



A Dataset and Benchmark for Large-scale Multi-modal Face Anti-spoofing

Shifeng Zhang^{1*}, Xiaobo Wang^{2*}, Ajian Liu³, Chenxu Zhao², Jun Wan¹, Sergio Escalera⁴, Hailin Shi², Zezheng Wang⁵, Stan Z. Li¹

¹CBSR, NLPR, CASIA, UCAS ²JD AI Research ³MUST ⁴CVC, UB ⁵JD Finance



LONG BEACH
CALIFORNIA
June 16-20, 2019

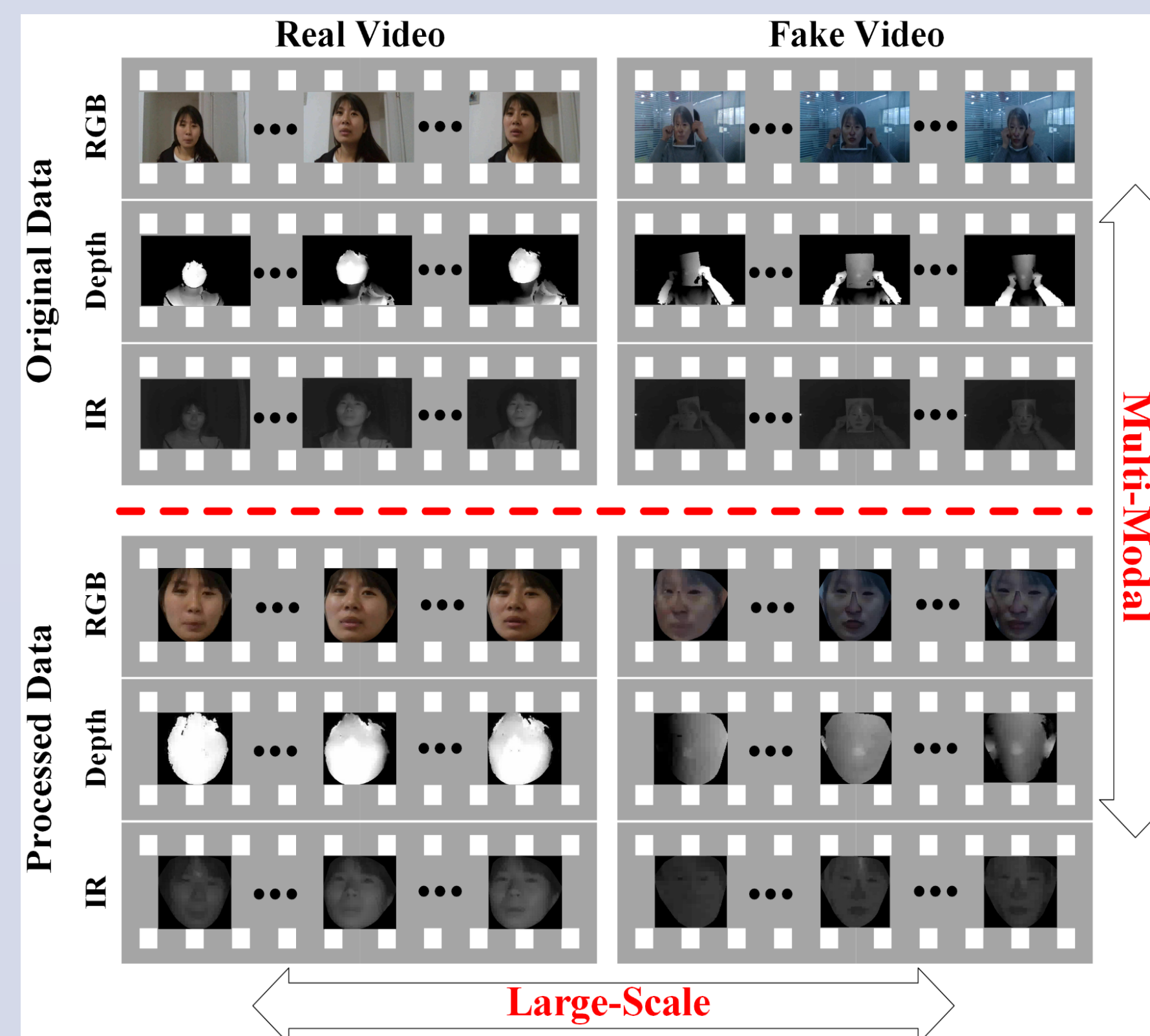
Motivation

Dataset	Year	# subjects	# videos	Camera	Modal	Attack
Replay-Attack	2012	50	1200	VIS	RGB	Print, 2 Replay
CASIA-MFSD	2012	50	600	VIS	RGB	Print, Replay
3DMAD	2013	17	255	VIS/Kinect	RGB/Depth	3D Mask
MSU-MFSD	2015	35	440	Phone/Laptop	RGB	Print, 2 Replay
Replay-Mobile	2016	40	1030	VIS	RGB	Print, Replay
Msspoof	2016	21	4704	VIS/NIR	RGB/IR	Print
Oulu-NPU	2017	55	5940	VIS	RGB	2 Print, 2 Replay
SiW	2018	165	4620	VIS	RGB	2 Print, 4 Replay
Ours	2018	1000	21000	RealSense	RGB/Depth/IR	Print, Cut

- Existing face anti-spoofing benchmark datasets have limited number of subjects (≤ 170) and modalities (≤ 2), hindering the further development.

Contribution

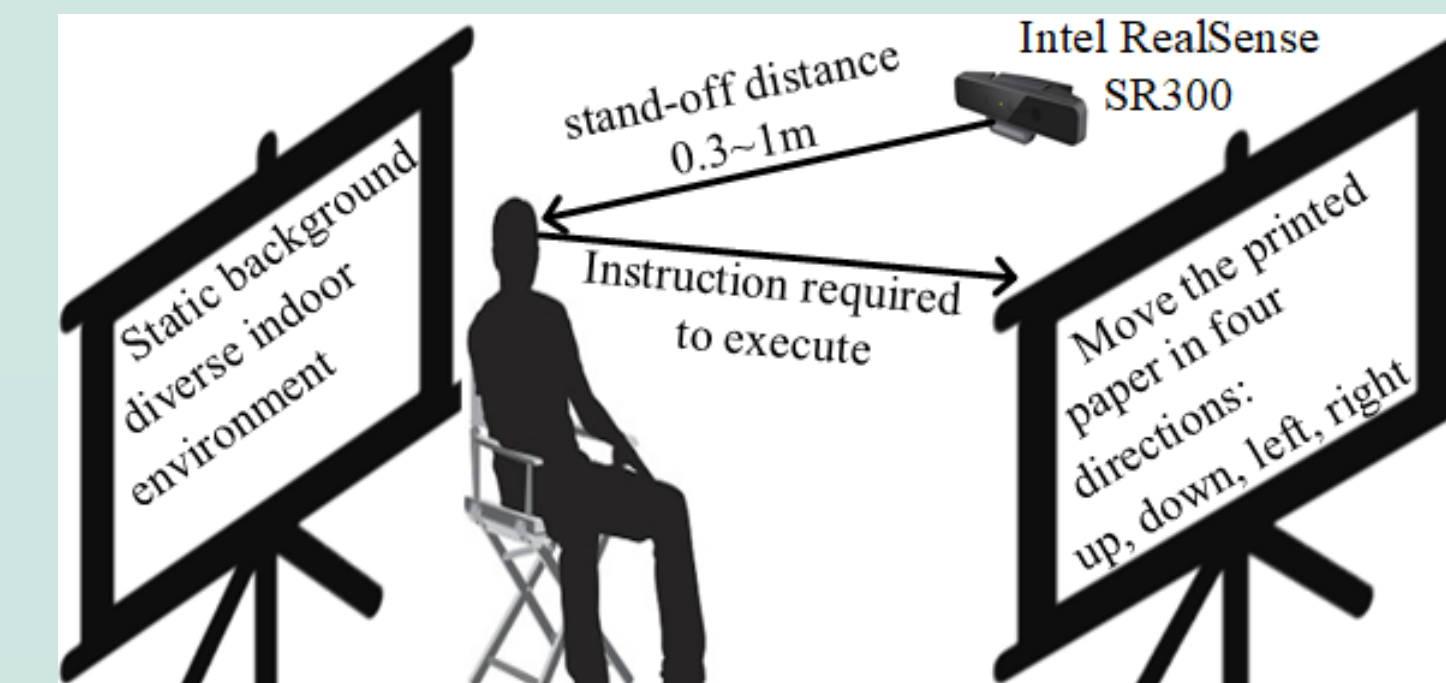
- Present a **large-scale** (1000 subjects) **multi-model** (3 modalities) face anti-spoofing dataset.
- Introduce a new multi-modal fusion method to merge three modalities.
- Conduct extensive experiments to verify its significance and generalization capability.



CASIA-SURF Dataset

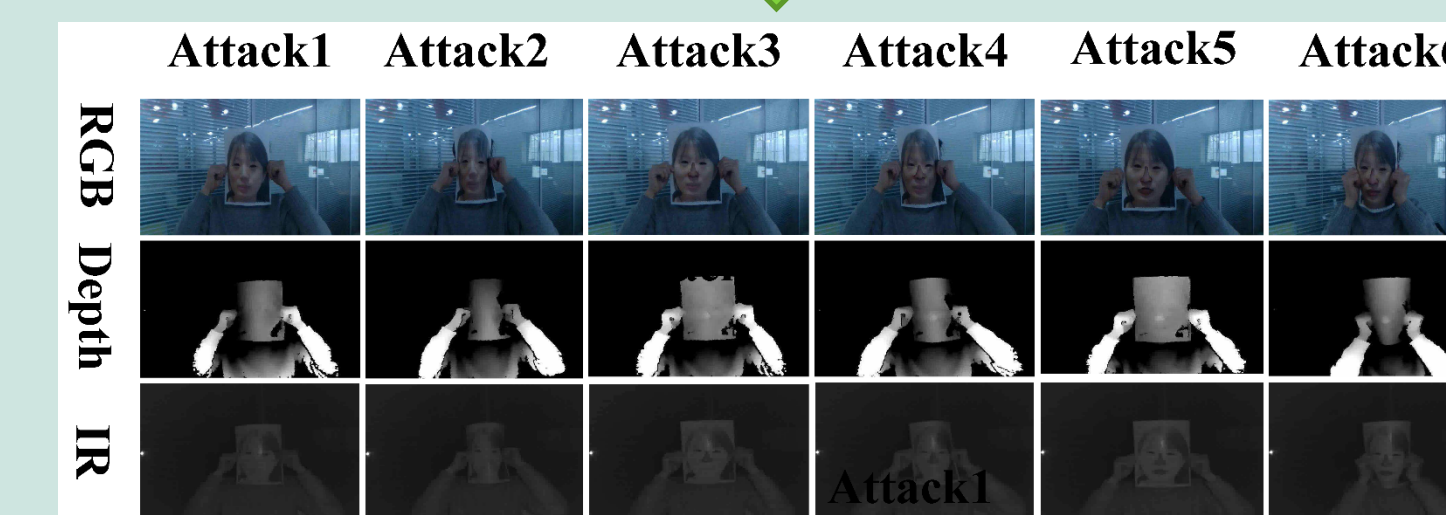
Acquisition Details

- Intel RealSense SR300 camera for RGB, Depth and Infrared (IR)
- Face angle: $\leq 30^\circ$
- Face distance: 0.3~1.0 m



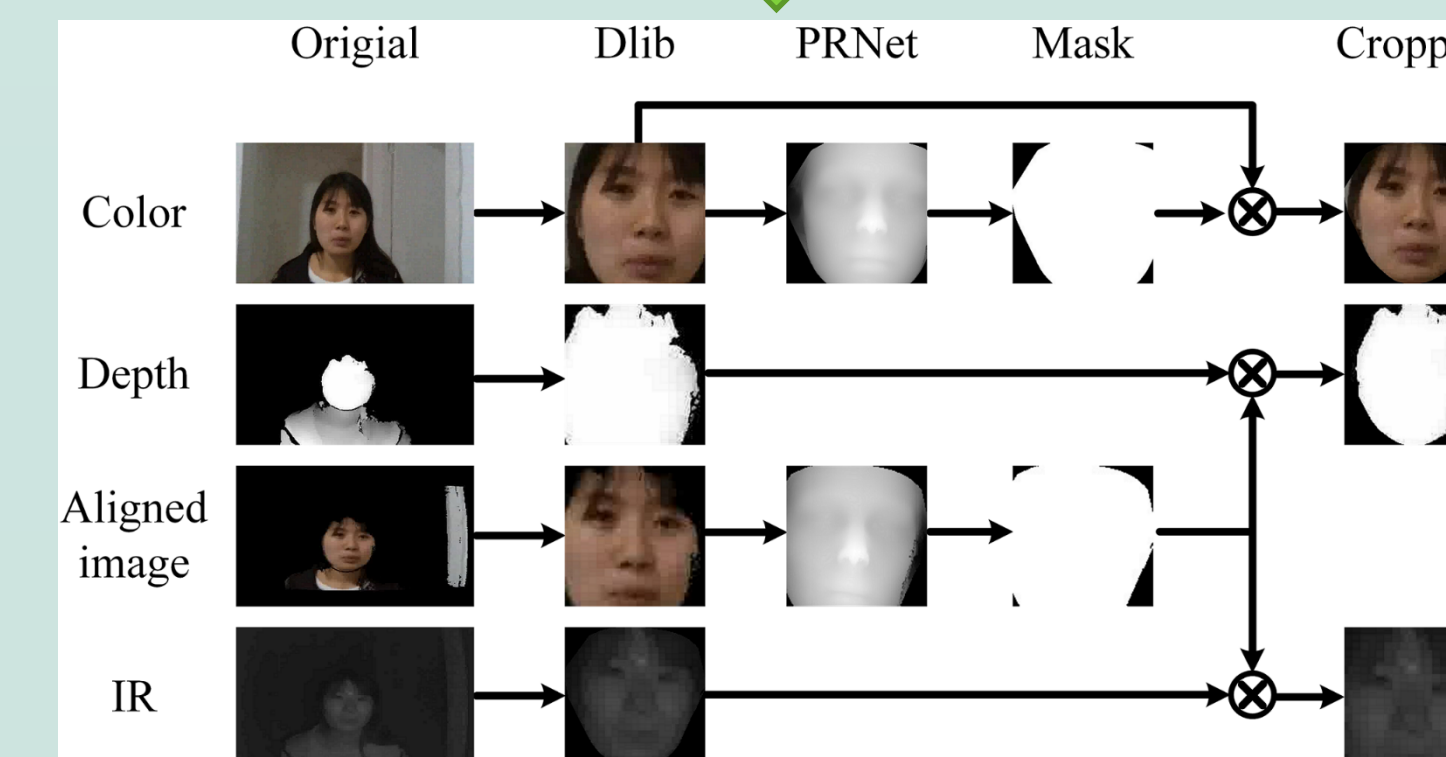
Six Attacks

- 2 Prints: flat and curved face
- 3 Cuts: eyes/nose/mouth region



Data Preprocessing

- Input original image to Dlib tool for face detection
- Apply PRNet to estimate face binary mask
- Get face area by pointwise product



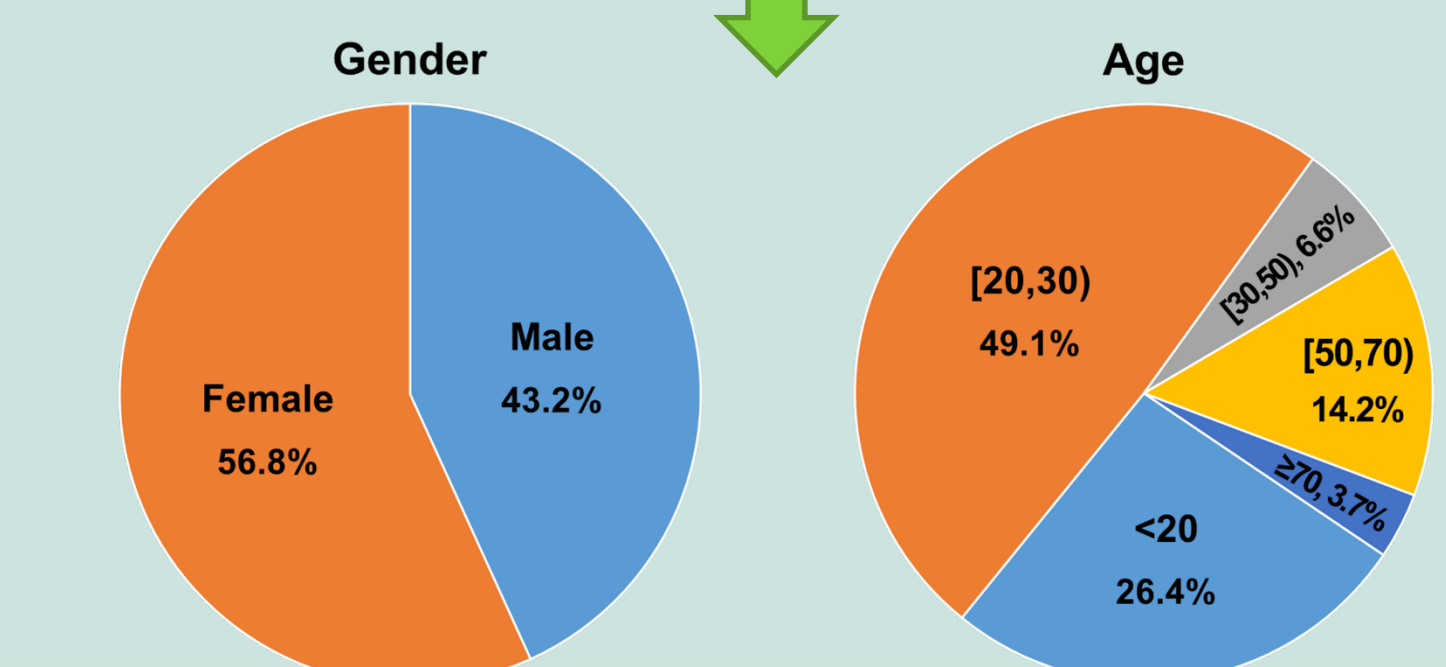
	Training	Validation	Testing	Total
#Obj.	300	100	600	1000
#Videos	6,300	2,100	12,600	21000
#Ori. img.	1,563,919	501,886	3,109,985	5,175,790
#Samp. img.	151,635	49,770	302,559	503,964
#Crop. img.	148,089	48,789	295,644	492522

Subset Dividing

- Split into train/val/test randomly

Statistics

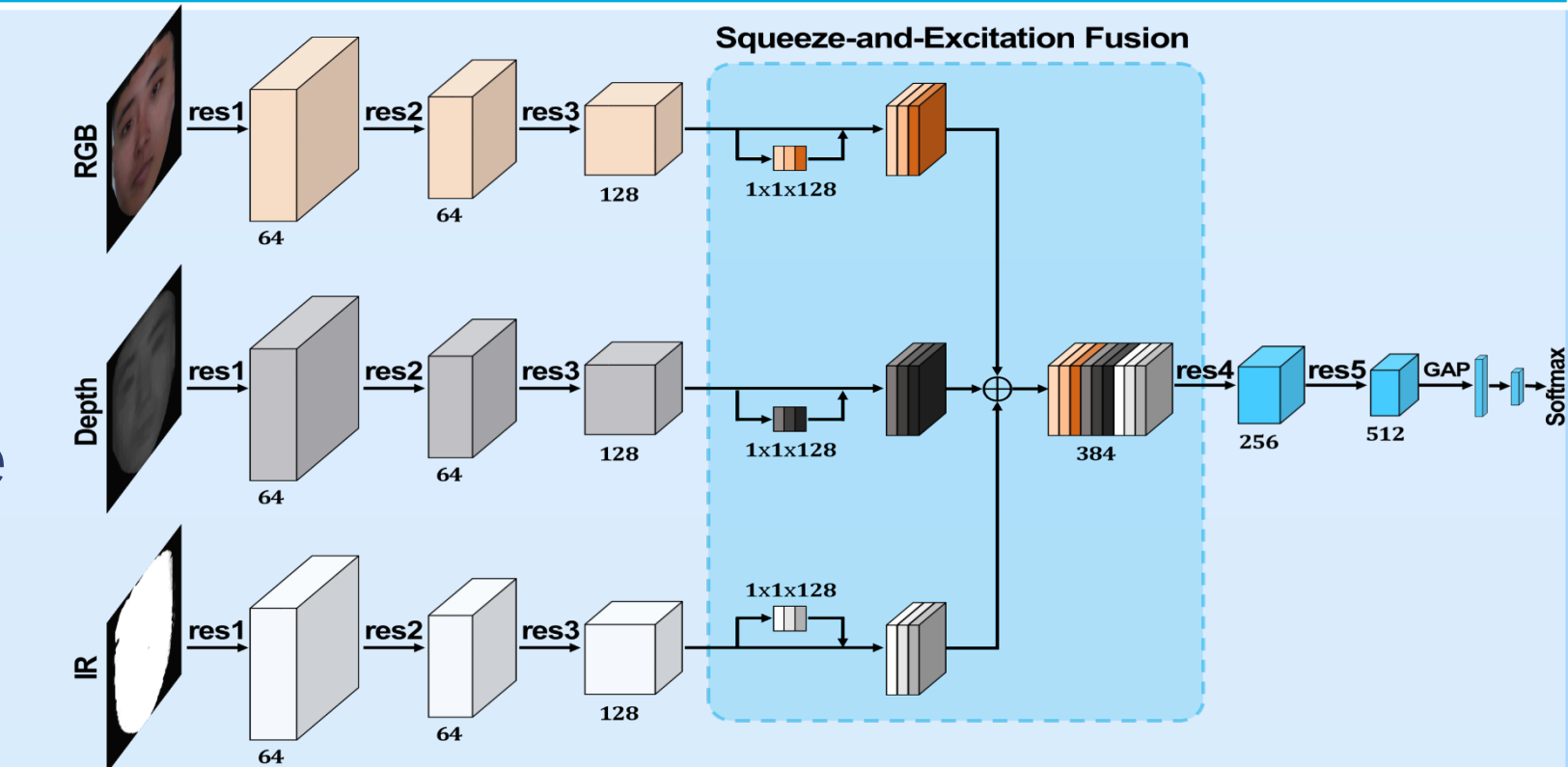
- 1000 subjects, 21000 videos, 492522 images, 3 modalities
- Balanced male to female ratio
- Reasonable age distribution



Experiments

Baseline: SEFusion

- Performs feature re-weighting to select more informative channels while suppressing less useful ones for each modal



Model Analysis

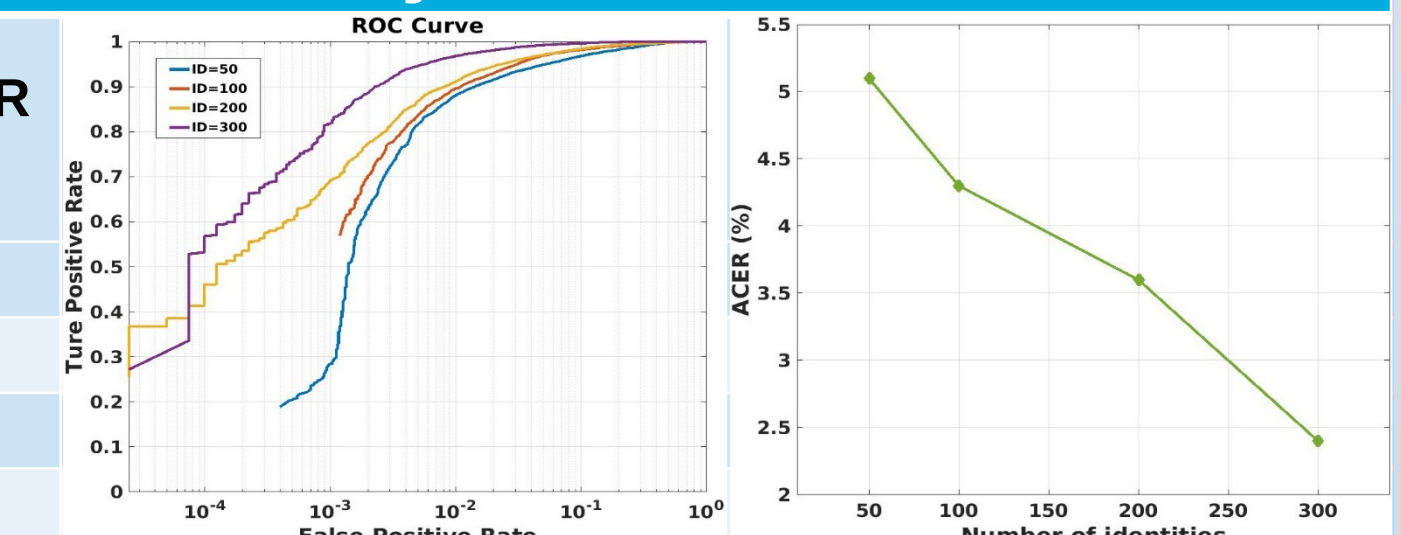
Method	TRP (%)			APCER (%)	NPCER (%)	ACER (%)
	@FPR=10 ⁻²	@FPR=10 ⁻³	@FPR=10 ⁻⁴			
Halfway fusion	89.1	33.6	17.8	5.6	3.8	4.7
Proposed fusion	96.7	81.8	56.8	3.8	1.0	2.4

Effect of Number Modalities

Modal	TRP (%)			APCER (%)	NPCER (%)	ACER (%)
	@FPR=10 ⁻²	@FPR=10 ⁻³	@FPR=10 ⁻⁴			
RGB	49.3	16.6	6.8	8.0	14.5	11.3
Depth	88.3	27.2	14.1	5.1	4.8	5.0
IR	65.3	26.5	10.9	15.0	1.2	8.1
RGB & Depth	86.1	49.5	10.6	4.3	5.6	5.0
RGB & IR	79.1	50.9	26.1	14.4	1.6	8.0
Depth & IR	89.7	71.4	24.3	1.5	8.4	4.9
RGB & Depth & IR	96.7	81.8	56.8	3.8	1.0	2.4

Effect of Number Subjects

# ID	TRP (%)			ACER (%)
	@FPR=10 ⁻²	@FPR=10 ⁻³	@FPR=10 ⁻⁴	
50	88.0	28.5	0	5.1
100	89.5	56.9	0	4.3
200	91.1	69.3	46.1	3.6
300	96.7	81.8	56.8	2.4



Generalization Capability

Prot.	Method	APCER (%)	BPCER (%)	ACER (%)	CASIA-MFSD dataset	Method	Training	Testing	HTER (%)
1	FAS-BAS	3.58	3.58	3.58		Motion	Replay-Attack	CASIA-MFSD	47.9
	FAS-TD-SF	1.27	0.83	1.05		LBP	Replay-Attack	CASIA-MFSD	57.6
	FAS-TD-SF-CASIA-SURF	1.27	0.33	0.80		Motion-Mag	Replay-Attack	CASIA-MFSD	47.0
2	FAS-BAS	0.57±0.69	0.57±0.69	0.57±0.69		Spectral cubes	Replay-Attack	CASIA-MFSD	50.0
	FAS-TD-SF	0.33±0.27	0.29±0.39	0.31±0.28		CNN	Replay-Attack	CASIA-MFSD	45.5
	FAS-TD-SF-CASIA-SURF	0.08±0.17	0.25±0.22	0.17±0.16		FAS-TD-SF	SiW	CASIA-MFSD	39.4
3	FAS-BAS	8.31±3.81	8.31±3.80	8.31±3.81		FAS-TD-SF	CASIA-SURF	CASIA-MFSD	37.3
	FAS-TD-SF	7.70±3.88	7.76±4.09	7.73±3.99					
	FAS-TD-SF-CASIA-SURF	6.27±4.36	6.43±4.42	6.35±4.39					