APPENDIX

Different Baselines We set other baselines of PA detector, including VGG16 and swin-transformer-tiny to validate the proposed method in addition to ResNet-18 used in the paper. As shown in Table 1, face attribute editing (*F.A.*) is adopted as the face related task (best case in the paper). Our proposed method can improve performance over both baselines, which further verified the flexibility of our method against the weaker and stronger architectures.

Table 1. Performance of Other PAD Baselines with or without Face-related Feature from FRT on Protocol-II.

	[O.M] to [C,I]			[C,I] to [O,M]		
	HTER(%)↓	AUC(%)↑	BPCER(%)↓	HTER(%)↓	AUC(%)↑	BPCER(%)↓
Vgg16	26.16	82.06	84.93	26.50	81.81	81.40
Vgg16 w/FRT	**19.18**	**89.13**	**62.22**	**15.32**	**92.24**	**55.53**
Swin-Transformer-tiny	32.89	66.95	100.00	35.76	67.27	84.19
Swin-Transformer-tiny w/ FRT	**28.33**	**78.31**	**89.17**	**31.40**	**72.45**	84.70

Different Face Related Tasks and Models We applied different face related tasks and models in our FRT-PAD model. As shown in Table 2, we also adopted ResNet34 and ResNet50 models in face recognition task (*F.R.*). Both architectures of *F.R.* brings similar improvement for PAD over the baseline detector (ResNet18), which proved that PA detector can be benefited from the different architectures in the face related tasks by following our solution.

On the other side, we applied other face related tasks in our method, i.e. face detection (*F.D.*) [1] and face localisation (*F.L.*) [2]. As listed in Table 2, consistent improvement can be observed, which further justifies the effectiveness of our method. However, the performance of our method with (*F.D.*) and (*F.L.*) are lower than the case of *F.R.*, *F.E.* and *F.A.*. One of the important reason is that (*F.D.*) and (*F.L.*) are trained to obtain location features rather than content features of faces. From this observation, the choice of the related face tasks is important. As an empirical result, (*F.A.*) performs the most suitable task to benefit PAD.

Table 2. Performance of FRT Using Other Models and Tasks on Protocol-II.

	[O.M] to [C,I]			[C,I] to [O,M]		
	HTER(%)↓	AUC(%)↑	BPCER(%)↓	HTER(%)↓	AUC(%)↑	BPCER(%)↓
Baseline (ResNet18)	25.65	79.14	95.93	28.14	79.05	81.34
Baseline w/ *F.R.* (ResNet18)	**18.17**	87.37	78.52	16.47	90.68	**62.81**
Baseline w/ *F.R.* (ResNet34)	19.86	86.74	74.51	**16.16**	**91.32**	64.51
Baseline w/ *F.R.* (ResNet50)	19.83	**88.04**	**73.05**	16.69	90.95	71.56
Baseline w/ *F.D.* (ResNet50)	21.47	85.75	79.47	18.60	89.32	65.04
Baseline w/ *F.L.* (MobileNetV2)	22.95	82.98	82.52	19.51	87.89	66.32

[1] Deng J,et al. Retinaface: Single-shot multi-level face localisation in the wild. CVPR 2020.
[2] Chen C. PyTorch Face Landmark: A Fast and Accurate Facial Landmark Detector.