

Clustered Object Detection in Aerial Images

笔记

1. 论文解析

1.1 研究动机

1.1.1 现代检测器的问题

1. 现代检测器难有效地利用外观信息来区分物体与周围背景或相似物体
2. 现代检测器面临两个挑战:(1)目标相对于图像通常具有较小的尺度;(2)目标在整个图像中的分布通常是稀疏且不均匀的。
3. 相关工作中尺度问题的解决忽略了目标的稀疏性，它们在执行检测时效率低下。
因此，在物体稀疏甚至没有物体的区域上，大量的计算资源被低效地应用

1.1.2 解决思路

1. 提高检测效率的一种直观的方法是将检测器集中在这些有大量对象的聚类区域上。
2. 将目标和聚类检测集成在一个统一的框架中来解决上述两个挑战

1.2 方法解析

1.2.1 Clustered object Detection (ClusDet) network

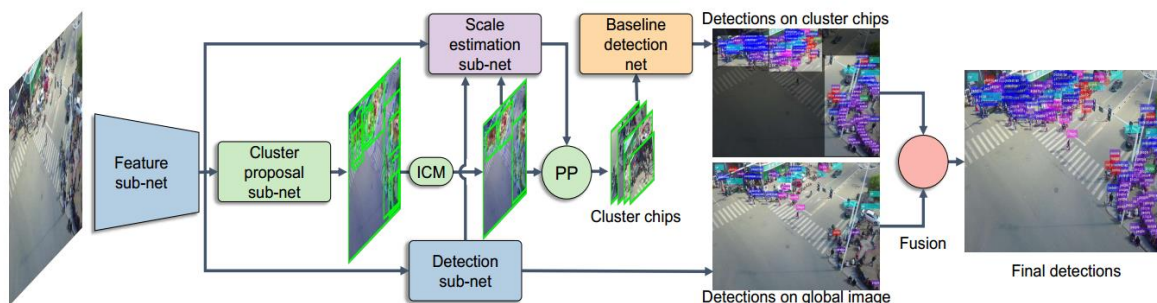


Figure 2: **Clustered object Detection (ClusDet) network**. The ClusDet network consists of three key components: (1) a cluster proposal sub-net (CPNet); (2) a scale estimation sub-net (ScaleNet); and (3) a dedicated detection network (DetecNet). CPNet serves to predict the cluster regions. ScaleNet is to estimate the object scale in the clusters. DetecNet performs detection on cluster chips. The final detections are generated by fusing detections from cluster chips and global image. The details of ICM (iterative cluster merging) and PP (partition and padding) are given in Section 3.

1. 航拍图像的检测包括三个阶段:聚类区域提取、聚类芯片精细检测和检测结果融合。

1.2.2 Cluster Region Extraction

1.2.2.1 Cluster Proposal Sub-network (CPNet)

1. CPNet 在航空图像的高级特征地图上工作，旨在预测集群的位置和规模
2. CPNet 将特征提取主干的高级特征映射作为输入，并利用两个子网分别进行回归和分类。
3. CPNet 用于提出聚类的候选区域
4. CPNet 的学习是有监督的过程

1.2.2.2 Iterative Cluster Merging (ICM)

1. 产生背景：我们观察到 CPNet 产生的初始簇是密集而杂乱的。这些密集而杂乱的聚类区域由于其高重叠和大尺寸，难以直接用于精细检测
2. 解决问题：我们提出了一个简单而有效的迭代聚类合并(ICM)模块来清理聚类

$\mathcal{B} = \{B_i\}_{i=1}^{N_{\mathcal{B}}}$ 表示 CPNet 检测到的 NB 类边界框集合

$\mathcal{R} = \{R_i\}_{i=1}^{N_{\mathcal{B}}}$ 表示相应的类分类得分

τ_{op} 表示合并的阈值

N_{max} ; 最大合并簇数

Algorithm 1: Iterative Cluster Merging (ICM)

Input: Initial cluster bounding boxes $\mathcal{B} = \{B_i\}_{i=1}^{N_{\mathcal{B}}}$, initial cluster scores $\mathcal{R} = \{R_i\}_{i=1}^{N_{\mathcal{B}}}$, threshold τ_{op} and maximum number of merged clusters N_{max} ;

Output: Merged clusters $\mathcal{B}' = \{B'_i\}_{i=1}^{N_{\mathcal{B}'}}$;

begin

$\mathcal{B}' \leftarrow \mathcal{B}$;

while $|\mathcal{B}'| > N_{\text{max}}$ **do**

$\mathcal{B}', \mathcal{R}' \leftarrow \text{NMM}(\mathcal{B}, \mathcal{R}, \tau_{\text{op}})$

if $|N'_{\mathcal{B}}| = |N_{\mathcal{B}}|$ **then**

break;

else

$\mathcal{B} \leftarrow \mathcal{B}'; \mathcal{R} \leftarrow \mathcal{R}'$;

end

end

$\mathcal{B}'' \leftarrow \{\}$;

for $i \leq \min(N_{\text{max}}, |\mathcal{B}'|)$ **do**

$\mathcal{B}'' \leftarrow \mathcal{B}'' \cup \{B'_i\}$;

end

$\mathcal{B}' \leftarrow \mathcal{B}''$;

end

目前的理解是，挑出一个置信度最高的检测框，找出某些检测框，与其相似度高于阈值的，进行一个合并（目前的理解是把它们删掉，这样直到挑完后的检测框数量剩下小于等于 N_{max} 的时候，停止），

合并算法这里用到 NMM,就是相似度大于阈值的和谐掉？

1.2.3 Fine Detection on Cluster Chip

规模估计子网络(ScaleNet)来估计芯片中物体的规模，避免了物体的极端规模降低检测性能。基于估计的尺度，ClusDet 在每个芯片上执行分区和填充(PP)操

作进行检测。

1.2.3.1 Scale Estimation Sub-network (ScaleNet)

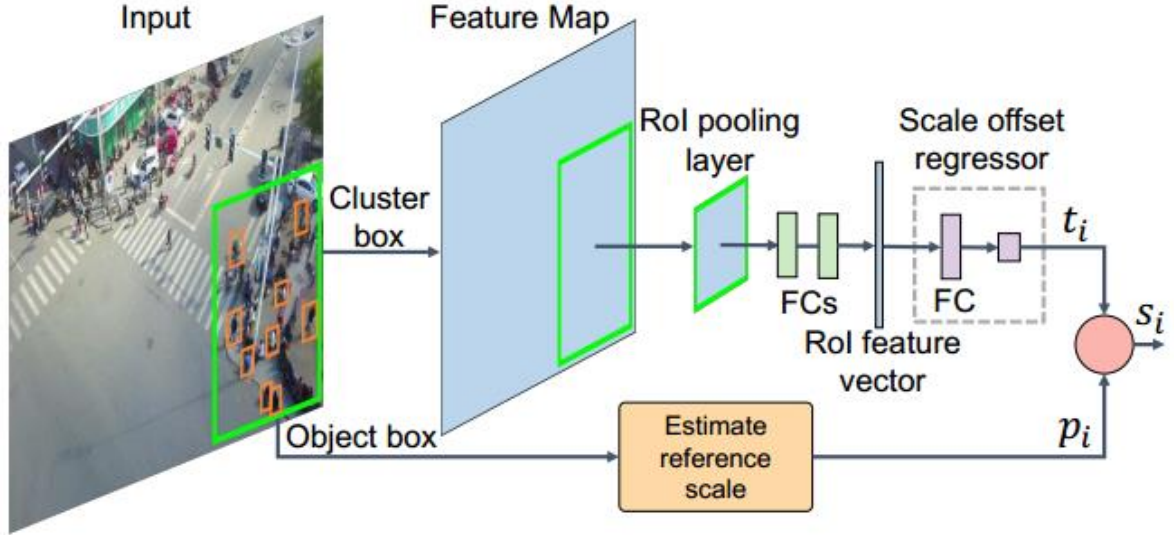


Figure 4: The architecture of the scale estimation network (ScaleNet). The cluster detections are projected to feature map space. Each cluster is pooled into a fixed-size feature map and mapped into a feature vector by fully connected layers (FCs). The network has an output per cluster, *i.e.*, the scale regression offset.

参数表示

p_i 表示检测物体的参考比例尺

s_i^* 表示第 i 个聚类中 groundtruth 框的平均规模

$$\mathcal{L}(\{t_i\}) = \frac{1}{M} \sum_i^M \ell_{\text{reg}}(t_i, t_i^*)$$

损失函数定义

$t_i^* = (p_i - s_i^*)/p_i$ 代表 相对比例偏移

$t_i = (p_i - s_i)/p_i$ 代表 估计相对尺度偏移

s_i 估计的规模

这个就是把聚类检测映射到投影空间，通过神经网络进行特征向量，结合把目标框输入而输出的估计参考比例尺，输出估计规模 S_i

1.2.3.2 Partition and Padding (PP)

S_i^{in}

超过一定范围，则按比例填充集群，否则分割集群（分割后两块区域相等），这里范围设置为[70,280].

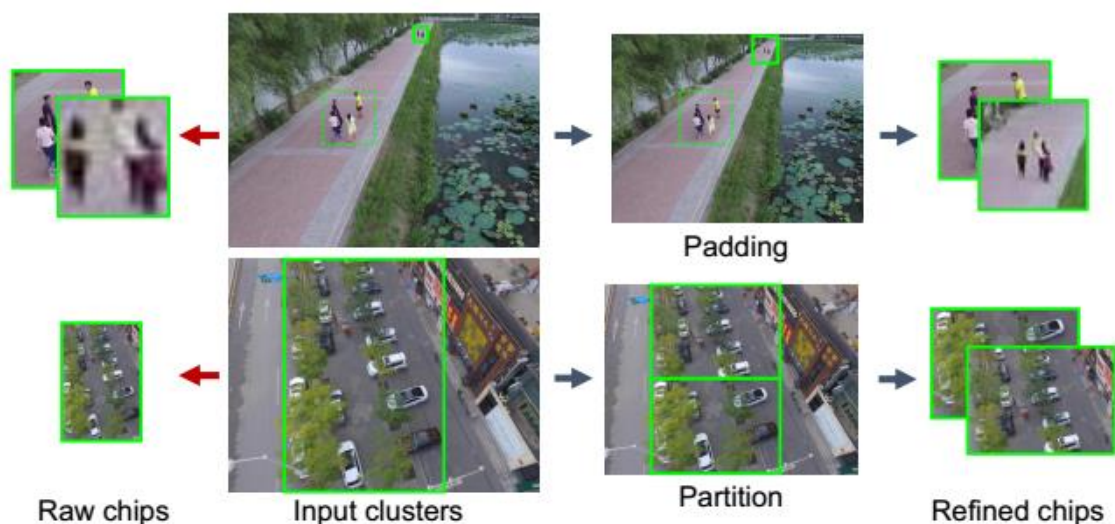
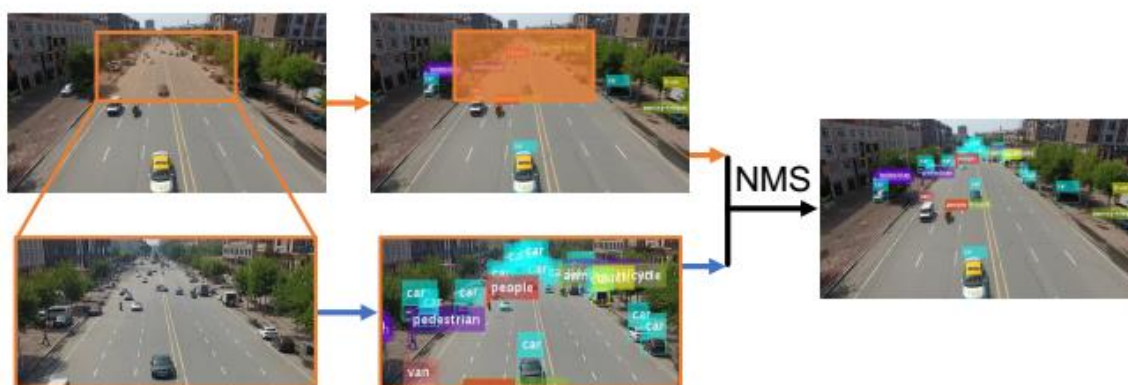


Figure 5: Illustration of the partition and padding (PP) process. The raw chips and refined chips are the input of detector without and with using PP, respectively.

1.2.4 Final Detection with Local-Global Fusion

在重新缩放集群芯片后，一个专用的基线检测网络(DetecNet)进行精细目标检测。

将集群芯片的局部检测结果与整幅图像的全局检测结果进行标准 NMS 后处理，得到航拍图像的最终检测结果



值得注意的是，现有的任何现代检测器都可以用于全局检测

1.3 讨论和思考

1.3.1 Clustered Object Detection in Aerial Images 优点

1. 提出了一种新颖的 clusterdet 网络，同时解决航空图像中目标检测的规模和稀疏性挑战。
2. 提出了一种有效的 ScaleNet 算法，以解决聚类芯片中尺度不均匀的问题，实现更好的精细检测
3. 在 VisDrone[37]、UAVDT[8]、DOTA[33]三个具有代表性的航空图像数据集上，以较少的计算量实现了最先进的性能。

1.3.2 讨论

1.3.3 思考

1.4 实验设计

1.4.1 Compared Methods

1. 在所有数据集上比较了我们的 clusterdet 方法和均匀图像分区(EIP)方法。
2. 将我们的方法与所有数据集上具有代表性的最先进方法进行了比较。

Table 3: The detection performance of the baselines and proposed method on the UAVDT [8] dataset.

Methods	backbone	#img	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
R-FCN[5]	ResNet50	15,069	7.0	17.5	3.9	4.4	14.7	12.1
SSD[23]	N/A	15,069	9.3	21.4	6.7	7.1	17.1	12.0
RON[16]	N/A	15,069	5.0	15.9	1.7	2.9	12.7	11.2
FRCNN[27]	VGG	15,069	5.8	17.4	2.5	3.8	12.3	9.4
FRCNN[27]+FPN[20]	ResNet50	15,069	11.0	23.4	8.4	8.1	20.2	26.5
FRCNN[27]+FPN[20]+EIP	ResNet50	60,276	6.6	16.8	3.4	5.2	13.0	17.2
ClusDet	ResNet50	25,427	13.7	26.5	12.5	9.1	25.1	31.2

Table 4: The detection performance of the baselines and proposed method on DOTA [33] dataset.

Methods	backbone	#img	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
RetinaNet[21]+FPN[20]+EIP	ResNet50	2,838	24.9	41.5	27.4	9.9	32.7	30.1
RetinaNet[21]+FPN[20]+EIP	ResNet101	2,838	27.1	44.4	30.1	10.6	34.8	33.7
RetinaNet[21]+FPN[20]+EIP	ResNeXt101	2,838	27.4	44.7	29.8	10.5	35.8	32.8
FRCNN[27]+FPN[20]+EIP	ResNet50	2,838	31.0	50.7	32.9	16.2	37.9	37.2
FRCNN[27]+FPN[20]+EIP	ResNet101	2,838	31.5	50.4	36.6	16.0	38.5	38.1
ClusDet	ResNet50	1,055	32.2	47.6	39.2	16.6	32.0	50.0
ClusDet	ResNet101	1,055	31.6	47.8	38.2	15.9	31.7	49.3
ClusDet	ResNeXt101	1,055	31.4	47.1	37.4	17.3	32.0	45.4

1.4.2 Evaluation Metric

使用 AP、AP50 和 AP75 作为度量精度的指标。

1. AP 是通过对所有类别取平均值来计算的。
2. AP50 和 AP75 在所有类别中按单个 IoU 阈值 0.5 和 0.75 计算。

1.4.3 Ablation Study (消融实验)

Methods	backbone	AP	AP_{50}	AP_{75}
RetinaNet[21]+FPN[20]	ResNet50	13.9	23.0	14.9
RetinaNet[21]+FPN[20]	ResNet101	14.1	23.4	14.9
RetinaNet[21]+FPN[20]	ResNeXt101	14.4	24.1	15.5
FRCNN[27]+FPN[20]	ResNet50	21.4	40.7	19.9
FRCNN[27]+FPN[20]	ResNet101	21.4	40.7	20.3
FRCNN[27]+FPN[20]	ResNeXt101	21.8	41.8	20.1
FRCNN[27]+FPN[20] ★	ResNeXt101	28.7	51.8	27.7
FRCNN[27]+FPN[20]+EIP	ResNet50	21.1	44.0	18.1
FRCNN[27]+FPN[20]+EIP	ResNet101	23.5	46.1	21.1
FRCNN[27]+FPN[20]+EIP	ResNeXt101	24.4	47.8	21.8
FRCNN[27]+FPN[20]+EIP★	ResNeXt101	25.7	48.4	24.1
ClusDet	ResNet50	26.7	50.6	24.7
ClusDet	ResNet101	26.7	50.4	25.2
ClusDet	ResNeXt101	28.4	53.2	26.4
ClusDet ★	ResNeXt101	32.4	56.2	31.6

1. 验 EIP 影响
2. 验 Cluster Detection 影响
3. 验 Effect of Scale Estimation 影响
4. 验超参数 TopN.影响