WHU_NERCMS at TRECVID2018: INS

Dongshu Xu, Longxiang Jiang, Xiaoyu Chai, Jin Chen,
Han Fang, Li Jiao, Jiaqi Li, Shichen Lu, and Chao Liang

National Engineering Research Center for Multimedia Software
Wuhan university, Wuhan, 430072, China
cliang@whu.edu.cn

Category

1. Introduction
2. Our approach
3. Results & conclusions

Introduction

TRECVID 2018 INS Task

- Given person name, example images and shots
- Given scene name, example images and shots
- Retrieve specific person in specific scene

Person
(Jane)

Scene
(cafe2)

Specific person in
specific scene

Category

1. Introduction
2. Our approach
3. Results & conclusions

Framework

1

MTCNN

Resnet50

Face features

f_face

2

Aligned reid

Reid features

f_reid

3

SSD

Local scene features

f_local_scene

4

Resnet 50

Global scene features

f_global_scene

5

Score fusion

Ranking list

Local scene retrieval

Framework

SSD

stage1

Input keyframes

SSD network

Initial pedestrian features

stage2

Query category

Input image

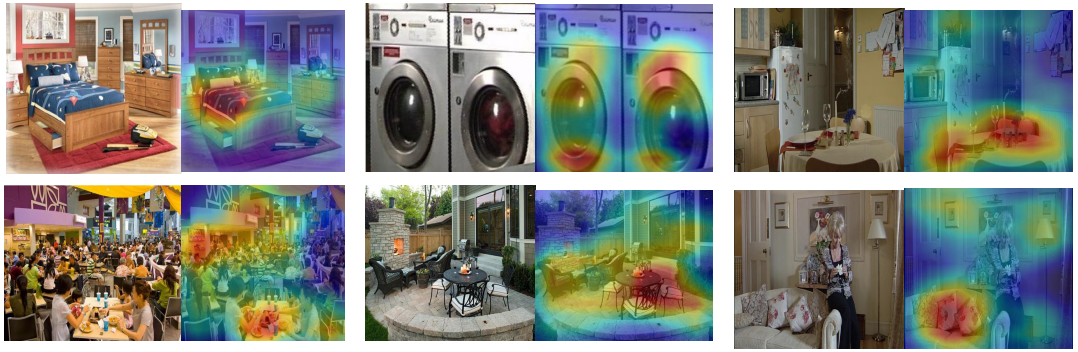
Trained SSD network

Expected results

Global scene retrieval

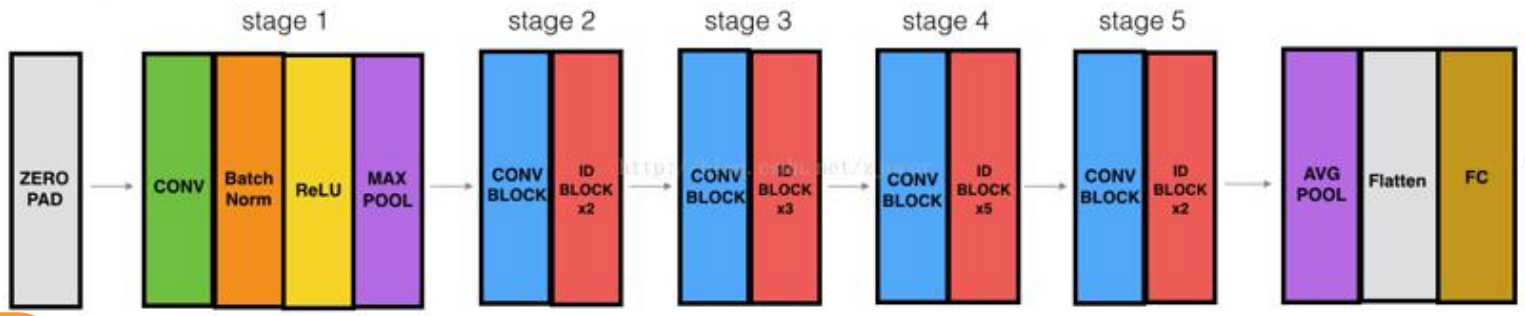
Places365-CNN

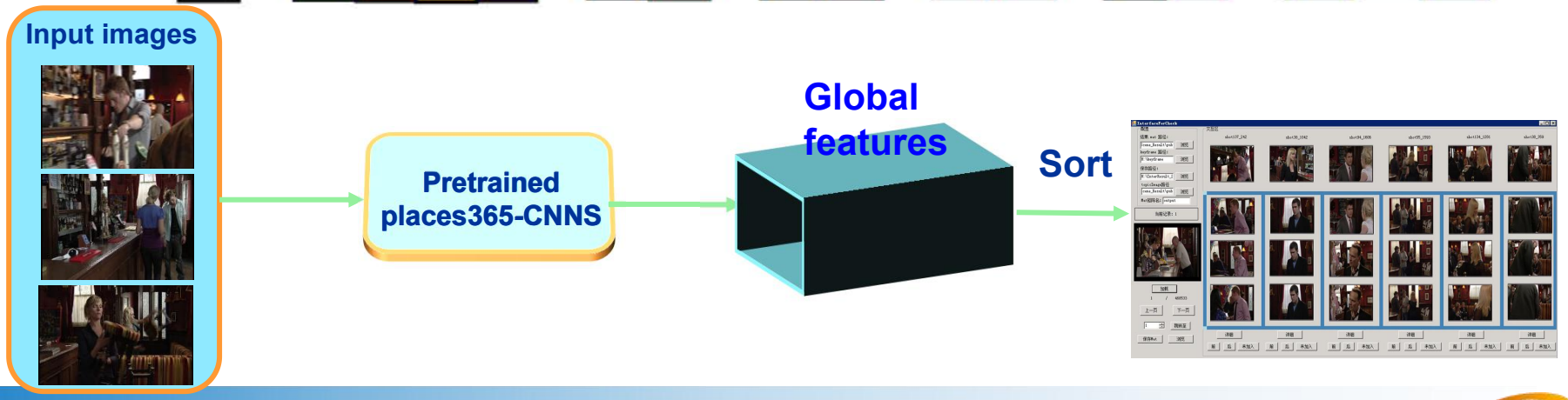
The dataset covers 365 image scenes and also provides pre-trained models for multiple network architectures.



Network

Resnet50





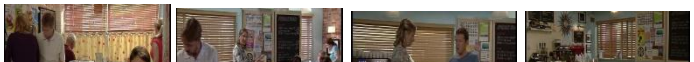
NERCMS

Training samples of scene retrieval

Training Dataset

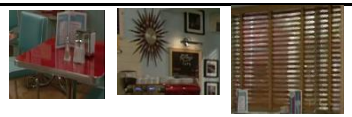
From different views:

cafe 2



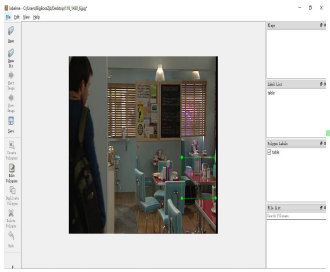
laun

From different objects:

Scene	Landmarks
Pub	
Cafe2	
Laun	
Market	

Datasets production

Keyframes are labelled with landmarks



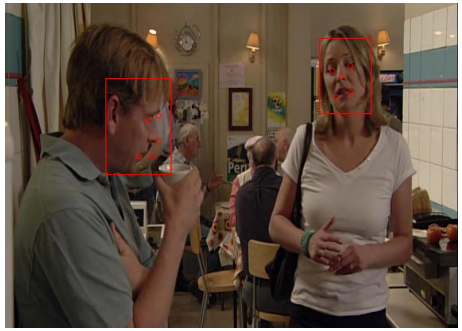
```
<?xml version="1.0"?>
- <annotation>
    <folder>pub</folder>
    <filename>1_1407_0</filename>
    <path>/home/jl/BBox_label/locations/pub/1_1407_0.jpg</path>
  - <source>
      <database>Unknown</database>
    </source>
  - <size>
      <width>768</width>
      <height>576</height>
      <depth>3</depth>
    </size>
    <segmented>0</segmented>
  - <object>
      <name>1</name>
      <pose>Unspecified</pose>
      <truncated>0</truncated>
      <difficult>0</difficult>
    - <bndbox>
        <xmin>478</xmin>
        <ymin>28</ymin>
        <xmax>753</xmax>
        <ymax>563</ymax>
```

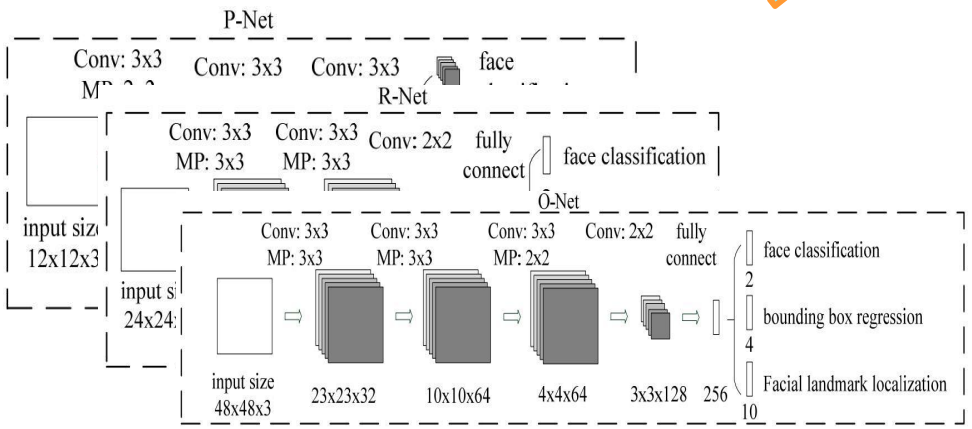
NERCMS

Face recognition

- Face Detection
- Face Alignment
- Feature Extraction
- Distance Measure

MTCNN





Face recognition

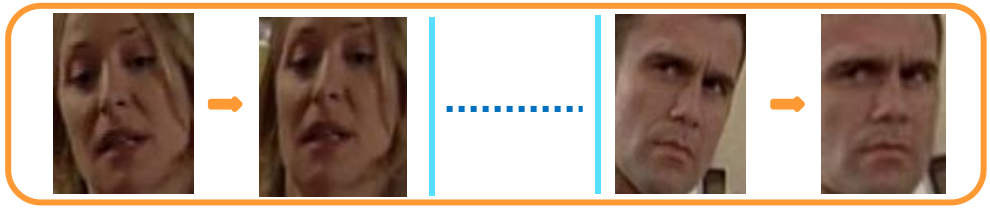
Face Detection

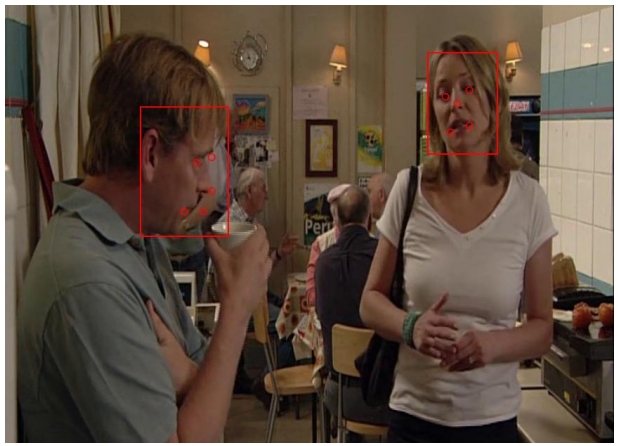
Face Alignment

Feature Extraction

Distance Measure

Similarity transformation





Face recognition

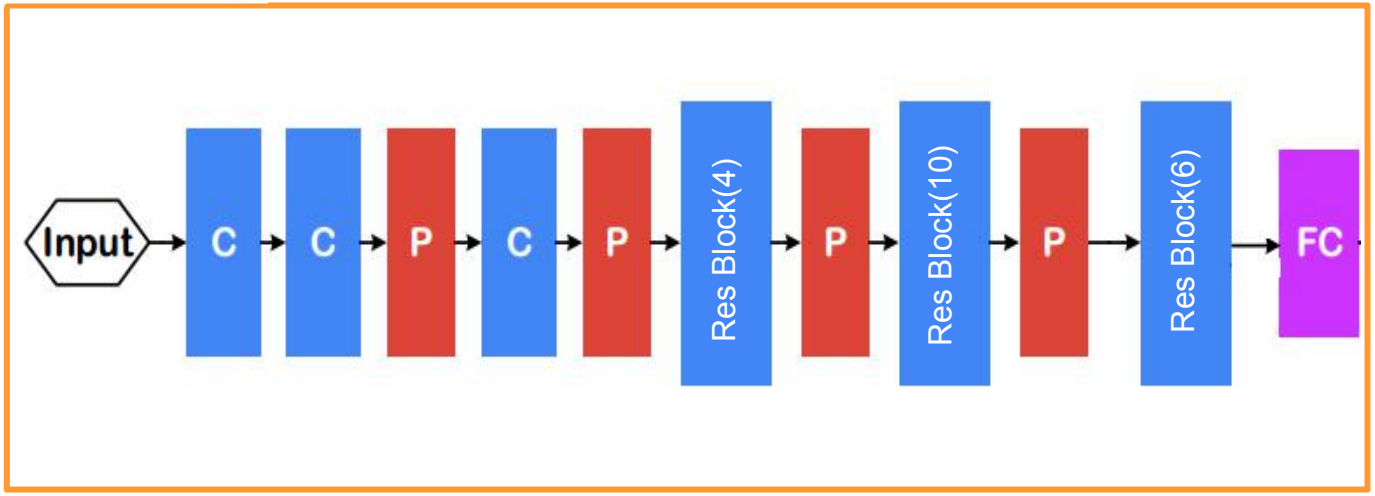
Face Detection

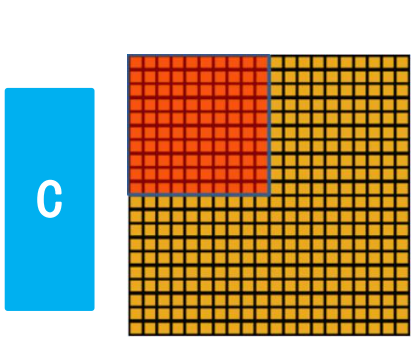
Face Alignment

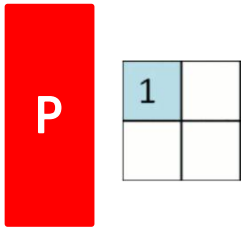
Feature Extraction

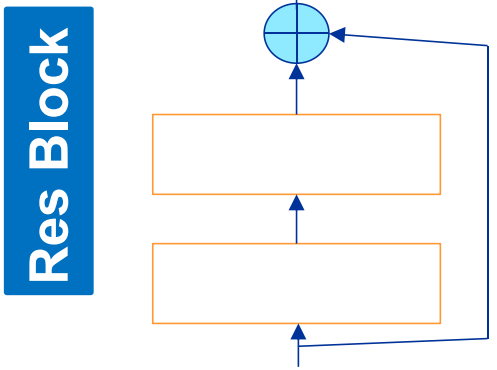
Distance Measure

Face-ResNet









Face recognition

Face Detection

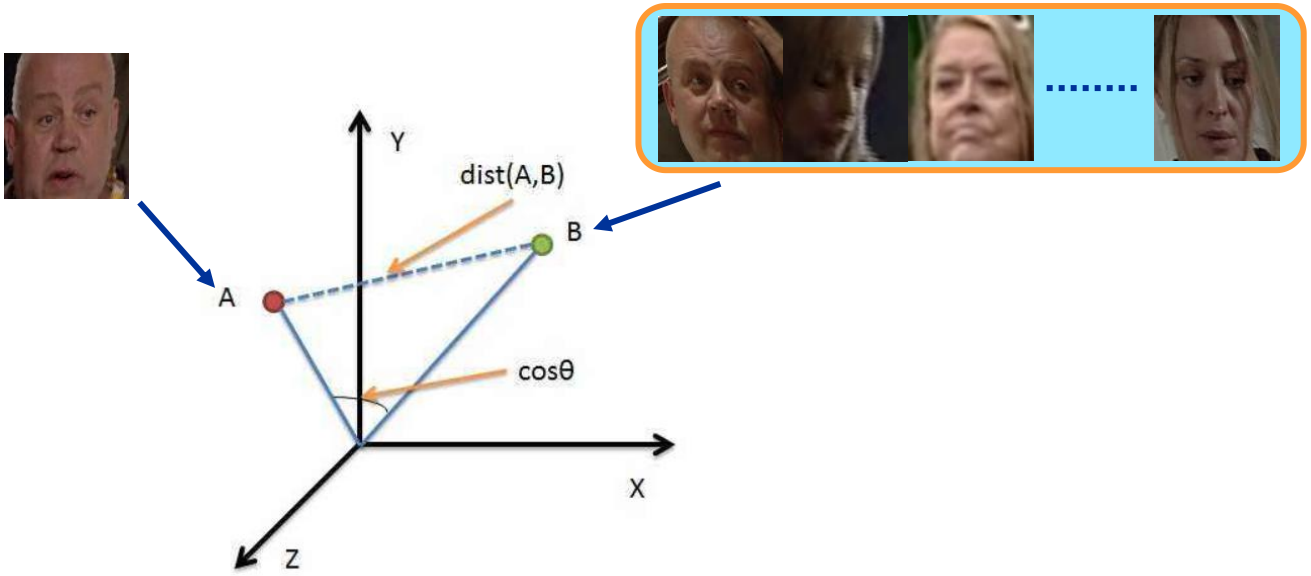
Face Alignment

Feature Extraction

Distance Measure

Cosine distance

$$similarity = \cos(\theta) = \frac{a.b}{\|a\|\|b\|}$$



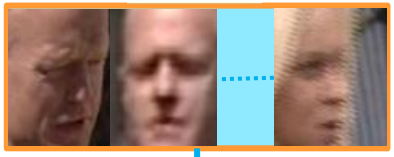
Face recognition

Pipeline

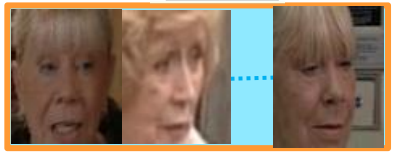
Gallery set

Topic identity

Shot 1



Shot n



Extended reference identity

map

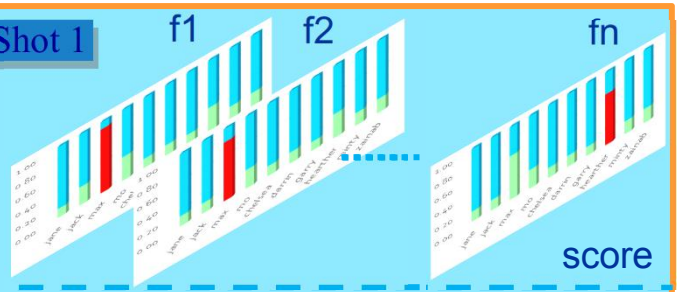
Cosine Distance

Shot 1

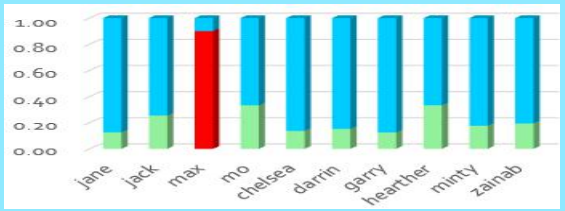
f1

f2

fn



score



Identity ***Max*** has the highest score

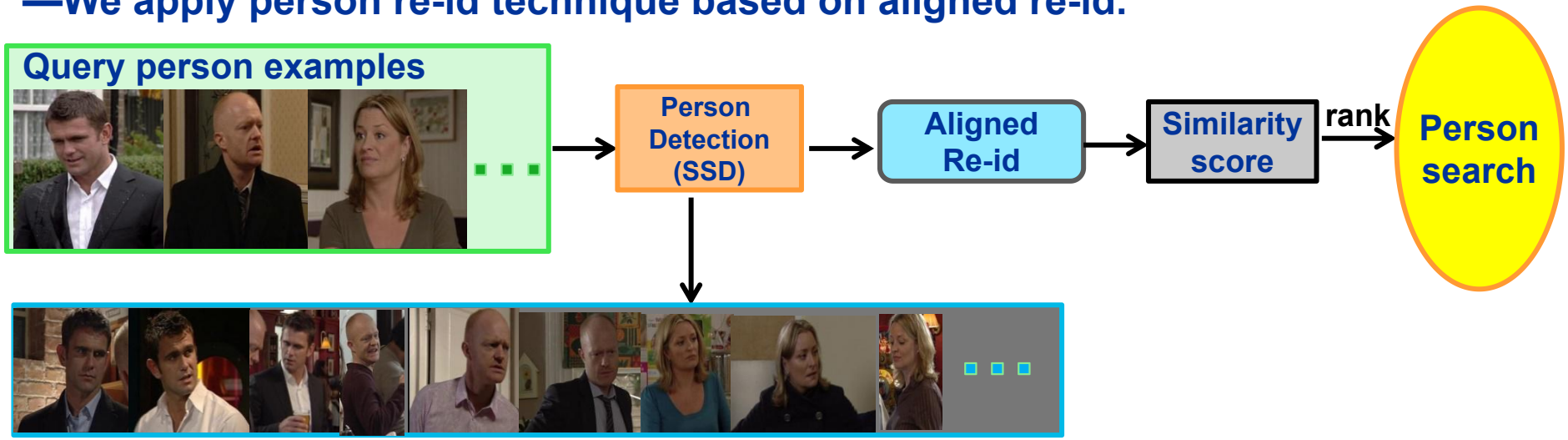
.....

```
for  i=1:n
{ processing the i-th shot
}
```

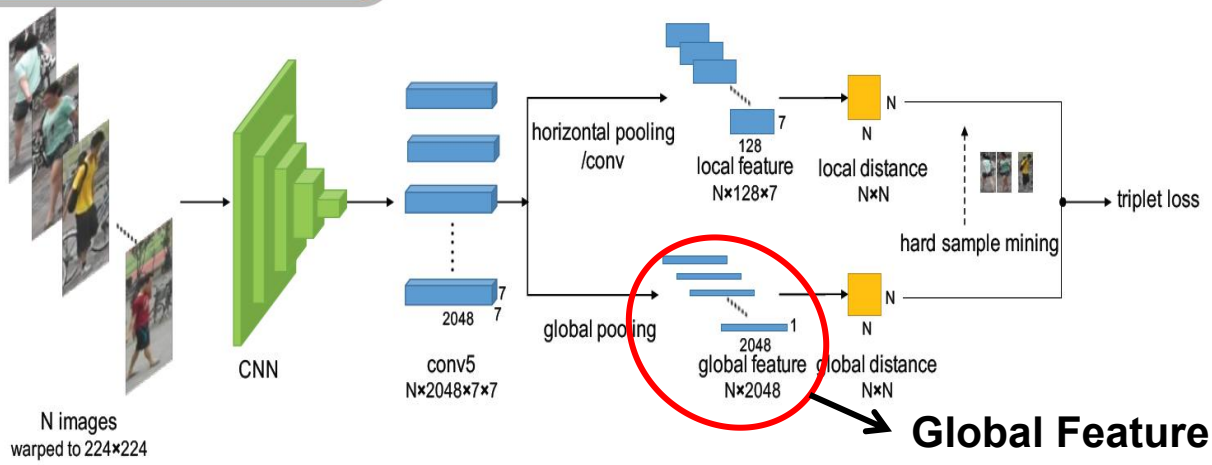
Person re-identification based person search

Person search

—We apply person re-id technique based on aligned re-id.



Aligned Re-id [1]

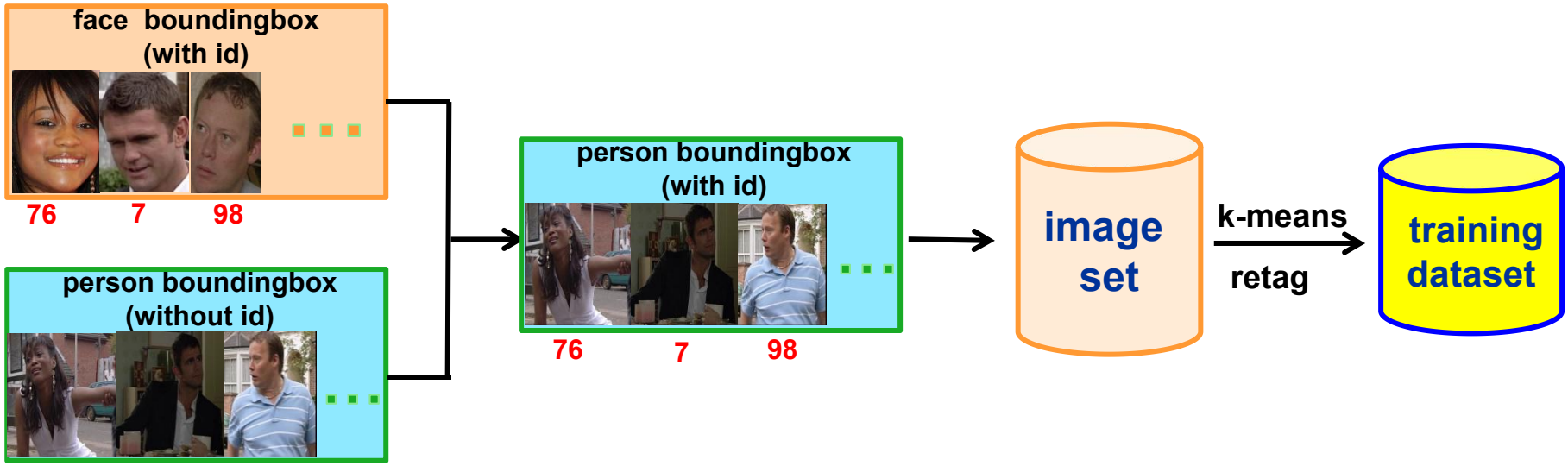


[1] X. Zhang, H.Luo, etc. AlignedReID: Surpassing Human-Level Performance in Person Re-Identification. arXiv:1711.08184v2, 2017

NERCMS

Person re-identification based person search

How to get training dataset



Details of training dataset

Number of images	Number of ids	Number of clusters
2,486,571	194	24864

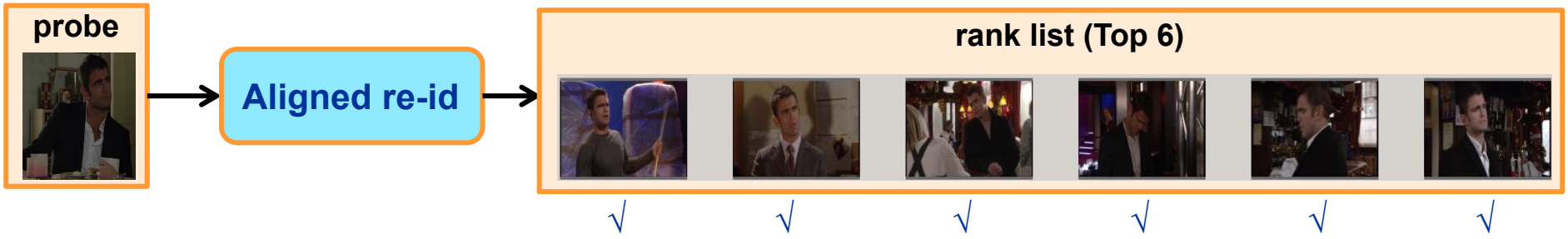
For example

id	numbers of images
1	161955
2	21942
3	671
4	3352
5	5074
6	81586
7	39448
8	8079
9	86527
...	...

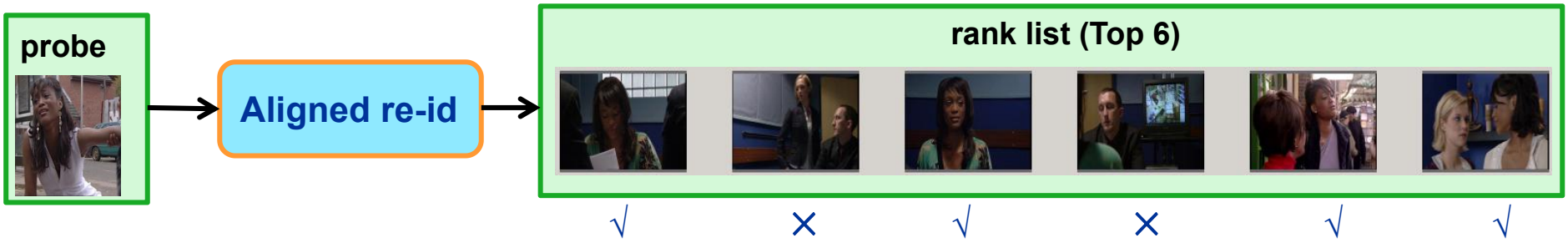
Person re-identification based person search

Visualization results

good



bad



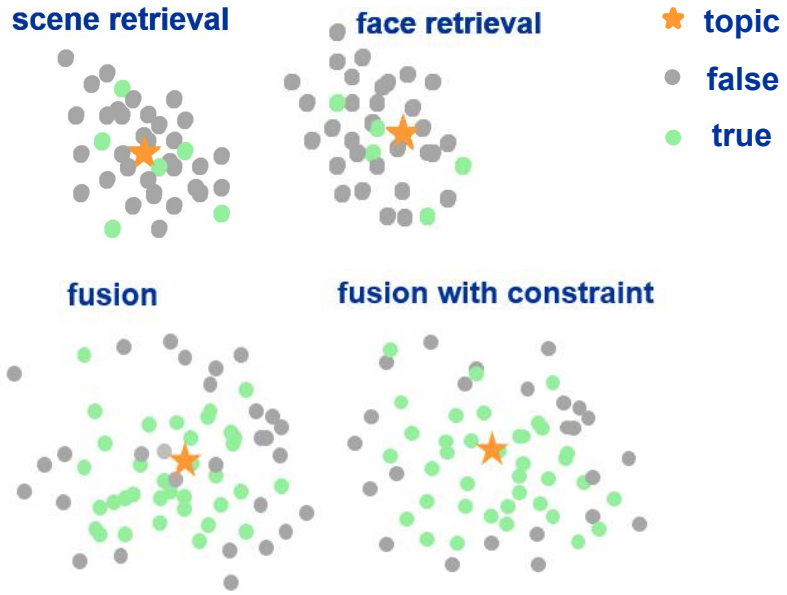
The reason for the bad query is that the clothes are too similar,

Score fusion

Weight based score fusion

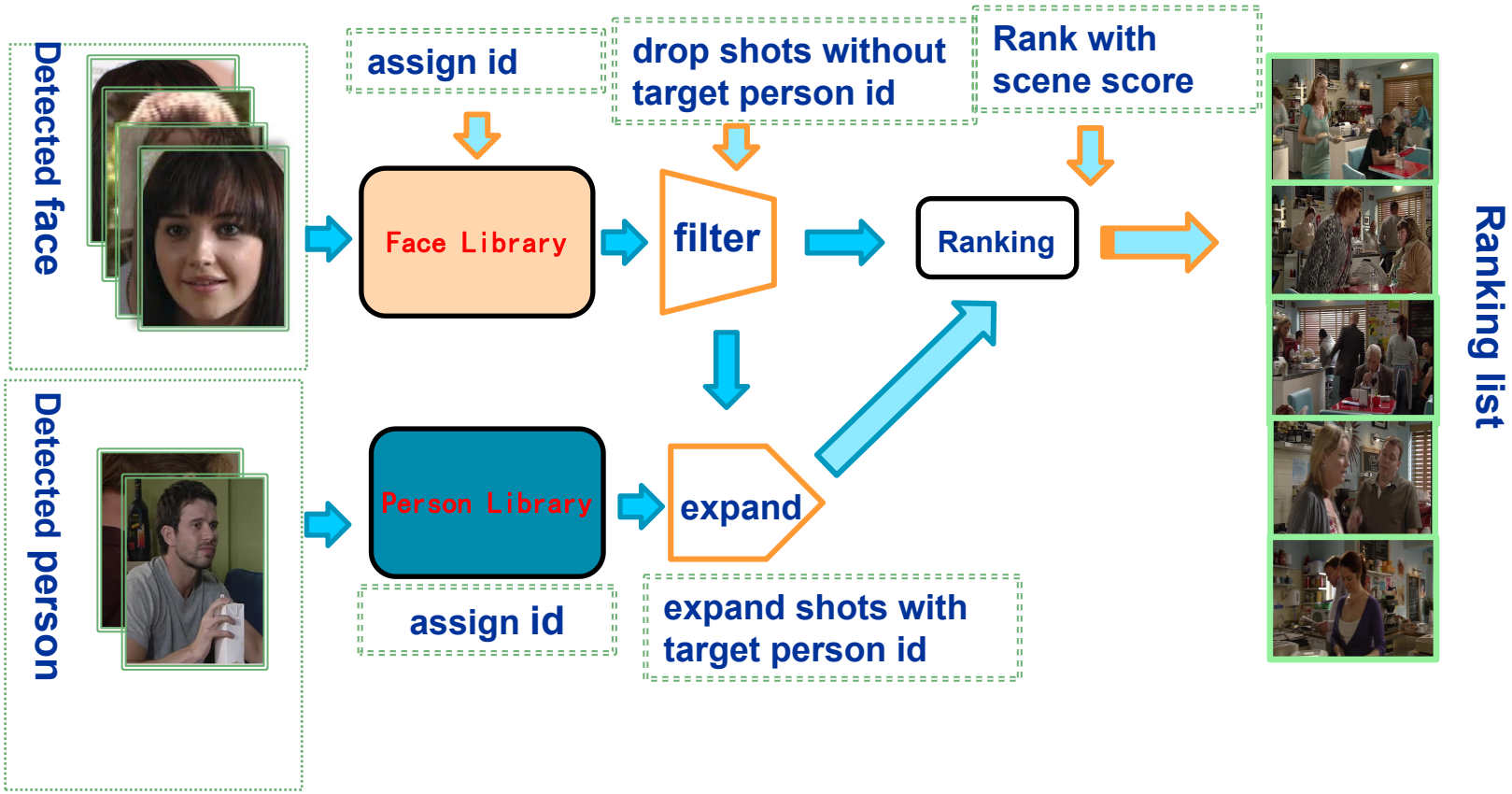
$$f = w_1 * f\_scene + w_2 * f\_face + \exp(- |f\_scene - f\_face|^2)$$

$$f = w_1 * f\_scene + w_2 * f\_face + w_3 * f\_reid + \exp(- |f\_scene - w_2/(w_2+w_3)f\_face - w_3/(w_2+w_3)f\_reid|^2)$$



Score fusion

- Face filter and person expansion



Category

1 Introduction

2 Our approach

3 Results & conclusions

Results & conclusions

Results

Score lists	Fusing method	Auto	Interactive
$f\_scene + f\_face$	weight	0.243	0.261
$f\_scene + f\_face + f\_reid$	weight	0.174	0.184
$f\_scene + f\_face$	filter	0.211	0.235
$f\_scene + f\_face + f\_reid$	filter+expansion	0.182	0.200

Analysis

- The ineffectiveness of reid:
  - IoU computation
  - Cluster strategy
- The effectiveness of fine-tuning:
  - Fine-tuned on some scenes

Results & conclusions

Conclusions

- The face recognition is a key method to identify person. New person search method should be introduced for person images with back and side views or in low resolution
- The training dataset of scene model needs more effective images including different views of positive and negative scenes.
- Score fusion and expansion method is useful to retrieve hard samples.


THANKS