广东工业大学章云团队：智能图像检测与识别成果简介

报告人：李东 副教授

Outline

- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

Outline

- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

1. 针对预训练的先验权重问题，在不使用额外的数据情况下采用注意力机制辅助迁移，以使网络更倾向于选择人脸面部肌肉运动的位置。
2. 通过Prompt Learning方法的启发，我们认为Label Embedding可以与AU一一对应，可以学习到AU之间的关联关系

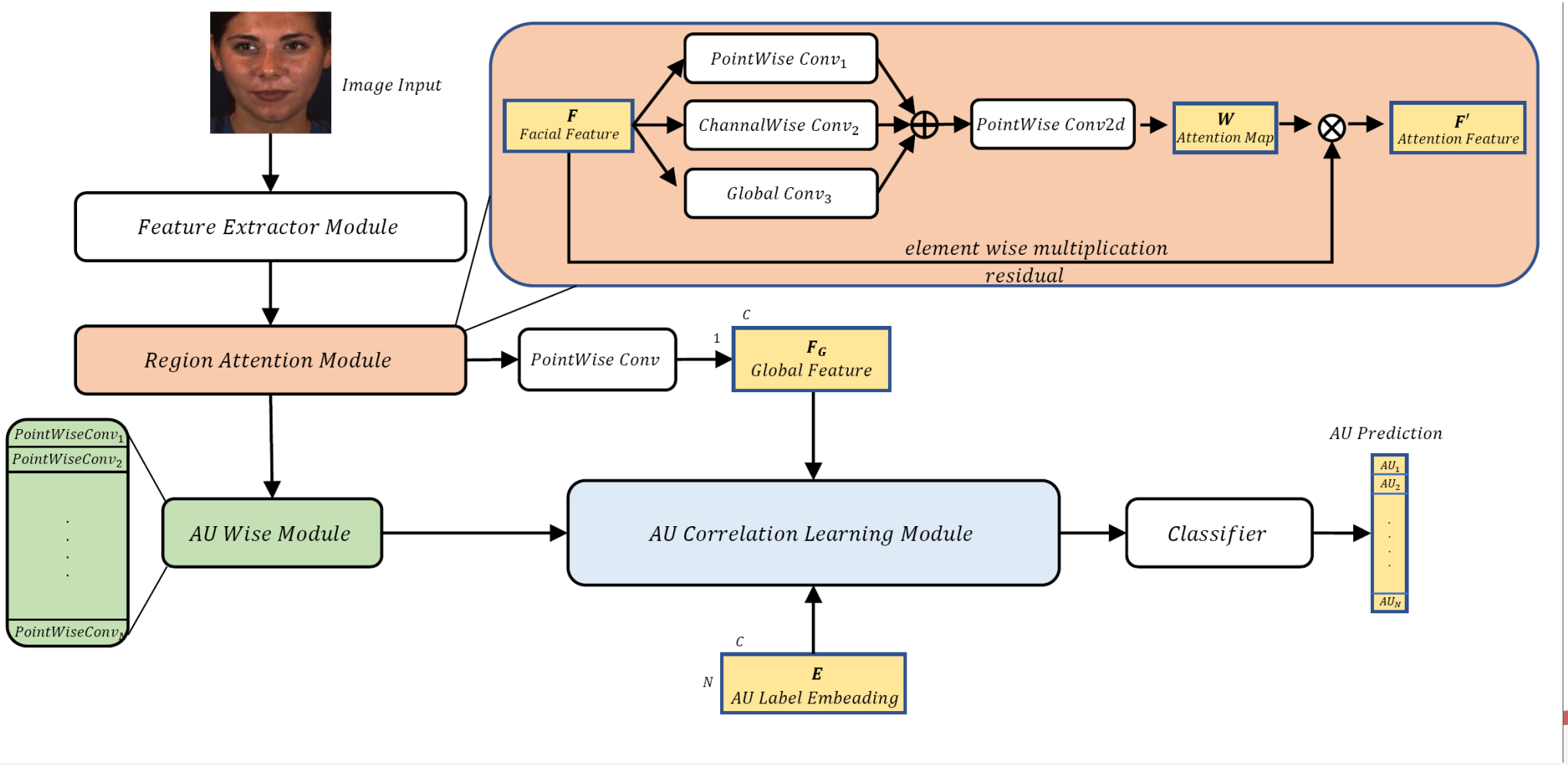


TABLE I

COMPARISONS OF OUR METHOD AND THE STATE-OF-THE-ART METHODS ON BP4D IN TERMS OF F1 SCORES(%). THE BEST RESULTS ARE SHOWN IN BOLD

Method	AU1	AU2	AU4	AU6	AU7	AU10	AU12	AU14	AU15	AU17	AU23	AU24	Avg.
LSVM	23.2	22.8	23.1	27.2	47.1	77.2	63.7	64.3	18.4	33.0	19.4	20.7	35.3
JPML	32.6	25.6	37.4	42.3	50.5	72.2	74.1	65.7	38.1	40.0	30.4	42.3	45.9
DRML	36.4	41.8	43.0	55.0	67.0	66.3	65.8	54.1	33.2	48.0	31.7	30.0	48.3
EAC-Net	39.0	35.2	48.6	76.1	72.9	81.9	86.2	58.8	37.5	59.1	35.9	35.8	55.9
DSIN	51.7	40.4	56.0	76.1	73.5	79.9	85.4	62.7	37.3	62.9	38.8	41.6	58.9
ARL	45.8	39.8	55.1	75.7	77.2	82.3	86.6	58.8	47.6	62.1	47.4	55.4	61.1
LP-Net	43.4	38.0	54.2	77.1	76.7	83.8	87.2	63.3	45.3	60.5	48.1	54.2	61.0
AU-GCN	46.8	38.5	60.1	80.1	79.5	84.8	88.0	67.3	52.0	63.2	40.9	52.8	62.8
SRERL	46.9	45.3	55.6	77.1	78.4	83.5	87.6	63.9	52.2	63.9	47.1	53.3	62.9
AU-RCNN	50.2	43.7	57	78.5	78.5	82.6	87	67.7	49.1	62.4	50.4	49.3	63.0
UGN-B	54.2	46.4	56.8	76.2	76.7	82.4	86.1	64.7	51.2	63.1	48.5	53.6	63.3
JÂA-Net	53.8	47.8	58.2	78.5	75.8	82.7	88.2	63.7	43.3	61.8	45.6	49.9	62.4
SEV-Net	**58.2**	50.4	58.3	**81.9**	73.9	**87.8**	87.5	61.6	52.6	62.2	44.6	47.6	63.9
HMP-PS	53.1	46.1	56.0	76.5	76.9	82.1	86.4	64.8	51.5	63.0	49.9	54.5	63.4
FAUDT	51.7	49.3	**61.0**	77.8	79.5	82.9	86.3	67.6	51.9	63.0	43.7	56.3	64.2
CISNet	54.8	48.3	57.2	76.2	76.5	85.2	87.2	66.2	50.9	65.0	47.7	**56.5**	64.3
D-PAttNet^{tt}	50.7	42.5	59.0	79.4	79.0	85.0	**89.3**	67.6	51.6	**65.3**	49.6	54.5	64.7
ME-GraphAU	53.7	46.9	59.0	78.5	**80.0**	84.4	87.8	67.3	52.5	63.2	**50.6**	52.4	64.7
(Ours)	57.0	**51.4**	57.1	77.9	79.1	84.3	89.0	**67.7**	**54.1**	62.9	50.3	54.7	**65.4**

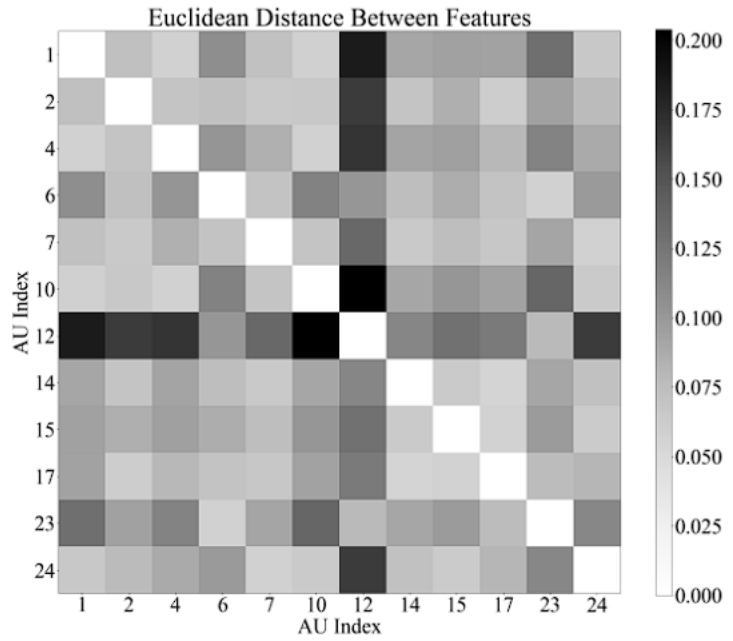


Fig. 6. The L2 distance between different AU embedding.

1. 左图为相比于目前的方法能够实现更快的训练速度和推理速度，采用较小的Backbone也能实现更高的精度
2. 右图为学习得到的AU Label Embedding两两之间计算的L2距离。根据BP4D数据集中的描述，AU12通常在特定表情中独立显示，因此AU12的嵌入与其他AU之间的距离相对较大。在眼睛附近激活的AU1、2、4、6和7之间的嵌入距离相对较小。由AU4代表的皱眉和由AU6代表的扬起脸颊在这些AU中具有最大的距离。此外，与通常出现在愉快表情中的AU6对应的扬起脸颊与与AU12对应的拉动嘴角的AU之间的距离相对较小。

Outline

- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

Periodic-Aware Network for Fine-grained Action Recognition， PRCV2023

1.首次提出周期性特征对于细分类动作识别的重要性。提出了一种称为“周期感知网络（PAN）”的架构，该架构可以有效区分具有不同重复次数的子类别。
2.设计了一个周期性特征提取模块（PFEM），通过时空变换（TSM）来表示周期性信息，并从构建的TSM中提取周期性特征。这些特征被融合到3D-CNN的低层细节特征和高层语义特征中，使网络的每个层级都能感知周期性信息。
3.我们提出了一种新颖的周期性融合模块（PFM），通过两步的挤压-激励过程来融合周期性特征和时空特征。

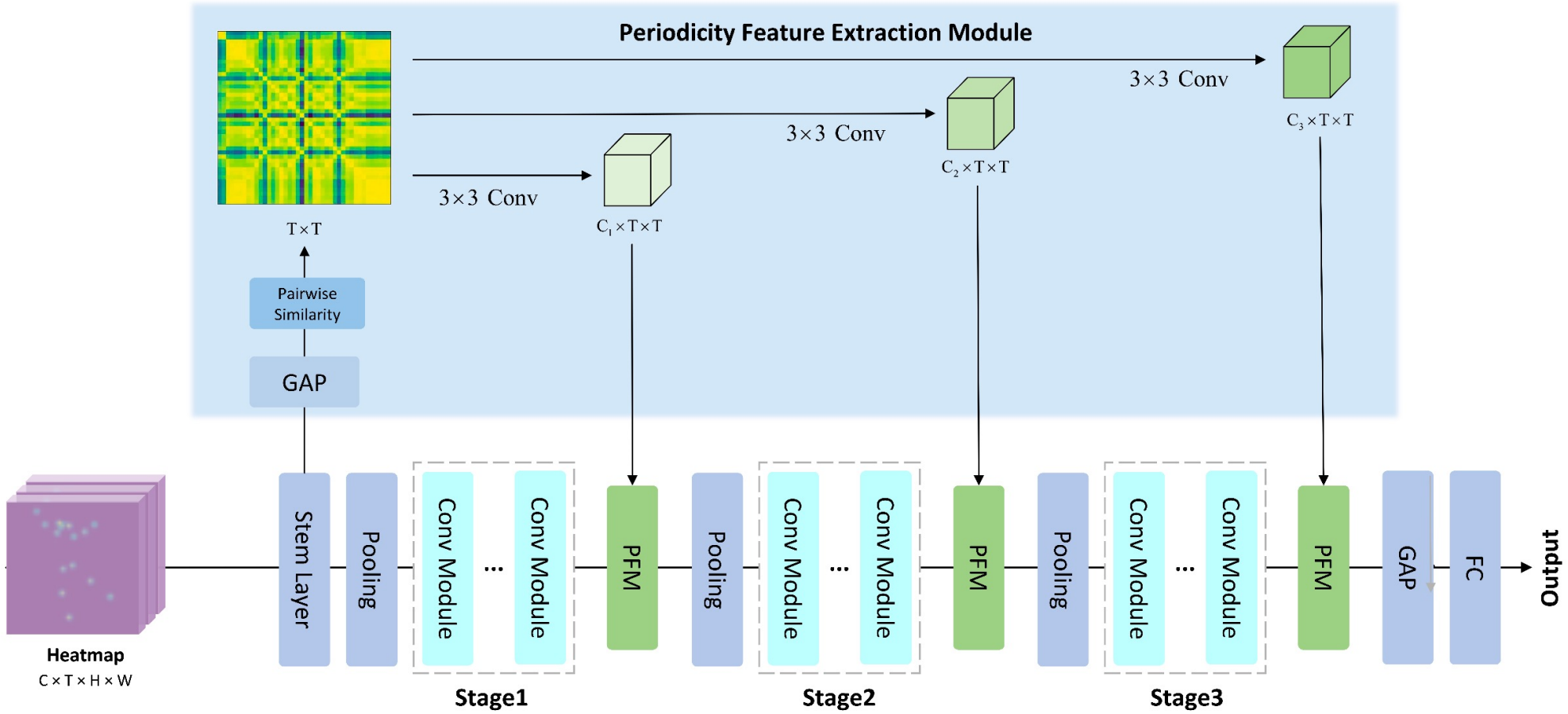


Fig. 1: Periodic-aware network Architecture instantiates with a general 3D-CNN architecture. The stem layer quickly downsamples the input feature maps with convolutions of a quite large kernel size. The module PFM is proposed to combine both periodicity and spatiotemporal information.

Table 2: Different stages fusion of periodicity feature extraction module

	Stage1	Stage2	Stage3	Mean-Top1(%)
SlowOnly				46.72
	✓			48.40
	✓	✓		49.44
PAN	✓	✓	✓	**50.39**

Table 3: Different squeeze methods for periodicity fusion module

Method	**Mean-Top1(%)**	**Top1(%)**
Base	46.72	55.94
Channel	**50.39**	**57.66**
Temporal	49.54	56.95

Table 4: Performance Comparison of Different Methods in FineGym & Diving48 benchmarks

Model	FineGym			Diving48-V2		
	Mean-Top1(%)	Top1(%)	Top5(%)	Mean-Top1(%)	Top1(%)	Top5(%)
ST-GCN[4]	86.40	90.62	99.34	34.84	39.49	76.65
ST-GCN++[28]	90.03	93.12	99.57	40.89	46.65	81.62
CTR-GCN[14]	90.46	93.15	99.59	38.34	45.63	78.73
MS-AAGCN[29]	87.58	91.35	99.33	35.25	43.10	75.43
DG-STGCN[7]	90.60	92.90	99.59	35.26	41.32	75.38
MS-G3D[9]	90.45	93.56	99.62	37.63	45.69	79.59
Pose-C3D-s[5]	92.07	94.85	**99.79**	41.66	52.03	86.19
Pose-X3D-s[5]	88.75	92.84	99.69	41.26	49.85	85.58
Pose-SlowOnly[5]	93.22	95.36	**99.79**	46.72	55.94	87.01
PAN-SlowOnly (ours)	**93.91**	**95.65**	**99.79**	**50.39**	**57.66**	**88.12**

Outline

- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

人脸活体检测

数据样例

如今，人脸识别技术在各行各业中的应用越来越广泛，给人们日常生活带来便利的同时，也面临着各类人脸欺诈攻击。一旦虚假人脸攻击成功，极有可能对用户造成重大损失，因此针对防欺诈的人脸活体检测技术的深入研究已经势在必行。

CFAT-2023 人脸活体检测挑战赛

CFAT 2023 复赛　　CFAT2023复赛B-数据集

排行榜（总榜）　企业排行榜　科研院所/院校排行榜

排名	参赛者	APCER	BPCER	ACER
1	MD-Face	0.11764(1)	0.11408(2)	0.11586(1)
2	KiwiTech	0.12685(2)	0.12176(6)	0.12431(2)
3	GDUT-G2-413	0.13954(4)	0.11014(1)	0.12484(3)
4	Yifan	0.14349(6)	0.11565(3)	0.12957(4)
5	FaceSystem	0.14612(7)	0.12465(7)	0.13539(5)
6	邮你更精彩	0.13966(5)	0.13157(9)	0.13561(6)
7	小mang	0.13870(3)	0.13387(10)	0.13629(7)
8	XSBANK	0.15330(8)	0.12836(8)	0.14083(8)
9	shenxiaohui	0.17293(10)	0.11786(4)	0.14539(9)
10	bleakie	0.17257(9)	0.11861(5)	0.14559(10)

挑战赛方案

1.数据增强

- CutMix

在消融实验中发现CutMix效果会优于Mixup，所以选择0.65的概率执行CutMix操作，0.3的概率执行Mixup操作

- RandomErasing

为了进一步缓解人脸被遮挡的情况，我们采用了Random Erasing的操作，但大概率的Random Erasing会导致检测效果更差，所以我们针对这个操作只有0.05的概率

2.模型选择

我们对一些常用的分类模型进行了训练，如ResNet，VanillaNet，ConvNeXt进行实验，我们发现在ResNet中ResNet18的效果优于ResNet50，无论是在验证集还是测试集。在这个实验的基础上我们认为针对这个任务选择大模型或许不是一个很好的选择，要么选择经典的小模型，要么自己对网络进行修改，在初赛中我们最后选择了ResNet34，而在复赛的时候给全连接层加上了Dropout0.5，以避免过拟合

3.训练策略

优化器方面，我们采用的Lion Optimizer 是23年谷歌新提出的优化器。我们在实验中发现，效果优于常规的AdamW和SGD。
而学习率策略方面，则采用OneCycle

Outline

- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

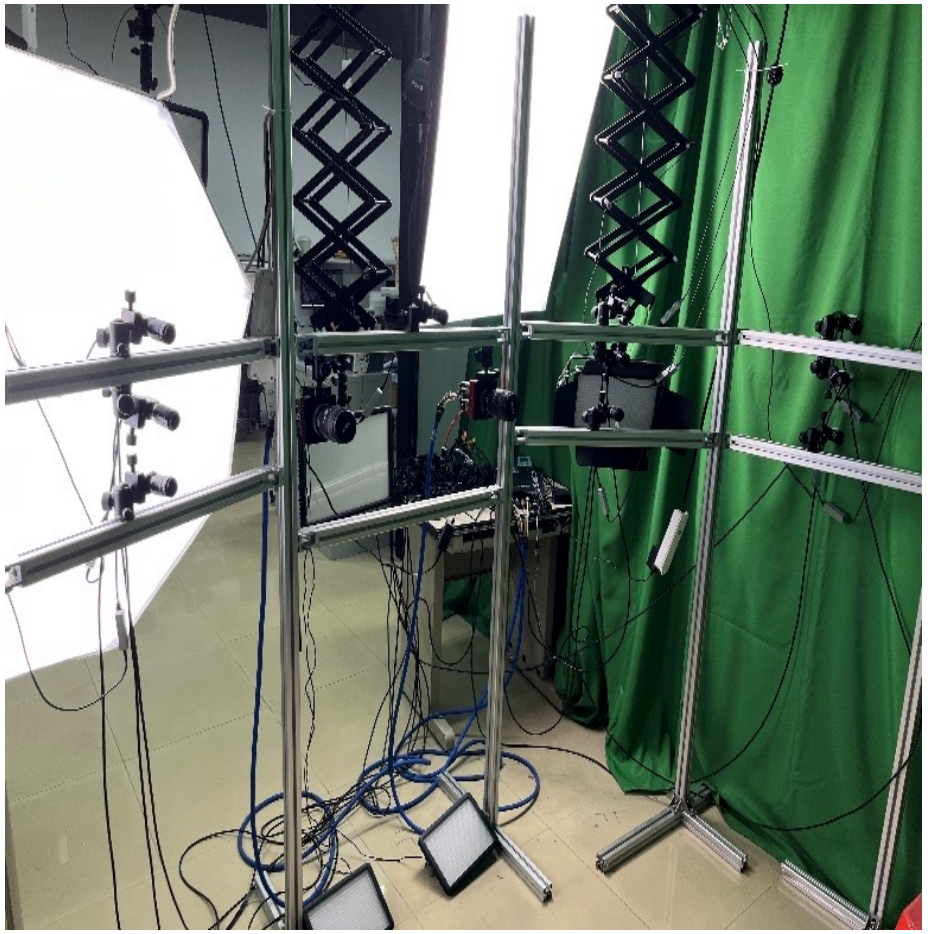
研究内容

1、建立适用于特征描述子提取网络训练的多视角人脸图像块数据集

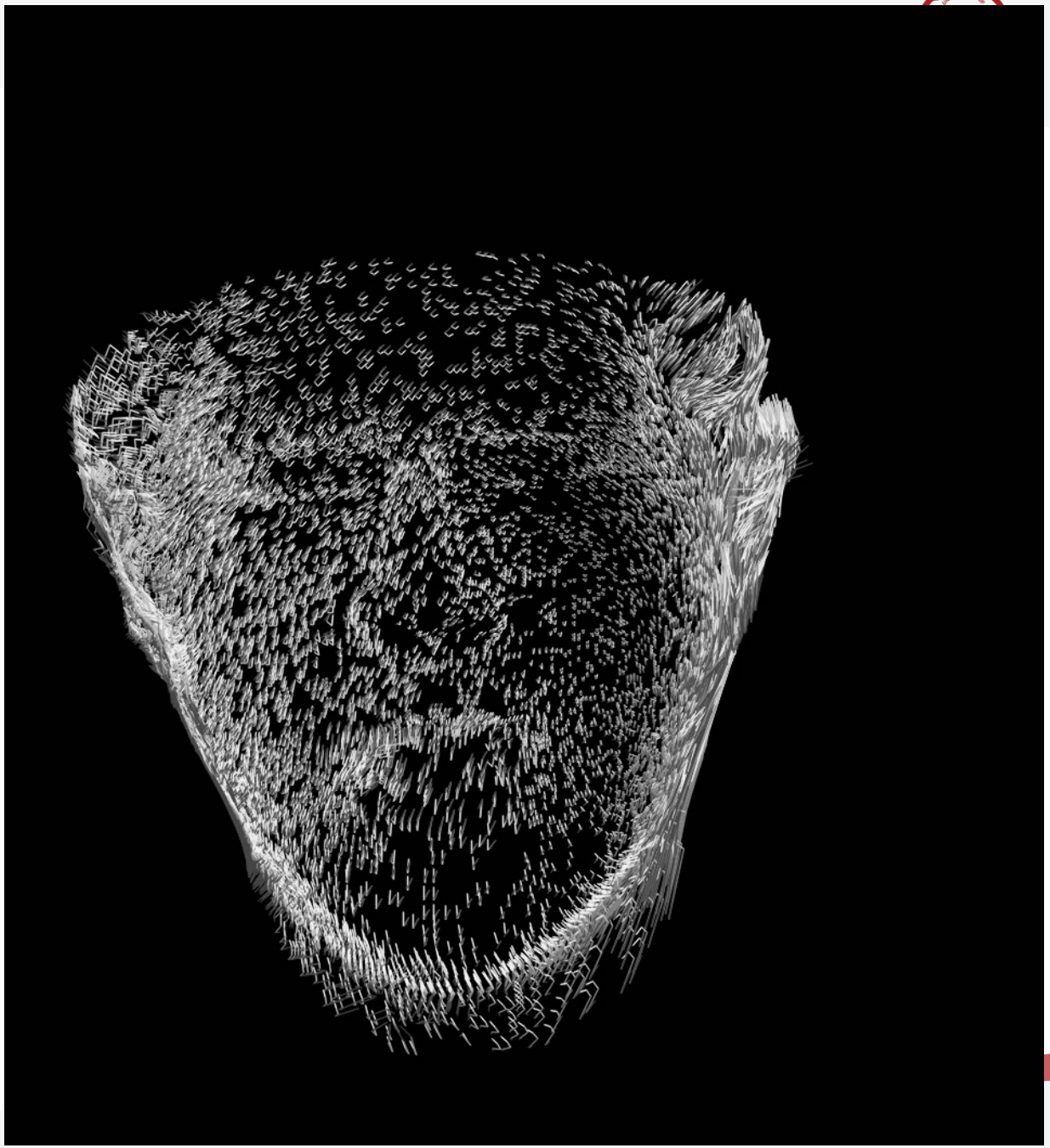
2、提出了一个具有反金字塔结构的中间特征图聚合模块（IMA），用于自适应地将特征信息纳入描述子;

3、提出了一种基于中间特征图白化变换的新型正则化技术，以消除由风格信息引入的噪声

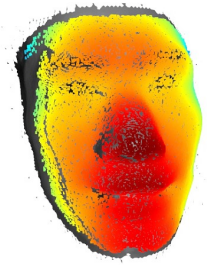
多视角人脸图像块数据集



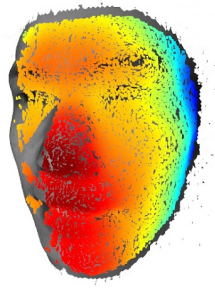
多视角人脸拍摄系统

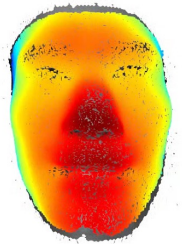


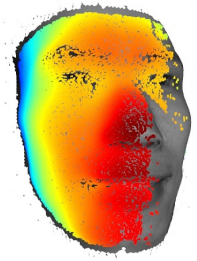
多视角人脸图像块数据集

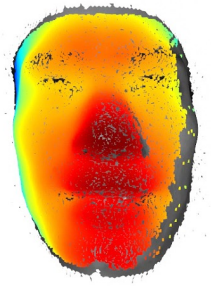



3D模型与
二维图像对应关系

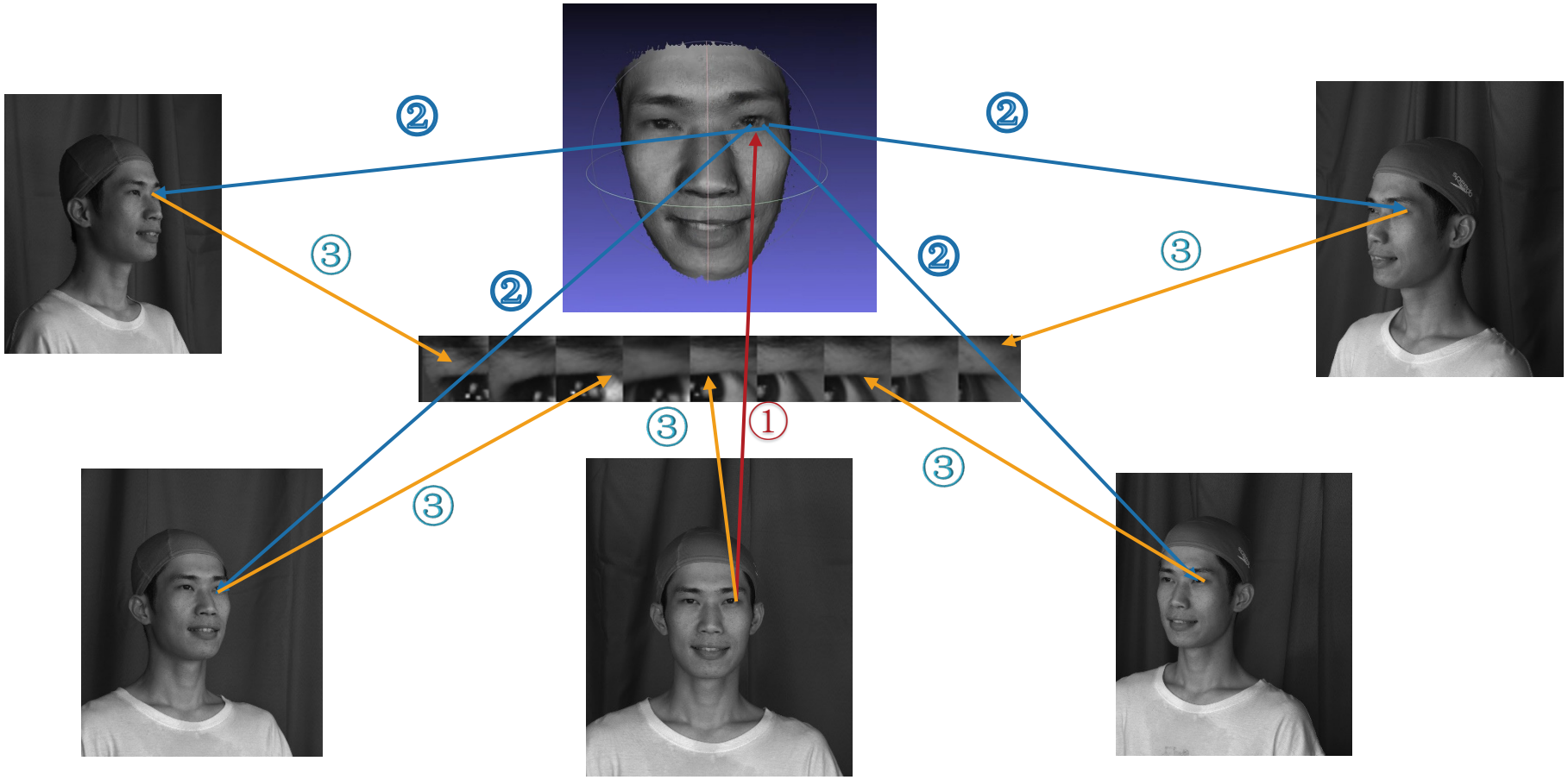




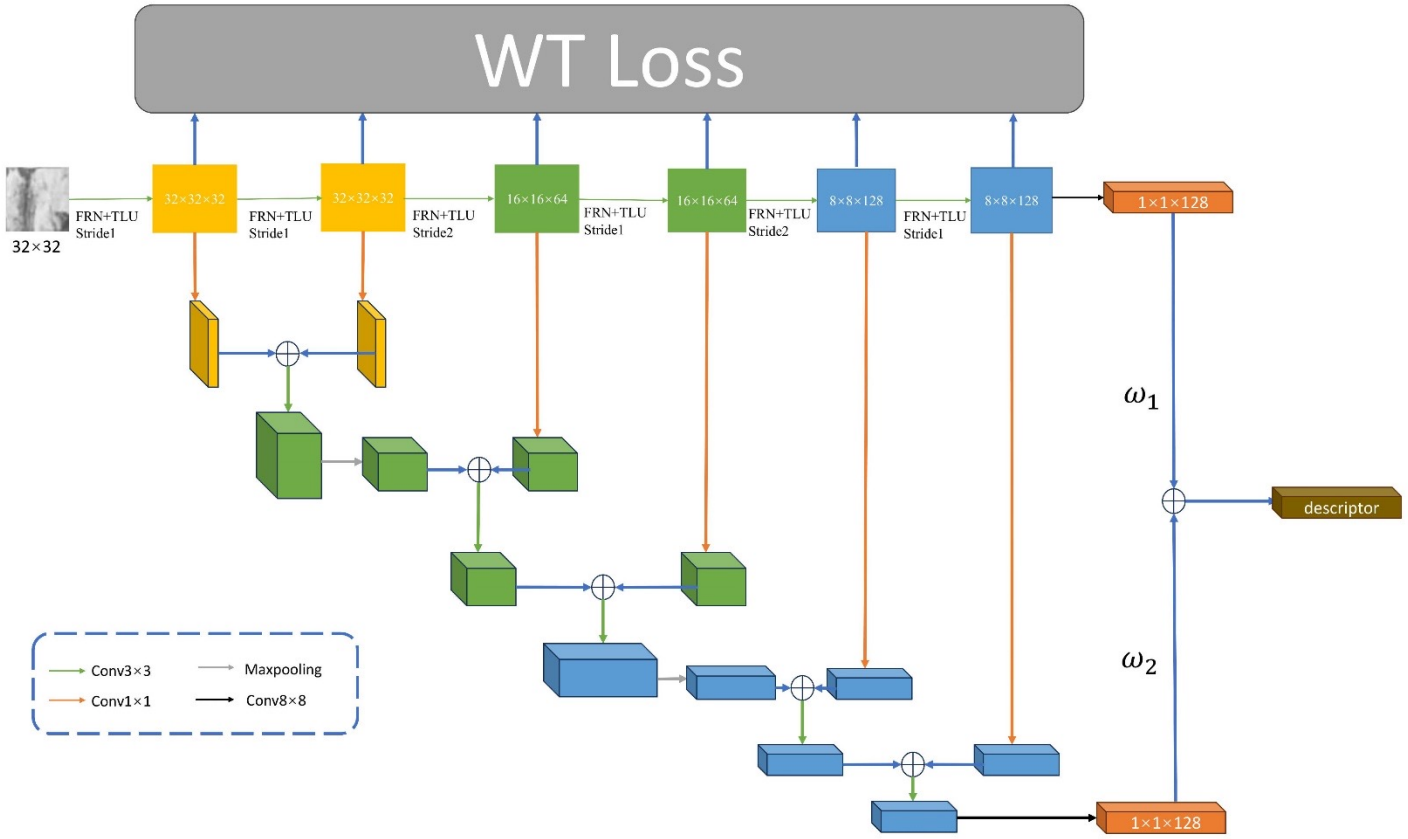




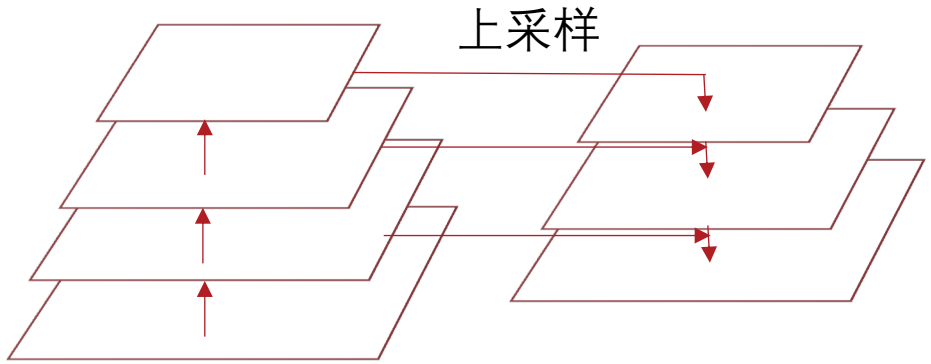

多视角人脸图像块数据集



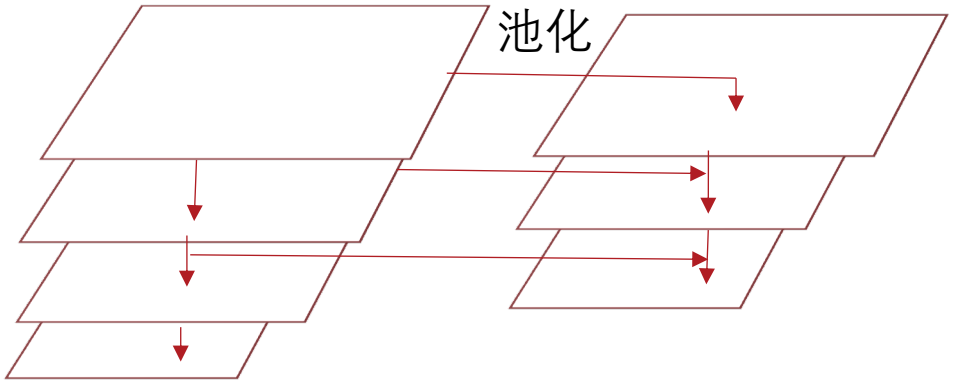
网络结构



中间特征图聚合结构图

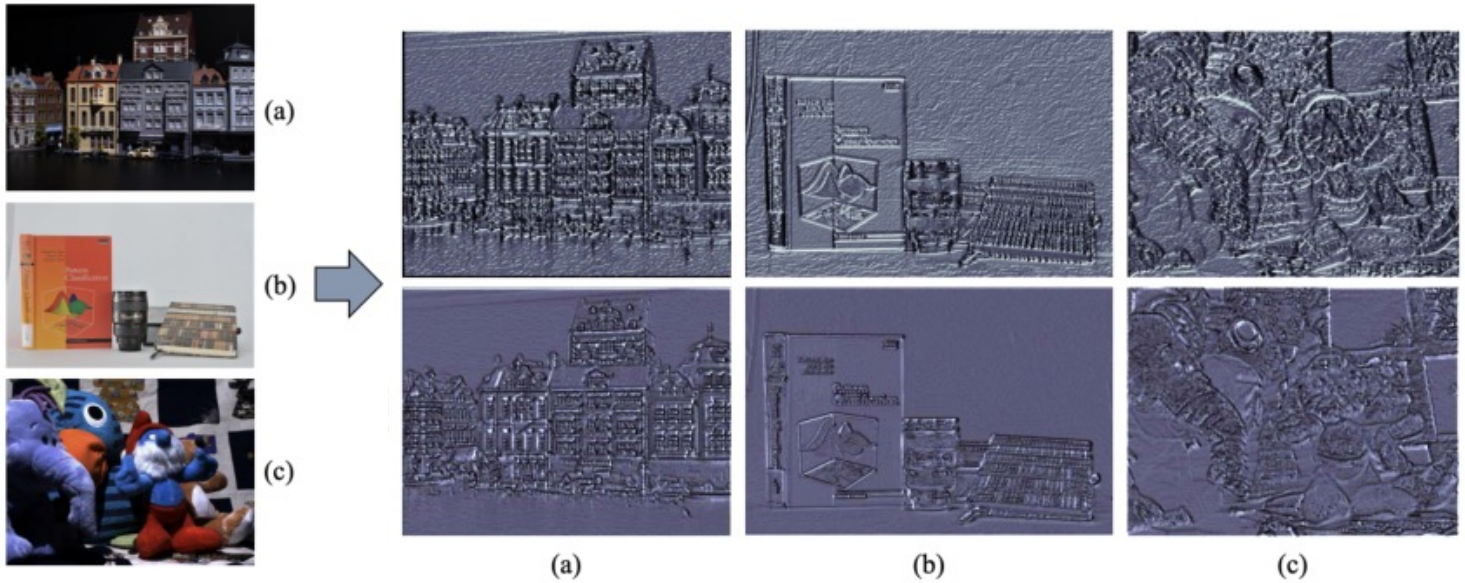


金字塔结构图



逆金字塔结构图

WT正则化



白化变换（WT）示意图

$$L_{Triplet} = \frac{1}{B}\sum_{i=1}^{B} max(0, 1 + s_H(a_i, a_i^+) - min(s_H(a_i, a_i^+), s_H(a_{n_{min}}, a_i^+)))^2,$$

$$s_H(\theta) = \alpha(1 - s(\theta)) + d(\theta).$$

$$\mathbf{\Sigma_s} = \frac{1}{HW}(\mathbf{X_s})(\mathbf{X_s})^\top \in \mathbb{R}^{C\times C},$$

$$\mathcal{L}_{\mathrm{IW}} = \mathbb{E}[\|\mathbf{\Sigma_s} \odot \mathbf{M}\|_1],$$

$$WTR_l = \begin{cases} \mathbb{E}\left[\left\|\Sigma_C \odot \mathbf{M}_{\mathrm{sty}}\right\|_1\right] & \text{if } \Sigma_C > \mu \\ 1 - \mathbb{E}\left[\|\Sigma_C \odot \mathbf{M}_{str}\|_1\right] & \text{otherwise} \end{cases}$$

$$WTR = \frac{1}{N}\sum_{l=1}^{N} WTR_l$$

损失函数计算表达式

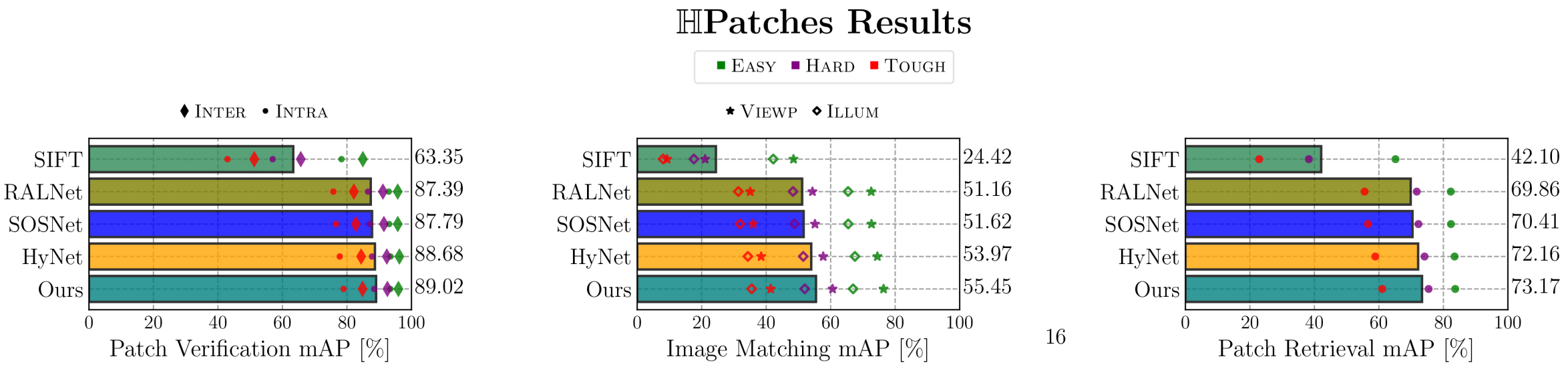
实验结果

在Brwon数据集上的实验结果

Train	Notredam	Yosemite	Liberty	Yosemite	Liberty	Notredam	Mean
Test	Liberty		Notredam		Yosemite		
SIFT [4]	29.84		22.53		27.29		26.55
DeepDesc [5]	10.90		4.40		5.69		6.99
MatchNet [6]	7.04	11.47	3.82	5.65	11.60	8.70	8.05
SNet [7]	6.39	8.43	6.61	2.83	6.61	5.57	5.28
L2Net [8]	2.36	4.70	0.72	1.29	2.57	1.71	2.23
HardNet [9]	1.49	2.51	0.53	0.79	1.96	1.84	1.51
HardNet-GOR[10]	1.48	2.43	0.51	0.79	1.76	1.53	1.41
DOAP[11]	1.54	2.62	0.43	0.87	2.00	1.21	1.45
RALNet[12]	1.30	2.39	0.37	0.67	1.52	1.31	1.26
SOSNet[3]	1.08	2.12	0.35	0.67	1.03	0.95	1.03
CDF[13]	1.21	2.01	0.39	0.68	1.51	1.29	1.18
HSD+[14]	1.19	1.91	0.37	0.64	1.38	1.14	1.11
MR3A[15]	1.47	2.09	0.50	0.77	1.69	1.75	1.38
MFD-Net[16]	1.21	2.10	0.40	0.74	1.85	1.77	1.35
HyNet[17]	0.89	1.37	0.34	0.61	0.88	0.96	0.84
HyNet-SOSR[17]	0.91	1.62	0.31	0.54	0.78	0.73	0.82
MSFASP(224-dim)[18]	0.99	1.92	0.34	**0.59**	1.29	0.92	1.01
MSFASP-SOSR(224-dim)[18]	0.93	1.88	0.30	0.62	0.95	0.62	0.88
Ours	0.91	1.74	0.30	0.65	0.78	0.62	0.82
Ours-SOSR	**0.85**	**1.74**	**0.24**	0.67	**0.66**	**0.60**	**0.79**

实验结果

在HPatches数据集上的实验结果

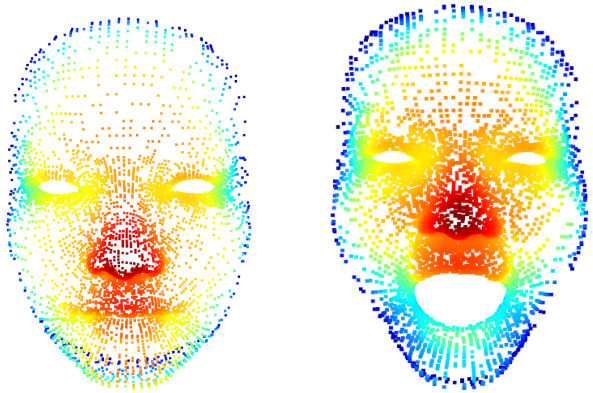


Outline

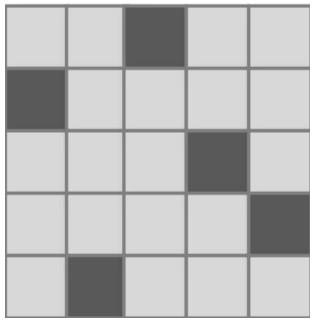
- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

问题分析

Input



Output



深度学习网络

$\mathbf{X} \in \mathbb{R}^{M\times 3}$　　$\mathbf{Y} \in \mathbb{R}^{N\times 3}$

输入描述：

M、N表示的是点云中点的数量；

3表示的是x, y, z三轴坐标；

以x为原点云、y目标点云

输出描述：

一个M*N的(0-1)矩阵，其中每一行或者每一列只有一个1。

矩阵表示了点云X中的每个点与点云Y中每个点的对应关系。

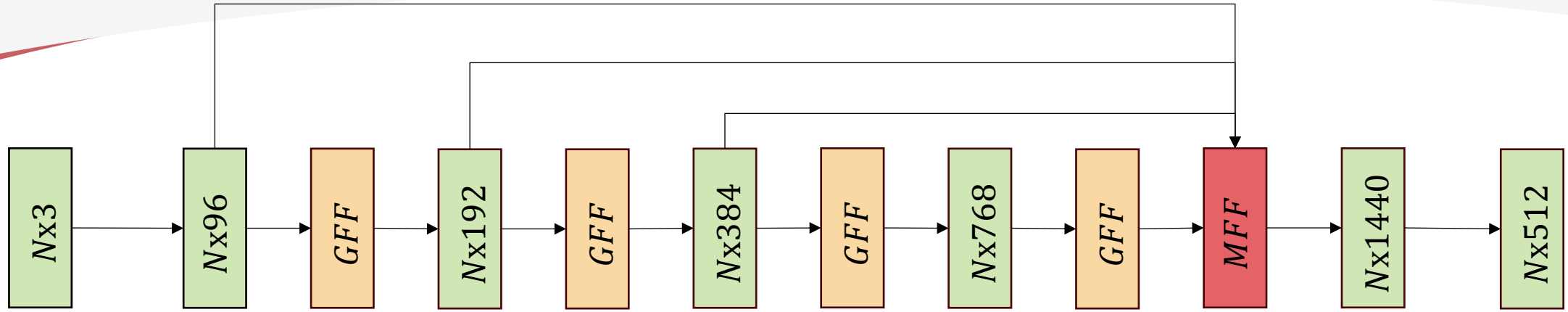
创新点

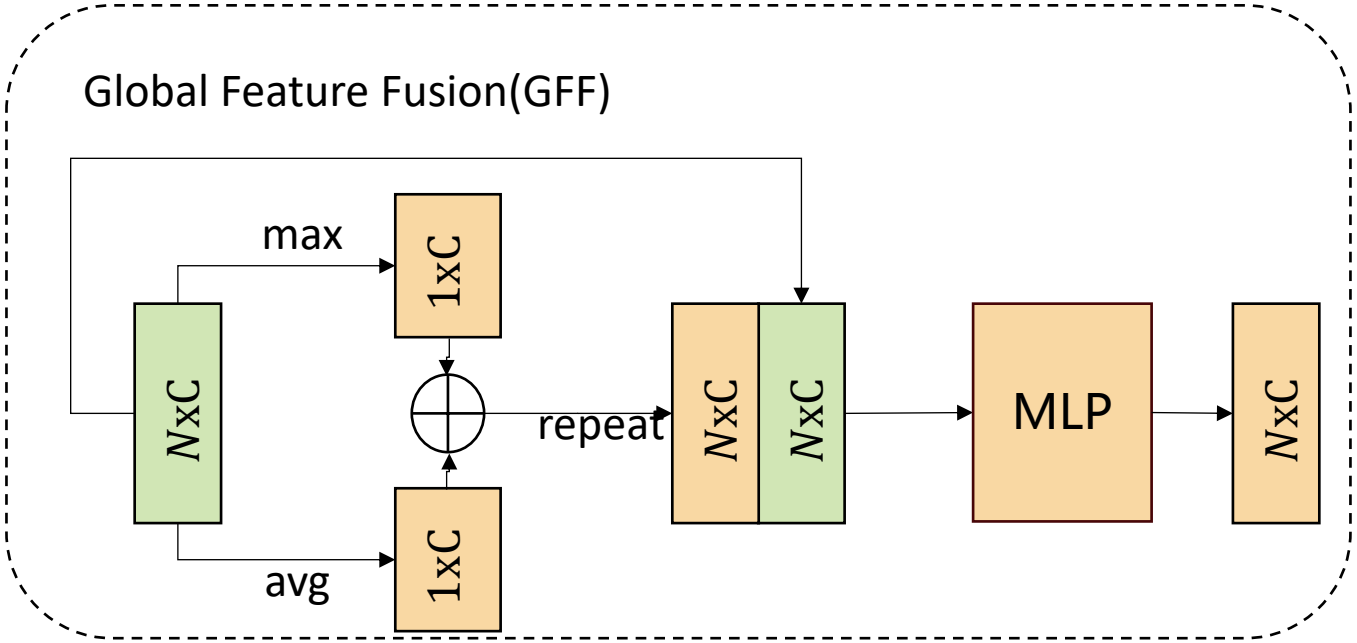
1、基于先前方法的骨干网络，设计全局特征融合模块，在逐点特征中融入全局信息，提升特征的表示能力。

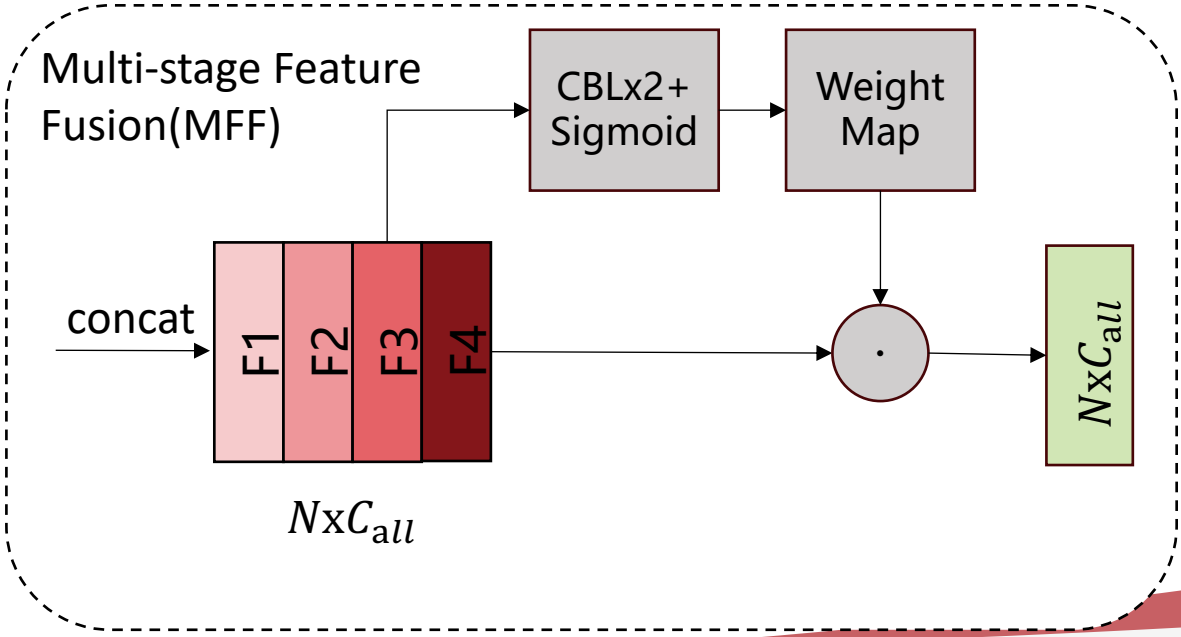
2、改进先前方法的多尺度特征融合模块，使用注意力机制使得网络可以自适应学习各个尺度特征的权重;

3、相比于原有方法, 在SHREC 和 TOSCA数据集上取得了显著的提升

网络结构


Nx3
Nx96
GFF
Nx192
GFF
Nx384
GFF
Nx768
GFF
MFF
Nx1440
Nx512


Global Feature Fusion(GFF)
max
1xC
NxC
repeat
NxC
NxC
MLP
NxC
1xC
avg


Multi-stage Feature Fusion(MFF)
CBLx2+ Sigmoid
Weight Map
concat
F1
F2
F3
F4
NxC_all

对比实验

表1 SHREC数据集性能比较

Method	ACC(↑)	Err(↓)
Diff-FMaps	4.0%	7.1
3D-CODED	2.1%	8.1
Elementary	2.3%	7.6
CorrNet3D	6.0%	6.9
DPC	17.7%	6.1
Ours	**20.93%**	**5.4**

对比实验

表2 TOSCA数据集性能比较

Method	ACC(↑)	Err(↓)
Diff-FMaps	-/-	-/-
3D-CODED	-/-	-/-
Elementary	-/-	-/-
CorrNet3D	0.3	32.7
DPC	34.7%	2.8
Ours	**37.3%**	**2.6**

消融实验

表3 SHREC数据集上的消融实验结果

baseline	GFF	MFF	ACC (↑)	Err(↓)
√			17.7%	6.1
	√		20.4%	5.6
		√	18.9%	5.9
√	√	√	**20.9%**	**5.4**

Outline

- 人脸表情（2D、3D、4D）
- 动作识别
- 人脸活体检测
- 图像特征提取
- 激光雷达点云处理
- SLAM

KILO: Robust Kinematics-Inertial-LiDAR Odometry for Dynamic Legged Robots

四足机器人在激光SLAM存在的问题

1. 由于足端与地面高频率的撞击，带来的冲击严重影响加速度计的精度。目前流行的LIO框架（如LIOSAM，FASTLIO2）都无法正常运行。
即：直接将IMU作为模型的输入，会有较大漂移

2. 四足机器人需要轻型，实时性高的SLAM系统

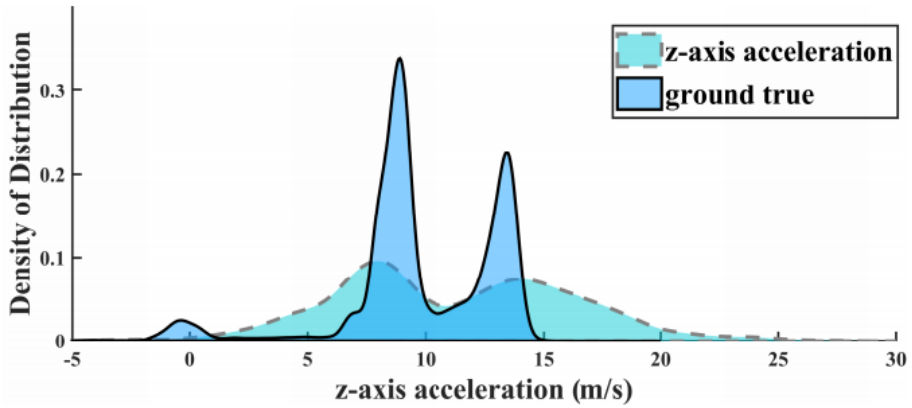
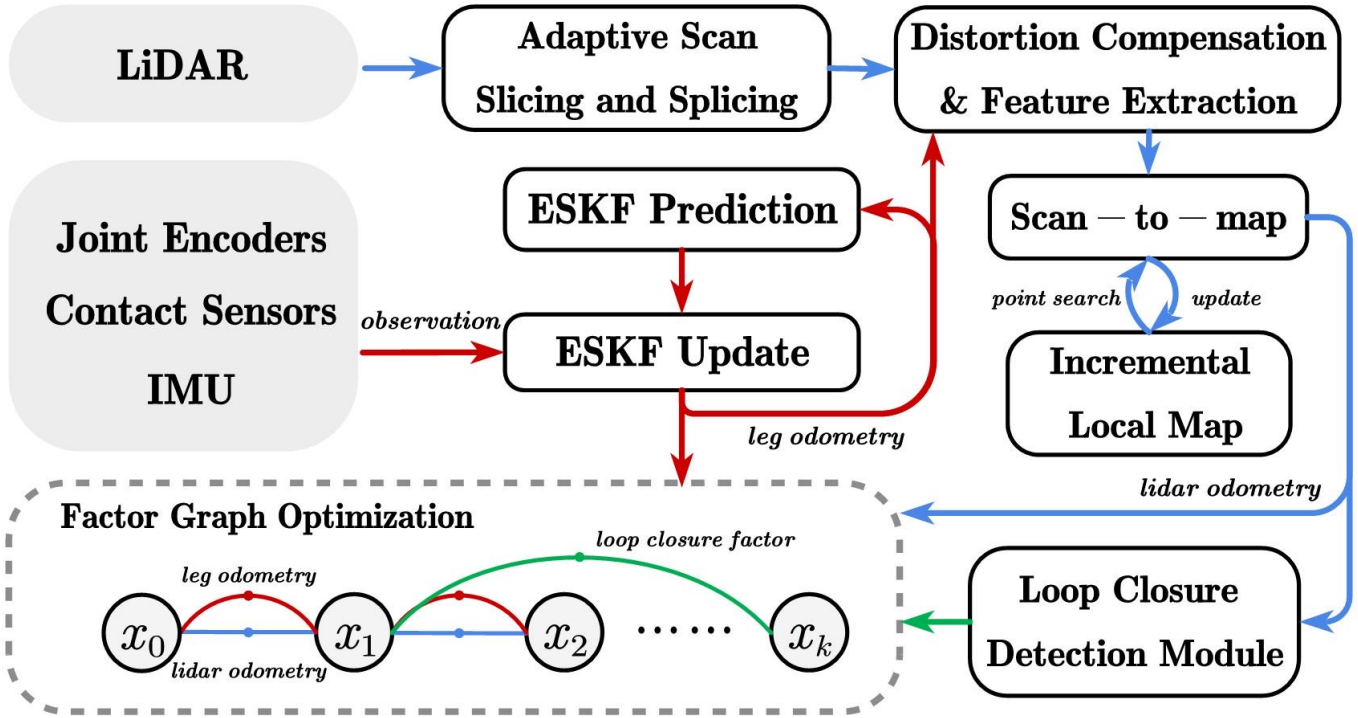
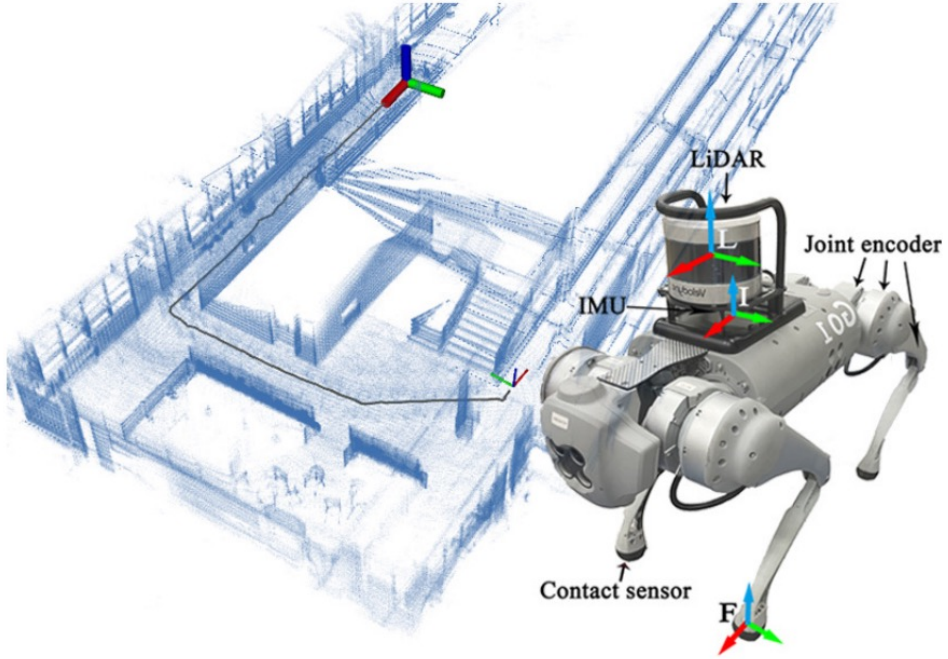


Fig. 2: Comparison of the distribution of accelerometer measurements along the Z-axis between the real environment and the simulation environment. Unitree Go1 with built-in IMU performs trot gait movement in real environment and Gazebo simulation platform respectively.





解决方案：

1. 一种融合腿足运动学和IMU的腿足里程计
（利用误差卡尔曼滤波ESKF，融合编码器、足端力传感器、IMU的滤波框架）

2. 自适应LiDAR scan slicing， 利用激光雷达采样特性，提高lidar odometry 的输出频率，降低误差累积

3. 利用因子图，紧耦合腿足里程计、雷达里程计、回环因子，提高全局一致性

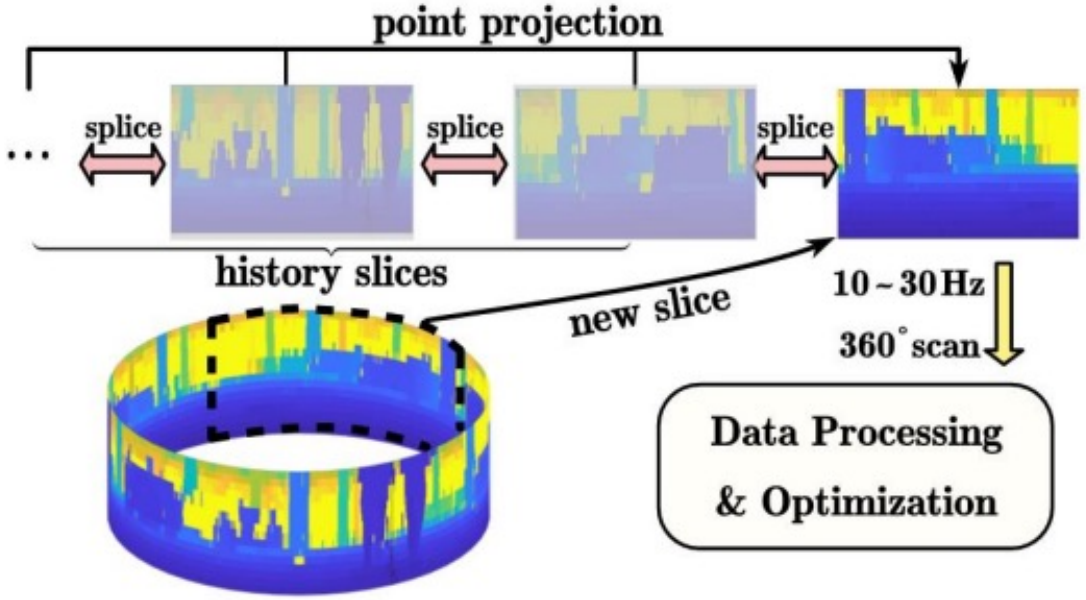


Fig. 4: ASSS

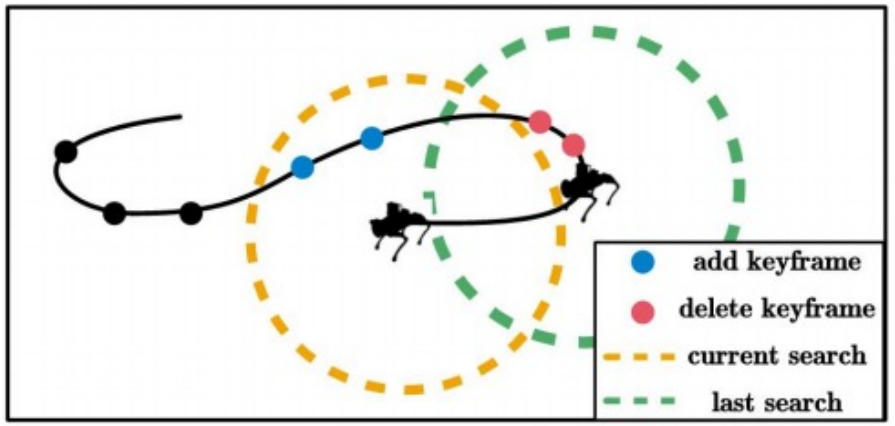


Fig. 5: robot-centric incremental map

TABLE I: Comparison of Relative pose Error (RPE) for our proposed method and other methods

Sequences (mean velocity [m/s])	Mean Euler Angle Error [deg] / Translation Error [m]							
	KF [3]	ESKF[4] (Ours)	A-LOAM	FAST-LIO2	POINT-LIO	LIO-SAM	LIO-SAM[5] (with ESKF)	KILO (Ours)
park[1]	0.76/0.75	0.97/0.67	0.74/0.18	1.06/0.28	1.09/0.24	×[3]	0.68/0.17	**0.13/0.06**
corridor[1]	0.54/1.08	1.1/0.98	0.88/0.23	1.03/0.26	1.00/0.23	×	0.78/0.20	**0.08/0.04**
indoor[2]	0.71/0.10	0.53/0.09	0.45/0.04	0.84/0.09	0.72/0.06	0.47/0.04	0.47/0.04	**0.07/0.03**
running[2]	0.67/0.17	1.24/0.17	1.08/0.08	2.11/0.18	1.73/0.09	1.10/0.09	1.03/0.10	**0.28/0.04**

The distance of RPE is set based on the sequence's distance ([1] RPE of 10 m, [2] RPE of 1 m). [3] × denotes the method failed.
[4] Our proposed kinematics-inertial leg odometry. [5] LIO-SAM modified by leg odometry instead of original IMU odometry.
Bold values represent the best result among all methods. Underline values represent the best result among kinematics-inertial methods

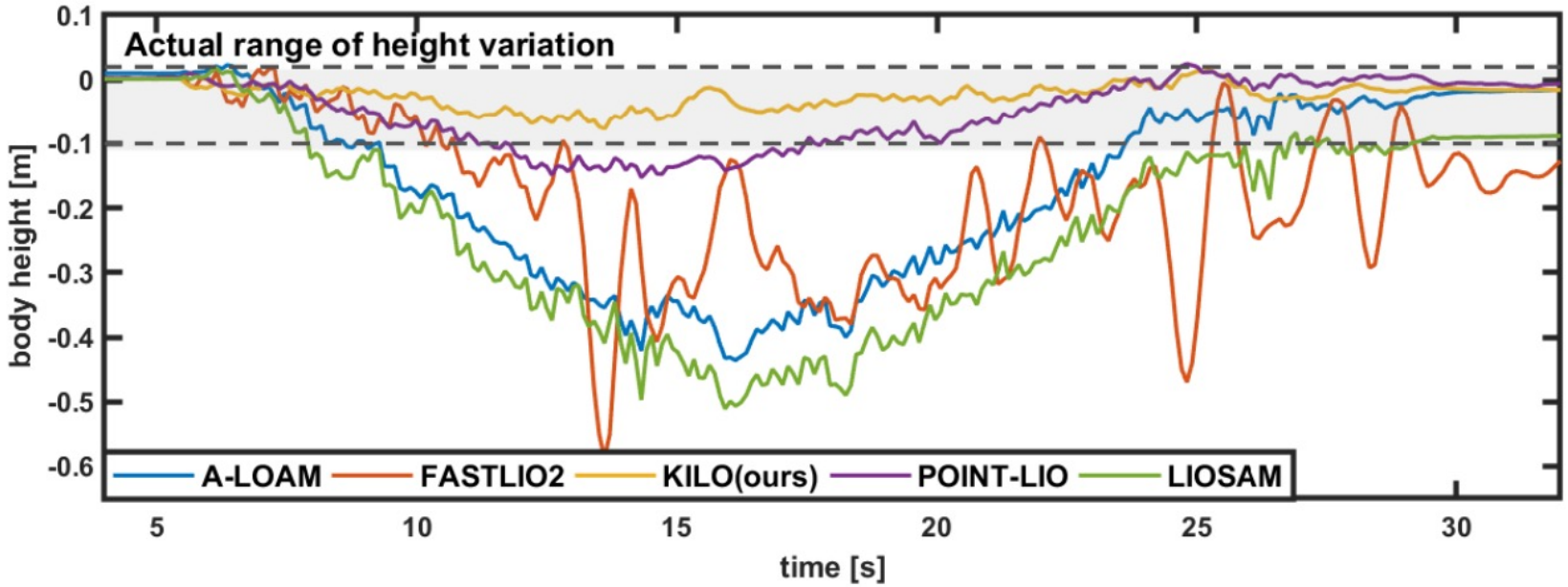


Fig. 7: Comparison of body height's variation for various methods while running on flat ground. The body height will drop when running, and will not exceed the gray range in the figure.

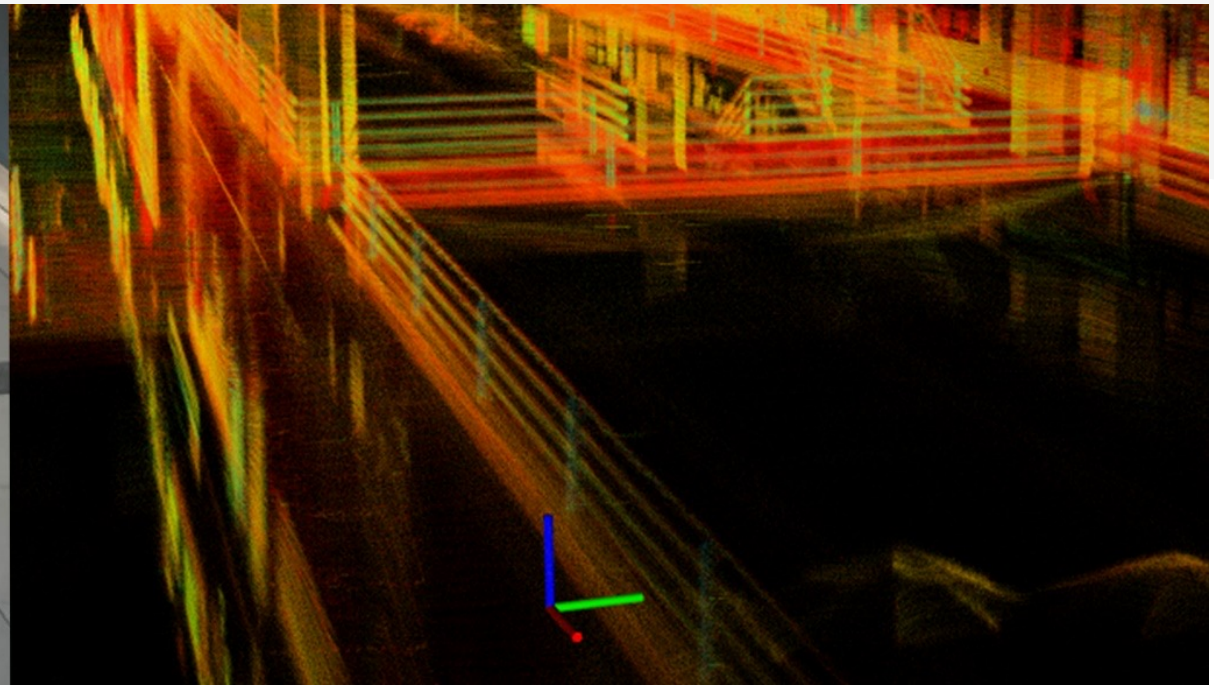


TABLE III: Comparison of methods on sequence *corridor*

End-to-End Error [m] / Time Consumption per Scan [ms]			
A-LOAM	LIO-SAM (with ESKF)	FAST-LIO2	KILO (Ours)
1.86/85.84	0.10/40.11	0.20/**6.21**	**0.04**/13.31

TABLE IV: Ablation analysis of Adaptive Scan Slicing and Splicing (ASSS) during high dynamic motion (sequence *running*)

Trajectory	Mean Absolute Trajectory Error (ATE) [m]	
Length [m]	KILO (w/s ASSS)	KILO (full)
29.85	0.0614	0.0546

广东工业大学
Guangdong University of Technology

报告人：李东 副教授

欢迎提问！