

A Survey and Benchmark Evaluation for Neural-Network-Based Lossless Universal Compressors Toward Multi-Source Data

Hui SUN, Huidong MA, Feng LING, Haonan XIE, Yongxia SUN
Liping YI, Men YAN, Cheng ZHONG, Gang WANG, Xiaoguang LIU

Frontiers of Computer Science, DOI: [10.1007/s11704-024-40300-5](https://doi.org/10.1007/s11704-024-40300-5)

Problems & Ideas

- **Problems:** In recent years, due to the emergence of hardware acceleration devices such as GPUs and TPUs, the performance bottleneck of neural networks (NN) has been overcome, making NN-based compression algorithms increasingly practical and popular. However, the research survey for the NN-based universal lossless compressors has not been conducted yet, and there is also a lack of unified evaluation metrics.
- **Ideas:** To address the above problems, in this paper, we present a holistic survey as well as benchmark evaluations.

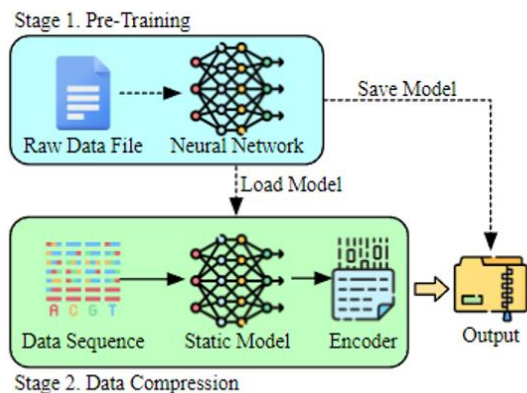


Fig. 1 The compression process of the static pre-training methods

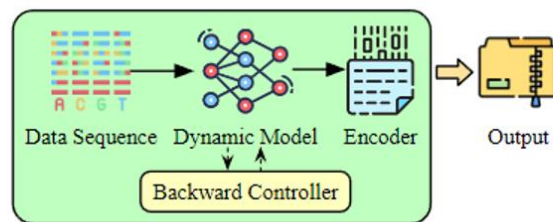


Fig. 2 Schematic of the compression process of the adaptive methods

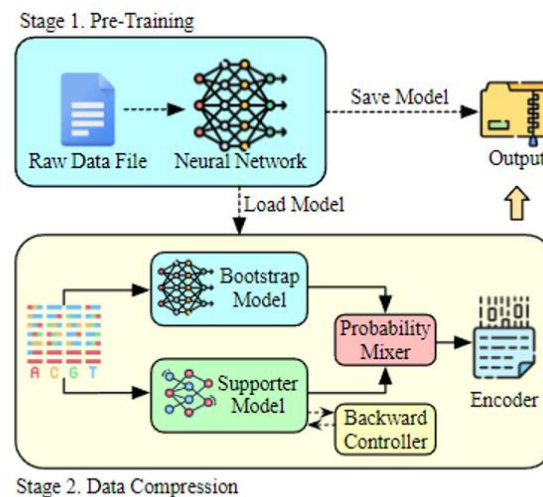


Fig. 3 The compression process of the semi-adaptive methods

Main Contributions

- Contributions:
 - We holistically reviewed NN-based universal compressors and unified the corresponding evaluation metrics.
 - We offered a comprehensive benchmark evaluation for NN-based compressors and conducted a thorough analysis of existing state-of-the-art solutions.
 - We identified the challenges of NN-based lossless compression and summarized potential research directions.

Table 1 Overall comparison of NN-based universal lossless compression methods

Method	Model Pre-training	Model Saving	Compression Ratio	Time Cost	Representative Works
Static Pre-training	✓	✓	★ ★ ★	★	Ref [78, 80, 81], DeepZip [82], DecMac [83], LLMZip [84]
Adaptive	✗	✗	★	★ ★ ★	TRACE [37], OREA [38], PAC [39], Cmix [85]
Semi-Adaptive	✓	✓	★★	★★	DZip (Combined Model) [79]

Notes. **Model Pre-training:** Whether or not the model \mathcal{M} requires pre-training. **Model Saving:** Whether or not the model need to be saved and included in the compressed file. **Compression Ratio:** A smaller value indicates better compression performance, detailed in the evaluation metrics section. The “✓” and “✗” represent “Yes” or “No”, and more instances of “★” indicate a stronger advantage.