上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

用于动态视觉感知的高效表征学习算法研究

姓名：谢斐

导师：马超

日期：2025/07

1 个人简介

2 已有研究工作

3 未来研究计划

4 总结与讨论

1 个人简介

2 已有研究工作

3 未来研究计划

4 总结与讨论

1.个人简介

谢斐 Xie Fei

- **性别：**男
- **出生年月：**1995/03/19
- **籍贯：**江西赣州
- **邮箱：**jaffe031@sjtu.edu.cn
- **电话/微信：**14779773579
- **学历：**博士生
- **学校：**上海交通大学
- **个人主页：**https://phiphiphi31.github.io/
- **专业：**计算机，人工智能
- **研究方向：**计算机视觉，视频分析，视觉基础模型

1.个人简介：教育/科研经历

教育经历

- 2022.09-2026.05 (预期), **博士研究生**, 计算机科学与技术, 马超教授，**上海交通大学**, 上海
- 2019.06 - 2022.04, **硕士**, 模式识别, **东南大学**, 杨万扣教授，南京
- 2016.09 - 2018.04, **硕士**, 电气工程, **英属哥伦比亚大学** (UBC), Prof. William Dunford, 温哥华
- 2012.06 - 2016.06, **学士**, 电气工程, **华中科技大学**, 武汉

实习经历

- 2024.04-2025.06, **华为**, **诺亚方舟实验室**, 王重道博士
- 2022.05-2023.04, **微软亚洲研究院(MSRA)**, Media Computing Group, 初磊博士，李嘉豪博士
- 2021.04-2022.02, **微软亚洲研究院(MSRA)**, Intelligent Media Group, 王春雨博士，曹越博士

1.个人简介：教育/科研经历

研究兴趣

我的研究方向集中于计算机视觉与通用人工智能领域，主要研究方向涵盖图像与视频中的视觉感知，包括：

- **视频目标跟踪** (Visual Tracking)
- **点云目标跟踪** (Tracking on Point Clouds)
- **视觉骨干网络设计** (Vision Backbone)

其中对以下方向具有特别的研究兴趣：

- **多模态大语言模型**（MLLM）
- **生成式人工智能**（AIGC）
- **基础模型架构设计**（Transformer、RNN 、 Mamba 、 DeltaNet）

1.个人简介：教育/科研经历

学术服务

- 会议审稿人：**CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML, MM, WACV, ACCV.**
- 期刊审稿人: **TPAMI, TNNLS, TMM, TCSVT, NN, IVC.**

其他

- 近年来，我有幸与多位学者合作研究，包括**杨万扣**教授（东南大学）、**张开华**教授（东南大学）、**左旺孟**教授（哈工大）、**王春雨**博士（MSRA）、**王光庭**博士（腾讯）、曾文军教授（东方理工校长）、**初磊**博士（MSRA）以及**李嘉豪**博士（MSRA） 。最近，我与**王重道**博士保持着密切的科研合作关系。
- 2025.07 谷歌学术引用超**1000**，单篇超**200**
- 2025.04 合作编写《**生成式人工智能**》高等教育AI产业系列丛书
- 2021.07 国际视觉跟踪挑战赛ICCV workshop-VOT2021 第二名
- 2023.05 微软亚洲研究院-明日之星“Stars of Tomorrow”from Microsoft Research Asia.
- 2022.06 东南大学优秀硕士学位论文 Excellent Master’s Thesis from Southeast University.
- 2019.06 研究生入学考试专业第一，初试430/500分.

1.个人简介：已发表论文

一作已发表在CCF-A会议期刊：

- **Fei Xie**, Chunyu Wang, Guangting Wang, Yue Cao, Wankou Yang, Wenjun Zeng.. Correlation-Aware Deep Tracking. **CVPR, 2022**
- **Fei Xie**, Lei Chu, Jiahao Li, Yan Lu, Chao Ma. VideoTrack: Learning To Track Objects via Video Transformer. **CVPR, 2023**
- **Fei Xie**, Zhongdao Wang, Chao Ma. DiffusionTrack: Point Set Diffusion Model for Visual Object Tracking. **CVPR, 2024**
- **Fei Xie**, Wankou Yang, Chunyu Wang, Lei Chu, Yue Cao, Chao Ma, Wenjun Zeng. Correlation-Embedded Transformer Tracking: A Single-Branch Framework. **T-PAMI, 2024**
- **Fei Xie**, Weijia Zhang, Zhongdao Wang, Chao Ma. QuadMamba: Learning Quadtree-based Selective Scan for Visual State Space Model. **NeurIPS, 2024**
- **Fei Xie**, Jiahao Nie, Yujin Tang, Wenkang Zhang, Hongshen Zhao. Mamba-Adaptor: State Space Model Adaptor for Visual Recognition. **CVPR, 2025**
- **Fei Xie**, Zhongdao Wang, Weijia Zhang, Chao Ma. PVMamba: Parallelizing Vision Mamba via Dynamic State Aggregation. **ICCV, 2025**

发明专利：

- 基于动态聚合的并行状态空间模型的特征提取方法和系统。马超，谢斐
- 一种基于点集扩散的视觉目标跟踪系统及方法。马超，谢斐，杨小康

1.个人简介：已发表论文

其他论文：

- Weijia Zhang, **Fei Xie,** Weidong Cai, Chao Ma. Knowledge Distillation via Virtual Relation Matching. **ICCV, 2025**
- Jiahao Nie, **Fei Xie**, Sifan Zhou, Xueyi Zhou, Dong-Kyu Chae, Zhiwei He. P2P: Part-to-Part Motion Cues Guide a Strong Tracking Framework for LiDAR Point Clouds. **IJCV2024**
- Jiahao Nie, Zhiwei He, Xudong Lv, Xueyi Zhou, Dong-Kyu Chae, **Fei Xie**. Towards Category Unification of 3D Single Object Tracking on Point Clouds. **ICLR, 2024**.
- Yujin Tang, Lu Qi, **Fei Xie,** Xiangtai Li, Chao Ma, Ming-Hsuan Yang. PredFormer: Transformers Are Effective Spatial-Temporal Predictive Learners. **arXiv preprint, 2024**
- Wenkang Zhang, **Fei Xie**, Tianyang Xu, Jiang Zhai, Wankou Yang. CRTrack: Learning correlation-refine network for visual object tracking. **Pattern Recognition, 2024**
- **Fei Xie**, Wankou Yang, Kaihua Zhang, Bo Liu, Guangting Wang, Wangmeng Zuo. Learning Spatio-Appearance Memory Network for High-Performance Visual Tracking, **ICCV, 2021, workshop**
- **Fei Xie**, Chunyu Wang, Guangting Wang, Yue Cao, Wankou Yang, Wenjun Zeng. Learning Tracking Representations via Dual-Branch Fully Transformer Networks. **ICCV, 2021, workshop**
- Xiaolou Sun, Qi Wang, **Fei Xie**, Zhibin Quan, Wei Wang, Hao Wang, Yuncong Yao, Wankou Yang, Satoshi Suzuki. Siamese transformer network: Building an autonomous real-time target tracking system for UAV. **Journal of Systems Architecture.2022**
- **Fei Xie**, Ning Wang, Yuncong Yao, Wankou Yang, Kaihua Zhang, Bo Liu. Hierarchical representations with discriminative meta-filters in dual path network for tracking. **PRCV2020**

1 个人简介

2 已有研究工作

3 未来研究计划

4 总结与讨论

2.已有研究工作：研究背景

应用场景

01 机器视觉

视觉感知技术用于质量控制和机器视觉引导，确保生产线上的产品质量，同时提高了自动化生产的精度和速度

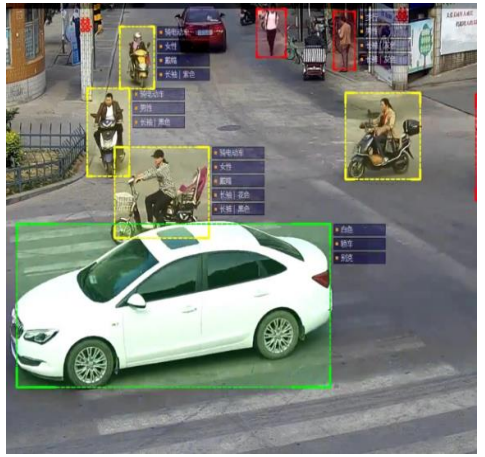
02 安防监控

在安防监控中，人脸识别和行为分析技术广泛应用于身份验证和异常行为检测，提高了公共安全管理的效率和精确度

03 自动驾驶

在自动驾驶领域，视觉感知用于环境感知和路径规划，帮助车辆检测并避开障碍物，识别道路标志和行人，从而实现安全驾驶

04 视频分析



视频内容分析对可视的监视摄像机视频图像进行分析，借助计算机的高速计算能力使用各种过滤器，准确判断人类的各种活动。

2.已有研究工作：研究背景

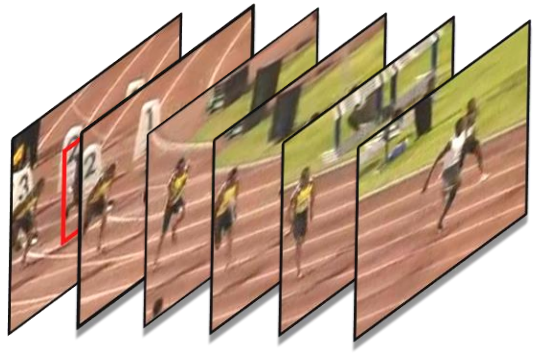
数据模态

01 图片

图片 Image

- 模态：2D 图像
- 形式：(R, G, B, ...)
- 优点：稠密，语义丰富
- 缺点：缺乏时序信息

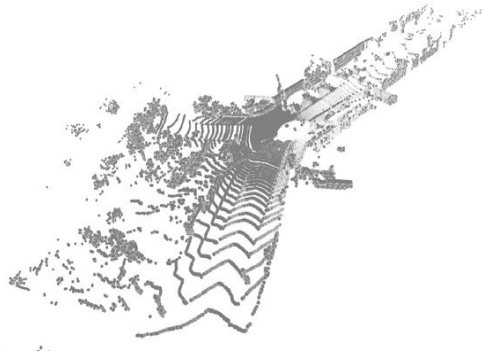
02 视频



视频 Video

- 模态：2D 时序图像
- 形式：(R, G, B, ...)
- 优点：时序建模，语义丰富
- 缺点：冗余信息过多

03 点云



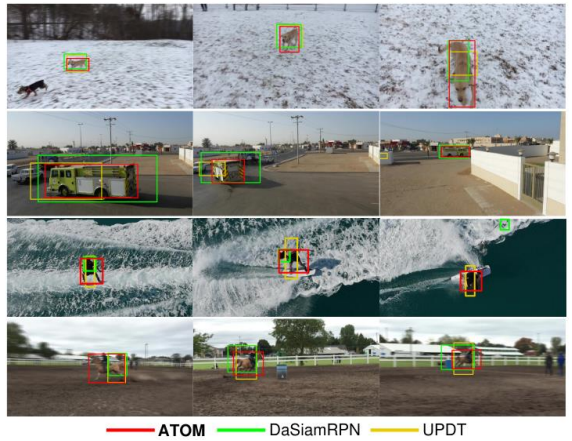
激光雷达 LiDAR

- 模态：3D点云
- 形式：（X, Y, Z, I, ...)
- 优点：定位精度高
- 缺点：点云稀疏、无序

2.已有研究工作：研究背景

任务应用

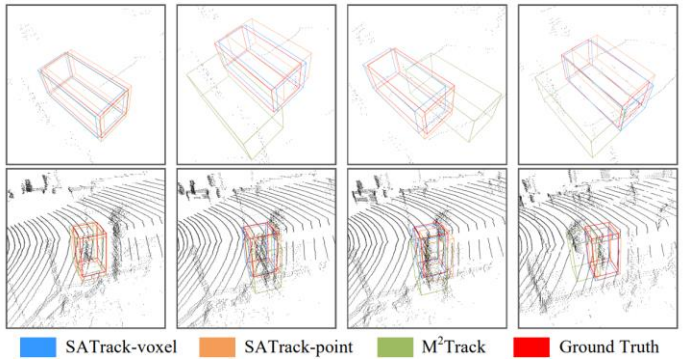
01 视觉目标跟踪



视频 Video

- 描述：在给定的视频序列中持续定位出每一帧的跟踪牧的位置和大小
- 输出：二维目标框

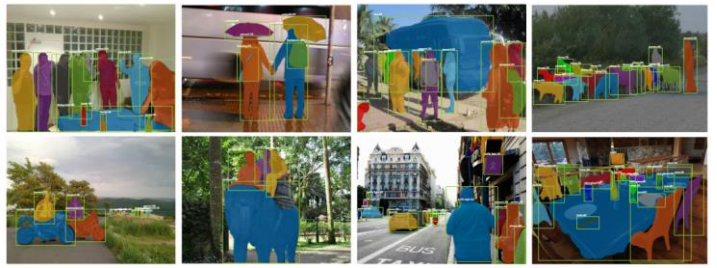
02 点云目标跟踪



点云序列 Point cloud

- 描述：在给定的点云序列中持续定位出每一帧的跟踪牧的位置和大小
- 输出：三维目标框

03 图片检测分割



图片 Image

- 描述：检测和分割出图片中所有物体的种类，大小和分割掩码
- 输出：种类，目标框，分割掩码

2.已有研究工作：研究背景

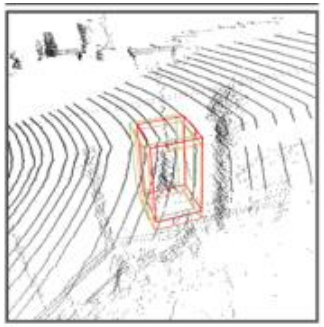
- 研究目标

研究更优的表征学习算法，以适配视频目标跟踪，点云目标跟踪，图片检测分割等下游子任务。

表征学习
(Representation Learning)

视觉目标跟踪
(Visual Object Tracking)

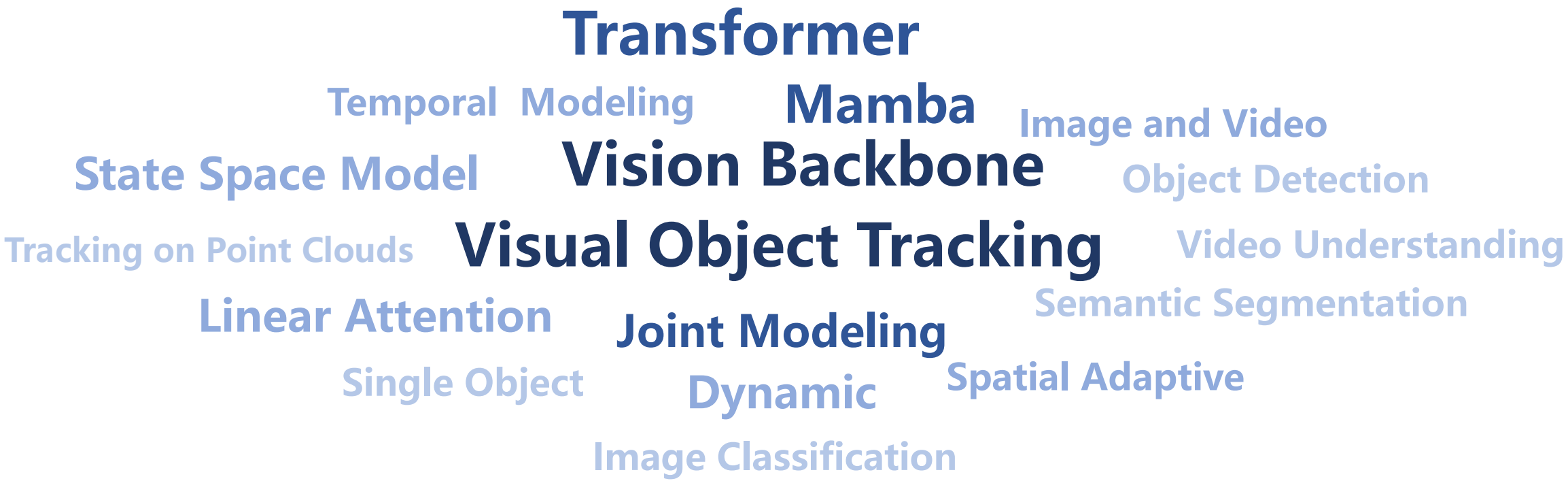
点云目标跟踪
(3D Object Tracking on Point Clouds)



图像检测和语义分割
(Object Detection and Semantic Segmentation)

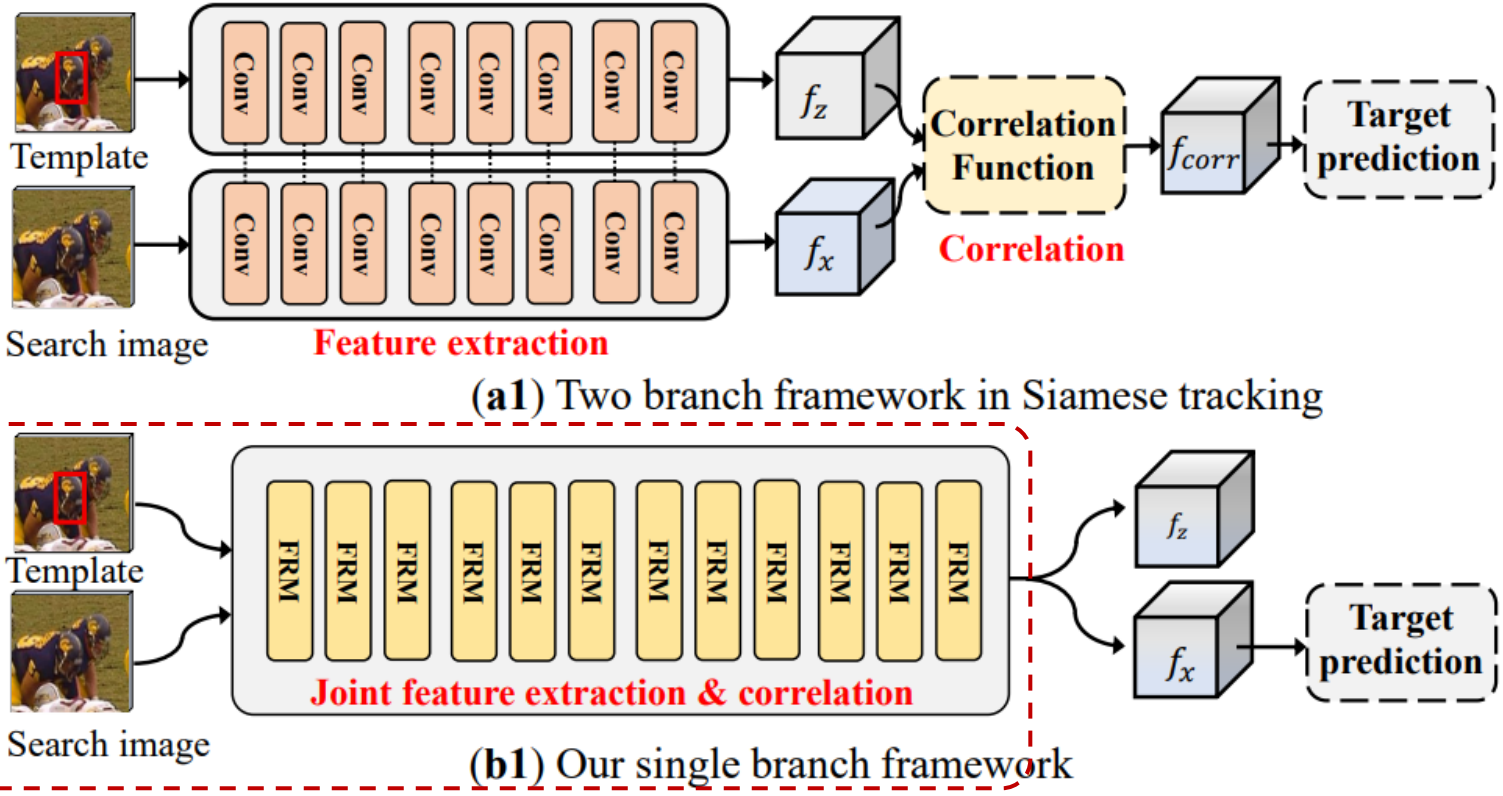
2.已有研究工作：研究主线

- **Object Tracking :** 基于时序的视频、点云目标跟踪分析方法
- **Mamba :** 基于状态空间模型的高效网络建模方法
- **关键词分布:**



2.已有研究工作：工作（1） SBT

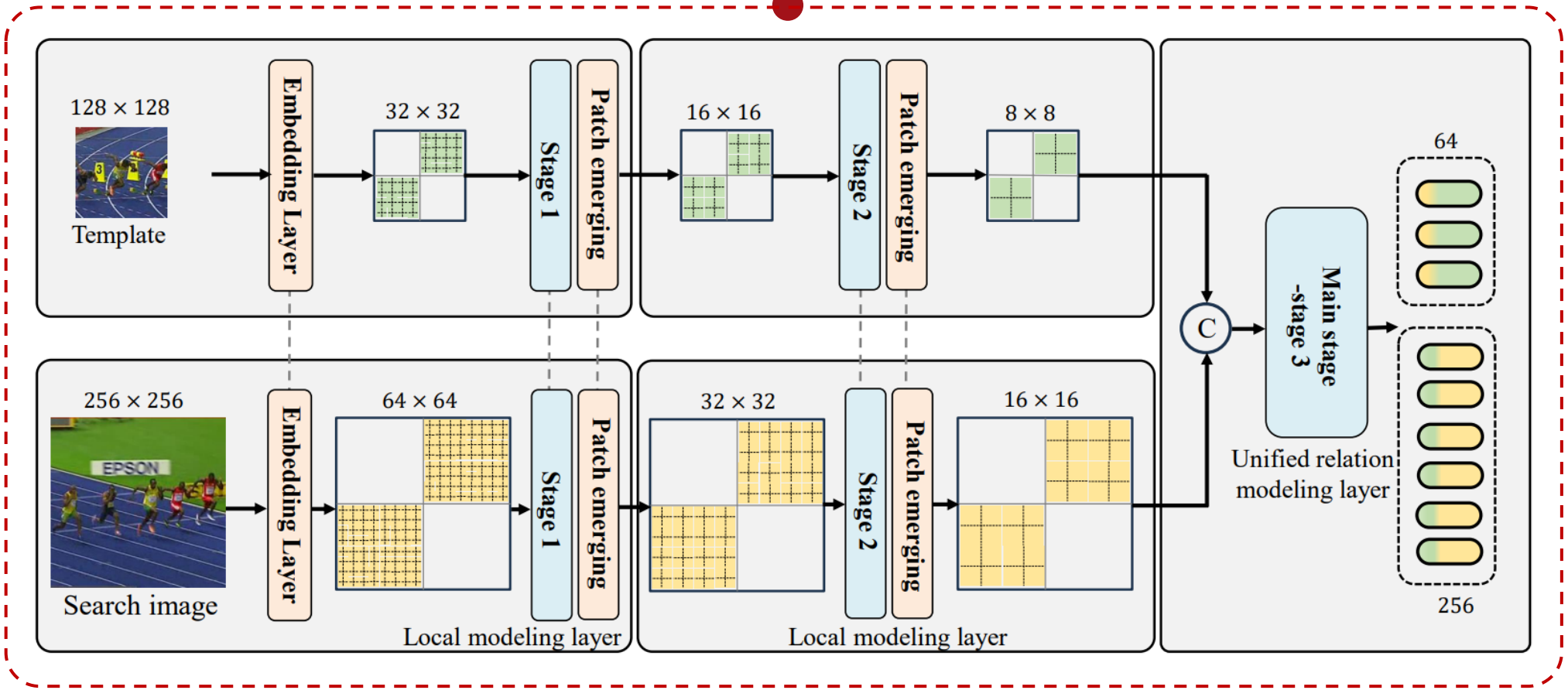
1. 提出了基于transformer的专用于视觉跟踪的特征提取网络，简化原先SiameRPN为代表的双分支框架，从而提出了一种全新的单分支跟踪框架。



Fei Xie, etc, Correlation-Aware Deep Tracking. CVPR, 2022

2.已有研究工作：工作（2） SuperSBT

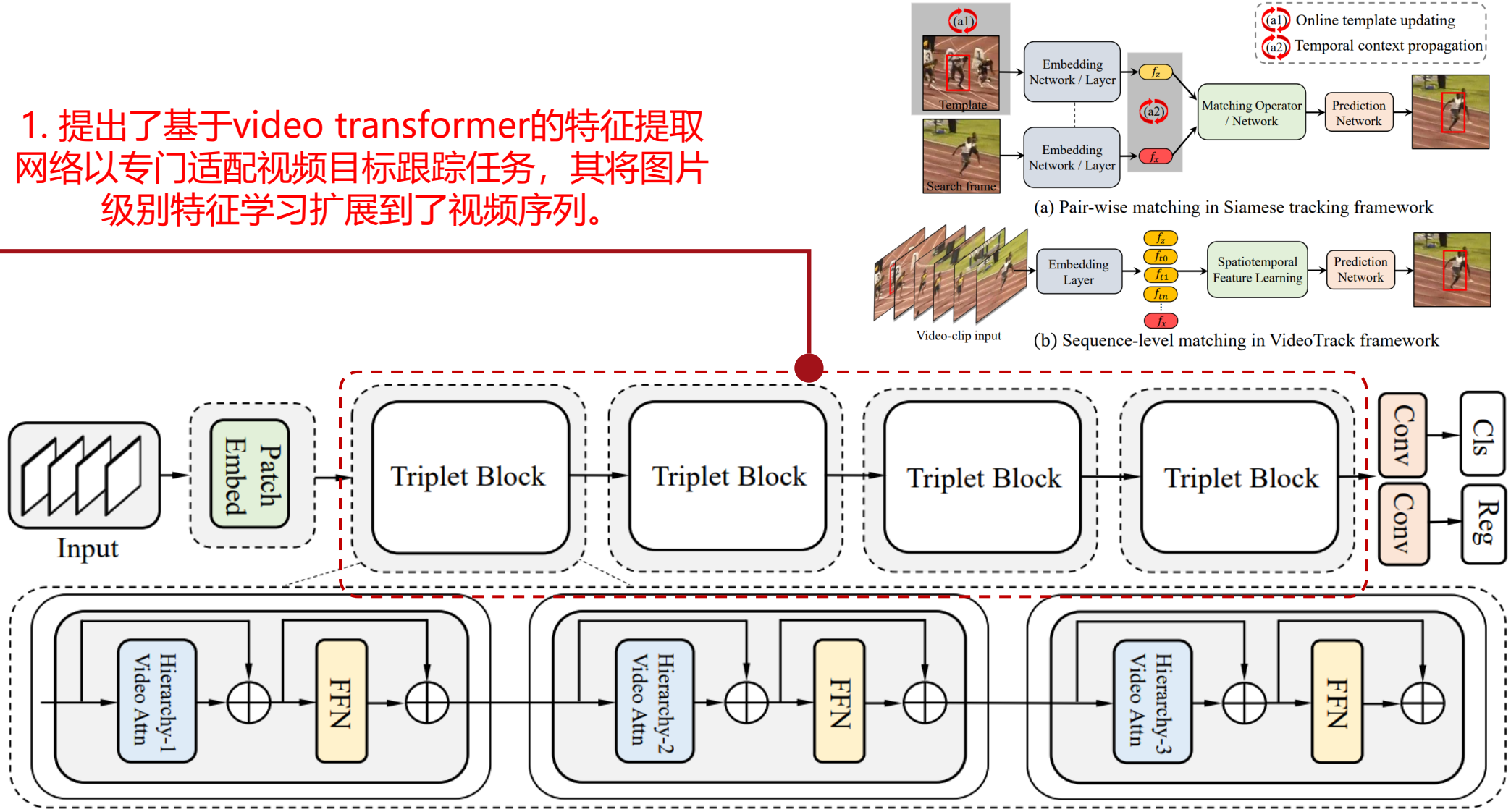
2. 对基于Transformer的特征提取网络针对跟踪任务进行了更加定制化的表征网络改进。



Fei Xie, etc, Correlation-Embedded Transformer Tracking: A Single-Branch Framework. T-PAMI 2024.

2.已有研究工作：工作（3）VideoTrack

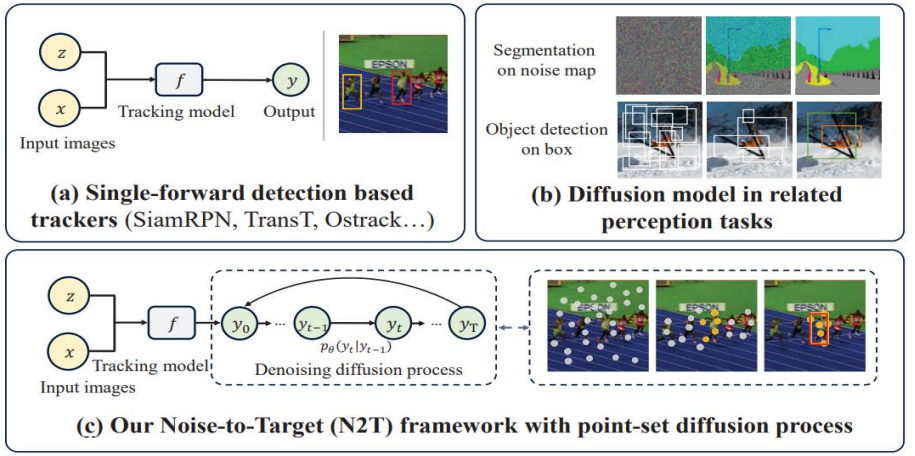
1. 提出了基于video transformer的特征提取网络以专门适配视频目标跟踪任务，其将图片级别特征学习扩展到了视频序列。

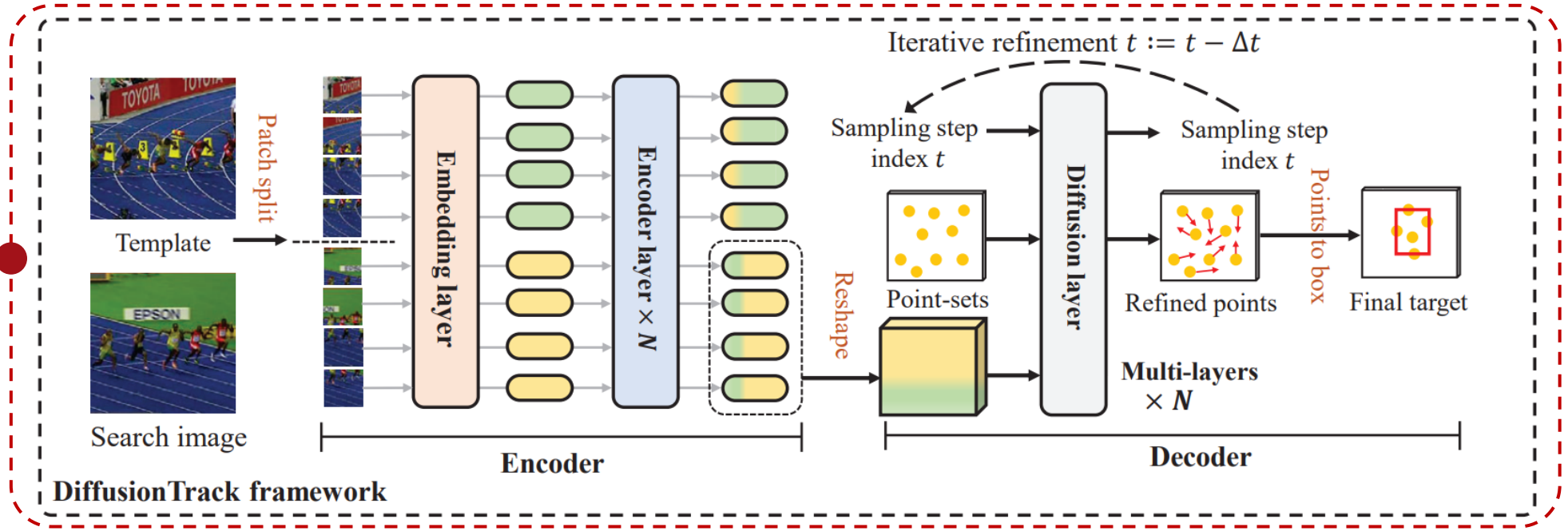


Fei Xie, etc, VideoTrack: Learning To Track Objects via Video Transformer. CVPR, 2023

2.已有研究工作：工作（4）DiffusionTrack

1. 采用基于扩散模型的生成方法对视觉目标跟踪进行更有效的表征学习。

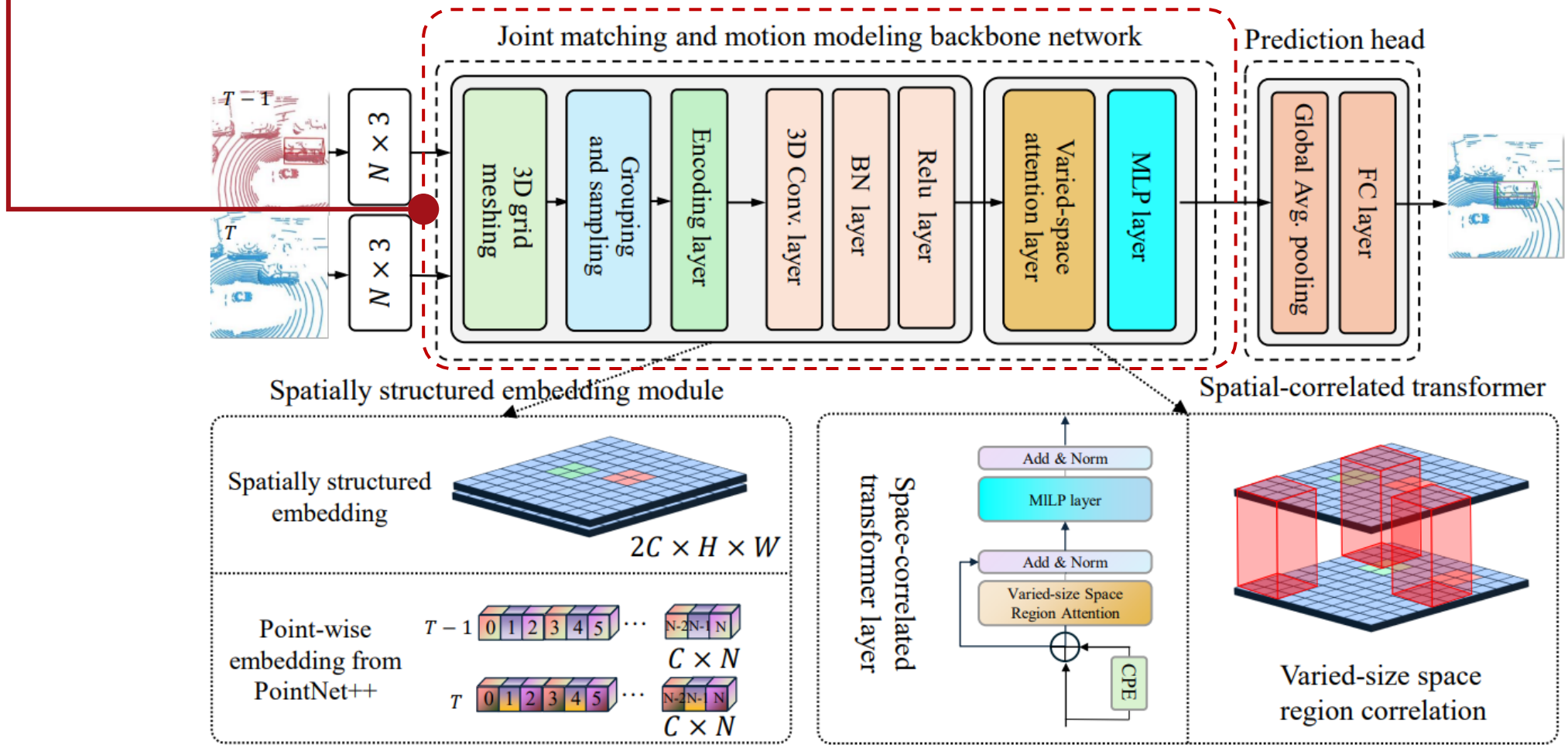




Fei Xie, etc, DiffusionTrack: Point Set Diffusion Model for Visual Object Tracking. CVPR, 2024

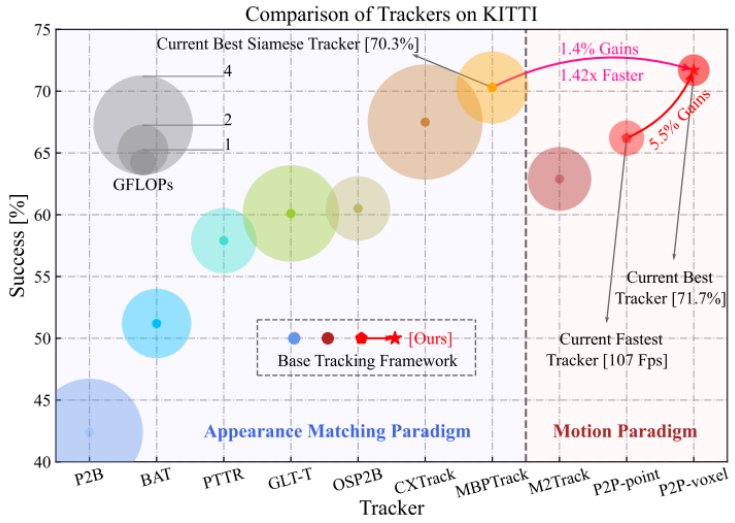
2.已有研究工作：工作（5）SCtrack

1. 采用基于可变区域的点云特征学习网络来对点云的形状和运动信息同时建模

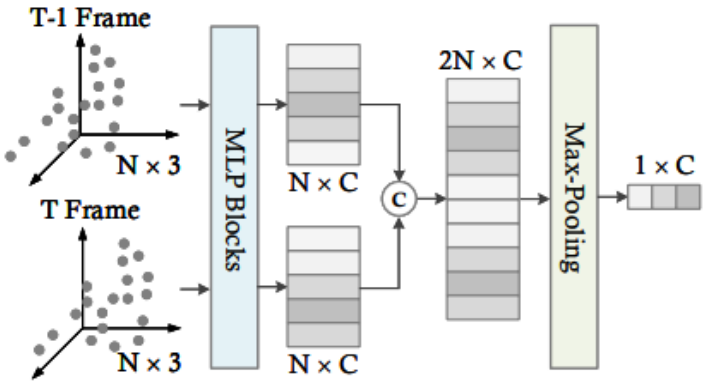


Fei Xie, etc, Space-Correlated Transformer: Jointly Explore the Matching and Motion Clues in 3D Single Object Tracking. ECCV, 2024.

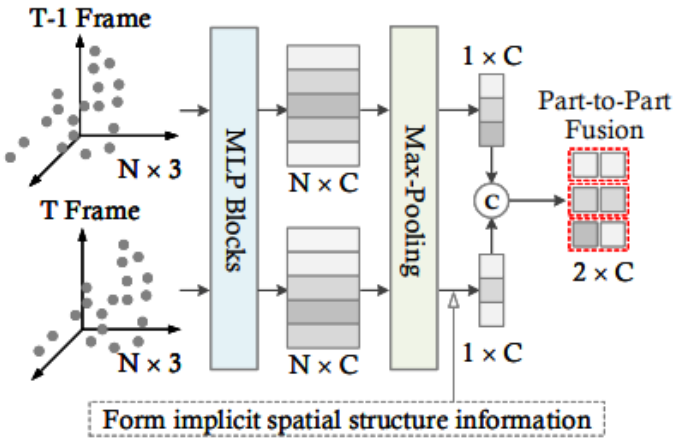
2.已有研究工作：工作（6）P2Ptrack



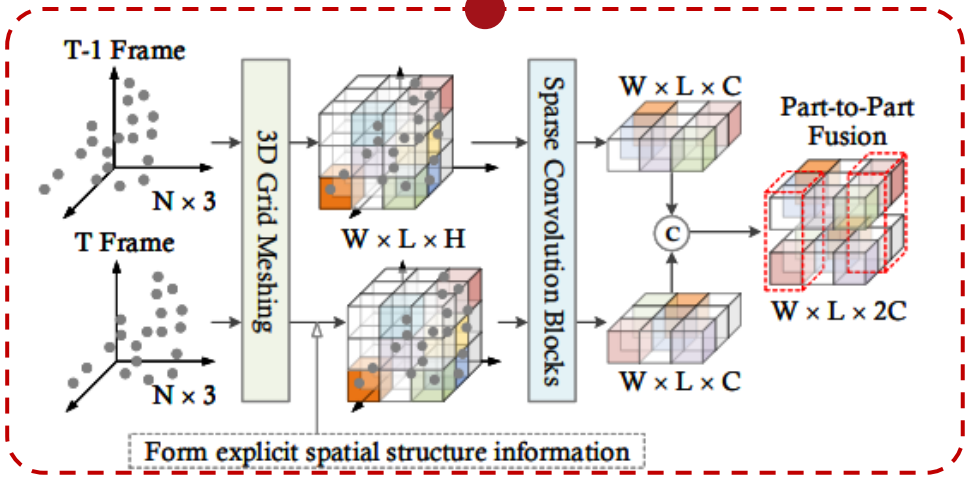
1. 采用voxel的表征对点云的形状和运动信息进行空间对齐，简化了模型流程



(b) M²Track-style motion modeling



(b) Implicit part-to-part motion modeling

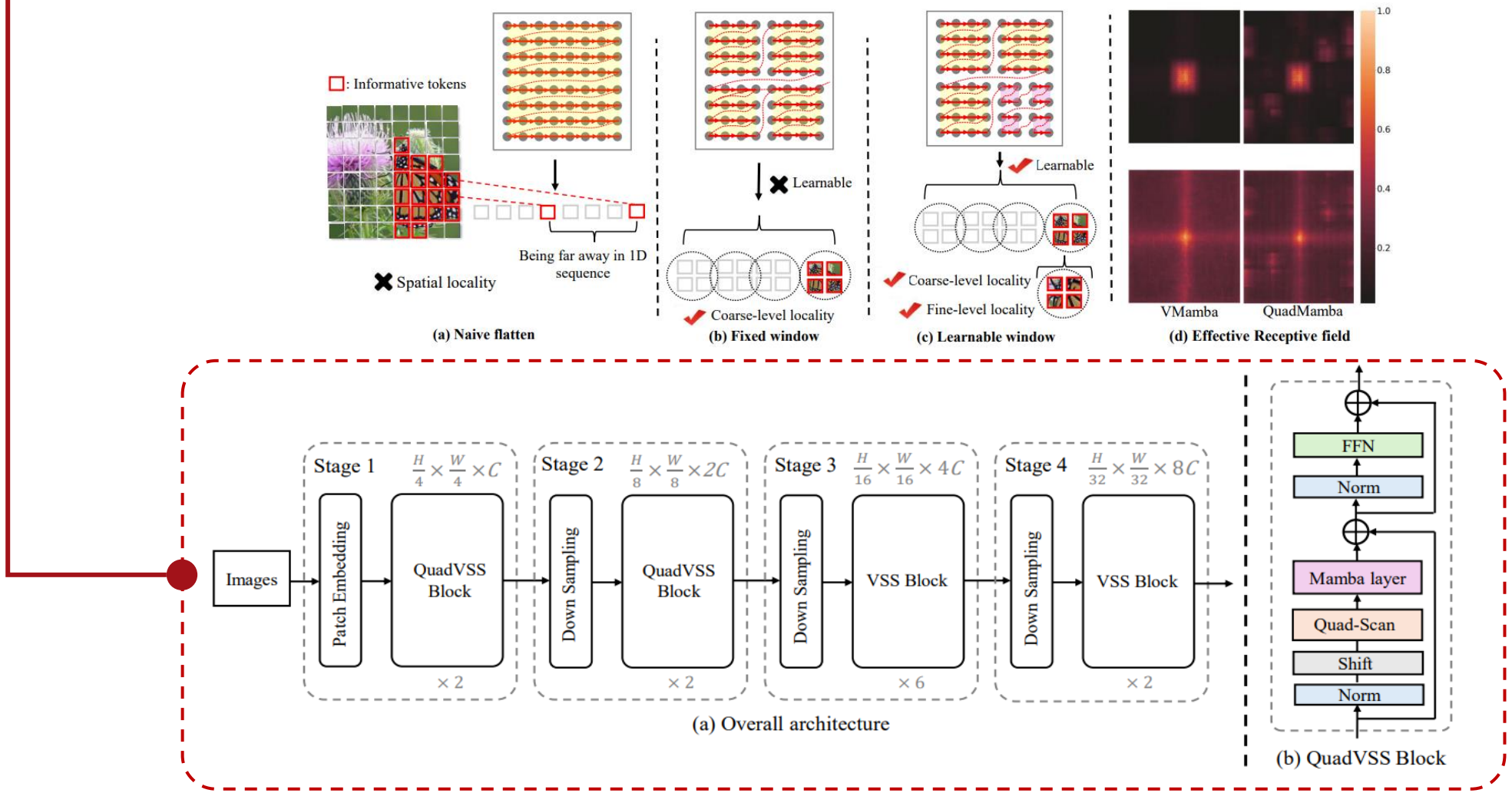


(c) Explicit part-to-part motion modeling

Jiahao Nie , Fei Xie , Sifan Zhou , Xueyi Zhou , Dong-Kyu Chae , Zhiwei He, P2P: Part-to-Part Motion Cues Guide a Strong Tracking Framework for LiDAR Point Clouds, IJCV, 2025

2.已有研究工作：工作（7）QuadMamba

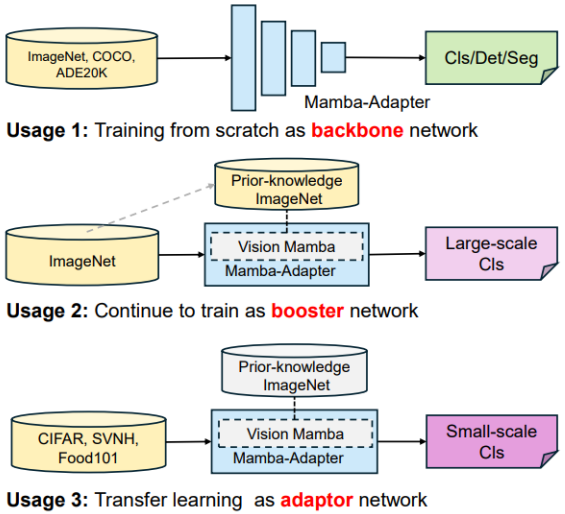
1. 基于四叉树的动态辨识的视觉Mamba图片特征提取网络。

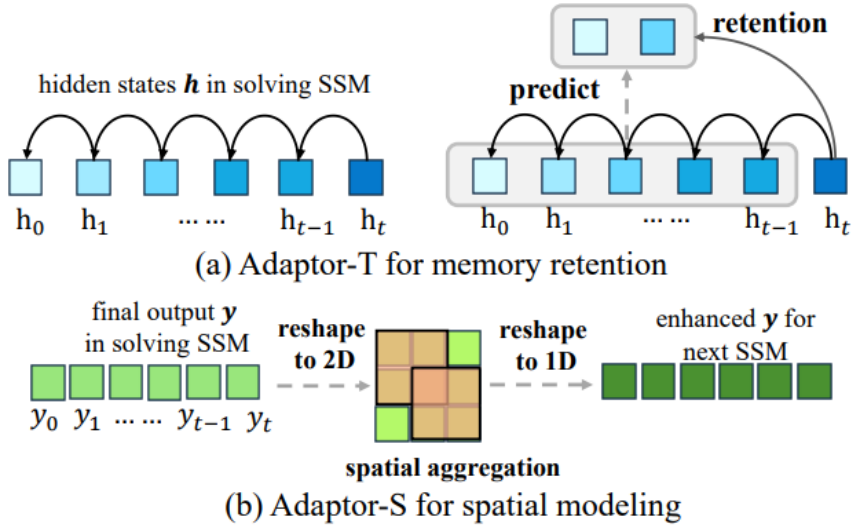


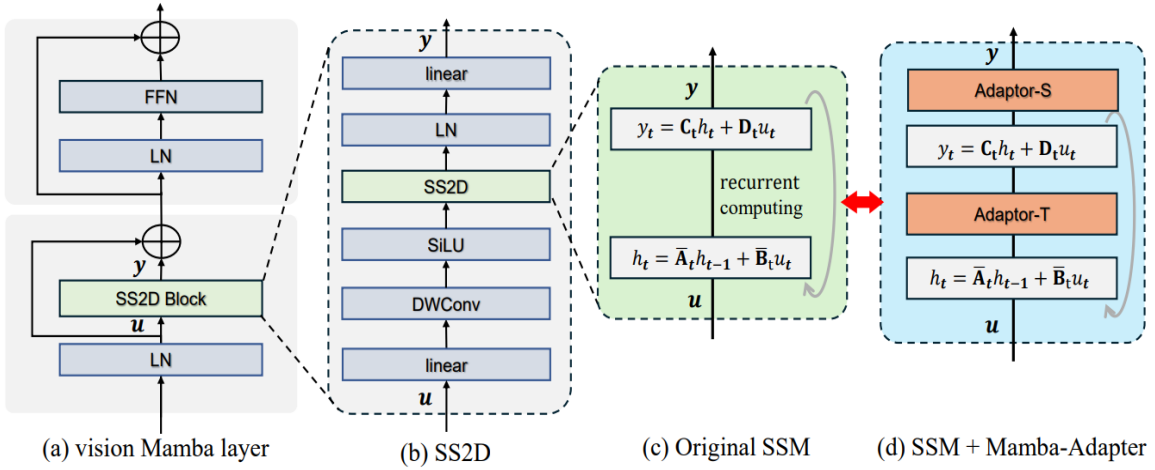
Fei Xie, etc,. QuadMamba: Learning Quadtree-based Selective Scan for Visual State Space Model. NeurIPS, 2024

2.已有研究工作：工作（8）Mamba-Adaptor

1. 改善Mamba模型中的空间结构先验不足和长距离遗忘问题。



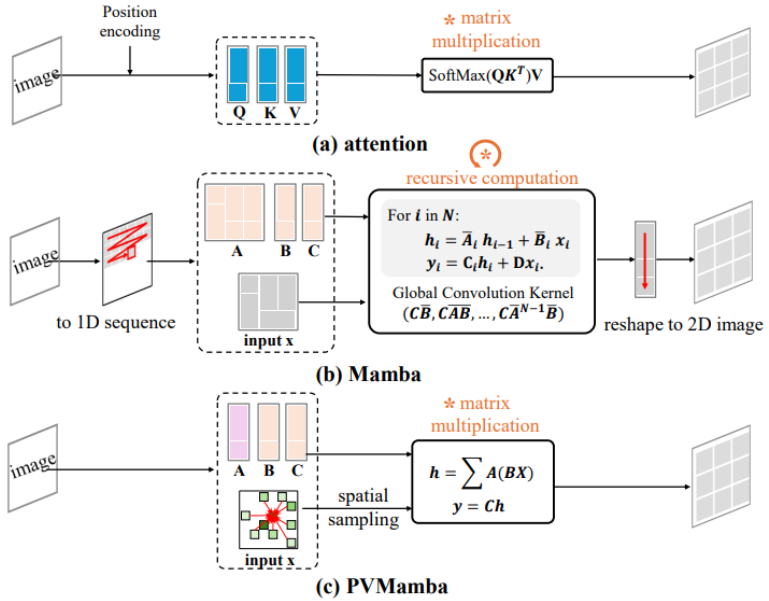


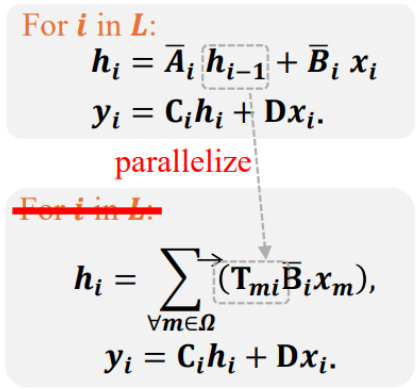


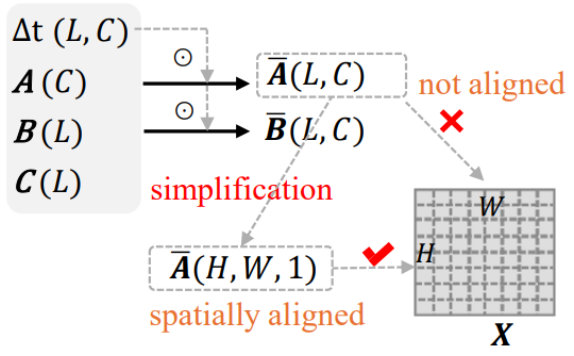
Fei Xie, etc,. Mamba-Adaptor: State Space Model Adaptor for Visual Recognition. CVPR, 2025

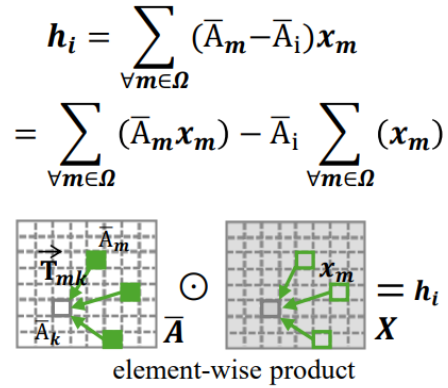
2.已有研究工作：工作（9）PVMamba

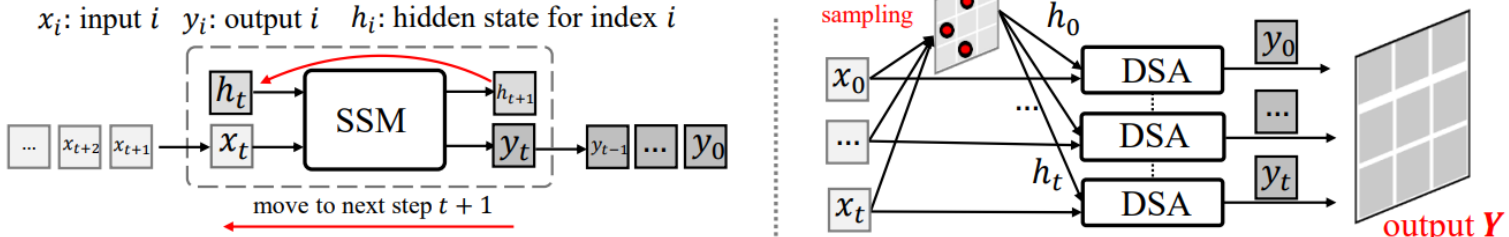
- 1. 克服Mamba的序列化限制，使用空间自适应算子适配二维图片数据。











Fei Xie, etc,. PVMamba: Parallelizing Vision Mamba via Dynamic State Aggregation. ICCV, 2025

1 个人简介

2 已有研究工作

3 未来研究计划

4 总结与讨论

3.未来研究计划：

- **研究方向一：多模态大模型(MLLM)**
  - 多模态大模型(MLLM)在视频理解等下游任务的应用和结合
  - 强化学习与VLA
- **研究方向二：基座模型结构**
  - 高效的线性注意力结构，如Mamba、Linear Attention、RNN
  - 高效网络结构在视觉方向的适配
- **研究方向三：生成式人工智能**
  - 基于Diffusion Transformer的模型层面的改进
  - 基于Flow Model的训练加速与改进

1 个人简介

2 已有研究工作

3 未来研究计划

4 总结与讨论

4.总结与讨论

- **工作的相关信息：**
  - 地理位置
  - 薪酬待遇
  - 工作时长强度
  - 发展空间
  - 计算资源
  - 其他待遇
- **工作的要求：**
  - 考核要求
  - 工作内容
  - 其他要求

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

谢　谢！