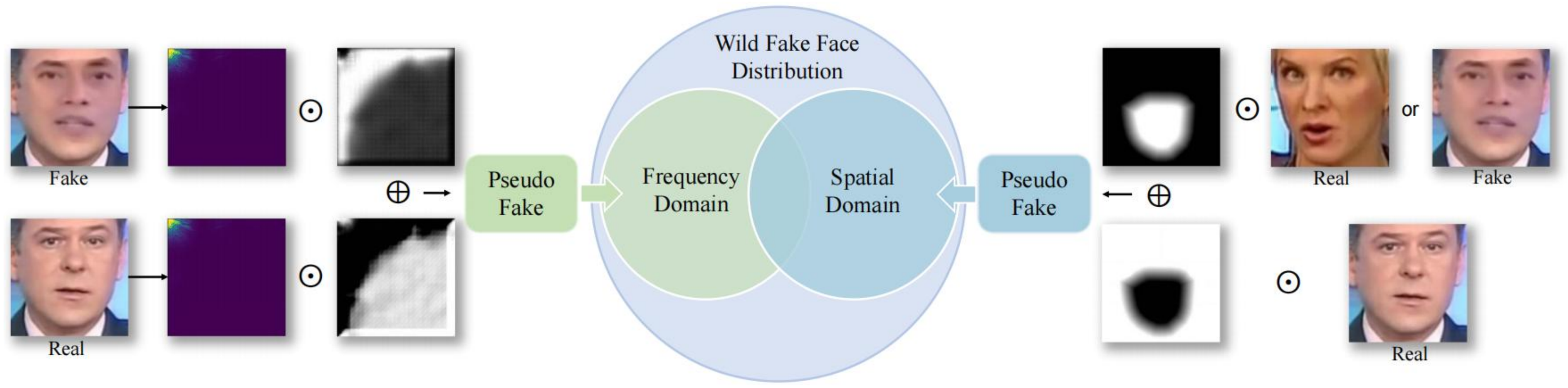
FreqBlender: Enhancing DeepFake Detection by Blending Frequency Knowledge

Hanzhe Li, Jiaran Zhou, Yuezun Li#, Baoyuan Wu, Bin Li, Junyu Dong

Corresponding author

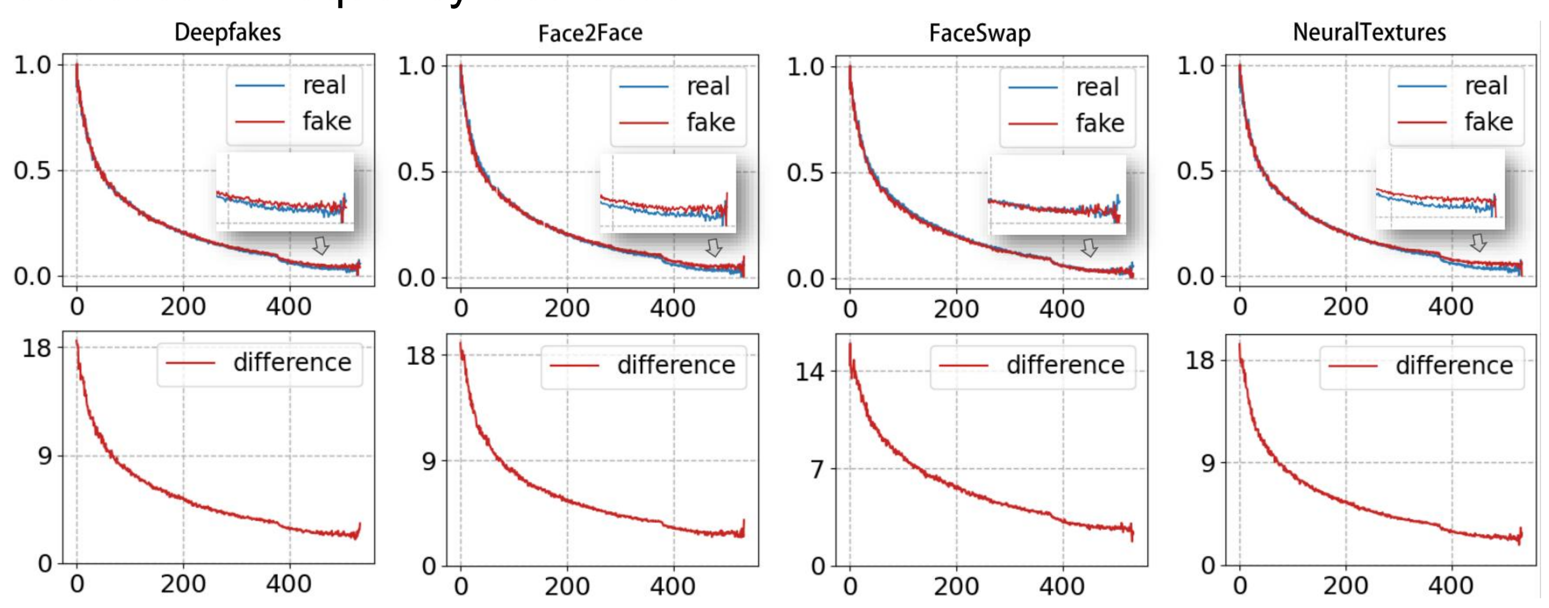
Introduction

❖ Key Idea: By leveraging the frequency knowledge, our method can generate pseudo-fake faces that closely resemble the distribution of wild fake faces.



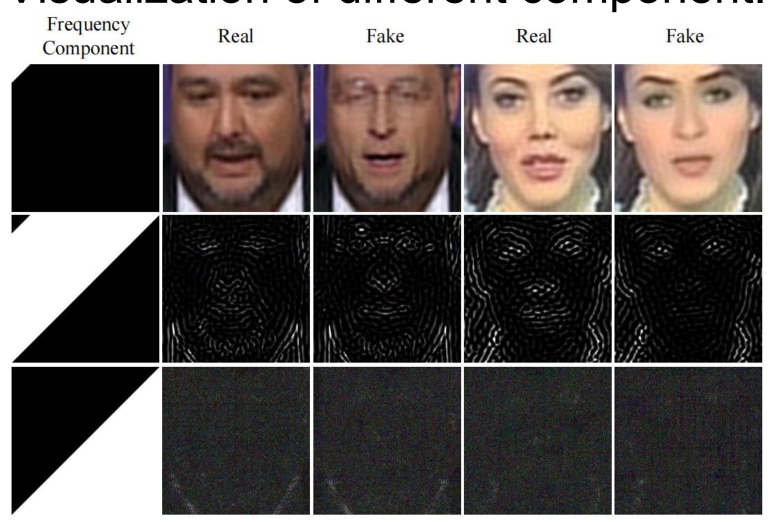
Preliminary Analysis

❖ Statistics of frequency distribution.

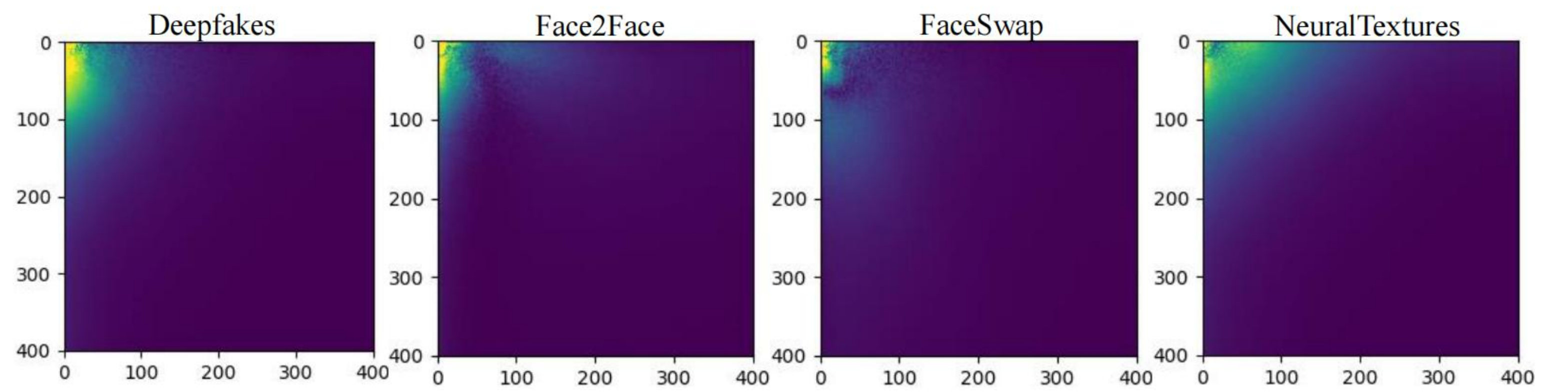


- The top part shows the frequency distribution of real and fakefaces.
- The bottom part shows the frequency difference between real and fake. The values on the vertical axis are logarithmic with 2.

❖ Visualization of different component.

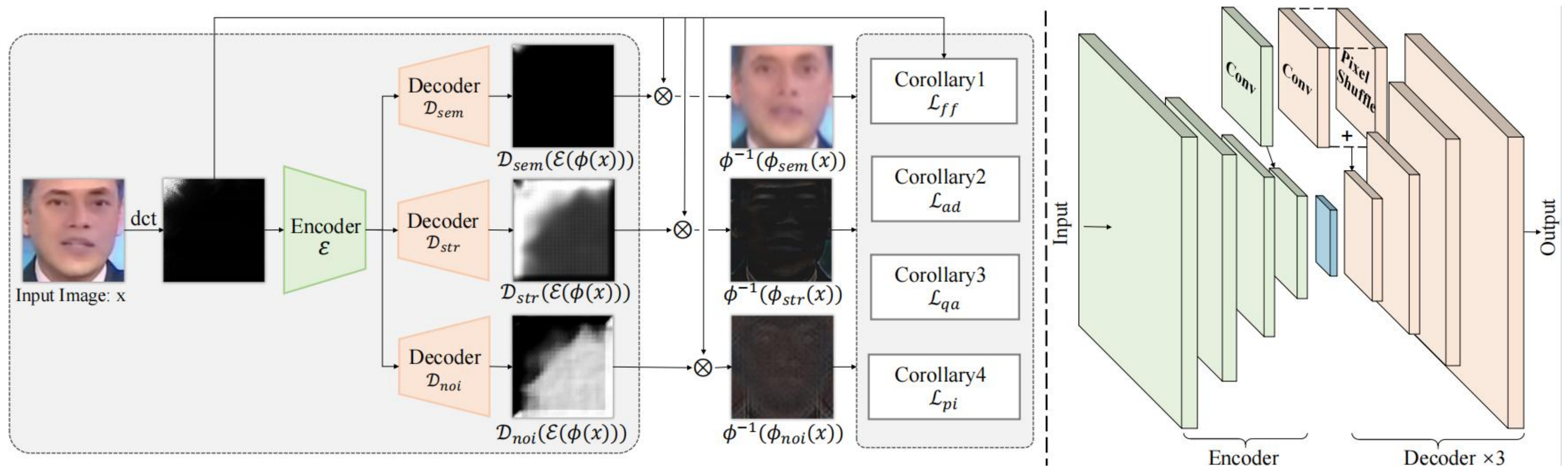


❖ Visualization of the frequency difference.



FreqBlender Methods

❖ Frequency Parsing Network.



❖ The objective Design for FPNet are designed based on following proposition:

- Semantic information can reflect the facial identity.

  Face Fidelity Loss: $\mathcal{L}_{ff}(x) = \|\mathcal{F}_f(\phi^{-1}(\phi_{sem}(x))) - \mathcal{F}_f(x)\|_2^2$

- Structural infromation serves as the carrier of forgery traces.

  Authenticity-determinative Loss: $\mathcal{L}_{ad}(x_r, x_f) = \frac{1}{|\mathcal{C}_r|}\sum_{x\in\mathcal{C}_r} CE(x,1) + \frac{1}{|\mathcal{C}_f|}\sum_{x\in\mathcal{C}_f} CE(x,0)$

- Noise information has minimal impact on visual quality.

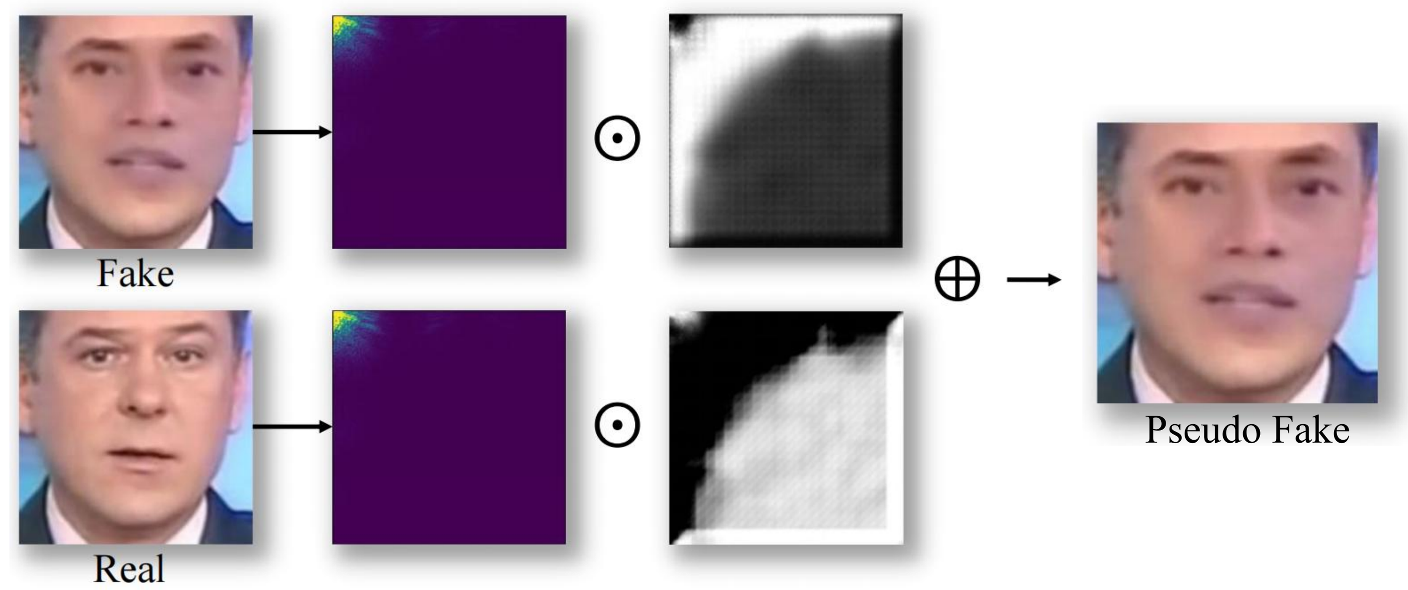
  Quality-agnostic Loss: $\mathcal{L}_{qa}(x) = \|x - \phi^{-1}(\phi_{sem}(x)) + \phi_{str}(x)\|_2^2$

- The preliminary analysis finding are generally applicable.

  Prior and Integrity Loss: $\mathcal{L}_{pi} = \|\mathcal{D}_{sem}(\mathcal{E}(\phi(x))) - m_{sem}\|_2^2 + \|\mathcal{D}_{str}(\mathcal{E}(\phi(x))) - m_{str}\|_2^2 + \|\mathcal{D}_{noi}(\mathcal{E}(\phi(x))) - m_{noi}\|_2^2 + \|\mathcal{D}_{sem}(\mathcal{E}(\phi(x))) + \mathcal{D}_{str}(\mathcal{E}(\phi(x))) + \mathcal{D}_{noi}(\mathcal{E}(\phi(x))) - 1\|_2^2$

Deployment of FreqBlender

❖ We can generate a pseudo-fake face by following methods.



$$Pseuso\ Fake\ Face = \phi^{-1}(\phi(x_r)\mathcal{D}_{sem}(\mathcal{E}(\phi(x_f))) + \phi(x_f)\mathcal{D}_{sem}(\mathcal{E}(\phi(x_f))) + \phi(x_r)\mathcal{D}_{sem}(\mathcal{E}(\phi(x_f))))$$

♞ Note that in our method, it is not necessary to perform the blending using wild faces. Instead, we can tacfully substitute wild fake faces with the pseudo-fake faces generated by existing spatital face blending methods.

Experiments

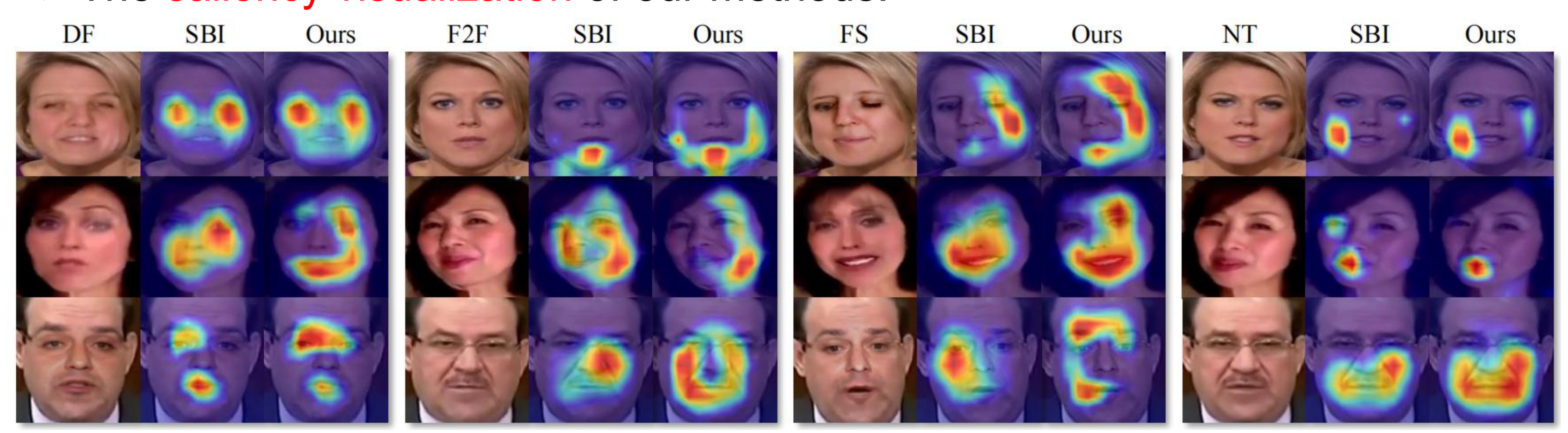
❖ The cross-dataset evaluation of different methods.

Method	Input Type	Training Set		Test Set AUC (%)			
		Real	Fake	CDF	DFDC	DFDCP	FFIW
Two-branch (ECCV'20) [38]	Video	✓	✓	76.65	-	-	-
DAM (CVPR'21) [9]	Video	✓	✓	75.3	-	72.8	-
LipForensics (CVPR'21) [1]	Video	✓	✓	82.4	73.50	-	-
FTCN (ICCV'21) [39]	Video	✓	✓	86.9	71.00	74.0	74.47
SST (CVPR'24) [24]	Video	✓	✓	89.0	-	-	-
DSP-FWA (CVPRW'19 [10])	Frame	✓	✓	69.30	-	-	-
Face X-ray (CVPR'20) [14]	Frame	✓	-	-	-	71.15	-
Face X-ray (CVPR'20) [14]	Frame	✓	✓	-	-	80.92	-
F3-Net* (ECCV'20) [30]	Frame	✓	✓	72.93	61.16	81.96	61.58
LRL (AAAI'21) [40]	Frame	✓	✓	78.26	-	76.53	-
FRDM (CVPR'21) [41]	Frame	✓	✓	79.4	-	79.7	-
PCL+I2G (ICCV'21) [15]	Frame	✓	-	90.03	67.52	74.37	-
DCL (AAAI'22) [42]	Frame	✓	✓	82.30	-	76.71	71.14
SBI* (CVPR'22) [16]	Frame	✓	-	92.94	72.08	85.51	85.99
SBI (CVPR'22) [16]	Frame	✓	-	93.18	72.42	86.15	84.83
TALL-Swin (ICCV'23) [22]	Frame	✓	✓	90.79	76.78	-	-
UCF (ICCV'23) [12]	Frame	✓	✓	82.4	**80.5**	-	-
BiG-Arts (PR'23) [25]	Frame	✓	✓	77.04	-	80.48	-
F-G (CVPR'24) [43]	Frame	✓	✓	74.42	61.47	-	-
LSDA (CVPR'24) [23]	Frame	✓	✓	83.0	73.6	81.5	-
FreqBlender (Ours)	Frame	✓	-	**94.59**	74.59	**87.56**	**86.14**

❖ The evaluation of different network architectures.

Method	CDF	FF++	DFDCP	FFIW	Avg
ResNet-50 [35] + SBI	84.82	95.39	73.51	81.67	83.85
ResNet-50 [35] + Ours	85.44	94.61	76.16	86.32	**85.63**
EfficientNet-b1 [37] + SBI	90.25	94.66	87.54	82.55	88.75
EfficientNet-b1 [37] + Ours	90.53	94.65	87.70	83.76	**89.16**
Xception [45] + SBI	87.00	91.40	75.68	70.24	81.08
Xception [45] + Ours	90.52	93.32	76.07	70.43	**82.59**
ViT [46] + SBI	85.85	96.09	87.71	86.05	88.92
ViT [46] + Ours	86.34	96.10	87.17	86.88	**89.12**
F3-Net [30]	84.94	93.42	79.29	73.42	82.77
F3-Net [30] + Ours	88.10	95.16	84.32	74.49	**85.52**
GFFD [41]	81.34	91.81	77.19	65.53	78.97
GFFD [41] + Ours	86.71	92.18	78.25	77.45	**83.65**

❖ The saliency visualization of our methods.



This work is supported in part by the National Natural Science Foundation of China (No.62402464), Shandong Natural Science Foundation (No.ZR2024QF035), and China Postdoctoral Science Foundation (No.2021TQ0314; No.2021M703036). Jiaran Zhou is supported by the National Natural Science Foundation of China (No.62102380) and Shandong Natural Science Foundation (No.ZR2021QF095, No.ZR2024MF083). Baoyuan Wu is supported by Guangdong Basic and Applied Basic Research Foundation (No.2024B1515020095), National Natural Science Foundation of China (No. 62076213), and Shenzhen Science and Technology Program (No. RCYX20210609103057050). Bin Li is supported in part by NSFC (Grant U23B2022, U22B2047).