Numerical Experiments on AI-Powered Data Compression

Objective of the Numerical Experiments

The primary goal of these numerical experiments is to develop and optimize AI-powered data compression techniques. The research aims to enhance efficiency and scalability by leveraging machine learning models to recognize patterns in data, thereby improving compression performance while maintaining or improving reconstructed **image** quality. The focus is on achieving competitive compression rates compared to existing standards such as JPEG while minimizing information loss.

Methodology

Theoretical Background

Data compression techniques are categorized into:

- **Lossless Compression**: Ensures no data loss and is ideal for text and database storage.
- **Lossy Compression**: Sacrifices some data to achieve higher compression ratios, commonly used for images, videos, and audio.
- **AI-Powered Compression**: Uses machine learning models to find optimal representations of data, improving compression ratios and decoded quality.

This research employs supervised learning techniques where the AI model minimizes reconstruction error using the Mean Squared Error (MSE) loss function. Activation functions such as ReLU and Sigmoid introduce non-linearity to help the model learn complex patterns. Different architectures, ranging from 2-layer to 5-layer encoder-decoder models, are evaluated for their impact on compression efficiency and output quality.

Implementation Details

Codebase Structure

- **/src**: Python scripts implementing compression algorithms.
- **/docs**: Documentation including setup guides and tutorials.
- **/examples**: Predefined cases demonstrating image compression.

Dataset and Tools

- **Dataset**: Flickr Faces Dataset Resized (open-source image dataset, 128x128 px resolution)
- **Tools & Libraries**: Python, PyTorch, NumPy, OpenCV, and compression techniques such as gzip and LZMA.

Model Training & Evaluation

- Training is conducted on datasets ranging from 30 to 52,000 images.

- Evaluation metrics include encoded file size, decoded image quality, and training time.
- Compression performance is tested against traditional methods like JPEG.

Results and Analysis

Version 1 (V1)

- **Architecture**: 2-layer Encoder-Decoder, model size 0.1MB
- **Performance**:
  - 30-image dataset: Encoded 100 KB → Decoded 26.4 KB
  - 52,000-image dataset: Encoded 100 KB → Decoded 30 KB
- **Training Time**:
  - 30-image dataset: 1 minute
  - 52,000-image dataset: 170 minutes
- **Findings**:
  - Decoded image quality is superior to JPEG.
  - Encoded image size is excessively large, requiring further optimizations.

Version 2 (V2)

- **Architecture**: 3-layer Encoder-Decoder, model size 0.4MB
- **Performance**:
  - 30-image dataset: Encoded 12.6 KB → Decoded 31.6 KB
  - 52,000-image dataset: Encoded 13.1 KB → Decoded 26.9 KB
- **Training Time**:
  - 30-image dataset: 30 seconds
  - 52,000-image dataset: 121 minutes
- **Findings**:
  - File size reduced significantly.
  - Quality requires further improvements to compete with JPEG.

Version 3 (V3)

- **Architecture**: 5-layer Encoder-Decoder, model size 4MB
- **Performance**:

    - 30-image dataset: Encoded 4.62 KB → Decoded 29 KB
    - 52,000-image dataset: Encoded 5.8 KB → Decoded 25 KB
- **Training Time**:
    - 30-image dataset: 1 minute
    - 52,000-image dataset: 250 minutes
- **Compression Enhancements**:
    - gzip: 5.89 KB (no loss)
    - LZMA: 5.58 KB (no loss)
    - LZMA + Quantization (FP16 → INT8): 2.9 KB (minimal loss)
- **Findings**:
    - Compression performance matches JPEG levels.
    - Image quality requires optimization.

Version 4 (V4) – Hypothesis Testing

- **Approach**: Introduced skip connections (U-Net style) to enhance quality.
- **Preliminary Results**:
    - Final loss remains unchanged (FN = 0.0039), indicating the need for further refinements.

Discussion of Findings in Relation to Information Theory

The experiments align with core information theory concepts:

- **Entropy and Redundancy**: The AI models learn efficient data representations, minimizing entropy while preserving essential features.
- **Rate-Distortion Tradeoff**: Achieving lower file sizes while maintaining image quality reflects this fundamental balance.
- **Kolmogorov Complexity**: The model approximates an optimal data representation, reducing unnecessary information.
- **Error Minimization**: The use of MSE loss directly correlates with Shannon's information loss measurement.

Conclusions and Possible Improvements

Key Takeaways

1. **AI-based compression achieves good quality** compared to traditional methods but requires optimization in encoding size.

2. **LZMA + Quantization effectively reduces file sizes** without significant loss.
3. **Deep architectures improve efficiency**, but training times increase significantly.
4. **Skip connections (V4) show potential** for improving quality while maintaining compression.

Future Work

- Optimize neural network layers and loss functions to balance quality and compression.
- Implement advanced quantization techniques to reduce model size further.
- Explore alternative architectures such as transformer-based compression models.
- Test different datasets to generalize the model's effectiveness across various image types.

By addressing these areas, AI-powered compression can become a viable alternative to conventional methods, offering efficiency gains in real-world applications.