Learning Deep Models for Face Anti-spoofing via Auxiliary Supervision

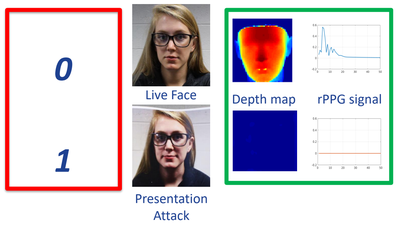
Yaojie Liu*, Amin Jourabloo*, Xiaoming Liu
Department of Computer Science and Engineering, Michigan State University, MI

Computer Vision Lab

CVPR 2018 SALT LAKE CITY • JUNE 18-22

Introduction

✗ **Binary Supervision**



✓ **Auxiliary Supervision**

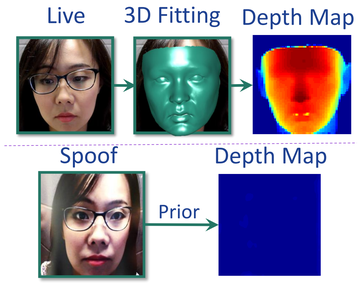
- Simply training a binary classifier leads to overfitting.
- There are auxiliary information that can help.
- We prefer a model that can explain its decision.

Auxiliary Supervisions

➢ Auxiliary spatial supervision: **Depth Map Estimation**

$$\mathbf{S}_F = \mathbf{S}_0 + \sum_{i=1}^{N_{id}} \alpha_{id}^i \mathbf{S}_{id}^i + \sum_{i=1}^{N_{exp}} \alpha_{exp}^i \mathbf{S}_{exp}^i$$

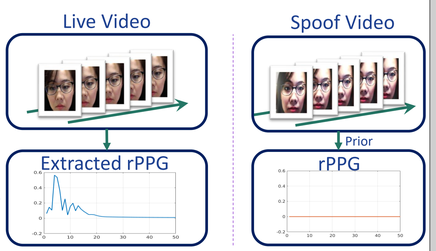
$$\mathbf{S} = s\mathbf{R}\mathbf{S}_F + \mathbf{t}$$



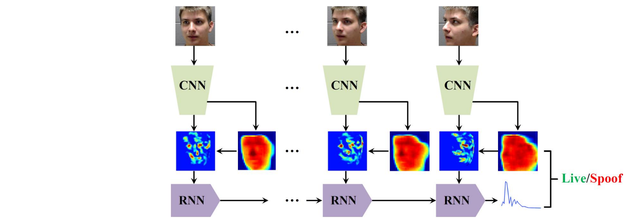
- Normalize the z values of the 3D vertices within [0, 1].
- Apply the Z-Buffer algorithm.

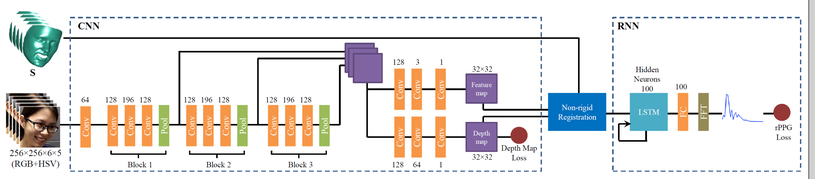
➢ Auxiliary temporal supervision: **rPPG Signal Estimation**

- rPPG signal provides temporal information about liveness of the face.
- Related to changes in the intensities of the face skin over time.
- Intensity changes are highly correlated with the blood flow.



Proposed Method

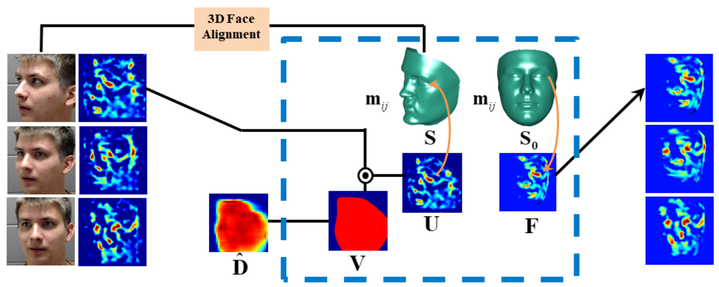




Non-rigid Registration Layer

➢ Non-rigid registration layer

- Register the CNN feature maps for LSTM training.
- Use 3D face shape to do non-rigid registration.
- Differentiable



$$\mathbf{V} = \hat{\mathbf{D}} \geq threshold.$$

$$\mathbf{F}(i,j) = \mathbf{U}(\mathbf{S}(\mathbf{m}_{ij},1), \mathbf{S}(\mathbf{m}_{ij},2))$$

➢ Classification score

- Compute the estimated depth map
- Compute the estimated rPPG signal

$$score = ||\hat{\mathbf{f}}||_2^2 + \lambda||\hat{\mathbf{D}}||_2^2$$

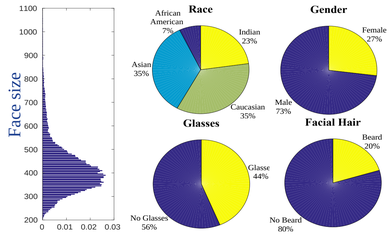
➢ Demo systems

- 3 FPS on phone
- 26 FPS on laptop

Experimental Results

➢ Spoofing in the Wild (SiW) Databases

- Contain 165 subjects
- Pose range from -90° to 90°
- Include print, replay, funny eye, and mannequin head attack
- Lighting variation
- Scene variation



Protocol	Subset	Subject #	Attack	APCER	BPCER	ACER
Various face pose and expression	Train	90	First 60 Frames	3.58%	3.58%	3.58%
	Test	75	All			
Cross spoof medium	Train	90	3 display	0.57±0.69%	0.57±0.69%	0.57±0.69%
	Test	75	1 display			
Cross PA	Train	90	print (display)	8.31±3.81%	8.31±3.80%	8.31±3.81%
	Test	75	display (print)			

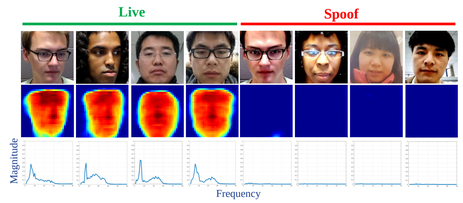
➢ Intra-test on Oulu

Protocol	Method	APCER	BPCER	ACER
Various illumination conditions	CPqD	2.9%	10.8%	6.9%
	GRADIANT	1.3%	12.5%	6.9%
	Proposed method	**1.6%**	**1.6%**	**1.6%**
Different spoof medium	MixedFASNet	9.7%	2.5%	6.1%
	Proposed method	**2.7%**	**2.7%**	**2.7%**
	GRADIANT	3.1%	1.9%	2.5%
Different camera devices	MixedFASNet	5.3±6.7%	7.8±5.5%	6.5±4.6%
	GRADIANT	2.6±3.9%	5.0±5.3%	3.8±2.4%
	Proposed method	**2.7±1.3%**	**3.1±1.7%**	**2.9±1.5%**
All above challenges	Massy_HNU	35.8±35.3	8.3±4.1%	22.1±17.6%
	GRADIANT	5.0±4.5%	15.0±7.1%	10.0±5.0%
	Proposed method	**9.3±5.6%**	**10.4±6.0%**	**9.5±6.0%**

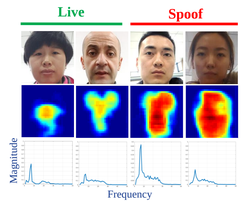
➢ Cross-test on CASIA and IDIAP

Method	Train	Test	Train	Test
	CASIA MFSD	Replay Attack	Replay Attack	CASIA MFSD
Spectral cubes	34.4%		50.0%	
CNN	48.5%		45.5%	
LBP	47.0%		39.6%	
Color Texture	30.3%		37.7%	
Proposed method	**27.6%**		**28.4%**	

➢ Successful cases



➢ Failure cases



➢ Acknowledgement

- This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017-17020200004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S.

* denotes equal contribution by the authors.