FACING-2

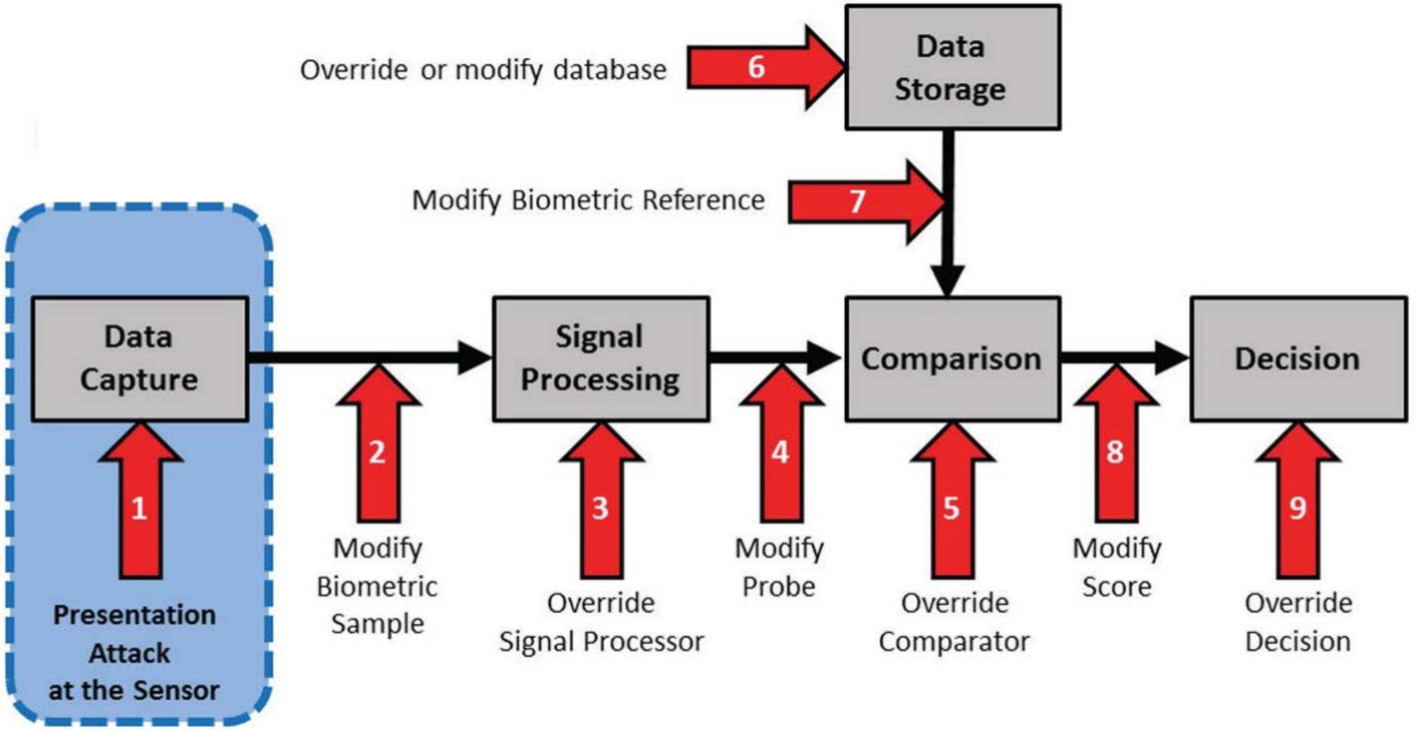
Task - Liveness Detection

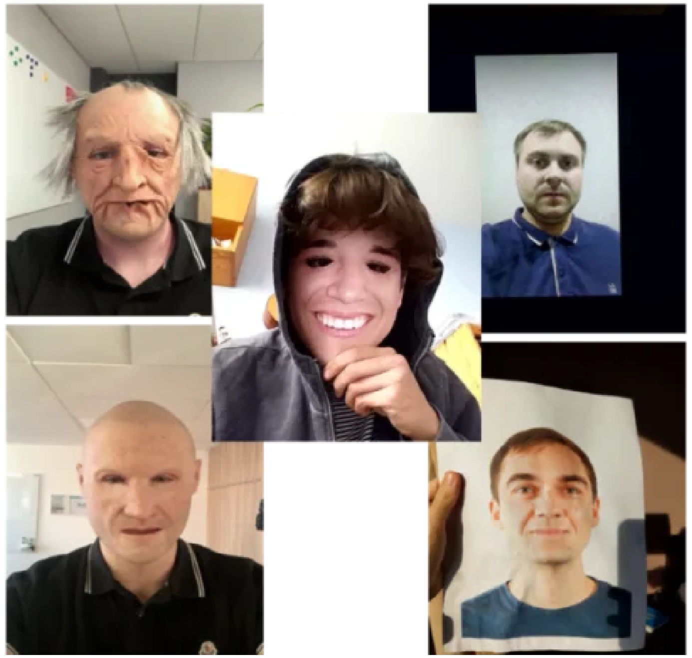
Diogo Nunes – diogo.nunes@isr.uc.pt

Coimbra | 2023-05-16

Overview

- Presentation Attacks (PAs) try to bypass facial recognition systems by impersonation of legitimate users, giving rising to a system security concern





Objectives

- The objective is then to develop a PAD/Liveness detection system capable of differentiate an attack from bonafide presentation, considering as requirements:
  - High portability;
  - Low resource consumption;
  - Based only on RGB images/videos;
  - High generalization performance.

Challenges

- The manifestation of spoof artifacts highly depends on:
  - Capturing device (resolution / distortion / image quality)
  - Illumination
  - Background
  - Spoof instrument and its particularities
    - Specific printer for **print attacks** (InkJet printer / laser printer / photograph printer)
    - Specific display device **for replay attacks** (resolution / distortion / image quality)
    - Mask material for **mask attacks** (paper / silicone / latex)

Challenges

- The problem is that the majority of PAD datasets comprise low variation in the mentioned aspects.
- Resulting in overfit to the training dataset/domain.
- Which leads to the use of techniques/solutions in the perspective of Domain Generalization

Domain Generalization for PAD

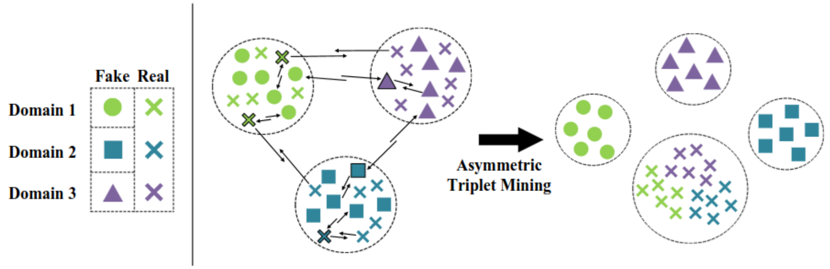
- A common approach is to try to filter out domain specific features (recurring to multiple training domains), this is commonly achieved in two ways:
  - Metric Learning
  - Adversarial Learning

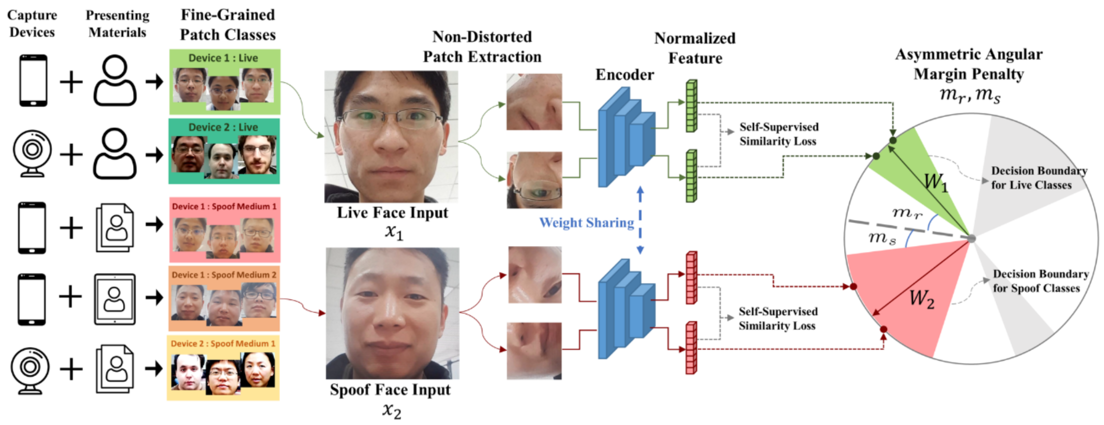
Metric Learning

- Recurs to the use of embedding loss functions to manipulate the feature space and cancel domain specific features.

[2]

[1]



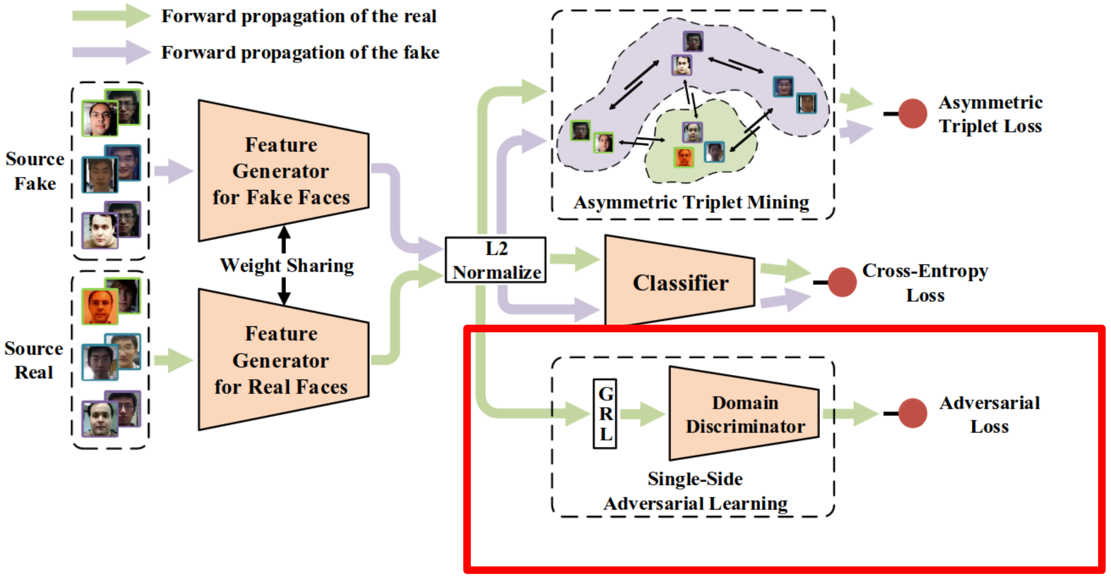


[1] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8484–8493, 2020

[2] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. PatchNet: A simple face anti-spoofing framework via fine-grained patch recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 20281–20290, 2022

Adversarial Learning

- A classifier/discriminator tries to predict the source domain of a given set of features, its success is a cost function for the feature extraction procedure
- In the optimal state, the discriminator is not able to predict the source domain, and thus, the dataset/domain specific features were "eliminated"



[1]

SoTA Results

Method	OCI →M		OMI →C		OCM →I		ICM →O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MMD-AAE [9]	27.08	83.19	44.59	58.29	31.58	75.18	40.98	63.08
MADDG [16]	17.69	88.06	24.50	84.51	22.19	84.99	27.98	80.02
SSDG-M [7]	16.67	90.47	23.11	85.45	18.21	94.61	25.17	81.83
DR-MD-Net [20]	17.02	90.10	19.68	87.43	20.87	86.72	25.02	81.47
RFMeta [17]	13.89	93.98	20.27	88.16	17.30	90.48	16.45	91.16
NAS-FAS [26]	19.53	88.63	16.54	90.18	14.51	93.84	13.80	93.43
D2AM [3]	12.70	95.66	20.98	85.58	15.43	91.22	15.27	90.87
SDA [21]	15.40	91.80	24.50	84.40	15.60	90.10	23.10	84.30
DRDG [11]	12.43	95.81	19.05	88.79	15.56	91.79	15.63	91.75
ANRL [10]	10.83	96.75	17.83	89.26	16.03	91.04	15.67	91.90
SSAN-M [22]	10.42	94.76	16.47	90.81	14.00	94.58	19.51	88.17
SSDG-R [7]	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
SSAN-R [22]	6.67	**98.75**	10.00	**96.67**	8.88	96.79	13.72	93.63
PatchNet [19]	7.10	98.46	11.33	94.58	13.40	95.67	11.82	95.07
SA-FAS [18]	**5.95**	96.55	**8.78**	95.37	**6.58**	**97.54**	**10.00**	**96.23**

However...

- A new article [4] accepted at CVPR2023 states that the results presented before are respective to the epoch that gave the best test performance, independent if it is the first epoch, second, or last.

- Also states, that the test performance should be assessed by the mean and std of the last few epochs, in order to: (1) prevent search biases; (2) reveal unstable training procedures; (3) mimic a realistic scenario where the test domain is not available, even as training stopping criteria.

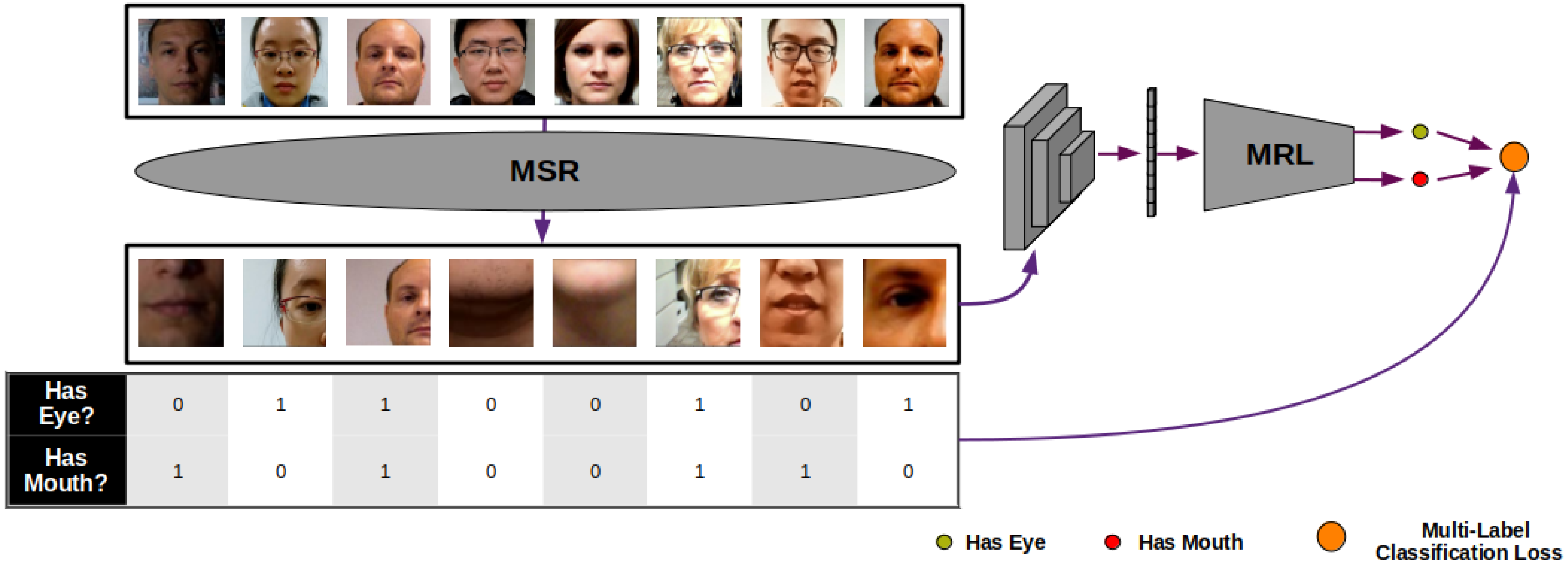
[4] Sun, Yiyou, et al. "Rethinking Domain Generalization for Face Anti-spoofing: Separability and Alignment." arXiv preprint arXiv:2303.13662 (2023).

With a new and fairer comparison setting

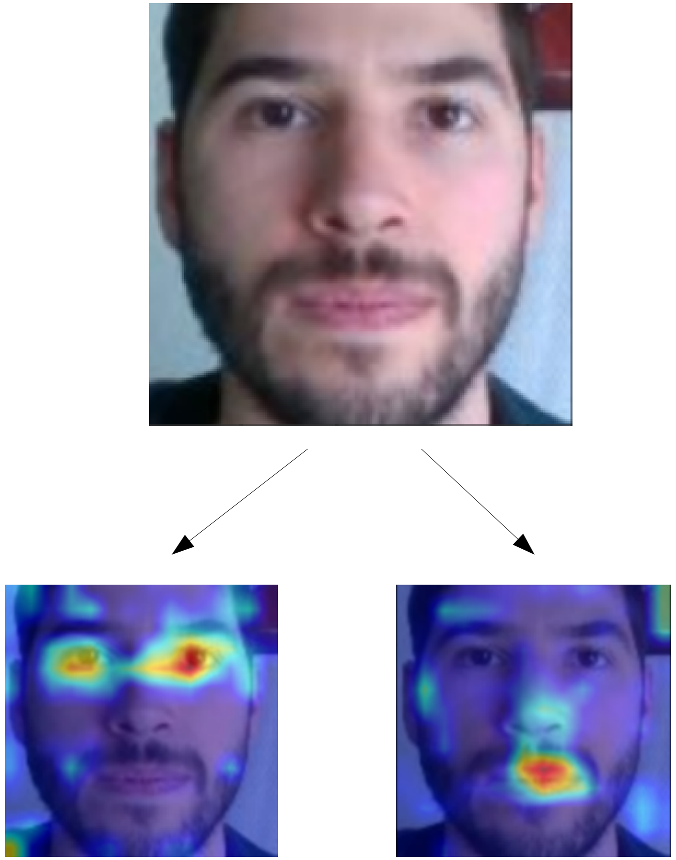
Method	OCI →M HTER/AUC/TPR@FPR=5%	OMI →C HTER/AUC/TPR@FPR=5%	OCM →I HTER/AUC/TPR@FPR=5%	ICM →O HTER/AUC/TPR@FPR=5%
SSDG-R [7]	$14.65^{1.21}$/ $91.93^{1.35}$/ $53.68^{2.56}$	$28.76^{0.89}$/ $80.91^{1.10}$/ $41.47^{2.68}$	$22.84^{1.14}$/ $78.67^{1.31}$/ $50.80^{5.95}$	$15.83^{1.29}$/ $92.13^{0.96}$/ $66.54^{4.00}$
SSAN-R [22]	$21.79^{3.68}$/ $84.06^{3.78}$/ $51.91^{4.28}$	$26.44^{2.91}$/ $78.84^{2.83}$/ $45.36^{4.29}$	$35.39^{8.04}$/ $70.13^{9.03}$/ $64.00^{2.70}$	$25.72^{3.74}$/ $79.37^{4.69}$/ $36.75^{5.19}$
PatchNet [19]	$25.92^{1.13}$/ $83.43^{0.87}$/ $38.75^{8.31}$	$36.26^{1.98}$/ $71.38^{1.89}$/ $19.22^{3.85}$	$29.75^{2.76}$/ $80.53^{1.35}$/ $54.25^{2.18}$	$23.49^{1.80}$/ $84.62^{1.92}$/ $36.39^{6.83}$
SA-FAS [18]	$\mathbf{14.36}^{1.10}$/ $92.06^{0.53}$/ $55.71^{4.82}$	$19.40^{0.66}$/ $88.69^{0.67}$/ $50.53^{3.60}$	$\mathbf{11.48}^{1.10}$/ $\mathbf{95.74}^{0.55}$/ $\mathbf{77.05}^{3.26}$	$\mathbf{11.29}^{0.32}$/ $\mathbf{95.23}^{0.24}$/ $\mathbf{73.38}^{1.64}$

- They found out that domain-invariant techniques cause the training procedure to be highly unstable and lead to a final solution with poor generalization power
- The authors of SA-FAS, on the other hand, encourage the domain separability, and focus on the alignment task, specifically, in the regularization between live-to-spoof transitions and enforcing the same transition direction for all domains.

Approach: Learning Face Regions



Has Eye?	0	1	1	0	0	1	0	1
Has Mouth?	1	0	1	0	0	1	1	0

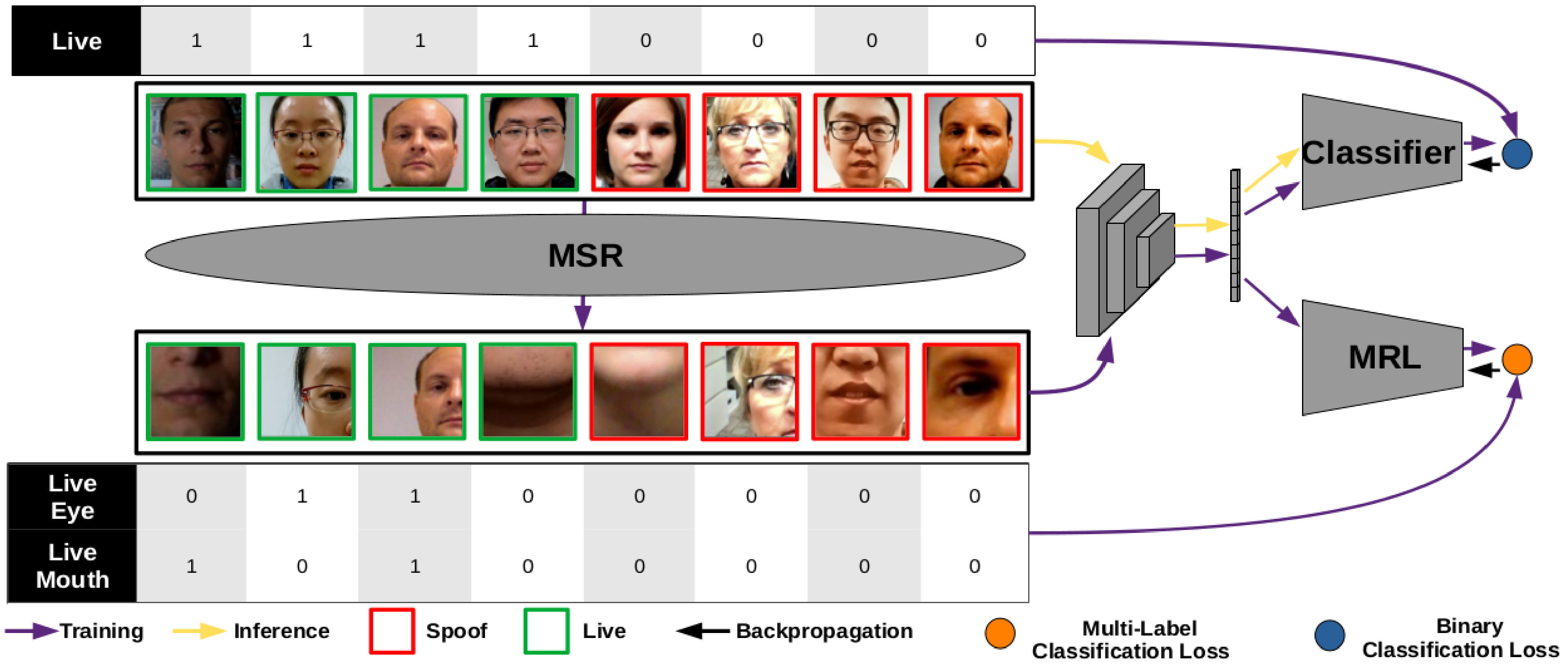


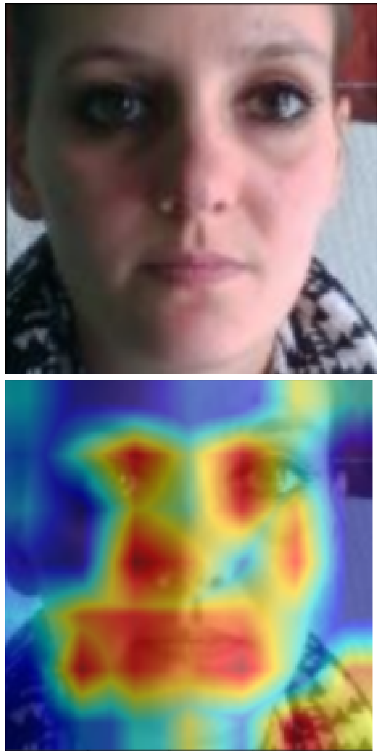
Transition to Liveness Detection:

Is the network able to learn what is a real/bonafide eye or mouth?

- The previous question to the network was if the image contains an eye or a mouth, or both
- **The new question is if the image contains a real eye or mouth**

Our Approach





Results

Method	OCI →M HTER/AUC/TPR@FPR=5%	OMI →C HTER/AUC/TPR@FPR=5%	OCM →I HTER/AUC/TPR@FPR=5%	ICM →O HTER/AUC/TPR@FPR=5%
SSDG-R [7]	$14.65^{1.21}$/ $91.93^{1.35}$/ $53.68^{2.56}$	$28.76^{0.89}$/ $80.91^{1.10}$/ $41.47^{2.68}$	$22.84^{1.14}$/ $78.67^{1.31}$/ $50.80^{5.95}$	$15.83^{1.29}$/ $92.13^{0.96}$/ $66.54^{4.00}$
SSAN-R [22]	$21.79^{3.68}$/ $84.06^{3.78}$/ $51.91^{4.28}$	$26.44^{2.91}$/ $78.84^{2.83}$/ $45.36^{4.29}$	$35.39^{8.04}$/ $70.13^{9.03}$/ $64.00^{2.70}$	$25.72^{3.74}$/ $79.37^{4.69}$/ $36.75^{5.19}$
PatchNet [19]	$25.92^{1.13}$/ $83.43^{0.87}$/ $38.75^{8.31}$	$36.26^{1.98}$/ $71.38^{1.89}$/ $19.22^{3.85}$	$29.75^{2.76}$/ $80.53^{1.35}$/ $54.25^{2.18}$	$23.49^{1.80}$/ $84.62^{1.92}$/ $36.39^{6.83}$
SA-FAS [18]	$\mathbf{14.36}^{1.10}$/ $92.06^{0.53}$/ $55.71^{4.82}$	$19.40^{0.66}$/ $88.69^{0.67}$/ $50.53^{3.60}$	$\mathbf{11.48}^{1.10}$/ $\mathbf{95.74}^{0.55}$/ $\mathbf{77.05}^{3.26}$	$\mathbf{11.29}^{0.32}$/ $\mathbf{95.23}^{0.24}$/ $\mathbf{73.38}^{1.64}$
IFRLL (ours)	$14.82^{0.60}$/ $\mathbf{93.40}^{0.72}$/ $\mathbf{74.31}^{1.01}$	$\mathbf{13.22}^{1.10}$/ $\mathbf{94.64}^{0.46}$/ $\mathbf{71.27}^{3.28}$	$18.53^{1.03}$/ $86.54^{0.48}$/ $66.36^{2.63}$	$20.21^{1.45}$/ $94.69^{0.28}$/ $73.06^{1.81}$

- SoTa improvement in 2 of the four protocols
- Limitations: resolution dependent performance

Next steps

- Analysis on more facial regions
- Exploration of resolution-invariant techniques
- Video-based face region solution

Questions, ideas, suggestions, ...