Spoofing Attack Detection with Convolutional Neural Networks

MARANESI ANDREA

MUSCATELLO ALESSANDRO

20/06/2022

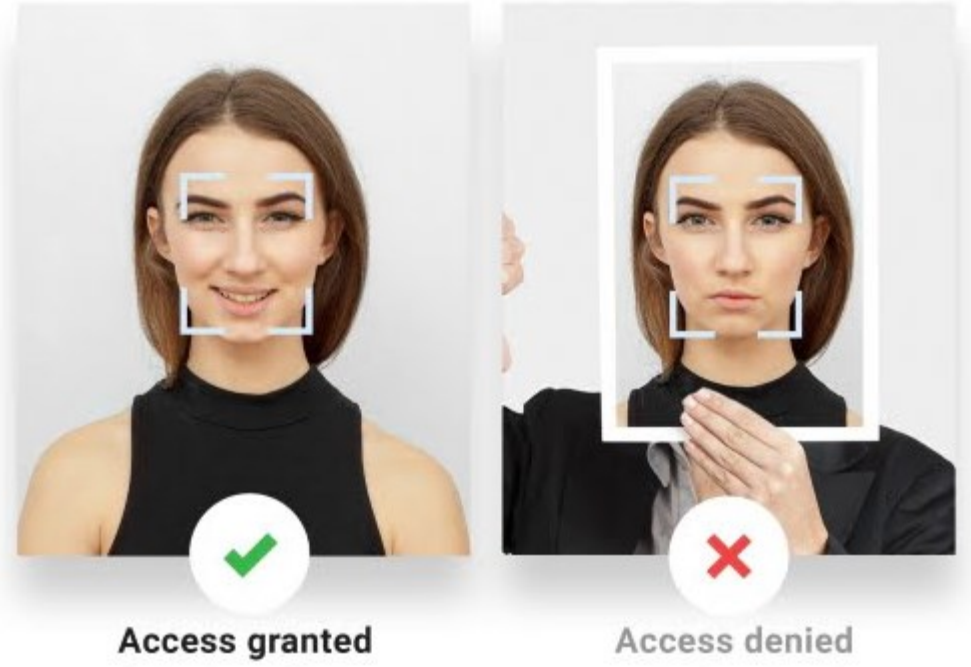
Introduction

Face recognition system are increasingly used since the last few years. They are simple and convenient to use and emerged as a secure approach for authentication.

Already in 2015 Google FaceNet allowed a correct face recognition with an accuracy of 99.96% [1].

Mobile devices often use those kind of system to authenticate users, so it’s important to implement some countermeasures to threat and spoofing-attacks.

Advances in computer graphics and the reduction of costs to produce high resolution masks are emerging as a potential risk that threatens even this countermeasures.



PAD

[1] SCHROFF, FLORIAN; KALENICHENKO, DMITRY; PHILBIN, JAMES. FACENET: A UNIFIED EMBEDDING FOR FACE RECOGNITION AND CLUSTERING. IN: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. 2015. P. 815-823.

State-of-the-art

	Paper	Method	Dataset	Training procedure
[1]	Learn Convolutional Neural Network for Face Anti-Spoofing	Alexet	CASIA Replay-Attack	Pretrained CNN
[2]	A Performance Evaluation of Convolutional Neural Networks for Face Anti Spoofing	Inception-v3, ResNet50, ResNet152	MSU-MFSD	Fine tuning, training from scratch, weight transfer
[3]	Liveness Detection with OpenCV	Self built CNN	Self built	Training from scratch

[1] Yang J, Lei Z, Li SZ (2014) Learn convolutional neural network for face anti-spoofing. CoRR. http://arxiv.org/abs/1408.5601, arXiv:1408.5601
[2] Nagpal C, Dubey SR (2018) A performance evaluation of convolutional neural networks for face anti spoofing. CoRR. https://arxiv.org/abs/1805.04176, arXiv:1805.04176
[3] https://pyimagesearch.com/2019/03/11/liveness-detection-with-opencv/

Why use CNN?

In initial years, the hand designed feature-based approaches were more common and utilized characteristics like texture-based features, motion-based features and depth-based features followed by SVM. Those methods were not efficient and recent works stated that CNN, due to their intrinsic characteristics, have been proven to be very effective for this task.

The CNN architecture already are known for getting very good results in different problems like object detection, semantic segmentation, image classification, biomedical analysis, and there are already several works also on the PAD task with very good results [1].

One significant reason for investing in face-PAD CNN research is the expansion of publicly shared datasets, which facilitates comparison of the performance of new PAD algorithms with existing baseline results.

[1] C. NAGPAL AND S. R. DUBEY, "A PERFORMANCE EVALUATION OF CONVOLUTIONAL NEURAL NETWORKS FOR FACE ANTI SPOOFING," 2019 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019, PP. 1-8, DOI: 10.1109/IJCNN.2019.8852422.

Dataset used – 1/5

- **Our Dataset**: it's made by two sets of images (bona fide and attackers): the first set contains images extracted from videos that shows 'real' people. Those videos are taken with a smartphone or downloaded from internet. The second set contains images extracted from videos taken with a laptop webcam that filmed the playback of the first set of videos on the smartphone screen (or viceversa).

Original frame

Photo of the original

Dataset used – 1/5

Our dataset characteristic	
Dataset dimensions	4656
Class distribution	50 / 50
Training / validation split	48 / 52
Training images	2238
Validation images	2418
Training class distribution	50 / 50
Validation class distribution	50 / 50

Dataset used – 2/5

- **Replay-Attack:** The Replay-Attack Database for face spoofing consists of 1300 video clips of photo and video attack attempts to 50 clients, under different lighting conditions. This Database was produced at the Idiap Research Institute, in Switzerland. After taking the bona fine images from the subjects, 20 attack videos were registered for each client (10 holding the camera 10 placing the camera on a stand). To the camera were shown different types of attacks:
  - 4 x mobile attacks using an image on a 480x320 screen
  - 4 x high-resolution screen attack using an image on a 1024x768 screen
  - 2 x hard-copy prints of image

Bona fide

Low-res attack

High-res attack

Printed

Dataset used – 2/5

Replay-Attack dataset characteristic	
Dataset dimensions	10.804
Class distribution	50 / 50
Training / validation split	82 / 18
Training images	8880
Validation images	1924
Training class distribution	50 / 50
Validation class distribution	50 / 50

Dataset used – 3/5

- **Multispectral-Spoof** contains face images and printed spoofing attacks recorded in Visible (VIS) and Near-Infrared (NIR) spectra for 21 identities. The attacks are made for each subject taking the best 3 VIS and 3 NIR images, printing them on a 600dpi black and white print and showing them on the camera.

The first row are images taken with in VIS, while the second one are images taken in NIR. The first column are real accesses. The second column are VIS attacks. The third column are NIR attacks.

Dataset used – 3/5

MSSPOOF dataset characteristic	
Dataset dimensions	4914
Class distribution	50 / 50
Training / validation split	67 / 33
Training images	3276*
Validation images	1638
Training class distribution	50 / 50
Validation class distribution	50 / 50

*it has been used offline data augmentation to increase the number of the training images

Dataset used – 4/5

- **3DMAD:** The 3D Mask Attack Database (3DMAD) is a biometric (face) spoofing database. It contains 76500 frames of 17 persons, recorded using Kinect for both real-access and spoofing attacks. Attacks are made with a masks of one of the 17 real subject.

Real (RGB)

Attack (RGB)

Attack (IR)

Masks

Dataset used – 4/5

3DMAD dataset characteristic	
Dataset dimensions	2.100
Class distribution	50 / 50
Training / validation split	63 / 37
Training images	1320
Validation images	780
Training class distribution	50 / 50
Validation class distribution	50 / 50

Dataset used – 5/5

- **CSMAD:** The Custom Silicone Mask Attack Dataset (CSMAD) contains presentation attacks made of six custom-made silicone masks. Each mask cost about 4000$. The dataset is designed for face presentation attack detection experiments.

The images in the dataset are in various types:

- Near Infrared (NIR) @ 860nm wavelength
- Thermal (long-wave infrared (LWIR))
- **Depth maps**
- **Color images**



Dataset used – 5/5

CSMAD dataset characteristic	
Dataset dimensions	2448
Class distribution	50 / 50
Training / validation split	66 / 34
Training images	1632
Validation images	816
Training class distribution	50 / 50
Validation class distribution	50 / 50

Our goal

Starting from a published shallow CNN architecture [1], our goal is to create a novel CNN that is able to detect various spoofing attacks made by presenters in form of image presentation via printed images or digital device screens.

Then improve the architecture using the latest tecniques in order to obtain a more accurate model [2].

Finaly adapt the architecture in order to use it with depth data taken from particular cameras, such as Microsoft Kinect camera, and further improve the results.

MAIN RESULTS

We introduced a more lightweight architecture for PAD compared to the most used models that are deep convolutional neural networks.

We improved the generalization performance compared to the original architecture

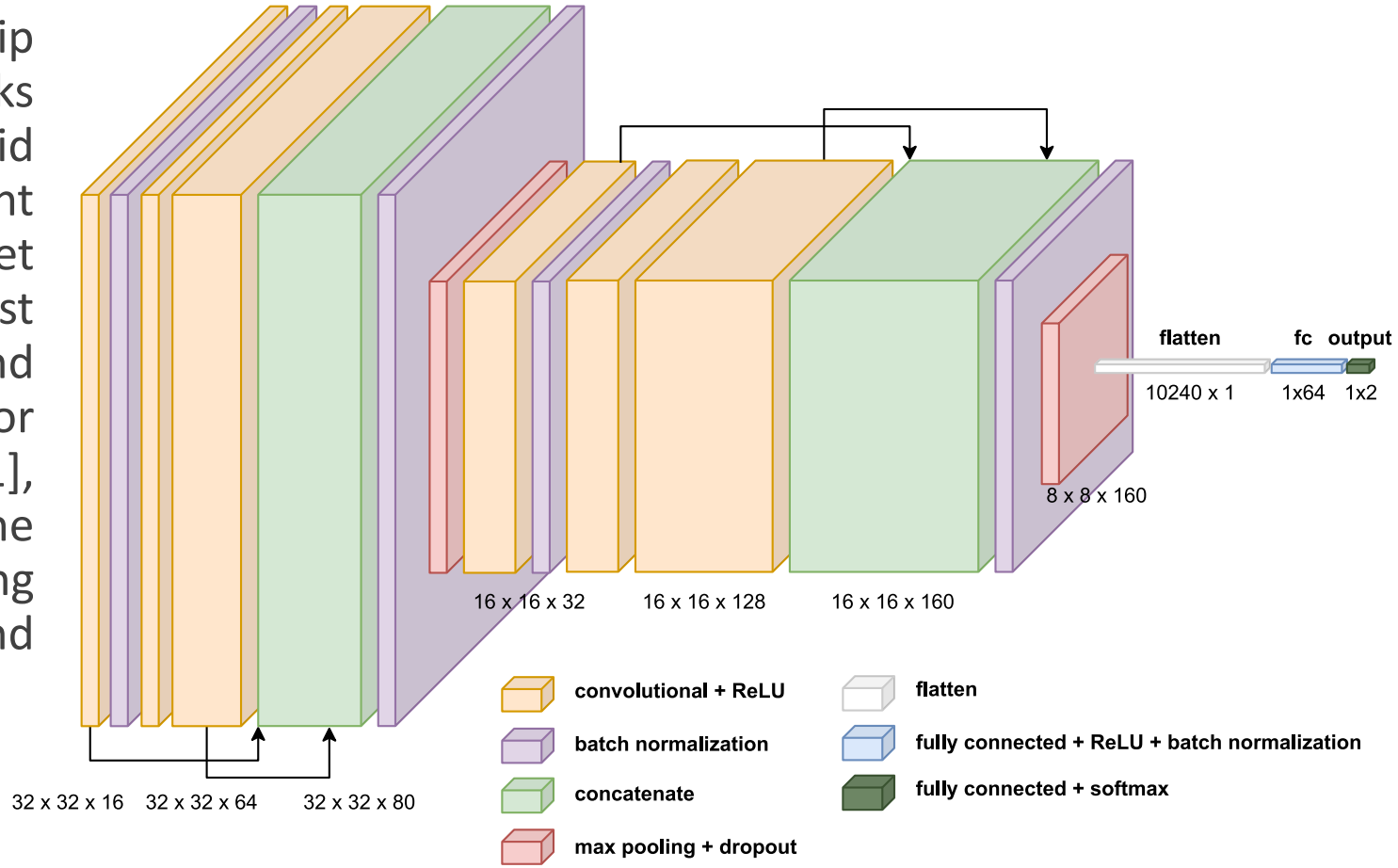
[1] PYIMAGESEARCH ARCHITECTURE: HTTPS://PYIMAGESEARCH.COM/2019/03/11/LIVENESS-DETECTION-WITH-OPENCV/
[2] C. NAGPAL AND S. R. DUBEY, "A PERFORMANCE EVALUATION OF CONVOLUTIONAL NEURAL NETWORKS FOR FACE ANTI SPOOFING," 2019 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019, PP. 1-8, DOI: 10.1109/IJCNN.2019.8852422.

AttackNet

OUR FIRST IMPROVEMENT TO THE NETWORK

Our first network – AttackNet

Since is known that using skip connection with residual blocks are a good practice to avoid explosion or vanishing gradient problem and the ResNet architecture is one of the most used type of architecture, and in fact is particularly suited for the spoofing-detection task [1], we wanted to improve the original implementation adding more convolutional layers and skip connections.



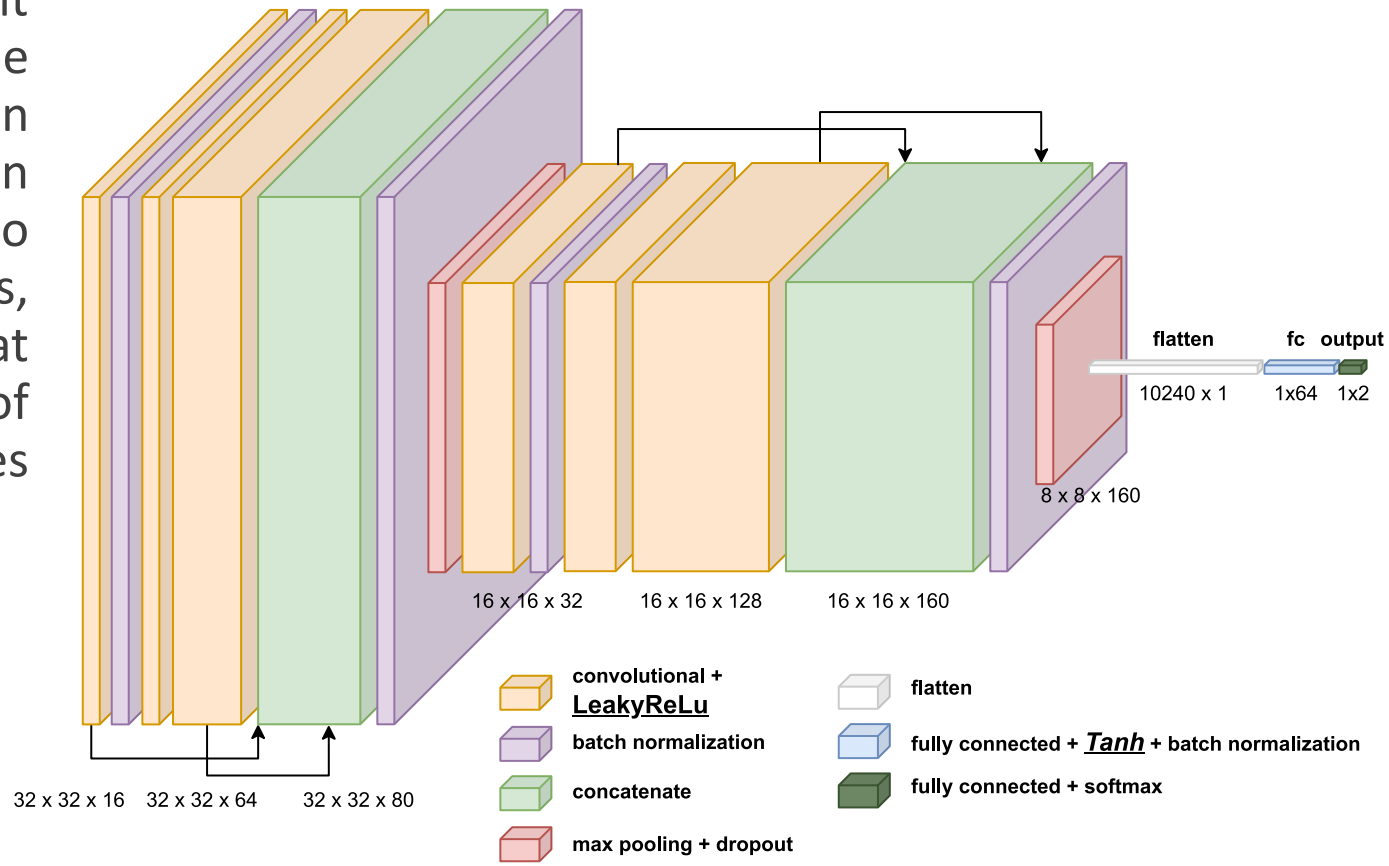
[1] C. NAGPAL AND S. R. DUBEY, "A PERFORMANCE EVALUATION OF CONVOLUTIONAL NEURAL NETWORKS FOR FACE ANTI SPOOFING," 2019 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019, PP. 1-8, DOI: 10.1109/IJCNN.2019.8852422.

AttackNet V2.1

A STEP FURTHER

AttackNet V2.1

After the first improvement we felt that the architecture still could be improved. Using BatchNormalization we already have a normalized input in the next layers, so we decided to adopt different activation functions, such as LeakyReLu and Tanh. In that way we can mitigate the loss of information in case of negative values in input.



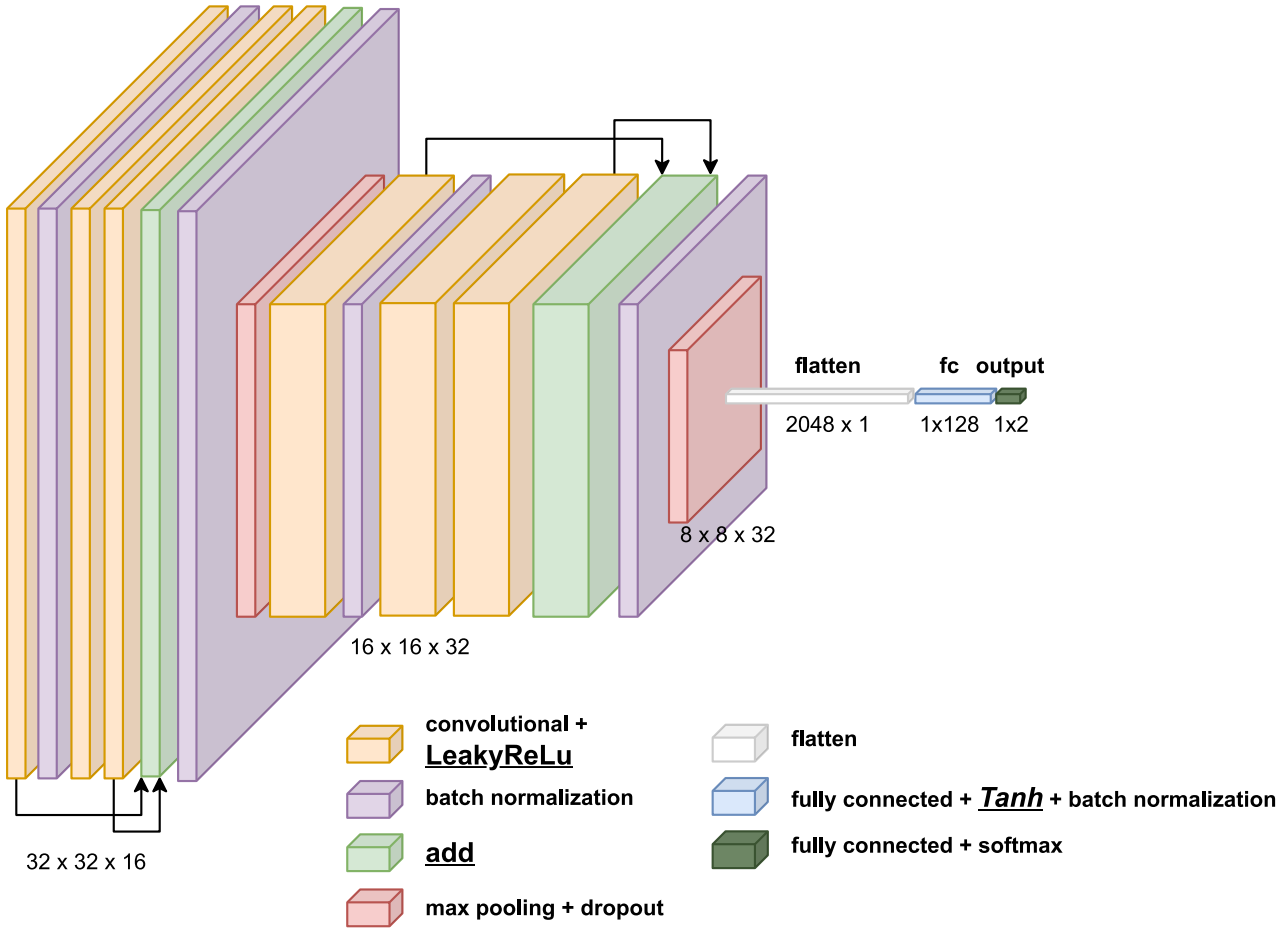
[1] C. NAGPAL AND S. R. DUBEY, "A PERFORMANCE EVALUATION OF CONVOLUTIONAL NEURAL NETWORKS FOR FACE ANTI SPOOFING," 2019 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019, PP. 1-8, DOI: 10.1109/IJCNN.2019.8852422.

AttackNet V2.2

SMALL CHANGES TO THE SKIP CONNECTION IMPLEMENTATION

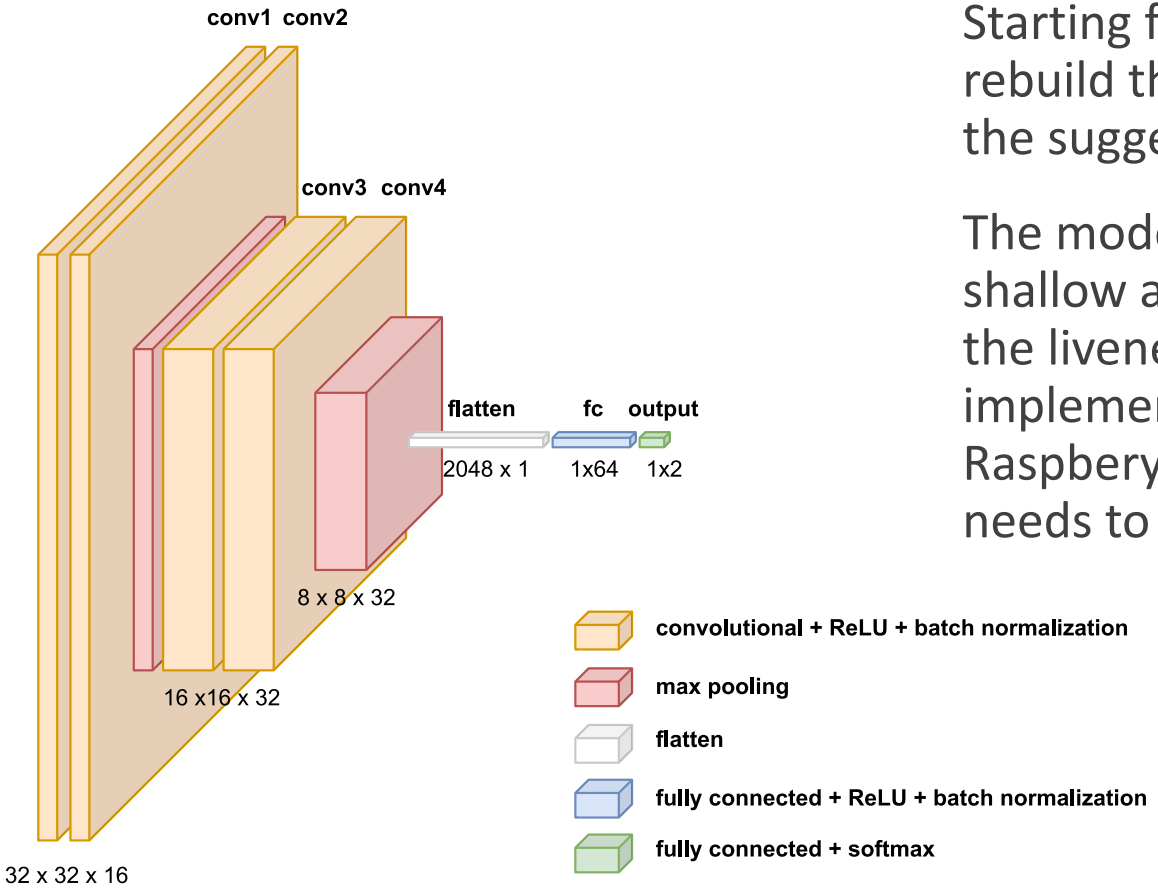
AttackNet V2.2

In this architectural version of our model we only changed the way on how the skip connection are implemented, in particular we utilized the **add** function of Keras instead of concatenate.



[1] C. NAGPAL AND S. R. DUBEY, "A PERFORMANCE EVALUATION OF CONVOLUTIONAL NEURAL NETWORKS FOR FACE ANTI SPOOFING," 2019 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019, PP. 1-8, DOI: 10.1109/IJCNN.2019.8852422.

EXP COMPARISON II



Starting from the PyImageSearch article [1] we rebuild the proposed architecture together with the suggested dataset.

The model presented could be considered a shallow architecture. The original idea was that the liveness recognition was going to be implemented on a low end system (like Raspbery) so the computational power required needs to be minimum.

[1] PYIMAGESEARCH ARCHITECTURE: HTTPS://PYIMAGESEARCH.COM/2019/03/11/LIVENESS-DETECTION-WITH-OPENCV/

Result Comparison

Network	Dataset	Opt.	Mom	LR	Epochs	BS	Acc.	Prec. B/A		Recall B/A		F1 Score	
LivenessNet	Our	SGD	0.5	1e-5	30	16	0.76	0.75	0.83	0.86	0.67	0.79	0.74
AttackNet V1	Our	SGD	0.5	1e-5	30	32	0.80	0.77	0.85	0.87	0.74	0.82	0.79
AttackNet V2.1	Our	Adam	X	1e-6	25	32	0.87	0.8	0.89	0.9	0.77	0.85	0.83
AttackNet V2.1	Replay-Att	Adam	X	1e-4	20	64	**0.96**	**0.99**	0.94	0.93	**0.99**	**0.96**	**0.96**
AttackNet V2.1	Replay-Att (FT)	Adam	X	1e-4	100*	256	**0.95**	**1.0**	**0.95**	**0.94**	**1.0**	**0.97**	**0.97**
LivenessNet	CSMAD (RGB)	SGD	0.5	1e-4	20	32	0.85	0.77	**1.0**	**1.0**	0.7	0.87	0.82
AttackNet V2.2	CSMAD (RGB)	Adam	X	1e-5	20	32	0.87	0.89	0.85	0.85	0.89	0.87	0.87
AttackNet V2.2	CSMAD (depth)	Adam	X	1e-5	32 *	32	0.89	0.88	0.91	0.91	0.87	0.89	0.89
AttackNet V2.2	3DMAD	Adam	X	1e-5	20	16	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
AttackNet V2.2	MS-Spoof	Adam	X	1e-4	15	16	**0.95**	0.92	**0.99**	**0.99**	**0.91**	0.95	0.95

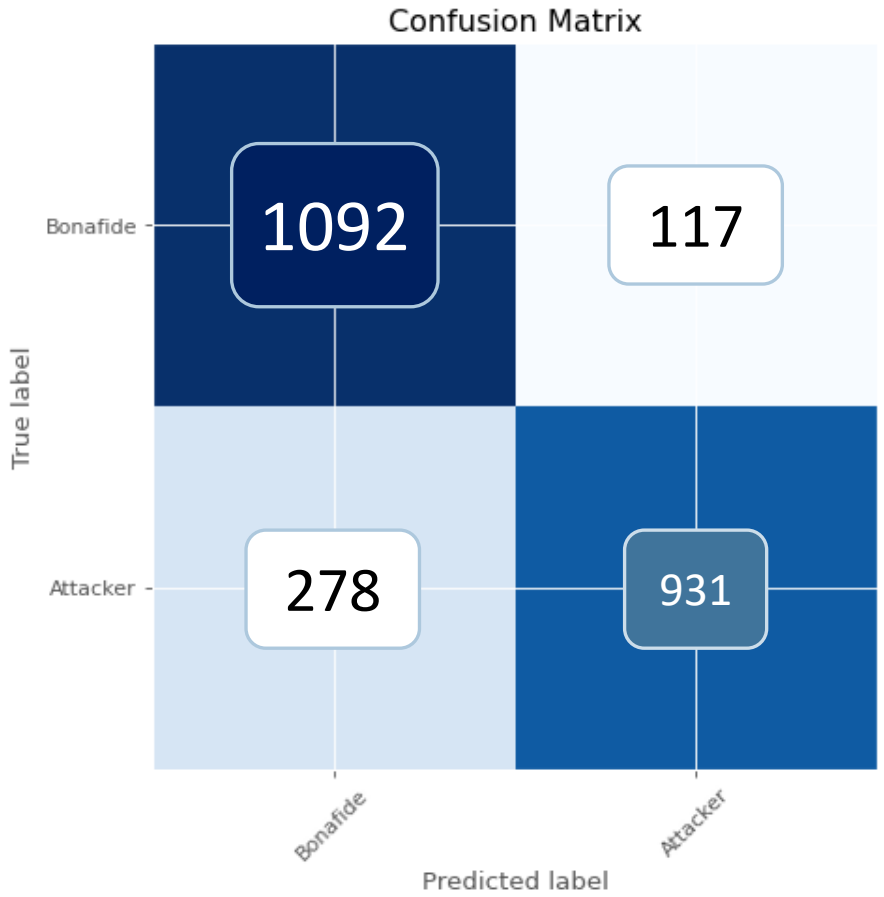
B/A = Bona fide / Attack
* using early stopping

FT = Fine tuning on AttackNet V2.1 on our dataset
RGB = RGB only

depth = using depth as 4° dimension

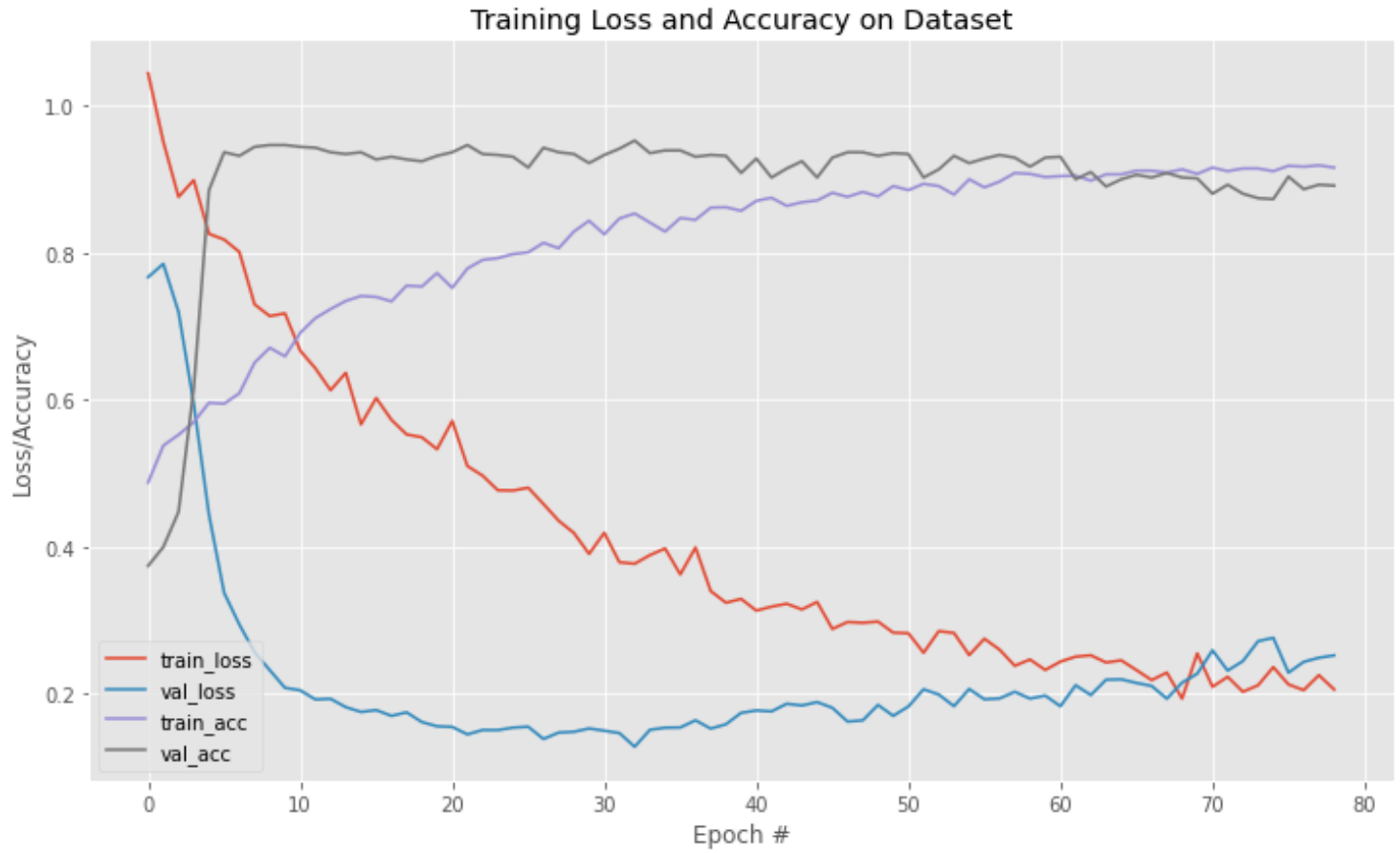
AttackNet V2.1 on our dataset

AttackNet V2.1 vs AttackNet (our dataset)

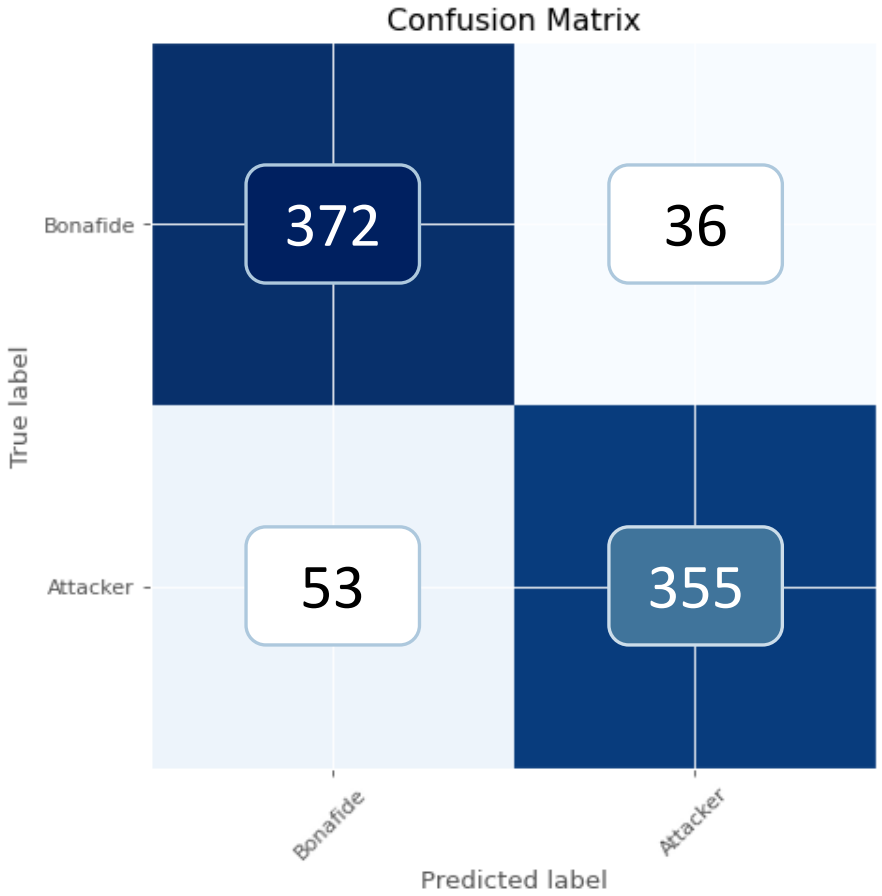


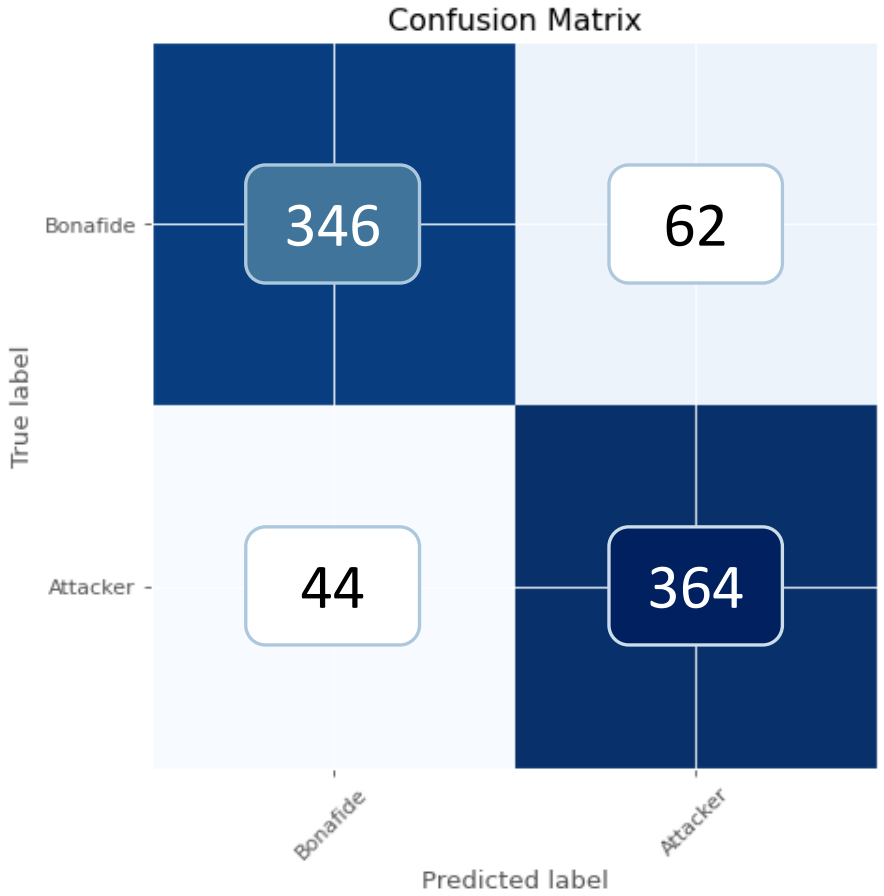
	Bonafide	Attacker
Precision	0.8	0.89
Recall	0.9	0.77
F1 Score	0.85	0.83

AttackNet V2.2 on CSMAD (with depth)



AttackNet V2.2 on CSMAD Depth vs. RGB





AttackNet V2.2 on CSMAD RGB vs. Depth

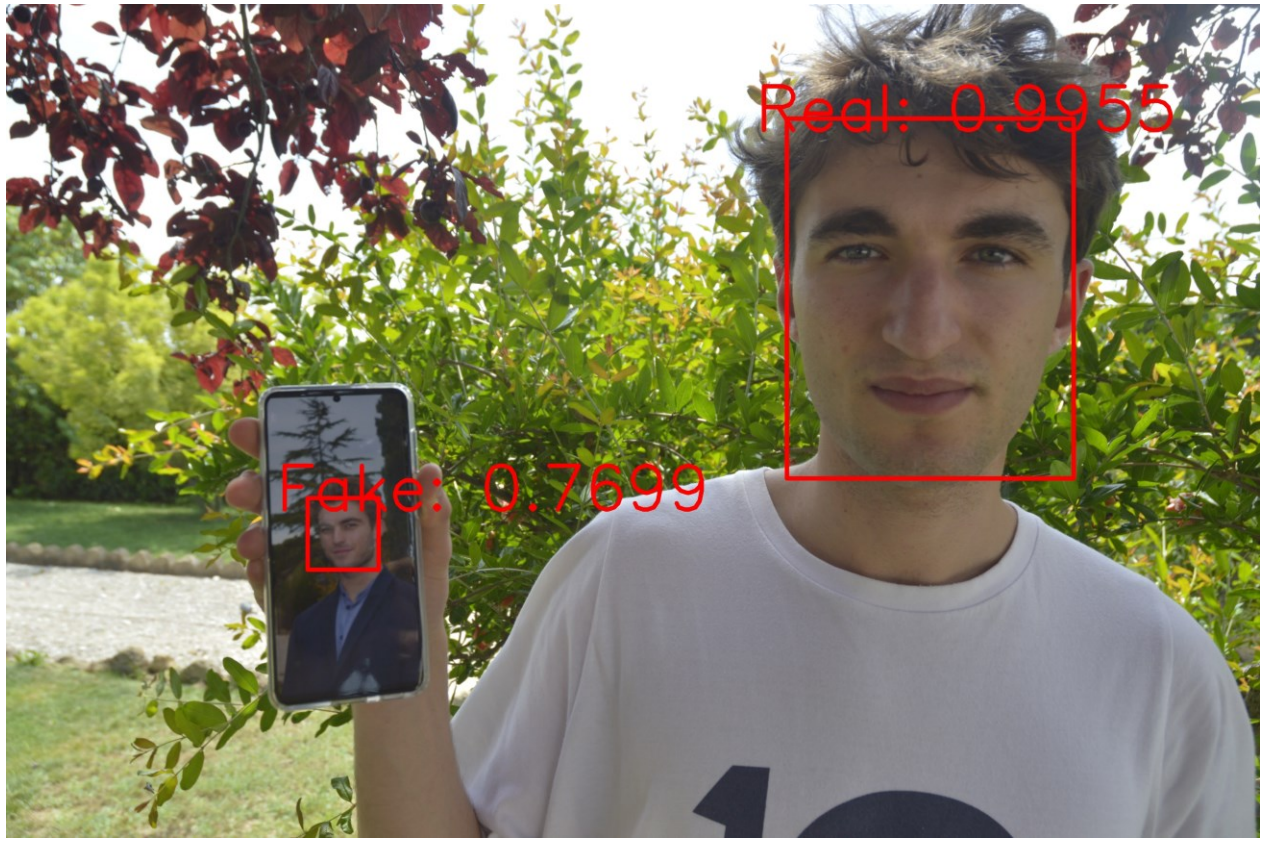
	Bonafide	Attacker
Precision	0.88	**0.91**
Recall	**0.91**	0.87
F1 Score	**0.89**	**0.89**

AttackNet V2.2
CSMAD Depth

	Bonafide	Attacker
Precision	**0.89**	0.85
Recall	0.85	**0.89**
F1 Score	0.87	0.87

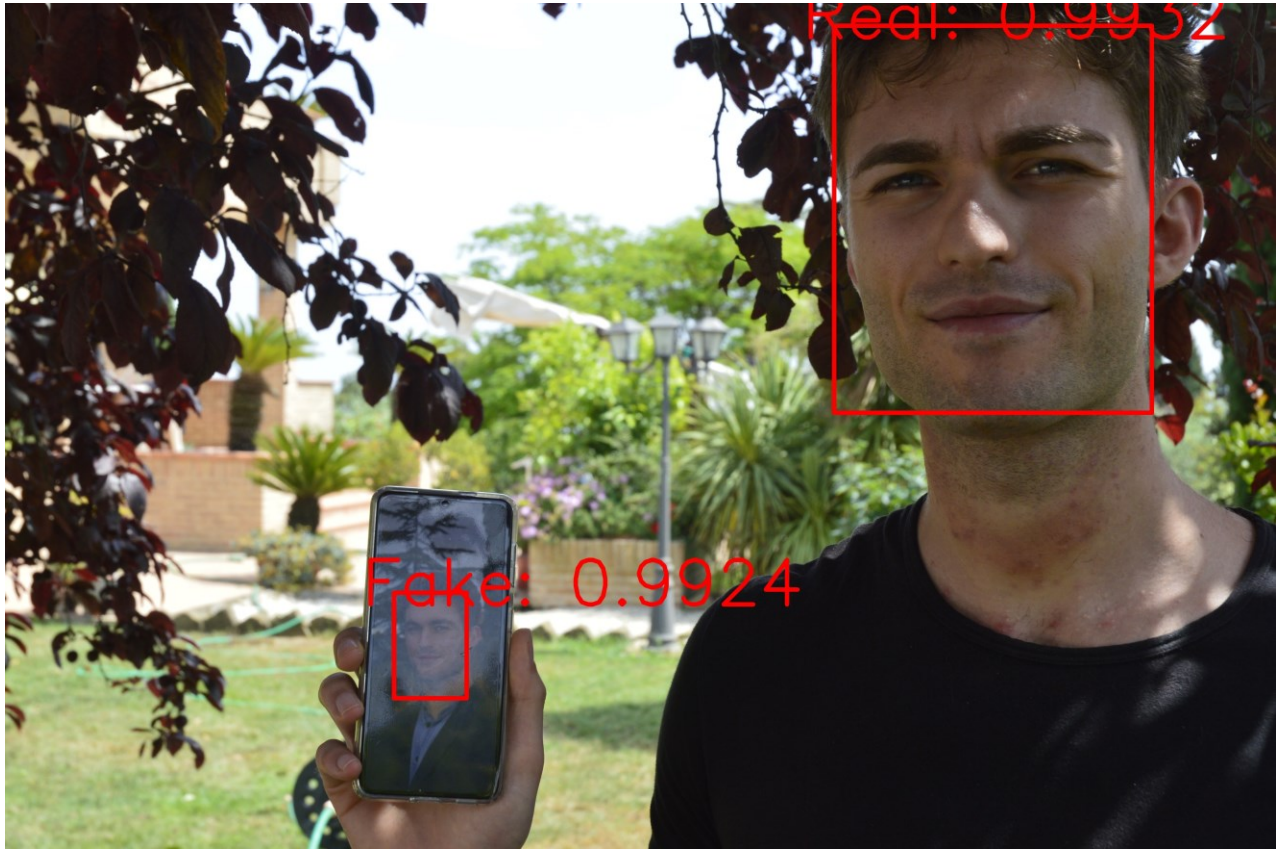
AttackNet V2.2
CSMAD RGB

Real-use case



*faces are extracted with OpenCV's SSD face detector without any tuning and then passed to AttackNet

Real-use case



*faces are extracted with OpenCV's SSD face detector without any tuning and then passed to AttackNet

Conclusions

Conclusion

The new model created is a good improvement over the initial architecture proposed. There aren't many known papers that used that kind of a approach with a total usage of a CNN architecture to obtain an Anti Spoofing Attack system, in fact the not perfect accuracy that the model gets (in all of it's forms) suggest that a mixed approach is recommended.

Evidence from results are that the addition on the fourth layer of the image depth it surely helps the Network to learn important feature.

In general we think that current dataset available are too small to train CNNs from scratch. In fact most of the works that involves using a CNN for Anti Spoofing task are using existing pre trained Face Recognition CNN using transfer learning.

Future work

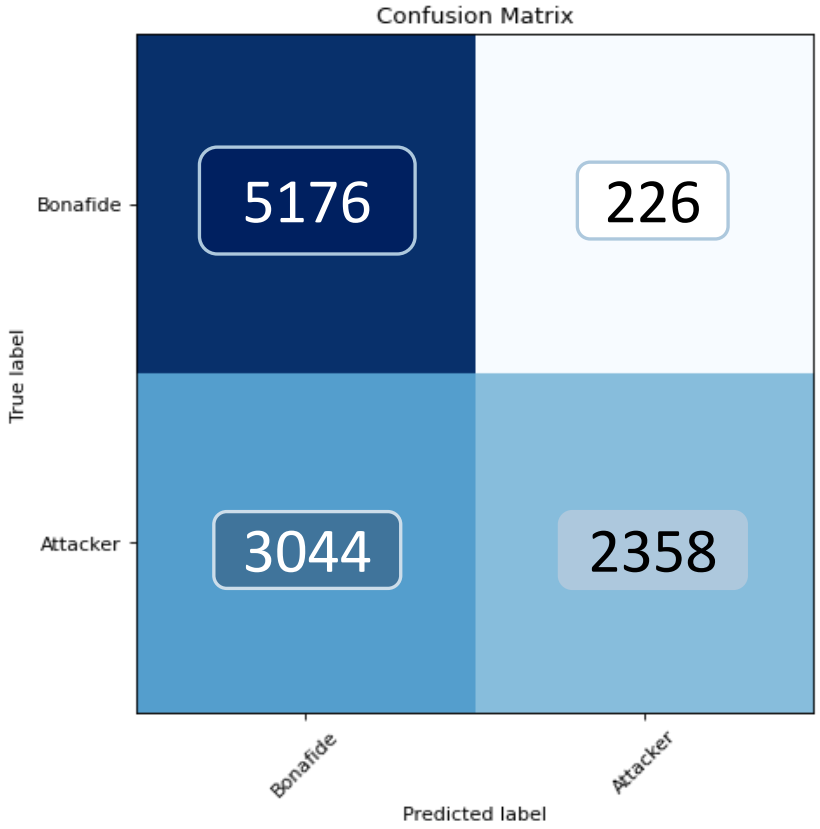
Future work may be to combine or utilize different dataset during the training phase, given the fact that the database shown in this work are old and possibly outdated.

Another interesting point could be to investigate on how the differnt network layers are extracting valuable feature from the images and give meaning to them.

Finally could be important to extend the network with more layers and try to define a new succesful convolutional neural network architecture.

Train and Test on different datasets

Training with AttackNet V2.1 on Our Dataset and Test on Replay-Attack (Acc. 0.71)



	Bonafide	Attacker
Precision	0.63	0.91
Recall	0.96	0.44
F1 Score	0.76	0.59