6 Appendix

A. Impact of Spatial Neighborhood Distance in DBO

The full version of deep bilateral operator (DBO) with spatial neighborhood distance term can be formulated as

$$\begin{aligned} DBO_{full}(\mathcal{F})_p &= \frac{1}{k}\sum_{q\in\mathcal{F}} g_{\sigma_s}(\|p-q\|)g_{\sigma_r}(|\mathcal{F}_p-\mathcal{F}_q|)\mathcal{F}_q, \\ with: \quad k &= \sum_{q\in\mathcal{F}} g_{\sigma_s}(\|p-q\|)g_{\sigma_r}(|\mathcal{F}_p-\mathcal{F}_q|). \end{aligned} \tag{9}$$

In this ablation study, the impact of spatial neighborhood distance $g_{\sigma_s}(\|p-q\|)$ would be evaluated. Here the default setting $\sigma_r^2 = 1.0$ is utilized. It can be seen from Fig. 8 that there are no improvements (2.4% and 2.1% ACER for with and without spatial neighborhood distance, respectively) when introducing spatial neighborhood distance term into BCN.

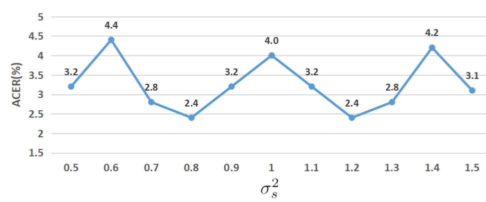

Fig. 8. Impact of σ_s in BCN.

Note that the ablation study about distance term σ_s here is not enough thus it might be a sub-optimal solution. The optimal hyperparameter setting could be found via strict grid search, which is one of our future works. Long-range spatial impact of distance term σ_s under large kernel size (e.g., 5x5 and 7x7) is also worth exploring in future.

B. Network Details of Multi-head Supervision

The detailed layers are illustrated in Fig. 9. With the supervision from three kinds of cues, the backbone network is able to learn more holistic material-based features.

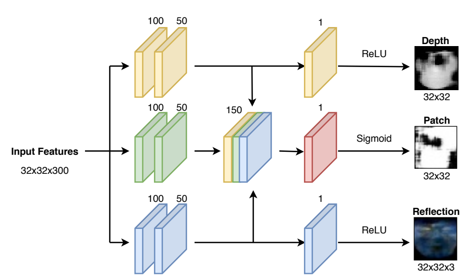

Fig. 9. Network structure of multi-head supervision. The number of filters are shown on top of each convolution, the size of all filters is 3×3 with stride 1.

C. Intra Testing Results on SiW

As shown in Table 6, the proposed method performs the best for all three protocols, revealing the excellent generalization capacity.

Table 6. The results of intra testing on three protocols of SiW [16].

Prot.	Method	APCER(%)	BPCER(%)	ACER(%)
1	Auxiliary [16]	3.58	3.58	3.58
	STASN [38]	–	–	1.00
	FAS-TD [37]	0.96	0.50	0.73
	BASN [32]	–	–	0.37
	Ours	0.55	0.17	**0.36**
2	Auxiliary [16]	0.57±0.69	0.57±0.69	0.57±0.69
	STASN [38]	–	–	0.28±0.05
	FAS-TD [37]	0.08±0.14	0.21±0.14	0.15±0.14
	BASN [32]	–	–	0.12±0.03
	Ours	0.08±0.17	0.15±0.00	**0.11±0.08**
3	STASN [38]	–	–	12.10±1.50
	Auxiliary [16]	8.31±3.81	8.31±3.80	8.31±3.81
	BASN [32]	–	–	6.45±1.80
	FAS-TD [37]	3.10±0.81	3.09±0.81	3.10±0.81
	Ours	2.55±0.89	2.34±0.47	**2.45±0.68**

D. Cross-type Testing on CASIA-MFSD, Replay-Attack and MSU-MFSD

In these cross-type testing, three datasets CASIA-MFSD [64], Replay-Attack [65] and MSU-MFSD [66] are utilized to perform intra-dataset cross-type testing between replay and print attacks. For instance, the second column 'Video' in Table 7 means that model should be trained from 'Cut Photo' and 'Wrapped Photo' while tested on 'Video'. Table 7 shows that our proposed method achieves the best overall performance (96.77% AUC), indicating the learned features generalized well among unknown attacks.

Table 7. The results of cross-type testings. The evaluation metric is AUC (%).

Method	CASIA-MFSD [64]			Replay-Attack [65]			MSU-MFSD [66]			Overall
	Video	Cut Photo	Wrapped Photo	Video	Digital Photo	Printed Photo	Printed Photo	HR Video	Mobile Video	
OC-SVM+BSIF [69]	70.74	60.73	95.90	84.03	88.14	73.66	64.81	87.44	74.69	78.68±11.74
SVM+LBP [63]	91.94	91.70	84.47	99.08	98.17	87.28	47.68	99.50	97.61	88.55±16.25
NN+LBP [70]	94.16	88.39	79.85	99.75	95.17	78.86	50.57	99.93	93.54	86.69±16.25
DTN [33]	90.0	97.3	97.5	99.9	**99.9**	99.6	**81.6**	99.9	97.5	95.9±6.2
AIM-FAS [10]	93.6	99.7	99.1	99.8	**99.9**	99.8	76.3	99.9	99.1	96.4±7.8
Ours	**99.62**	**100.00**	**100.00**	**99.99**	99.74	**99.91**	71.64	**100.00**	**99.99**	**96.77±9.99**

E. Cross-dataset Testing on CASIA-MFSD and Replay-Attack

As shown in Table 8, our proposed method has 16.6% HTER on protocol CR, outperforming all prior state-of-the-arts. For protocol RC, we also achieve comparable performance with 36.4% HTER.

Table 8. Cross-dataset testing between CASIA-MFSD and Replay-Attack.

Method	Protocol CR (HTER)		Protocol RC (HTER)	
	Train	Test	Train	Test
	CASIA-MFSD	Replay-Attack	Replay-Attack	CASIA-MFSD
STASN [38]	31.5%		30.9%	
Color Texture [5]	30.3%		37.7%	
FaceDs [13]	28.5%		41.1%	
Auxiliary [16]	27.6%		28.4%	
BASN [32]	23.6%		29.9%	
FAS-TD [37]	17.5%		**24.0%**	
Ours	**16.6%**		36.4%	