
南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

THE UNIVERSITY OF SYDNEY

max planck institut informatik

Hierarchical Online Instance Matching for Person Search

Di Chen, Shanshan Zhang, Wanli Ouyang, Jian Yang, Bernt Schiele

11 Feb. 2020

Contents

- Background
- Motivation
- Our Solution
- Summery

Contents

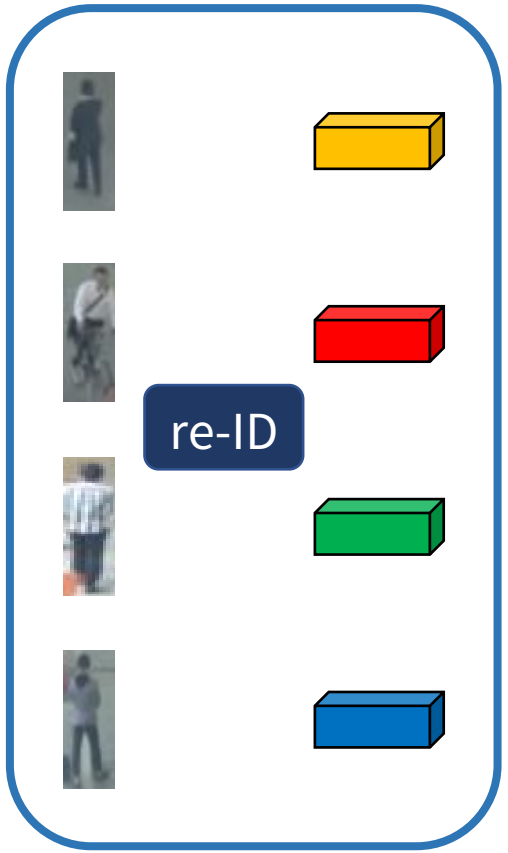
- Background: video surveillance
- Motivation
- Our Solution
- Summery

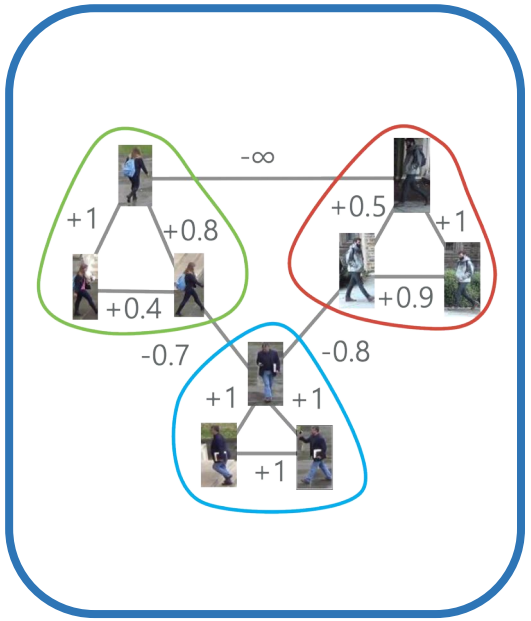
Our Goal: Multi-target Multi-Camera Tracking

Credit: DukeMTMC

- Spatial position (bounding box)
- Identity-aware
- Intra/Inter-camera trajectory

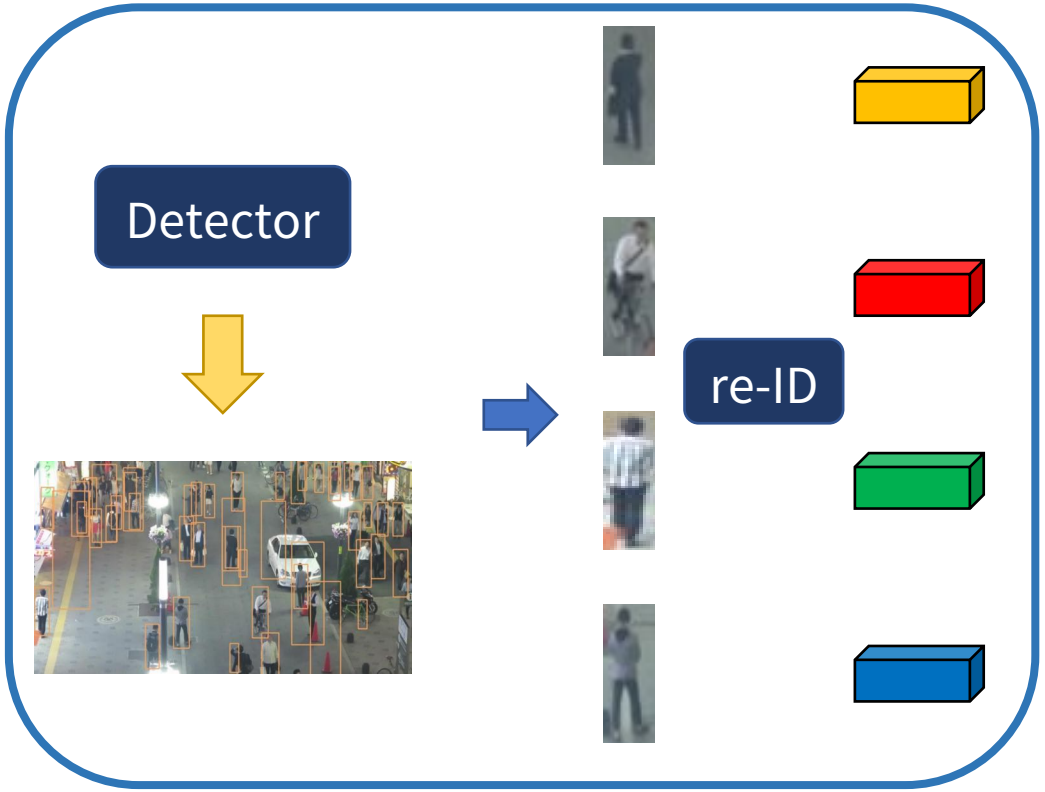
A Typical Pipeline

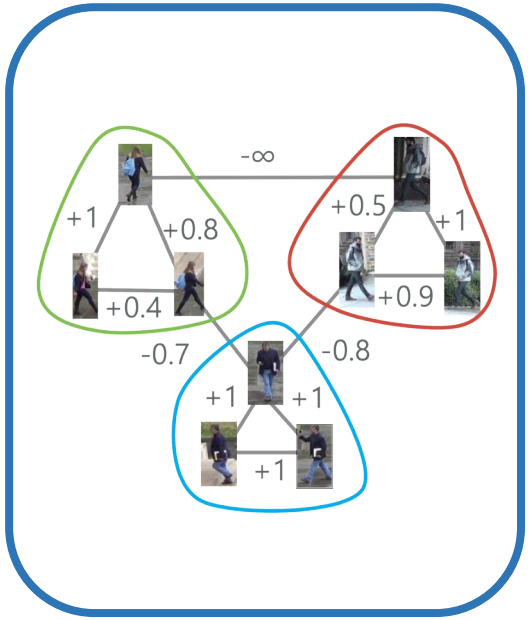




Video Streams → Person Detection → Discriminative Feature → Association → Trajectories

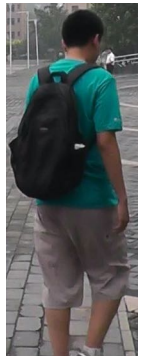
Where Person Search lies

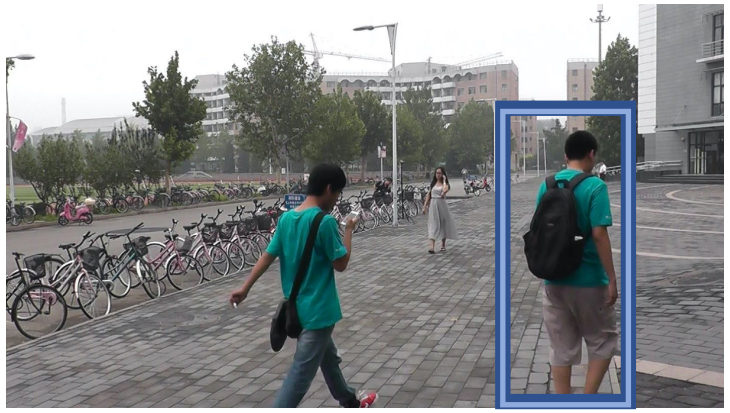




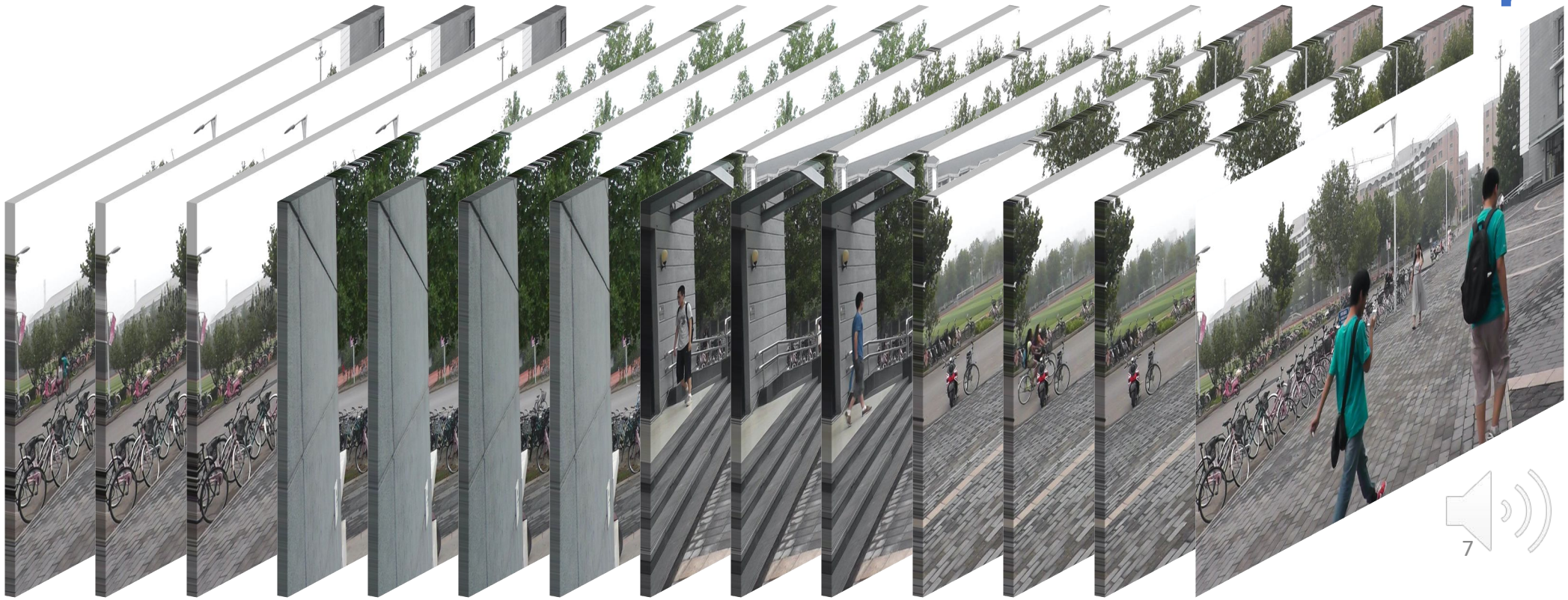
Video Streams → Person Search = Detection + Re-ID → Association → Trajectories

Probe/Query
Cam 1





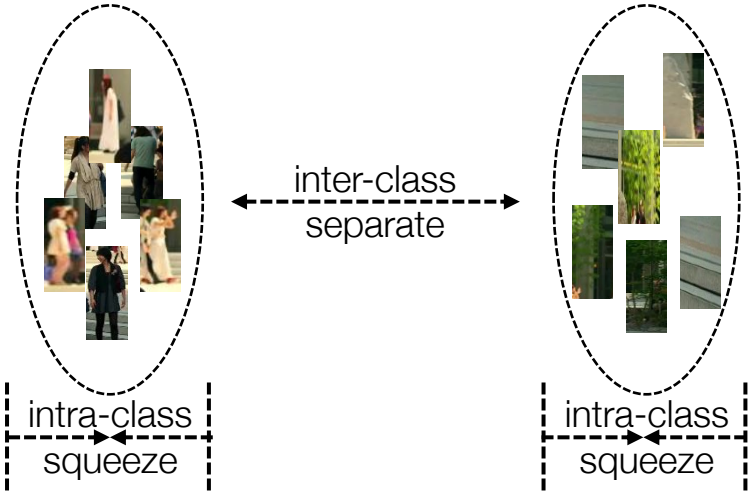
Gallery
Cam 2, 3, …

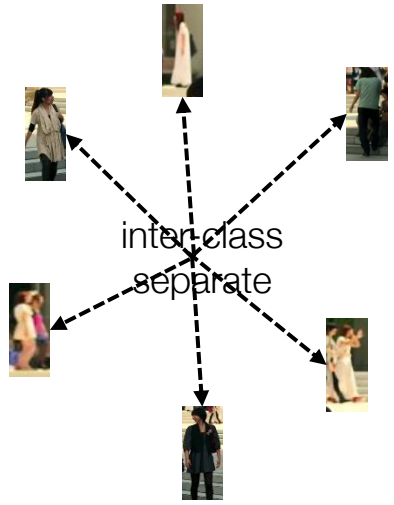


Contents

Inharmonious Combination

- Contradictory Objectives
  - **Detection**: Squeeze persons together, regardless of identity
  - **Re-ID**: Separate persons, identity-aware





Inharmonious Combination

- Contradictory Objectives
  - **Detection**: Squeeze persons together, regardless of identity
  - **Re-ID**: Separate persons, identity-aware

- Distinct Focusing Feature
  - Detection:
    - Favors figure shape
    - Robust to color changes
    - Robust to part details, etc.
  - Re-ID
    - Robust to figure shape
    - Favors color changes
    - Favors part details, etc.

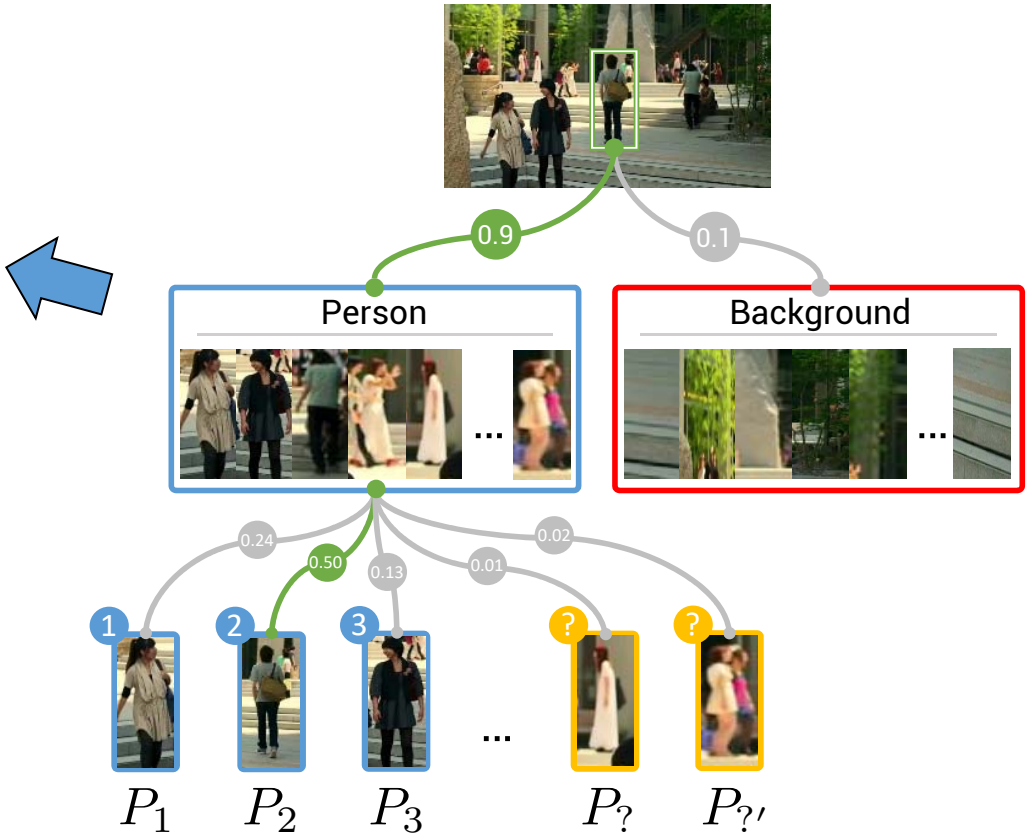
Contents

- Background
- Motivation
- Our Solution: HOIM & SMR
- Summery

Our Solution: HOIM

---- Explicitly model the Hierarchical relationship between Det & Re-ID

$$P_{\text{person}} = \sum_{\text{id}=1}^{N} P_{\text{id}} + \sum_{\text{id}\in\{?\}} P_{\text{id}}$$



$$P_{\text{bg}} = \sum_{j\in\{\text{bg}\}} P_j$$

Hierarchical Softmax

---- Hierarcahy ONE: Detection

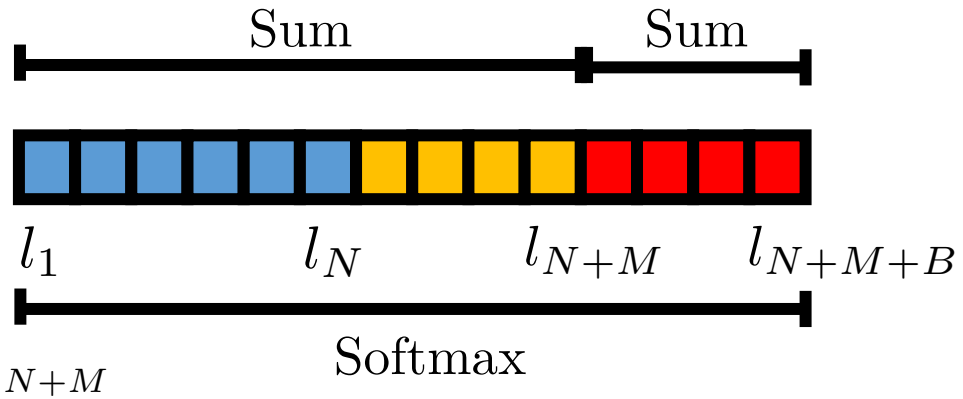
Given a feature vector $\mathbf{f}$ for a proposal,

The identity-wise linear projection:

$$\mathbf{L} = \mathbf{W}\mathbf{f} \in \mathbb{R}^{(N+M+B)\times 1}$$

Then normalize: $\mathbf{P} = \mathrm{Softmax}(\mathbf{L})$

where $p_i = \dfrac{e^{l_i/\tau}}{\sum_{j=1}^{N+M+B} e^{l_j/\tau}}$



Then the probability of this proposal being person: $p(\text{is person}) = \sum_{i=1}^{N+M} p_i$

The probability of this proposal being background: $p(\text{is bg}) = \sum_{i=N+M+1}^{N+M+B} p_i$

The detection sub-loss: $\mathcal{L}_{\text{det}}(\mathbf{f}) = -y\log(p(\text{is person})) - (1-y)\log(p(\text{is bg}))$

Hierarchical Softmax

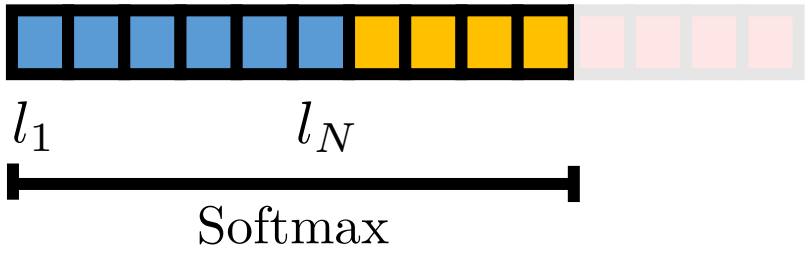
---- Hierarcahy TWO: Re-ID classification

Given a feature vector $\mathbf{f}$ for a proposal,

The identity-wise linear projection:

$$\mathbf{L} = \mathbf{W}\mathbf{f} \in \mathbb{R}^{(N+M+B)\times 1}$$

Slice $\mathbf{L}$ into $\mathbf{L}' = [l_1, l_2, \ldots, l_N, l_{N+1}, \ldots, l_{N+M}]$



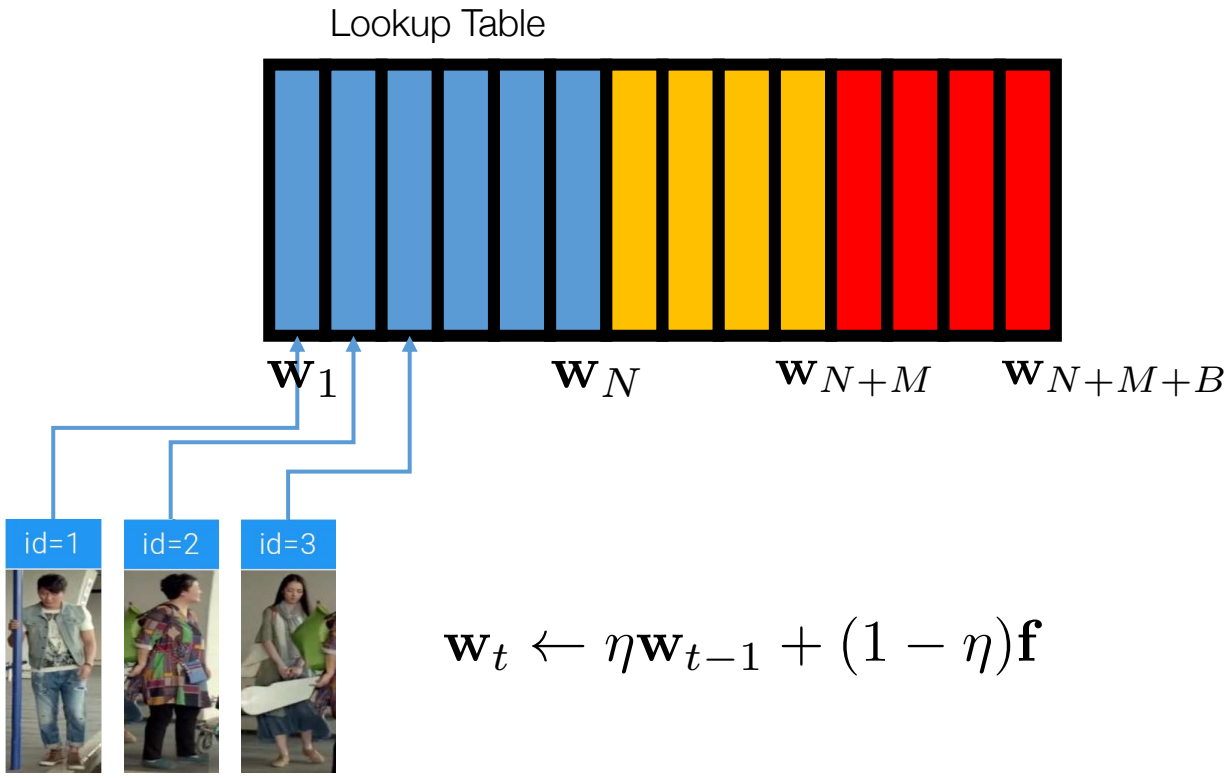
Normalize: $\mathbf{P}' = \mathrm{Softmax}(\mathbf{L}')$

where $p(\text{is identity } i \mid \text{is person}) = \dfrac{e^{l_i/\tau}}{\sum_{j=1}^{N} e^{l_j/\tau} + \sum_{j=N+1}^{N+M} e^{l_j/\tau}}$

The re-ID sub-loss: $\mathcal{L}_{\text{re-ID}}(\mathbf{f}) = -\mathbf{1}_{\text{id}=i} \log(p(\text{is identity } i \mid \text{is person}))$

The whole hierarchical loss: $\mathcal{L}_{\text{HOIM}} = \mathcal{L}_{\text{det}} + \lambda \mathcal{L}_{\text{re-ID}}$

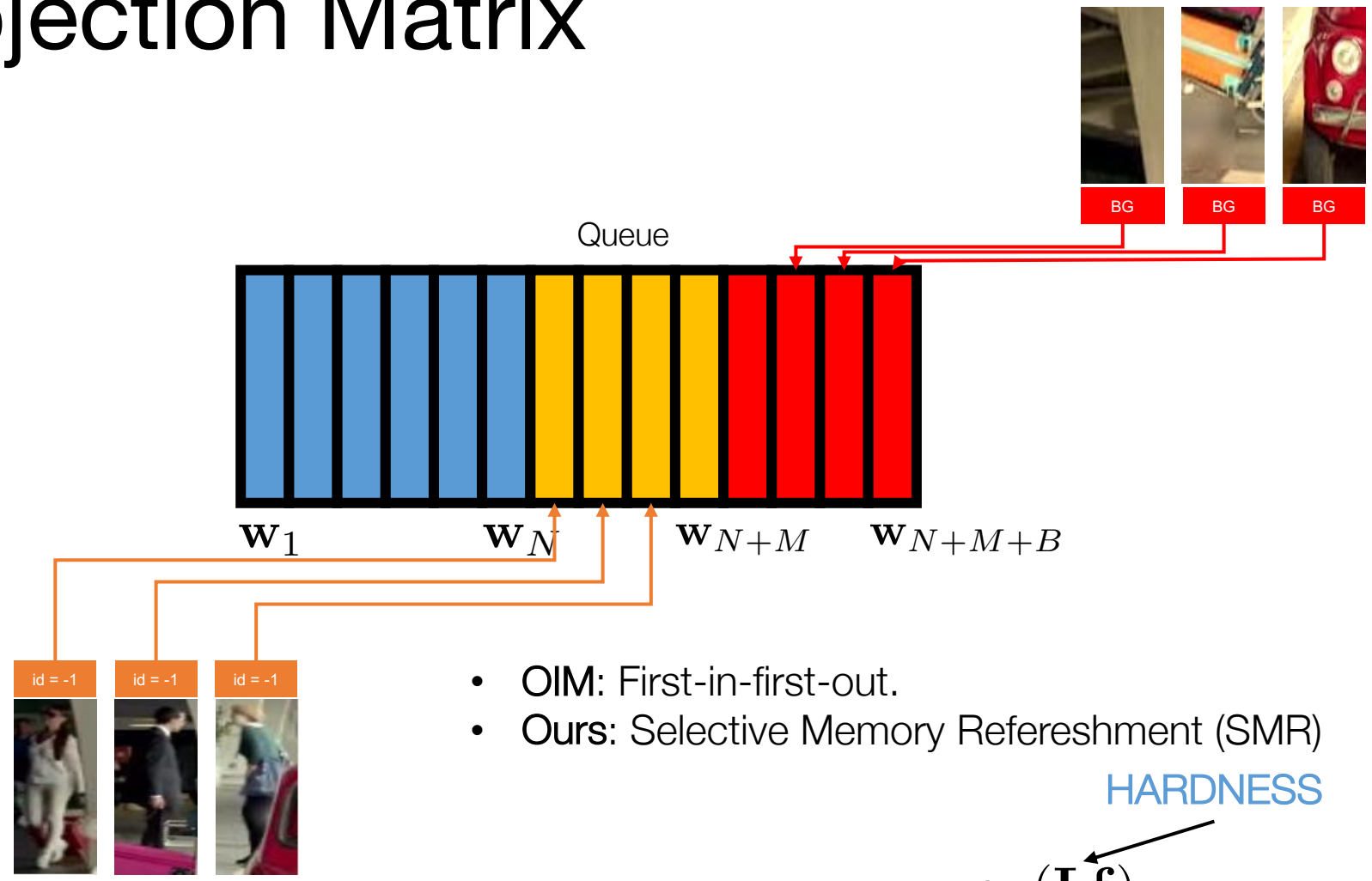
The Projection Matrix



$$\mathbf{w}_t \leftarrow \eta \mathbf{w}_{t-1} + (1-\eta)\mathbf{f}$$

Xiao, Tong, et al. "Joint detection and identification feature learning for person search." in CVPR 2017

The Projection Matrix



- **OIM**: First-in-first-out.
- **Ours**: Selective Memory Refereshment (SMR)

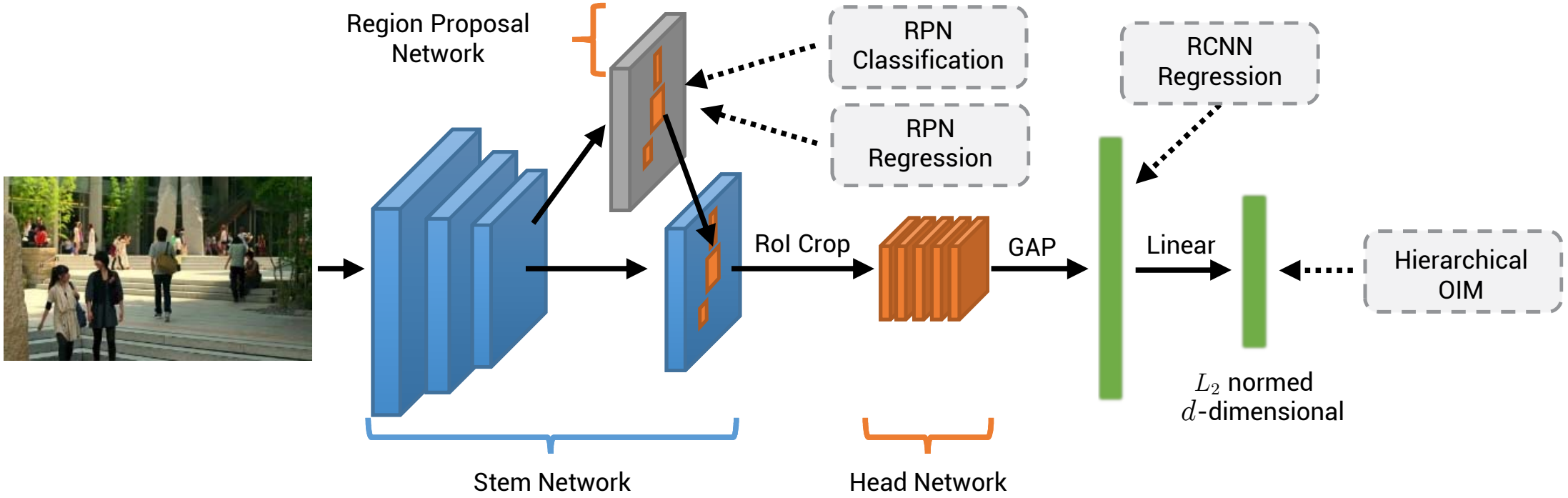
HARDNESS

$$\omega = \frac{\max(\mathbf{Lf})}{\max(\mathbf{Qf}) + \epsilon} \cdot k^{\iota}$$

DIVERSITY

MORTALITY

The Whole Architecture



Analytical Experiments

- Ablation Study:
  - HOIM with SMR improves over a strong baseline

Method	**mAP**(%)	**top-1**(%)	Δ(%)
OIM	75.5	78.7	
OIM-base	83.6	87.4	
+ Focal Loss	85.1	87.6	(+1.5, +0.2)
+ SMR	85.5	88.2	(+0.4, +0.6)
HOIM	**89.7**	**90.8**	(+4.2, +2.6)

Analytical Experiments

- Ablation Study:
  - HOIM and SMR improves over a strong baseline.

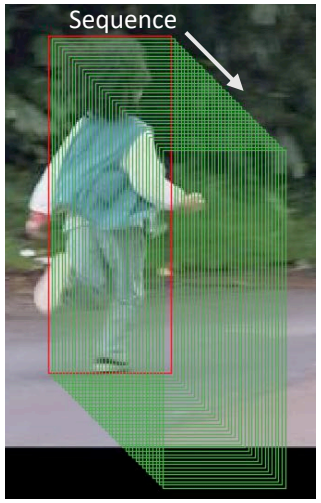
- Model Inspection:
  - HOIM has better detection accuracy than OIM.

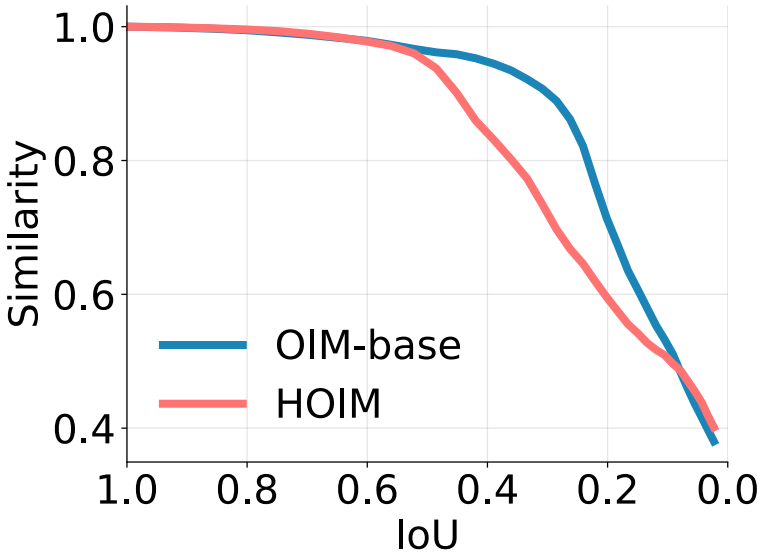
Method	RPN		Faster R-CNN	
	Recall(%)	**AP**(%)	**Recall**(%)	**AP**(%)
detector	89.27	69.07	93.12	87.02
OIM-base	-9.1	-21.69	-12.01	-11.18
HOIM	-0.73	-12.66	-1.36	-1.35

Analytical Experiments

- Ablation Study:
  - HOIM and SMR improves over a strong baseline.

- Model Inspection:
  - HOIM has better detection accuracy than OIM.
  - HOIM embeddings are more discriminative under false detections

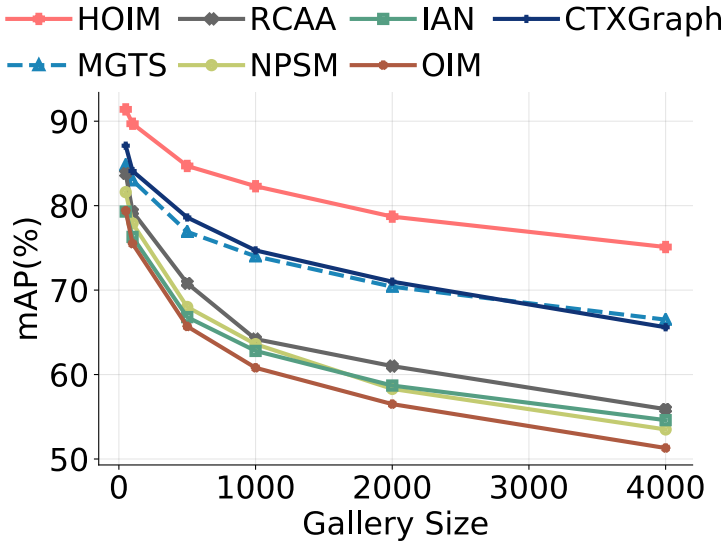




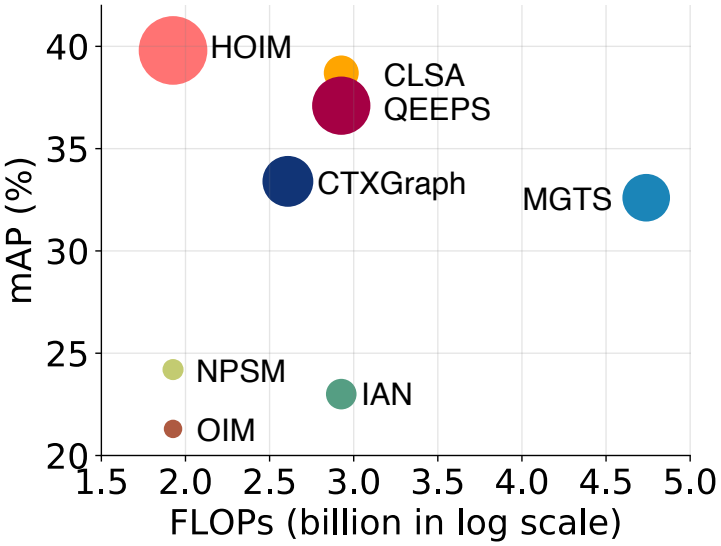
Performance Comparison

Method	CUHK-SYSU		PRW	
	mAP	top-1	mAP	top-1
DPM + IDE w. CWS	-	-	20.5	48.3
CNN + MGTS	83.0	83.7	32.6	72.1
CNN + CLSA	87.2	88.5	38.7	65.0
OIM	75.5	78.7	21.3	49.9
IAN	76.3	80.1	23.0	61.9
NPSM	77.9	81.2	24.2	53.1
RCAA	79.3	81.3	-	-
CTXGraph	84.1	86.5	33.4	73.6
QEEPS	88.9	89.1	37.1	76.7
Ours	**89.7**	**90.8**	**39.8**	**80.4**

- HOIM outperforms both two-stage methods (upper block) and one-stage methods (lower block).



- Performance advantage is consistent over different gallery sizes.



- HOIM excels on both speed and accuracy.

Contents

- Background
- Motivation
- Our Solution
- Summery

Summery

The task:

Person Search = Pedestrian Detection + Re-ID

The Motivation:

Conflict objectives for detection and re-ID

Distinct focusing feature

The Solution:

Use the task prior: hierarchical structure between detection and re-ID

Selective Memory Refreshment for softmax weight updating

Experiments:

HOIM has better detection accuracy than baseline

HOIM embeddings are more discriminative under false detections

Balanced speed and accuracy

Summery

The task:

Person Search = Pedestrian Detection + Re-ID

The Motivation:

Conflict objectives for detection and re-ID

Distinct focusing feature

The Solution:

Use the task prior: hierarchical structure between detection and re-ID

Selective Memory Refreshment for buffer updating

Experiments:

HOIM has better detection accuracy than baseline

HOIM embeddings are more discriminative under false detections

Balanced speed and accuracy

Thank you!

For more questions: `dichen@mpi-inf.mpg.de`