

MadEye: Boosting Live Video Analytics Accuracy with Adaptive Camera Configurations

Mike Wong and Murali Ramanujam, Princeton University; Guha Balakrishnan, Rice University; **Ravi Netravali**, Princeton University

NSDI 2024

汇报人：陈宇杰

作者团队背景

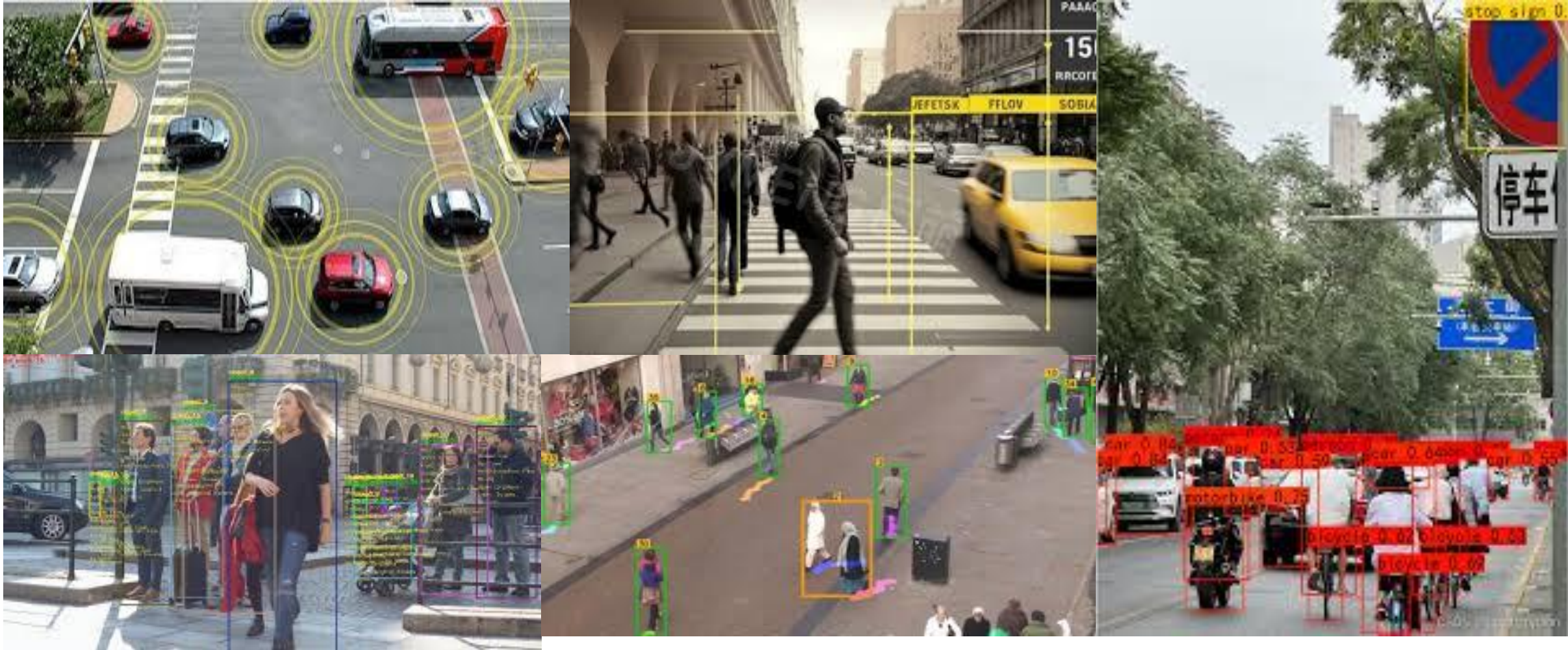
RESEARCH INTERESTS

- Computer systems and networks
- ML+systems
- distributed applications

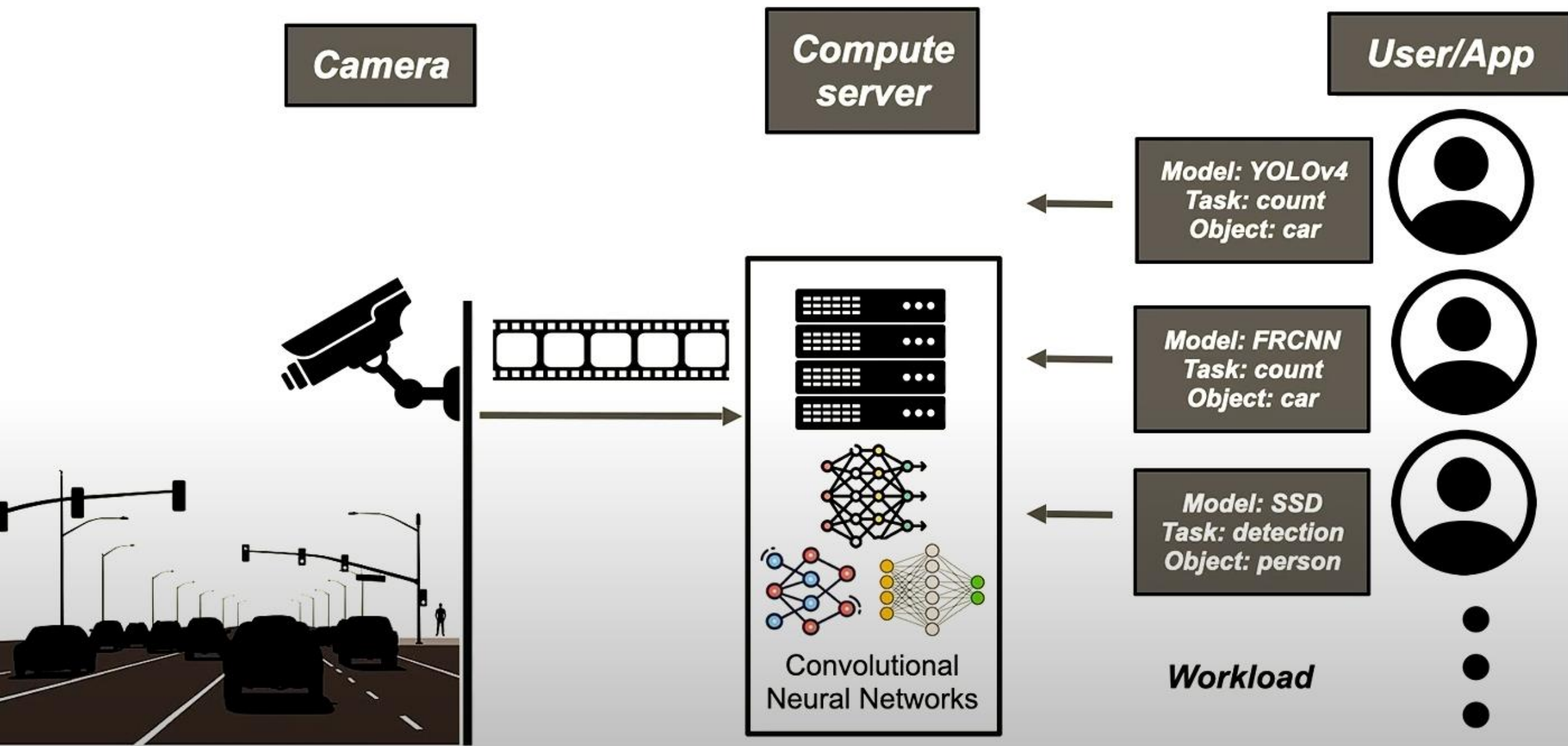


<https://www.cs.princeton.edu/~ravian/>

Video analytics background



Video analytics background



Related work

提高实时视频分析管道的资源效率和准确性

- 针对准确性感知的推理配置、编码或外观参数调整

[Accmpeg: Optimizing video encoding for accurate video analytics.\(MLsys 2022\)](#)

- 过滤掉冗余内容

[Spatula: Efficient cross-camera video analytics on large camera networks. \(SEC 2020\)](#)

- 使用成本更低的模型

[Ekya: Continuous Learning of Video Analytics Models on Edge Compute Servers.\(NSDI 22\)](#)

- 改进作业调度

[Gemel: Model Merging for Memory-Efficient, Real-Time Video Analytics at the Edge.\(NSDI 23\)](#)

Related work

提高实时视频分析管道的资源效率和准确性

- 针对准确性感知的推理配置、编码或外观参数调整

[Accmpeg: Optimizing video encoding for accurate video analytics.\(MLsys 2022\)](#)

- 过滤掉冗余内容

[Spatula: Efficient cross-camera video analytics on large camera networks. \(SEC 2020\)](#)

- 使用成本更低的模型

[Ekya: Continuous Learning of Video Analytics Models on Edge Compute Servers.\(NSDI 22\)](#)

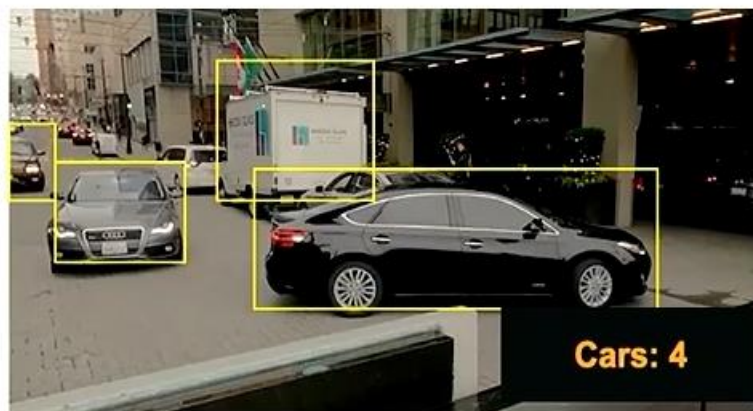
- 改进作业调度

[Gemel: Model Merging for Memory-Efficient, Real-Time Video Analytics at the Edge.\(NSDI 23\)](#)

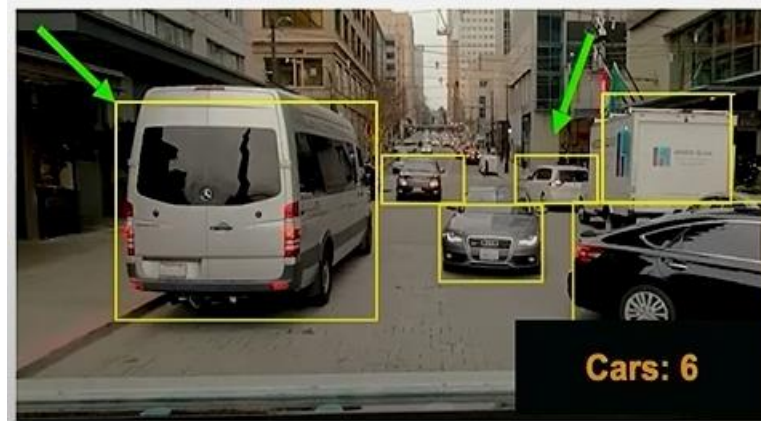
本质上：专注于优化固定、预设的摄像头部署

Motivation

平移

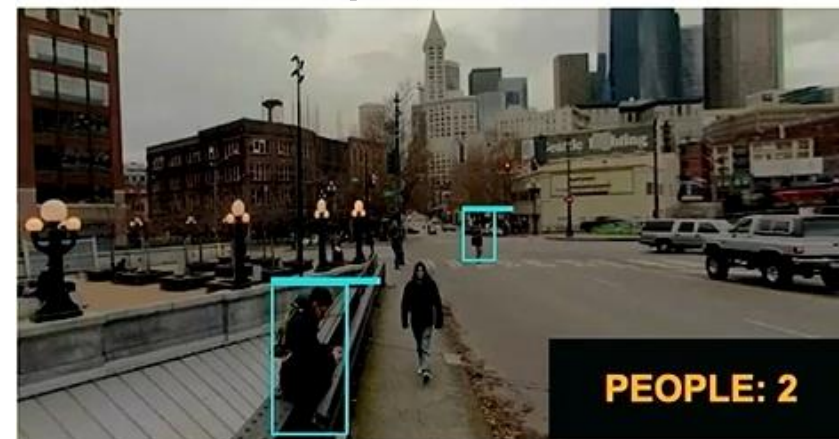


↓ 30°

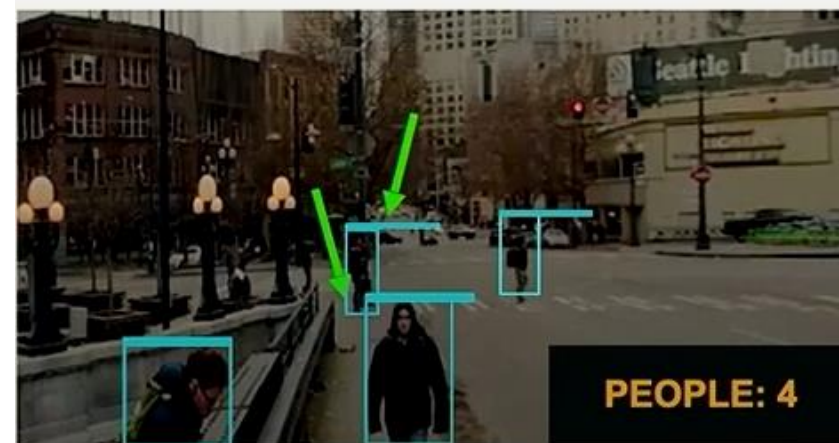


Cars(SSD)

缩放



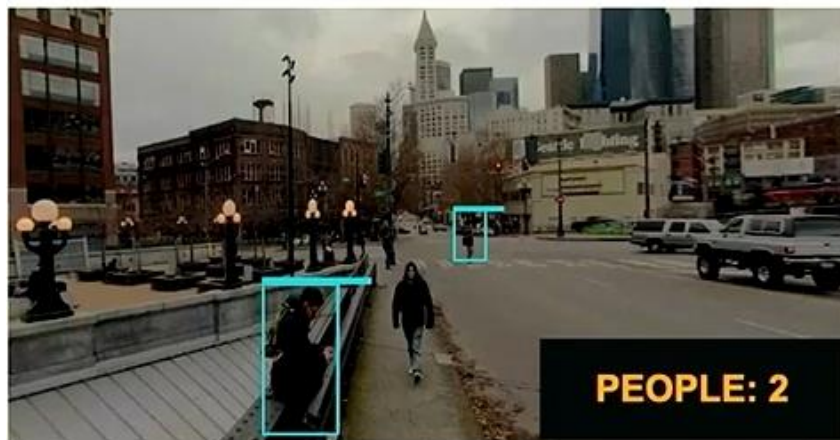
↓ 2X



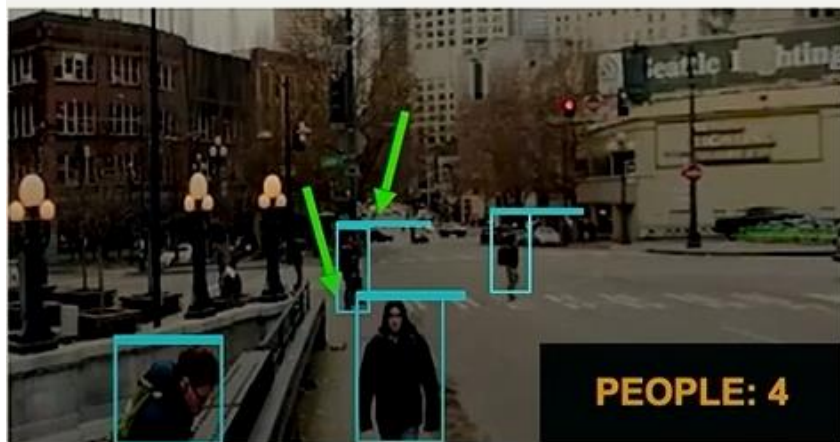
People (SSD)

Motivation

缩放



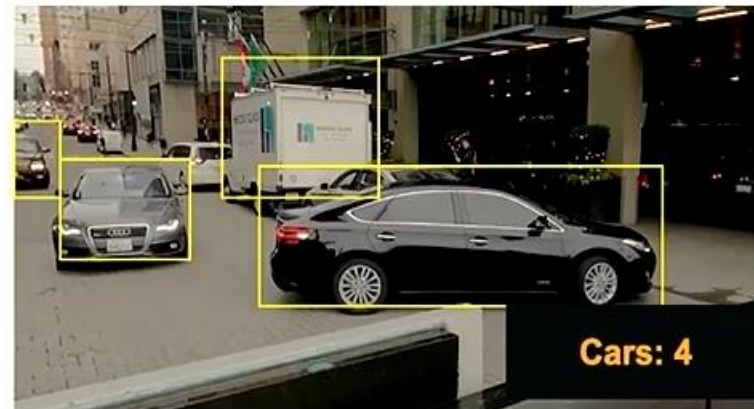
↓ 2X



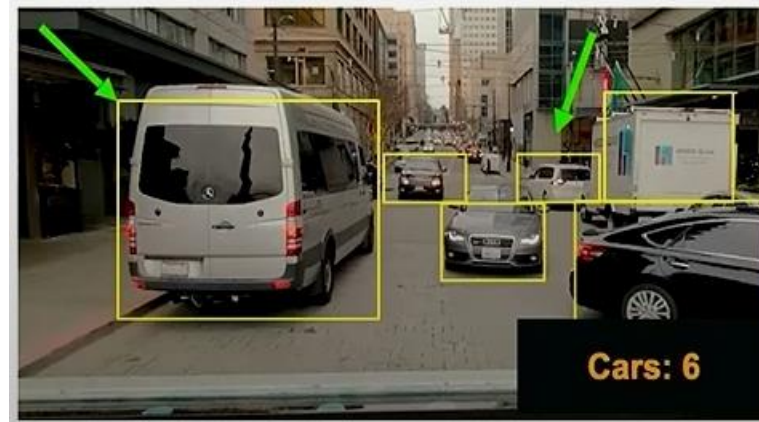
People (SSD)

摄像头的旋转和缩放对查询产生很大影响。

平移



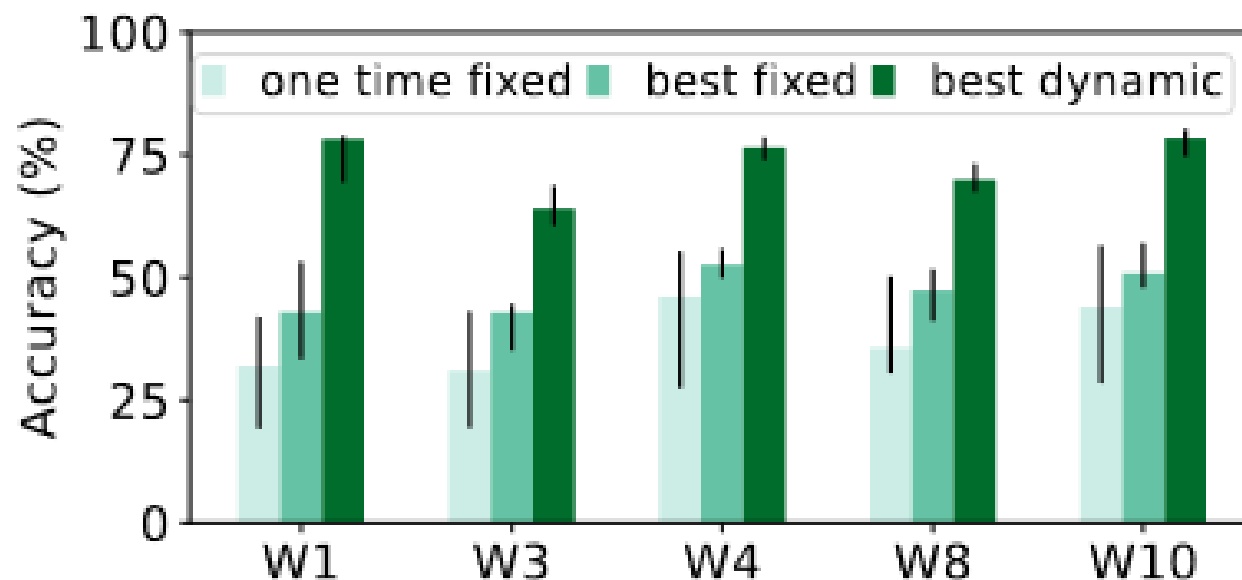
↓ 30°



Cars(SSD)

Motivation

Opportunities with Tuning Camera Orientations



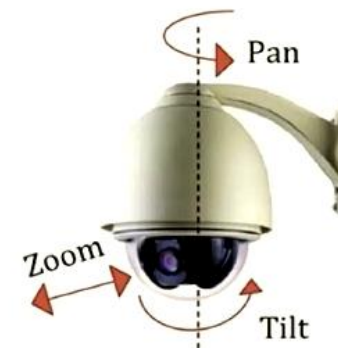
one time fixed: 在 $t = 0$ 时选择最佳方向并在整个视频中保持。

best fixed: 使用预言机知识来选择最大化平均工作负载准确度的最佳单一方向。

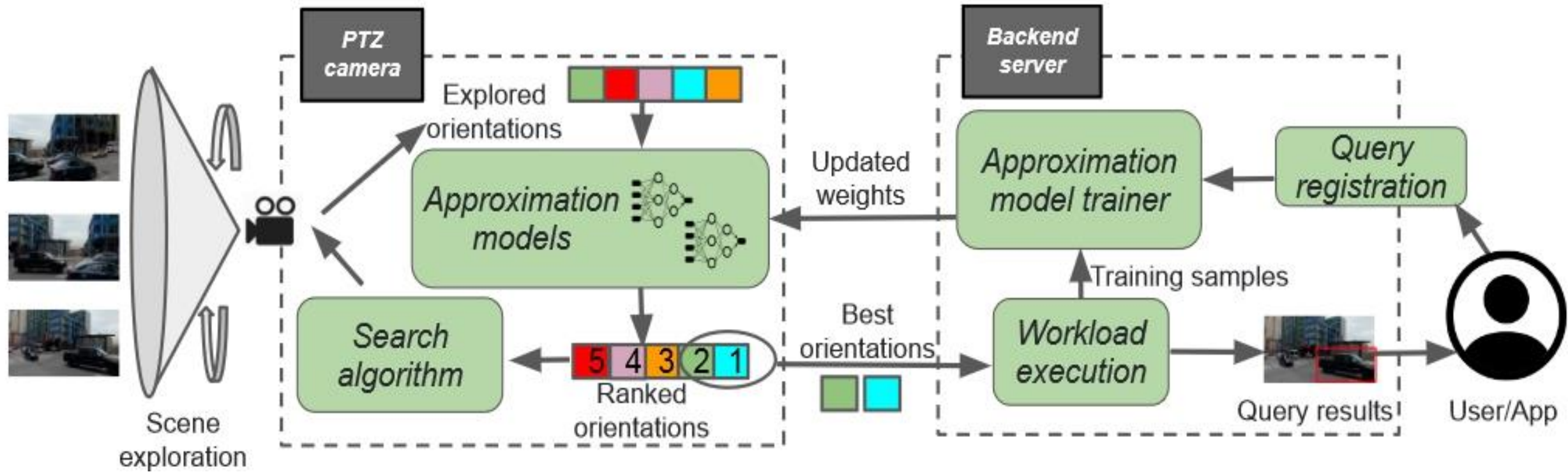
best dynamic: 选择视频中每帧的最佳方向。

PTZ cameras

PTZ相机支持平移（Pan）、俯仰（Tilt）和变焦（Zoom）
可以以每秒高达 600° 的速度旋转

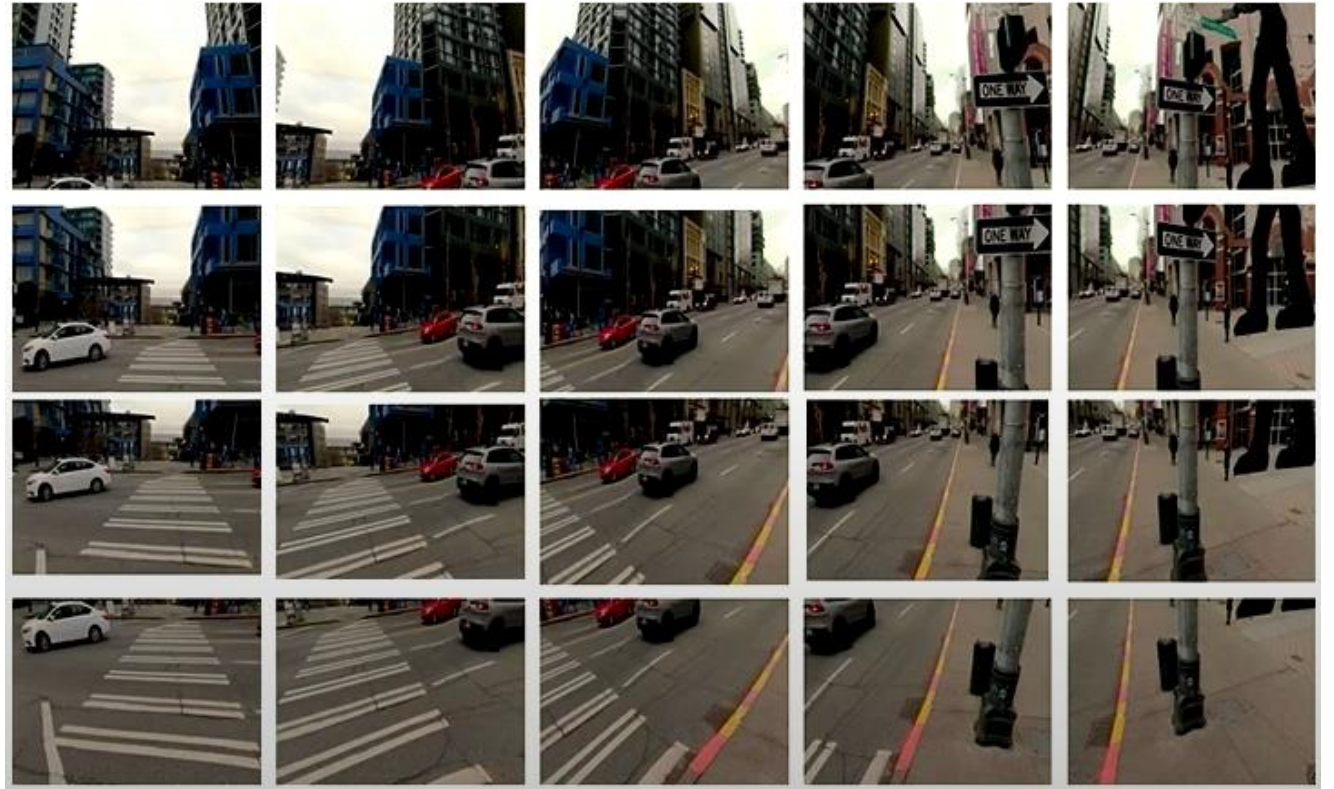
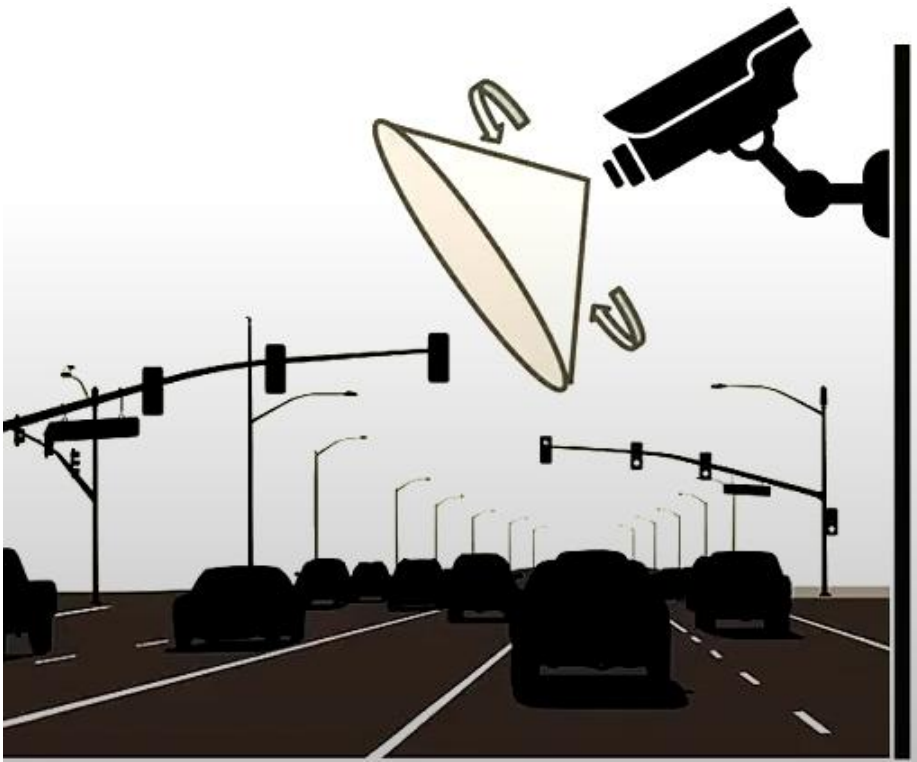


Design



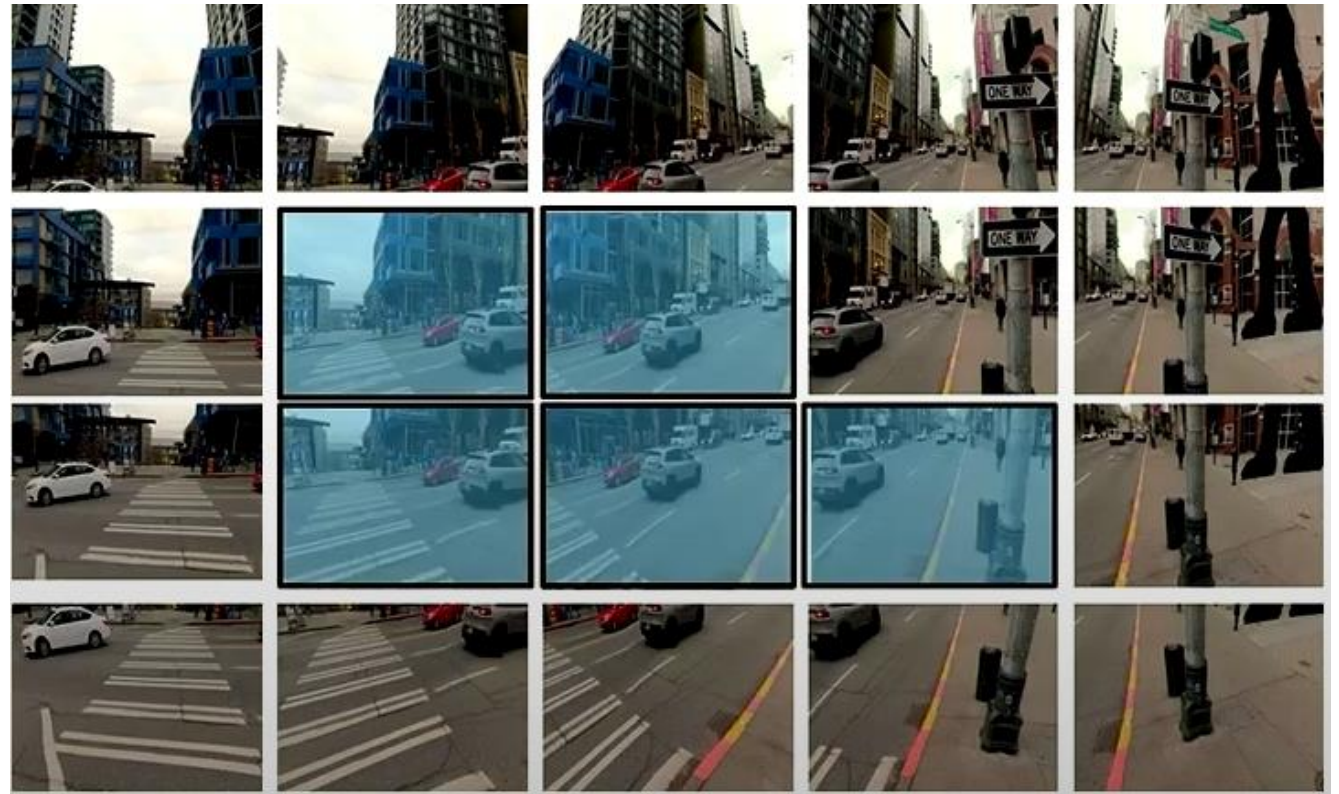
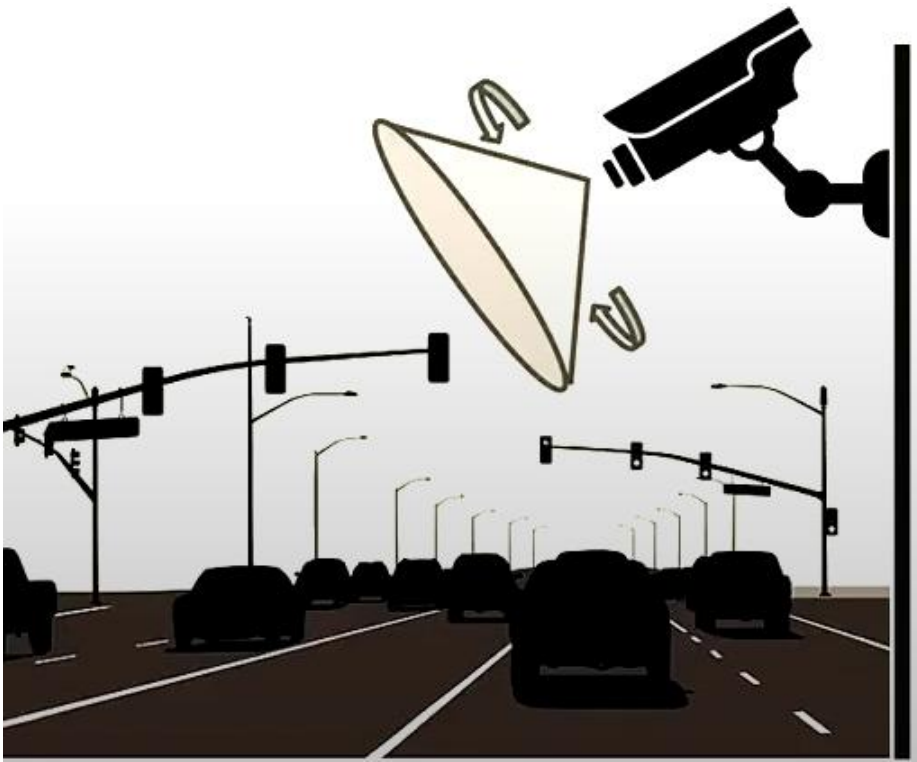
Design

Exploring and Ranking Orientations



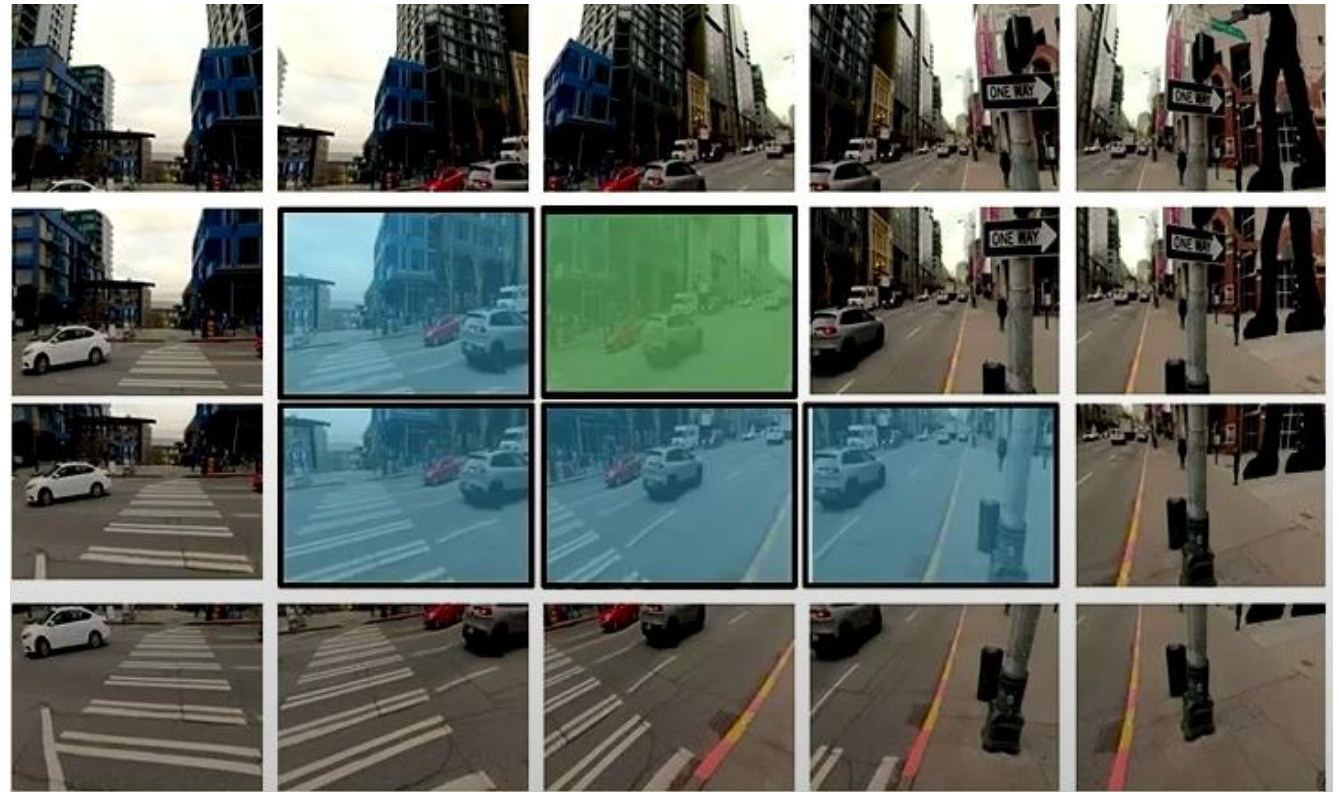
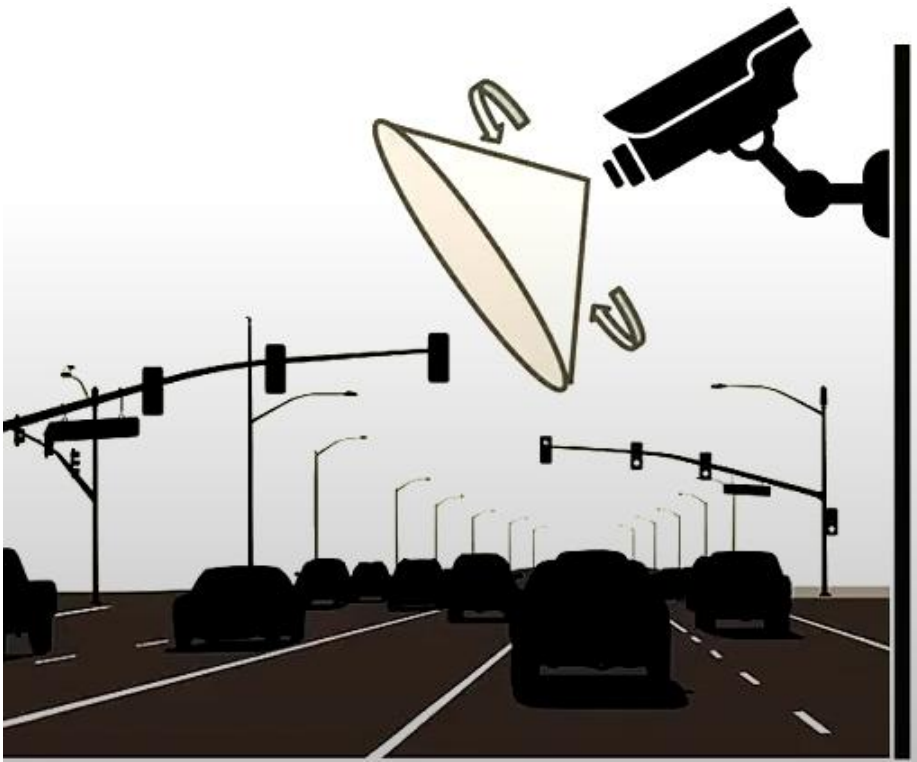
Design

Exploring and Ranking Orientations



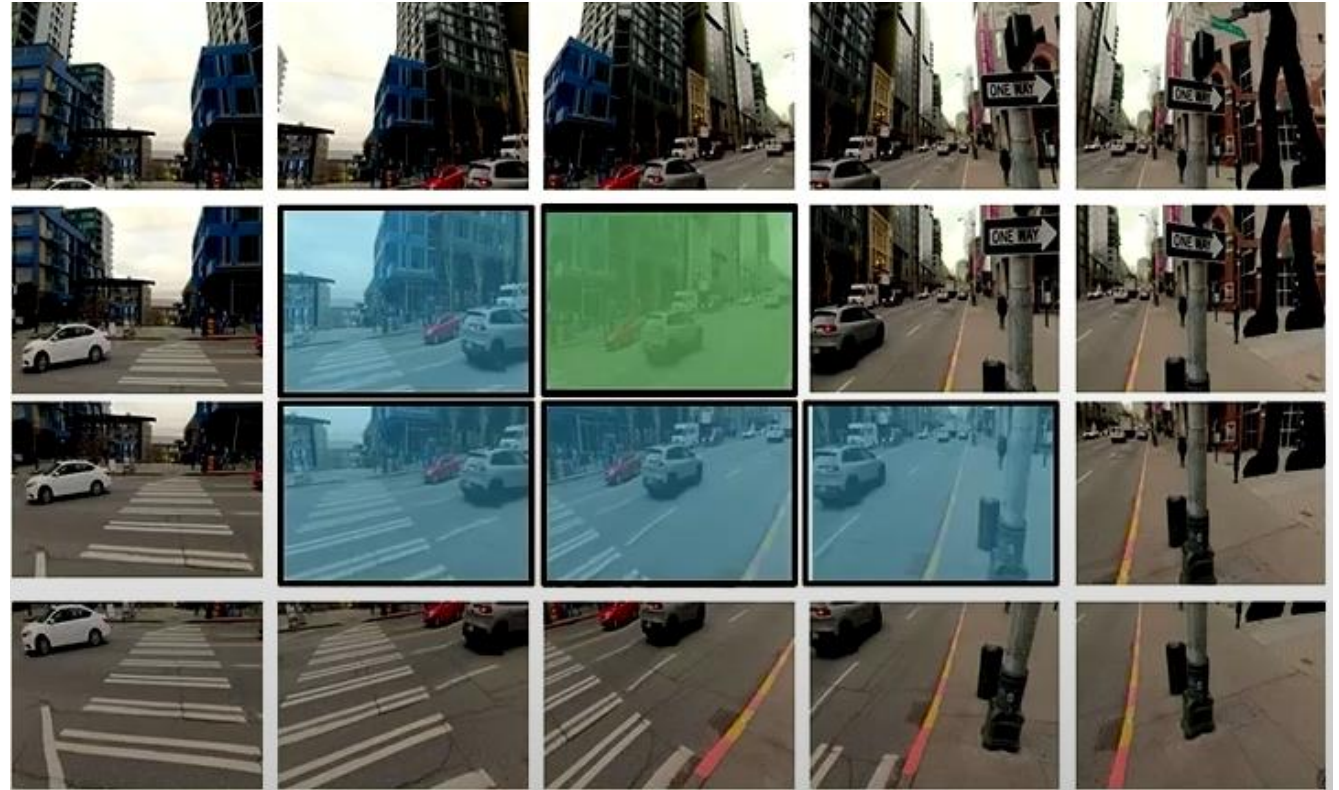
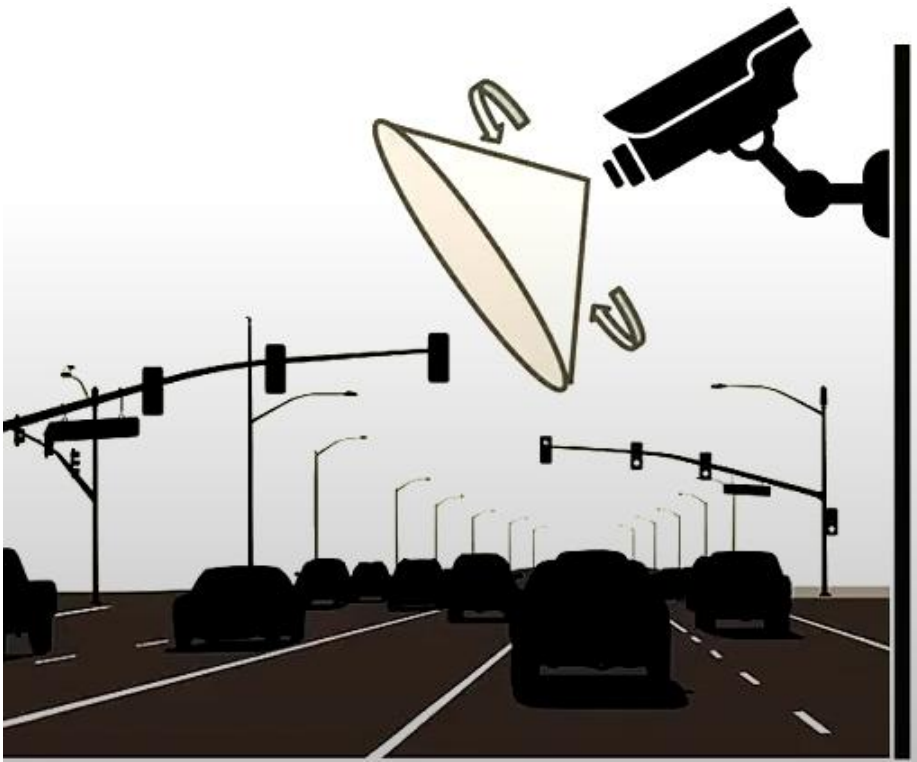
Design

Exploring and Ranking Orientations

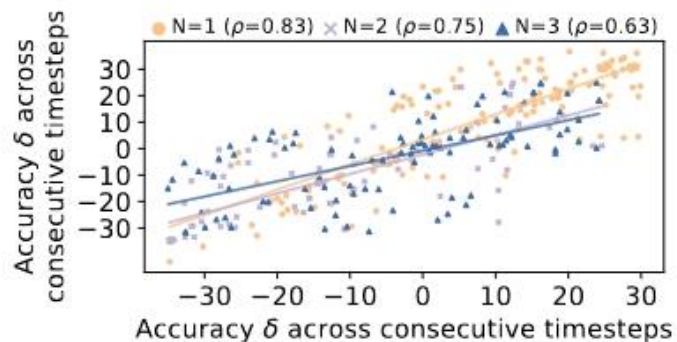
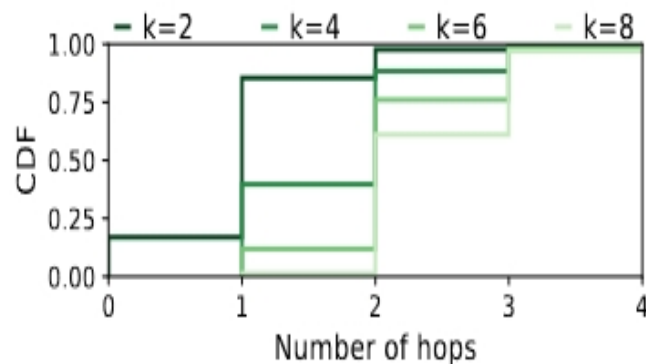
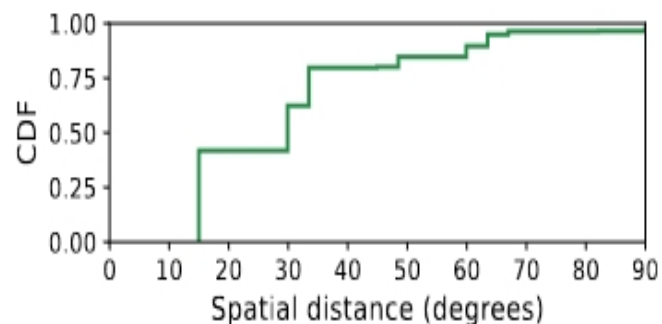


Design

Exploring and Ranking Orientations



Exploring and Ranking Orientations

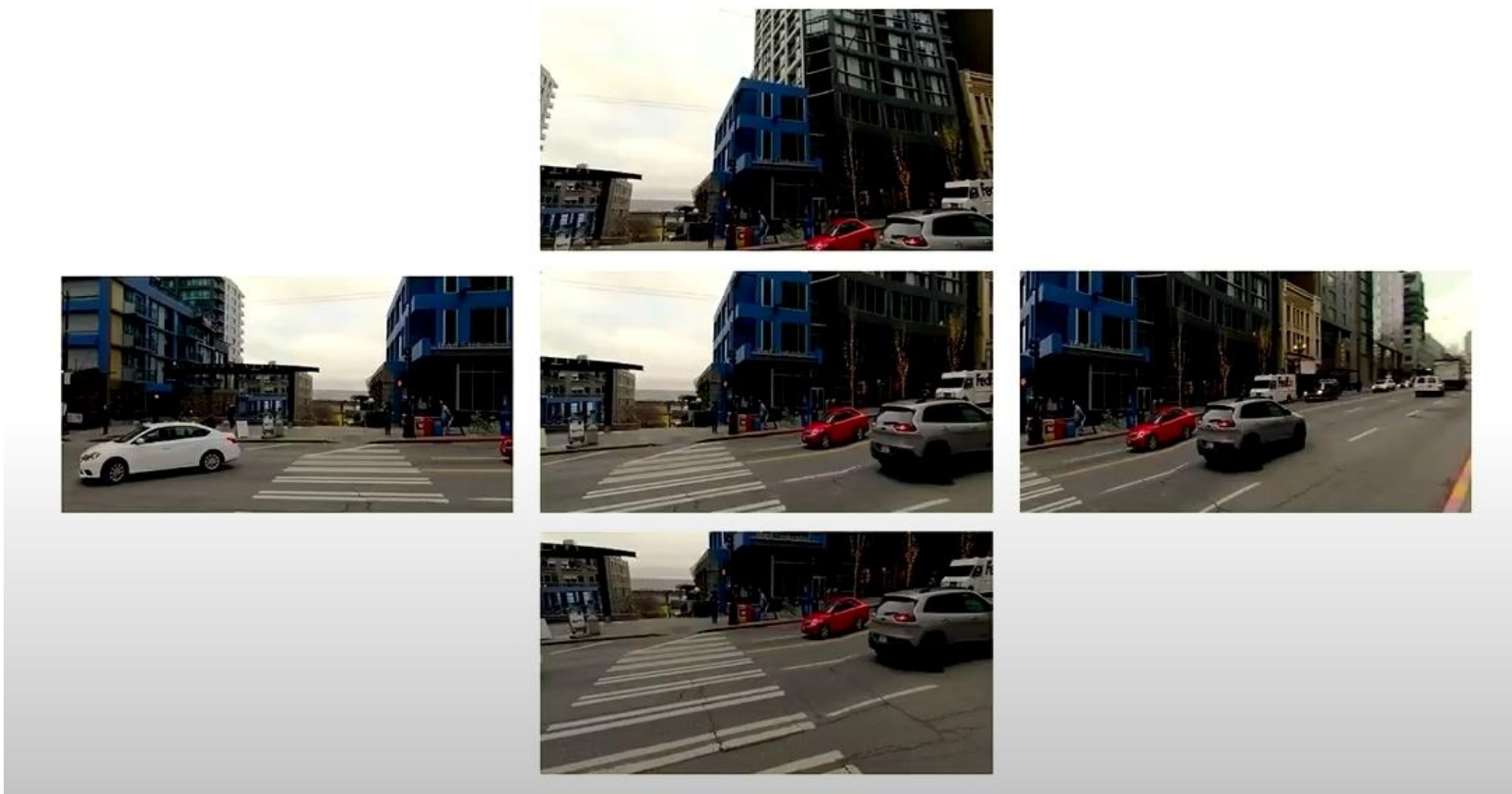


- 最佳方位的快速变化在空间维度上较慢
- 任何时刻的最佳表现方位在空间上聚集
- 邻近方位的准确度通常会同步变化

Design

Search algorithm

考虑每个时间步长中连续方向，逐步替换表现较差的方向，同时替换为我们可以可靠预测下一时间步长的邻近方向。



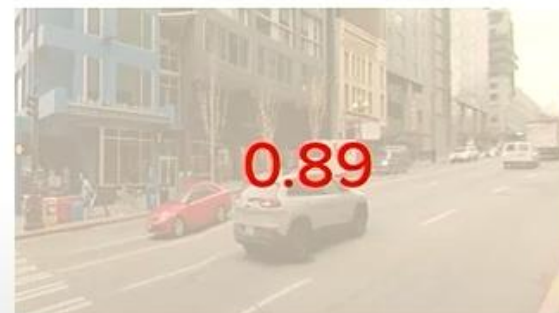
Design

Search algorithm

MadEye 使用一个值标记上一个时间步中的每个方向，该值指示在下一个时间步中取得成果的可能性。



计算指数加权移动平均:
(1)预测精度值
(2)精度值之间的变化量



**Scores to predict
how an orientation
will perform**

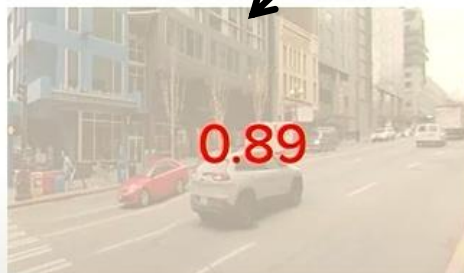
Design

Search algorithm

Worst orientation



Best orientation



- 最优与最差比率超过阈值
- 最优方向的邻居尚未处于形状中
- 删除最差方向不会破坏连续性

**Scores to predict
how an orientation
will perform**

Design

Search algorithm



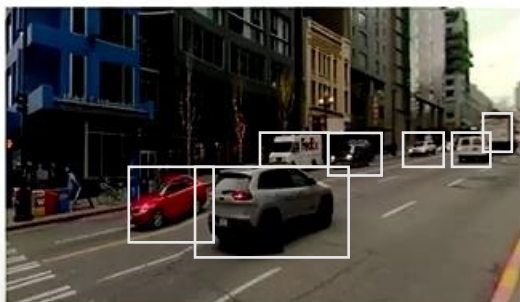
Neighbors of
best
orientation



Best orientations
are typically
clustered

Design

Search algorithm



Neighbors of
best
orientation



Best orientations
are typically
clustered

对于每个候选邻居，计算两个值的比率：

(1) 到最佳方向中心的距离。

(2) 到最佳方向中所有边界框质心的距离。

Design

Search algorithm



处理图片缩放

- 初始变焦因子选择：

当一个新的方向被添加到形状中时，MadEye从最低的变焦因子开始，获得该方位的整体视图。

- 计算边界框和质心距离：

MadEye使用近似模型生成的边界框，计算每个边界框与所有边界框质心之间的平均距离。

- 选择变焦因子：

将这些距离值与每个变焦因子覆盖的区域进行比较，选择一个合适的变焦因子。

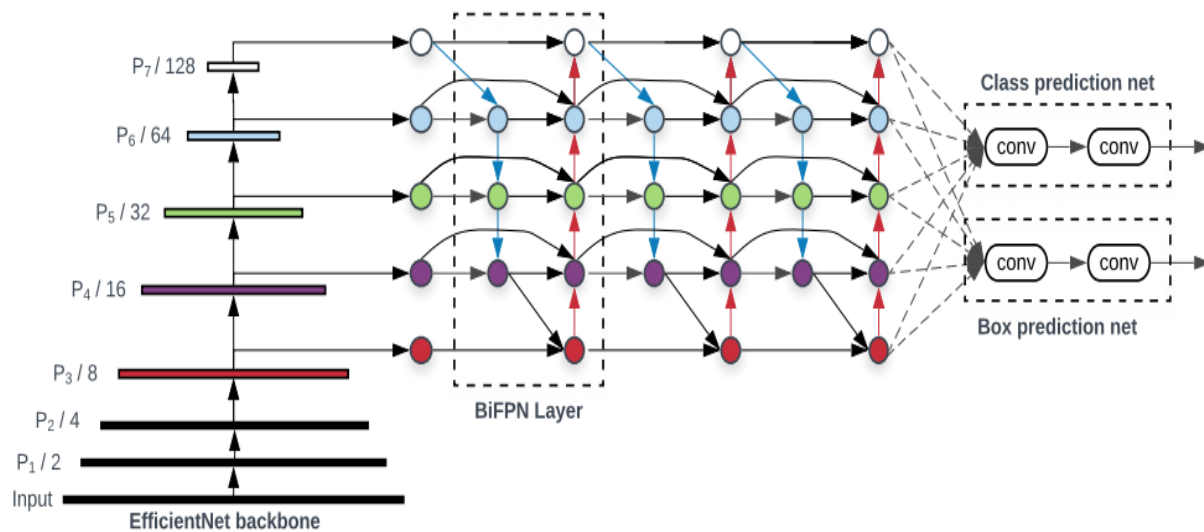
- 自动缩小视野：

MadEye会在3秒后自动缩小视野，从而避免错过新进入该方位的对象。

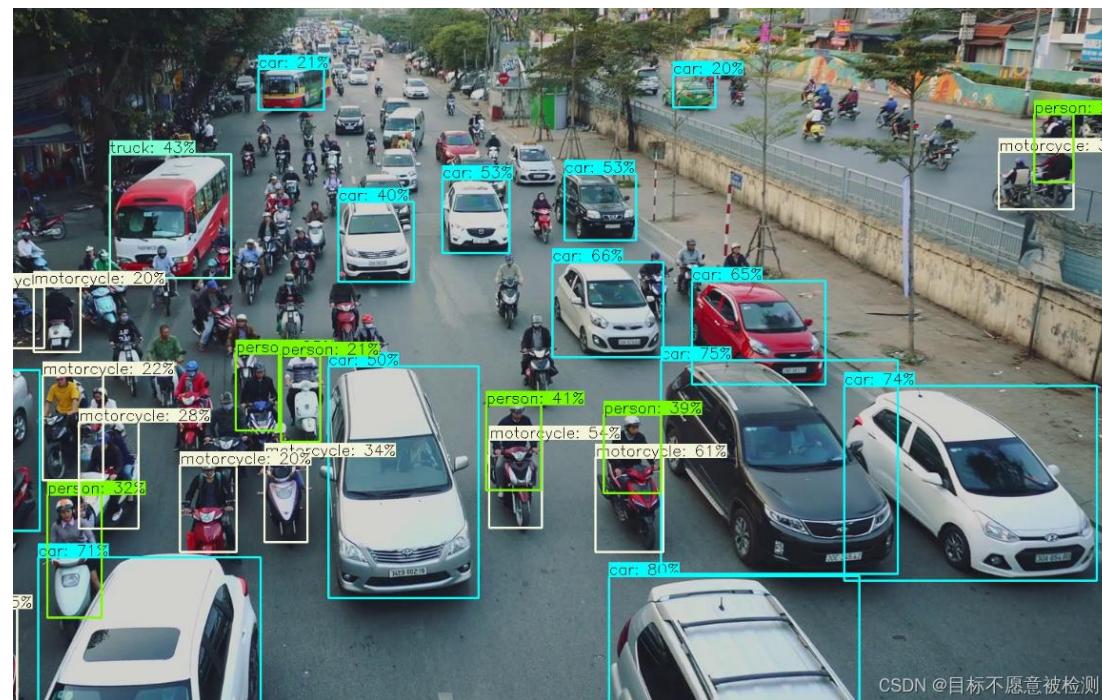
Design

Approximation models

- PTZ摄像机计算资源限制
- 对缩放和旋转的不同工作负载具有不同敏感性

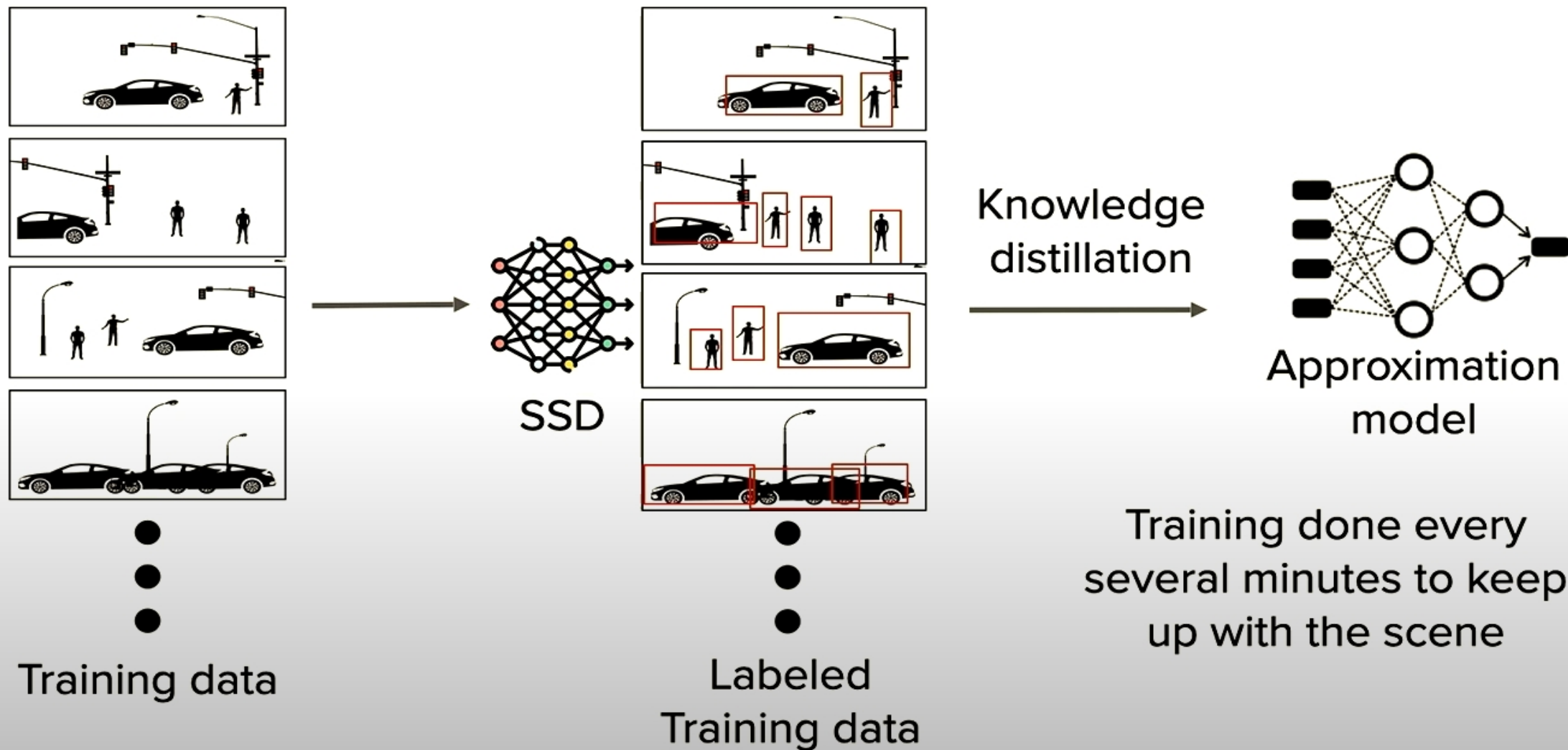


EfficientDet 模型结构



Design

Training approximation models



Continually Training Approximation Models

MadEye系统通过持续学习（Continual Learning）来应对场景动态变化导致的模型准确度下降问题

初始训练

数据集准备：目标场景1000张历史图像

预训练模型：在Pascal VOC数据集上预训练的EfficientDet模型

微调：仅微调用于边界框和类别预测的权重，初始持续40个epoch

持续训练

120秒/次

数据不平衡问题

数据集：最近120个最优方向选择的样本（1个样本/秒）和200个历史图像（平均值）

历史图像选择方法：

邻近方向+远距离方向(样本数量指数递减)

估计工作负载的准确性

使用approximation models完成计数任务（Counting）

估计方法：计算每个方向的对象数量，将其与所有探测方向中数量最大的方向进行比较。

计算每个方向对象数量与最大对象数量的比值。



数量	1
准确率	1/6

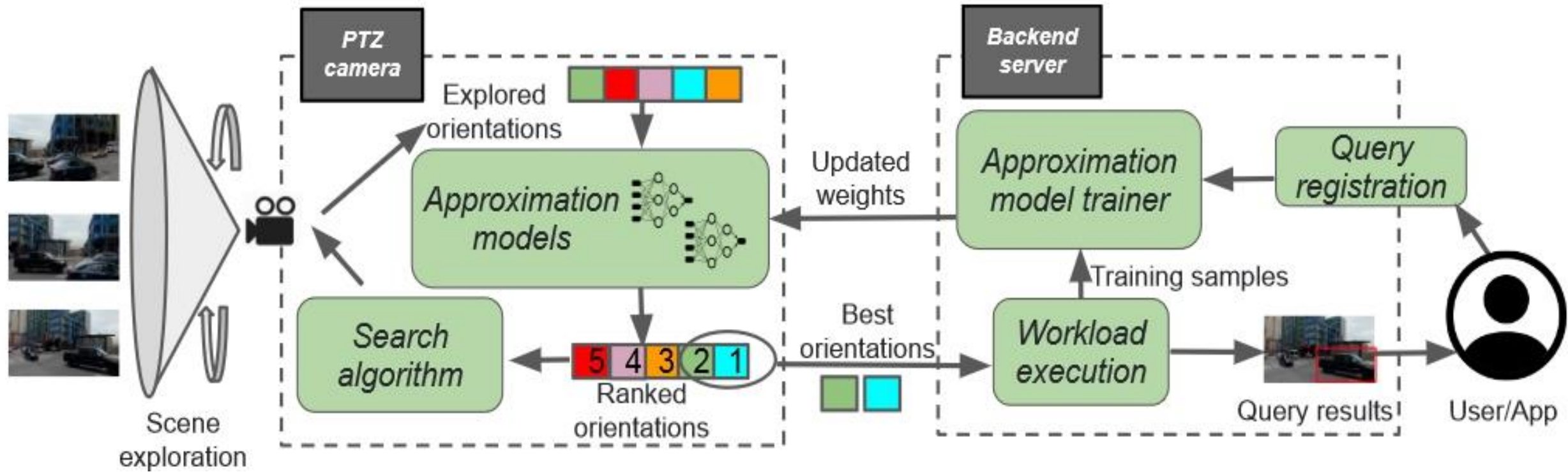


数量	3
准确率	3/6



数量	6
准确率	6/6

Design



Evaluation

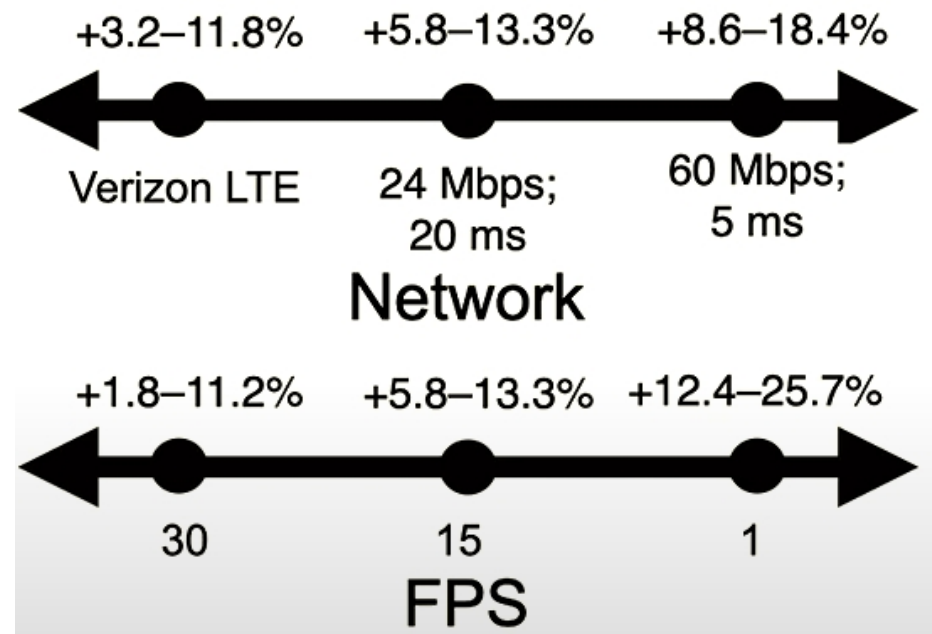
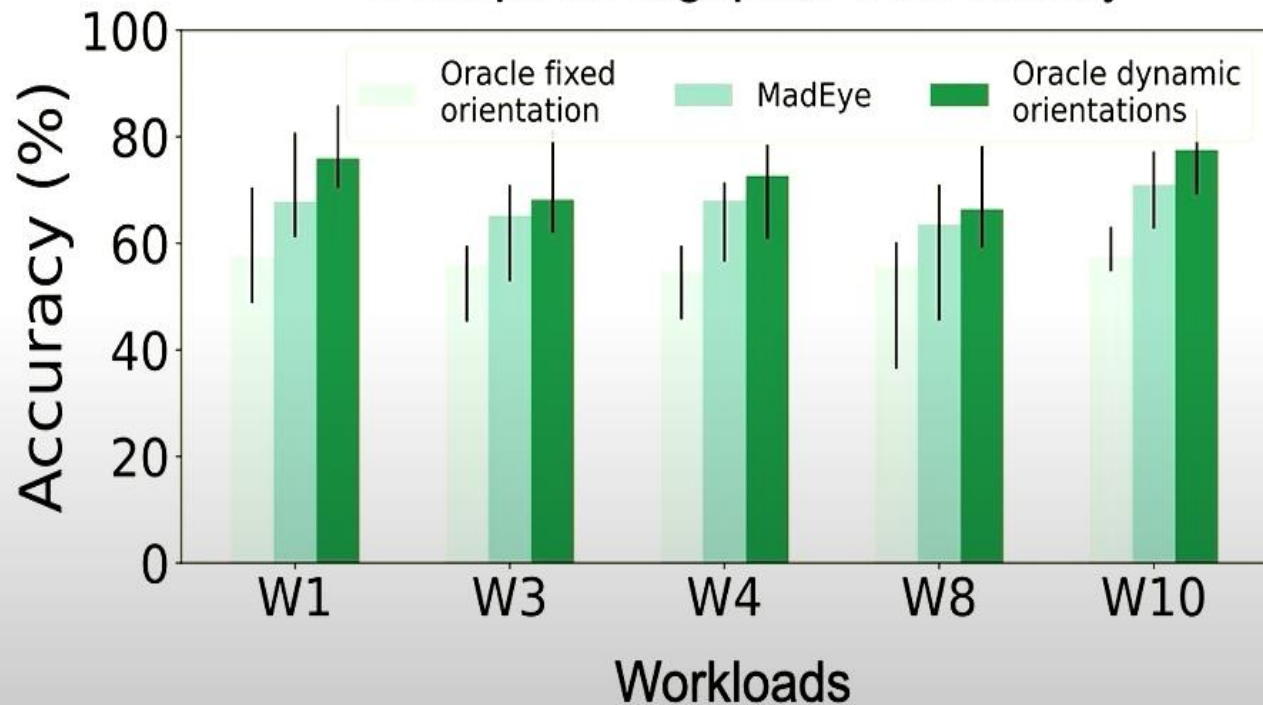
- 边缘级计算设备Jetson Nano:
 - 128核Maxwell GPU
 - 四核ARM CPU, 时钟速度为1.43 GHz
 - 4 GB内存
- 服务器:
 - NVIDIA RTX 2080 Ti GPU (8 GB RAM)
 - 18核Intel Xeon 5220 CPU (2.2 GHz, 125 GB RAM)
- 模拟的Mahimahi网络:
 - 固定带宽: 24-60 Mbps
 - 延迟: 5-20 ms
- 其他说明
 - 默认相机旋转速度: 每秒400°

主要评估在真实边缘硬件和模拟网络上运行, 而非使用已部署的PTZ摄像机。

Evaluation

对不同工作负载的实验表明，在相同的资源使用情况下，相比于固定方向方案，MadEye 的准确率提高了 2.9-25.7%

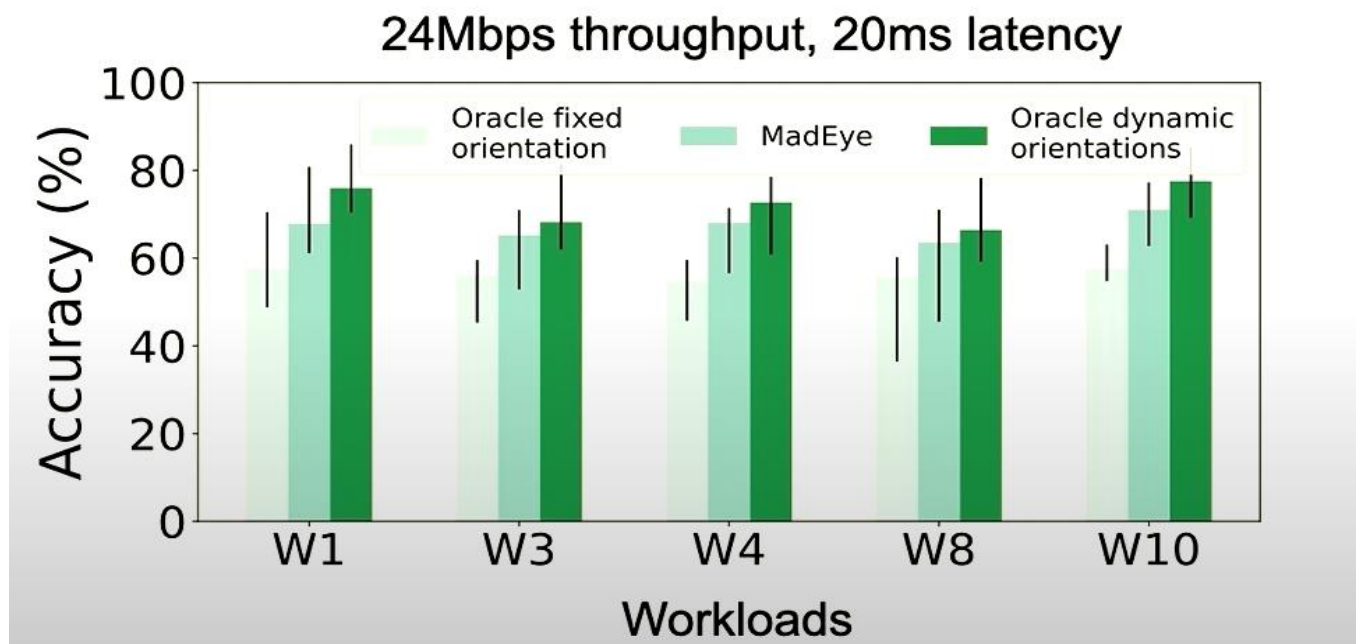
24Mbps throughput, 20ms latency



高网络速度+低帧速率 --> 更多方向探索

评估实验结果与best dynamic方案存在一定的差距

- 优化方向选择逻辑
- 综合考虑距离较远方向的图像



请老师和同学们批评指正！

汇报人：陈宇杰