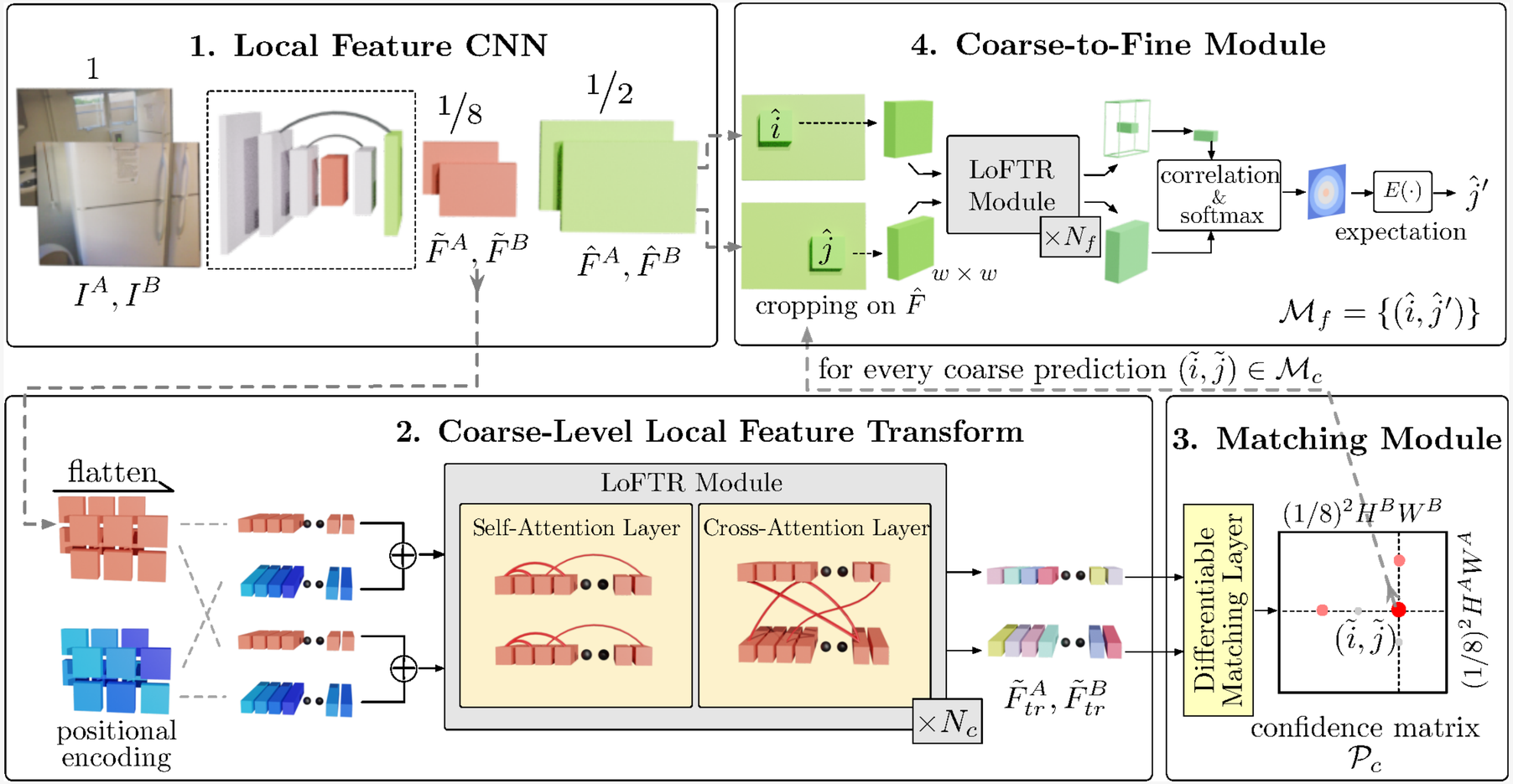


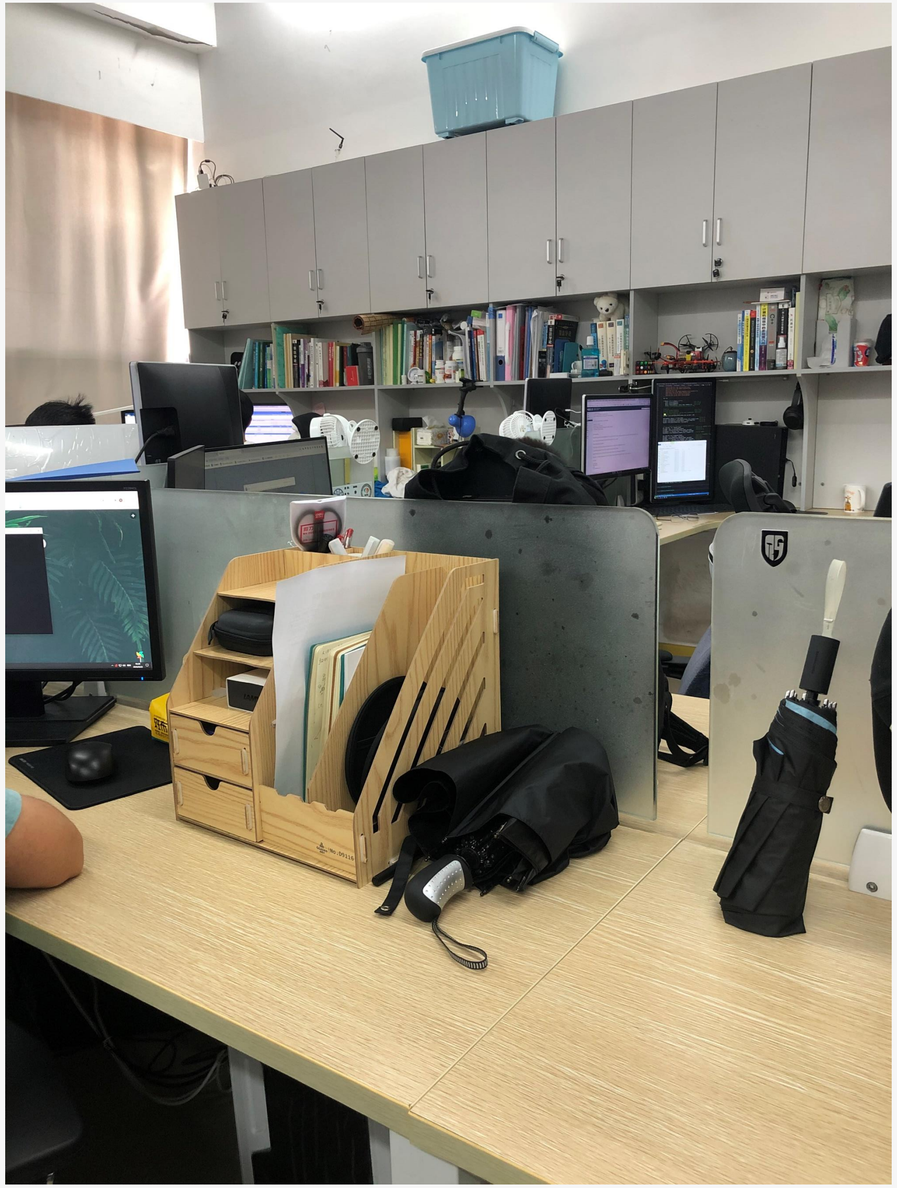
LoFTR: Detector-Free Local Feature Matching with Transformers

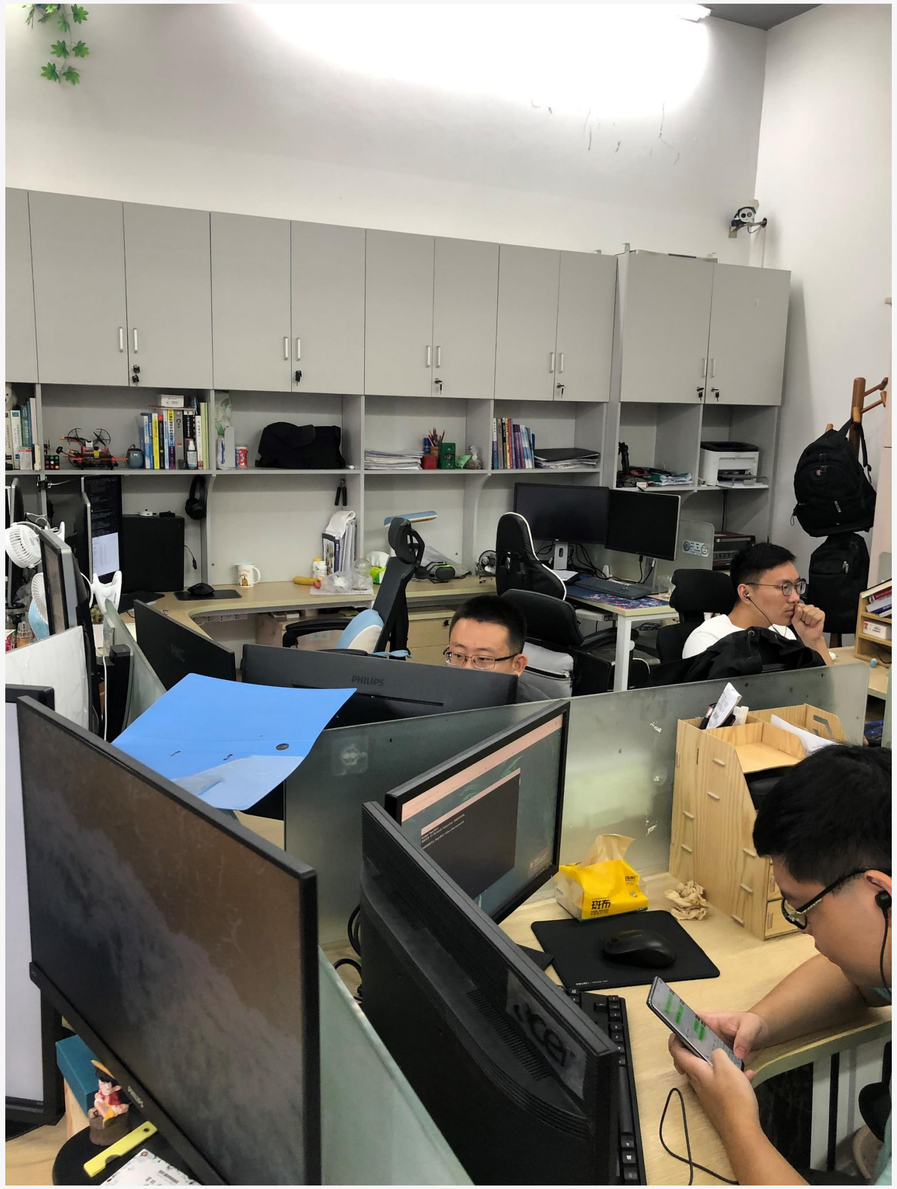
CVPR 2021

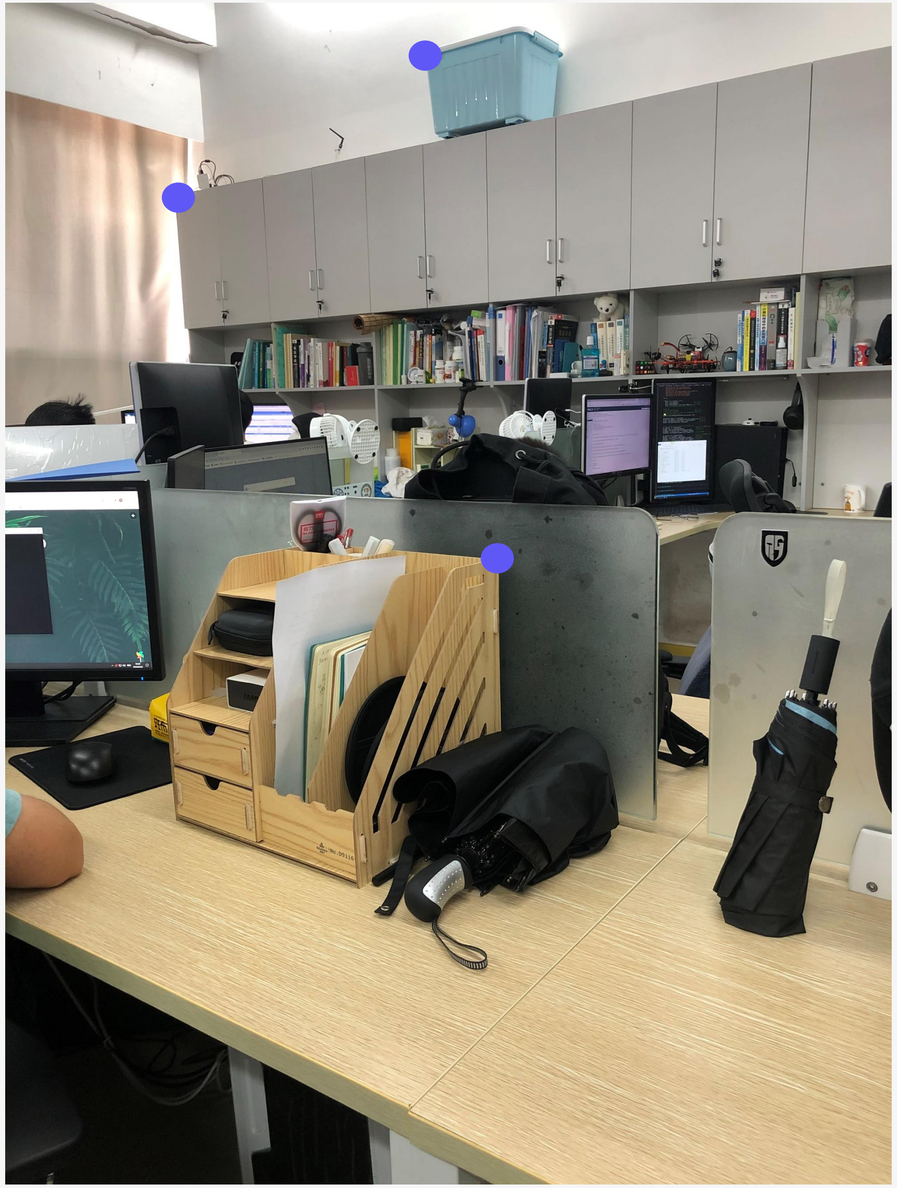
Jiaming Sun[1,2*], Zehong Shen[1*], Yuang Wang[1*], Hujun Bao[1], Xiaowei Zhou[1]

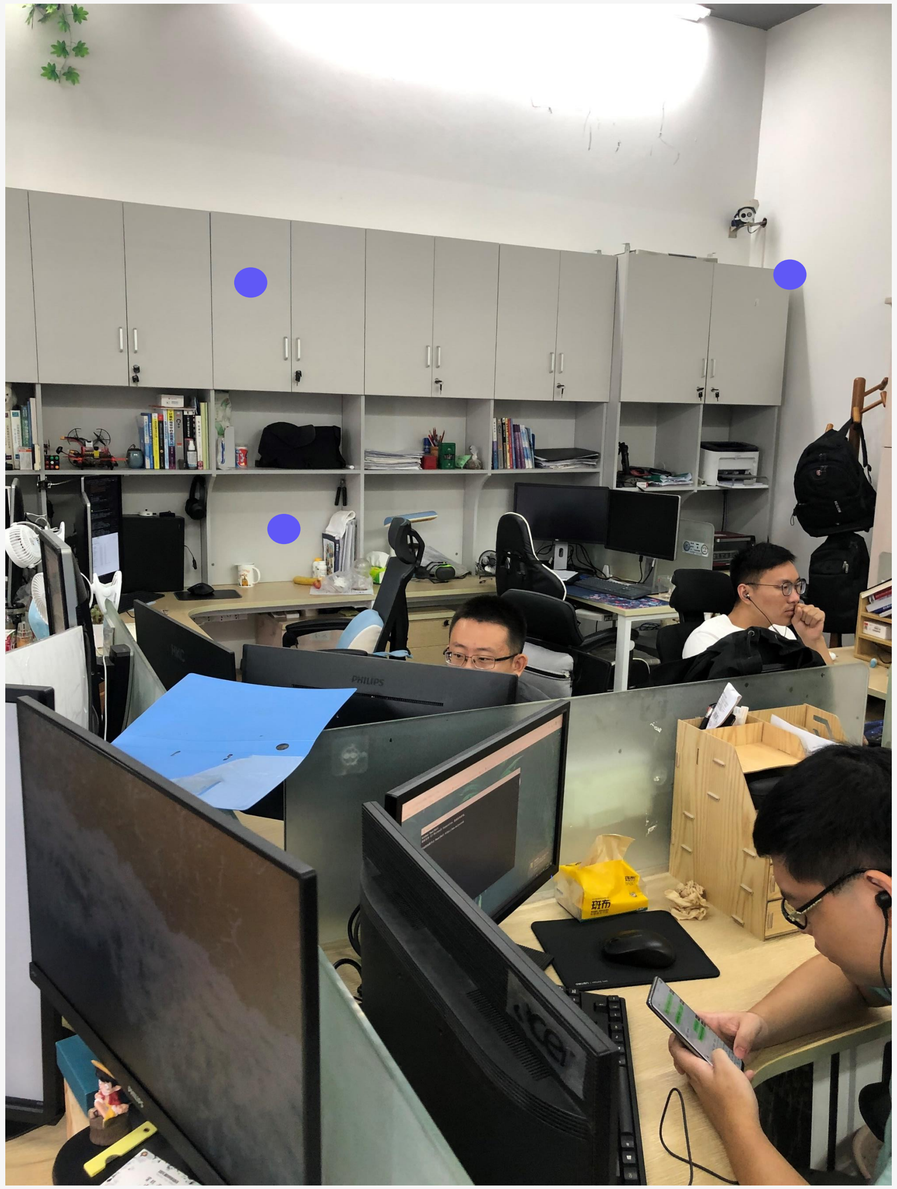
[1]State Key Lab of CAD & CG, Zhejiang University [2]SenseTime Research

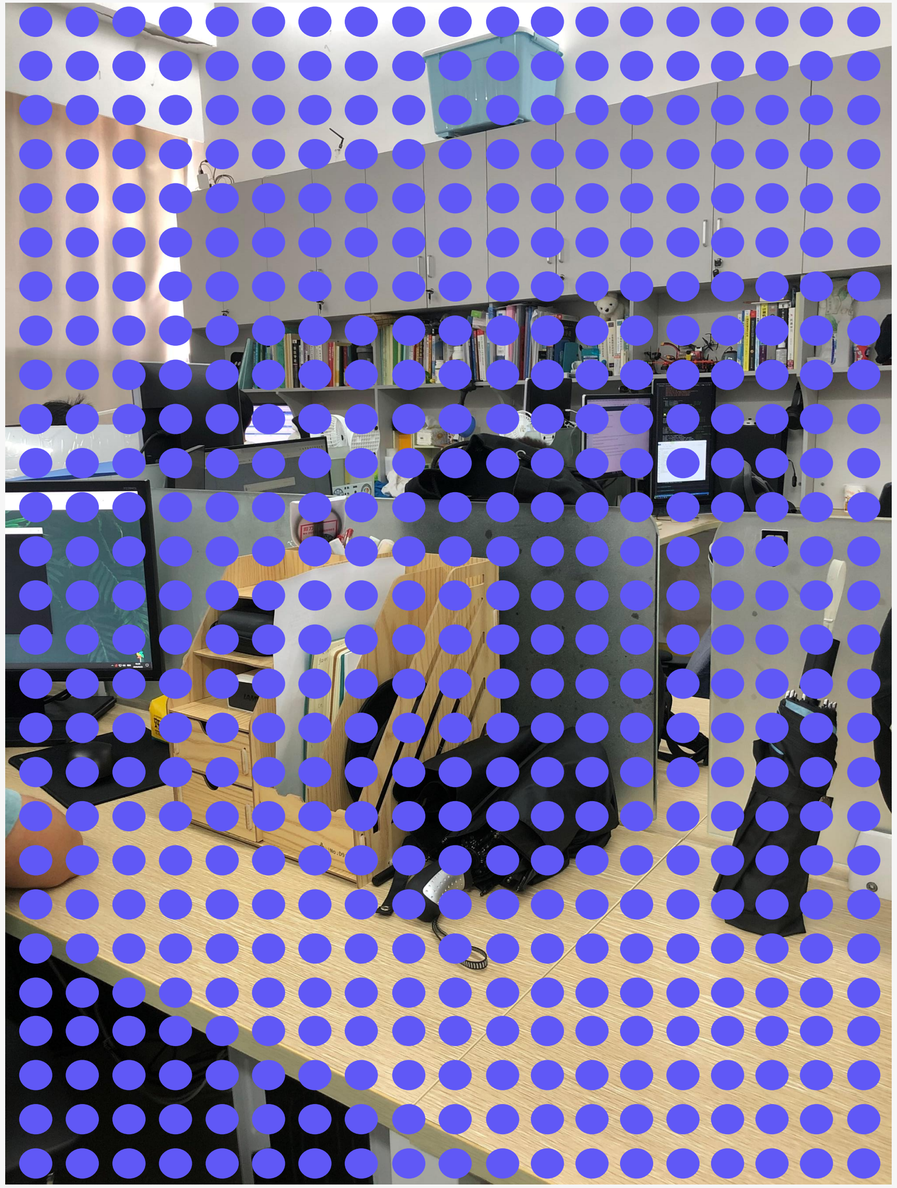
[*] denotes equal contribution

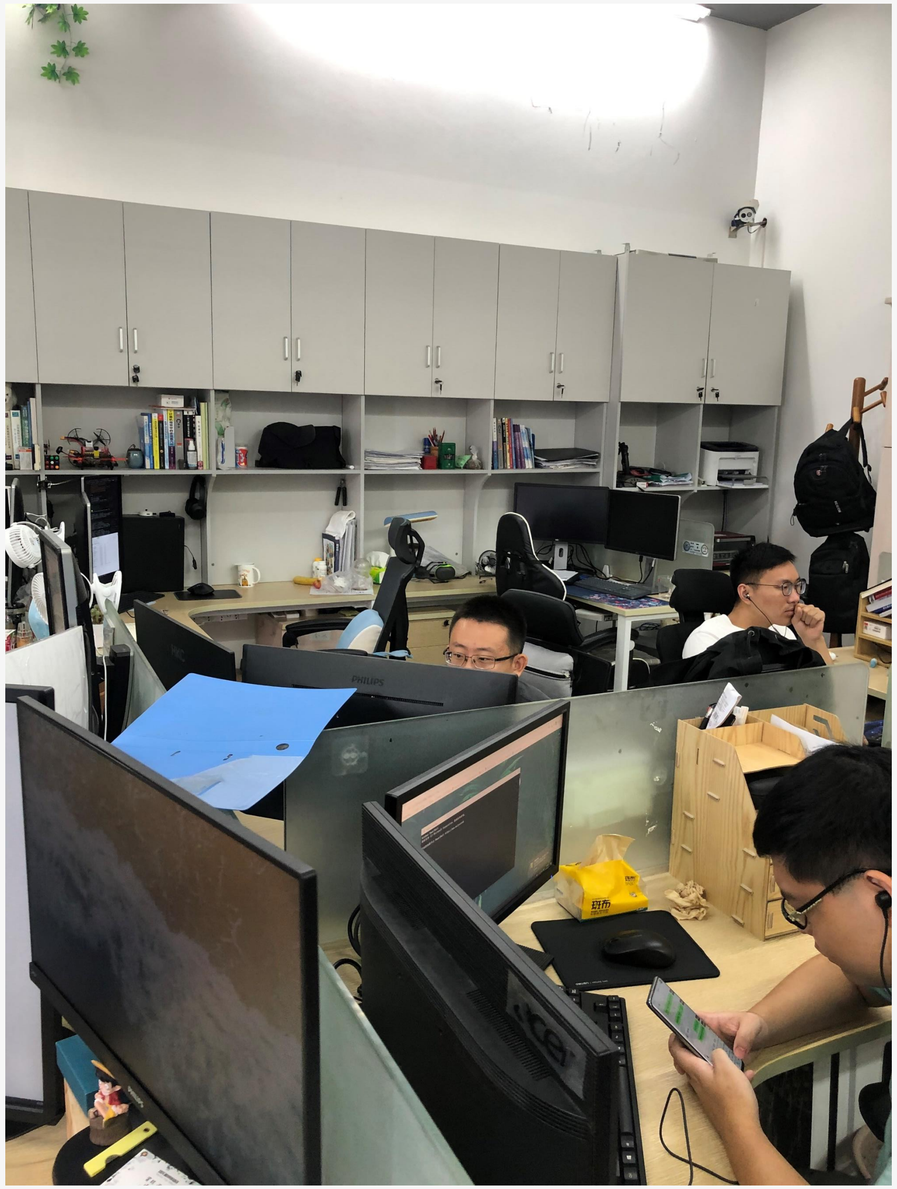




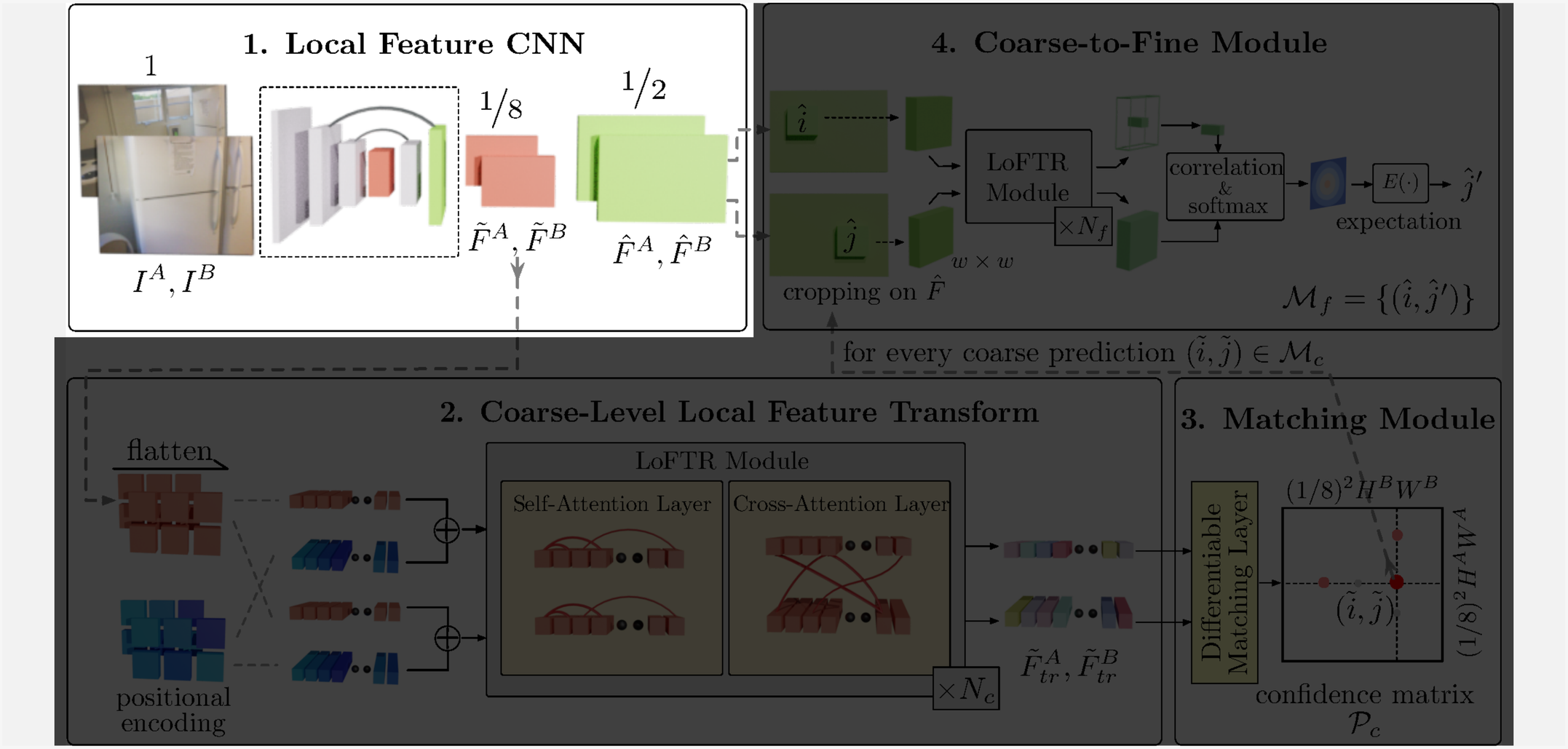


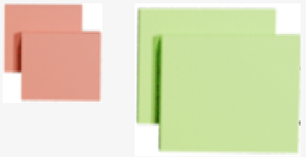





PHILIPS

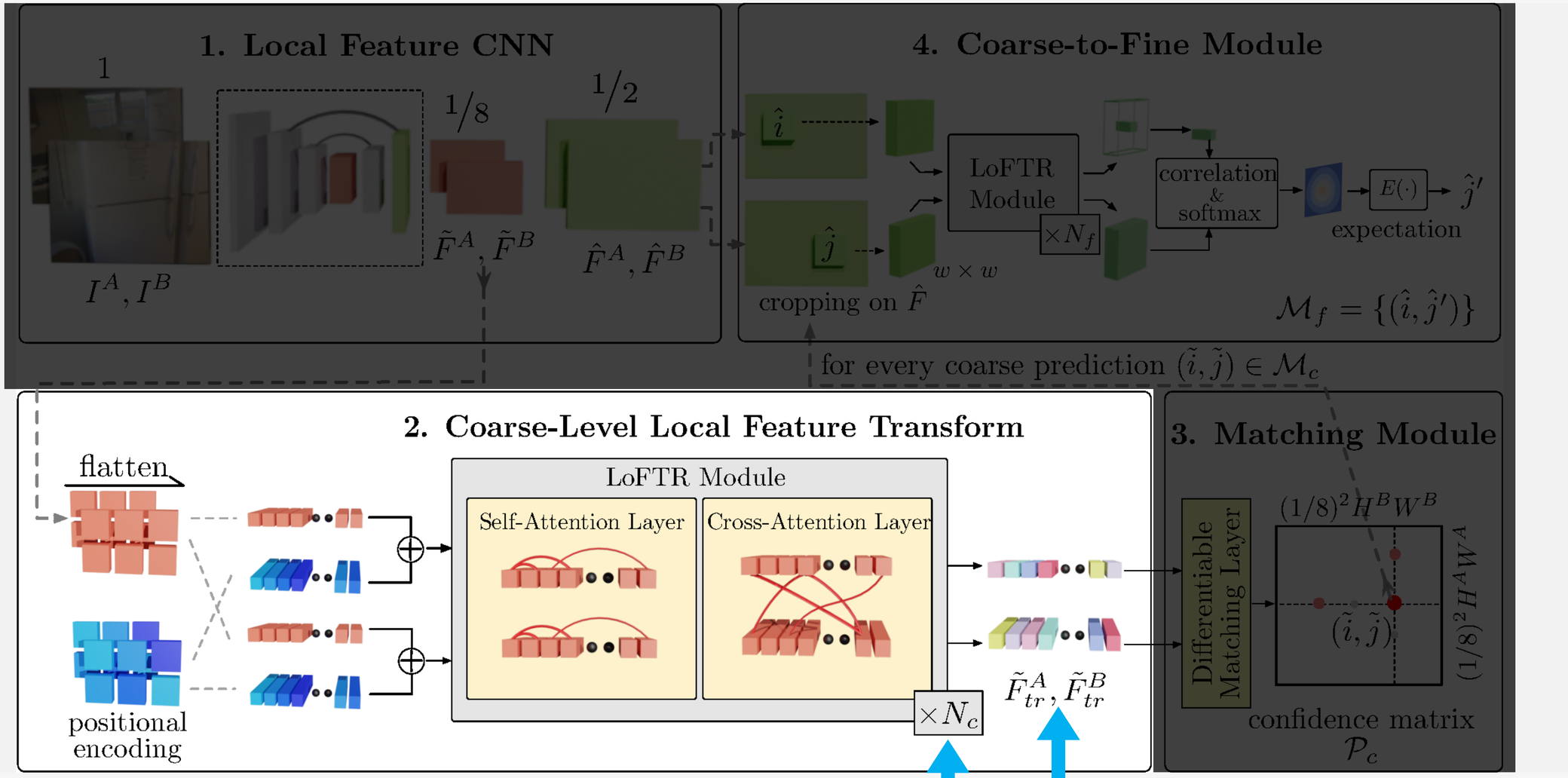
Extract two-level feature maps: Coarse-level and Fine-level





Two layers correspond two images

Apply self-attention and cross-attention on positional encoding features from two images

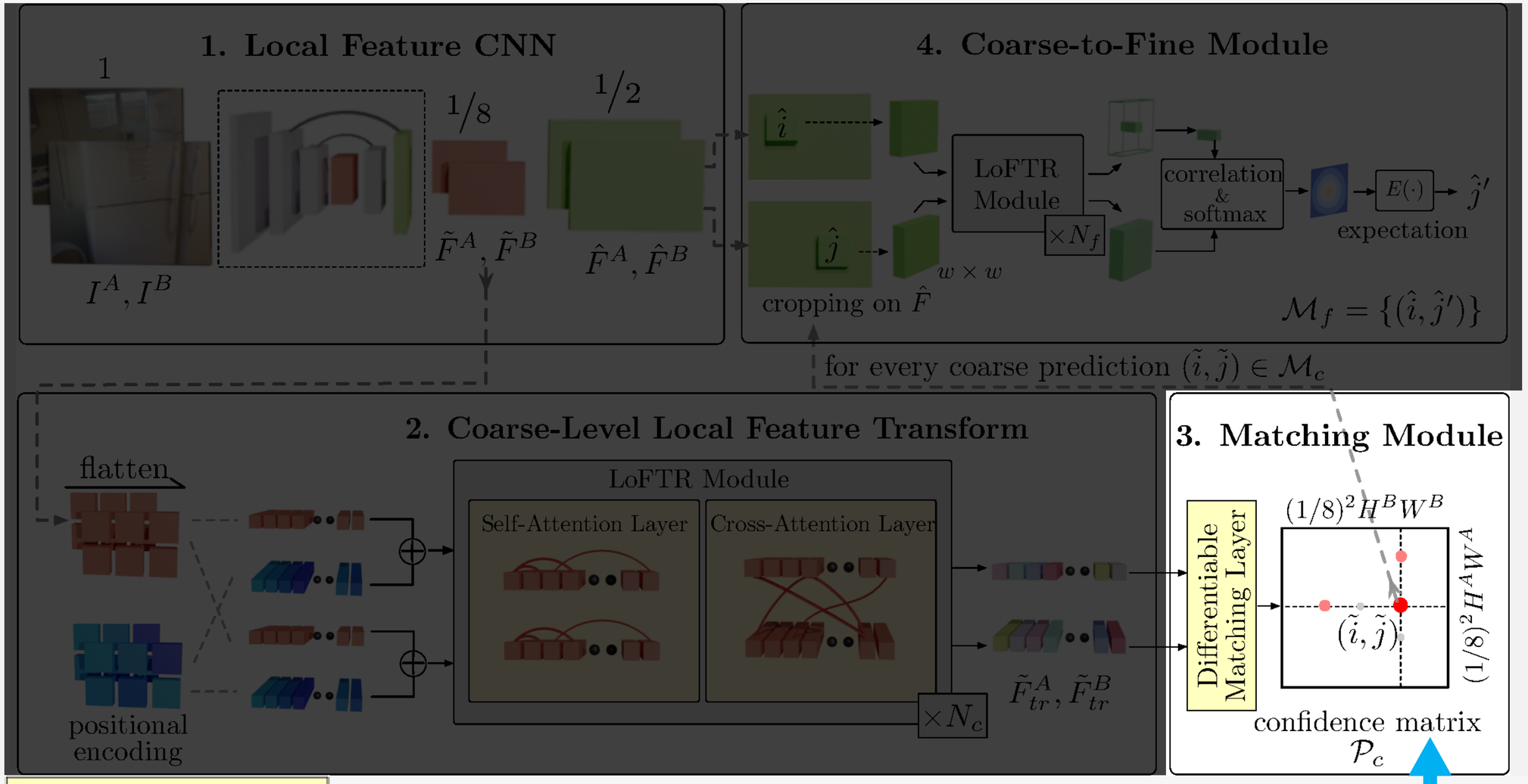


Give each point a feature vector

This process will loop N_c times

The features mixed up with position, self- and cross-image information

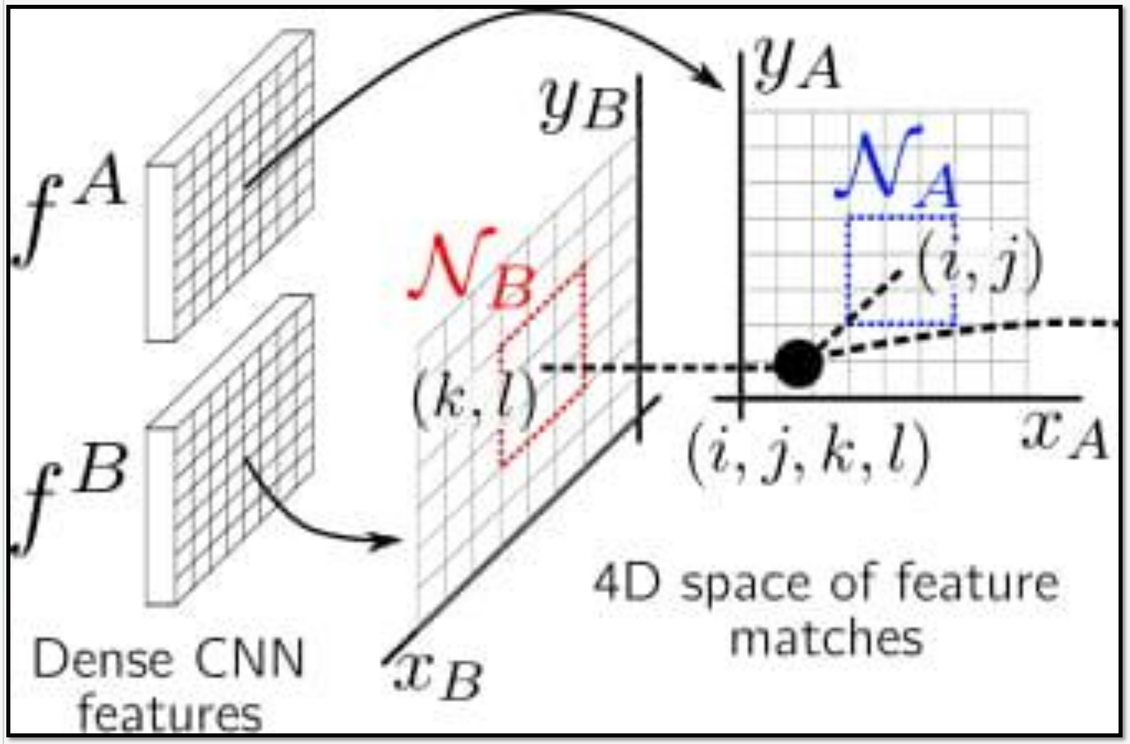
Matching all Coarse point pairwise

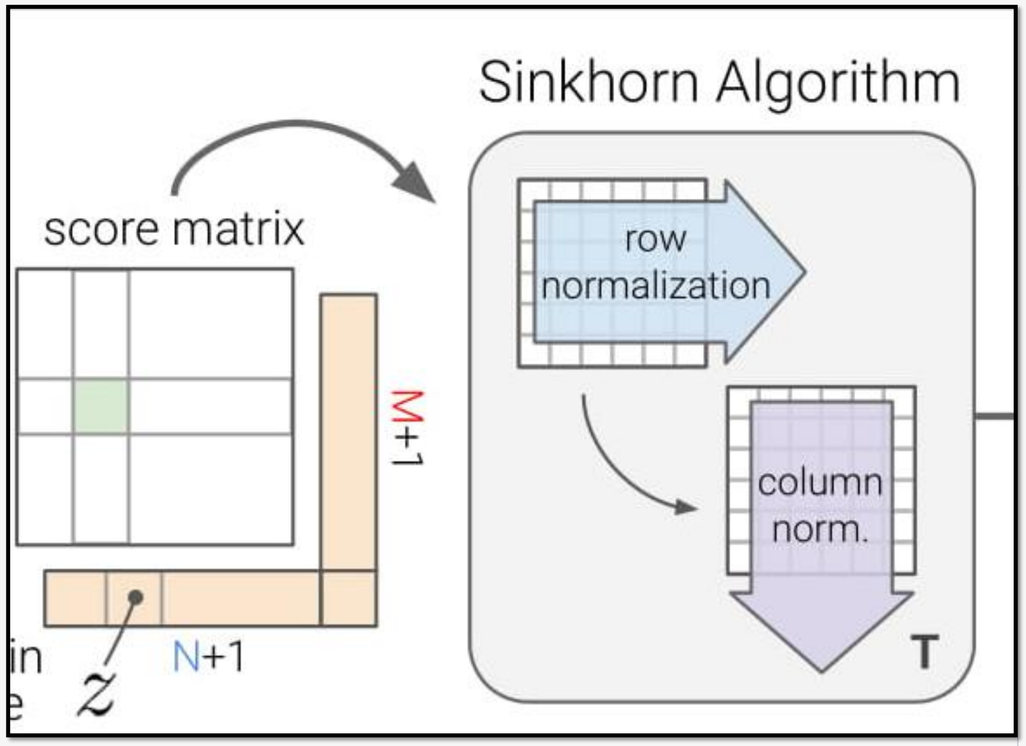


Differentiable Matching Layer

1. DS - Dual Softmax
2. OT - Optimal Transport

After the original image is reduced to1/8, the matching probability of pairwise point in two images



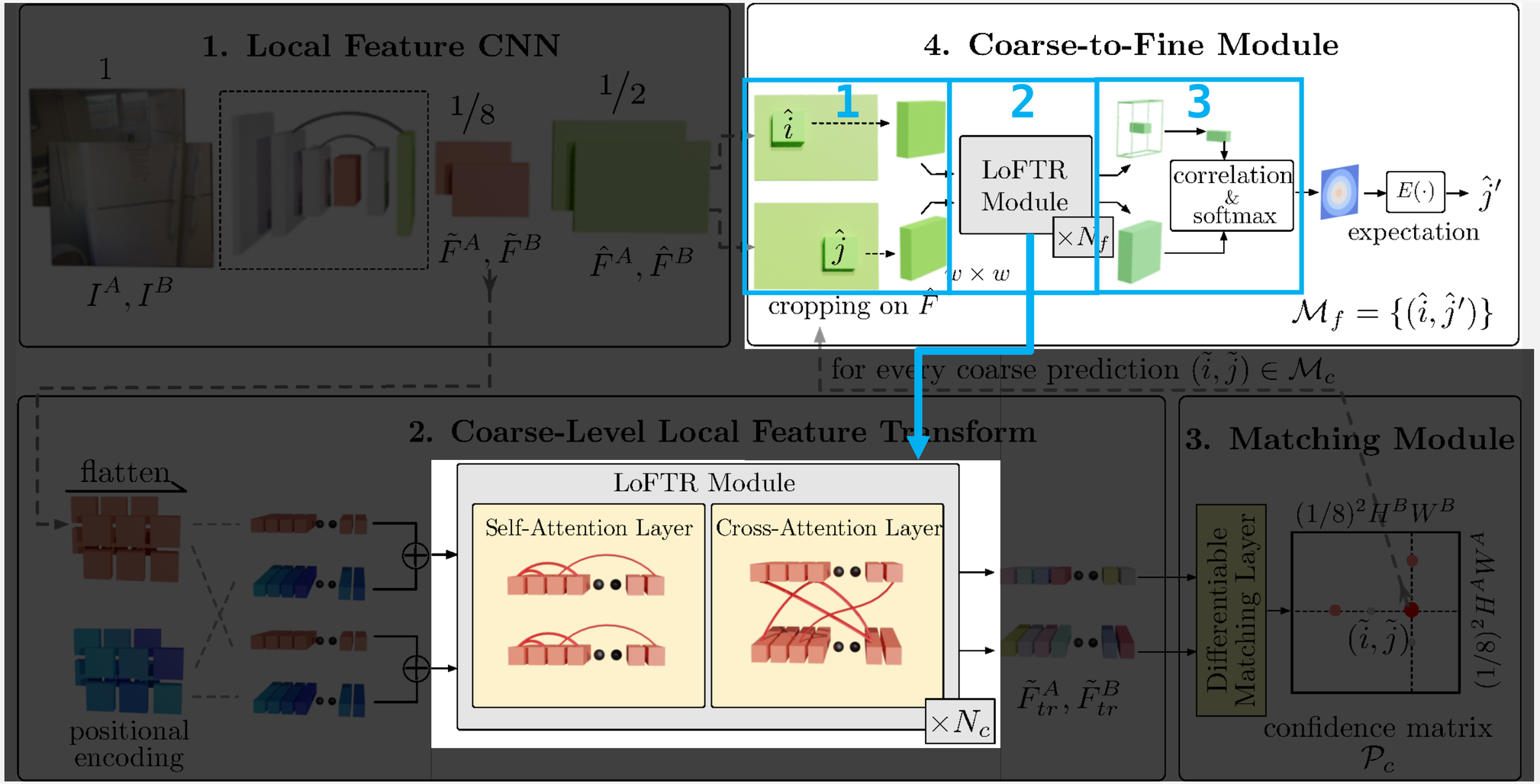


DS: Dual Softmax **OT: Optimal transport**

	DS: Dual Softmax	OT: Optimal transport
Could Obtain Matching Matrix	√	√
Matching Matrix Form (Bold represents the original from it be proposed)	4D / 2D	4D / 2D
Implement Method	Multiply then Dual Softmax	Sinkhorn Algorithm

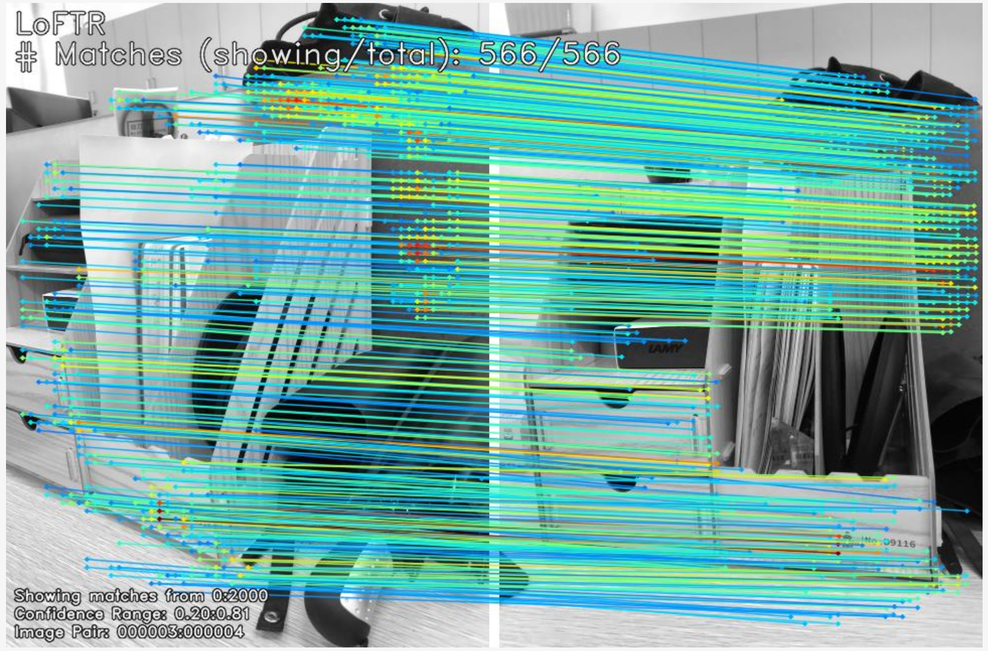
Matching Fine point

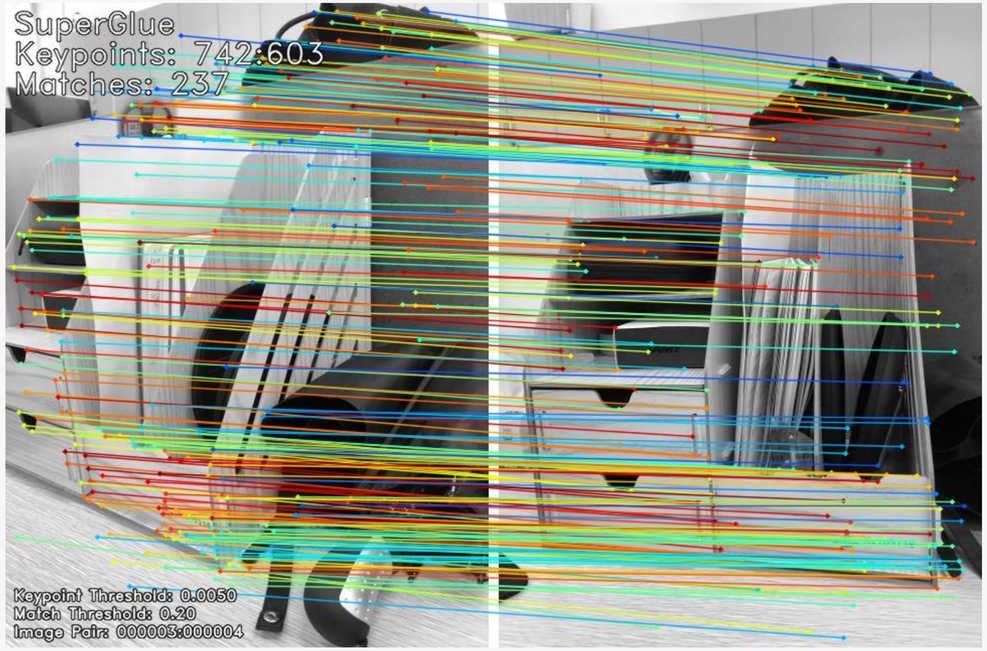
This is why te keypoints in visualization are arranged like a grid



1. Crop small $w \times w$ $(w = 5)$ feature patchs
2. Transform it
3. use left center point in left patch to find corresponding point in right patch


LoFTR
Showing matches from 0:800
Matches: 800
Confidence Range: 0.86:0.98
Image Pair: 000000:000001
SuperGlue
Keypoints: 295:295
Matches: 295
Keypoint Threshold: 0.0550
Match Threshold: 0.20
Image Pair: 000000:000001


LoFTR
Matches (showing/total): 566/566
Showing matches from 0:2000
Confidence Range: 0.20:0.81
Image Pair: 000003:000004

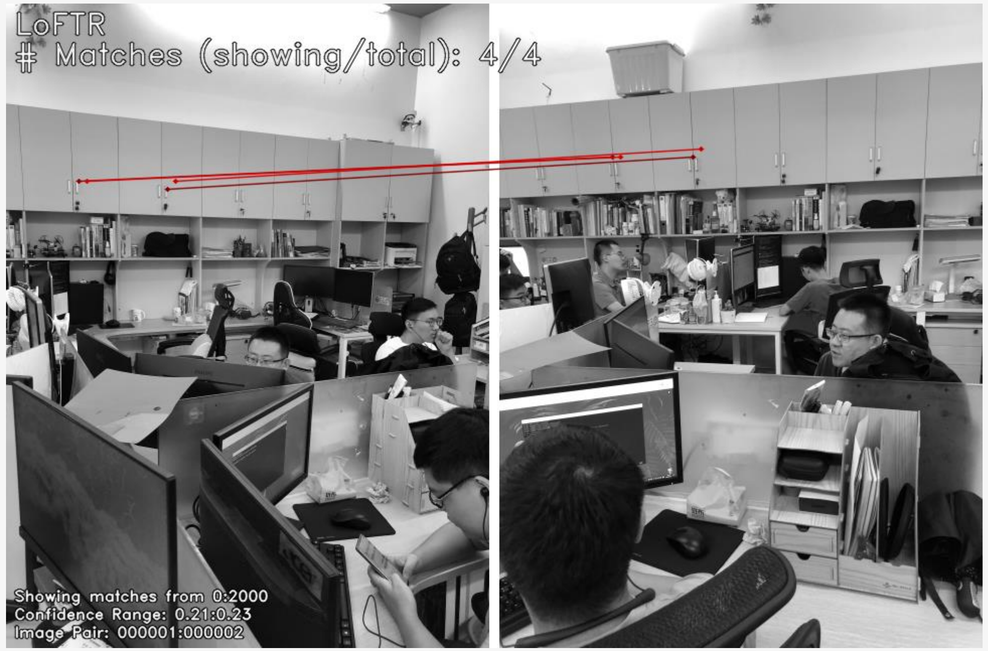

SuperGlue
Keypoints: 742:603
Matches: 237
Keypoint Threshold: 0.0050
Match Threshold: 0.20
Image Pair: 000003:000004


LoFTR
Matches (showing/total): 3/3
Showing matches from 0:2000
Confidence Range: 0.20:0.27
Image Pair: 000004:000005


SuperGlue
Keypoints: 603:736
Matches: 0
Keypoint Threshold: 0.0050
Match Threshold: 0.20
Image Pair: 000004:000005


LoFTR
Matches (showing/total): 5/5
Showing matches from 0:2000
Confidence Range: 0.21:0.32
Image Pair: 000000:000004


SuperGlue
Keypoints: 993:603
Matches: 103
Keypoint Threshold: 0.0050
Match Threshold: 0.20
Image Pair: 000000:000004

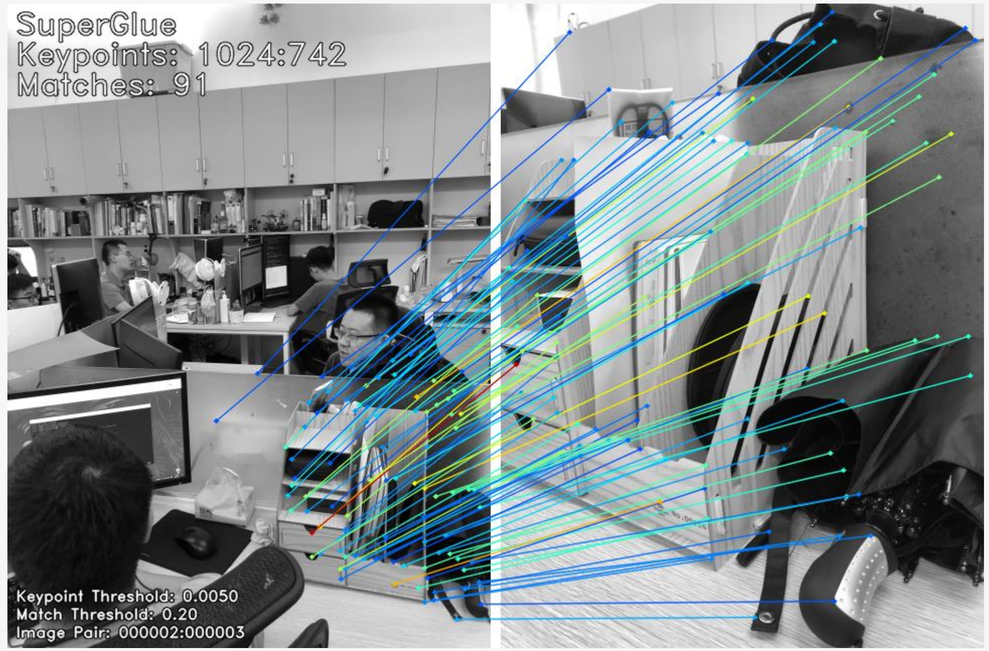

LoFTR
Matches (showing/total): 4/4
Showing matches from 0:2000
Confidence Range: 0.21:0.23
Image Pair: 000001:000002


SuperGlue
Keypoints: 1024:1024
Matches: 293
Keypoint Threshold: 0.0050
Match Threshold: 0.20
Image Pair: 000001:000002


LoFTR
Matches (showing/total): 14/14
Showing matches from 0:2000
Confidence Range: 0.21:0.39
Image Pair: 000001:000005


SuperGlue
Keypoints: 1024:736
Matches: 96
Keypoint Threshold: 0.0050
Match Threshold: 0.20
Image Pair: 000001:000005


LoFTR
Matches (showing/total): 1/1
Showing matches from 0:2000
Confidence Range: 0.21:0.21
Image Pair: 000002:000003


SuperGlue
Keypoints: 1024:742
Matches: 91
Keypoint Threshold: 0.0050
Match Threshold: 0.20
Image Pair: 000002:000003

	优点	缺点
LoFTR	在无纹理区域具有较好的表现 Ps. 寻找无纹理区域的匹配是反直觉的, 在此之前, 领域内一直避免去寻找无纹理上的匹配,是因为我们觉得无纹理区域无法提供可靠的关键点	关键点的精度不会太高,并且在无纹理之外的情况, 表现一般般,甚至较差,就好像一个四条腿的动物, 一条腿特别粗, 而其他三条腿却特别细
SuperGlue	精度在任何情况下都有不错的表现, 除了无纹理区域	在无纹理情况下表现一般, 就好像一个四条腿的动物, 三条腿都比较壮, 而有一条腿比较平常