Face Anti-Spoofing with Out-of-distribution Detection

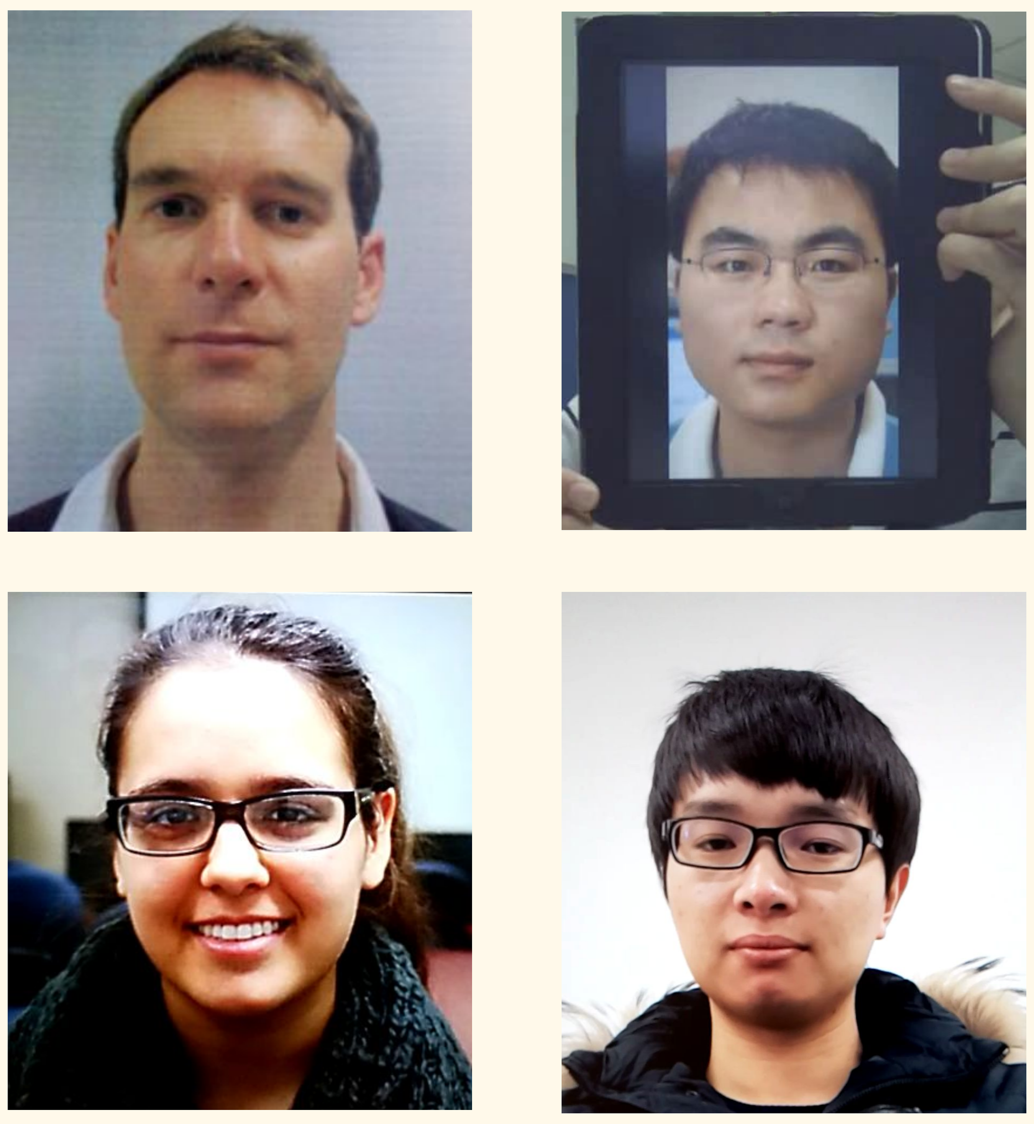
Bc. Petr Češka
2024

Introduction

Motivation: Attackers tirelessly produce new types of spoofing attacks. Model needs to be ready to face them.

Goal: Improve Vision Transformer-based face anti-spoofing model's ability to detect unknown attacks.

Method: Applying out-of-distribution (OOD) detection to filter out images that are too different from the model's training dataset.



Figures 1-4: Spoof images from different dataset. From left: MSU-MFSD, Replay-Attack, CASIA-FASD and OULU-NPU.

Proposed method

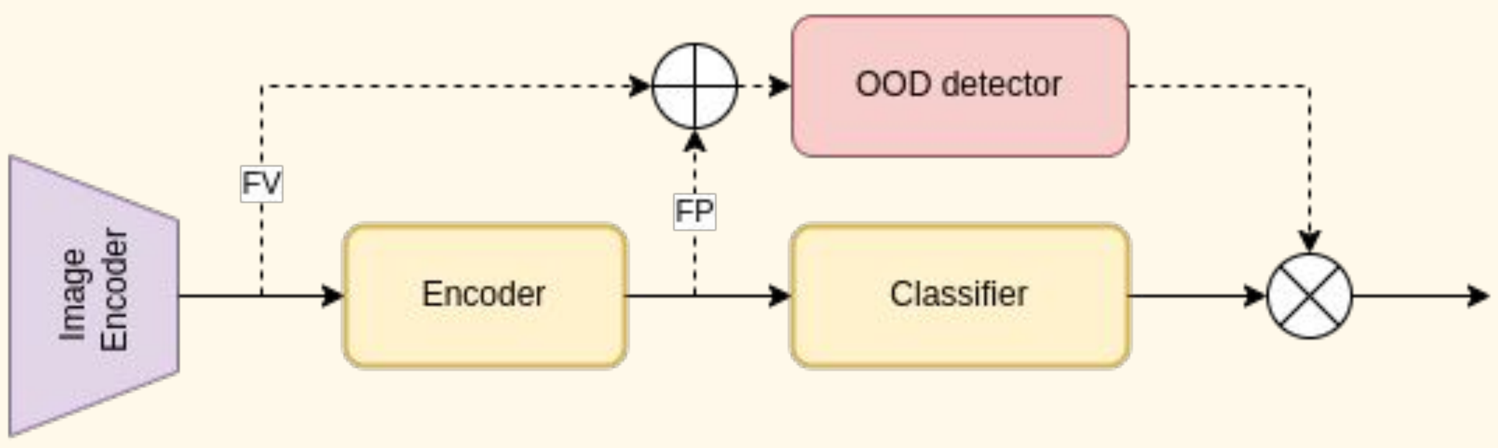


Figure 5: Proposed OOD detection for FLIP-V model.

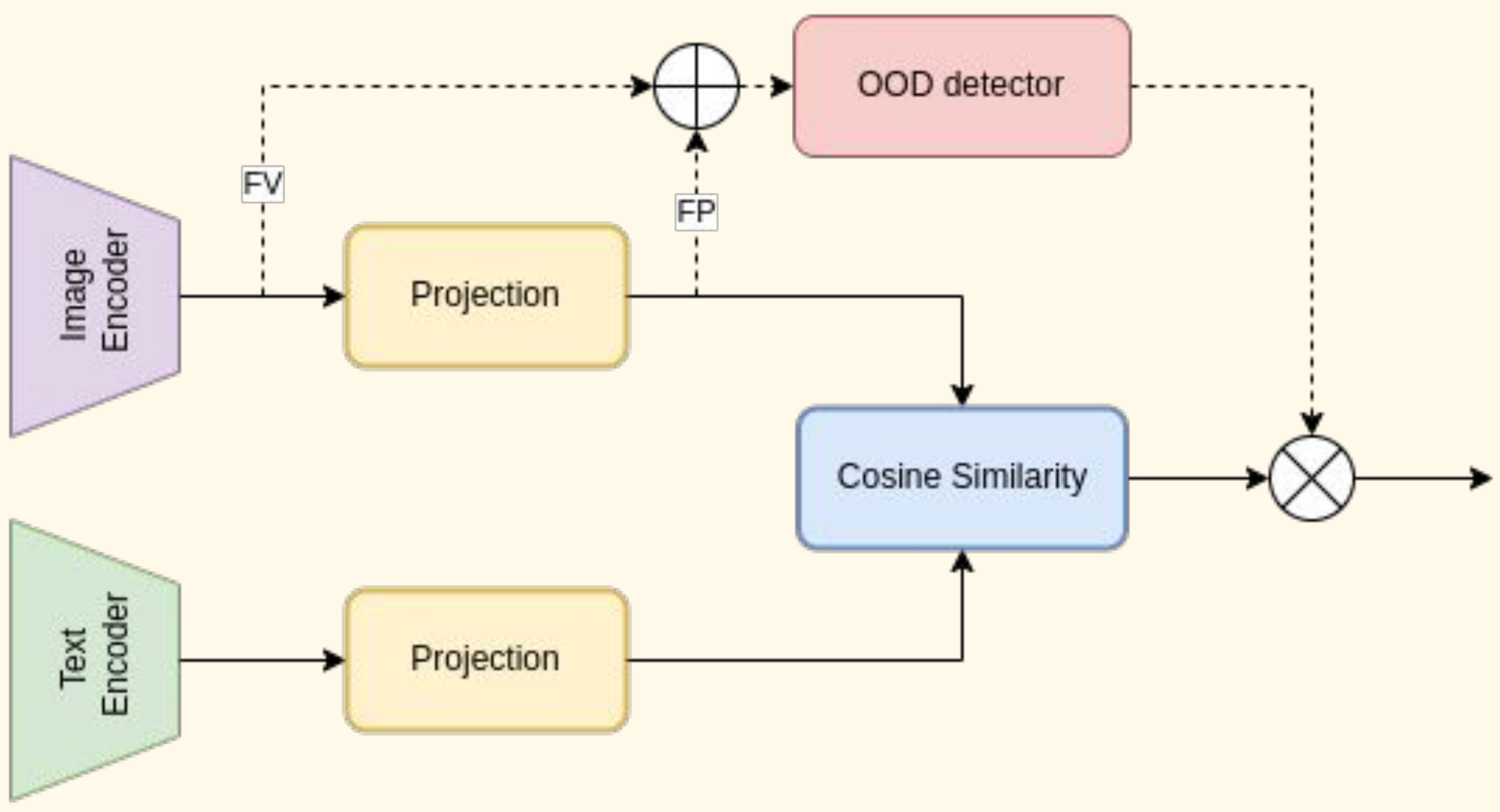


Figure 6: Proposed OOD detection for FLIP-IT and FLIP-MCL model.

Evaluation

Model	$OCI \rightarrow M$		$OMI \rightarrow C$		$OCM \rightarrow I$		$ICM \rightarrow O$		Average	
	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
SSAN-R	6.67	98.75	10.00	96.67	8.88	96.79	13.72	93.63	9.82	96.46
PatchNet	7.10	98.46	11.33	94.58	13.40	95.67	11.82	95.07	10.91	95.95
GDA	9.20	98.00	12.20	93.00	10.00	96.00	14.40	92.60	11.45	94.90
DiVT-M	2.86	99.14	8.67	96.62	3.71	99.29	13.06	94.04	7.08	97.27
ViT	**1.58**	**99.68**	5.70	98.91	9.25	97.15	7.47	98.42	6.00	98.54
FLIP-V	3.79	99.31	1.27	99.75	4.71	98.80	4.15	98.76	3.48	99.16
FLIP-IT	5.27	98.41	**0.44**	**99.98**	**2.94**	**99.42**	3.61	99.15	3.07	99.24
FLIP-MCL	4.95	98.11	0.54	**99.98**	4.25	99.07	**2.31**	**99.63**	**3.01**	**99.20**

Figure 7: Comparing chosen FLIP models with other anti-spoofing models.

Method	FLIP-V		FLIP-IT		FLIP-MCL	
	Type	AUROC	Type	AUROC	Type	AUROC
Energy	FV (N)	0.7635	FP (NL)	0.6891	FP (NL)	0.6489
Energy+React	FP (B)	0.7756	FP (B)	0.6372	FP (L)	0.7017
GradNorm	FV (B)	0.7884	FP (L)	0.7381	FV (NL)	0.7236
KL-Matching	FV (N)	0.8962	FV (L)	0.8924	FP (L)	0.7568
MSP	FP (B)	0.7203	FP (L)	0.6891	FP (L)	0.6489
Mahalanobis	FV (L)	0.9517	FV (L)	0.9540	FV (L)	0.8923
MaxLogit	FP (B)	0.7203	FP (L)	0.6891	FP (L)	0.6489
Rel. Mahalanobis	FV (L)	0.9721	FV (B)	0.9765	FV (B)	0.9568
Residual	FV (L)	0.9452	FV (L)	0.9439	FV (L)	0.8796
ViM	FV (L)	0.9454	FV (L)	0.9439	FV (L)	0.8797

Figure 8: Table showing best AUROC that each OOD detection methods achieved for each model. Type shows which features were used to reach this AUROC.

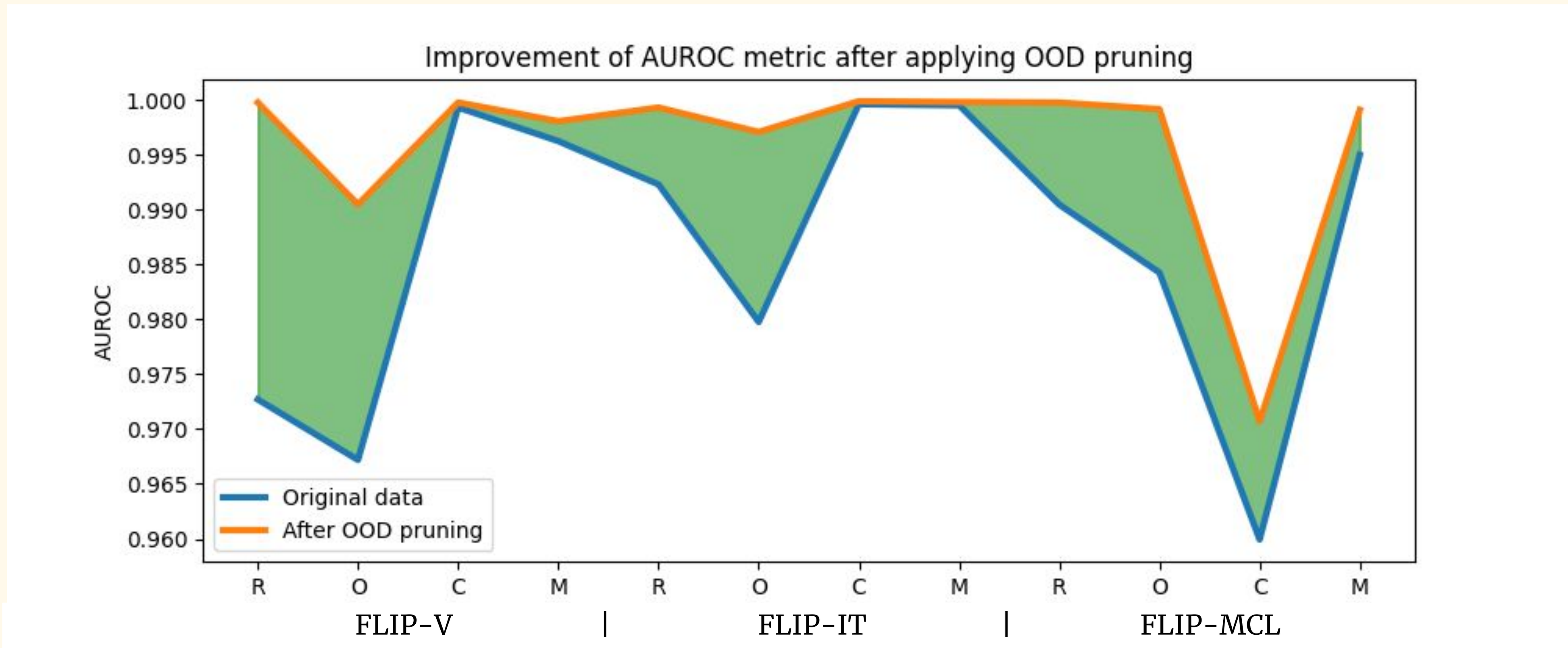


Figure 9: Table showing best improvement in accuracy of models after pruning testing data based on OOD detection. It is shown for models FLIP-V, FLIP-IT and FLIP-MCL where R, O C, M notes which dataset was used for testing (Replay-Attack, OULU-NPU, CASIA-FAS and MSU-MFSD respectively).

Conclusion

OOD detection was successful with auroc 0.9721, 0.9765 and 0.9568 on models FLIP-V, FLIP-IT and FLIP-MCL respectively. Model auroc was increased by 0.97 % in average.

Acknowledgement

I would like to thank my supervisor Ing. Jakub Špaňhel for his help and guidance and for providing me with all the training datasets.