Discussion 1

Time:

2020-06-17 23:00

Main Target:

1. 汇总文献阅读工作，具体内容见附录
2. 安排下一阶段工作

Process：

1. 汇报 superpoint 文献， by guyi
2. 汇报 NetVLAD 文献， by suixin
3. 安排下阶段工作：

（1）汇总 Discriptor 的相关研究工作, 着重分析 SIFT. (by suixin)

（2）运行 inverting 开源代码，（by liqiang）

（3）深入研究 SuperPoint 并运行 superpoint 开源代码(by guyi)

[附录]

Article 1： Leveraging Deep Visual Descriptor for Hierarchical Efficient Localization

Target problem：

1. precise pose estimates despite operating in large and changing environments.
2. Control the Computation Cost；

Main idea:

propose to leverage recent advances in deep learning to perform an efficient hierarchical localization : firstly, localizing at the map level using learned image-wide global descriptors; secondly, estimating a precise pose from 2D-3D matches computed in the candidate places only.

Algorithm Framework:

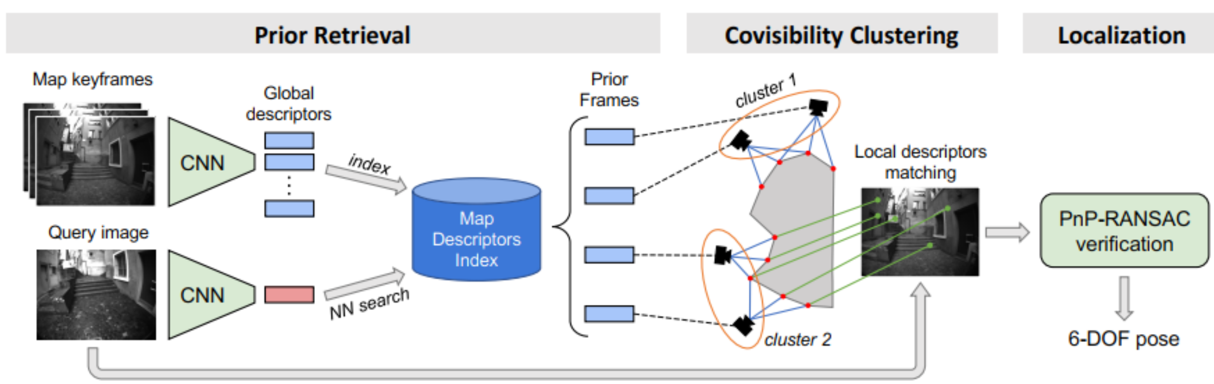
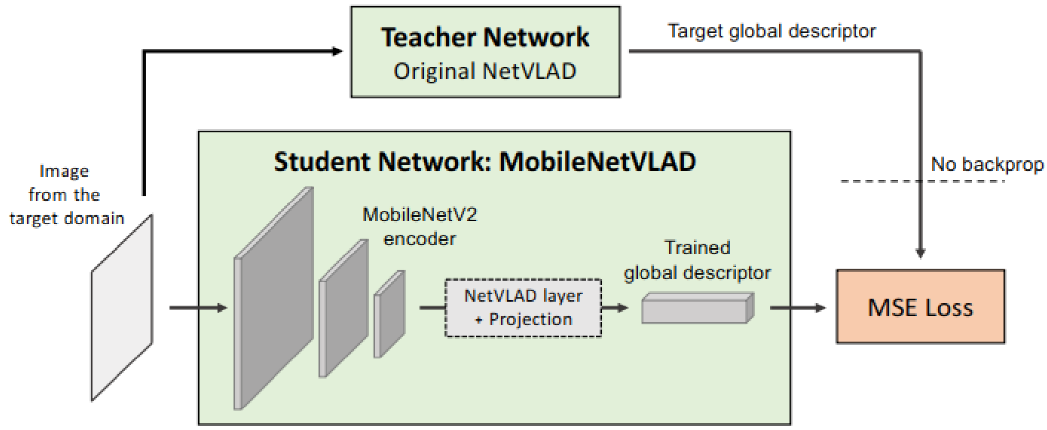


Figure 1: **Overview of our hierarchical localization system.** For a given query image, a coarse search first finds candidate keyframes in the map using global descriptors learned by a CNN. These prior frames are clustered into places, and a local search over expensive local descriptors is performed for each of them until a valid 6-DoF pose is estimated.

(1) Prior Retrieval



KeyPoint : teacher-student Model

1. the student network as a scaled-down version of NetVLAD using a MobileNet encoder followed by a smaller VLAD layer, and name it MobileNetVLAD.
2. A learned linear fully-connected layer down-projects the student descriptors so that their dimension matches with the target descriptors.

(2) Covisibility Clustering

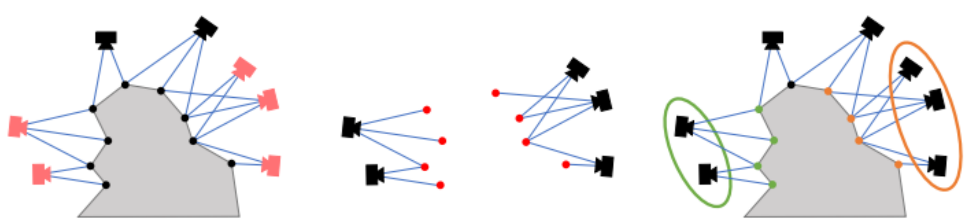


Figure 3: **Example of covisibility clustering.** 1) Five prior frames (red) are retrieved among all the map keyframes. 2) These, along with the 3D points that they observe (red), correspond to the vertices of a bipartite graph, whose edges (blue) correspond to the 2D-3D observations. 3) Two connected components are found (green and orange). The query image is subsequently matched to each set of corresponding points, starting from the orange component as it contains the most keyframes.

keyPoint: The prior frames are clustered named place based on the 3D structure covisibility。

(3) Local matching and Localization

Keypoint: using high-dimensional non-binary descriptors (local descriptor_)

1. compute local descriptor matches between the 3D points that it contains and keypoints in the query image.

2. 2D-3D， PnP Problem

Article 2 From Coarse to Fine: Robust Hierachical Localization at Large Scale

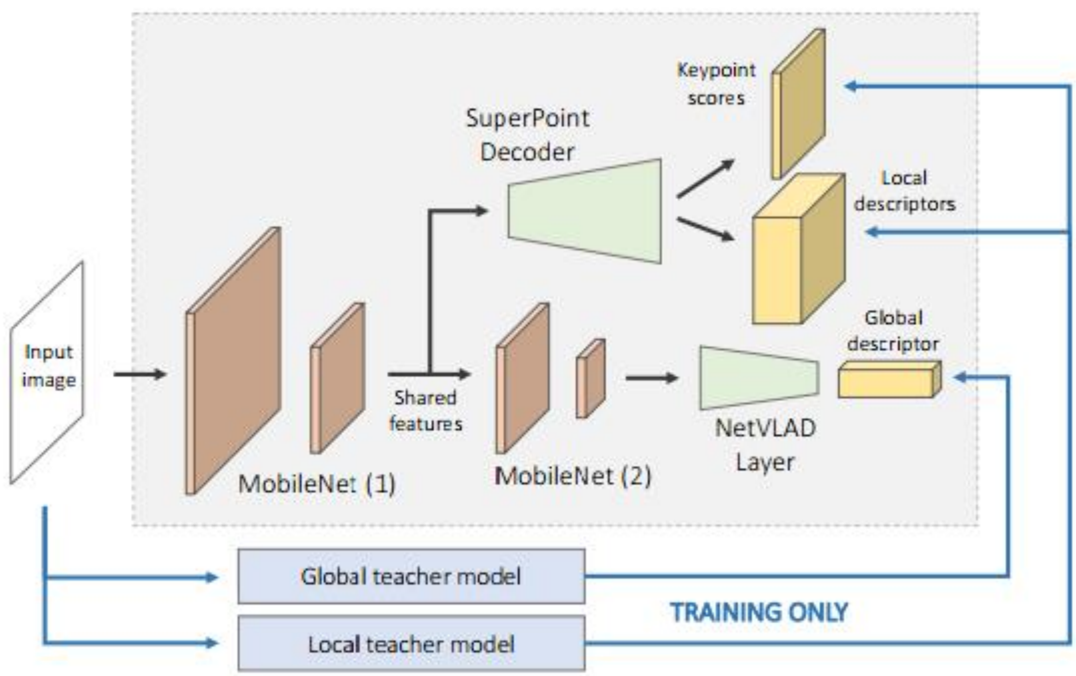
Target Problem:

1. Localization in Large Scale Environment
2. Real-time

Main idea：

1. Propose HF_net; a hierarchical localization approach based on a monolithic CNN that simultaneously predicts local features and global descriptors for accurate 6-DoF localization.
2. Hierarchical Localization

Algorithm FramWork



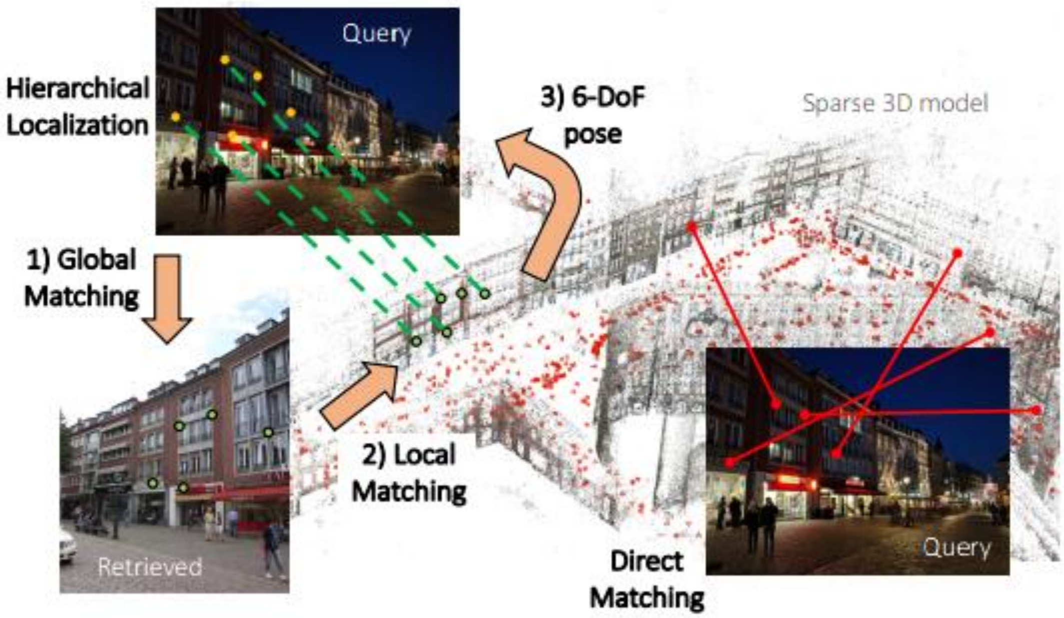
Main point :

1 Single Shot , 3 Heads

2 Teacher – student model

3 Getting Local descriptor with deep learning(MoblieNetV2)

PiPLine



1. Prior Retrieval

   Using global descriptor to match, and using KNN to find candidates prior frames;

2. Co-visibility clustering(the same as article 1)

3. Local feature Match

   Form 2d local descriptor to 3d point contained in the place. Getting 6-DoF by solving 2D-3D problem.

Training process tricks:

1. Data augmentation
2. Muti-task distillation

Three experiment

1. Local Feature Descriptor

   （1） dataset: HPatches, SfM

   （2） result

	HPatches		SfM	
	Rep.	MLE	Rep.	MLE
DoG	0.438	1.00	0.284	1.20
Harris	0.531	1.18	0.511	1.46
SuperPoint	0.496	1.04	0.508	1.46
LF-Net	0.466	1.14	0.448	1.46

2. Large Scale Localization

   (1) Dataset:

   The Aachen Day-Night dataset;

   the CMU Seasons dataset;

   The RobotCar Seasons dataset

   (2) Result

	Aachen		**RobotCar**				**CMU**	
	day	night	dusk	sun	night	night-rain	urban	suburban
m	.25/.50/5.0	0.5/1.0/5.0	.25/.50/5.0	.25/.50/5.0	.25/.50/5.0	.25/.50/5.0	.25/.50/5.0	.25/.50/5.0
deg	2/5/10	2/5/10	2/5/10	2/5/10	2/5/10	2/5/10	2/5/10	2/5/10
AS	57.3 / 83.7 / **96.6**	19.4 / 30.6 / 43.9	44.7 / 74.6 / **95.9**	25.0 / 46.5 / 69.1	0.5 / 1.1 / 3.4	**1.4** / 3.0 / 5.2	**55.2** / 60.3 / 65.1	20.7 / 25.9 / 29.9
CSL	52.3 / 80.0 / 94.3	24.5 / 33.7 / 49.0	**56.6 / 82.7 / 95.9**	**28.0** / 47.0 / 70.4	0.2 / 0.9 / 5.3	0.9 / 4.3 / 9.1	36.7 / 42.0 / 53.1	8.6 / 11.7 / 21.1
DenseVLAD	0.0 / 0.1 / 22.8	0.0 / 2.0 / 14.3	10.2 / 38.8 / 94.2	5.7 / 16.3 / 80.2	**0.9** / 3.4 / **19.9**	1.1 / 5.5 / **25.5**	22.2 / 48.7 / 92.8	9.9 / 26.6 / 85.2
NetVLAD	0.0 / 0.2 / 18.9	0.0 / 2.0 / 12.2	7.4 / 29.7 / 92.9	5.7 / 16.5 / **86.7**	0.2 / 1.8 / 15.5	0.5 / 2.7 / 16.4	17.4 / 40.3 / 93.2	7.7 / 21.0 / 80.5
HF-Net (ours)	**75.7** / **84.3** / 90.9	**40.8 /55.1 / 72.4**	22.1 / 69.0 / 94.4	26.5 / **58.5** / 86.1	0.7 / **4.6** / 14.8	**1.4** / **9.3** / 20.5	39.6 / **89.7 / 95.7**	**30.9 / 73.3 / 86.6**
NV+SP	79.7 / 88.0 / 93.7	40.8 / 56.1 / 74.5	22.8 / 70.1 / 96.2	26.3 / 66.1 / 92.6	1.8 / 11.9 / 31.3	1.8 / 12.3 / 26.6	40.4 / 90.6 / 97.5	31.5 / 75.6 / 91.0
NV+SIFT	82.8 / 88.1 / 93.1	30.6 / 43.9 / 58.2	55.6 / 83.5 / 95.3	46.3 / 67.4 / 90.9	4.1 / 9.1 / 24.4	2.3 / 10.2 / 20.5	63.9 / 71.9 / 92.8	28.7 / 39.0 / 82.1
SMC	-	-	53.8 / 83.0 / 97.7	46.5 / 74.6 / 95.9	6.2 / 18.5 / 44.3	8.0 / 26.4 / 46.4	75.2 / 82.1 / 87.7	44.6 / 53.9 / 63.5

3. Runtime Evaluation

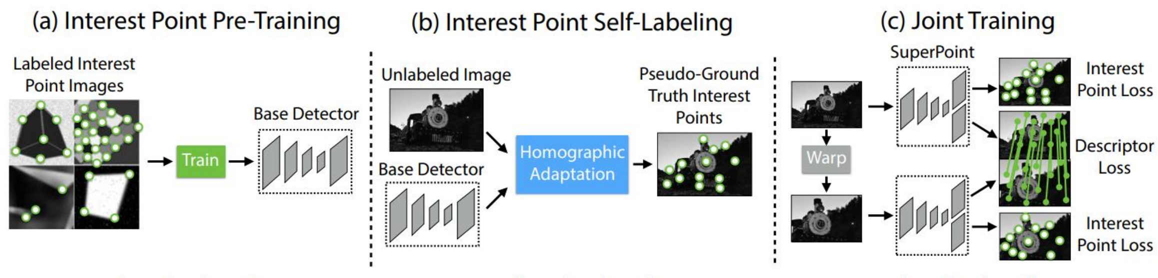
   Result

		HF-Net (ours)						**AS**			**NV+SIFT**			
		HF-Net	Global search	Covisibility	Local search	PnP+RANSAC	**Total**	SIFT	Loc.	**Total**	SIFT	NetVLAD	Loc.	**Total**
Aachen	Day	15	51	5	163	9	**243**	263	112	**375**	263	92	1264	**1356**
	Night	15	52	5	170	18	**260**	263	132	**395**	263	92	1563	**1655**
RobotCar	Dusk	19	53	1	58	4	**135**	189	283	**472**	189	91	294	**385**
	Night	19	55	1	103	38	**216**	189	1021	**1210**	189	91	554	**645**

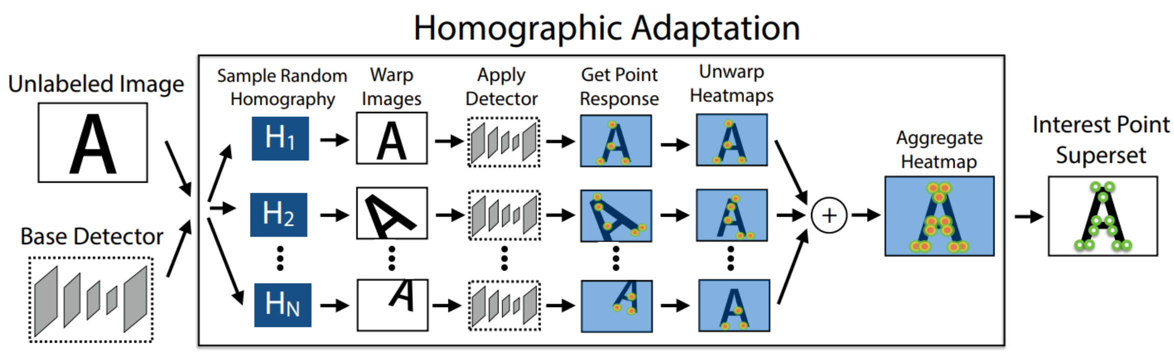
Article 3 SuperPoint: Self-Supervised Interest Point Detection and Description

By Guyi

Point1 PipeLine



Point 2



Point 3

Article 4 NetVLAD: CNN architecture for weakly supervised place recognition

By SuiXin