"tokens":
["the", "water", "surface", "extends", "over", "the", "image", "with", "ships", "flowing", "through", "the", "top",
"right", "center", "and", "bottom", "left", "parts"],
"raw": "The water surface extends over the image, with ships flowing through the top-right, center, and
bottom-left parts.",
"imgid": 14108, "sentid": 70544}],
"sentsids": [70540,70541,70542,70543,70544]
}
```

JSON 对象表示一张遥感图像及其描述信息，主要包含以下字段：

1、filename

含义：图像文件的名称。

作用：用来唯一标识图像文件，可以用于加载或查找图像。

2、imgid

含义：图像的唯一 ID，这里是 14108。

作用：在数据库或数据集中标识该图像，便于索引和关联。

3、split

含义：数据集划分标记，这里是 "train"。

作用：标识该图像属于训练集还是测试集。

4、sentences

含义：对应的自然语言描述列表，每个元素描述图像中的场景。

每个元素包含：

tokens：分词后的句子。例如 ["a", "sar", "image", "reveals", "an", "open",

"water", "body", ...]。

raw: 完整的原始描述句子。例如 "A SAR image reveals an open water body with five large, elongated cargo ships located in the top-right, center, and bottom-left regions."

imgid: 对应父图像的 imgid，用于关联句子和图像。

sentid: 每条描述的唯一 ID，例如 70540，便于索引和引用。

5、sentids

含义：列出所有描述句子的 ID [70540, 70541, 70542, 70543, 70544]。

作用：快速获取该图像下所有句子的 ID，便于批量处理或索引。

数据预处理文件 `create_input_files.py`:

该脚本可以把本数据集一键转成图像描述模型能直接使用的标准格式，参数修改：`--karpathy_json_path` 指向 json 标注文件，`--image_folder` 指向图片大文件夹，`--output_folder` 输出文件夹，`--captions_per_image` 保持 5，每张图片队形的描述数量，`--min_word_freq` 设 3~5 控制词表大小，`--max_len` 最大长度。运行后得到 TRAIN/VAL/TEST 三个 hdf5 图片文件+对应的 CAPTIONS、CAPLENS 和 WORDMAP.json。

致谢：感谢中国科学院空天信息创新研究院目标认知与应用技术重点实验室吴有明、刁文辉、索玉玺、孙显等人构建的 FAIR-CSAR-V1.0 数据集，为本研究提供了关键数据基础。在此基础之上，我们进一步开展了文本标注与扩展工作，以形成本数据集 (https://radars.ac.cn/web/data/getData?dataType=FAIR_CSAR.)。