KI DATA TOOLING
The Data Kit for Automotive AI

Adaptive Bitrate Quantization Scheme Without Codebook For Learned Image Compression

Jonas Löhdefink, Jonas Sitzmann, Andreas Bär, Tim Fingscheidt | TUBS

Acquisition, Curation & Refinement of Real Data

Motivation

Integer rounding (INT) is still a wide-spread quantization method in learned image compression. It uses a scalar, uniform codebook.

Problem: Training/inference mismatch due to replacement of uniformly distributed noise addition instead of quantization during training.

Our One-Hot Max (OHM) quantization:

- Vector quantization along new dimension
- Codebook: not required, whereas prior art requires a codebook: [1]
- No training / inference mismatch
- Three bitrate control schemes supported: fixed, variable, adaptive

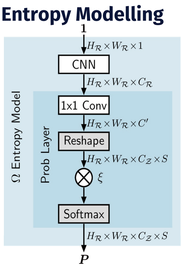
Bitrate Control Schemes:

Fixed bitrate: The quantizer discretizes the continuous feature space. The bitrate is determined by feature space size and the number of quantization levels.

Variable bitrate: Entropy coding losslessly reduces the bitrate further. The entropy model determines the bitrate and is jointly trained with the autoencoder. The rate-distortion trade-off is determined during training.

Adaptive bitrate: The bitrate is controlled via an input parameter. The rate-distortion trade-off is determined during inference.

Entropy Modelling



The entropy model predicts the probability of quantized symbols at each feature map location. The bitrate equals the cross entropy between the real and estimated distribution

Figure 1: Entropy model for OHM quantization allows direct estimation of probabilities for each reconstruction vector. The architecture is similar to a quantizer: It adds a new dimension and the final softmax. (© TUBS)

New One-Hot Max (OHM) Quantization

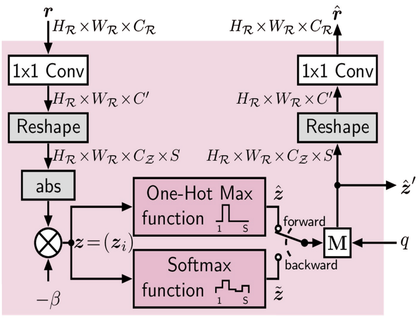


Figure 2: New one-hot max quantization where a new dimension is introduced during reshaping. Forward: One-hot max function replaces the maximum value with 1 and all remaining values with 0, Backward: Softmax function generates a smooth distribution summing up to 1. M: Feature map masking. (© TUBS)

Feature Map Masking for Adaptive Bitrates

Feature map masking (block "M" in Fig. 2) removes parts of the feature space channel-wisely. The model learns a hierarchy in the feature space channel dimension, i.e., low indexed channels contain coarse structures and high indexed channels contain additional fine-grained details. A single model generates the complete rate-distortion curve. During training, a hyperparameter is sampled from a distribution that balances training of low and high indexed feature maps.

Conclusions

We propose a new learnable quantizer scheme without the use of a codebook and without training/inference mismatch. It includes an adaptive bitrate compression system by feature map masking during inference.

MNIST: OHM quantizer exceeds INT quantizer in almost all conditions.

References:

[1]: E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, and L. V. Gool : "Soft-to-Hard Vector Quantization for End-to-End Learning Compressible Representations", in Proc. of NeurIPS 2017

[2]: J. Löhdefink, J. Sitzmann, A. Bär, T. Fingscheidt: "Adaptive Bitrate Quantization Scheme Without Codebook for Learned Image Compression", in Proc. of CVPR-Workshops 2022

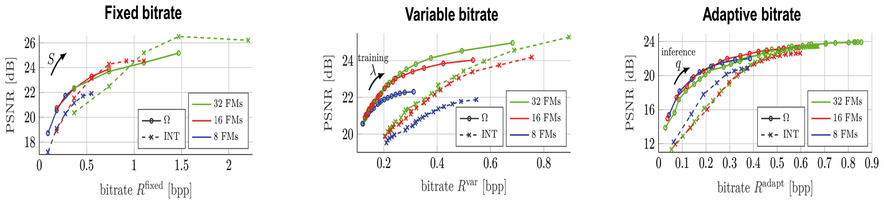


Figure 3: Evaluation on MNIST: For the fixed bitrate, OHM quantizer is better at a low bitrate, INT quantization is better at higher bitrates. For the variable bitrate, OHM quantization is better at all bitrates. For the adaptive bitrate, OHM quantization is better at all bitrates as it is better suited to learn a hierarchical structured feature space. KODAK results in [2]. (© TUBS)

Partners

External partners

For more information contact:
t.fingscheidt@tu-bs.de

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

www.ki-datatooling.de @KI_Familie KI Familie

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag