NEURAL VIDEO CODING

BACKGROUND

The main principle of video codec is that, for the current frame to be coded, the codec will find the relevant context information (e.g., various predictions as context information) from previously reconstructed frames to reduce the spatial-temporal redundancy. The more relevant the context information is, the higher bitrate saving is achieved. Traditional video coding (e.g., from H.261 to H.266) extracts and utilizes context information from various hand-crafted coding modes. At present, neural video codec (NVC) is proposed for extraction and utilization of context information in an automatic-learned manner, which will bring more flexibility and improve coding efficiency.

WO2023/278068A1 describes a context-based image coding solution. According to the solution, a reference image of a target image is obtained. A contextual feature representation is extracted from the reference image , the contextual feature representation characterizing contextual information associated with the target image. Conditional encoding or conditional decoding is performed on the target image based on the contextual feature representation. In this way, the enhancement of the performance is achieved in terms of the reconstruction quality and the compression efficiency.

“Hybrid Spatial-Temporal Entropy Modelling for Neural Video Compression” by JIAHAOLIET AL, ARXIV.ORG, CORNELL, UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 13 JULY 2022 (2022-07-13), XP091269974 proposes a powerful entropy model which efficiently captures both spatial and temporal dependencies. In particular, the Authors introduce the latent prior which exploits the correlation among the latent representation to squeeze the temporal redundancy. Meanwhile, the dual spatial prior is proposed to reduce the spatial redundancy in a parallel-friendly manner. In addition, the Authors’ entropy model is also versatile. Besides estimating the probability distribution, the Authors’ entropy model also generates the quantization step at spatial-channel-wise. This content-adaptive quantization mechanism not only helps the Author’s codec achieve the smooth rate adjustment in single model but also improves the final rate-distortion performance by dynamic bit allocation. Experimental results show that, powered by the proposed entropy model, the Authors’ neural codec can achieve 18.2% bitrate saving on UVG dataset when compared with H.266 (VTM) using the highest compression ratio configuration. It makes a new milestone in the development of neural video codec. The codes are at http://github.com/microsoft/DCVC.

“Exploring Long & Short Range Temporal Information for Learned Video Compression” by HUAIRUI WANG ET AL, ARXIVORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 7 AUGUST 2022 (2022-08-07),XP 091289721 focuses on exploiting the unique characteristics of video content and further exploring temporal information to enhance compression performance . Specifically, for long-range temporal information exploitation, the Authors propose temporal prior that can update continuously within the group of pictures (GOP) during inference. In that case, temporal prior contains valuable temporal information of all decoded images within the current GOP. As for short-range temporal information, the Authors propose a progressive

guided motion compensation to achieve robust and effective compensation. In detail, the Authors design a hierarchical structure to achieve multi-scale compensation. More importantly, the Authors use optical flow guidance to generate pixel offsets between feature maps at each scale, and the compensation results at each scale will be used to guide the following scale's compensation. Sufficient experimental results demonstrate that the Authors' method can obtain better RD performance than scale-of-the-art video compression approaches. The code is publicly available on : https://github.com/Huairui/LSTVC.

SUMMARY

The invention is set out in the appended set of claims.

According to implementations of the subject matter described herein, a solution for neural video coding is proposed. In this solution, estimated motion information of a target frame in a video, reference feature information and a reference reconstructed frame of a reference frame for the target frame are obtained. Context information for the target frame is determined using a context extraction model and based on the estimated motion information, the reference reconstructed frame, and the reference feature information. In a conversion between the target frame and a bitstream of the video, a reconstructed target frame of the target frame is generated using a frame coding model and based on at least the context information. In this way, richer context information can be extracted for coding and thus the coding efficiency can be improved.

The Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. The Summary is neither intended to identify key features or essential features of the subject matter described herein, nor is it intended to be used to limit the scope of the subject matter described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a block diagram of an example environment in which various implementations of the subject matter described herein can be implemented;

FIG. 2 illustrates a schematic block diagram of a neural video codec system in accordance with some implementations of the subject matter described herein;

FIG. 3 illustrates a schematic block diagram of example architecture of a context extraction model in accordance with some implementations of the subject matter described herein;

FIG. 4 illustrates a schematic block diagram of example architecture of an offset predictor in the context extraction model in accordance with some implementations of the subject matter described herein;