Stochastic Optimization of Vector Quantization Methods in Application to Speech and Image Processing

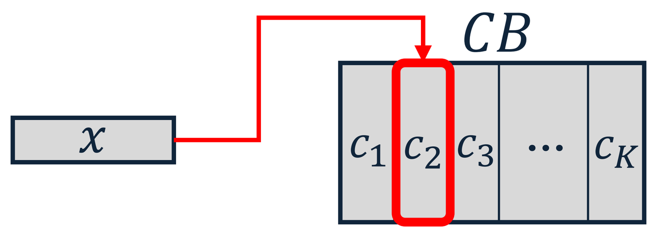
Aalto University School of Electrical Engineering

Mohammad Hassan Vali   Tom Bäckström

Department of Information and Communications Engineering, Aalto University, Finland

1. Vector Quantization (VQ)

- A data compression technique similar to k-means algorithm
- Quantizes the input vector x to the closest codeword within the codebook (CB)

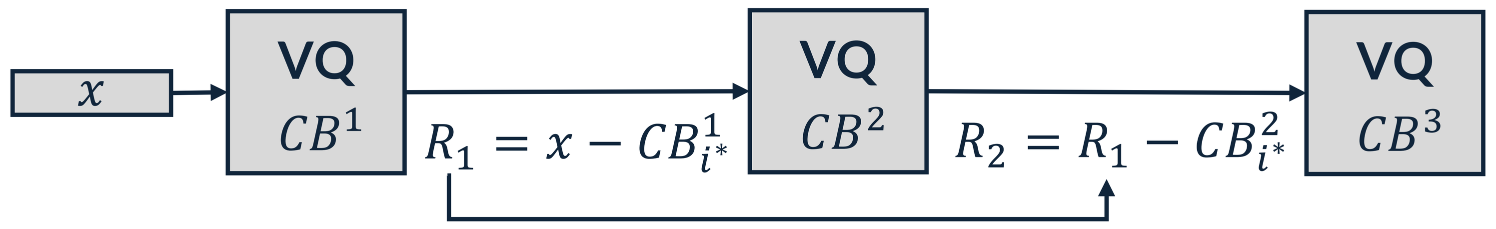


$$x_{quantized} = c_{i^*} \, ; \, i^* = \arg\min_i \|x - c_i\|^2 \, ; i \in \{1, \ldots, K\} \quad (1)$$

- **Challenge:** computationally complex for a large codebook
- **Solution:** employ variants of VQ; Residual VQ, Product VQ, and Additive VQ

2. Residual Vector Quantization (RVQ)

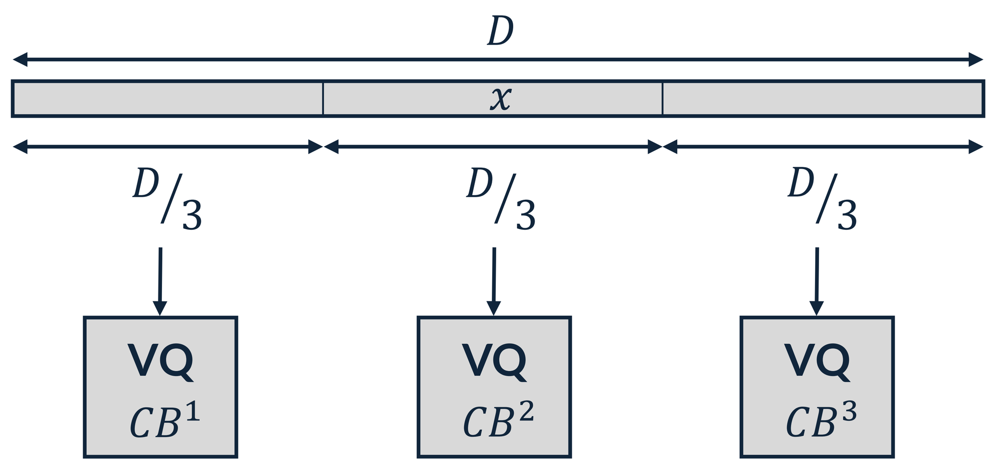
- Quantizes the input vector x by M consecutive VQ modules
- Quantizes x as a summation of M codewords
- Suppose $M = 3$:



$$x_{quantized} = CB^1_{i^*} + CB^2_{i^*} + CB^3_{i^*}$$

3. Product Vector Quantization (PVQ)

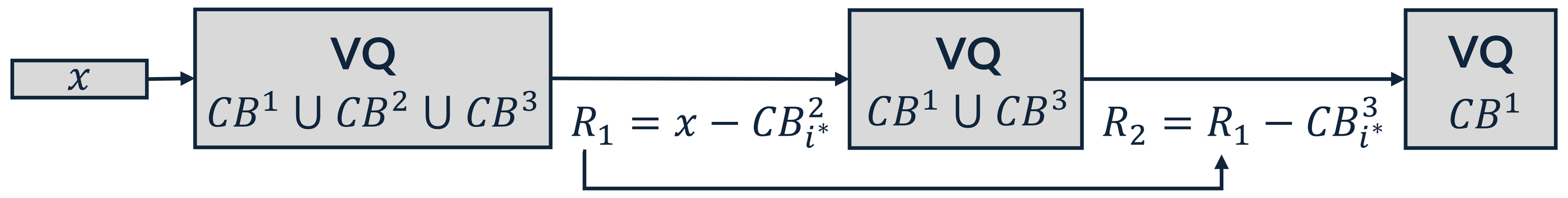
- Splits the input vector x of dimension D to M independent subspaces of dimension D/M
- Applies M independent VQ modules to the existing subspaces
- Quantizes x as a concatenation of M closest codewords
- Suppose $M = 3$:



$$x_{quantized} = concatenate[CB^1_{i^*}, CB^2_{i^*}, CB^3_{i^*}]$$

4. Additive Vector Quantization (AVQ)

- Applies beam searching [1] to find the closest codewords
- Quantizes the input vector x by M consecutive VQ modules
- Quantizes x as a summation of M codewords
- Suppose $M = 3$:



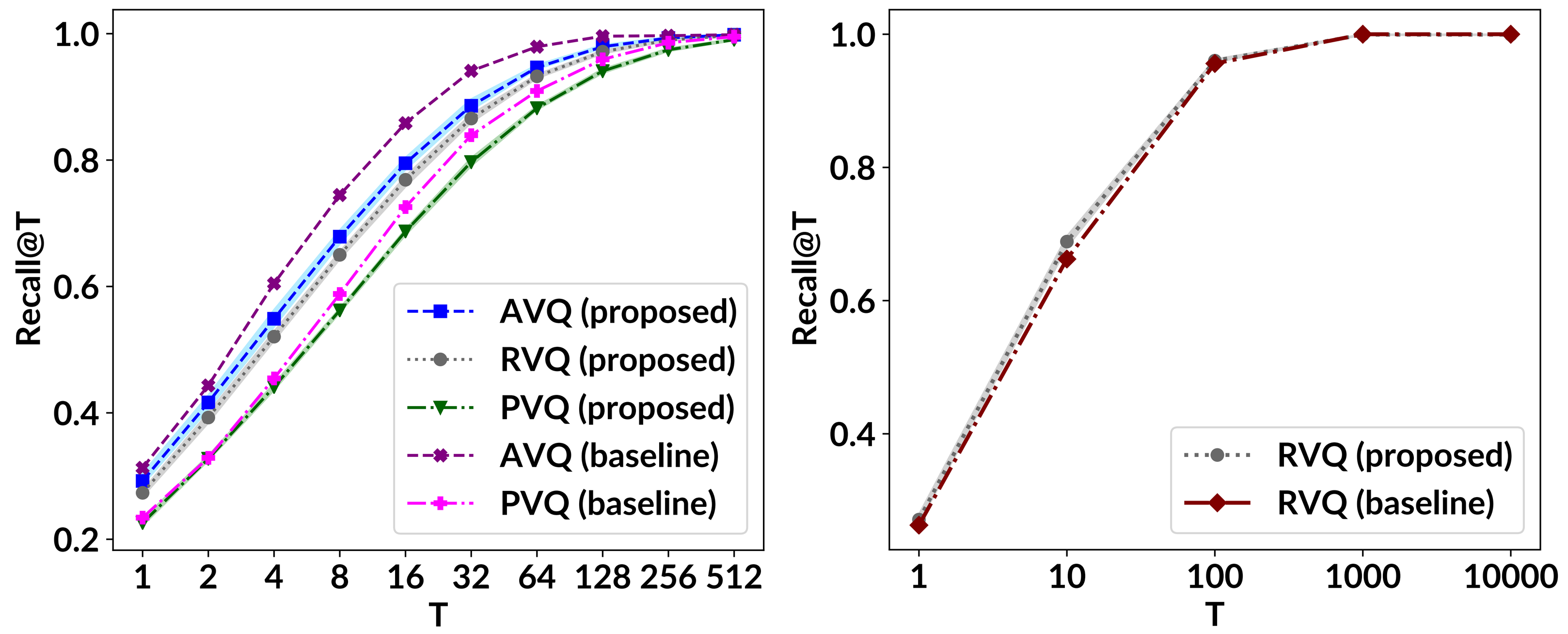
$$x_{quantized} = CB^2_{i^*} + CB^3_{i^*} + CB^1_{i^*}$$

5. Codebooks Optimization

- **Traditional approach:** Optimization by k-means algorithm
- **Problem with machine learning optimization:** argmin function in Eq. 1 is not differentiable
- **Solutions:**
  1. Proposed: Noise Substitution in Vector Quantization (NSVQ) [2]. models the quantization error by noise addition
  2. Conventional: Straight Through Estimator (STE) [3]. copies the gradients over VQ module ($VQ_{gradient} = 1$)
- **Advantages of NSVQ over STE:** 1) More accurate gradients. 2) Faster convergence. 3) No additional hyper-parameter tuning for VQ training.

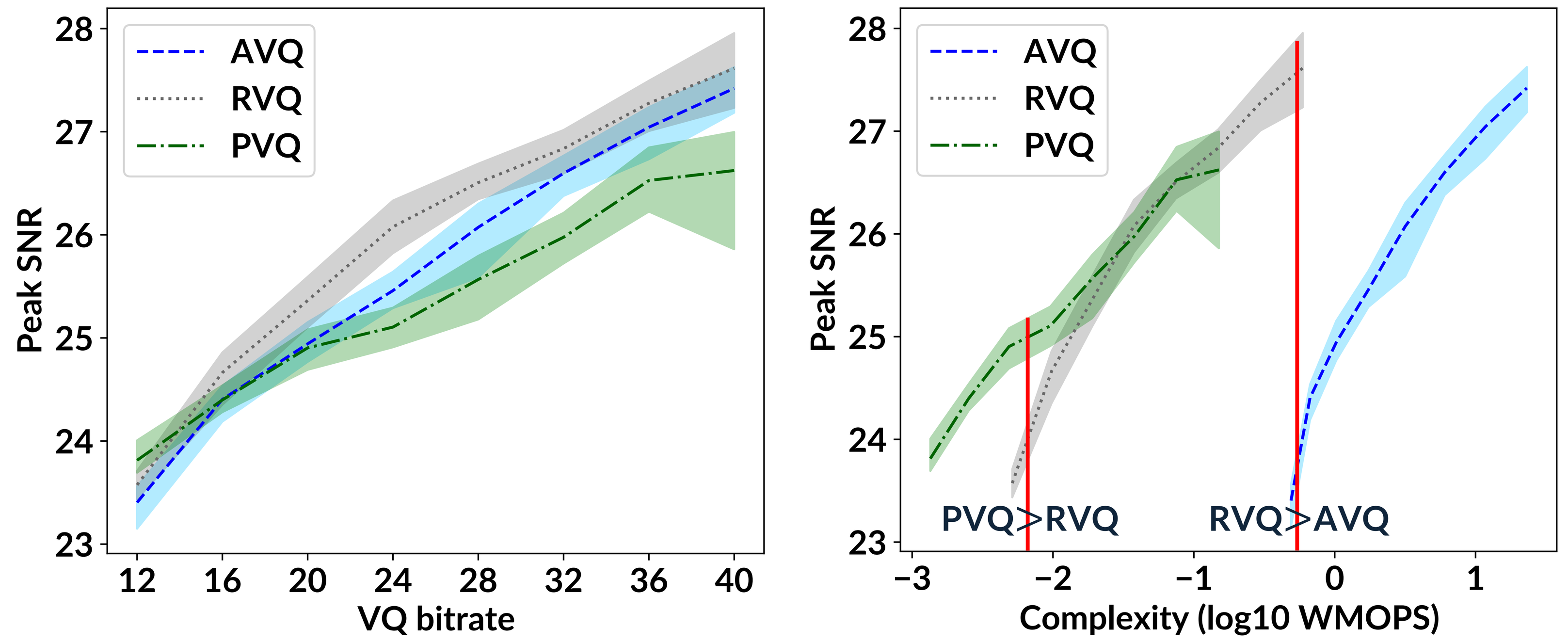
6. Experiments: Approximate Nearest Neighbor (ANN) Search

- Compress the SIFT1M dataset (128-D image descriptors)
- Evaluate recall metric: whether the actual nearest neighbor (from groundtruth) exists in the T computed nearest neighbors



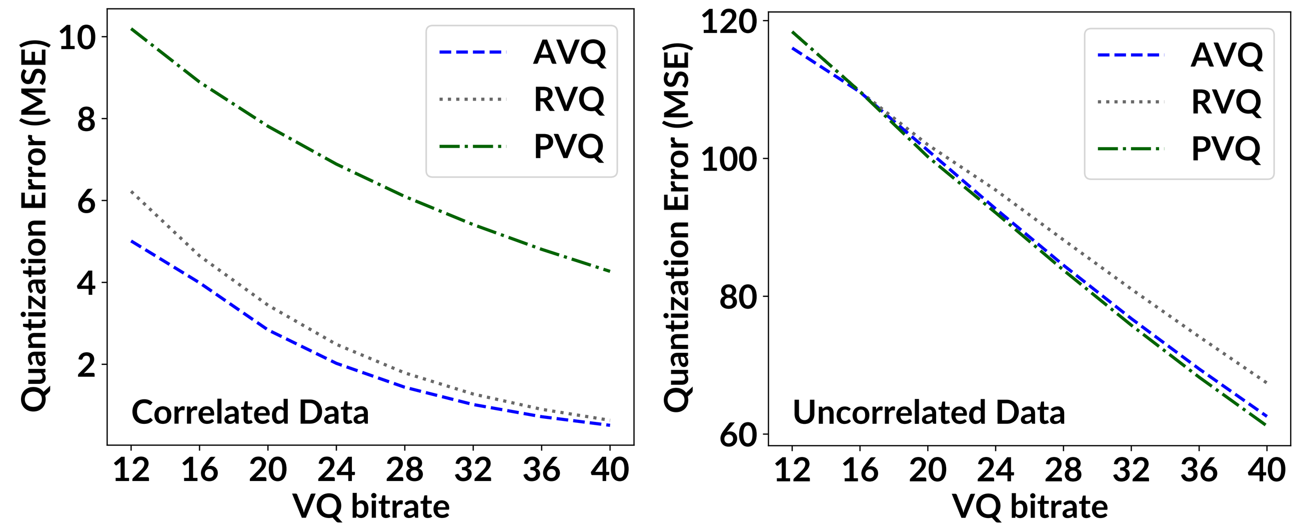
7. Experiments: Image Compression using VQ-VAE

- Compress CIFAR10 dataset with vector quantized variational autoencoder (VQ-VAE)
- Evaluate peak signal to noise ratio (Peak SNR) and complexity



8. Experiments: Toy Example Datasets

- Compress correlated and uncorrelated toy datasets
- Evaluate accuracy with respect to correlation in the data



9. Conclusions

- Variants of VQ are desirable for higher bitrates and dimensions
- Machine learning optimization of codebooks using our recently proposed NSVQ technique [2]
- Achieve comparable results to the baselines in ANN search
- Study the trade-offs between bitrate, accuracy, complexity
- Using our open source implementation [4] enables choosing the most suitable VQ method

References

[1] A. Babenko and V. Lempitsky, "Additive Quantization for Extreme Vector Compression," in *Proc. CVPR*, 2014.
[2] M. H. Vali and T. Bäckström, "NSVQ: Noise Substitution in Vector Quantization for Machine Learning," *IEEE Access*, 2022.
[3] Y. Bengio, N. Léonard, and A. Courville, "Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation," *arXiv preprint arXiv:1308.3432*, 2013.
[4] https://gitlab.com/speech-interaction-technology-aalto-university/vq-variants.