RNNSC: Recurrent Neural Network-Based Stereo Compression Using Image and State Warping

M. Shahzeb Khan Gul, Hamid Suleman, Michel Bätz, and Joachim Keinert

Fraunhofer Institute for Integrated Circuits IIS
Am Wolfsmantel 33, Erlangen, 91058, Germany
{muhammad.gul,sulemahd,michel.baetz,joachim.keinert}@iis.fraunhofer.de

Stereo images are used in various applications, such as autonomous driving, surveillance, robotics, and 3D-TV. Those images are captured by two horizontally adjacent cameras, capturing a scene from two different points of view. In this work, we propose an end-to-end trainable recurrent neural network (RNN) for stereo image compression, which we call RNNSC. The RNN allows variable compression rates without retraining of the network due to the iterative nature of the recurrent units. The proposed method makes use of the redundancies, to reduce the overall bit rate. Each image in the stereo pair has its separate encoder and decoder network similar to [1]. We propose to share the mutual information between the stereo pair networks by warping the hidden states of one codec network to the other with the help of disparity information that is coded and transmitted independently via JPEG2000. Moreover, we also improve the quality of the shared mutual information by eliminating wrong information by estimating and applying occlusion maps which are computed with a convolutional neural network without direct supervision. The proposed method outperforms all tested image codecs on MS-SSIM, a perceptual metric capturing the structural quality of an image, as shown in Table 1.

Methods	Bit rate savings % MS-SSIM	PSNR
RNNSC	**15.84**	15.00
HEVC Inter	3.11	**62.09**
BPG	-5.55	60.73
JPEG2000	-39.81	41.24

(a) VKitti2

Methods	Bit rate savings % MS-SSIM	PSNR
RNNSC	**12.24**	11.72
HEVC Inter	-11.52	43.17
BPG	-6.22	**50.48**
JPEG2000	-46.96	34.16

(b) Kitti2012

Table 1: Bjøntegaard rate-difference on MS-SSIM and PSNR for the VKitti2 [2] & Kitti2012 [3] test datasets. The rate differences are relative to the RNN-based single image compression baseline. The best result for each metric is highlighted.

References

[1] George Toderici, Damien Vincent, Nick Johnston, Sung Jin Hwang, David Minnen, Joel Shor, and Michele Covell, "Full resolution image compression with recurrent neural networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 5306–5314.

[2] Yohann Cabon, Naila Murray, and Martin Humenberger, "Virtual KITTI 2," in *arXiv preprint arXiv:2001.10773*, 2020.

[3] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2012.

This work was supported by the EU's H2020 ITN Programme through the MSCA Grant Agreement (RealVision) under Grant 765911.

2022 Data Compression Conference (DCC) | 9781-6654-7893-9/22/$31.00 © 2022 IEEE | DOI: 10.1109/DCC52660.2022.00066