

Towards Real-time Video Compressive Sensing on Mobile Devices – Supplementary Material

Anonymous Authors

1 LIMITATION DISCUSSION

While our proposed MobileSCI model can achieve real-time reconstruction on an iPhone 15 in the main paper, there is still room for further improvement, which we discuss as follows.

First of all, it is well-known that inference speed highly depends on the specific mobile platform. For example, it is typically required to quantize a deep learning-based model to 8-bit to make sure that the method can inference at real-time speed on the Hexagon DSP of a Qualcomm Snapdragon processor. Moreover, some operations such as the 3D Deconv layers cannot be deployed on the mobile devices. In the paper, we mainly evaluate the method on the mainstream mobile devices (e.g., iPhone). Advancing the proposed algorithm on more mobile devices, especially those with less computing power, is a worthy next step.

Second, although our MobileSCI method can achieve superior performance on the *gray-scale* testing data, it is also important to capture and then retrieve *color* scene in real-world applications.

Finally, implementing a real-time mobile video snapshot compressive imaging (SCI) system involves a co-design of miniaturized hardware and mobile-friendly reconstruction algorithm, which poses a big challenge to the researchers in the computational imaging community. The paper currently mainly focuses on the algorithm axis. More systematical optimization of hardware-algorithm co-design is also worth exploring in the future.

Table T1: Effect of the Batch Norm (BN) layers in EfficientFormer V2 and MobileNet V2, where “w/o BN” and “w/ BN” denote disabling and enabling the BN layers in the deep learning-based models, respectively.

Method	PSNR	SSIM	Params (M)	FLOPs (G)
EfficientFormer V2 (w/o BN) [1]	33.82	0.956	12.207	142.38
EfficientFormer V2 (w/ BN) [1]	33.43↓	0.951	12.271	142.70
MobileNet V2 (w/o BN) [2]	33.61	0.955	8.186	133.17
MobileNet V2 (w/ BN) [2]	33.26↓	0.951	8.214	133.40

2 MORE REMARKS ON EXPERIMENTS

More details about Tab. 3 in the manuscript: In Tab. 3 of the manuscript, we conduct ablation study on different mobile efficient modules. For EfficientFormer V2 and MobileNet V2, we report their

performance without Batch Norm (BN) layers in the manuscript due to their better performance on the video SCI reconstruction task. Specifically, as shown in Tab. T1, employing the BN layers in EfficientFormer V2 and MobileNet V2 will bring a 0.39dB and 0.35dB drop in PSNR value, respectively.

More details about Tab. 4 in the manuscript: Here, we discuss more about the network setting in Tab. 4 of the manuscript. In the baseline model, we put six convolutional units with channel number as 64 in the Conv Block (CB) 1–4 and the Bottleneck Block(BB). Following this, keeping the same testing time on the same NVIDIA RTX 3090 GPU, we sequentially replace the convolutional units in the baseline model with our proposed feature mixing block (FMB).

Specifically, *i*) In the second row of Tab. 4 in the manuscript, we replace the convolutional units in the BB of the baseline model with 1 FMB. *ii*) In the third row of Tab. 4 in the manuscript, we sequentially replace the convolutional units in the BB and the CB 2&3 of the baseline model with 1 and 5 FMBs, respectively. *iii*) In the last row of Tab. 4 in the manuscript, we sequentially replace the convolutional units in the BB, the CB 2&3, and the CB 1&4 of the baseline model with 1, 4, and 4 FMBs, respectively.

REFERENCES

- [1] Yanyu Li, Ju Hu, Yang Wen, Georgios Evangelidis, Kamyar Salahi, Yanzhi Wang, Sergey Tulyakov, and Jian Ren. 2023. Rethinking vision transformers for mobilenet size and speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16889–16900.
- [2] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4510–4520.

Permission to make digital or hard copies of all or part of this work for personal or commercial use, by users registered with ACM, is granted by ACM Publishing Group for users registered with ACM. This permission is granted on the basis that the copiers pay the stated per-copy fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. For all other use, permission should be sought from ACM or the author(s).

Unpublished working draft. Not for distribution. This document is preliminary and should not be used for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM MM, 2024, Melbourne, Australia

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>