3D Tensor Auto-encoder with Application to Video Compression

YANG LI, School of IoT Engineering (School of Information Security), Jiangsu Vocational College of Information Technology, China GUANGCAN LIU, YUBAO SUN, and QINGSHAN LIU, Nanjing University of Information Science and Technology, China SHENGYONG CHEN, Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, China

Auto-encoder has been widely used to compress high-dimensional data such as the images and videos. However, the traditional auto-encoder network needs to store a large number of parameters. Namely, when the input data is of dimension n, the number of parameters in an auto-encoder is in general O(n). In this article, we introduce a network structure called 3D Tensor Auto-Encoder (3DTAE). Unlike the traditional auto-encoder, in which a video is represented as a vector, our 3DTAE considers videos as 3D tensors to directly pass tensor objects through the network. The weights of each layer are represented by three small matrices, and thus the number of parameters in 3DTAE is just $O(n^{1/3})$. The compact nature of 3DTAE fits well the needs of video compression. Given an ensemble of high-dimensional videos, we represent them as 3DTAE networks plus some small core tensors, and we further quantize the network parameters and the core tensors to get the final compressed data. Experimental results verify the efficiency of 3DTAE.

 $\label{eq:CCS} \mbox{Concepts:} \bullet \mbox{Information systems} \rightarrow \mbox{Data encoding and canonicalization}; \mbox{Multimedia and multimodal retrieval}; \bullet \mbox{Computing methodologies} \rightarrow \mbox{Image compression}; \bullet \mbox{Computer systems organization} \rightarrow \mbox{Neural networks};$

Additional Key Words and Phrases: Video compression, tensor, auto-encoder, ADAM

© 2021 Association for Computing Machinery.

1551-6857/2021/05-ART48 \$15.00

https://doi.org/10.1145/3431768

ACM Trans. Multimedia Comput. Commun. Appl., Vol. 17, No. 2, Article 48. Publication date: May 2021.

This work was supported in part by the New Generation AI Major Project of Ministry of Science and Technology of China under Grant 2018AAA0100601, in part by National Natural Science Foundation of China (NSFC) under Grant 62020106004, Grant 61532009, Grant 61672292, and Grant 61703304, in part by Higher Vocational Education Teaching Fusion Production Integration Platform Construction Projects of Jiangsu Province under Grant No.2019(26), in part by "Qing Lan Project" Teaching Team in Colleges and Universities of Jiangsu Province under Grant No.2017(15), in part by High Level of Jiangsu Province Key Construction Project Fund under Grant No.2017(17), in part by Research Project of Jiangsu Vocational College of Information Technology under Grant 1007202028(001).

Authors' addresses: Y. Li, School of IoT Engineering (School of Information Security), Jiangsu Vocational College of Information Technology, No 1 Qianou Road, Wuxi, China 214153; email: liyang_19901222@163.com; G. Liu, Y. Sun, and Q. Liu, Nanjing University of Information Science and Technology, 219 Ningliu Road, Nanjing, China 210044; emails: {gcliu, ybsun, qsliu}@nuist.edu.cn; S. Chen, Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, No 319 Binshui West Road, Tianjin, China 300384; email: csy@tjut.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed 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 ACM 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.