Learning-based Visual Compression

Other titles in Foundations and Trends® in Computer Graphics and Vision

Video Summarization Overview
Mayu Otani, Yale Song and Yang Wang
ISBN: 978-1-63828-078-1

A Comprehensive Review of Modern Object Segmentation Approaches
Yuanbo Wang, Unaiza Ahsan, Hanyan Li and Matthew Hagen
ISBN: 978-1-63828-070-5

Deep Learning for Image/Video Restoration and Super-resolution
A. Murat Tekalp
ISBN: 978-1-68083-972-2

Deep Learning for Multimedia Forensics
Irene Amerini, Aris Anagnostopoulos, Luca Maiano and Lorenzo Ricciardi Celsi
ISBN: 978-1-68083-854-1

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art
Joel Janai, Fatma Güney, Aseem Behl and Andreas Geiger
ISBN: 978-1-68083-688-2

Discrete Graphical Models - An Optimization Perspective
Bogdan Savchynskyy
ISBN: 978-1-68083-638-7

Learning-based Visual Compression

Ruolei Ji
Arizona State University
ruoleiji@asu.edu

Lina J. Karam
Lebanese American University
Arizona State University
lina.karam@lau.edu.lb
karam@asu.edu

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

R. Ji and L. J. Karam. *Learning-based Visual Compression*. Foundations and Trends® in Computer Graphics and Vision, vol. 15, no. 1, pp. 1–112, 2023.

ISBN: 978-1-63828-113-9

Foundations and Trends® in Computer Graphics and Vision
Volume 15, Issue 1, 2023

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2023, Volume 15, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Contents

Learning-based Visual Compression

Ruolei Ji[1] and Lina J. Karam[1,2]

[1] *Arizona State University, USA; ruoleiji@asu.edu, karam@asu.edu*
[2] *Lebanese American University, Lebanon; lina.karam@lau.edu.lb*

ABSTRACT

Visual compression is an application of data compression to lower the storage and/or transmission requirements for digital images and videos. Due to the rapid growth in visual data transmission demand, more efficient compression algorithms are needed. Considering that deep learning techniques have successfully revolutionized many visual tasks, learning-based compression algorithms have been explored over the years and have been shown to be able to outperform many conventional compression methods. This survey provides a review of various visual compression algorithms, both end-to-end learning-based image compression approaches and hybrid image compression approaches. Some learning-based video compression methods are also discussed. In addition to describing a wide range of learning-based image compression approaches that have been developed in recent years, the survey describes widely used datasets, presents recent standardization efforts, and discusses potential research directions.

Ruolei Ji and Lina J. Karam (2023), "Learning-based Visual Compression", Foundations and Trends® in Computer Graphics and Vision: Vol. 15, No. 1, pp 1–112. DOI: 10.1561/0600000101.

1

Introduction

In recent years, the demand for visual media has been growing exponentially. According to the 2019 CISCO Visual Network Index (VNI) forecast update (Nowell, 2019), the global IP video traffic will be 82% of all IP traffic in 2022, up from 73% in 2016 (CISCO, 2016). Among the large amount of visual traffic over the Internet, high-resolution visual content constitutes an increasingly large percentage. Despite the increase in the average broadband speed, about 1.9-fold from 2016 to 2021, the growth rate of visual data, approximately 3-fold from 2016 to 2021, is much higher than the broadband growth rate (CISCO, 2016; Nowell, 2019). With such a rapid growth of digital visual media traffic, there is a growing need for image/video compression approaches that can achieve much higher compression ratios than the ones obtained using existing conventional image/video compression methods, while maintaining a high visual quality.

Most conventional lossy image compression methods, *e.g.,* JPEG (Wallace, 1992), WebP (Google, 2015), JPEG2000 (Taubman and Marcellin, 2002), BPG[1] (Bellard, 2018), HEVC-based intra coding (Sullivan

[1]This corresponds to a container for HEVC-based intra coding, Main 4:4:4 16 intra profile.

et al., 2012), and VVC-based intra coding (Ohm and Sullivan, 2018) are built based on a transform coding based framework (Goyal, 2001), where an invertible transform module is used to map image pixel intensities or predicted pixel residuals into a latent representation at the encoder. The latent representation is then quantized to produce a compact representation. An entropy encoder is later employed for coding the quantized latent representation. At the decoder, an inverse transform module is applied to the entropy decoded quantized data to recover a lossy image.

Although the conventional compression algorithms have been widely used and achieved promising results, researchers are working on using learning-based approaches for image/video compression to help further improve the compression performance. Such increased interest in learning-based compression stems from the fact that, over the last decade, deep-learning-based approaches have achieved huge success in a variety of visual tasks including but not limited to classification, segmentation, object detection, and super-resolution.

References

Adams, M. (2006). "JasPer Software Reference Manual (Version 1.900.0". *ISO/IEC JTC 1/SC 29/WG 1, N 2415*. Dec. URL: https://www.ece.uvic.ca/~frodo/jasper/.

Agustsson, E., D. Minnen, N. Johnston, J. Balle, S. J. Hwang, and G. Toderici. (2020). "Scale-Space Flow for End-to-End Optimized Video Compression". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8503–8512.

Agustsson, E. and R. Timofte. (2017). "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and study". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 126–135.

Agustsson, E., M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool. (2019). "Generative Adversarial Networks for Extreme Learned Image Compression". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 221–231.

Akbari, M., J. Liang, and J. Han. (2019). "DSSLIC: Deep Semantic Segmentation-based Layered Image Compression". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2042–2046. DOI: 10.1109/ICASSP.2019.8683541.

Amestoy, T., A. Mercat, W. Hamidouche, D. Menard, and C. Bergeron. (2019). "Tunable VVC Frame Partitioning Based on Lightweight Machine Learning". *IEEE Transactions on Image Processing*. 29: 1313–1328.

Andrews, D. F. and C. L. Mallows. (1974). "Scale Mixtures of Normal Distributions". *Journal of the Royal Statistical Society: Series B (Methodological)*. 36(1): 99–102.

Arbeláez, P., M. Maire, C. Fowlkes, and J. Malik. (2011). "Contour Detection and Hierarchical Image Segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 33(5): 898–916. DOI: 10.1109/TPAMI.2010.161.

Ascenso, J. and E. Upenik. (2021). "White Paper on JPEG AI Scope and Framework". *ISO/IEC JTC1/SC29/WG1, N90049*. Jan.

Ascenso, J., P. Akyzi, M. Testolina, A. Boev, and E. Alshina. (2019). "Performance Evaluation of Learning based Image Coding Solutions and Quality Metrics". *ISO/IEC JTC 1/SC29/WG1 N85013, 85th JPEG Meeting*. Nov.

Asuni, N. and A. Giachetti. (2014). "TESTIMAGES: a large-scale archive for testing visual devices and basic image processing algorithms". In: *STAG: Smart Tools & Apps for Graphics*. 63–70.

Ballé, J., V. Laparra, and E. P. Simoncelli. (2016a). "END-TO-END OPTIMIZED IMAGE COMPRESSION". *arXiv preprint arXiv:1611.01704*.

Ballé, J., V. Laparra, and E. P. Simoncelli. (2016b). "End-to-end optimization of nonlinear transform codes for perceptual quality". In: *2016 Picture Coding Symposium (PCS)*. 1–5. DOI: 10.1109/PCS.2016.7906310.

Ballé, J., D. Minnen, S. Singh, S. J. Hwang, and N. Johnston. (2018). "VARIATIONAL IMAGE COMPRESSION WITH A SCALE HYPERPRIOR". *arXiv preprint arXiv:1802.01436*.

Bellard, F. (2018). "BPG Image Format (2018)". URL: https://bellard.org/bpg/.

Benjak, M., H. Meuel, T. Laude, and J. Ostermann. (2021). "Enhanced Machine Learning-Based Inter Coding for VVC". In: *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE. 021–025.

Bjøntegaard, G. (2001). "VCEG-M33: Calculation of Average PSNR Differences between RD curves". *Video Coding Experts Group (VCEG)*: 1520–9210.

Bossen, F. *et al.* (2013). "Common test conditions and software reference configurations". *JCTVC-L1100*. 12(7).

Boyce, J., K. Suehring, X. Li, and V. Seregin. (2018). "JVET Common Test Conditions and Software Reference Configurations". *Joint Vdieo Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-J1010-v1*. Apr.

Cai, C., G. Lu, Q. Hu, L. Chen, and Z. Gao. (2019). "Efficient Learning Based Sub-pixel Image Compression". In: *CVPR Workshops*. 4.

Cao, C., M. Preda, and T. Zaharia. (2019). "3D Point Cloud Compression: A Survey". In: *The 24th International Conference on 3D Web Technology*. 1–9.

Carreira, J. and A. Zisserman. (2017). "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset". In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6299–6308.

Chen, Z., K. Fan, S. Wang, L. Duan, W. Lin, and A. C. Kot. (2019). "Toward Intelligent Sensing: Intermediate Deep Feature Compression". *IEEE Transactions on Image Processing*. 29: 2230–2243.

Cheng, Z., H. Sun, M. Takeuchi, and J. Katto. (2020). "Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7939–7948.

Choi, Y., M. El-Khamy, and J. Lee. (2019). "Variable Rate Deep Image Compression With a Conditional Autoencoder". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3146–3154.

CISCO. (2016). "CISCO VNI Complete Forecast Highlights 2021".

Dai, Y., D. Liu, and F. Wu. (2017). "A Convolutional Neural Network Approach for Post-Processing in HEVC Intra Coding". In: *International Conference on Multimedia Modeling*. Springer. 28–39.

Dang-Nguyen, D.-T., C. Pasquini, V. Conotter, and G. Boato. (2015). "RAISE: A Raw Images Dataset for Digital Image Forensics". In: *Proceedings of the 6th ACM multimedia systems conference*. 219–224.

De Queiroz, R. L. and P. A. Chou. (2016). "Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform". *IEEE Transactions on Image Processing*. 25(8): 3947–3956.

De Queiroz, R. L. and P. A. Chou. (2017). "Transform Coding for Point Clouds Using a Gaussian Process Model". *IEEE Transactions on Image Processing*. 26(7): 3507–3517.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. (2009). "ImageNet: A Large-Scale Hierarchical Image Database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. DOI: 10.1109/CVPR.2009.5206848.

Djelouah, A., J. Campos, S. Schaub-Meyer, and C. Schroers. (2019). "Neural Inter-Frame Compression for Video Coding". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6421–6429.

Dong, C., Y. Deng, C. C. Loy, and X. Tang. (2015). "Compression Artifacts Reduction by a Deep Convolutional Network". In: *Proceedings of the IEEE International Conference on Computer Vision*. 576–584.

Dong, C., C. C. Loy, K. He, and X. Tang. (2014). "Learning a Deep Convolutional Network for Image Super-Resolution". In: *European Conference on Computer Vision*. Springer. 184–199.

Duan, L., J. Liu, W. Yang, T. Huang, and W. Gao. (2020). "Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics". *IEEE Transactions on Image Processing*. 29: 8680–8695.

Egiazarian, K., J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. (2006). "New Full-Reference Quality Metrics Based on HVS". In: *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*. Vol. 4.

Fan, Y., H. Sun, J. Katto, J. Ming'E, *et al.* (2020). "A Fast QTMT Partition Decision Strategy for VVC Intra Prediction". *IEEE Access*. 8: 107900–107911.

Fan, Y., Y. Huang, and J. Peng. (2013). "Point Cloud Compression Based on Hierarchical Point Clustering". In: *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE. 1–7.

Fatima, A., W. Hussain, and S. Rasool. (2021). "Grey is the new RGB: How good is GAN-based image colorization for image compression?" *Multimedia Tools and Applications*. 80(3): 3775–3791.

Fischer, K., F. Brand, C. Herglotz, and A. Kaup. (2020). "Video Coding for Machines with Feature-Based Rate-Distortion Optimization". In: *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 1–6.

Flickr. (2022). "Flickr homepage". URL: https://flickr.com.

Garcia, D. C. and R. L. de Queiroz. (2018). "Intra-Frame Context-Based Octree Coding for Point-Cloud Geometry". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE. 1807–1811.

Geiger, A., P. Lenz, C. Stiller, and R. Urtasun. (2013). "Vision Meets Robotics: The KITTI Dataset". *The International Journal of Robotics Research*. 32(11): 1231–1237.

Google. (2015). "WebP Compression Study". URL: https://developers.google.com/speed/webp/docs/webp_study.

Goyal, V. (2001). "Theoretical Foundations of Transform Coding". *IEEE Signal Processing Magazine*. 18(5): 9–21. DOI: 10.1109/79.952802.

Gumhold, S., Z. Kami, M. Isenburg, and H.-P. Seidel. (2005). "Predictive Point-Cloud Compression". In: *ACM SIGGRAPH 2005 Sketches*. 137–es.

Habibian, A., T. v. Rozendaal, J. M. Tomczak, and T. S. Cohen. (2019). "Video Compression with Rate-Distortion Autoencoders". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7033–7042.

He, T., S. Sun, Z. Guo, and Z. Chen. (2019). "Beyond Coding: Detection-driven Image Compression with Semantically Structured Bit-stream". In: *2019 Picture Coding Symposium (PCS)*. 1–5. DOI: 10.1109/PCS48520.2019.8954525.

Hoang, T. M. and J. Zhou. (2021). "Recent Trending on Learning Based Video Compression: A Survey". *Cognitive Robotics*. 1: 145–158.

Hu, Y., S. Yang, W. Yang, L.-Y. Duan, and J. Liu. (2020). "Towards Coding for Human and Machine Vision: A Scalable Image Coding Approach". In: *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 1–6.

Hu, Z., G. Lu, and D. Xu. (2021). "FVC: A New Framework Towards Deep Video Compression in Feature Space". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1502–1511.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. (2017). "Densely Connected Convolutional Networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

Huang, H., I. Schiopu, and A. Munteanu. (2019). "Deep Learning Based Angular Intra-Prediction for Lossless HEVC Video Coding". In: *2019 Data Compression Conference (DCC)*. IEEE. 579–579.

Huang, Y., J. Peng, C.-C. J. Kuo, and M. Gopi. (2006). "Octree-Based Progressive Geometry Coding of Point Clouds." In: *PBG@ SIGGRAPH*. 103–110.

Iizuka, S., E. Simo-Serra, and H. Ishikawa. (2017). "Globally and Locally Consistent Image Completion". *ACM Transactions on Graphics (ToG)*. 36(4): 1–14.

ITU. (2011). "Recommendation ITU-R BT.601-7: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios". URL: https://www.itu.int/rec/R-REC-BT.601/.

Johnston, N., D. Vincent, D. Minnen, M. Covell, S. Singh, T. Chinen, S. J. Hwang, J. Shor, and G. Toderici. (2018). "Improved Lossy Image Compression with Priming and Spatially Adaptive Bit Rates for Recurrent Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4385–4393.

Kathariya, B., L. Li, Z. Li, J. Alvarez, and J. Chen. (2018). "Scalable point cloud geometry coding with binary tree embedded quadtree". In: *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 1–6.

Kay, W., J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, *et al.* (2017). "The Kinetics Human Action Video Dataset". *arXiv preprint arXiv:1705.06950*.

Kingma, D. P. and M. Welling. (2013). "Auto-Encoding Variational Bayes". *arXiv preprint arXiv:1312.6114*.

Kitago, M. and M. Gopi. (2006). "Efficient and Prioritized Point Subsampling for CSRBF Compression." In: *PBG@ SIGGRAPH*. Citeseer. 121–128.

Krasin, I., T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, A. Veit, *et al.* (2017). "OpenImages: A public dataset for large-scale multi-label and multi-class image classification". *Dataset available from https://github.com/openimages*. 2(3): 18.

Krishnaraj, N., M. Elhoseny, M. Thenmozhi, M. M. Selim, and K. Shankar. (2020). "Deep learning model for real-time image compression in Internet of Underwater Things (IoUT)". *Journal of Real-Time Image Processing*. 17(6): 2097–2111.

Kudo, S., S. Orihashi, R. Tanida, and A. Shimizu. (2019). "GAN-based Image Compression Using Mutual Information Maximizing Regularization". In: *2019 Picture Coding Symposium (PCS)*. 1–5. DOI: 10.1109/PCS48520.2019.8954548.

Ladune, T., P. Philippe, W. Hamidouche, L. Zhang, and O. Déforges. (2020). "Binary Probability Model for Learning Based Image Compression". In: *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2168–2172. DOI: 10.1109/ICASSP40776.2020.9053997.

Larson, E. C. and D. M. Chandler. (2010). "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy". *Journal of electronic imaging*. 19(1): 011006.

Le Callet, P. and F. Autrusseau. (2005). "Subjective Auality Assessment IRCCyN/IVC Database".

Lee, J., S. Cho, and S.-K. Beack. (2018). "Context-adaptive Entropy Model for End-to-end Optimized Image Compression". In: *International Conference on Learning Representations*.

Lee, J., S. Cho, and M. Kim. (2019). "An End-to-End Joint Learning Scheme of Image compression and Quality Enhancement with Improved Entropy Minimization". *arXiv preprint arXiv:1912.12817*.

Li, J., B. Li, J. Xu, R. Xiong, and W. Gao. (2018a). "Fully Connected Network-based Intra Prediction for Image Coding". *IEEE Transactions on Image Processing*. 27(7): 3236–3247.

Li, M. (2019). "A Better Color Space Conversion Based on Learned Variances for Image Compression". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 5.

Li, M., C. Xia, J. Hu, Z. Huang, Y. Zhang, D. Chen, J. Zan, G. Li, and J. Nie. (2019). "VimicroABCnet: An Image Coder Combining A Better Color Space Conversion Algorithm and A Post Enhancing Networ." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 5.

Li, M., W. Zuo, S. Gu, D. Zhao, and D. Zhang. (2018b). "Learning Convolutional Networks for Content-weighted Image Compression". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3214–3223.

Li, T., M. Xu, R. Tang, Y. Chen, and Q. Xing. (2021). "DeepQTMT: A Deep Learning Approach for Fast QTMT-Vased CU Partition of Intra-Mode VVC". *IEEE Transactions on Image Processing*. 30: 5377–5390.

Li, X. and S. Ji. (2020). "Neural Image Compression and Explanation". *IEEE Access*. 8: 214605–214615. DOI: 10.1109/ACCESS.2020.3041416.

Lin, C., J. Yao, F. Chen, and L. Wang. (2020a). "A Spatial RNN Codec for End-to-End Image Compression". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13269–13277.

Lin, J., D. Liu, H. Li, and F. Wu. (2020b). "M-LVC: Multiple Frames Prediction for Learned Video Compression". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3546–3554.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. (2014). "Microsoft COCO: Common Objects in Context". In: *European conference on computer vision*. Springer. 740–755.

Liu, D., Y. Li, J. Lin, H. Li, and F. Wu. (2020a). "Deep Learning-based Video Coding: A Review and A Case Study". *ACM Computing Surveys (CSUR)*. 53(1): 1–35.

Liu, H., T. Chen, P. Guo, Q. Shen, and Z. Ma. (2019). "Gated Context Model with Embedded Priors for Deep Image Compression". *arXiv preprint arXiv:1902.10480*.

Liu, J., G. Lu, Z. Hu, and D. Xu. (2020b). "A Unified End-to-End Framework for Efficient Deep Image Compression". *arXiv preprint arXiv:2002.03370*.

Liu, Z., P. Luo, X. Wang, and X. Tang. (2015). "Deep Learning Face Attributes in the Wild". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 3730–3738.

Lotter, W., G. Kreiman, and D. Cox. (2016). "Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning". *arXiv preprint arXiv:1605.08104*.

Lu, G., C. Cai, X. Zhang, L. Chen, W. Ouyang, D. Xu, and Z. Gao. (2020). "Content Adaptive and Error Propagation Aware Deep Video Compression". In: *European Conference on Computer Vision*. Springer. 456–472.

Lu, G., W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao. (2019a). "DVC: An End-to-End Deep Video Compression Framework". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11006–11015.

Lu, M., T. Chen, H. Liu, and Z. Ma. (2019b). "Learned Image Restoration for VVC Intra Coding". In: *CVPR Workshops*. 4.

Ma, C., D. Liu, X. Peng, L. Li, and F. Wu. (2019a). "Convolutional Neural Network-based Arithmetic Coding for HEVC Intra-predicted Residues". *IEEE Transactions on Circuits and Systems for Video Technology*. 30(7): 1901–1916.

Ma, C., D. Liu, X. Peng, and F. Wu. (2018). "Convolutional Neural Network-based Arithmetic Coding of DC Coefficients for HEVC Intra Coding". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE. 1772–1776.

Ma, D., F. Zhang, and D. Bull. (2021). "BVI-DVC: A Training Database for Deep Video Compression". *IEEE Transactions on Multimedia*.

Ma, H., D. Liu, R. Xiong, and F. Wu. (2019b). "A CNN-Based Image Compression Scheme Compatible with JPEG-2000". In: *2019 IEEE International Conference on Image Processing (ICIP)*. 704–708. DOI: 10.1109/ICIP.2019.8803835.

Marpe, D., H. Schwarz, and T. Wiegand. (2003). "Context-based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard". *IEEE Transactions on circuits and systems for video technology*. 13(7): 620–636.

Mentzer, F., E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool. (2018). "Conditional Probability Models for Deep Image Compression". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4394–4402.

Mercat, A., M. Viitanen, and J. Vanne. (2020). "UVG Dataset: 50/120fps 4K Sequences for Video Codec Analysis and Development". In: *Proceedings of the 11th ACM Multimedia Systems Conference*. 297–302.

Merry, B., P. Marais, and J. Gain. (2006). "Compression of Dense and Regular Point Clouds". In: *Proceedings of the 4th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*. 15–20.

Minnen, D., J. Ballé, and G. Toderici. (2018). "Joint Autoregressive and Hierarchical Priors for Learned Image Compression". *Advances in neural information processing systems*. 31.

Minnen, D. and S. Singh. (2020). "Channel-Wise Autoregressive Entropy Models for Learned Image Compression". In: *2020 IEEE International Conference on Image Processing (ICIP)*. 3339–3343. DOI: 10.1109/ICIP40778.2020.9190935.

MPEG. (2020). "Evaluation Framework for Video Coding for Machines". *ISO/IEC JTC1/SC29/WG11, W19366*. Apr.

Nowell, M. (2019). "Cisco VNI Forecast Update".

Ochotta, T. and D. Saupe. (2004). *Compression of Point-Based 3D Models by Shape-Adaptive Wavelet Coding of Multi-Height Fields*.

Ohm, J.-R. and G. J. Sullivan. (2018). "Versatile video coding–towards the next generation of video compression". In: *Picture Coding Symposium*. Vol. 2018.

Park, W.-S. and M. Kim. (2016). "CNN-Based In-Loop Filtering for Coding Efficiency Improvement". In: *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. 1–5. DOI: 10.1109/IVMSPW.2016.7528223.

Park, W. and M. Kim. (2021). "Deep Predictive Video Compression Using Mode-Selective Uni- and Bi-Directional Predictions Based on Multi-Frame Hypothesis". *IEEE Access*. 9: 72–85. DOI: 10.1109/ACCESS.2020.3046040.

Pauly, M. and M. Gross. (2001). "Spectral Processing of Point-Sampled Geometry". In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. 379–386.

Perry, S. (2021). "JPEG Pleno Point Cloud - Use Cases and Requirements v1.4". *ISO/IEC JTC1/SC29/WG1 N92015, 92th Meeting, Online*. July.

Ponomarenko, N., F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin. (2007). "On Between-Coefficient Contrast Masking of DCT Basis Functions". In: *Proceedings of the Third International Workshop on Video Processing and Quality Metrics*. Vol. 4. Scottsdale USA.

Rippel, O. and L. Bourdev. (2017). "Real-Time Adaptive Image Compression". In: *International Conference on Machine Learning*. PMLR. 2922–2930.

Rissanen, J. and G. Langdon. (1981). "Universal Modeling and Coding". *IEEE Transactions on Information Theory*. 27(1): 12–23. DOI: 10.1109/TIT.1981.1056282.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.* (2015). "ImageNet Large Scale Visual Recognition Challenge". *International journal of computer vision*. 115(3): 211–252.

Santurkar, S., D. Budden, and N. Shavit. (2018). "Generative Compression". In: *2018 Picture Coding Symposium (PCS)*. 258–262. DOI: 10.1109/PCS.2018.8456298.

Schaefer, G. and M. Stich. (2003). "UCID: An Uncompressed Color Image Database". In: *Storage and Retrieval Methods and Applications for Multimedia 2004*. Vol. 5307. SPIE. 472–480.

Schiopu, I., H. Huang, and A. Munteanu. (2019). "CNN-Based Intra-Prediction for Lossless HEVC". *IEEE Transactions on Circuits and Systems for Video Technology*. 30(7): 1816–1828.

Sheikh, H., M. Sabir, and A. Bovik. (2006). "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms". *IEEE Transactions on Image Processing*. 15(11): 3440–3451. DOI: 10.1109/TIP.2006.881959.

Sheikh, H. (2005). "LIVE image quality assessment database release 2". *http://live.ece.utexas.edu/research/quality*.

Sneyers, J. and P. Wuille. (2016). "FLIF: FREE LOSSLESS IMAGE FORMAT BASED ON MANIAC COMPRESSION". In: *2016 IEEE International Conference on Image Processing (ICIP)*. 66–70. DOI: 10.1109/ICIP.2016.7532320.

Sullivan, G. J., J.-R. Ohm, W.-J. Han, and T. Wiegand. (2012). "Overview of the High Efficiency Video Coding (HEVC) Standard". *IEEE Transactions on Circuits and Systems for Video Technology*. 22(12): 1649–1668. DOI: 10.1109/TCSVT.2012.2221191.

Sun, H., C. Liu, J. Katto, and Y. Fan. (2020). "An Image Compression Framework with Learning-based Filter". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 152–153.

Tao, W., F. Jiang, S. Zhang, J. Ren, W. Shi, W. Zuo, X. Guo, and D. Zhao. (2017). "An End-to-End Compression Framework Based on Convolutional Neural Networks". In: *2017 Data Compression Conference (DCC)*. 463–463. DOI: 10.1109/DCC.2017.54.

Taubman, D. and M. Marcellin. (2002). "JPEG2000: Standard for Interactive Imaging". *Proceedings of the IEEE*. 90(8): 1336–1357. DOI: 10.1109/JPROC.2002.800725.

Thanou, D., P. A. Chou, and P. Frossard. (2016). "Graph-Based Compression of Dynamic 3D Point Cloud Sequences". *IEEE Transactions on Image Processing*. 25(4): 1765–1778.

Theis, L., W. Shi, A. Cunningham, and F. Huszár. (2017). "Lossy Image Compression with Compressive Autoencoders". *arXiv preprint arXiv:1703.00395*.

Thomee, B., D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. (2016). "YFCC100M: The New Data in Multimedia Research". *Communications of the ACM*. 59(2): 64–73.

Toderici, G., S. M. O'Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar. (2015). "Variable Rate Image Compression with Recurrent Neural Networks". *arXiv preprint arXiv:1511.06085*.

Toderici, G., W. Shi, R. Timofte, L. Theis, J. Balle, E. Agustsson, N. Johnston, and F. Mentzer. (2020). "Workshop and Challenge on Learned Image Compression (CLIC2020)". URL: http://www.compression.cc.

Toderici, G., D. Vincent, N. Johnston, S. Jin Hwang, D. Minnen, J. Shor, and M. Covell. (2017). "Full Resolution Image Compression with Recurrent Neural Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5306–5314.

UGC. (2022). "YouTube UGC Dataset". URL: https://media.withyoutube.com/.

Van den Oord, A., N. Kalchbrenner, L. Espeholt, K. Kavukcuoglu, O. Vinyals, and A. Graves. (2016). "Conditional Image Generation with PixelCNN Decoders". In: *Advances in Neural Information Processing Systems*. Vol. 29. 4790–4798.

Van Oord, A., N. Kalchbrenner, and K. Kavukcuoglu. (2016). "Pixel Recurrent Neural Networks". In: *International Conference on Machine Learning*. PMLR. 1747–1756.

VTL. (2022). "VIdeo Trace Library (VTL homepage)". URL: http://trace.eas.asu.edu/yuv/index.html.

Wainwright, M. J. and E. P. Simoncelli. (1999). "Scale Mixtures of Gaussians and the Statistics of Natural Images". In: *Advances in Neural Information Processing Systems (NIPS)*. Vol. 12. 855–861.

Wallace, G. (1992). "The JPEG Still Picture Compression Standard". *IEEE Transactions on Consumer Electronics*. 38(1): xviii–xxxiv. DOI: 10.1109/30.125072.

Wang, H., W. Gan, S. Hu, J. Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo. (2016). "MCL-JCV: a JND-based H. 264/AVC Video Quality Assessment Dataset". In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 1509–1513.

Wang, T., M. Chen, and H. Chao. (2017). "A Novel Deep Learning-Based Method of Improving Coding Efficiency from the Decoder-End for HEVC". In: *2017 Data Compression Conference (DCC)*. 410–419. DOI: 10.1109/DCC.2017.42.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. (2004). "Image Quality Assessment: From Error Visibility to Structural Similarity". *IEEE transactions on image processing*. 13(4): 600–612.

Wang, Z., E. P. Simoncelli, and A. C. Bovik. (2003). "Multiscale Structural Similarity for Image Quality Assessment". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 2. Ieee. 1398–1402.

WG1. (2022). "Final Call for Proposals for JPEG AI". *ISO/IEC JTC 1/SC29/WG1, N100095, 94th JPEG Meeting, Online*. Jan.

Wiegand, T., G. Sullivan, G. Bjontegaard, and A. Luthra. (2003). "Overview of the H.264/AVC Video Coding Standard". *IEEE Transactions on Circuits and Systems for Video Technology*. 13(7): 560–576. DOI: 10.1109/TCSVT.2003.815165.

Wu, C.-Y., N. Singhal, and P. Krahenbuhl. (2018). "Video Compression Through Image Interpolation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 416–431.

Xia, S., K. Liang, W. Yang, L.-Y. Duan, and J. Liu. (2020). "An Emerging Coding Paradigm VCM: A Scalable Coding Approach Beyond Feature and Signal". In: *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 1–6.

Xue, T., B. Chen, J. Wu, D. Wei, and W. T. Freeman. (2019). "Video Enhancement with Task-Oriented Flow". *International Journal of Computer Vision*. 127(8): 1106–1125.

Yan, N., D. Liu, H. Li, B. Li, L. Li, and F. Wu. (2018). "Convolutional Neural Network-based Fractional-pixel Motion Compensation". *IEEE Transactions on Circuits and Systems for Video Technology*. 29(3): 840–853.

Yang, R., F. Mentzer, L. V. Gool, and R. Timofte. (2020). "Learning for Video Compression with Hierarchical Quality and Recurrent Enhancement". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6628–6637.

Yang, R., M. Xu, T. Liu, Z. Wang, and Z. Guan. (2019). "Enhancing Quality for HEVC Compressed Videos". *IEEE Transactions on Circuits and Systems for Video Technology*. 29(7): 2039–2054. DOI: 10.1109/TCSVT.2018.2867568.

Yim, C. and A. C. Bovik. (2010). "Quality Assessment of Deblocked Images". *IEEE Transactions on Image Processing*. 20(1): 88–98.

Yu, A. and K. Grauman. (2014). "Fine-Grained Visual Comparisons with Local Learning". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, K., W. Zhu, and Y. Xu. (2018). "Hierarchical Segmentation Based Point Cloud Attribute Compression". In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 3131–3135.

Zhang, R., J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros. (2017). "Real-Time User-Guided Image Colorization with Learned Deep Priors". *arXiv preprint arXiv:1705.02999*.

Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. (2017). "Pyramid Scene Parsing Network". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2881–2890.

Zhou, B., A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. (2014). "Learning Deep Features for Scene Recognition using Places Database". *Advances in Neural Information Processing Systems*. 27.

Zhou, B., H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. (2017). "Scene Parsing through ADE20K Dataset". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 633–641.

Zhou, L., C. Cai, Y. Gao, S. Su, and J. Wu. (2018). "Variational Autoencoder for Low Bit-rate Image Compression". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Zhu, L., S. Kwong, Y. Zhang, S. Wang, and X. Wang. (2019). "Generative Adversarial Network-Based Intra Prediction for Video Coding". *IEEE Transactions on Multimedia*. 22(1): 45–58.