SaFace: Towards Scenario-aware Face Recognition via Edge Computing System

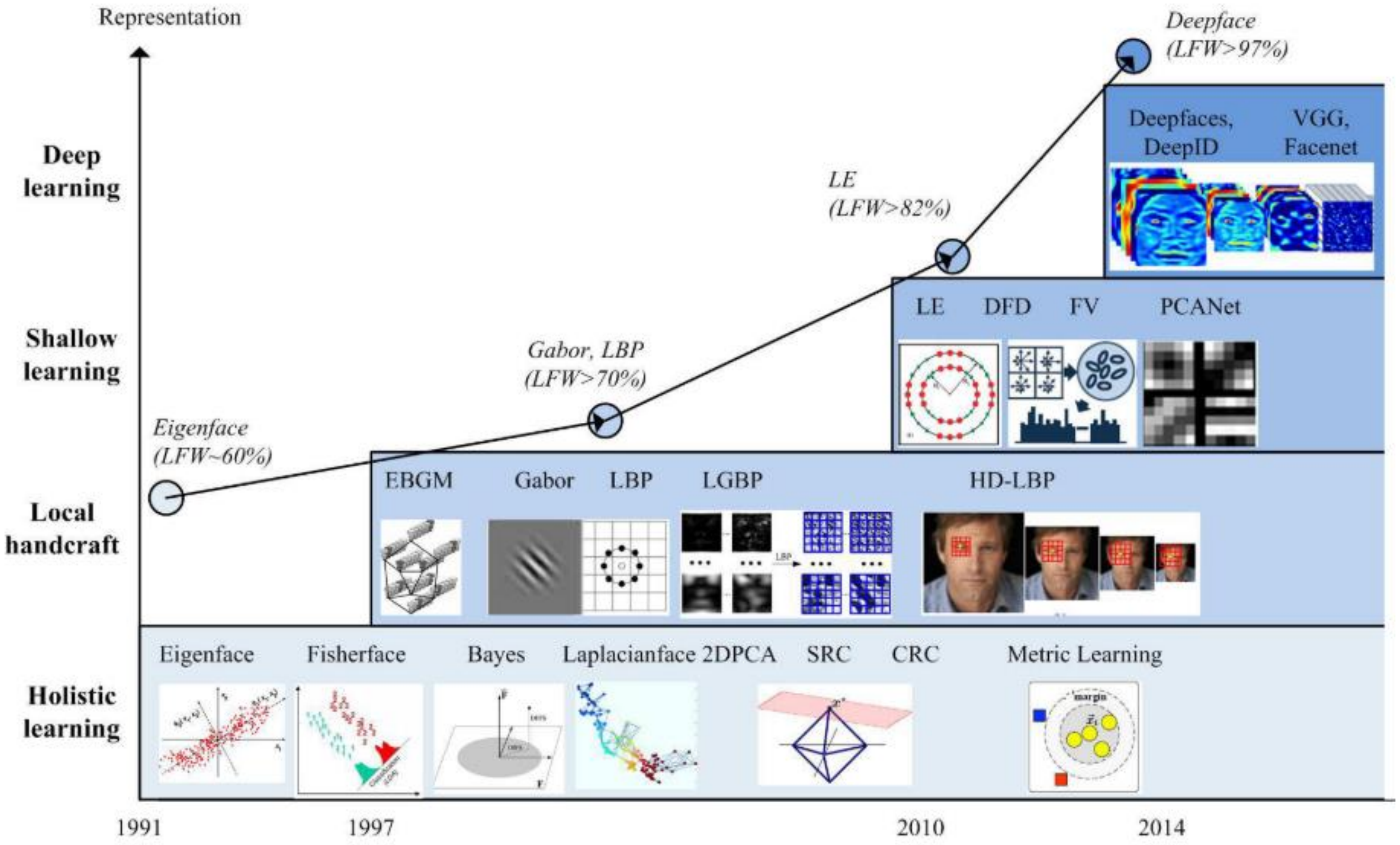
Zhe Zhou[1,2] Bingzhe Wu[1] Zheng Liang[1] Guangyu Sun[1,2] Chenren Xu[1] Guojie Luo[1,2]

[1]Peking University, China

[2]Advanced Institute of Information Technology, Peking University, China

Background

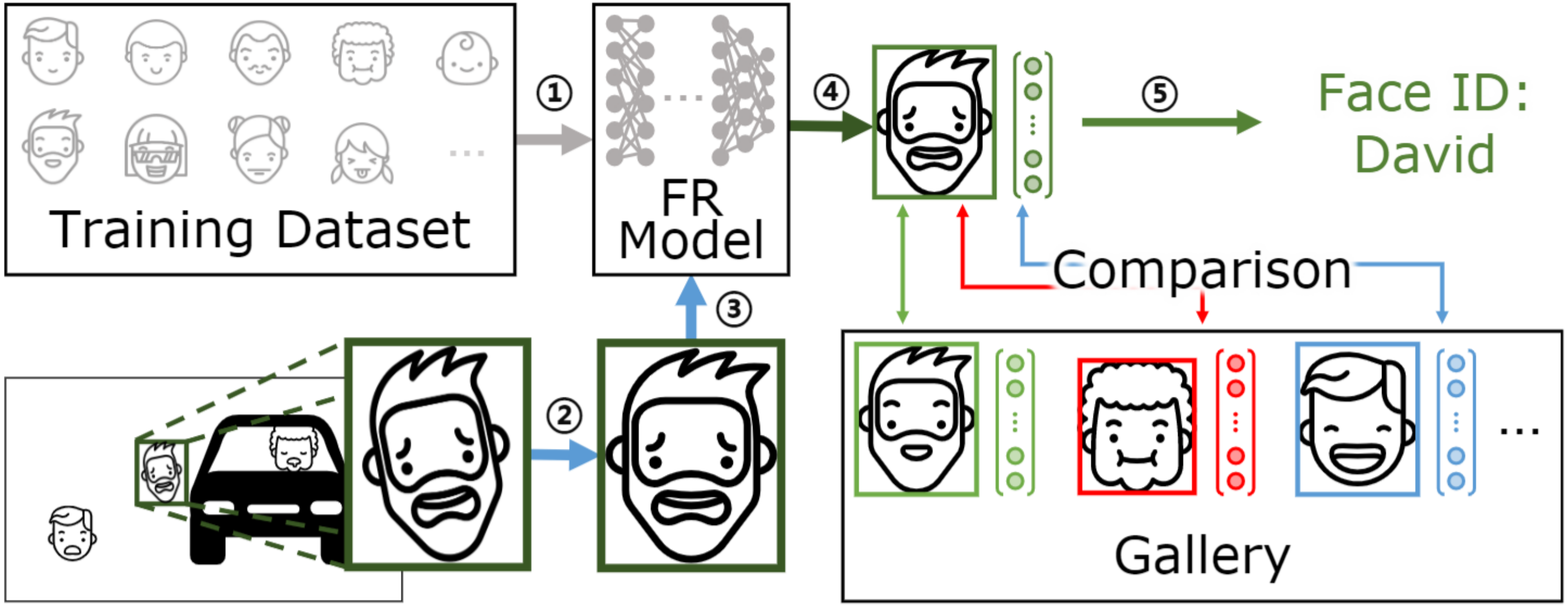
- Deep-learning based FR: outperforms humans in LFW benchmark.



Wang et al. *Deep Face Recognition: A Survey*

Background

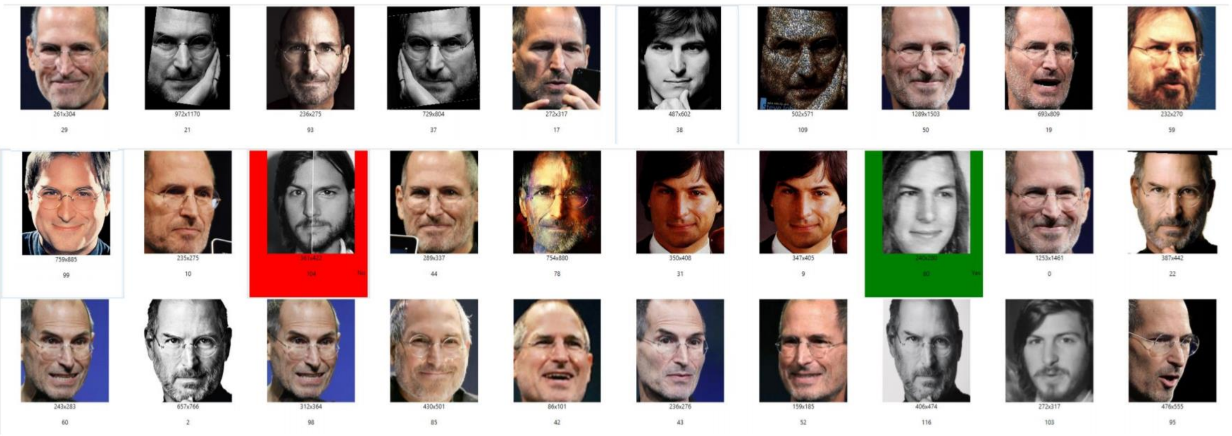
- Basic face recognition (FR) flow:



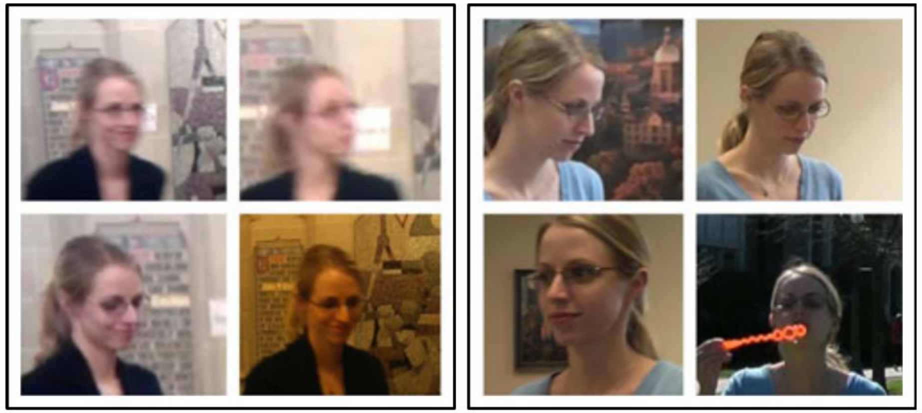
①： FR model training ②： Face detection and alignment ③: Feeding probes into FR model ④: Extracting face representations. ⑤: Comparing and determine the identity.

Motivations

- Deploying FR in real-world scenarios is still challenging:
  - Vast variances between training data and test data.
    - Head poses
    - Illumination
    - Visual quality
  - May result in significant accuracy drop!



MS-Celeb-1M dataset.



Faces in different deployed scenarios[1]

[1]Ding et al. *Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition*

Motivations

- How to build a robust FR system in real-world scenarios?
  - Collect more training data from the target scenario and then fine-tune the FR models.
  - Need to label training data!
    - Labor-intensive.
    - Can not scale in reality.

- Our solution:
  - Use unsupervised online learning to adapt the targeted scenarios.
  - Leverage edge computing paradigm to natively solve the scalability issue.

Unsupervised Online-learning

- Generate training data from the deployed scenario automatically.

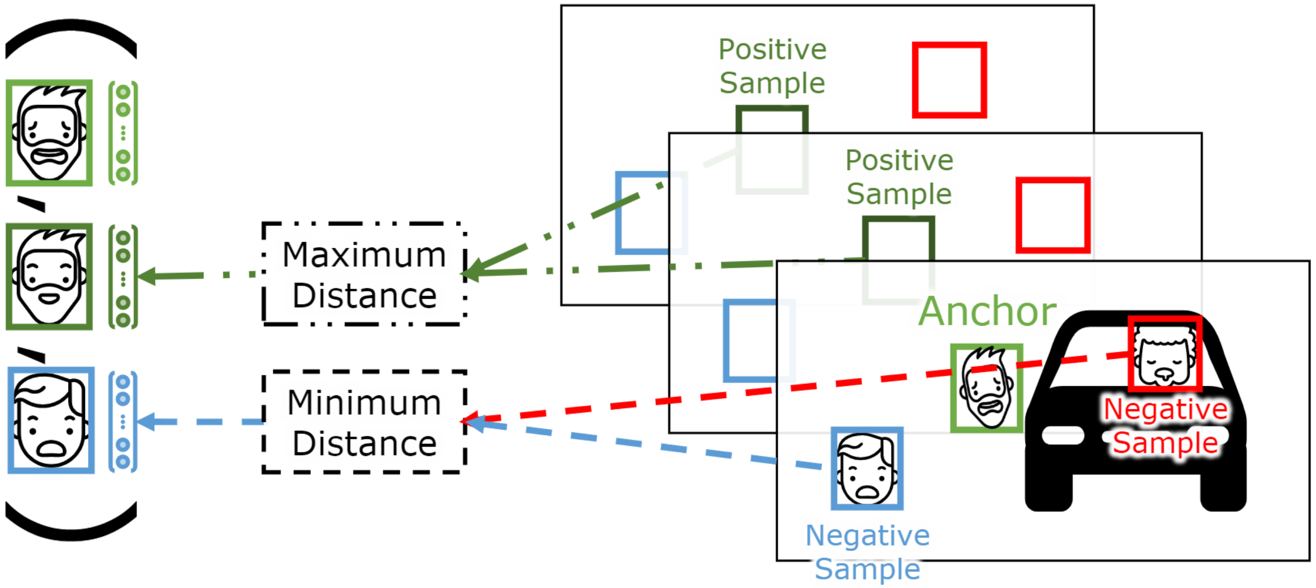


Figure 4: Example of triplets generation.

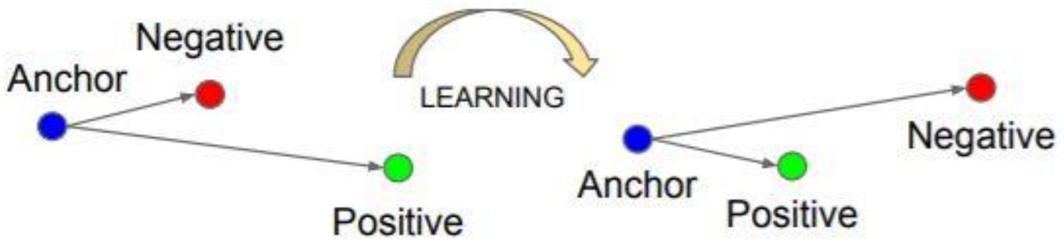
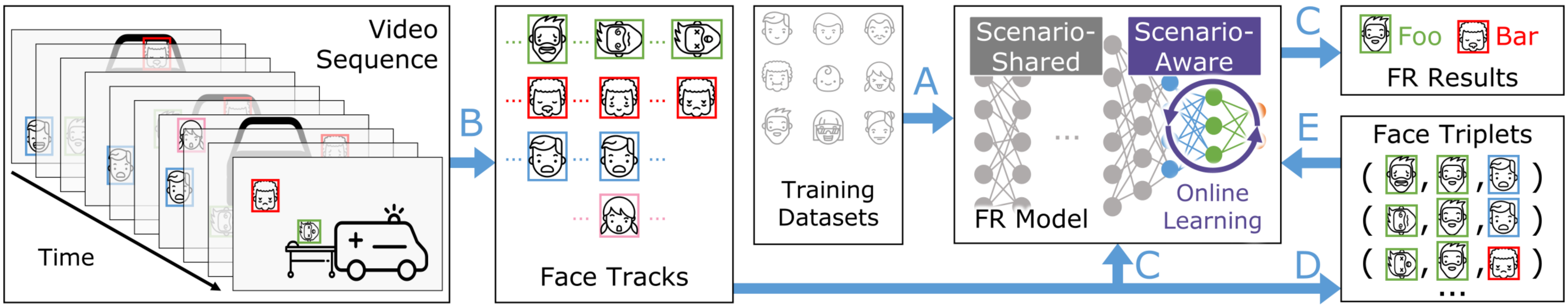


Illustration of Triplet Loss [1]

[1] Schroff et al. Facenet: *A unified embedding for face recognition and clustering*

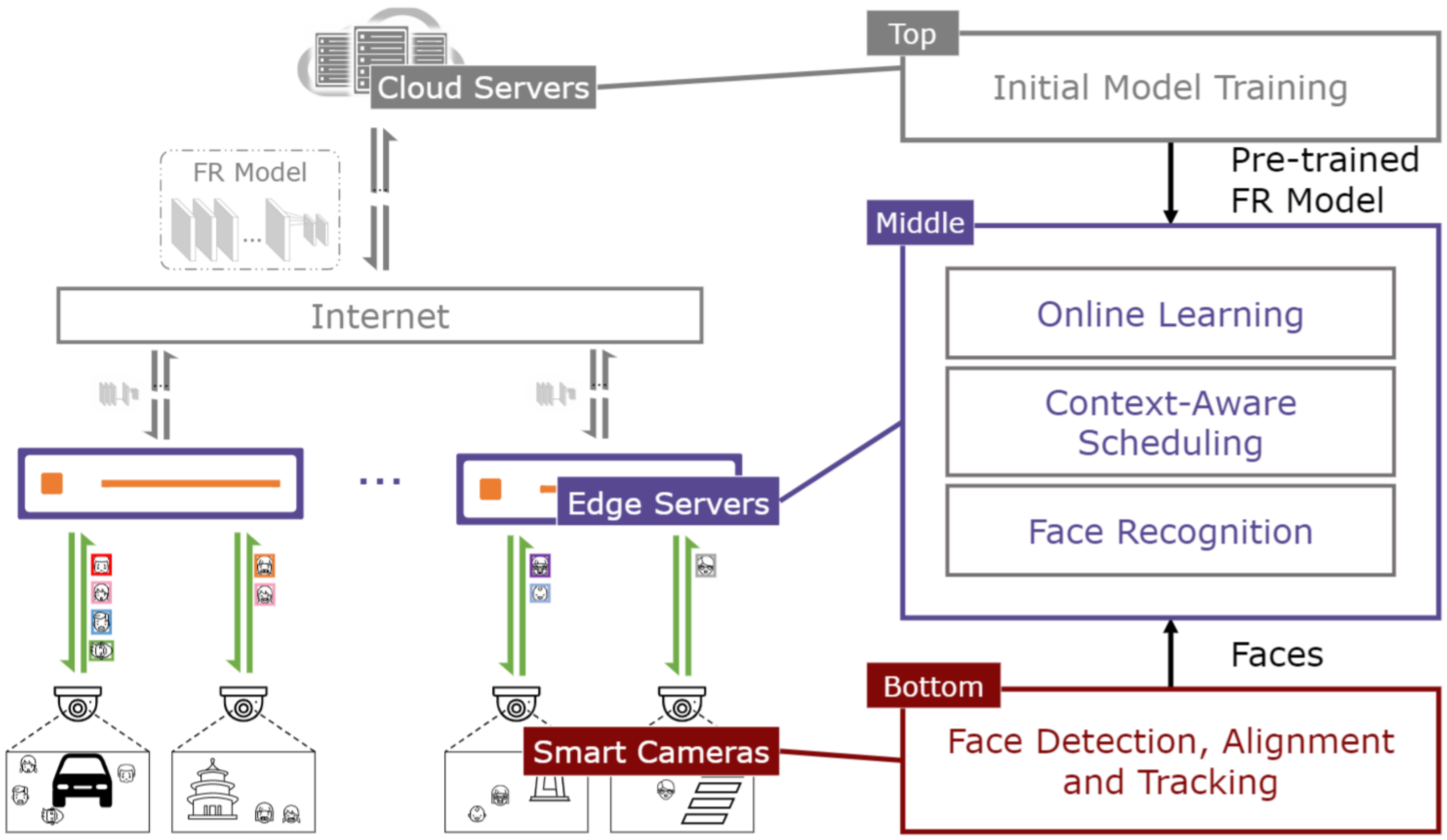
SaFace System

- SaFace workflow:
  - (A) Model pre-training
  - (B) Face detection& tracking
  - (C) FR inference
  - (D) Triplet generation
  - (E) Online learning



SaFace System

- System overview



Scenario-aware Stage

- Context-aware scheduling

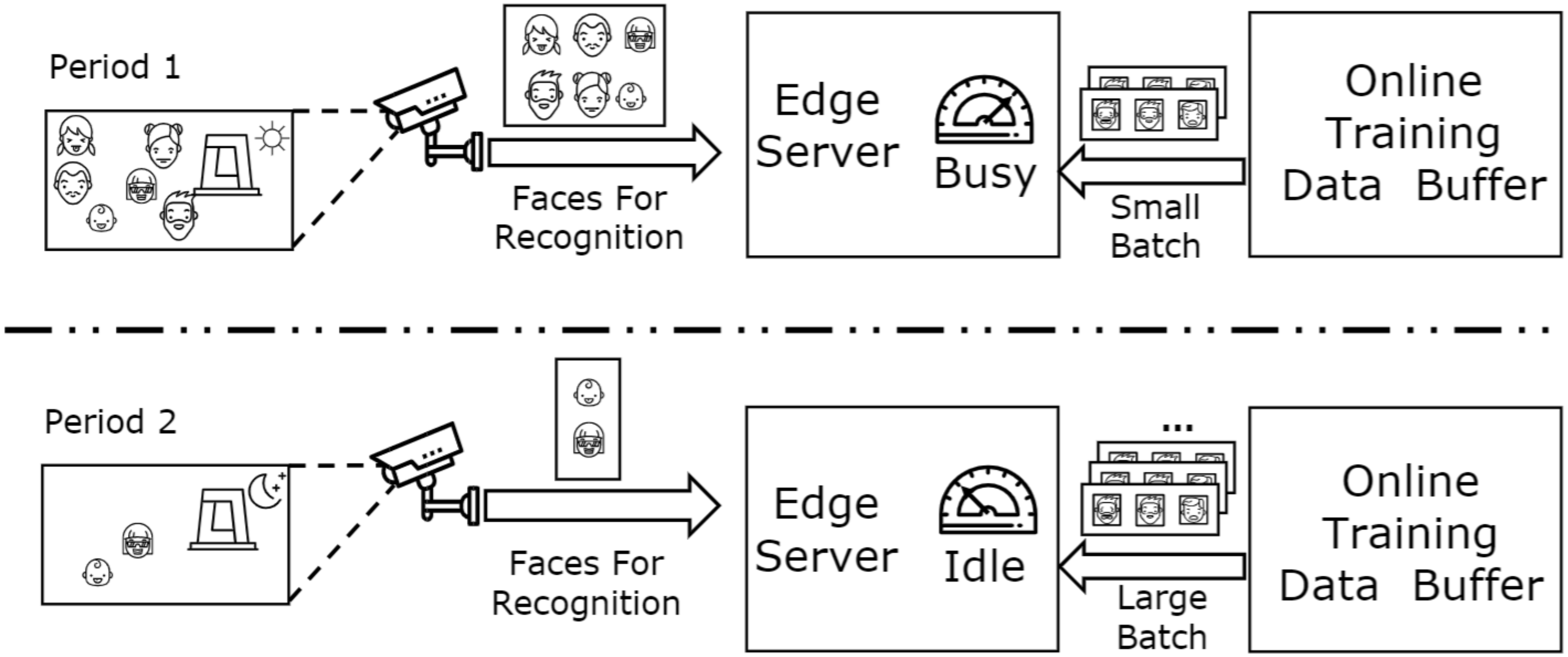


Figure 5: Context-aware Scheduling

Scenario-aware Stage

- Context-aware scheduling
  - R_C : Video frames rate.
  - N_C: The maximum number of cameras.
  - N_{Pmax} : Maximum number of probes contained in a frame.
  - N_E : Maximum number of probes can be processed in a time interval $\Delta t = 1/R_C$.

$$N_E \geq N_C \times N_{P_{max}}$$

  - B_{max} : Maximum batch size.
  - α : A pre-defined coefficient to adjust effective computation utilization.
  - B_t : Optimal runtime batch size of online-learning.

$$B^t = \max(0, B_{max} \times (1 - \alpha \frac{\sum_{i=1}^{N_C} N_{P_i}}{N_E}))$$

Prototype

- System prototype
  - Camera node: Hisilicon Hi3516CV500 IP Camera.
  - Edge node: A desktop PC with Intel i7-6700k CPU and Nvidia GTX1080 GPU.
  - Cloud: A GPU server with 4x GTX1080Ti.
- Communication
  - TP-Link WDR5620 router.
  - 100Mbps LAN.

Node	Platform	Processor	Computing power	Storage	RAM	GPU memory
IoT	HiSilicon Hi3516CV500 IP Camera Soc	2x ARM Cortex-A7+ 1 NPU	500GOPS	4GB	1GB	/
Edge	Desktop PC with GTX1080 GPU	Intel i7-6700k	8.8 TFLOPS	1TB	16GB	8GB
Cloud	GPU server with 4x GTX1080TI	Intel i9-7960x	46 TFLOPS	2TB	64GB	44GB

Evaluation

- Dataset visualization

Pang et al. *Cross-domain adversarial feature learning for sketch re-identification*.

Evaluation

- Baseline algorithm:
  - SphereFace[1]
- Accuracy improvement with online-learning.

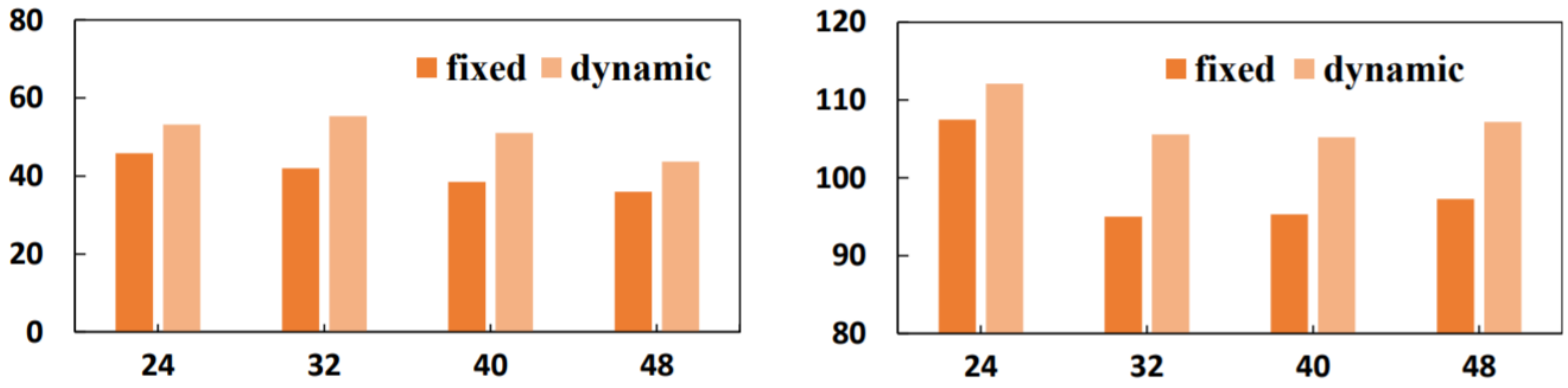
Table 1: Face verification accuracy (%).

Model	*Scenario1*		*Scenario2*	
	Before	After	Before	After
MobileNet	95.70	**96.12**	92.69	**93.51**
Sphere20	96.22	**97.13**	94.71	**96.20**
ResNet50	96.74	**97.33**	95.62	**96.43**

[1] Deng et al. *Arcface: Additive angular margin loss for deep face recognition*.

Evaluation

- Context-aware scheduling VS. Fixed batch size.



(a) Scenario1 (b) Scenario2

Figure 9: The comparison of context-aware scheduling (denoted as dynamic) and the strategy that uses fixed batch size (denoted as fixed). The x-axis is the fixed batch size, while the y-axis represents the throughput (triplets/min)

Evaluation

- Partial Fine-tuning

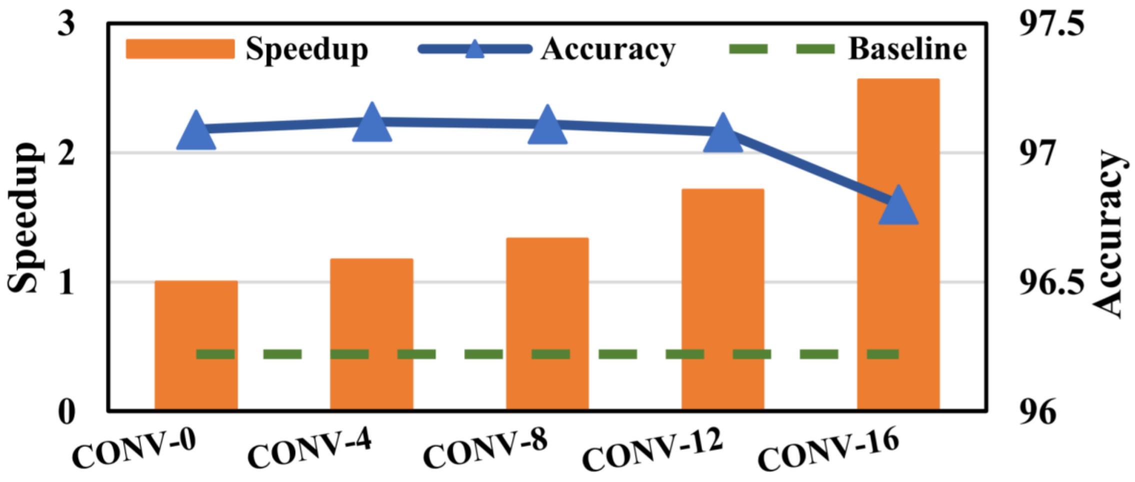


Figure 7: Speed-accuracy trade-off (#Scenario1)

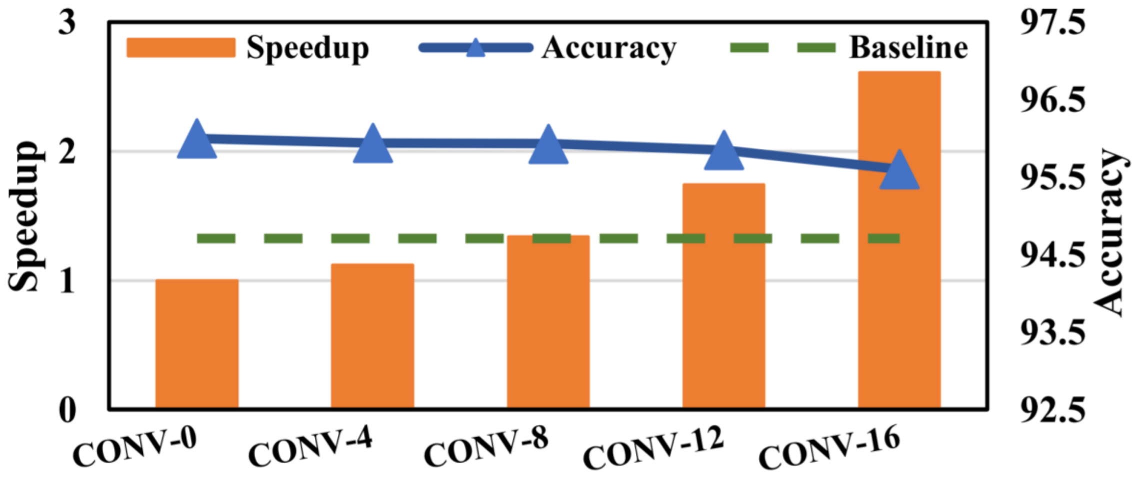


Figure 8: Speed-accuracy trade-off (#Scenario2)

Discussion & Future work

- Generality of SAFACE
  - SAFACE workflow can generalize to many other identification tasks.
- Better Offloading Strategy
  - Offload detection or tracking tasks to edge?
- Different Training Modes
  - Always-on or periodical training?
- Evaluate in More Realistic Scenarios

Thank you!