SUPPLEMENTARY MATERIALS

1. MORE COMPARISON RESULTS

1.1. Implementation Details

The evaluation results on the **TFG-LombardGRID** dataset are reported in Table 1. We follow the same cross-domain evaluation protocol described in the main paper: only *2* forged segments per zero-shot TFG method are used as seen-forgery set, while all personalized TFG forgeries are treated as unseen. For each identity, only *50* genuine samples are provided for enrollment training. Each input sample is a 10-frame clip at 25 FPS, with a 100×50 mouth ROI extracted using Dlib.

All detectors are trained with AdamW (learning rate 0.005) for about 35 epochs. We set $\lambda_{\mathrm{rec}} = 1$, $\lambda_{\mathrm{ortho}} = 0.1$, and fix $k_r = 3$, $k_f = 2$ for W_r and W_f. This configuration is used across all experiments. Performance is evaluated by the Area Under the ROC Curve (AUC).

Table 1: AUC results on the proposed TFG-LombardGRID dataset.

Method	**Unseen forgeries**			**Mean-U**	**Seen forgeries**						**Mean**
	PersonaTalk	ER-NeRF	MimicTalk		TalkLip	Wav2Lip	IP-LAP	FOMM	SimSwap	Human	
LipForensics	0.5416	0.5874	0.5924	0.5738	0.5750	0.6894	0.5511	0.6920	0.7377	0.8217	0.6432
LipInc	0.7447	0.7539	0.9065	0.8017	0.8549	0.8632	0.8548	0.8357	0.9571	0.9952	0.8629
ProDet	0.6697	0.7864	0.9449	0.8003	0.8877	0.9372	0.9569	0.9996	0.9992	0.9941	0.9084
SA-DTH	0.7031	0.7008	0.9406	0.7815	0.9039	0.9296	0.9119	0.9331	0.9797	0.9992	0.8891
CIDE	0.7951	0.6881	0.8405	0.7746	0.9107	0.9280	0.9112	0.9855	0.9982	0.9995	0.8952
TD-VSA	0.5407	0.6568	0.8371	0.6782	0.8142	0.7157	0.5423	0.9845	0.9978	1.0000	0.7877
Siamese	0.7359	0.9486	0.9100	0.8648	0.9513	0.9754	0.9894	0.9998	0.9982	1.0000	0.9454
+ *Ours*	**0.8462**	**0.9487**	0.9275	**0.9075**	**0.9675**	**0.9809**	0.9850	0.9987	0.9973	0.9974	**0.9610**

2. REFERENCE LINK

We provide a backbone checkpoint to facilitate the reproducibility of the results and support evaluation using the code in this repository. Link: Baidu Cloud link (pwd: n6ax).