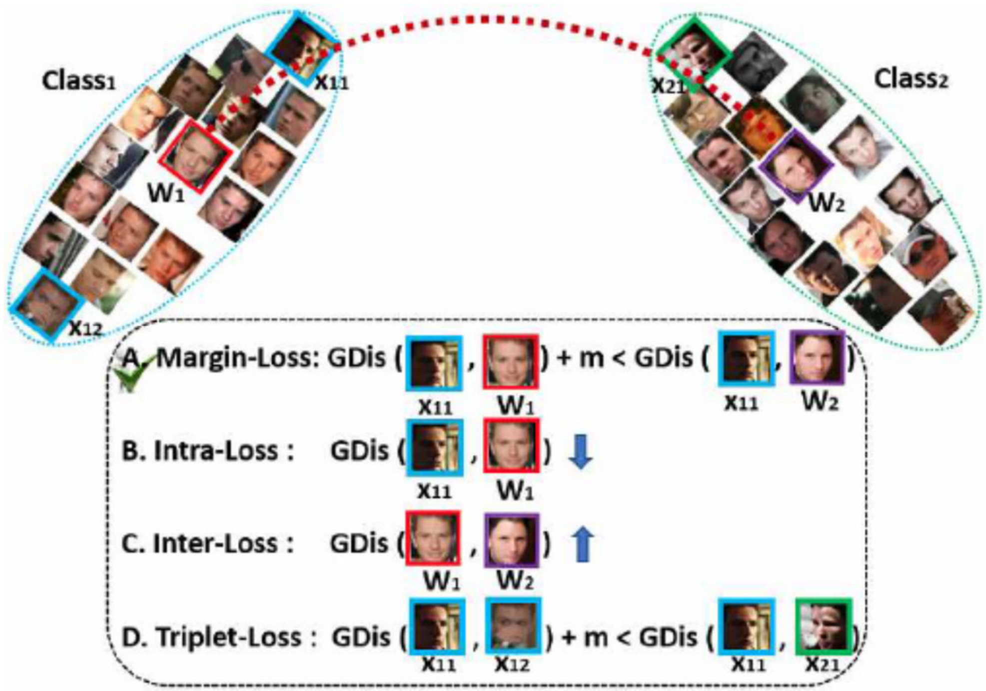
WSB 15-Jan-2020

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Biometrics in Surveillance Videos

Brian Lovell



ADVANCED SURVEILLANCE

Preface

- I have been working on reliable face recognition and detection from surveillance and mobile videos for 20 years
- During this time face recognition has advanced tremendously in terms of performance and some say it is now the Golden Age of Face Recognition
- Apple’s iPhone X and later phones have demonstrated to the public the sheer convenience of face recognition based authentication; on the other hand deployments of public facial recognition systems are currently raising enormous concerns worldwide.

Some Applications for Face Recognition

- Border Control
  - Cooperative and Strongly Controlled Enrolment and Capture
- Mobile Access Control and Banking (Apple iPhone X)
  - Cooperative and Weakly-Controlled Enrolment and Capture
- Chokepoint Face Recognition (crowd surveillance)
  - Non-Cooperative and Weakly-Controlled Enrolment and Capture
- All of these SOTA applications are CNN powered and give great benchmark performance
- It seems like everything is done - what’s left for us to research?

Face Recognition for Border Control

Face Recognition for Border Control

Cooperative Facial Verification

Airport smart gates, border control, access control

•Known reference image – e.g. passport photo
•Very high resolution
•Perfect artificial lighting
•Multiple high quality cameras or single height adjustable
•No movement, no glasses, no expression allowed
•One person at a time
•Photo based not video based
•Cooperative Subject – the subject wants to be recognised
•One-to-one match – verification only, not one-to-many recognition

Many Commercial Solutions available, fully tested by NIST

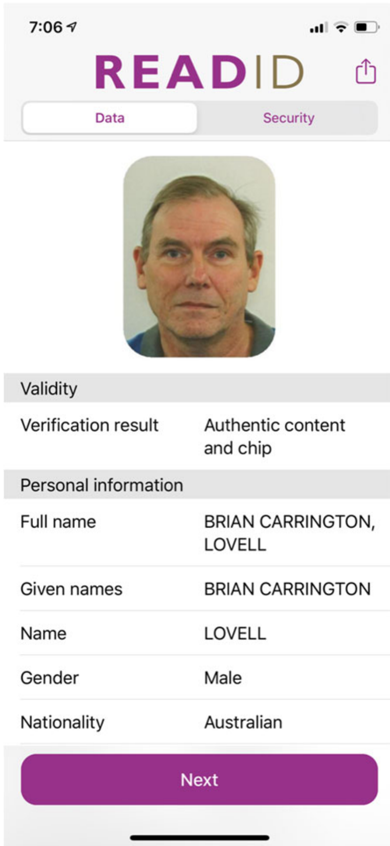
Australia was first in the World with Face for Border Control
Rollout in 2007 at BNE Airport

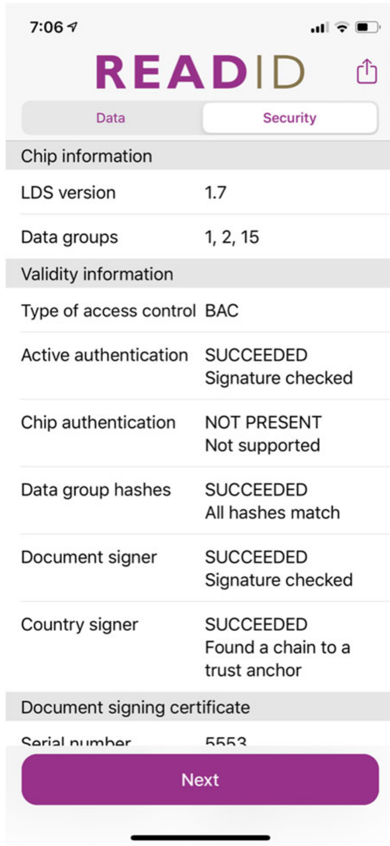
SmartGate

- Are these two faces the same person?
- Primarily used for passenger facilitation not security
- Now used for Australian Departures as well
- Similar Tech is in use in UK, NZ etc

Australian Customs and Border Control is now working on Digital Passports, so passengers can cross national borders without any paperwork. (Initiative Announced at ICB2018)

Photo and Passport Information is Stored in the RFID Chip





Apple only allowed this access in IOS 13 released late 2019 at the request of UK government.

Apple is at it again

- Way back in 2015, Apple Vice President Eddy Cue told us that replacing passports was one of the company's ambitions. Governments are already exploring use of Apple's devices to replace driving licenses.
- Apple in 2018 enabled use of its devices as digital ID at student campuses across the U.S.. This use of device as ID may also provide Apple with real usage data to help prove its systems work and can be trusted to do so — even by governments.

https://www.computerworld.com/article/3388300/your-iphone-will-be-your-passport.html

Face Recognition for Mobile Devices

iPhone X 2D and 3D Face Recognition

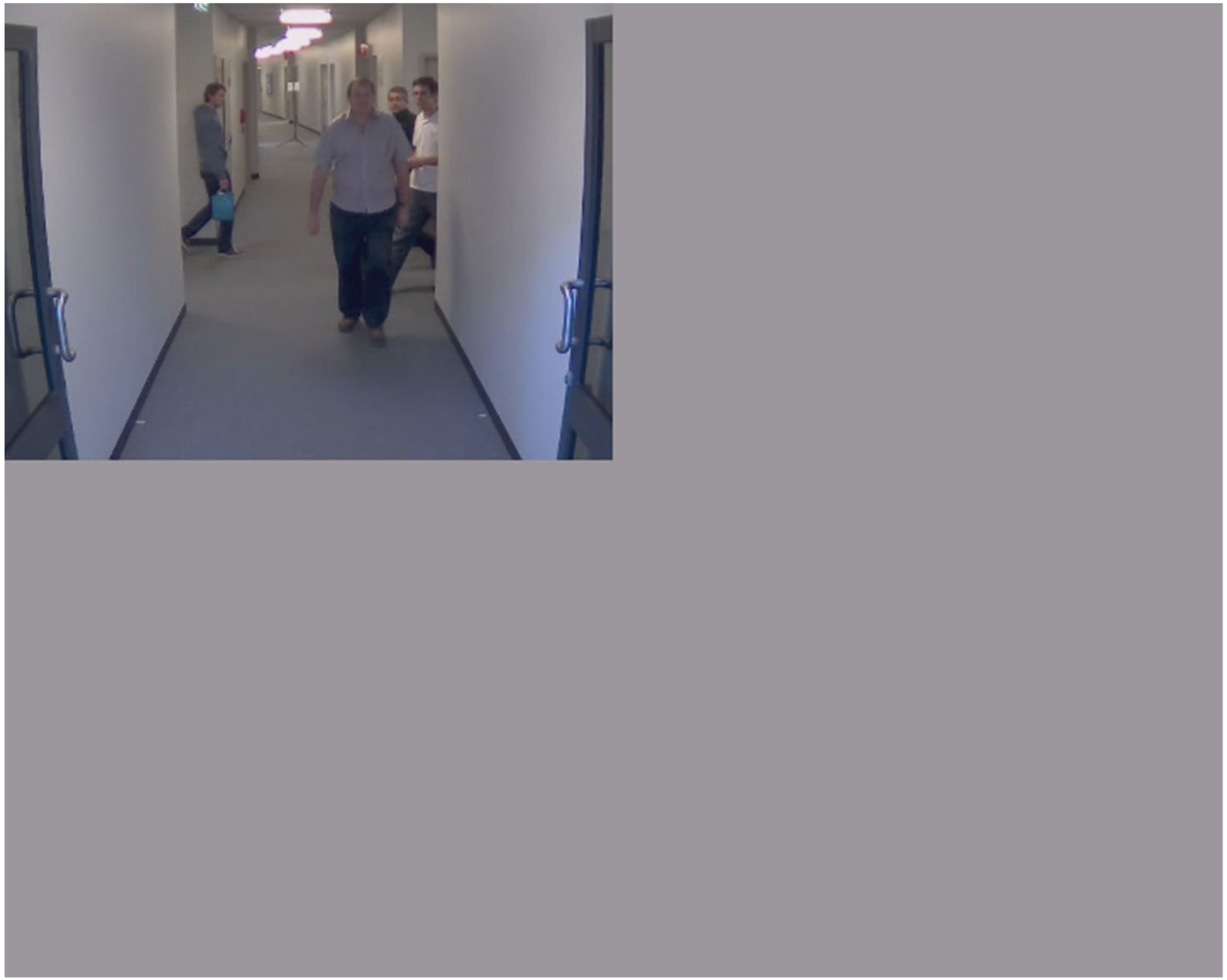
- Time of flight proximity sensor
- powers up other sensors
- IR dot projector
- IR Flood Illuminator
- IR camera
- Works at night with IR illumination

3D is mostly for anti spoofing not recognition accuracy.

Anyone know of a practical 2D Anti-spoof technique?

Chokepoint Face Recognition from Video

2011: Chokepoint Identification

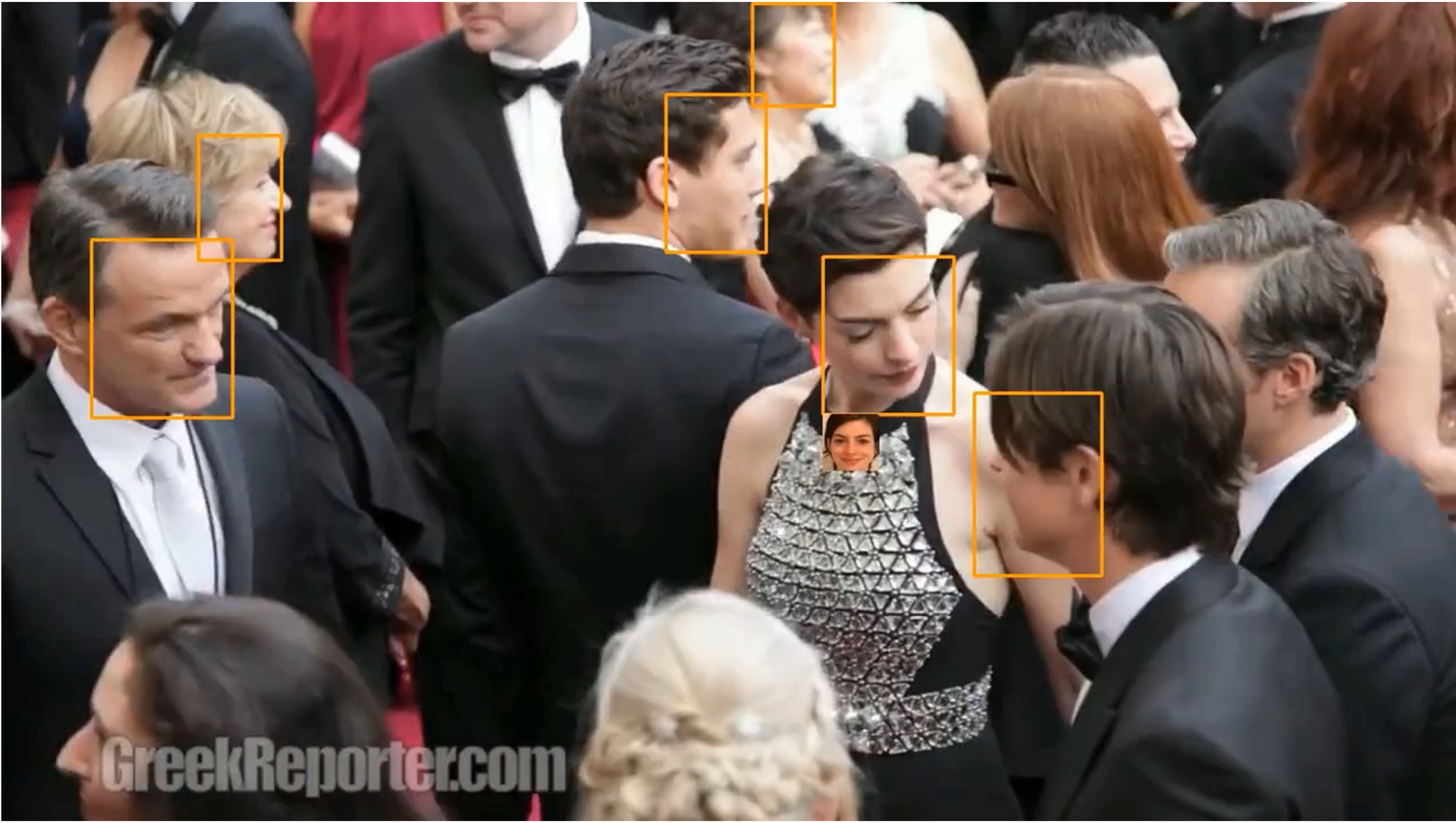


Notes

- Detection is Viola-Jones Cascade based (Pittpatt)
- Recognition is Bag of Words based
- No CNNs
- Multiprocessed using GPUs and Robot Operating System (ROS)
- We won the IFSEC Major Category of **CCTV System of the Year** for **Face Recognition in a Crowd in 2011** in Birmingham
- Chokepoint simulates persons walking down an Aerobridge and was intended to address the Undocumented Passenger Problem.
- Chokepoint Dataset released to community.

https://zenodo.org/record/815657#.XhP2nPxS-Uk

So what are we working on now?


GreekReporter.com


0.8533665
GreekReporter.com
0.8198014
GreekReporter.com
0.90810394
GreekReporter.com
0.66789776
GreekReporter.com

Faces in a Milling Crowd

- The problem with CCTV face recognition in many common situations is that people simply do not look at the camera, but we would still like to identify them.
- The Chokepoint scenario addresses this issue because people tend to look straight ahead when walking in a crowd
- This assumption applies to aerobridges, borders, concierge situations, but not to cocktail parties, conferences, shopping centres, check in areas.
- We would like to have much better performance under common non-cooperative conditions where people do not look at the camera.

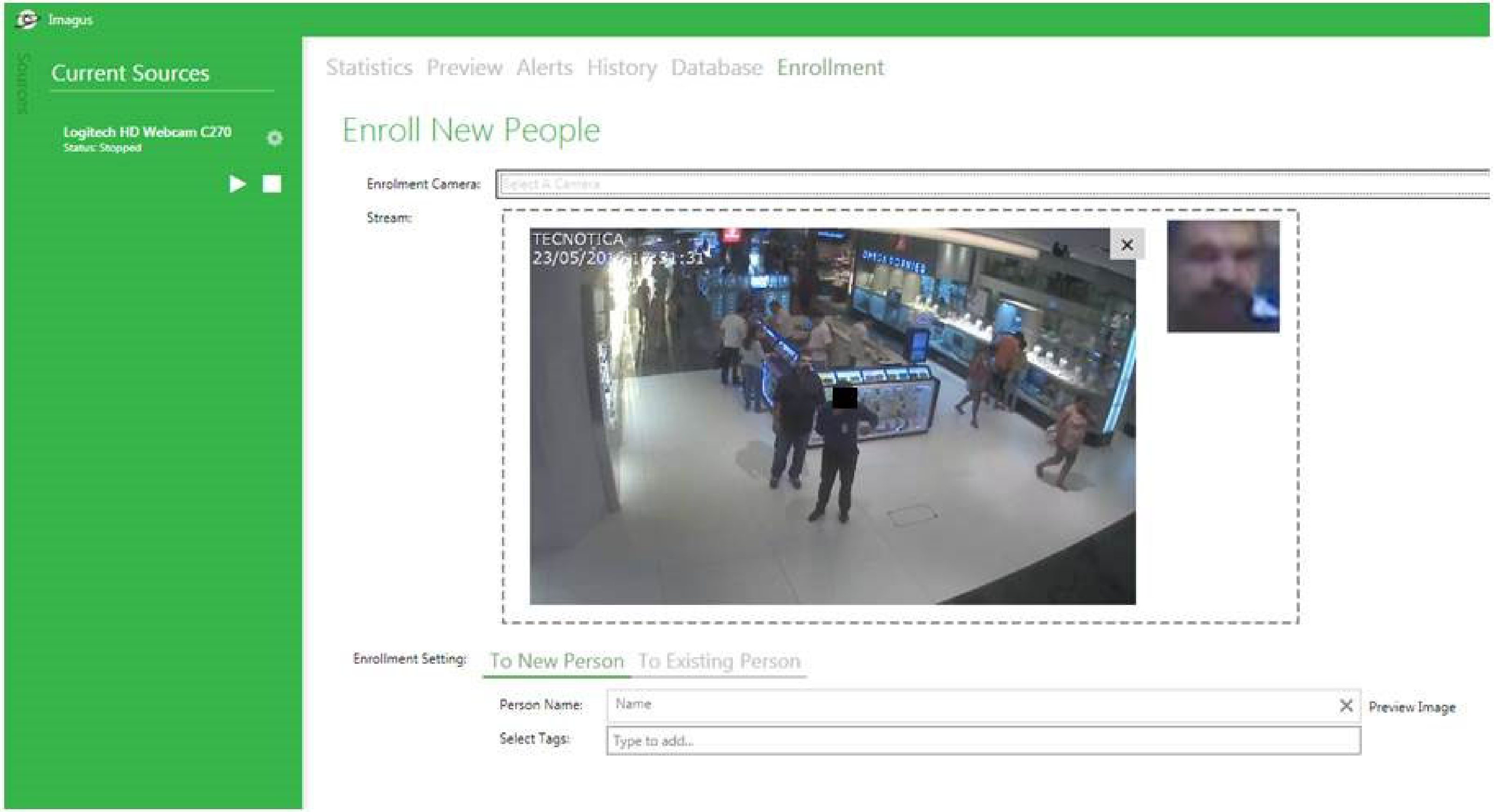
The Practical Problem In a Nutshell

- Due to computational requirements, face recognition from surveillance video has mainly used the Viola-Jones Cascade Face Detector on the Front End
- If we want to recognize faces in video at extreme angles, we must use CNN based detectors which are much slower and cannot easily handle the huge camera resolutions (5Mp or more) and multiple streams
- All decoding and detection must take place in GPU

Case Studies - Face Harvesting in Shopping Malls and Clubs

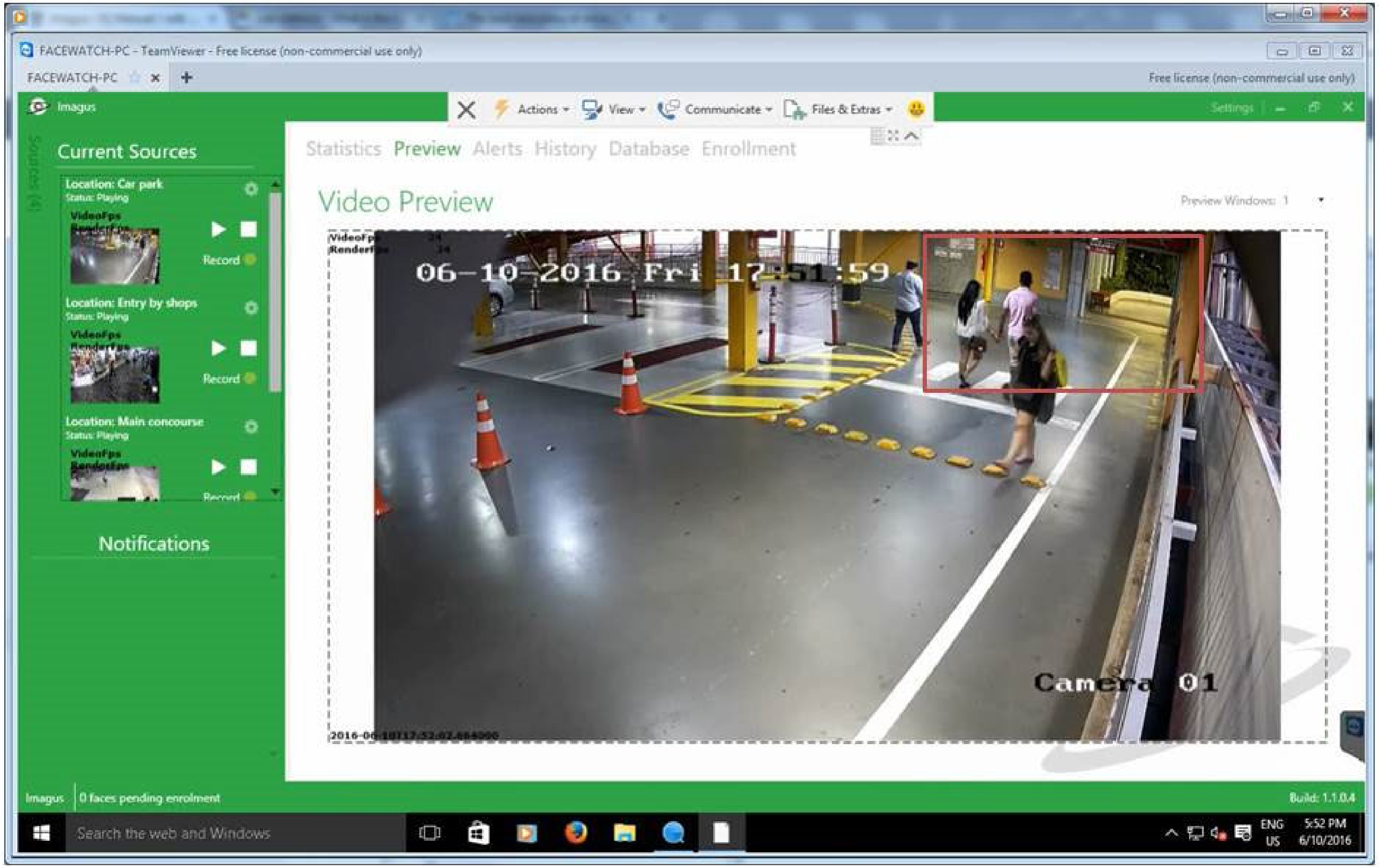
Typical CCTV Cameras in Sao Paulo Mall – Useless for Face Harvesting

Existing CCTV Cameras



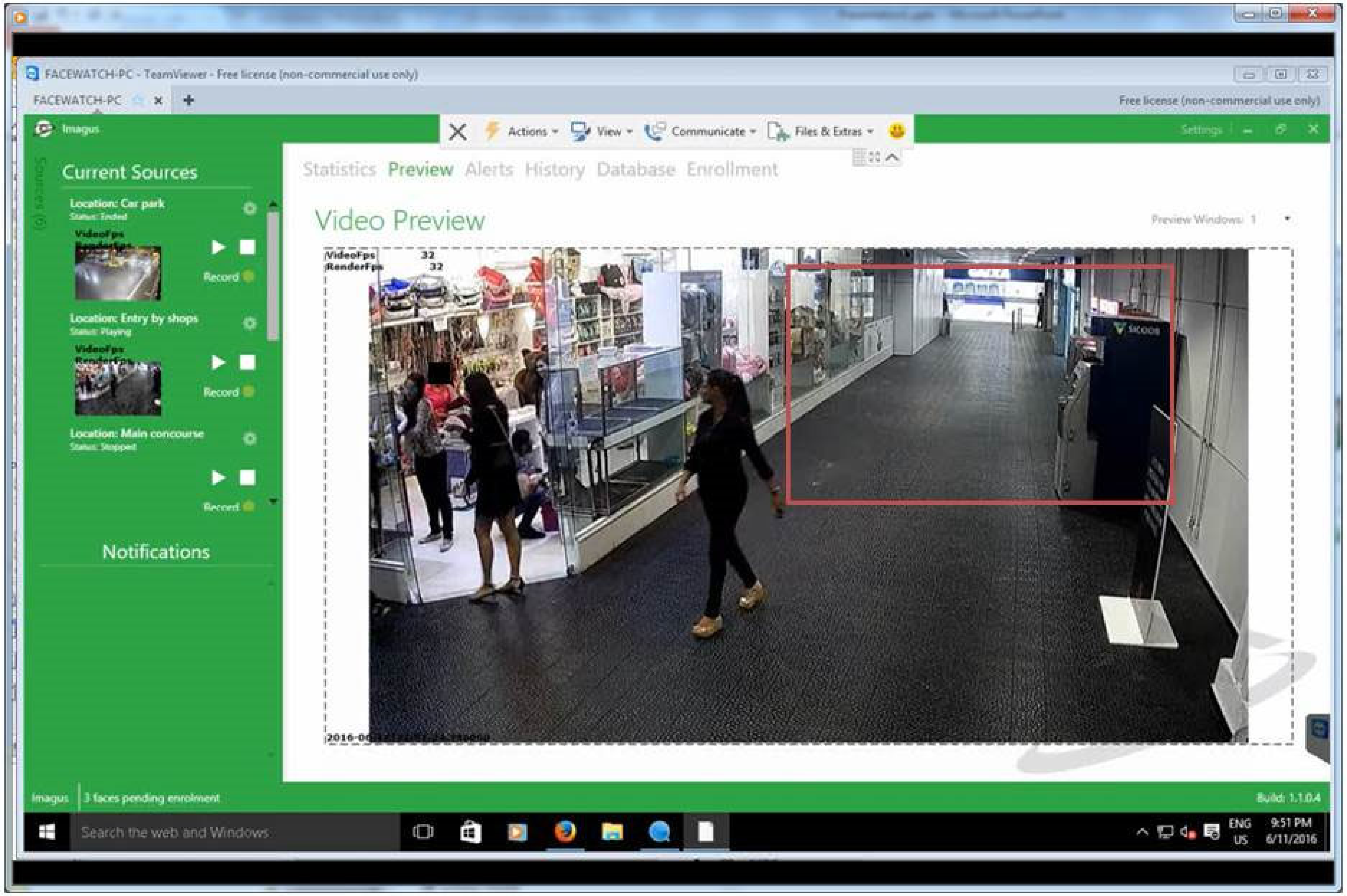
Not enough resolution. Slant angle is excessive.

Need More Focal Length



Problems with New Carpark Camera

Need Better Located Cameras



Problems with Corridor Camera

Issues Encountered with Camera-Based Face Detection

- Low Cost
- About 60s latency in camera based detection
- Poor detection rates, many bad images
- Large data rates due to full frame image size
- Hard to demonstrate live
- Hard to know what is going wrong
- Low rate of face harvesting as people often do not look at camera
- Some good matches and low false alarm rates

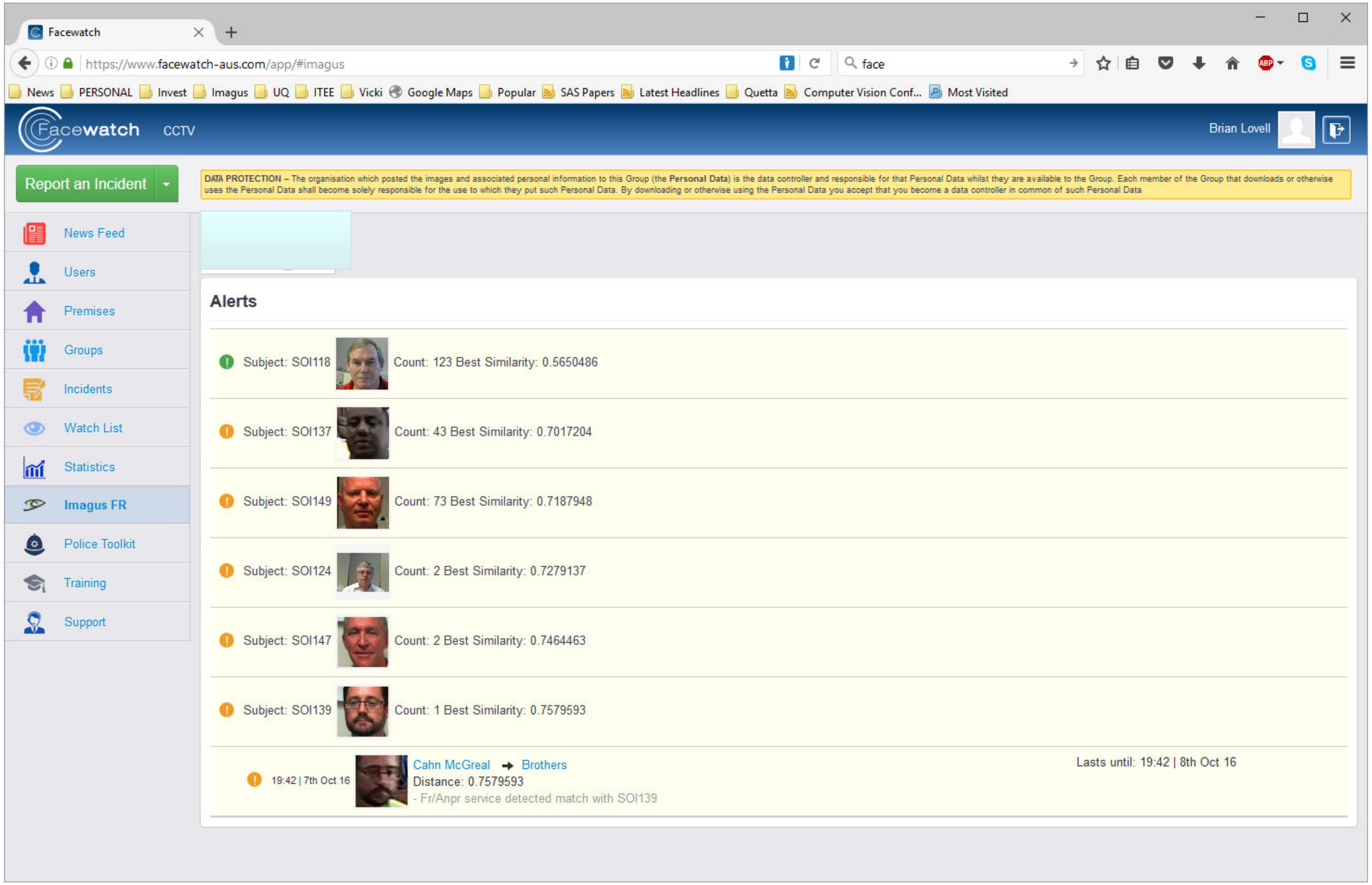
Issues Encountered in Video Appliance based detection

- Much better face harvesting due to greater number of frames
- People still do not look at camera
- Motion blur issues on almost all faces
- Strong H264 artefacts obscuring faces
- Much lower latency (2s)
- Instant local feedback and alerts
- Potential for a practical system once camera positioning issues were sorted

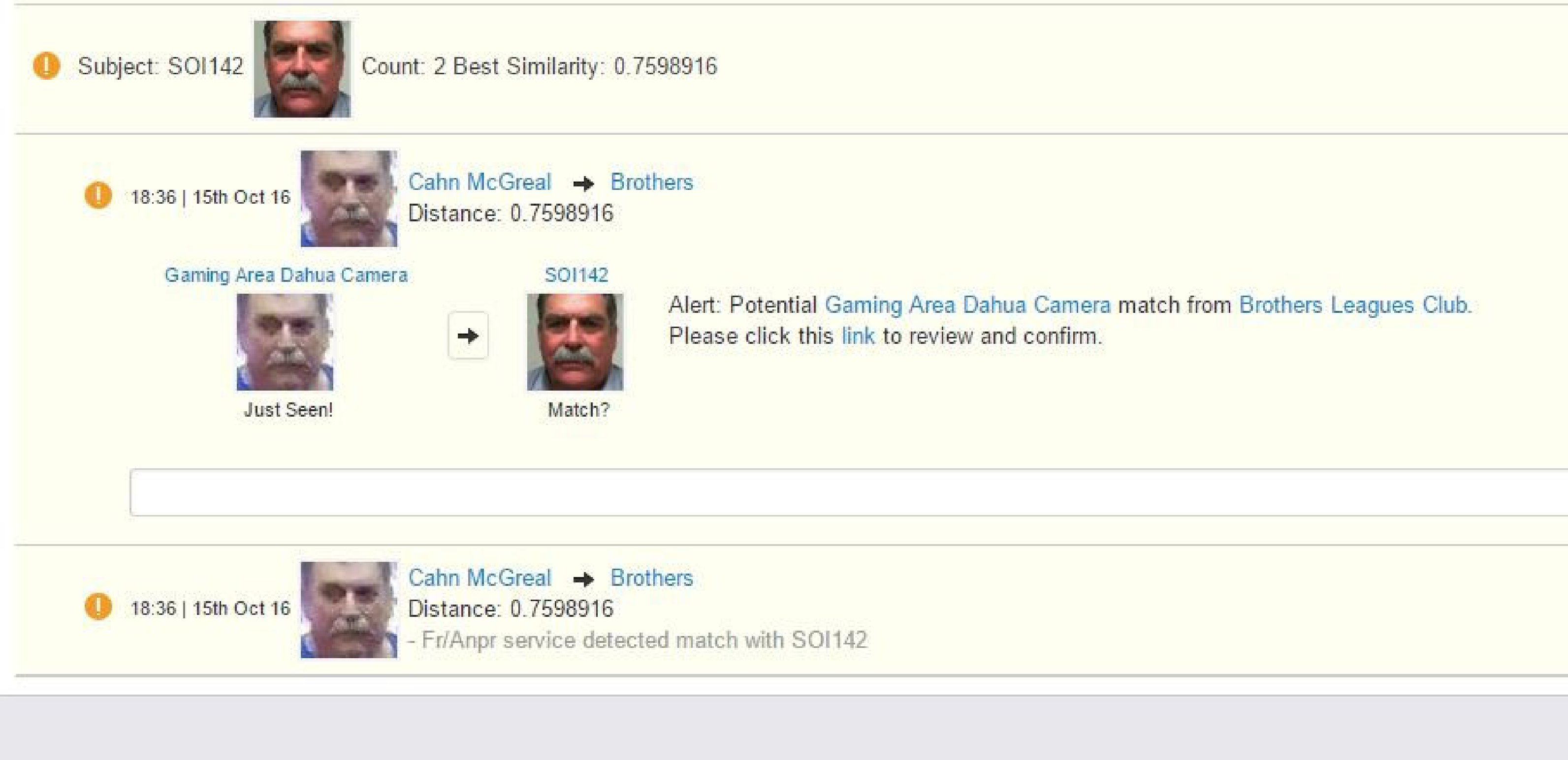
Brothers Leagues Club Network

- Deployed similar system at Brother’s Leagues Club
- Much easier due to local access, no time zone issues, and language
- Good positioning of cameras near eye level
- 3 cameras to cover foyer from a variety of angles
- System working quite well with regular alerts, but difficult to setup due to positioning and tuning of cameras

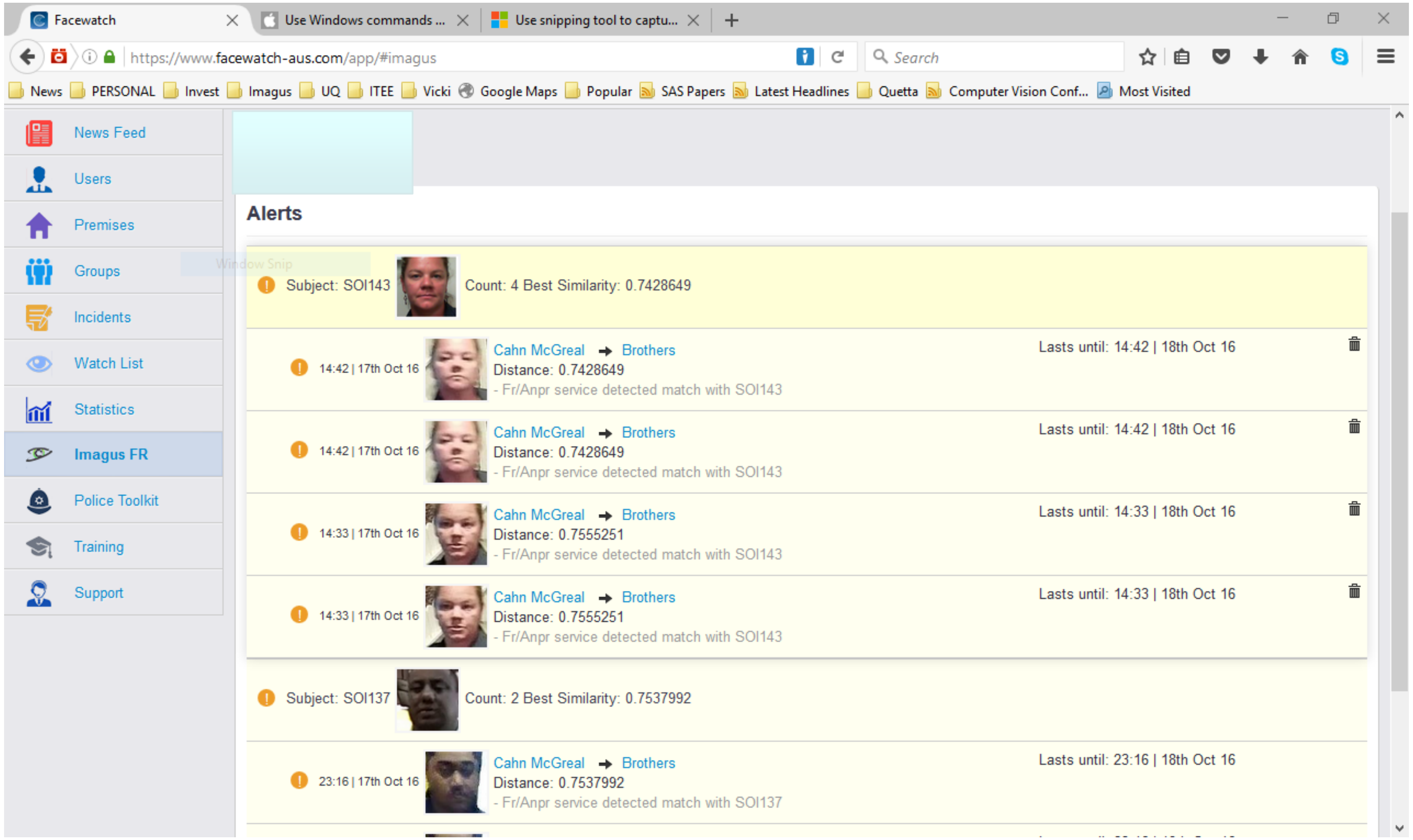
Person Alerts – Marketing Manager



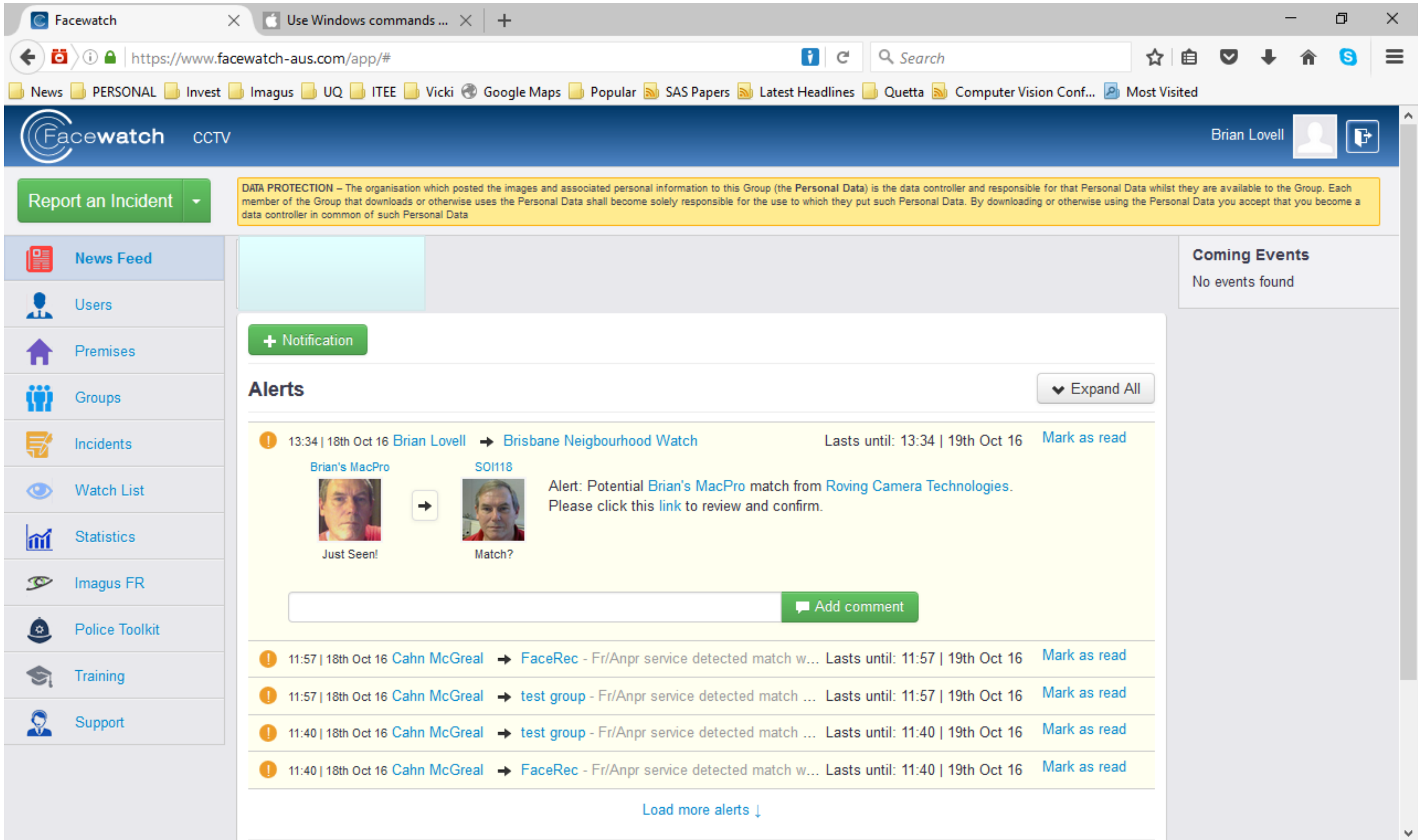
Another Match – General Manager



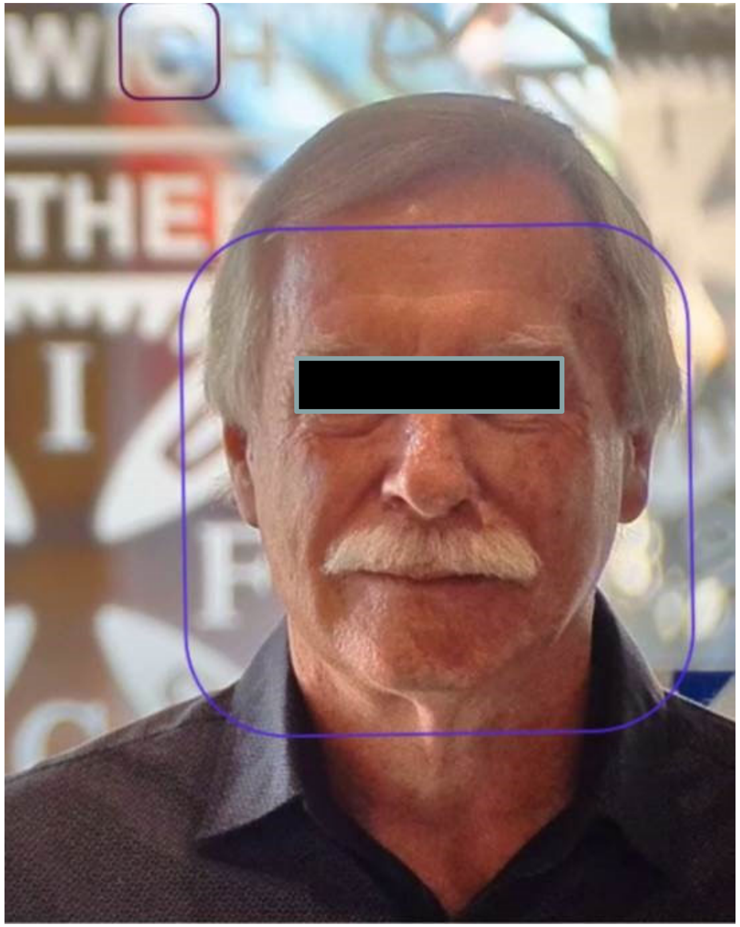
Daily Alerts



Alerting on Me



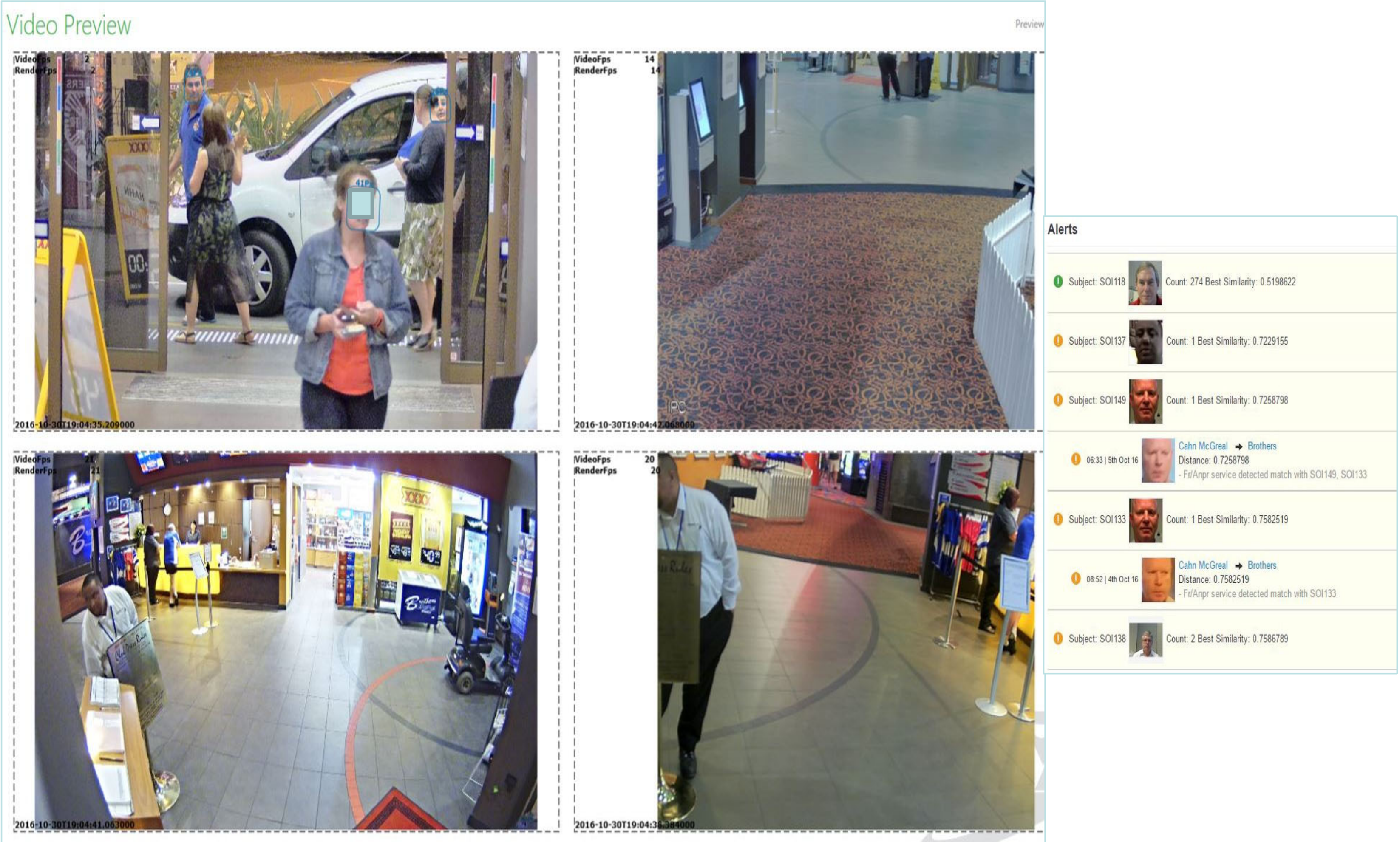
Best Camera for Doorway Installed in October 2016



We tried 15 models of camera and could not get detection on the doorway due to backlight issues.

This model was installed in October 2016 and replaces 3 others.

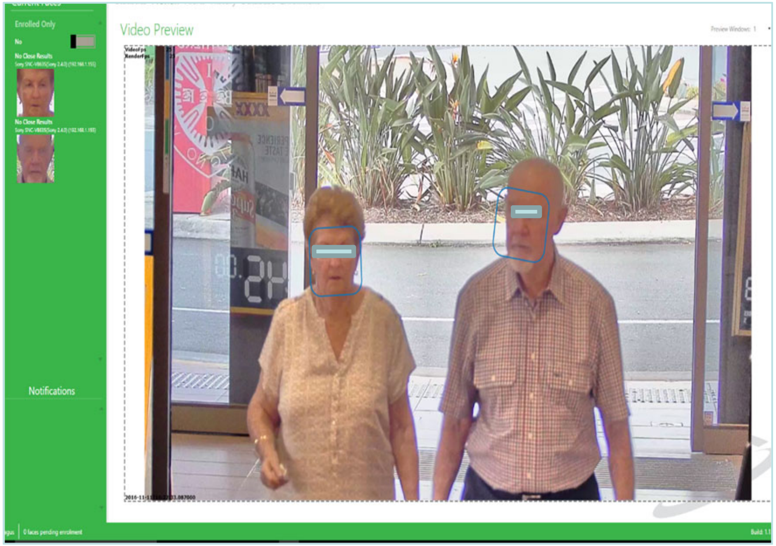
Case Study - Leagues Club

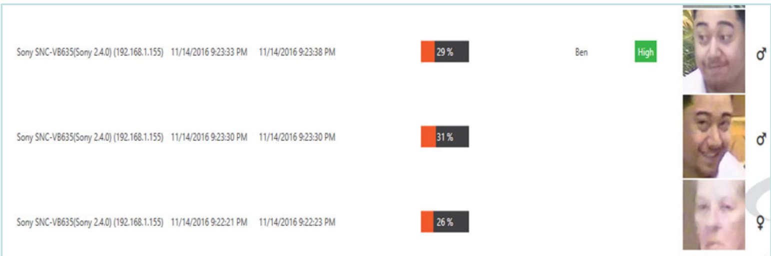


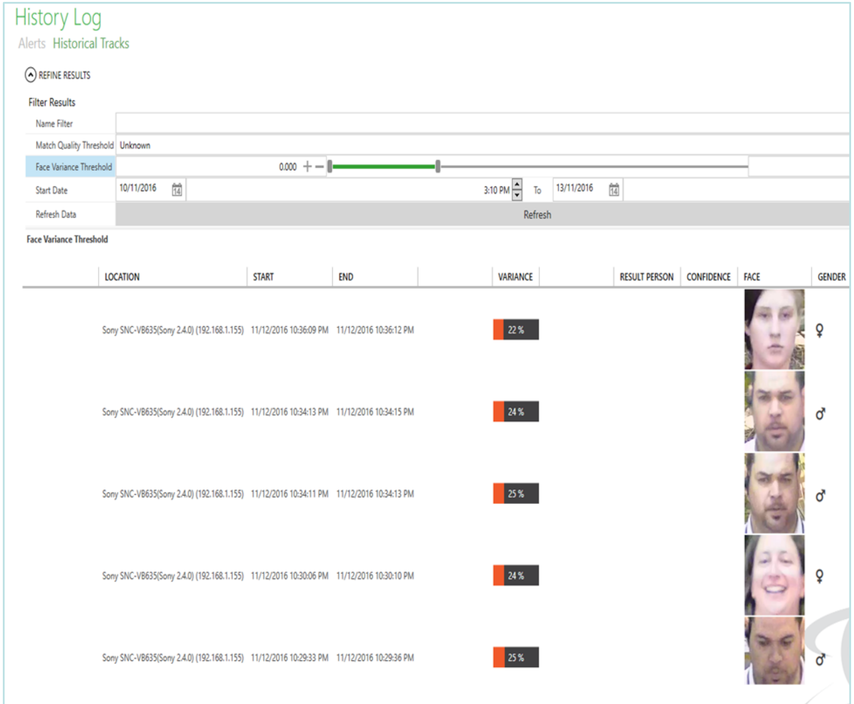
PC Platform

Leagues Club

PC Platform







Lessons Learned (1)

- Users want a single machine to handle a huge number of CCTV Streams
- CCTV Streams may be 5Mpixel or greater
- Changing Camera Positioning or settings can be a big problem in banking or similar environments
- H265 codecs preferentially blur (compress) faces as these tend to move
- CCTV installers don’t want to move cameras or change lenses

Lessons Learned (2)

- Many cameras are installed to collect profile shots
- People spend much of their lives looking at phone screens
- Eye levels cameras suffer from increased obscuration
- Identification performance decreases with increased angle of elevation
- How do we make a system that is better suited to identifying people in milling crowds?
- Build better wide angle detectors and recognisers!

Face Recognition Pipeline

Face Recognition Pipeline

- Face Detector (Cascade, MTCNN, Tiny Face)
- Face Aligner and Normalizer (deprecated with CNN)
  - For frontal, align eyes, and rescale to say 96x96
  - For non-frontal, what do you align?
- Face Recogniser
  - Turn face image into a feature vector or embedding
  - Often use Nearest Neighbour to classify
  - Scaling problems for large galleries

Face Detection

Comments

- Detection is the main computation bottleneck for video surveillance as CNN based methods are very accurate but very slow compared to Cascade.
- MTCNN (SPL 2015) is still close to state of the art for detection and many implementations are available
  - Jointly finds faces and five feature points

Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks

Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, *Senior Member, IEEE*, and Yu Qiao, *Senior Member, IEEE*

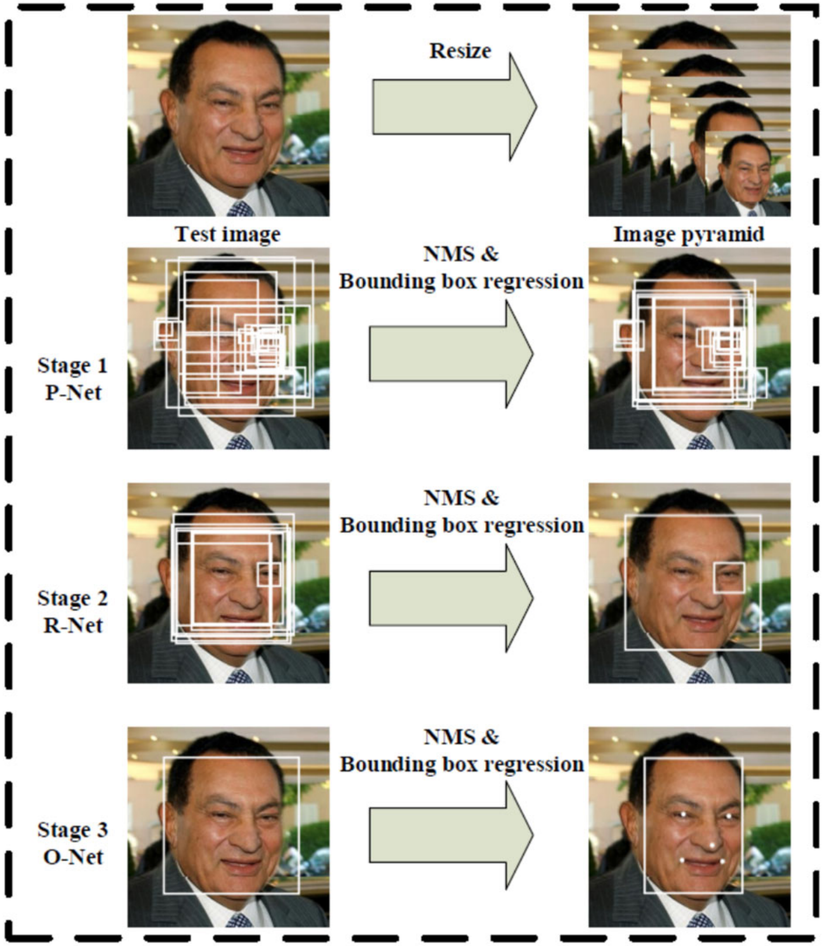
Viola-Jones Cascade

Most Impactful CVPR Paper of 2001. Awarded at CVPR2011

AWARD CVPR 2011 Longuet-Higgins Prize **Date:** *June 25, 2011*
Awarded to: Paul A. Viola and Michael J. Jones
Awarded for: "Rapid Object Detection using a Boosted Cascade of Simple Features"
Awarded by: *Conference on Computer Vision and Pattern Recognition (CVPR)*
MERL Contact: Michael Jones
Research Area: Machine Learning

Due for a well-earned retirement

MTCNN



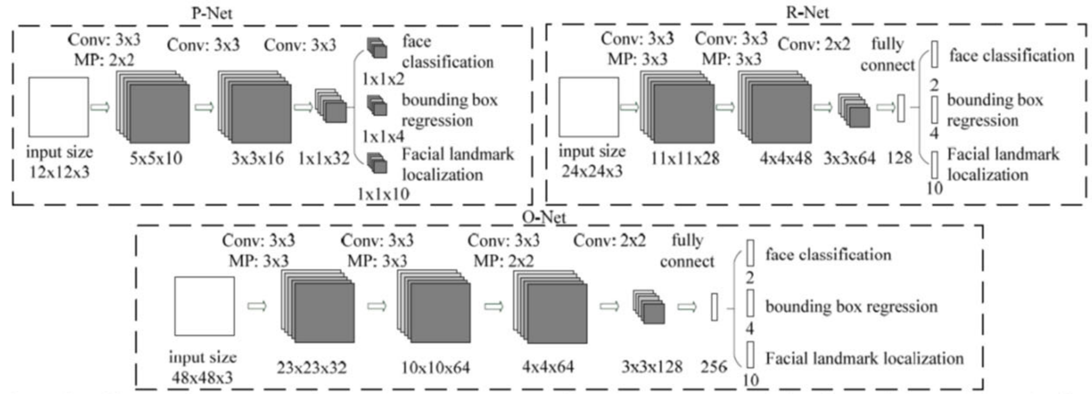
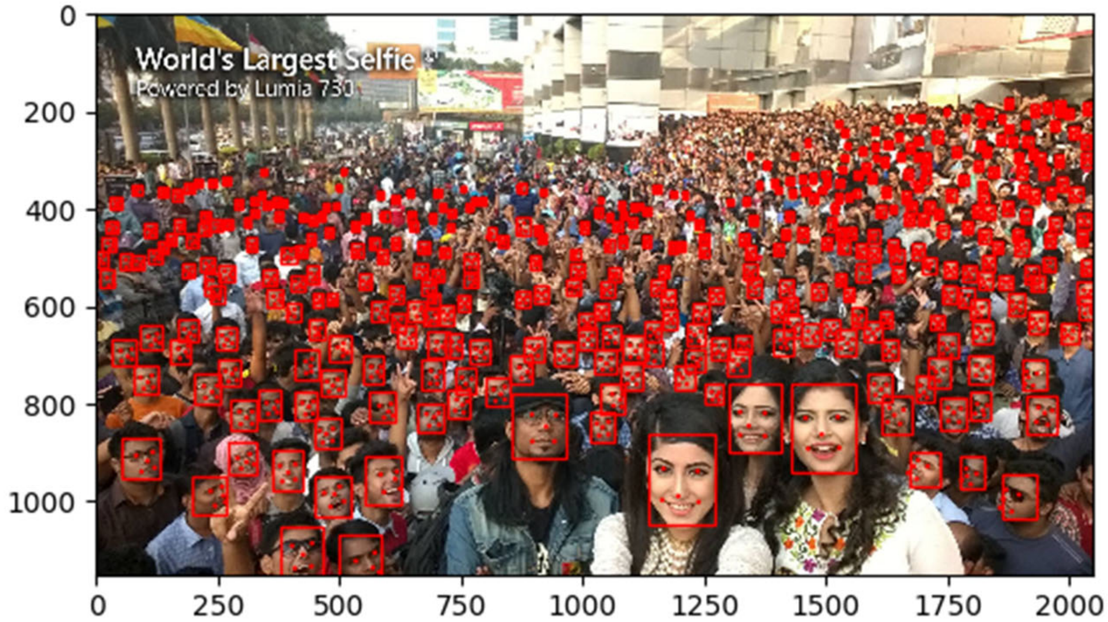


Fig. 2. The architectures of P-Net, R-Net, and O-Net, where "MP" means max pooling and "Conv" means convolution. The step size in convolution and pooling is 1 and 2, respectively.

Cascade

- Proposal Network
- Refinement Network
- Output Network

Comparison of Cascade vs MTCNN

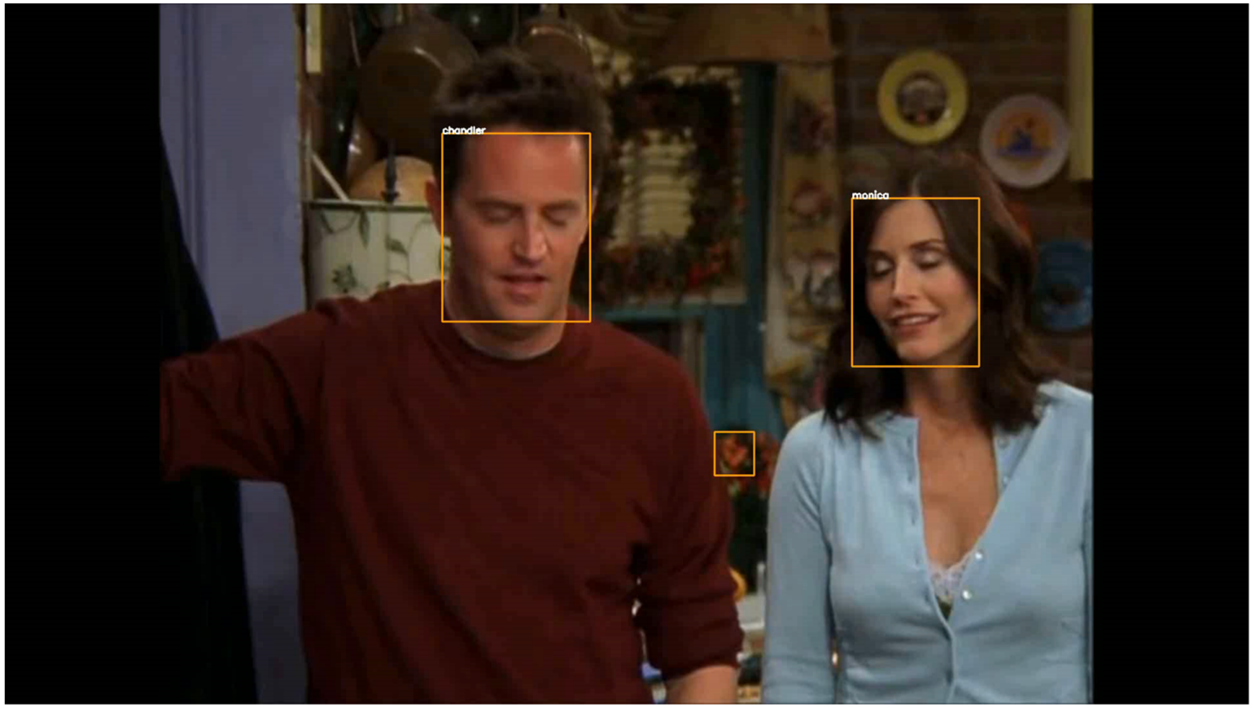


Fast, can handle large images, embedded
Even raspberry PI, in OpenCV2

Slow, handles high pose angles, provides
Facial features

Examples of Recognition with Open Source Software

- MTCNN followed by FaceNet



Comments

- Note recognition and detection at over 90 degree head pose
- Significant recognition error rates at high angles
- Slow to process in real-time due mainly to MTCNN speed
- How do we scale to multiple 5Mp CCTV streams?
- High number of **false detections** and this problem will undermine confidence in our system

Decreasing False Detection Rates in High Angle Face Detectors

Why do we care about false positives (false alarms)?

1. In a surveillance video, the majority of video frames are occupied by the background (i.e., non-faces) , which increases the probability of generating false detections

2. if too many false alarms are raised, users might lose confidence and turn off the security system

Slide Credit: Siqi Yang

Photo credit: http://www.a3con.com.br/ler-noticia/portaria-virtual-segura

Why do we care about false alarms?

3. Increased CPU load: face detection is the initial step

Face Detection

An initial step

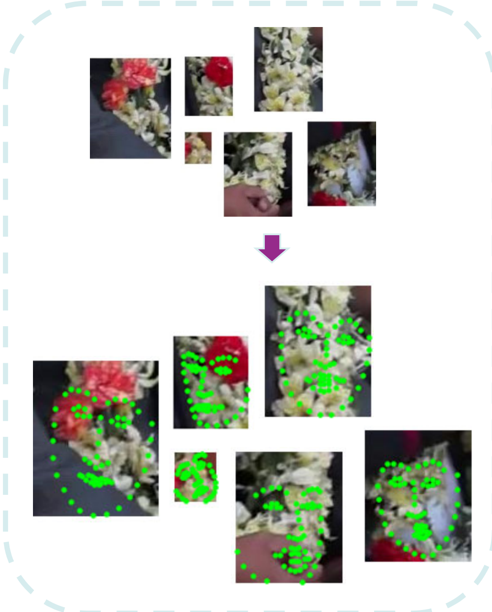
→ Face Alignment

→ Face Recognition

→ Face Expression Analysis

→ …

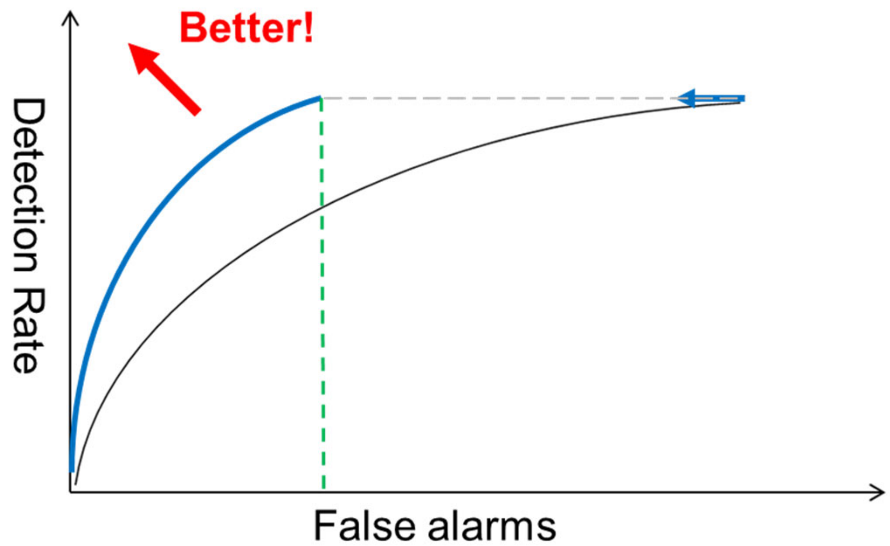
- 4. All face detectors generate false positives



Face alignment results on false detections by using the Dlib library: http://dlib.net/.

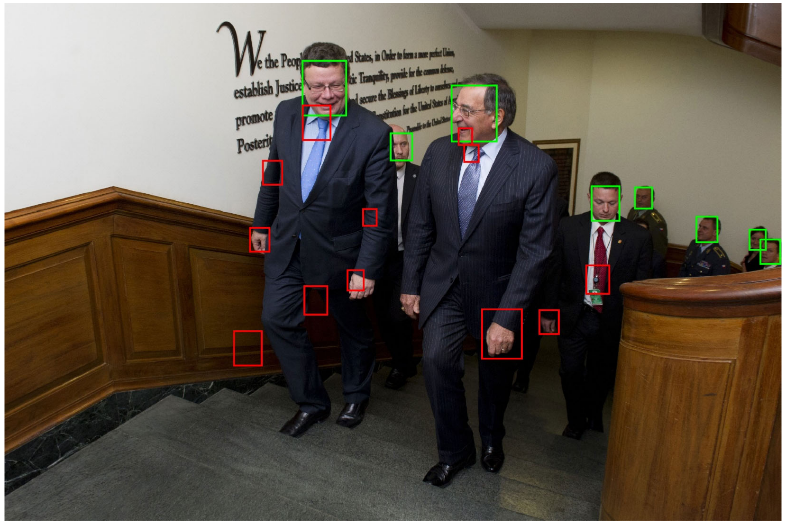
Reducing FPs improves detectors

- Aim: reducing false positives while maintaining SOTA true positive rates

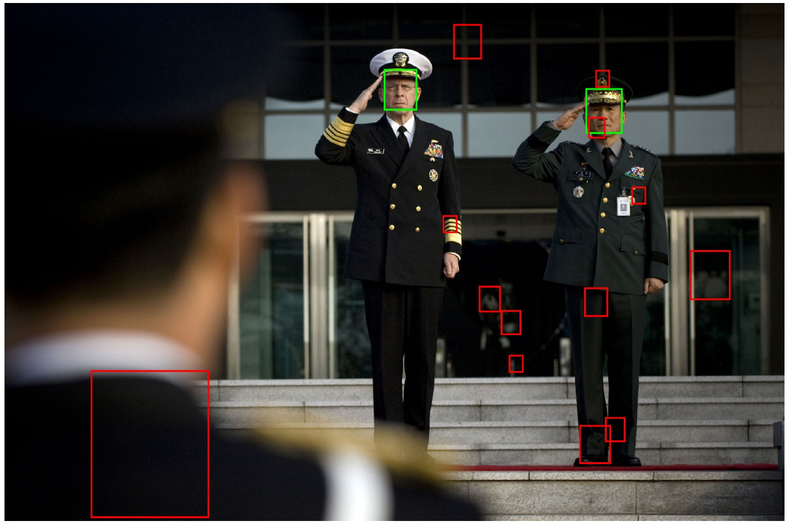


What's wrong with this face detector?

- We run the face detector, Normalised Pixel Differences (NPD) [Liao *et al.,* 2016], on the IJB-A dataset [Klare *et al.,* 2015].



12 False Positives

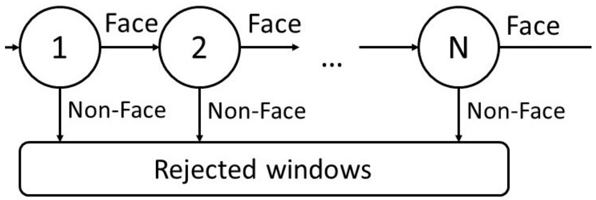


13 False Positives

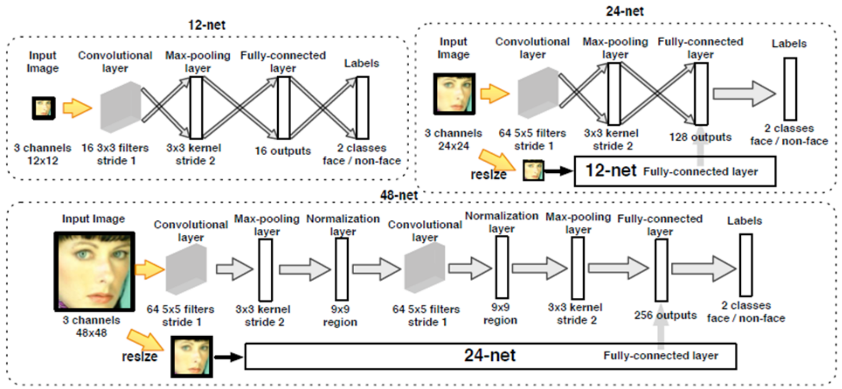
[1] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," in PAMI, 2016.
[2] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA JANUS Benchmark A," in CVPR, 2015.

Related work to reduce false positives

- **Cascade structure**



Viola and Jones (2001)



Cascade CNN (Li et al., 2015)

- **Bootstrapping or hard negative mining**
  - Online Hard Example Mining (OHEM), (Shrivastava et al., 2016)

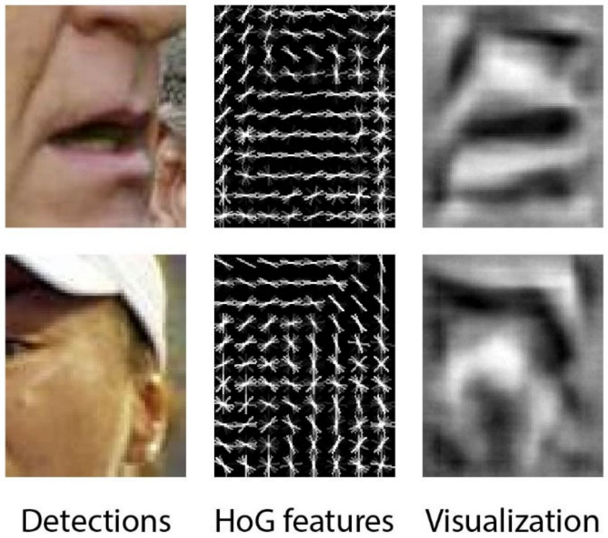
Shortcomings:

- Due to the features, classifiers and training samples, every face detector has its own theoretical limits
- The effort to train a new face detection model is enormous, e.g., large training datasets and some face detectors do not provide open source training code.

[1] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in CVPR, 2001.
[2] Li et al. "A convolutional neural network cascade for face detection." *in CVPR* 2015.
[3] Shrivastava et al.,"Training region-based object detectors with online hard example mining." IN CVPR, 2016.

Hard Face/non-Face (HFnF) Problem

- **Visualization of False Positives in HOG Feature Space**
- Method: Hoggles [Vondrick *et al.,* 2013]
    - Observe that false positives have a **face-like structure** in the feature space



[1] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba, “Hoggles: Visualizing object detection features,” in ICCV, 2013.

Cascading Face Detectors

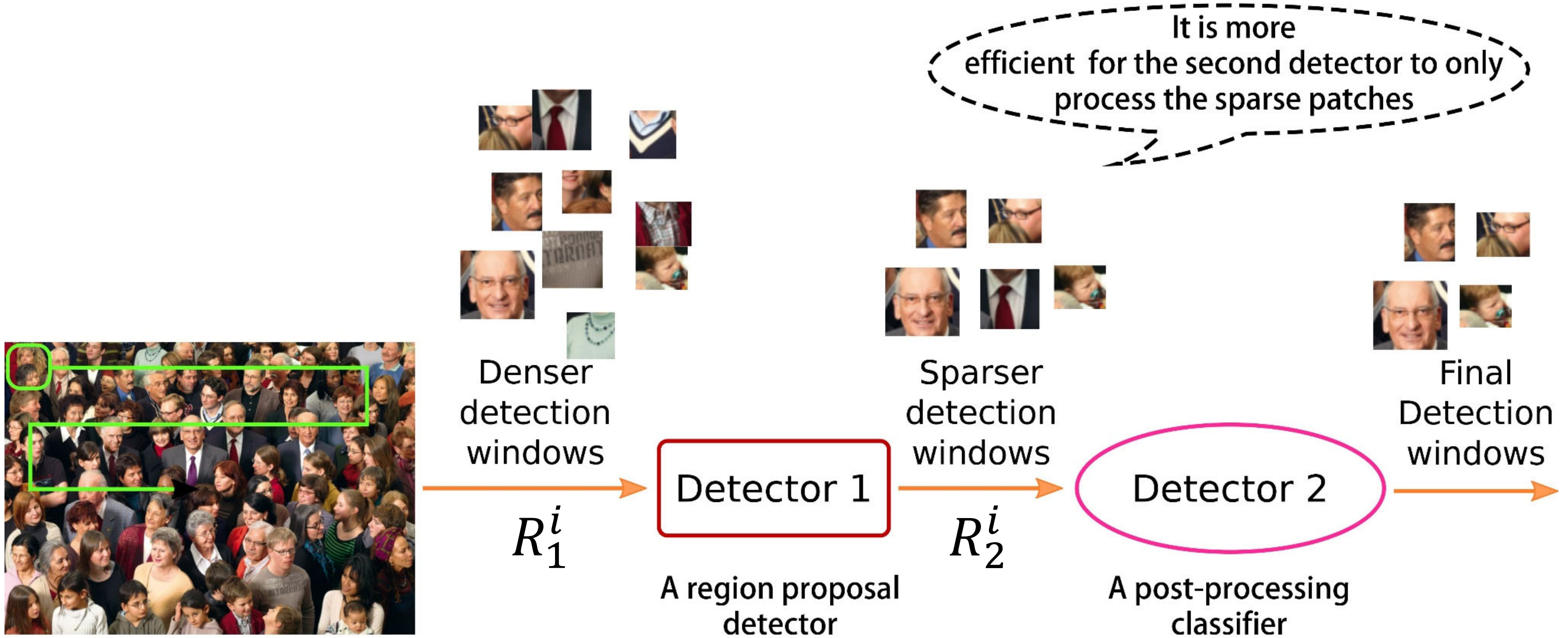
Inspiration

Siqi Yang, Arnold Wiliem, and Brian C. Lovell, **It takes two to tango: Cascading off-the- shelf face detectors,** *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Biometric Workshop*, 2018

Analogy: Combination of Cascaded Systems Using Different Technologies

Cyclonic Filter → Foam Filter → HEPA Filter

Two-stage Cascade Framework

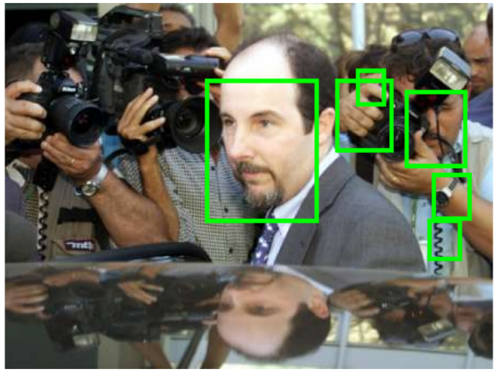


Insights

- *The cardinality of the set of input regions of the second detector is always far smaller than the cardinality of proposals in the first detector.*
- *Two different feature sets are applied.*

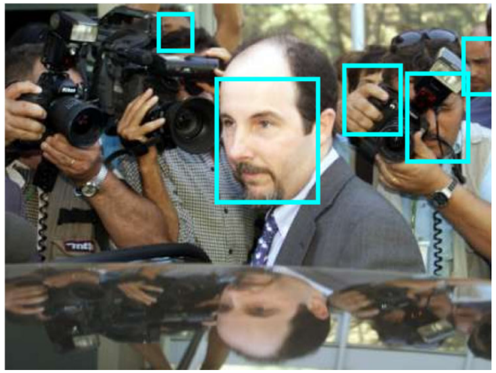
Detection Results of Current Detectors

NPD (Liao et al., 2016)



HeadHunter (Mathias et al., 2014)

MTCNN (Zhang et al., 2016)

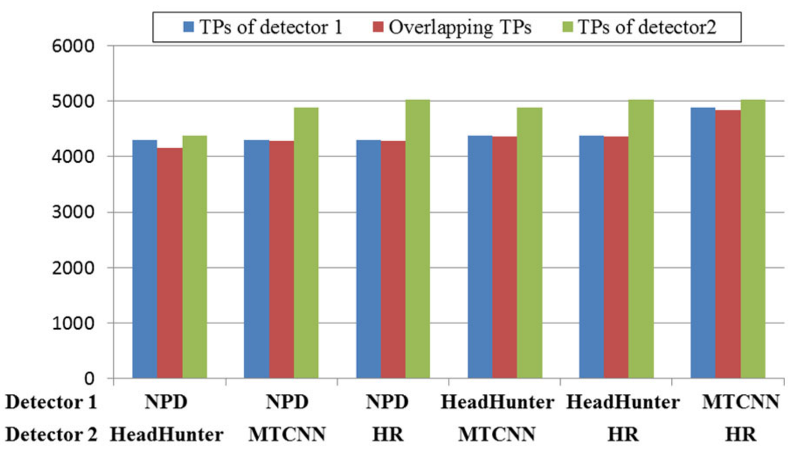


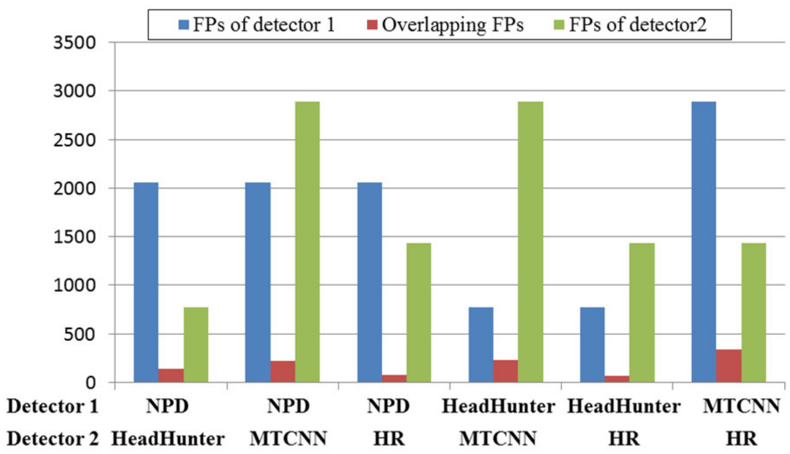
HR (Hu et al., 2017)

[1]S. Liao, A. K. Jain, and S. Z. Li. “A fast and accurate unconstrained face detector”. In PAMI, 2016..
[2] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. “Face detection without bells and whistles.” In ECCV, 2014.
[3] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. “Joint face detection and alignment using multitask cascaded convolutional networks.” In SPL, 2016.
[4] P. Hu and D. Ramanan. “Finding tiny faces.” In CVPR, 2017.

Correlation and Diversity

- The overlapping of true and false positives





Only a small number of false positives are detected by both detectors, whereas a majority of true positives overlap

Correlation and Diversity

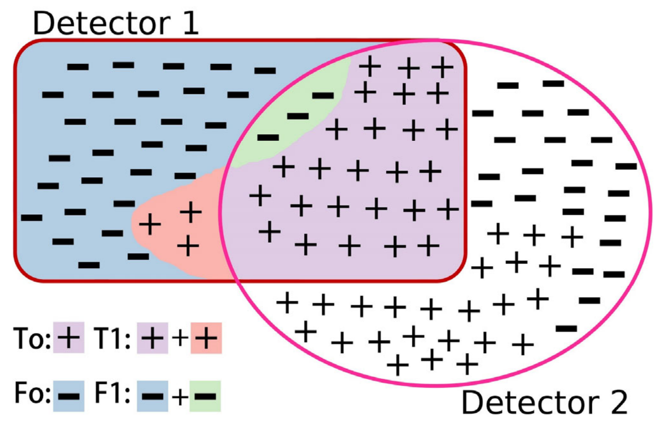
- Evaluation metrics

    - Correlation of true positives:

$$c_{2\to1}^{T} = \frac{|T_O|}{|T_1|},$$

    - Diversity of false positives:

$$d_{2\to1}^{F} = 1 - \frac{|F_O|}{|F_1|},$$



Cascade Properties

1. Correlation of true positives:

$$c^T_{2\to 1} \approx 1$$

2. Diversity of false positives:

$$d^F_{2\to 1} \approx 1$$

3. Detector runtime:

   Use the faster detector in the first stage to achieve an overall fast speed with low false detections

Experiments

Method	CPU time (SPF*)			TPR (FPPI#=0.1)
	1st stage	2nd stage	total time	
VJ [24]	0.271	-	0.271	0.462
NPD [15]	0.678	-	0.678	0.801
NPD-HeadHunter	0.678	988	988.678	0.810
NPD-MTCNN	0.678	0.073	0.751	**0.841**
NPD-HR	0.678	2.678	3.356	**0.841**
HeadHunter [18]	1961	-	1961	0.834
HeadHunter-NPD	1961	0.404	1961.404	0.819
HeadHunter-MTCNN	1961	0.116	1961.116	**0.889**
HeadHunter-HR	1961	3.648	1964.648	**0.889**
MTCNN [30]	0.355	-	0.355	0.919
MTCNN-NPD	0.355	0.220	0.575	0.843
MTCNN-HeadHunter	0.355	456	456.355	0.882
MTCNN-HR	0.355	3.496	3.851	**0.930**
HR [4]	17.687	-	17.687	**0.943**
HR-NPD	17.687	0.170	17.857	0.839
HR-HeadHunter	17.687	794	811.687	0.886
HR-MTCNN	17.687	0.076	17.763	0.930

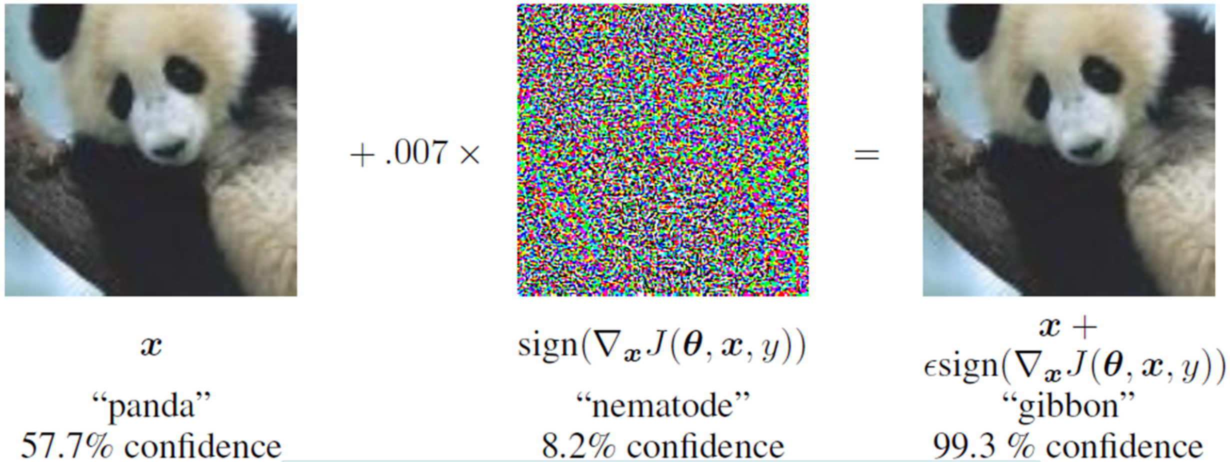
* SPF–Seconds Per Frame # FPPI–False Positives Per Image

False positives of a face detector could be reduced by 90% whilst still maintaining high true positive detection rate of HR.

Aside: Adversarial Attacks on Face Detectors

Adversarial Attack

- Adversarial perturbations are imperceptible perturbations that can change the neural network output significantly
- Fast Gradient Sign Method (FGSM) (Goodfellow et al., 2015)



$$X^{adv} = X + \alpha \cdot sign(\nabla_x \ell(f_\theta(X), y^{true}))$$

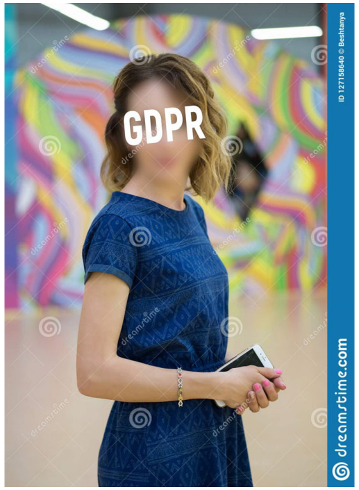
Slide Credit: Siqi Yang

[1] Goodfellow, I.J., Shlens, J., Szegedy, C.: "Explaining and harnessing adversarial examples." In ICLR, 2015.

Motivations

1. Security
   - Persons might hide themselves from being detected by surveillance cameras

2. Privacy
   - Biometric data might be utilized without the consent of the users
   - General Data Protection Regulation (GDPR) in Europe
   - To avoid faces being detected when uploaded to the servers



3. A better understanding of neural networks

Photo credit: https://www.dreamstime.com/gdpr-woman-hiding-her-face-inscription-general-data-protection-regulation-gdpr-woman-hiding-her-face-inscription-image127158640

Related Works

- Prior works in adversarial perturbation generation are applied to
  - Image classification (Goodfellow et al. 2015, Moosavi-Dezfooli et al. 2016, Carlini and Wagner 2017, Moosavi-Dezfooli et al. 2017),
  - Semantic segmentation (Metzen et al. 2017)
  - object detection (Xie et al. 2017)

- We adopt two adversarial perturbation generation methods:
  1. FGSM: $X^{adv} = X + \alpha \cdot sign(\nabla_x \ell(f_\theta(X), y^{true}))$
  2. Deepfool : $\arg\min_{\xi_i} \|\xi_i\|_2 \quad \text{subject to } f(\boldsymbol{X}_i) + \nabla f(\boldsymbol{X}_i)^T \xi_i = 0$

- To contrast these methods with our work, we categorize them as **IM**age based **P**erturbation (**IMP**) methods

[1] I. J. Goodfellow, J. Shlens, and C. Szegedy..: "Explaining and harnessing adversarial examples." In ICLR, 2015.
[2] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard..: "Deepfool: a simple and accurate method to fool deep neural networks." In CVPR, 2016.
[3] N. Carlini and D. Wagner. "Towards evaluating the robustness of neural networks." In IEEE Symposium on Security and Privacy, 2017.
[4] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard..: "Universal adversarial perturbations." In CVPR, 2017.
[5] J. H. Metzen, M. C. Kumar, T. Brox, and V. Fischer. "Universal adversarial perturbations against semantic image segmentation." In ICCV, 2017.
[6] C. Xie, J. Wang, Z. Zhang, Y. Zhou, L. Xie, and A. Yuille. "Adversarial examples for semantic segmentation and object detection." In ICCV, 2017.

The Challenge

- **An attack in object detection is more difficult than in Image Classification**
  - Need to ensure all region proposals associated with the object/instance are successfully attacked

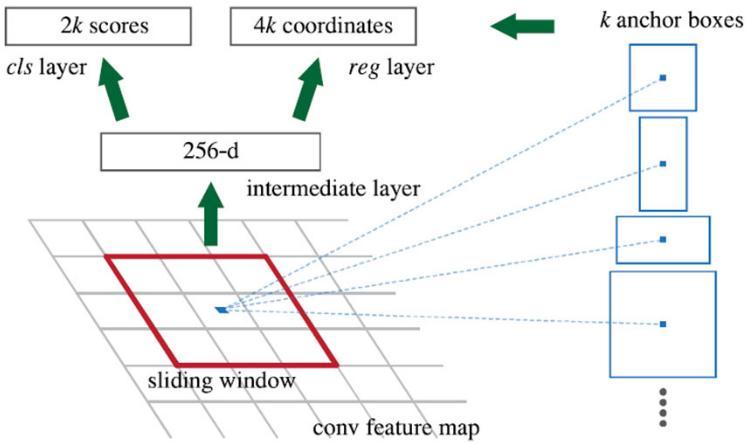
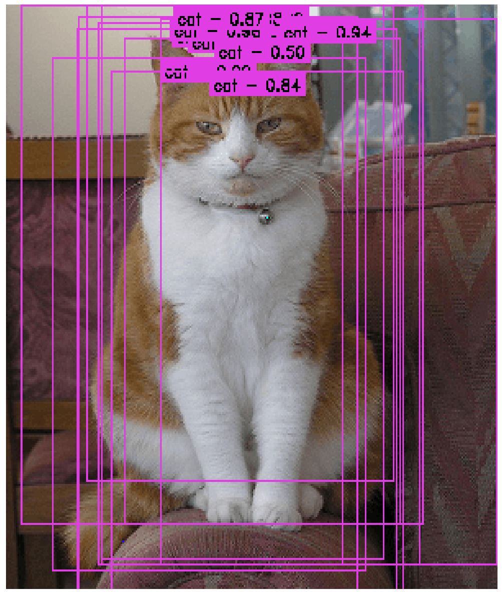


Figure. Region proposal network in Faster R-CNN (Ren et al. 2015)

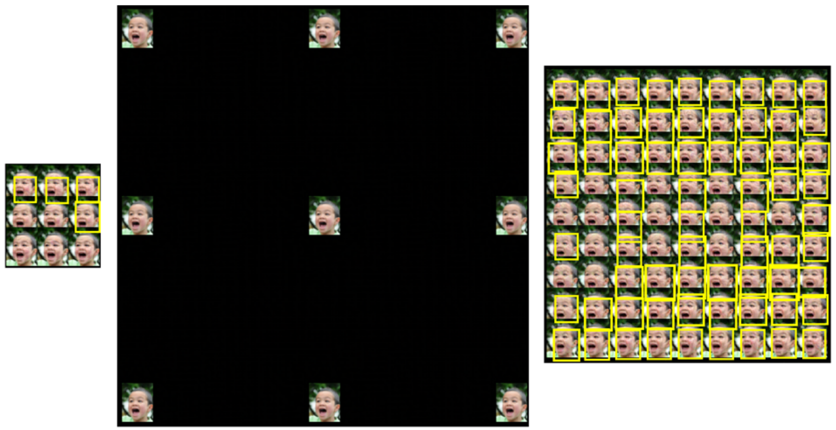


[1] Ren S, He K, Girshick R, Sun J. “Faster r-cnn: Towards real-time object detection with region proposal networks.” In NIPS, 2015.
[2] C. Xie, J. Wang, Z. Zhang, Y. Zhou, L. Xie, and A. Yuille. “Adversarial examples for semantic segmentation and object detection.” In ICCV, 2017.
Photo credtic: https://deepsense.ai/region-of-interest-pooling-explained/

Instance Perturbation Interference (IPI) Problem

- The attack success rate drops when the number of faces increases
- With the same number of faces, the attack success rate decreases as the distances among faces increases

Num of Faces	Distance	Attack Success Rate (%)
1	40	100
9	40	51.5
	160	56
	240	63.9
64	40	18.3



(a) (b) (c)

Explanations of the IPI Problem

- **Theoretical Receptive Field (TRF)**
    - A set of pixels in the input image that impact the neuron decision

- **The Distribution of Impact within TRF**:
    - In CNN, the distribution of **impact** within the TRF follows a 2D **Gaussian** distribution (Luo et al., 2016):

- **Effective Receptive Field (ERF)**
    - A fraction of TRF, where pixels have **significant impact** to the neuron decision (Luo et al., 2016)

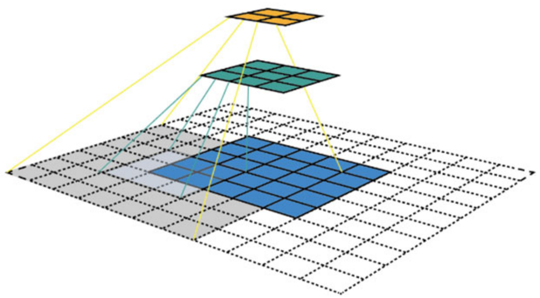


Figure. Theoretical Receptive Field

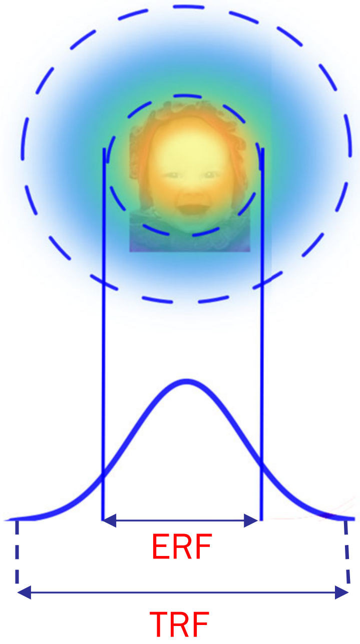


Figure. Distribution of Impact

[1] Luo, W., Li, Y., Urtasun, R., Zemel, R.: “Understanding the effective receptive field in deep convolutional neural networks.” In NIPS, 2016.
Photo credit: https://medium.com/mlreview/a-guide-to-receptive-field-arithmetic-for-convolutional-neural-networks-e0f514068807

Explanations of the IPI Problem (cont.)

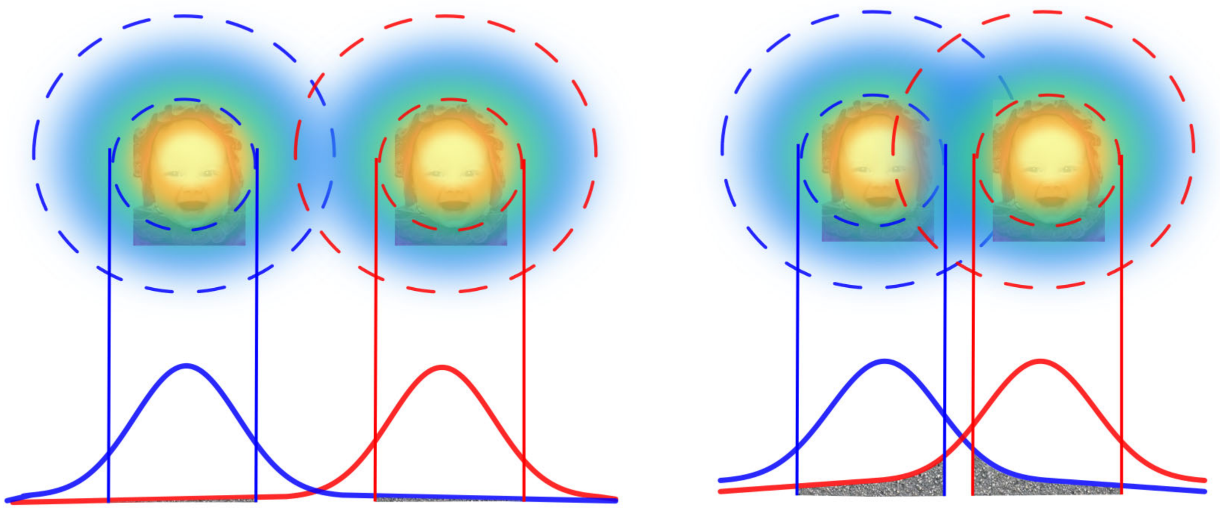
- **Our adversarial perturbation is a 2D Gaussian distribution**

$$\nabla_X L(f_\theta(X, t_c), -1) = \frac{\partial L(f_\theta(X, t_c), -1)}{\partial f_\theta(X, t_c)} \boxed{\frac{\partial f_\theta(X, t_c)}{\partial X}}$$

In CNNs, the distribution of impact within the Theoretical Receptive Field follows a 2D Gaussian distribution (Luo et al., 2016) $\frac{\partial f_\theta(X, t_c)}{\partial X}$

[1] Luo, W., Li, Y., Urtasun, R., Zemel, R.: "Understanding the effective receptive field in deep convolutional neural networks." In NIPS, 2016.

Explanations of the IPI Problem (cont.)



Perturbations overlap with the neighboring face ERF might disrupt the attack

Proposed Method: Localized Instance Perturbation (LIP)

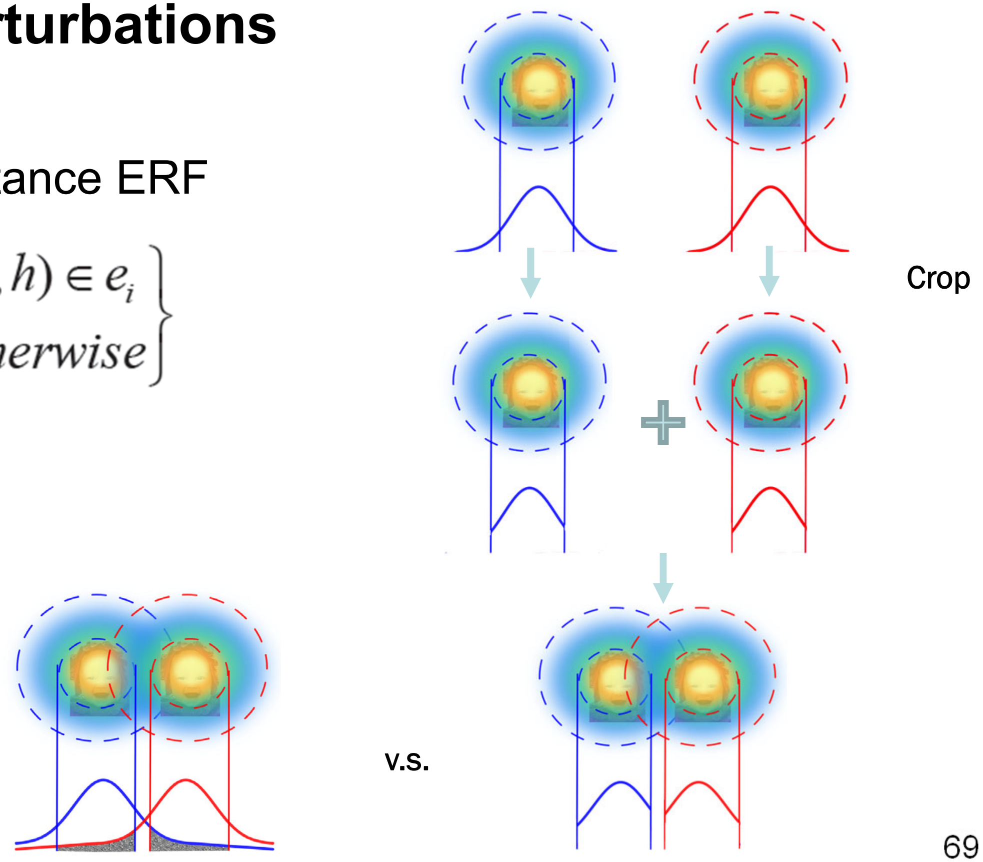
- **Aim: eliminating the interfering perturbations**

- 1. Perturbation cropping according to the instance ERF

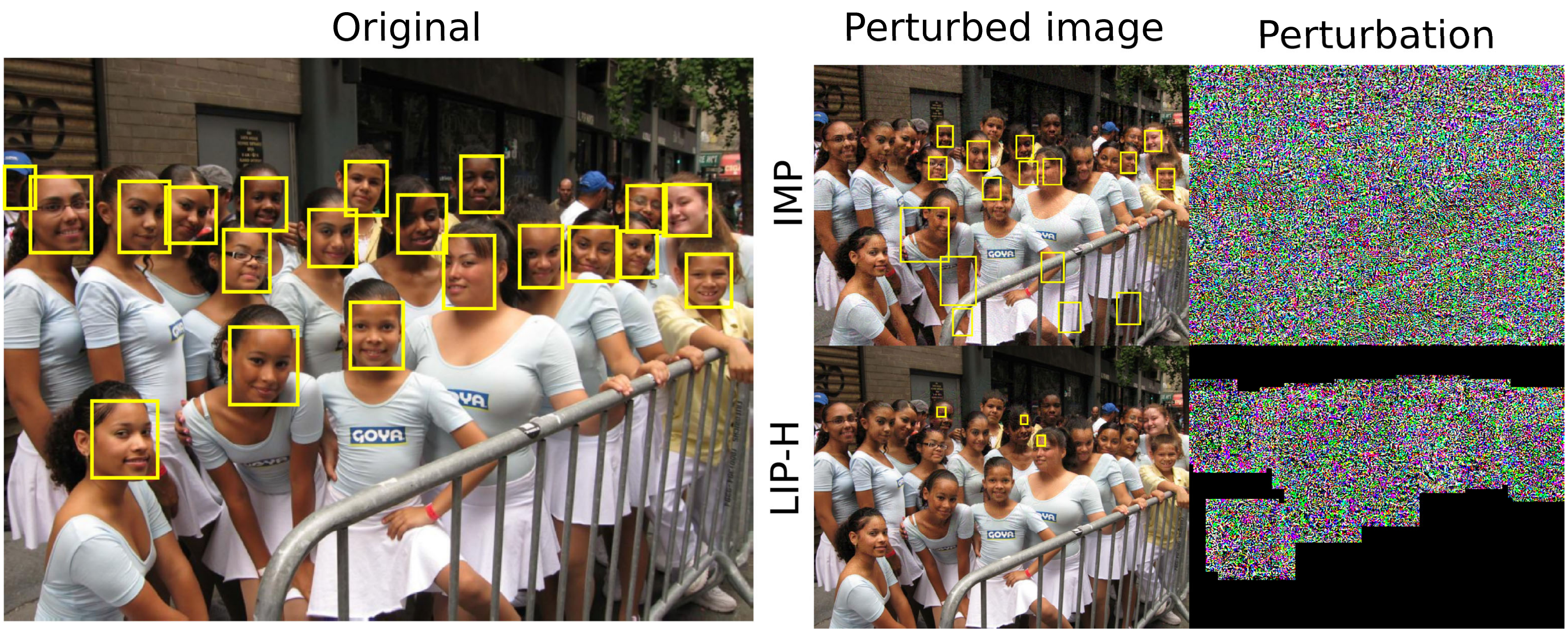
$$R_{m_i} = C_{e_i} \cdot \nabla_X L_{m_i} \quad \text{, where } C_{e_i}(w,h) = \begin{cases} 1, (w,h) \in e_i \\ 0, otherwise \end{cases}$$

- 2. Individual instance perturbation
    - processing each instance separately

$$R = \sum_{i=1}^{N} C_{e_i} \cdot \nabla_X L_{m_i}$$



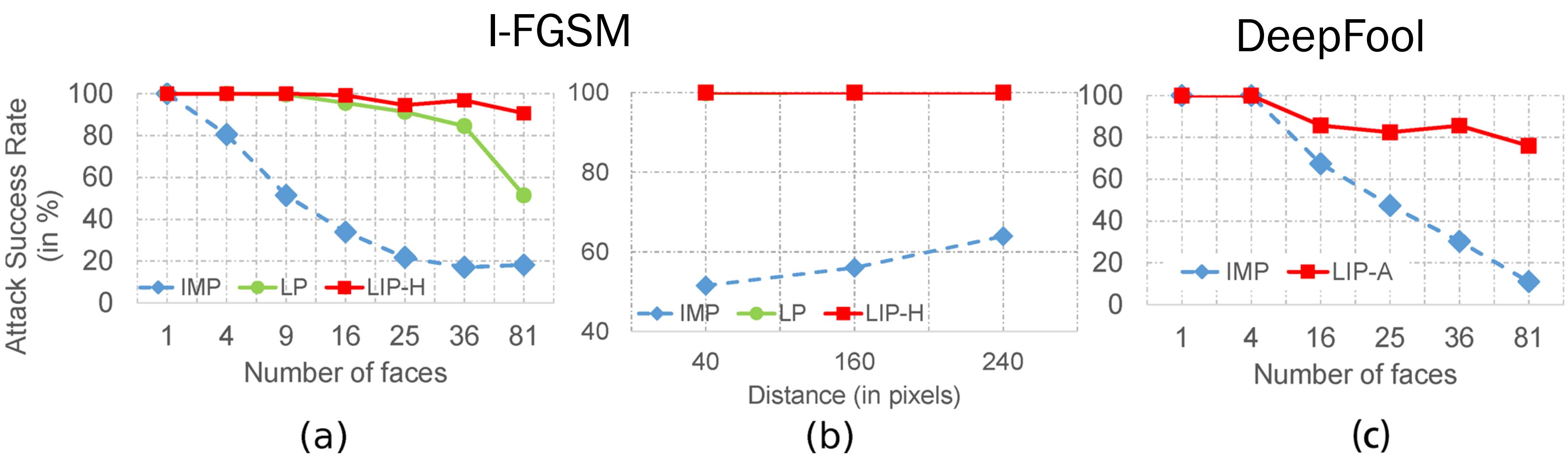
Example



Siqi Yang, Arnold Wiliem, Shaokang Chen and Brian C. Lovell, **Using LIP to Gloss Over Faces in Single-Stage Face Detection Networks,** *European Conference on Computer Vision (ECCV)*, 2018.

Evaluation on Synthetic Images

- The effect of number of faces
- The effect of distance between faces



[1] Kurakin, A., Goodfellow, I., Bengio, S.: “Adversarial examples in the physical world.” In ICLR workshop, 2017
[2] Moosavi-Dezfooli, S.M., Fawzi, A., Frossard, P.: “Deepfool: a simple and accurate method to fool deep neural networks.” In CVPR, 2016.

Evaluation on Face Detection Dataset

- We perform attacks on the pre-trained face detector, HR (Hu et al., 2017), on WIDER FACE dataset (Lin et al., 2014)

Perturbations		none	I-FGSM				DeepFool	
			IMP	LP	LIP-A	LIP-H	IMP	LIP-A
Detection Rate	easy	92.4	46.2	30.1	28.2	**26.5**	50.6	43.2
	medium	90.7	50.7	34.7	32.2	**31.1**	54.4	40.0
	hard	77.3	45.9	29.3	**23.6**	26.6	46.5	25.8
Attack Success Rate	easy	–	50.0	67.4	69.5	**71.3**	45.3	53.2
	medium	–	44.1	61.7	64.5	**65.7**	40.0	56.4
	hard	–	40.6	62.1	**69.5**	65.6	39.6	66.6

[1] Hu, P., Ramanan, D. “Finding tiny faces.” In CVPR, 2017
[2] Yang, S., Luo, P., Loy, C.C., Tang, X. “Wider face: A face detection benchmark.” In CVPR, 2015.

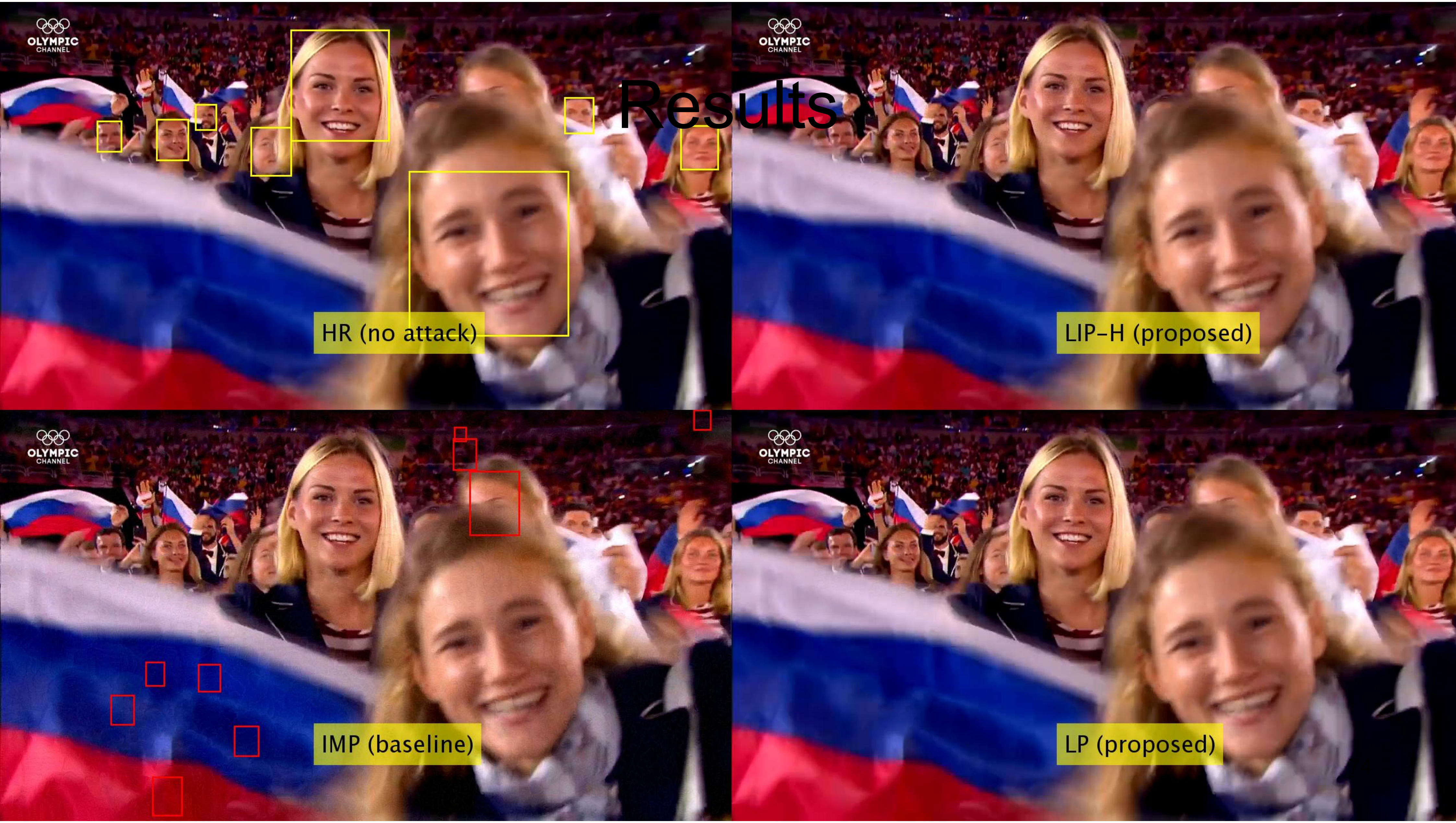
Evaluation on Object Detection Dataset

- We perform attacks on the pre-trained object detector, Faster-RCNN (Ren et al., 2015), on COCO2017 dataset (Lin et al., 2014)

Perturbations	IMP	LP
Average Recall	7.9	**2.2**
Average Precision	6.9	**1.9**

[1] Ren, S., He, K., Girshick, R., Sun, J. "Faster r-cnn: Towards real-time object detection with region proposal networks." In NIPS, 2015
[2] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., Zitnick, C.L. "Microsoft coco: Common objects in context." In ECCV, 2014.

Results

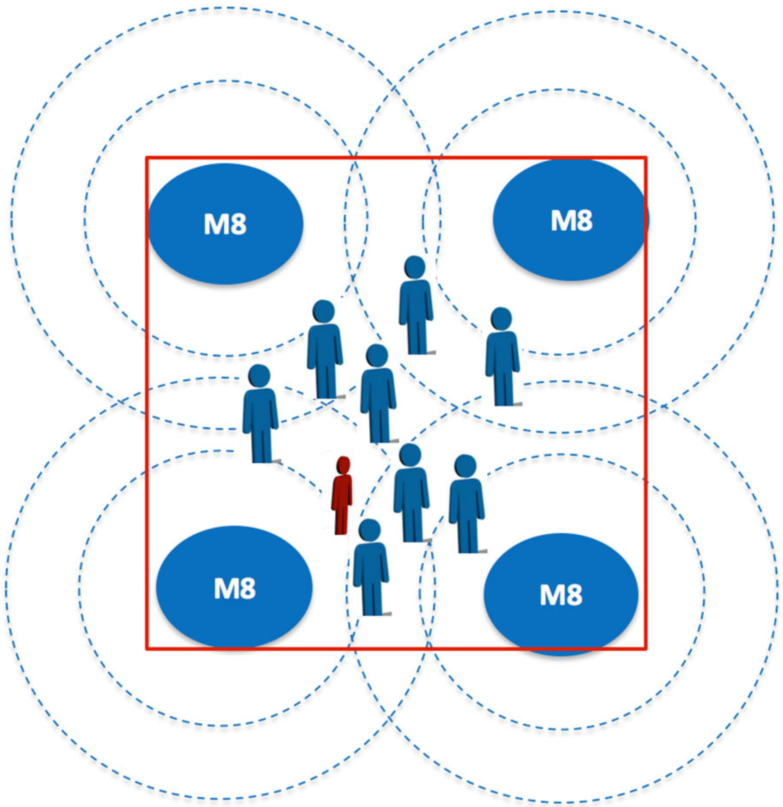

OLYMPIC CHANNEL
HR (no attack)
LIP-H (proposed)
IMP (baseline)
LP (proposed)

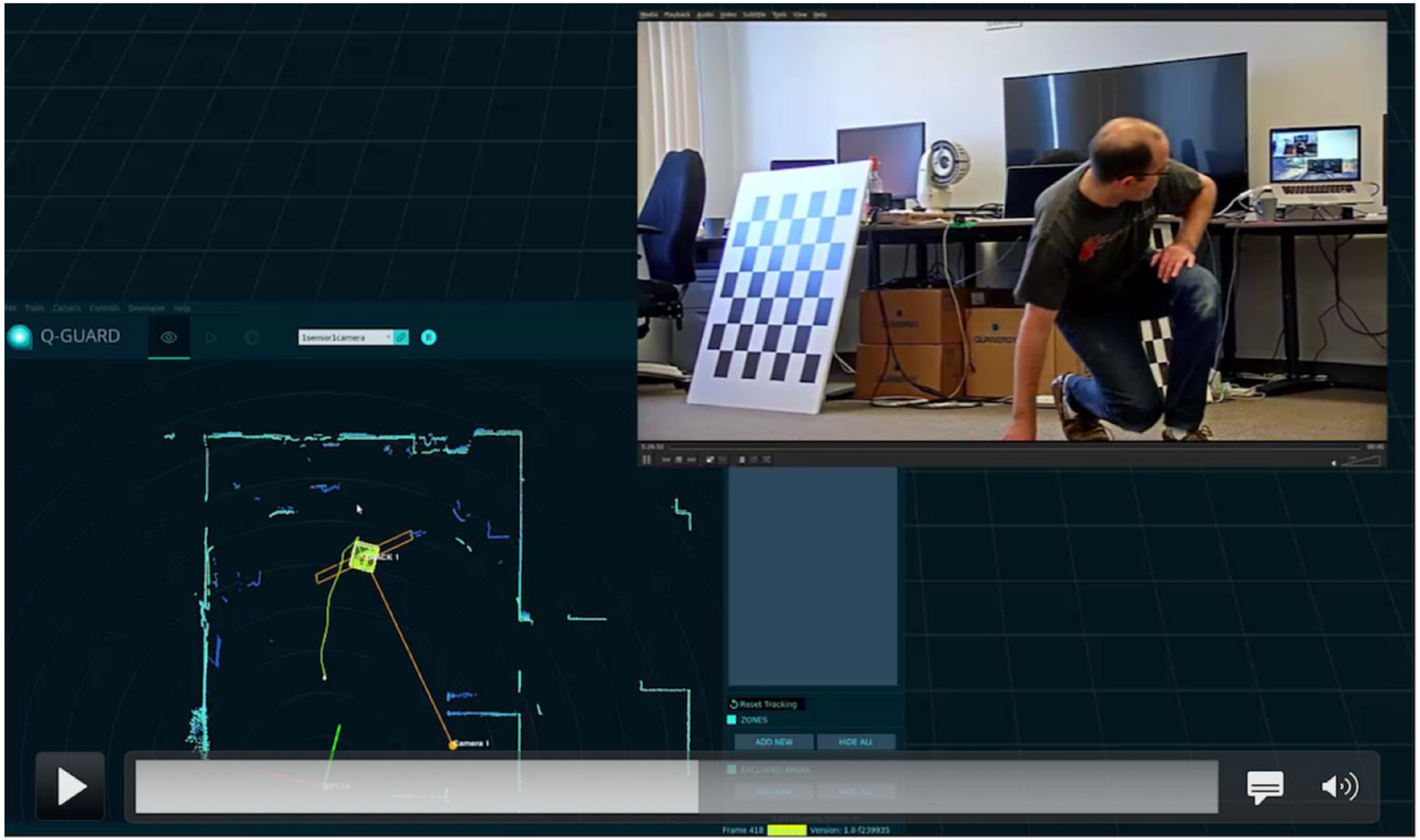
Results (cont.)

Wide Area Face Recognition with LIDAR


QUANERGY

- Line of sight technology
- Multiple sensors see each person, vehicle or other moving objects
- Machine Learning algorithms classify objects in sub-second
- Location of object known to 3cm of accuracy and uniquely tracked
- <u>Compliments</u> the existing CCTV solutions and adds more functionality re video analytics
- Add video analytics for fully automated PTZ camera control, business rules (facial recognition) for automated alerts.




Q-GUARD
Reset Tracking
ZONES
ADD NEW
HIDE ALL

Quanergy provides both hardware and software enabling a full system solution


QP1428

The sensors are monitoring traffic at one of the city's busiest intersections at Grenfell and Pulteney Streets for six days, tracking the movements of every car, bus, cyclist and even pedestrian that travels through.

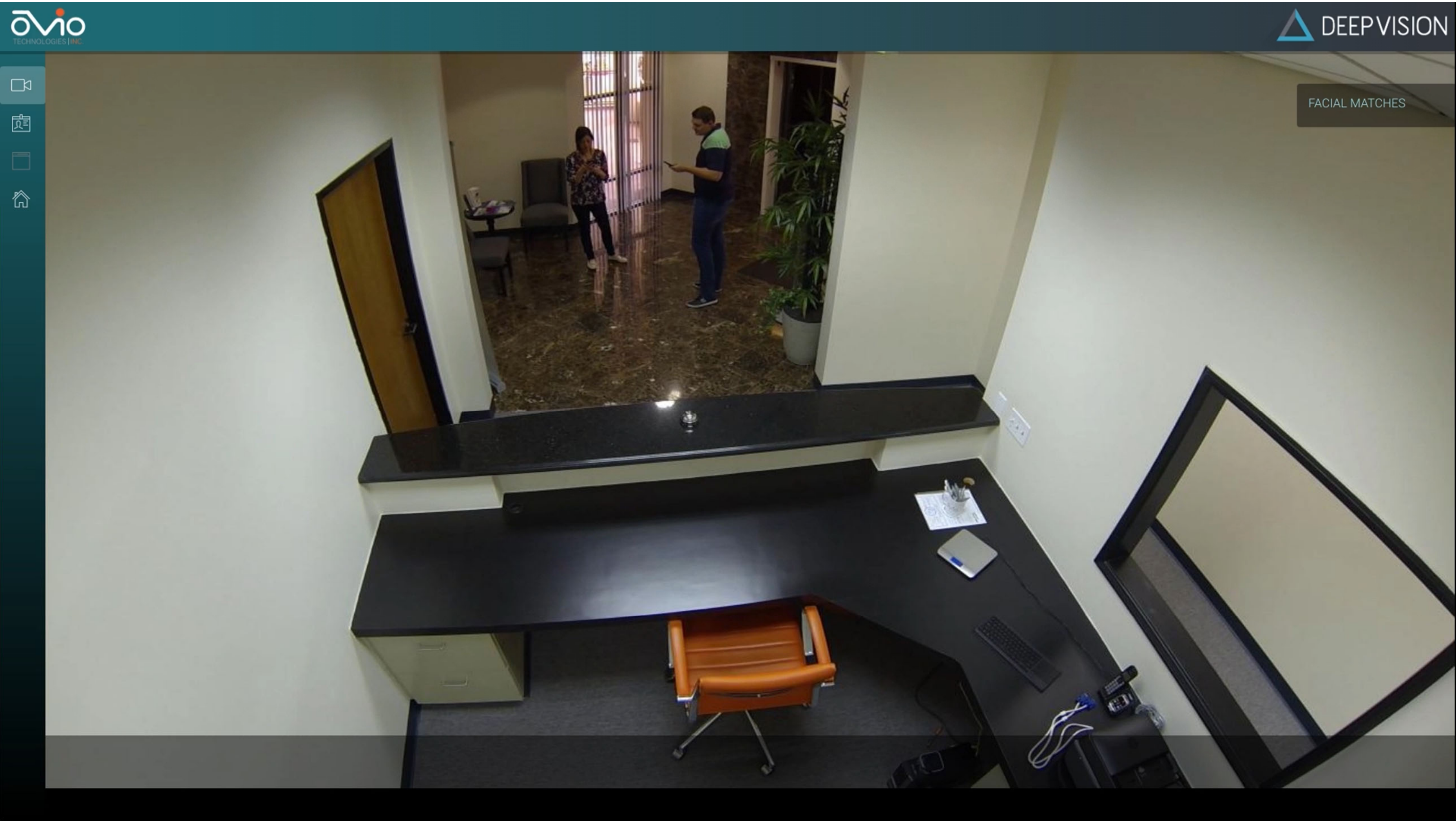
The aim is to relieve congestion, not through physically changing the road, but by improving signalling.

Real time data can be used to change traffic signals to fit current conditions. (9NEWS)

Wide Angle Face Capture

Large Angle Recognition

- CNN-based face recognition systems require a huge amount of training data at large pose angles.
- How do we collect such data in large volumes?
- Difficult to scrape from the internet as most people upload low pose angle data and we have difficulty detecting and recognizing high pose faces



TECHNOLOGIES | INC.
DEEPVISION
FACIAL MATCHES

3D Modelling to Generate Synthetic Faces

Wide Angle Face Recognition with ArcFace

ArcFace (CVPR2019)

Proposes new loss function on a hyper-spherical embedding that is easy to compute and yields very high performance

Jiankang Deng and Jia Guo and Stefanos Zafeiriou, ArcFace: Additive Angular Margin Loss for Deep Face Recognition, url = http://arxiv.org/abs/1801.07698

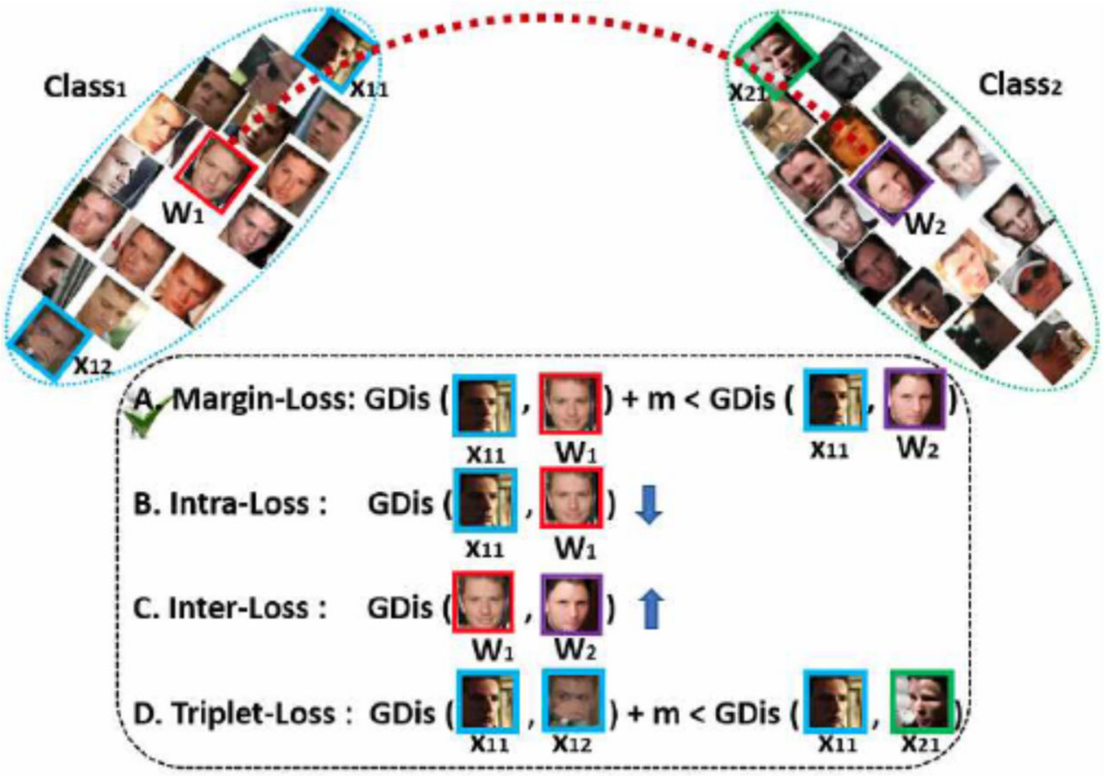
Training Deep Face CNNs

- Several methods
- Softmax loss (Closed set only as fixed number of classes/identities)
- Triplet loss
  - Learn an embedding followed by NN (Open Set)
  - Also called Deep Metric Learning
- If your embedding is good, there is little need for a fancy MLP on the output
- Modern Trend – more CNN layers and fewer FC layers

Problems with Triplet Loss

- Aim is for faces of same persons to be close in feature space (embedding) and for faces of different persons to be at least a distance d apart to provide sufficient margin.
- Problems
  - Need to perform semi-hard data mining
  - Combinatorial explosion in large databases

Gesodesic Loss Functions



Experiments show that Margin Loss A is the most effective

Angular Margins

- If we normalize weights such that the L2 norm is 1, and we normalise the embeddings such that the L2 norm is s, all embeddings lie in a hypersphere of radius s.
- Related earlier works, SphereFace and CosFace

[15] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *CVPR*, 2017. 1, 2, 3, 4, 5, 6, 7

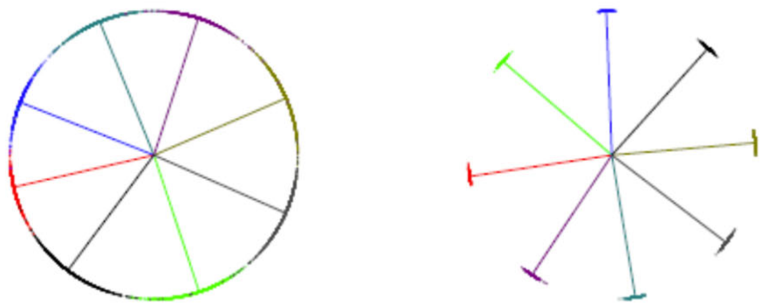
[35] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Zhifeng Li, Dihong Gong, Jingchao Zhou, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *CVPR*, 2018. 1, 2, 3, 5, 6, 7

Hyperspherical Embedding

For simplicity, we fix the bias $b_j = 0$ as in [15]. Then, we transform the logit [24] as $W_j^T x_i = \|W_j\| \|x_i\| \cos\theta_j$, where θ_j is the angle between the weight W_j and the feature x_i. Following [15, 35, 34], we fix the individual weight $\|W_j\| = 1$ by l_2 normalisation. Following [26, 35, 34, 33], we also fix the embedding feature $\|x_i\|$ by l_2 normalisation and re-scale it to s. The normalisation step on features and weights makes the predictions only depend on the angle between the feature and the weight. The learned embedding features are thus distributed on a hypersphere with a radius of s.

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}. \quad (2)$$

MNIST Embeddings



(a) Softmax (b) ArcFace

Figure 3. Toy examples under the softmax and ArcFace loss on 8 identities with 2D features. Dots indicate samples and lines refer to the centre direction of each identity. Based on the feature normalisation, all face features are pushed to the arc space with a fixed radius. The geodesic distance gap between closest classes becomes evident as the additive angular margin penalty is incorporated.

Comparison of Loss Functions

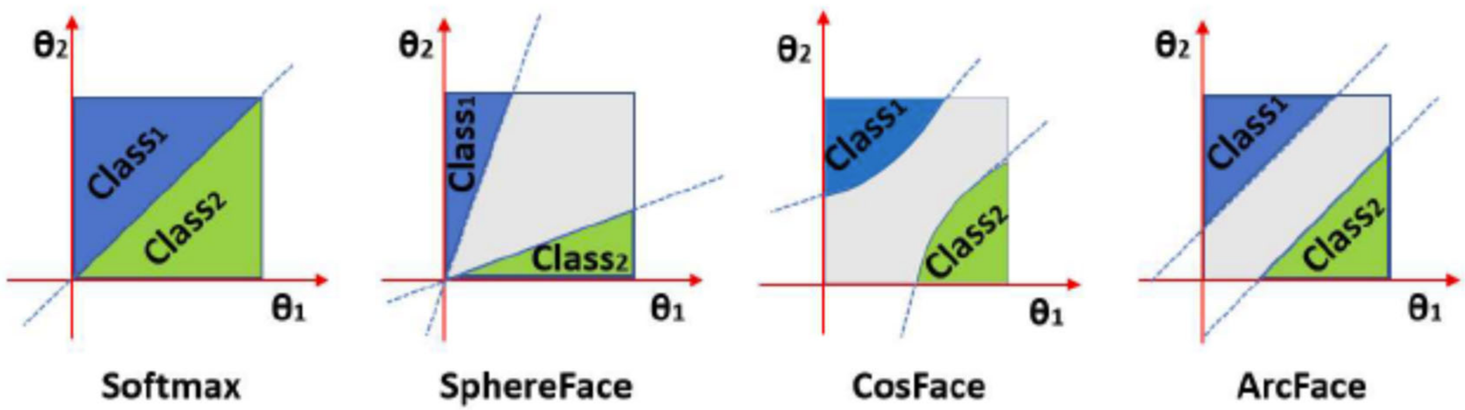


Figure 5. Decision margins of different loss functions under binary classification case. The dashed line represents the decision boundary, and the grey areas are the decision margins.

Thorough Evaluation of ArcFace

Datasets	#Identity	#Image/Video
CASIA [41]	10K	0.5M
VGGFace2 [3]	9.1K	3.3M
MS1MV2	85K	5.8M
MS1M-DeepGlint [1]	87K	3.9M
Asian-DeepGlint [1]	94 K	2.83M
LFW [10]	5,749	13,233
CFP-FP [28]	500	7,000
AgeDB-30 [19]	568	16,488
CPLFW [44]	5,749	11,652
CALFW [45]	5,749	12,174
YTF [38]	1,595	3,425
MegaFace [12]	530 (P)	1M (G)
IJB-B [37]	1,845	76.8K
IJB-C [18]	3,531	148.8K
Trillion-Pairs [1]	5,749 (P)	1.58M (G)
iQIYI-VID [17]	4,934	172,835

Table 1. Face datasets for training and testing. "(P)" and "(G)" refer to the probe and gallery set, respectively.

Loss Functions	LFW	CFP-FP	AgeDB-30
ArcFace (0.4)	99.53	95.41	94.98
ArcFace (0.45)	99.46	95.47	94.93
ArcFace (0.5)	**99.53**	**95.56**	**95.15**
ArcFace (0.55)	99.41	95.32	95.05
SphereFace [15]	99.42	-	-
SphereFace (1.35)	99.11	94.38	91.70
CosFace [35]	99.33	-	-
CosFace (0.35)	99.51	95.44	94.56
CM1 (1, 0.3, 0.2)	99.48	95.12	94.38
CM2 (0.9, 0.4, 0.15)	99.50	95.24	94.86
Softmax	99.08	94.39	92.33
Norm-Softmax (NS)	98.56	89.79	88.72
NS+Intra	98.75	93.81	90.92
NS+Inter	98.68	90.67	89.50
NS+Intra+Inter	98.73	94.00	91.41
Triplet (0.35)	98.98	91.90	89.98
ArcFace+Intra	99.45	95.37	94.73
ArcFace+Inter	99.43	95.25	94.55
ArcFace+Intra+Inter	99.43	95.42	95.10
ArcFace+Triplet	99.50	95.51	94.40

Table 2. Verification results (%) of different loss functions ([CASIA, ResNet50, loss*]).

What Cardinality of Embedding is required?

$$\mathbb{E}[\theta(W_j)] \to n^{-\frac{2}{d-1}}\Gamma(1+\frac{1}{d-1})(\frac{\Gamma(\frac{d}{2})}{2\sqrt{\pi}(d-1)\Gamma(\frac{d-1}{2})})^{-\frac{1}{d-1}},$$

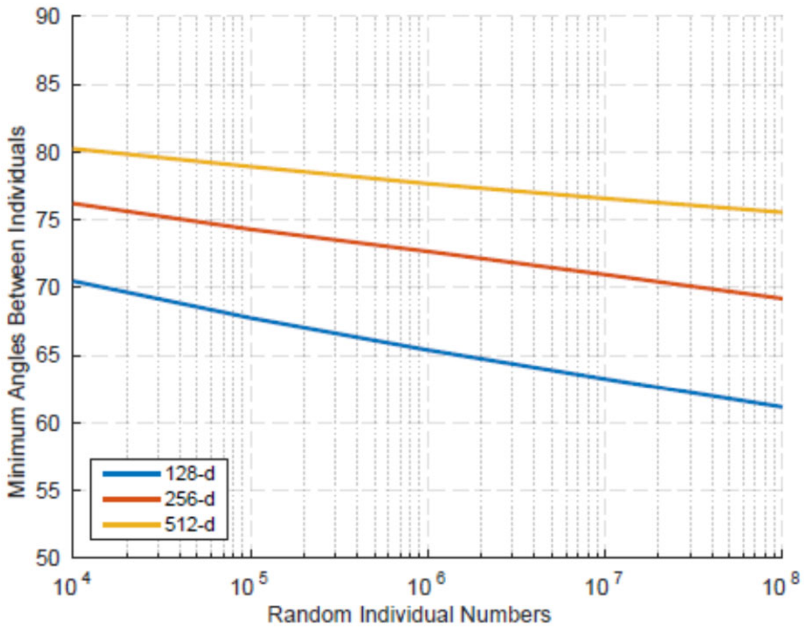


Figure 12. The high-dimensional space is so large that the mean of the nearest angles decreases slowly when the class number increases exponentially.

512-d features easily
Handles 100 million
Person databases

[5] J. S. Brauchart, A. B. Reznikov, E. B. Saff, I. H. Sloan, Y. G. Wang, and R. S. Womersley. Random point sets on the spherehole radii, covering, and separation. *Experimental Mathematics*, 2018. 10

Speculation

- Are hyperspherical embeddings **always** better than non-hyperspherical?
- If so, this could make CNNs considerably easier to understand as we could ignore non-hyperspherical embeddings

Conclusion

- Face recognition performance is now incredibly good
- Need to use temporal information in videos to reduce errors and false alarms
- Detection is the surveillance bottleneck in terms of computation as video is very high resolution (5MP is common now)