

日期：2022.12.31

报告人：李号元

记录人：李号元

工作汇报：Wasserstein Adversarial Alignment and Graph Refinement for Cross-Modal Visual Geo-localization

任务：给定一个SAR查询图，在光学底图中找到对应位置的patch。

challenges：

1. 光学和SARpatch之间的模态差异，导致特征难以度量。
2. 任意方位SAR图像和光学patches之间存在位置偏移，直接进行检索的难度高。

solutions：

1. 提出coarse-to-fine的基于检索框架解决跨模态图像匹配问题
2. 针对异源图像模态差异，提出最小化模态分布间Wasserstein距离的对抗训练方法以减少模态差异。
3. 针对任意位置SAR图像的匹配问题，结合参考图地理信息提出基于图结构的可学习检索优化模块。

问题：

王宇轩：

1. triplet本身作为度量学习的损失可以对样本进行度量，为什么还要用对抗学习？

答：在所有样本之间没有模态的差异情况下，triplet可以在一定程度上将样本映射到度量空间，但是如果检索的图片之间存在模态差异，而triplet需要度量不同域之间的距离，而不同域之间特征分布是不一样的，此时难以进行有效度量，需要用对抗学习将两个不同域的特征分布先映射到相同的特征表示，再进行度量可以获得更好的效果。计算两个域嵌入特征之间的wasserstein discrepancy是700，说明了两个域之间存在特征分布差异，通过对抗的方式减少Wasserstein距离，可以有效减少模态之间的差异，最终实验表明检索精度也会提高。

张瑞祥：

1.在失败案例中，图像可能是检索到了相近场景光学图，这种有没有考虑用临近的图像作为正样本监督？

答：vigor的semi-loss就是使用了临近的图像作为正样本，但是该方法在这个问题的实验精度没有提高，同时会增加训练成本。

2.会不会存在临近的参考图像作为负样本的情况？

答：因为在训练中使用了batch之间计算正负样本，每一对图像中选择SAR图像对应重叠度最高的光学参考图作为匹配对，由于数据量很大，出现临近的参考图的概率比较低，几乎可以忽略作为负样本的情况。

3.水体这一类不包含信息的图像是否要进行处理？水体图像的噪声导致几乎无法进行有效的匹配。

答：在现实情况中，数据集会包含一定的噪声样本，这种噪声会导致最终精度的下降，但占比较少，因此没有特意去除。

4.图像的尺寸有点小，难提供检索需要的建筑信息。

答：这个确实是个问题，当时为了保证数据集的大规模，就选择减少图像的尺寸来保证图像patch数量。

余老师：

1. 如果要解决无人机SAR图像定位问题，SAR的图像尺寸太小，SAR的成像机理不会出现这么小的图像，和遥感现实应用脱离。
2. 要解决的背景问题和提出的方法之间关联性不足，地理定位的相关工作和要解决的问题关联不大，主要是SAR和光学的patch匹配问题在实际应用上立不住。街景和卫星图的匹配是有实际应用的，因为现实中有直接用街景成像的应用，但是实际上并不会出现像这个问题里面这么小的SAR图像patch。
3. 用GNN优化检索排序的模块中，即使是top10的精度也很低，再对这个精度很低的初始排序进行优化的意义不大。

答：后续需要对不同的top-K进行实验，尽量选择更大的K保证初始检索的精度足够高。

4. 对抗训练模块的逻辑没有解释清楚，图表的解释和逻辑要链接恰当。

5. 难点中要突出位置偏差这个点，并提出对应的解决问题，针对这个点的具体解决操作。但是在工作中方法并没有和这个难点很好地关联上。
6. 图结构边构成的实际意义是什么？如果没有具体的构图的物理意义，这个图的模块很容易立不住，为什么图模型更好？我用常规mlp/cnn对匹配对做卷积是不是可以得到同样的结果？

答：图边的结构是由gps的 (x, y) 距离计算得到K临近的邻居连接边。后续实验应该对比GNN网络和MLP直接预测之间的精度。

杨老师：

1. 这个工作应该更专注于解决图像patch匹配问题本身，扯太多定位的背景容易挖坑，对引入的motivation要描述清楚。主要解决的是光SAR模态之间的差异。或者更专注于低空无人机的SAR图像，重点还是在检索这个任务比较好。
2. 对抗训练的目的一般是为了增加鲁棒性，不是为了增加精度，要考虑清楚对抗训练良方法的设计目的，在后续实验中要体现出对抗训练提升模型的泛化性。

答：triplet loss可以度量样本中的距离，wasserstein距离可以理解为针对模态差异进一步减少特征分布之间的差异，同时让网络在cross-area的检索中具有泛化性。

3. meter-level的精度太低了

答：这个指标可以换成曲线或者是换一个更合理的指标。

4. OSDataset是用在光-SAR配准上的，而且分辨率是1m的256尺寸图像，我们的数据集是0.5m的200尺寸图像，显示的内容太小了，要考虑实际应用对应的图像尺寸。并且SAR通道多不一定是优势，SAR注重目标的结构信息，单通道反而可以更接近光学图像。
5. 针对这个问题能否转换成one-shot的场景检测问题？看一下有没有相关工作。

答：讨论后认为one-shot的工作不适合这个任务，one-shot的单个样本需要较为稳定而且有规则的特征，同时该样本需要再训练中作为support集。考虑到SAR图像之间变化较大，因此不适合使用one-shot解决这个问题。