



蝴蝶检测和细粒度识别竞赛 - 答辩

答辩人：赵祈杰

队伍名：VDIG (A299)

单位：北京大学计算机科学技术研究所

实验室：VDIG实验室

Qijie Zhao

Visual Data Interpreting and Generation Lab
Institute of Computer Science & Technology, Peking University

Homepage: qijiezhao.github.io

Mail: zhaoqijie@pku.edu.cn



一. 队伍介绍

- 队名：VDIG (A299)
- 姓名：赵祈杰，倪烽，王勇涛
- 单位：北京大学，计算机科学技术研究所，VDIG lab
- 成绩：检测（第二），识别（第二）



赵祈杰
北京大学计算机科学技术研究所
硕士二年级学生
研究方向：深度学习与计算机视觉，
目标检测，语义分割，行为识别，
迁移学习等



倪烽
北京大学计算机科学技术研究所
实习生
研究方向：深度学习与计算机视觉，物
体识别与目标检测



导师：王勇涛
北京大学计算机科学技术研究所
副研究员
研究方向：复杂文档图像理解，深度学
习与计算机视觉



二. 赛题理解

➤ 问题描述:

- **目标检测和物体细粒度分类** (object detection, fine-grained classification)
- 计算机视觉, 图像内容理解, 机器学习领域的前沿课题
- 关于生物和视觉算法的交叉科学研究

蝴蝶模式照



蝴蝶生态照





二. 赛题理解

➤ 相关研究:

◆ 目标检测(按执行策略划分):

- 1, 传统算法: Haar+adaboost, DPM等
- 2, 深度学习-单步法: YOLO(v1, v2, v3), SSD, RetinaNet, RefineDet等
- 3, 深度学习-双步法: Faster R-CNN, R-FCN, Mask R-CNN, FPN等
- 4, 深度学习-多步法: Cascade R-CNN等

◆ 细粒度分类(按方法类型划分):

- 1, 使用通用CNN(深度卷积神经网络)直接进行细粒度分类: 所有分类网络皆可使用
- 2, 基于部位检测与对齐: Part-based RCNN, Pose Normalized Nets, PS-CNN, Deep LAC等
- 3, 基于网络集成的方法: 划分数据集为相似的子集或直接使用多个神经网络来提高, 双线性融合特征等。
- 4, 使用注意力模型定位区分性强的区域: Two-level attention, FCN attention等



二. 赛题理解

➤ 相关benchmark:

◆ 目标检测:

- 1, Pascal VOC, 通用
 - 2, MS-COCO, 通用
 - 3, KITTI, 车和行人等
 - 4, Wider Face, 人脸等
- ...

◆ 细粒度分类:

- 1, Stanford Dogs, Cars dataset.
 - 2, Oxford Flowers Dataset.
- ...



三. 数据观察

- 竞赛提供的训练集包括：**蝴蝶生态照片集**，**蝴蝶模式照片集**。
 - **生态照**：在野外由相机拍得。每张照片至少包含一只蝴蝶，背景变化大，蝴蝶相对较小且很难被发现。**721**张图，有位置和类别的标注信息。



存在图像间的域差异

- **模式照**：扫描中国蝶类志的蝴蝶图片而得。**4270**张/**1176**类，涵盖生态照的所有种类。每图只含一只蝴蝶。有多级类别信息（科属种等），分类精确到种。无背景。
- 问题解析：
 - 细粒度识别问题（fine-grained classification）
 - 少样本学习问题（few-shot learning）
 - 域适应问题（domain adaptation）



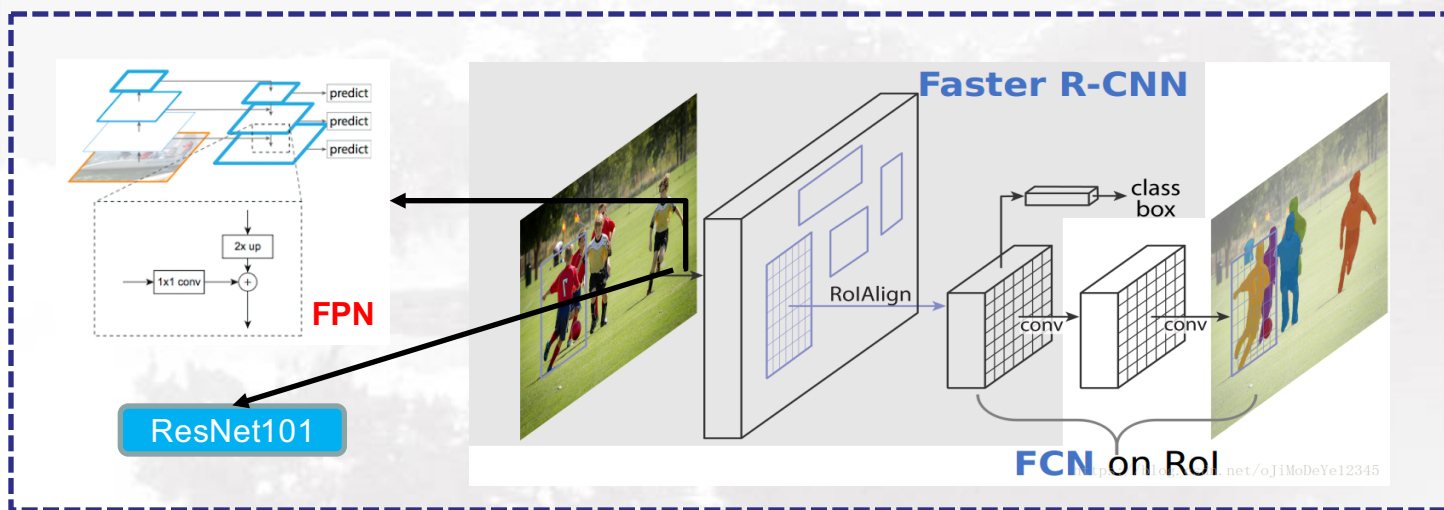
四. Task1方案

➤ 任务1：蝴蝶位置检测

● 从**问题分析**到**方案选择**:

- 1, 目标检测问题的定位精准要求: 例如会用IoU来计算定位精准度
- 2, 存在小目标蝴蝶的检测情况: 图中很小的蝴蝶可能在降采样的特征图上丢失太多信息
- 3, 训练数据集有限: 可能让模型收敛效果差

Feature Pyramid Network meets RoI align: Enhancing Detection for Small Objects





四. Task1方案

□ 解决方案

- ◆ Mask R-CNN - 检测框架
- ◆ ResNet101 - 基础网络
- ◆ FPN - 特征金字塔网络
- ◆ MS-COCO - 预训练
- ◆ Multi-scale train - 多尺度训练
- ◆ Early-stop – 训练策略
- ◆ Momentum+SGD – 优化策略



四. Task1实验

➤ 实验平台:

- OS: Linux Ubuntu 16.04
- Memory、GPU: 64G、2 x Titan 1080Ti
- Platform、DL framework: Keras2.0.5、TensorFlow1.3.0、Anaconda3.6.3

➤ 实验步骤和结果:

- 将721张图像按9:1分成训练集（654）和验证集（67），其中保证每张验证集图像只有一只蝴蝶（模拟测试集情况，以方便验证中间结果，调参和可视化）。然后我们在训练集上训练，验证集上做验证。
- 最终在67张图的验证集上评测，IoU=0.5时的average precision(AP)达到0.9403，IoU=0.7时AP达到0.8955。
- 将721张图作为训练集，然后在竞赛委员会提供的测试集上得到的结果：mIoU=0.8487，在最后的榜单上，排名第二。



五. Task2方案

➤ 任务2：蝴蝶细粒度识别

- 训练集图像和测试集图像均为生态照，但是训练集图像极少，且存在变化差异较大的背景。

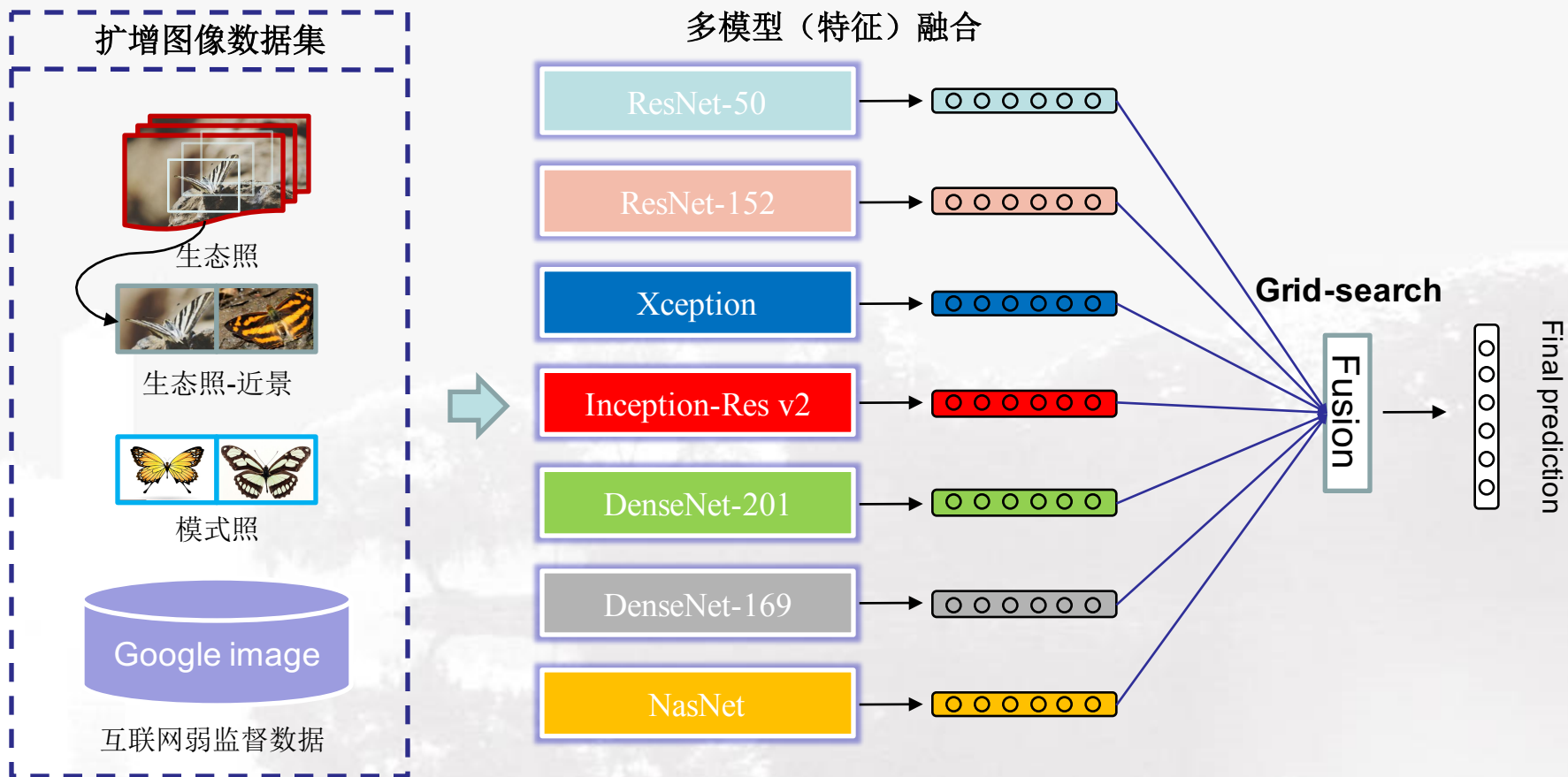
➤ 问题分析：

- 训练样本较少（721张图/94类，平均每类不足8张图），考虑借用模式照
- 背景范围大，可以考虑用检测器切图
- 类间差距细微，考虑多模型（特征）融合



五. Task2方案

DEME: Data Expanding and Model Ensemble for Butterfly fine-grained recognition





五. Task2方案

单模型结果

- 生态照按9:1分成训练集图像 (S_train) 和验证集图像 (S_val)

1. [Baseline]

用ResNet50训练S_train, 并在S_val上验证, 得到准确率0.73。

2. [训练模式照是否能带来提升]

从模式照挑选S_train中出现过类别的图像, 组成新的训练集, 用ResNet50训练, 并在S_val上做验证, 准确率0.80。

3. [训练切割图是否能带来提升]

用目标检测器拿到S_train的近景切割图, 组成新的训练集, 用ResNet50训练, 并在S_val上做验证, 准确率0.85。

4. [验证切割图是否能带来提升]

用目标检测器拿到S_val的近景切割图, 用ResNet50把上一步的训练集训练后在S_val上面验证, 得到准确率0.94。

5. [验证集结果的融合]

1. 调整 (4) 和 (5) 的融合权重, 得到1:2的最好比例, 准确率0.96

| ResNet50 | 生态图 | Google | 模式图 | 切割图 | 验证集 | 分类精度 |
|----------|-----|--------|-----|-----|----------------|------|
| √ | √ | | | | S_val | 0.73 |
| √ | √ | √ | | | S_val | 0.71 |
| √ | √ | | √ | | S_val | 0.80 |
| √ | √ | | √ | √ | S_val | 0.85 |
| √ | √ | | √ | √ | Scrop_val | 0.94 |
| √ | √ | | √ | √ | S_val+Scro_val | 0.96 |



五. Task2实验

多模型结果

| 网络模型 | 输入大小 | 验证集精度 | 验证集crop精度 | 融合比例 | 最后精度 |
|-------------|------|-------|-----------|------|--------------|
| ResNet50 | 224 | 0.847 | 0.945 | 1:2 | 0.958 |
| ResNet152 | 224 | 0.861 | 0.930 | 1:1 | 0.944 |
| Xception | 299 | 0.875 | 0.889 | 1:2 | 0.902 |
| Inc_Res v2 | 299 | 0.875 | 0.930 | 1:1 | 0.930 |
| DenseNet201 | 224 | 0.833 | 0.861 | 1:1 | 0.917 |
| DenseNet169 | 224 | 0.861 | 0.875 | 1:1 | 0.917 |
| NasNet | 331 | 0.875 | 0.847 | 2:1 | 0.917 |

- ◆ 由于比赛有运行时间限制，故我们最后选择了**ResNet50, ResNet152, InceptionResNetV2, NasNet**四个网络在测试集上做模型融合。这四个模型在验证集（val+val_crop）上融合后的精度达到了**0.9722**（72张仅错2张）
- ◆ 测试集的最后精度是**0.89**，排名第二。（测试集和验证集存在一定的分布偏差，且验证集数量较少，导致我们调整的参数存在一定的过拟合）



六. 创新点

1, 目标检测方面, 我们先分析数据, 以及数据标注情况, 然后选择了当前最好检测器之一的Mask R-CNN作为基准, 结合ResNet101, FPN等架构来提升小目标难目标检测能力。

2, 对于Few shot learning和Fine-grained classification结合的问题, 我们提出了一个扩增数据集的方案:

- 原图 (目标+背景)
- 切割图 (等同于引入了注意力机制)
- 生态图 (没有背景的目标图)
- 爬取谷歌图像 (引入了弱监督的数据)

并且做了可靠的实验, 验证了部分扩增数据集的作用。

3, 用Late-fusion进行特征 (模型) 融合, 并使用grid-search找到最好的线性权重, 有效地提升了分类精度。



七. 速度与效率

➤ 验证集表现 (测试显卡: Titan 1080Ti, Batch_size=1)

• 蝴蝶检测 (Butterfly Detection) :

- 检测精度: IoU=0.5, mAP=0.9403; IoU=0.7, mAP=8955
- 检测速度: 6.5张图/秒

• 蝴蝶分类 (Butterfly Fine-grained Classification) :

单模型 (ResNet152)

- 分类精度: 0.861
- 分类速度: 40张图/秒

多模型 (ResNet50, ResNet152, Inc_ResV2, NasNet) + 双验证

- 分类精度: 0.972
- 分类速度: 1张图/秒



八. 可视化结果





九. 困难与收获

➤ 困难:

数据集特别小，很容易造成过拟合。但是深度学习技术在实际应用中却会经常面对这样的挑战。

如何利用模式照，是一项挑战，模式照具有详细的标注信息，难点还包括如何把知识迁移过来。

➤ 收获:

合理的分配验证集，能决定最后结果的上限。验证集太大，会挤压训练集，导致训练出来的结果不够鲁棒。验证集太小，又会导致验证集上调得的参数过拟合。

深度学习 \approx 数据 + 模型，用深度学习解决问题，就从数据和模型两方面入手。



十. 总结

- ◆ 即斯坦福（认）车，百度识狗，京东识猪等挑战之后，本次竞赛引入了蝴蝶的检测和细粒度识别两项任务。本次竞赛对技术落地有更多考量，所以提供的数据集也和常见现实情况类似，极具挑战。
- ◆ 蝴蝶的检测+识别任务，可以泛化到更多的动植物。这些技术将给很多行业带来便利，甚至带来革命性的改变。



谢谢!

提问环节