BTI Institute

Borders • Trade • Immigration

A Department of Homeland Security Center of Excellence

UNIVERSITY of HOUSTON

EDGE: The "Eye in the Woods" Image-based Human Detection and Recognition System

Ioannis A. Kakadiaris
University of Houston

Other personnel: Ha Le, Charles Livermore, Lei Shi, Christos Smailis, Yuhang Wu, Xiang Xu (Ph.D. Candidates); Ioannis Konstantinidis (Staff)

Presentation Goals

- Knowledgeable
- Excited

Project Overview: Knowledge Gap

- Identify border crossers (if there is available information) to enable dispatchers to alert agents in the field

https://www.youtube.com/watch?v=BWJbazGLhH4

Project Overview

End-users of this research

- Enforcement Systems Division: Dispatchers
- DHS strategy officials: Analysts

Main outcomes and benefits end-user desires

- **Pain Points**
  - Limited matching success matching images where a person's face is partially visible due to pose or illumination
  - Lack of filtering of images where movement occurs but there is no human present

- **Gain Desired**
  - A method that will accurately identify individuals in real-time to enable the dispatchers to alert agents in the field

Project Overview: Goal

Design, develop, and evaluate a system to automatically detect and match human subjects from trail camera images.

Project Overview: Objectives (2018- 2019)

1. Design and develop algorithms for matching images in the visible domain

2. Acquire, curate and annotate images by trail cameras (VIS and NIR) in non-urban environments

3. Deploy a private cloud-based software system for software evaluation

Project Overview: Objectives (2019- 2020)

1. Evaluate FR matching on the acquired data
2. Design, develop, and evaluate algorithms to match images in the NIR domain
3. Design, develop, and evaluate algorithms to classify and tag the acquired images (VIS and NIR) by:
   (i) human presence or no human presence, and
   (ii) if human presence is detected, classify **whether or not the image is usable** for face recognition;
4. Design, develop, and evaluate algorithms to extract (for each image tagged to have human presence) the following information:
   (i) direction of movement of the individual or group, and
   (ii) carry-weapon / carry-load / no-carry;

Objective 1: Matching Facial Images

1. Face Detection
2. Illumination Enhancement
3. Occlusion-Aware Matching
4. Compact Pose-Robust Template Generation

Objective 1.1: Face Detection - LR

Two-stage Detectors

- Increased network complexity difficulty of network convergence due to the multiscale training

One-stage Detectors

- Sub-optimal feature selection for the detection task due to
  - noise introduced by choosing simple combinations of different levels of features
  - Use of pre-trained model based on only classification task only instead of the classification and regression tasks needed

- Introduction of gridding due to Dilated Convolutions

- Low computational efficiency because of computationally heavy modules for fusing different levels of features

Objective 1.1: Face Detection - LR

Two-stage Detectors

- Increased network complexity difficulty of network convergence due to the multiscale training

SSFD+

One-stage Detectors

- Sub-optimal feature selection for the detection task due to
  - noise introduced by choosing simple combinations of different levels of features
  - Use of pre-trained model based on only classification task only instead of the classification and regression tasks needed

- Introduction of gridding due to Dilated Convolutions

- Low computational efficiency because of computationally heavy modules for fusing different levels of features

Objective 1.1: Face Detection - LR

Two-stage Detectors

- Increased network complexity difficulty of network convergence due to the multiscale training

SANet

One-stage Detectors

- Sub-optimal feature selection for the detection task due to
  - noise introduced by choosing simple combinations of different levels of features
  - Use of pre-trained model based on only classification task only instead of the classification and regression tasks needed

- Introduction of gridding due to Dilated Convolutions

- Low computational efficiency because of computationally heavy modules for fusing different levels of features

Objective 1.1: Face Detection - LR

Two-stage Detectors

- Increased network complexity difficulty of network convergence due to the multiscale training

SEFD

One-stage Detectors

- Sub-optimal feature selection for the detection task due to
  - noise introduced by choosing simple combinations of different levels of features
  - Use of pre-trained model based on only classification task only instead of the classification and regression tasks needed
- Introduction of gridding due to Dilated Convolutions
- Low computational efficiency because of computationally heavy modules for fusing different levels of features

Face Detection: Quantitative Results

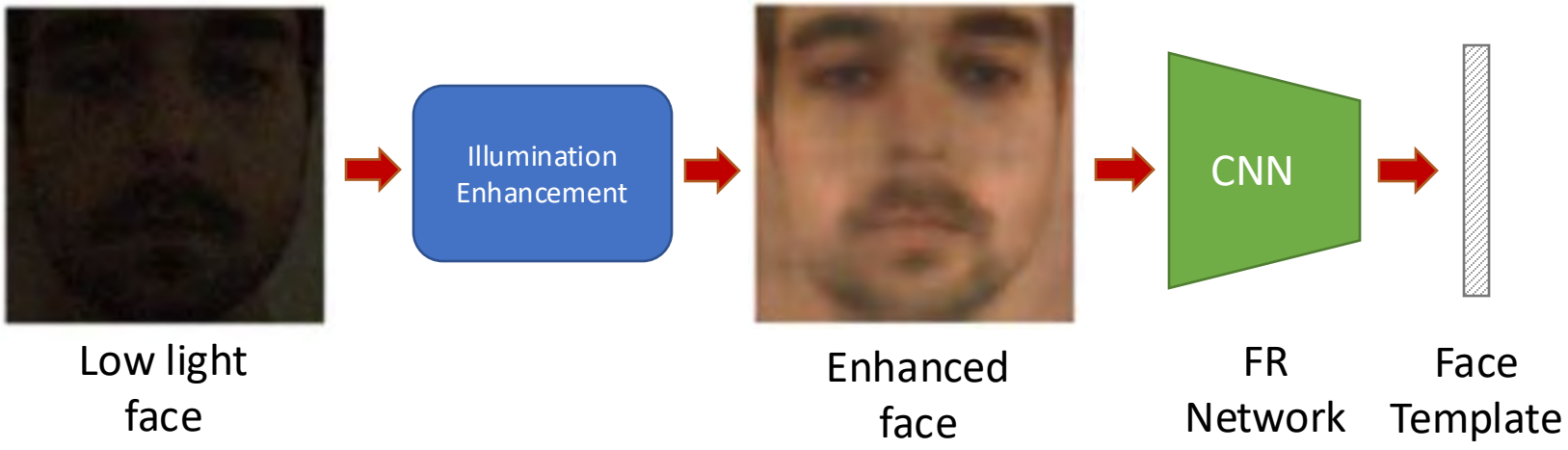
SSFD+ | SANet | SEFD

Objective 1.2: Illumination Enhancement

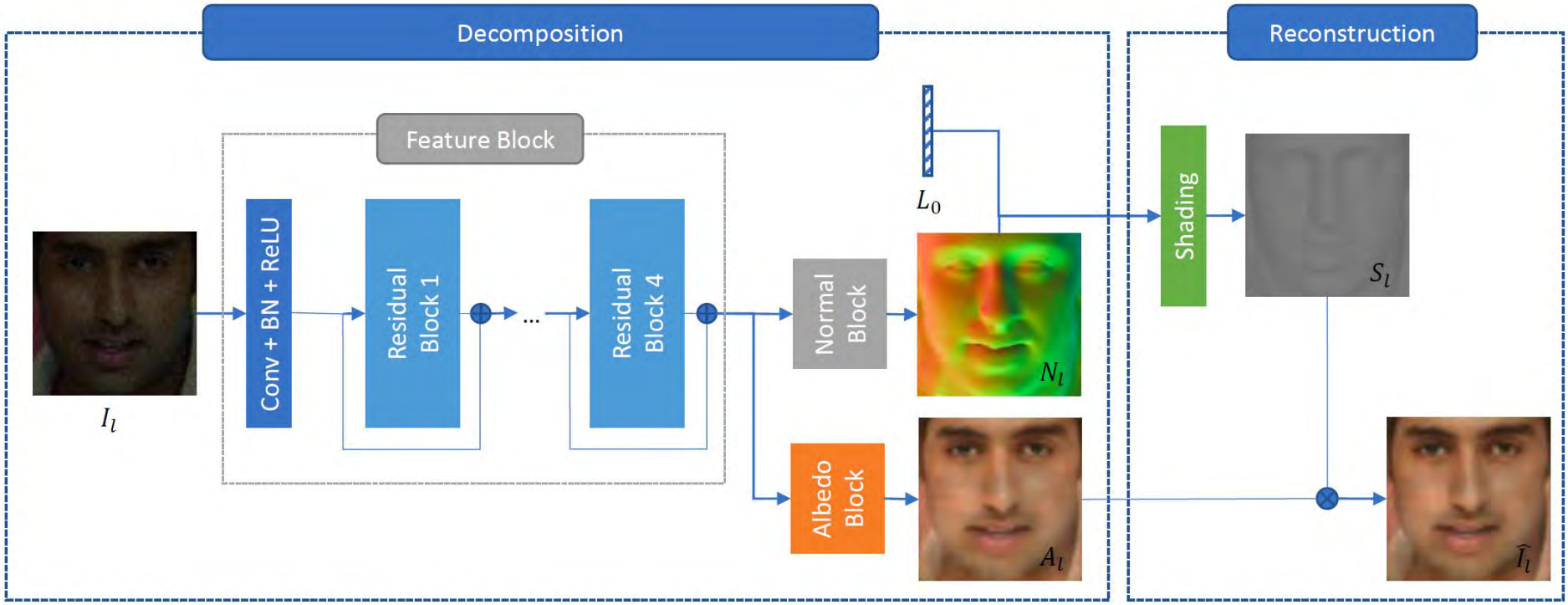
- **SeLENet**: Low light face enhancement for extreme dark face images
- **FRADA**: Illumination-invariant template learning
- **FRnR**: Improves FRADA to be faster and more accurate

SeLENet: Low Light Illumination Enhancement

- Face images acquired in poor lighting conditions suffer reduced recognition performance



SeLENet: Architecture

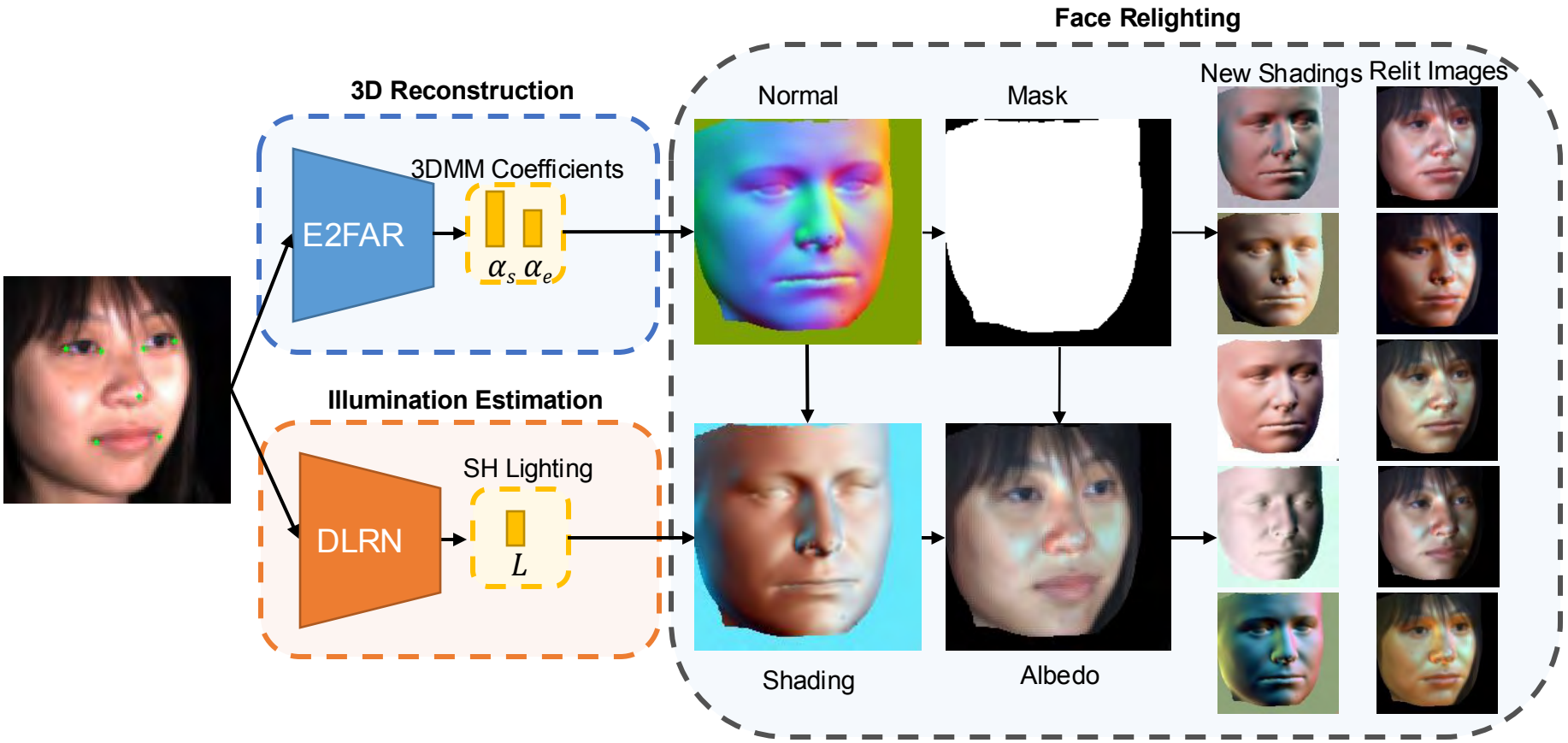


SeLENet: Results

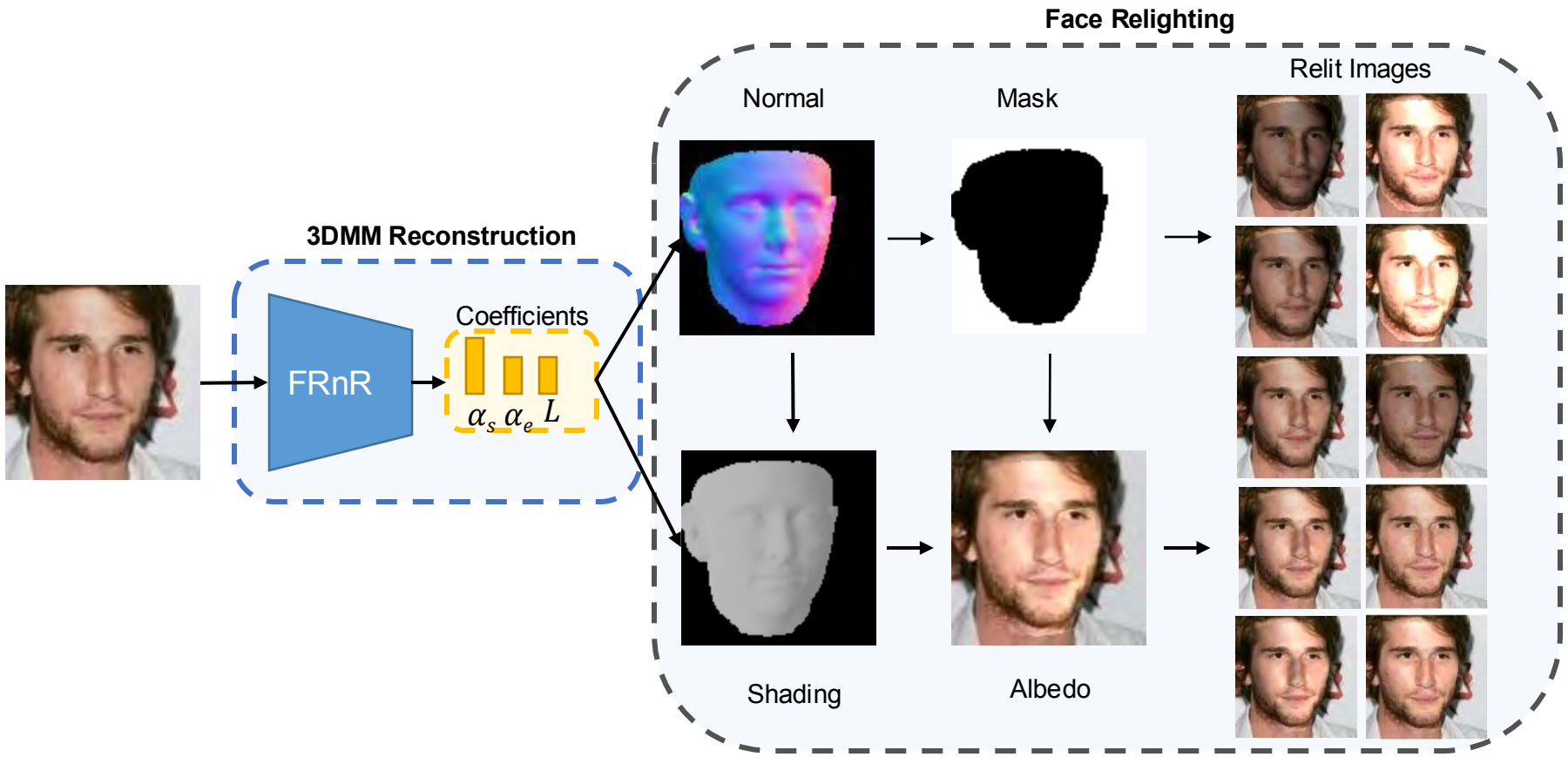
SeLENet: Performance Metrics

Method	CFP-FF										
	Orig.	01	02	03	04	05	06	07	08	09	10
N/A	99.43	96.14	98.00	98.46	98.66	98.94	98.96	99.03	99.10	99.17	99.34
RetinexNet	98.94	86.37	89.86	91.24	91.47	91.27	92.13	93.16	95.43	96.44	98.07
Li's	99.00	95.60	97.69	98.16	98.34	98.30	98.43	98.53	98.60	98.46	98.81
SeLENet-	97.21	96.09	96.74	96.94	97.24	97.39	97.57	97.66	97.50	97.61	97.6
SeLENet	**99.47**	**99.17**	**99.49**	**99.47**	**99.57**	**99.56**	**99.63**	**99.59**	**99.60**	**99.64**	**99.61**

FRADA


Face Relighting
3D Reconstruction
3DMM Coefficients
E2FAR
α_s α_e
Illumination Estimation
SH Lighting
DLRN
L
Normal
Mask
New Shadings
Relit Images
Shading
Albedo

FRnR

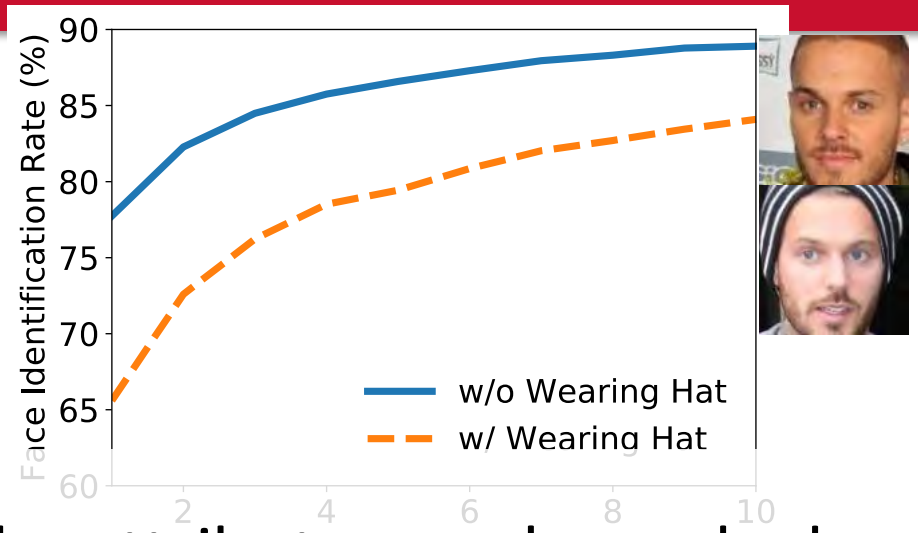

Face Relighting
Relit Images
Normal
Mask
3DMM Reconstruction
FRnR
Coefficients
α_s α_e L
Shading
Albedo

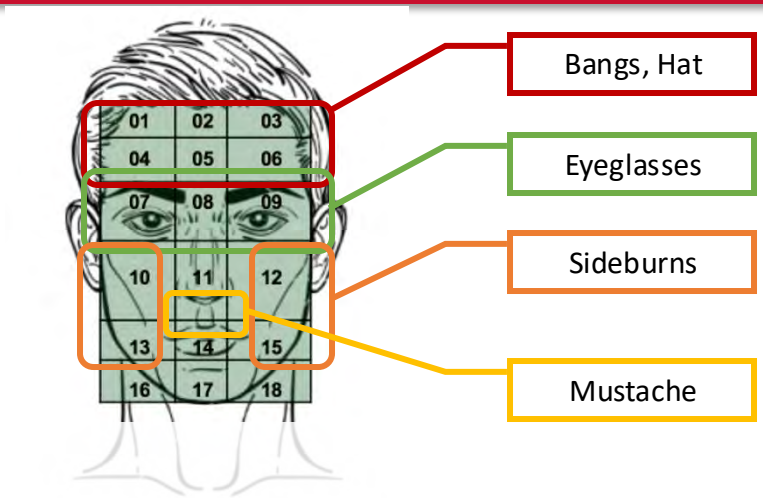
FRnR: Quantitative Results

Method	UHDB31.R0128				UHDB31.R0256			
	I01	I03	I05	Mean	I01	I03	I05	Mean
ArcFace	87.60	90.32	87.51	88.48	88.38	90.64	88.46	89.16
ArcFace-CA	87.81	90.96	88.14	88.97	88.10	91.91	88.74	89.58
ArcFace-FRADA	89.09	**92.55**	88.07	89.90	**90.61**	92.04	89.37	90.67
ArcFace-FRnR	**89.17**	92.04	**89.54**	**90.25**	90.47	**92.23**	**91.17**	**91.29**

FRnR improved ArcFace in all set of illuminations.

Objective 1.3: Occlusion-Aware Matching

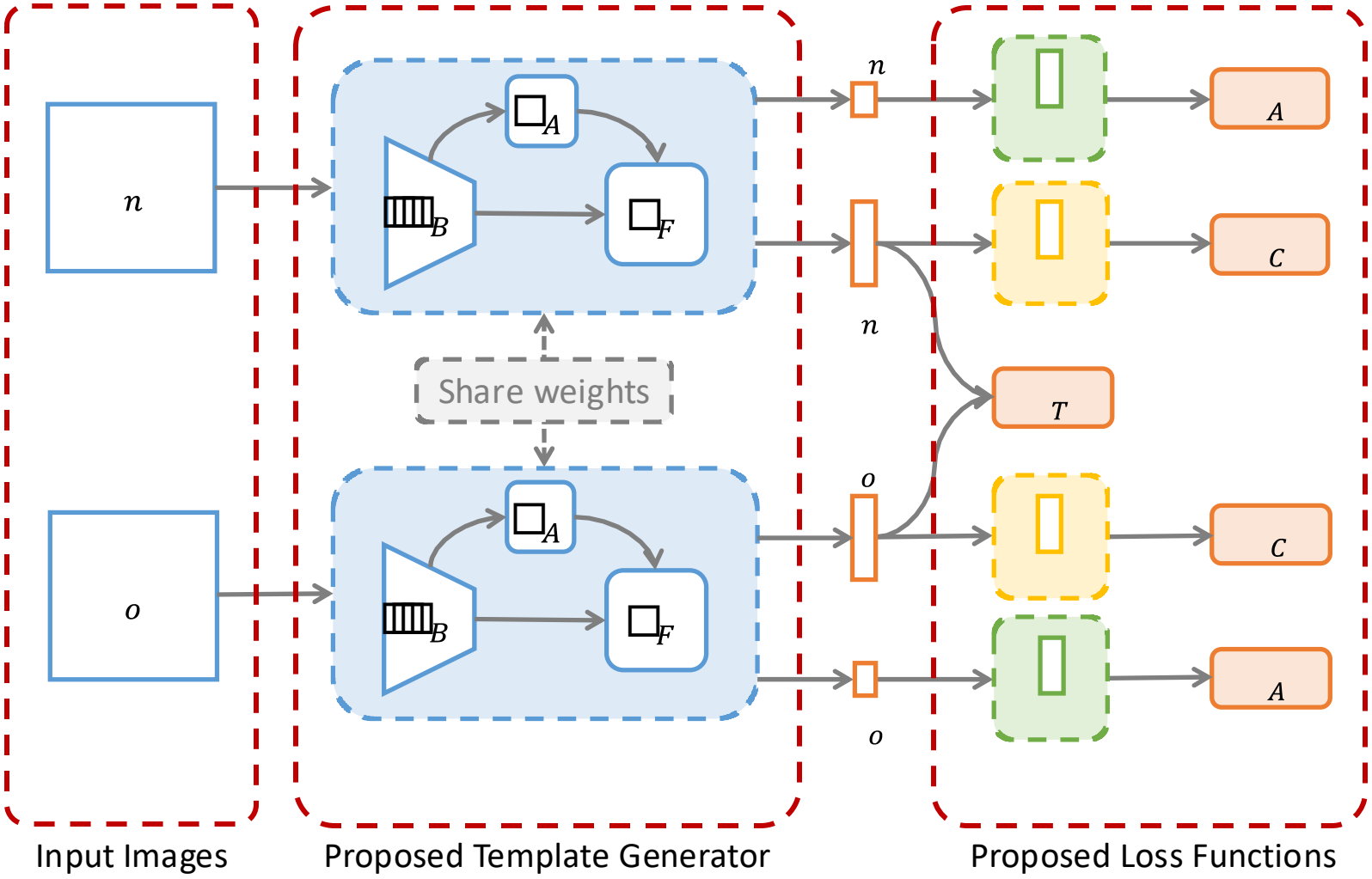




The attributes can be ranked according to the rank-1 identification rate degradation as follows: **Eyeglasses** (18%) > **Wearing Hat** (12%) > **Bangs** (7%) > **Sideburns** (3%) > **Mustache** (2%).

- Challenges
  - Identity information degradation
  - Occlusion imbalance

OREO: Overview

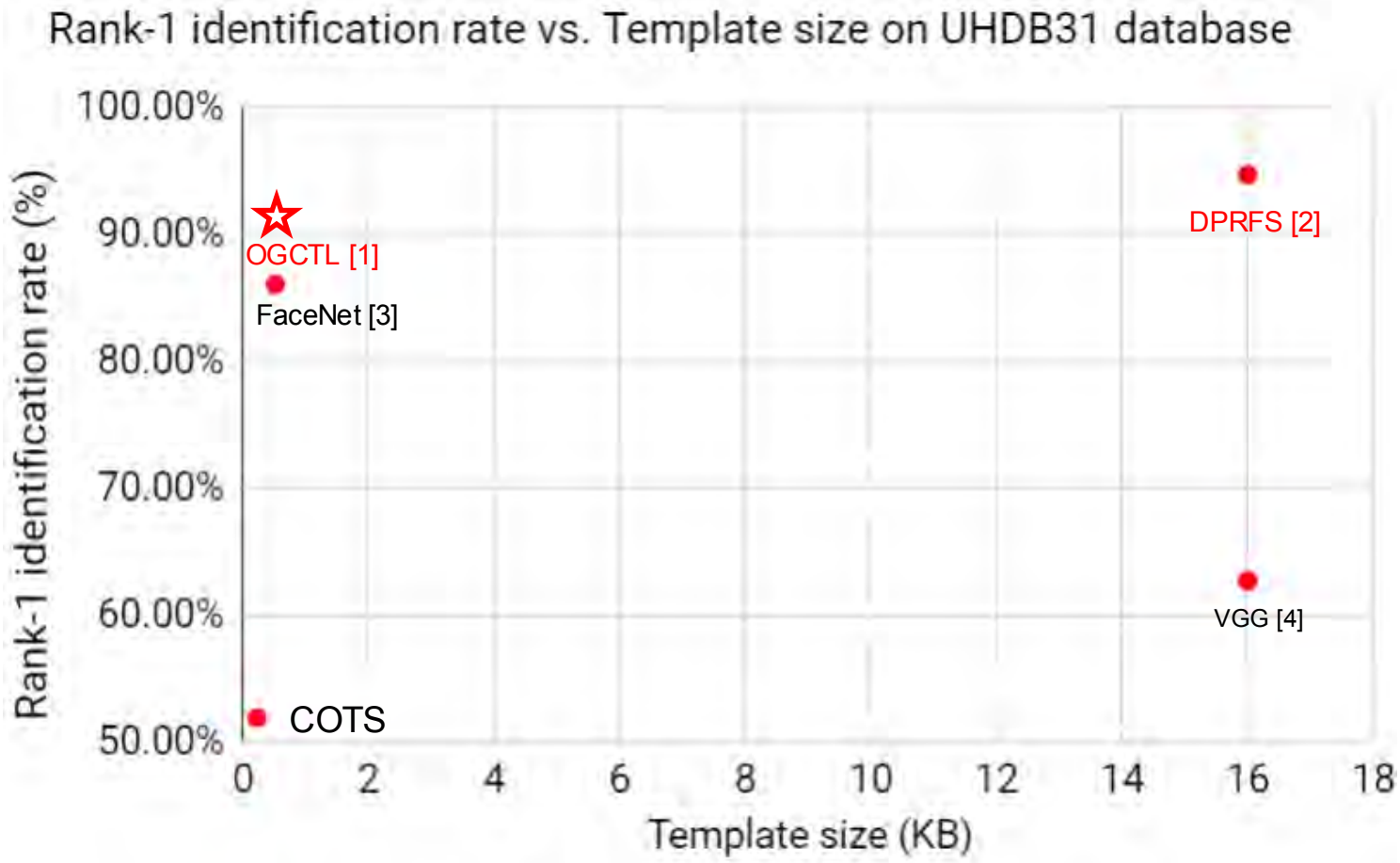


Objective 1.3: Performance metrics

Method	1: N Identification					
	TPIR (%) @ FPIR			Retrieval Rate (%)		
	10^{-3}	10^{-2}	10^{-1}	Rank-1	Rank-5	Rank-10
GOTS	2.66	5.78	15.60	37.85	52.50	60.24
FaceNet	20.58	32.40	50.98	69.22	79.00	81.36
VGGFace	26.18	45.06	62.75	78.60	86.00	89.20
ArcFace	**70.90**	**81.98**	**87.63**	92.25	94.31	95.30
ResNeXt-101	53.66	71.50	82.47	91.88	95.51	97.29
OREO	65.47	77.11	85.92	**93.76**	**96.68**	**97.74**

Comparison of the face verification and identification performance of different methods on the IJB-C dataset using 1: 1 identification protocol

Objective 1. 4: Performance vs template size on UHDB31



References

[1] Y. Wu, I.A. Kakadiaris, "Occlusion-guided compact template learning for ensemble deep network- based pose invariant face recognition," In Proc. IAPR international conference on biometrics, Crete, Greece, June 4-7, pp. 1-8, 2019.

[2] X, Xiang, H. A. Le, P. Dou, Y. Wu, and I. A. Kakadiaris. "Evaluation of a 3D-aided pose invariant 2D face recognition system." In Proc. IEEE International Joint Conference on Biometrics, Denver, Colorado pp. 446-455. 2017.

[3] S. Florian, D. Kalenichenko, and J. Philbin. "Facenet: A unified embedding for face recognition and clustering." In Proc. IEEE conference on computer vision and pattern recognition, pp. 815-823, Boston, MA, June 7-12, 2015.

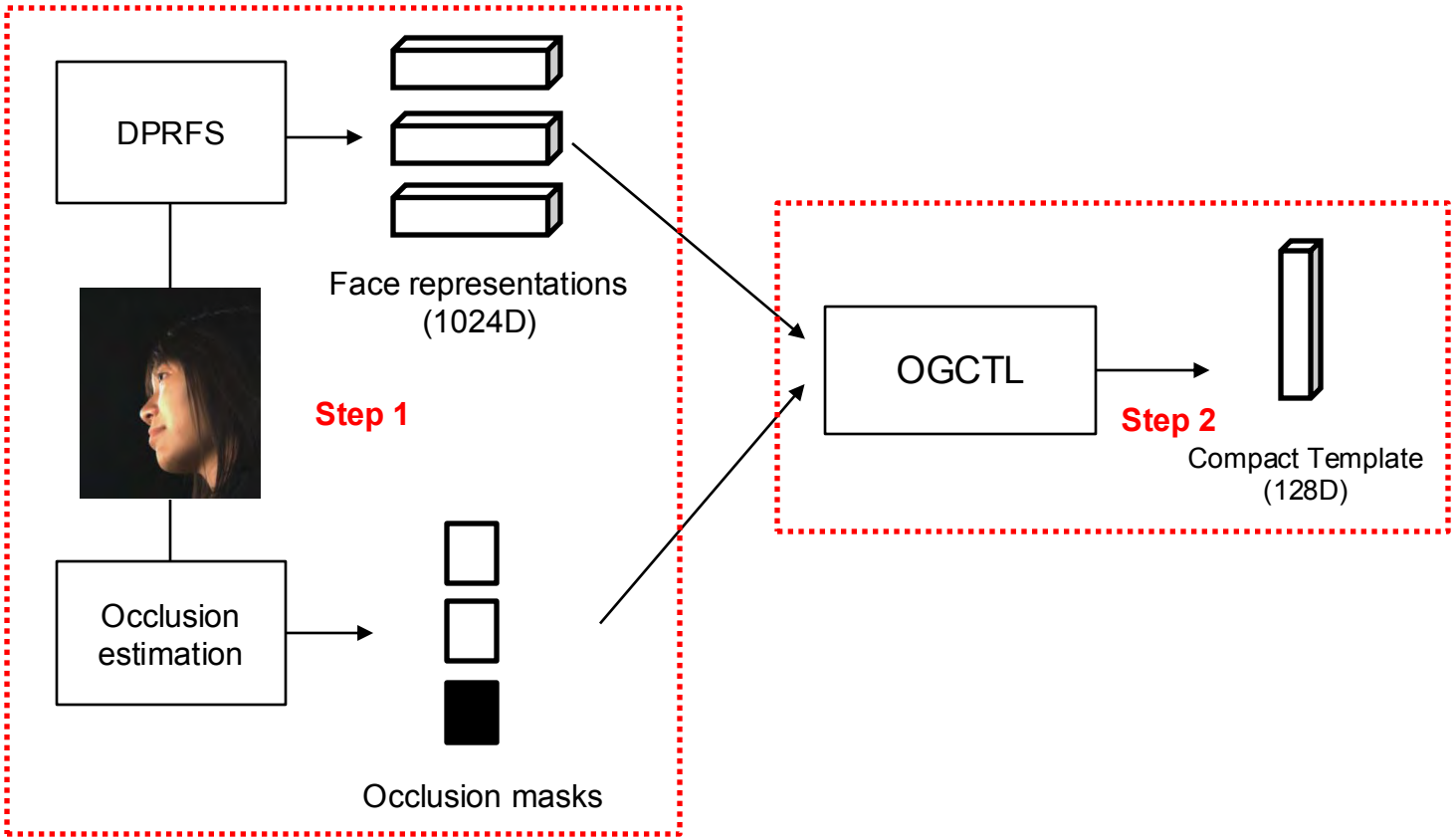
[4] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In Proc. British Machine Vision Conference, pp. 1–12, Swansea, UK, September 7-10 2015

Objective 1.4: Compact Pose-Robust Template Generation

Accomplishments

- A complete pipeline for extract pose-robust facial template (4096D) from VIS image
- A module to extract more compact facial template (128D) from the existing template (4096D)
- Faces under visible light condition can be described with a 4096D or 128D facial template

OGCTL: Occlusion-guided Compact Template Learning



Objective 1.4: Performance Metrics

- Face verification performance on IARPA Janus IJB-C database

TAR (True Accept Rate %) against the FAR (False positive rate %)

	Tem. Size	FAR=10^{-5}	FAR=10^{-4}	FAR=10^{-3}	FAR=10^{-2}	FAR=10^{-1}
COTS-1	-	0.090	0.160	0.320	0.620	0.800
FaceNet [1]	**0.5KB**	0.330	0.490	0.660	0.820	0.920
VGG-CNN [2]	16KB	0.430	0.600	0.750	0.860	0.950
DPRFS [3]	16KB	0.310	0.461	0.638	0.807	0.939
OGCTL	**0.5KB**	**0.608**	**0.737**	**0.839**	**0.918**	**0.975**

- Face matching speed

  Without compact template learning (DPRFS): 15K faces per second

  With compact template learning (OGCTL): 1M faces per second

[1] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. *Proc. Computer Vision and Pattern Recognition*, pages 815-823, June 2015.

[2] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *Proc. British Machine Vision Conference*, pages 1-12, Swansea, UK, September 7-10 2015.

[3] X. Xu, H. Le, P. Dou, Y. Wu, and I. A. Kakadiaris. Evaluation of 3D-aided pose invariant 2D face recognition system. In *Proc. International Joint Conference on Biometrics*, pages 446-455, Denver, Colorado, October 2017.

Objective 1: Performance Metrics

- 2018 – 2019
  - within schedule
  - within budget
- 2019 – 2020
  - Rank-1 > COTS Rank-1

Objective 2: Data Acquisition & Curation

1. Data Acquisition

2. Annotation Tool

3. Curation & Inter-observer Agreement

Objective 2.1: Data Acquisition

Day time at Lynn Eusan Park

Night time at Lynn Eusan Park

Objective 2.1: Data Acquisition (2)

- **Number of subjects**

Day Session	Night Session
197	67

- **Number of images**

Type	Cam 1	Cam 2	Cam 3	Cam 4
Day - VIS	685	620	719	673
Night - NIR	187	181	218	229

- **Hardware**
  - Four trail cameras
  - Four light meters

Objective 2.1: VIS Images

Camera 1

Camera 2

Camera 3

Camera 4

Objective 2.1: NIR Images

Camera 1

Camera 2

Camera 3

Camera 4

Objective 2.2: Data Annotation Tool

Designed, developed, and evaluated a GUI to perform annotation of VIS and NIR images

- **Cloud based deployment** (AWS EC2)
- **Multiple user accounts**, image and annotations upload, management of user gallery
- **Manual annotation** of human, object and facial bounding boxes as well as pose and facial landmarks
- **Integration of automated** person detection and pose estimation results to speed up the annotation process (AlphaPose Algorithm)
- **Action & Attribute Labels** (Default: Carrying Tube, Carrying Backpack, Wearing Hat, Wearing Woolen Cap)
- **Customizable** Class Labels and Pose Key point formats

Objective 2.2: Data Annotation Tool

Objective 2.3: Inter-observer Agreement

- Two Human Annotators refined the Person & Face Detection as well as Pose and Facial Key-point Proposals or provided annotations when the tool failed.

	InstaANT	Annotator 1	Annotator 2
AlphaPose	-	0.99	0.99
Annotator 1	0.99	-	0.99
Annotator 2	0.99	0.99	-

Table 1: Pearson's Correlation Coefficient for mean Bounding Box Area between Alpha Pose and two Human Annotators

	InstaANT	Annotator 1	Annotator 2
AlphaPose	-	0.99	0.99
Annotator 1	0.99	-	0.99
Annotator 2	0.99	0.99	-

Table 2: Pearson's Correlation Coefficient for mean Keypoint Euclidian distance from the origin of the Person's BBox, between Alpha Pose and two Human Annotators

Objective 2: Performance Metric

- ✓Achieve >70% inter-annotator agreement for image annotation

Objective 2: Deliverables

- Current status
  - Image Annotation Tool delivered within budget and within schedule
  - Curated data on schedule and within budget to be delivered by June 30, 2019

- Possible problems
  - N/A

Objective 3: Cloud Infrastructure

1. Setup & Deployment

2. Training

Objective 3: Setup & Deployment

- Jupyter notebooks for demo
  - AWS Rekognition demo with S3
  - 2 EDGE Pipeline notebooks

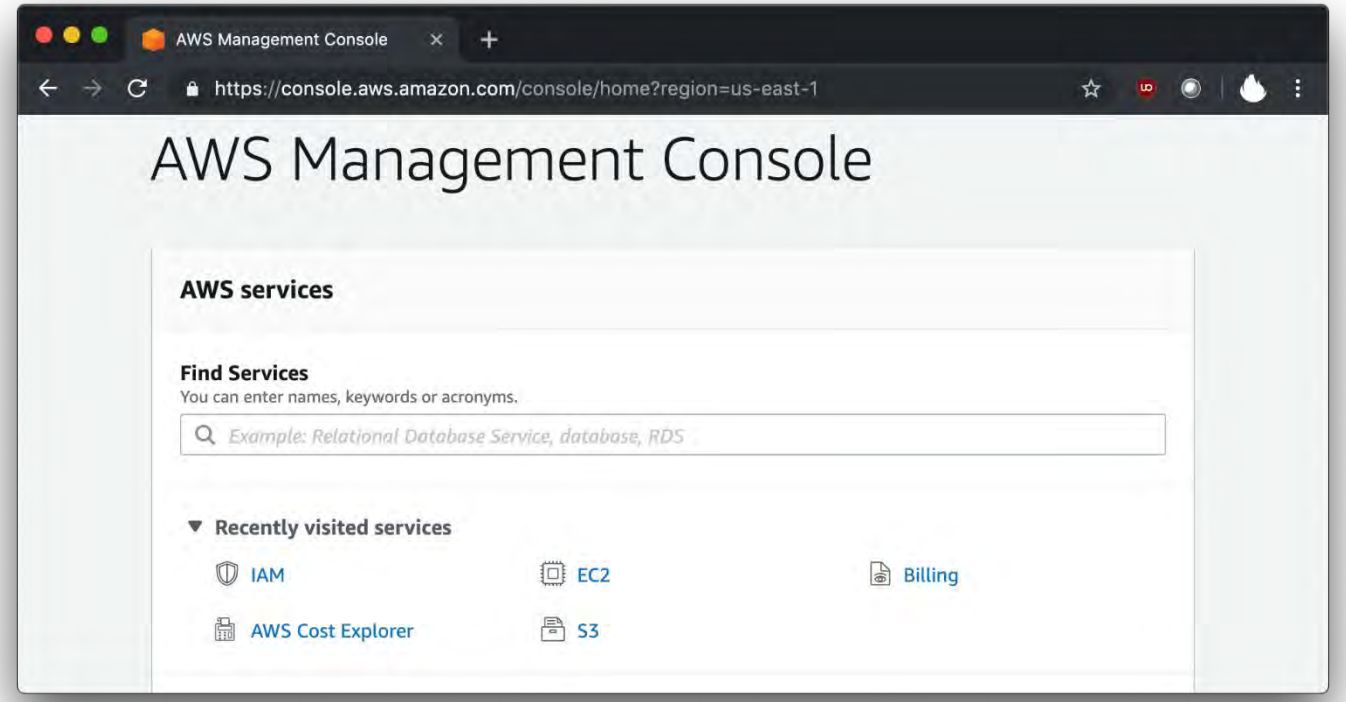
- EDGE facial recognition code for Jupyter notebooks also deployed on EC2 instance for direct inspection and experimentation.

Objective 3: Training

- Created AWS demo account and authentication tokens for access to EC2 instance and programmatic access to AWS resources
- Provided access tokens to specified individuals
- Demo access included:
  - Investigation of cloud resources
  - Use of early EDGE pipeline code
  - Access to AWS Rekognition

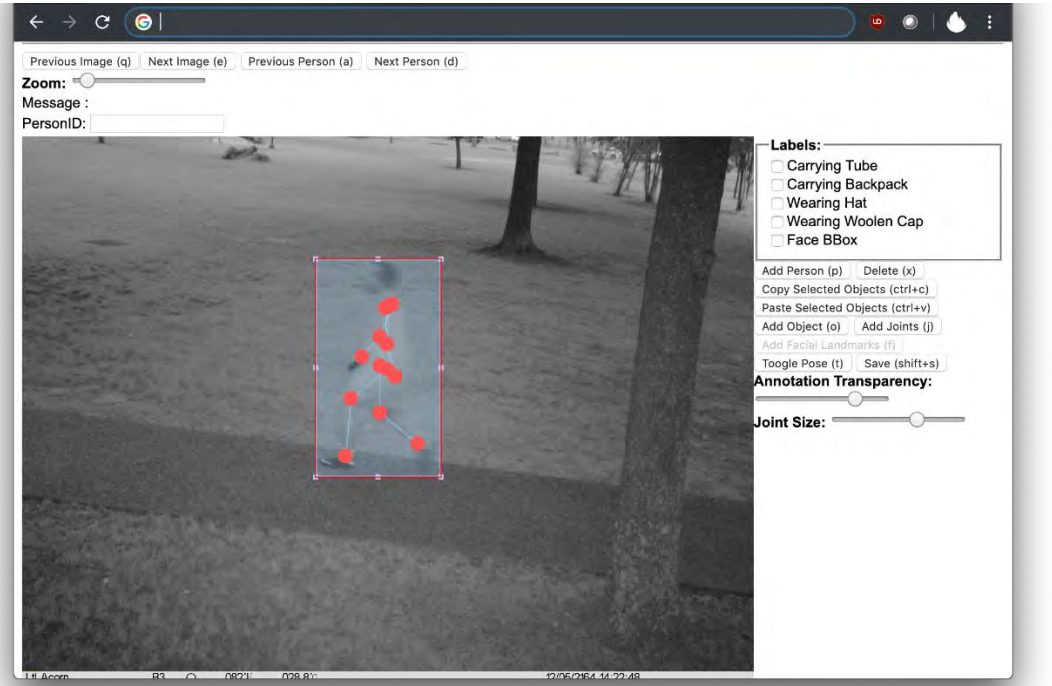
Objective 3: Deliverables – FR Code

- Interactive exploration
- Demonstration video
- Report



Objective 3: Deliverables – Annotation Tool

- Interactive exploration
- Demonstration video
- Report



Objective 3: Deliverables

- Current status
  - Private cloud infrastructure delivered within budget and within schedule
  - Data and Annotation tool to be delivered by 6/30

- Possible problems
  - N/A

Publications

1. C. Smailis, Ha Le, I.A. Kakadiaris, “InstANT: Image ANnotation Tool and a database for VIS-NIR Face and Carrying Action Recognition,” in Proc. International Joint Conference on Biometrics, Houston, USA, September 27-30, 2020. (In Preparation)
2. Y. Wu, I.A. Kakadiaris, “Occlusion-guided compact template learning for ensemble deep network- based pose invariant face recognition,” In Proc. IAPR international conference on biometrics, Crete, Greece, pp. 1-8, June 4-7, 2019.
3. Y. Wu, I.A. Kakadiaris. " Three-Dimensional Face Alignment Using A Convolutional Point-Set Representation”, IEEE Transactions on Biometrics, Behavior, and Identity Science, 2019. (Under Review)
4. H. Le and I. A. Kakadiaris. Illumination-invariant Face Recognition with Deep Relit Face Images. In Proc. Winter Conference on Applications of Computer Vision, Waikoloa Village, Hawaii, Jan. 2019.
5. X. Xu, H. Le, and I. A. Kakadiaris. On the Importance of Feature Aggregation for Face Reconstruction. In Proc. Winter Conference on Applications of Computer Vision, Waikoloa Village, Hawaii, Jan. 2019.

Publications (2)

6. X. Xu, N. Sarafianos, and I. A. Kakadiaris. On Improving the Generalization of Face Recognition in the Presence of Occlusions, In Proc. IEEE International Conference on Computer Vision, Seoul, Korea, Oct. 27 - Nov. 2, 2019 (Under Review).
7. X. Xu and I.A. Kakadiaris. Open Source Face Recognition Performance Evaluation Package, In Proc. International Conference on Image Processing, Taipei, Taiwan, Sept. 22-25, 2019
8. L. Shi, X. Xu, and I. A. Kakadiaris, Smoothed Attention Network for Single Stage Face Detector, In Proc. International Conference on Biometrics, Crete, Greece, Jun. 4-7, 2019.
9. L. Shi, X. Xu, and I. A. Kakadiaris, A Simple and Effective Single Stage Face Detector, In Proc. International Conference on Biometrics, Crete, Greece, Jun. 4-7, 2019.
10. X. Xu, H. Le, and I. A. Kakadiaris. On the Importance of Feature Aggregation for Face Reconstruction, In Proc. Winter Conference on Applications of Computer Vision, Waikoloa Village, Hawaii, Jan. 7 - 11, 2019.

Publications (3)

11. H. Le and I. A. Kakadiaris. Improved face reconstruction and relighting for illumination-invariant face recognition. IEEE Transaction on Biometrics, Behavior, and Identity Science (In preparation).
12. H. Le and I. A. Kakadiaris. Semi-supervised low light face enhancement for mobile face unlock. In Proc. International Conference On Biometrics, Crete, Greece, Jun. 2019.
13. L. Shi, X. Xu, and I. A. Kakadiaris, SSFD: A Face Detector using a Single-scale Feature Map, In Proc. The IEEE International Conference on Biometrics: Theory, Applications, and Systems, Los Angeles, CA, Oct. 22-25, 2018.
14. L. Shi, X. Xu and I. A. Kakadiaris. “SSFD+: A Two-Stage Robust Face Detector.” IEEE Transactions on Biometrics, Behavior, and Identity Science. (In Press)

Project Overview: Objectives (2018- 2019)

- ✓ Design and develop algorithms for matching images in the visible domain
- ✓ Acquire, curate and annotate images by trail cameras (VIS and NIR) in non-urban environments
- ✓ Deploy a private cloud-based software system for software evaluation

Additional Opportunities

- CBP Data Breach
  - Develop tools to help investigators locate dark web marketplaces
- Face Recognition
  - for minors
  - for vehicle biometric exit
  - predict face using image of hands and speech
- Video analytics
  - for drone footage
- Artificial intelligence
  - predicting adverse events
  - using Immigration data

BTI Institute

Borders • Trade • Immigration

A Department of Homeland Security Center of Excellence

Questions and Answers

UNIVERSITY of HOUSTON