International Journal of Computer Vision (2023) 131:2933–2958
https://doi.org/10.1007/s11263-023-01844-4

Deep Unfolding for Snapshot Compressive Imaging

Ziyi Meng[1] · Xin Yuan[2] · Shirin Jalali[3]

Received: 25 September 2020 / Accepted: 26 June 2023 / Published online: 8 July 2023

Abstract
Snapshot compressive imaging (SCI) systems aim to capture high-dimensional (≥ 3D) images in a single shot using 2D detectors. SCI devices consist of two main parts: a hardware encoder and a software decoder. The hardware encoder typically consists of an (optical) imaging system designed to capture compressed measurements. The software decoder, on the other hand, refers to a reconstruction algorithm that retrieves the desired high-dimensional signal from those measurements. In this paper, leveraging the idea of deep unrolling, we propose an SCI recovery algorithm, namely GAP-net, which unfolds the generalized alternating projection (GAP) algorithm. At each stage, GAP-net passes its current estimate of the desired signal through a trained convolutional neural network (CNN). The CNN operates as a denoiser projecting the estimate back to the desired signal space. For the GAP-net that employs trained auto-encoder-based denoisers, we prove a probabilistic global convergence result. Finally, we investigate the performance of GAP-net in solving video SCI and spectral SCI problems. In both cases, GAP-net demonstrates competitive performance on both synthetic and real data. In addition to its high accuracy and speed, we show that GAP-net is flexible with respect to signal modulation implying that a trained GAP-net decoder can be applied in different systems. Our code is available at https://github.com/mengziyi64/GAP-net.

Keywords Compressive imaging · Compressive sensing · Deep learning · Generative alternating projection · Snapshot · Convolution neural network · Convergence · Denoising

1 Introduction

Recent advances in artificial intelligence and robotics have resulted in an unprecedented demand for computationally efficient high-dimensional (HD) data capture and processing devices. However, existing optical sensors usually can only directly capture two-dimensional (2D) signals. Capturing 3D (*e.g.*, spatial-temporal) or higher (spatial, spectral, temporal and polarization, etc) dimensional signals remains challenging since sensors that directly perform 3D data acquisition do not yet exist.

In recent years, snapshot compressive imaging (SCI) systems that employ 2D detectors to capture HD (≥ 3D) signals have proven to be a promising solution to this challenge. Different from conventional cameras, SCI systems perform sampling on a set of consecutive images—video frames (*e.g.*, CACTI (Llull et al., 2013)) or spectral channels (*e.g.*, CASSI (Wagadarikar et al., 2008))—in accordance with the sensing matrix and integrate these sampled signals along time or spectrum to obtain the final compressed measurements. Using this technique, SCI systems (Gehm et al., 2007; Hitomi et al., 2011) [49] can capture high-speed motion (Qiao et al., 2020a, b, 2021a; Yuan et al., 2014; Yuan & Pang, 2016) or high-resolution spectral information (Wagadarikar et al., 2009; Yuan et al., 2015; Meng et al., 2020a, b, 2021a), with low memory, low bandwidth, low power and potentially low cost (Lu et al., 2020; Qiao et al., 2021b).

There are two main components in an SCI system: hardware encoder and software decoder (Yuan et al., 2021). The hardware encoder is typically an (optical) imaging system

Communicated by Jiri Matas.

✉ Xin Yuan
xyuan@westlake.edu.cn

Ziyi Meng
mengziyi64@163.com

Shirin Jalali
shirin.jalali@rutgers.edu

1 Kuaishou Technology, Beijing 100085, China

2 Research Center for Industries of the Future and School of Engineering, Westlake University, Hangzhou 310030, Zhejiang, China

3 Electrical and Computer Engineering Department, Rutgers University, Piscataway, NJ 08854, USA