EQODEC: A Carbon-Aware Deep Learning Framework for Sustainable Video Compression

Ruszed Ayad, Ian Jure Macalisang, Honey Angel Pabololot, and Christine Joy Sorronda

Abstract. Growing demand for online video has intensified the environmental impact of data storage and transmission, motivating environmentally sustainable compression methods. This study introduces a carbon-aware deep learning framework (EQODEC) that integrates estimated energy use and CO_2 emissions directly into the training objective of a neural video codec. The approach employs a spatiotemporal autoencoder with ConvGRU-based temporal modeling and differentiable quantization, optimized through a multi-objective loss balancing quality, bitrate, and carbon cost. Experiments on the Vimeo-90K dataset across five independent replicates demonstrate consistent sustainability gains, achieving a higher Energy Efficiency Score than a baseline model, while preserving perceptual quality and incurring no additional computational or emissions overhead.

Keywords: Carbon-aware learning; Neural video compression; Green AI; Energy-efficient deep learning; Spatiotemporal autoencoder; Sustainability metrics; Energy Efficiency Score (EES)

1 Background

The rapid expansion of online video services has made video data the single largest contributor to global internet traffic, accounting for more than 80% of all transmitted information (Digital Age, 2023). This surge in demand has placed immense pressure on data centers, which consume massive amounts of energy to store, process, and deliver video content. As a result, the video streaming industry has become a significant source of carbon emissions, contributing to the environmental footprint of digital technologies (Afzal et al., 2024; The Shift Project, 2019).

Traditional codecs such as H.264 and H.265 have achieved impressive compression ratios through handcrafted engineering, yet they are approaching their practical limits, particularly as high-resolution formats (4K, 8K, VR) become mainstream (ImageKit, 2022). Recent advances in deep learning have opened new possibilities for neural video compression, where models such as autoencoders, GANs, and transformers learn to compress data by discovering efficient latent representations (Chen et al., 2024). However, most of these approaches prioritize reconstruction quality and bitrate efficiency, without explicitly considering their environmental impact (MIT News, 2025).

This project introduces EQODEC or the Energy-Quality Optimized Codec, a carbon-aware deep learning framework designed to make video compression both intelligent and sustainable. Unlike traditional methods, EQODEC integrates environmental cost metrics—including estimated energy use and CO_2 emissions—directly into its learning objective. By jointly optimizing for reconstruction quality, bitrate reduction, and sustainability, the model aims to reduce data center load and network energy consumption. This approach aligns with global initiatives in Green AI, promoting computational efficiency and environmental responsibility in deep learning applications (Freitag et al., 2021; Goldverg et al., 2024).

2 Objectives

1. Design a carbon-aware video compression framework that balances efficiency and environmental impact.

2. Develop an end-to-end deep learning model with an energy-aware loss to encourage eco-friendly compression.

3. Evaluate compression quality, bitrate efficiency, and carbon footprint, comparing against a baseline neural codec.

3 Methodology

The EQODEC framework was developed as an environmentally conscious deep learning system for neural video compression. Its methodology integrates dataset construction, preprocessing, model architecture, carbon-aware optimization, and a comprehensive evaluation pipeline that accounts for both conventional quality metrics and sustainability-oriented measures. Unlike traditional neural compression systems that optimize solely for distortion or rate, EQODEC embeds carbon awareness directly into its training objective, enabling the model to learn representations that are simultaneously compact, high quality, and environmentally economical.

3.1 Dataset

The experiments were conducted using the Vimeo-90K Septuplet dataset, which contains 91,701 seven-frame video sequences. Each sequence is sampled from diverse natural scenes with significant variation in illumination, motion patterns, and textures, providing an ideal foundation for learning spatiotemporal redundancies. The uniform resolution of 448×256 ensures consistent feature map alignment across sequences, simplifying batch processing and temporal modeling. The dataset is widely regarded as a standard benchmark for tasks involving frame prediction, interpolation, and compression, as its septuplet structure captures fine-grained short-range motion and realistic dynamics essential for temporal modeling in video codecs.

A reproducible 1% subset was selected to reduce computational cost while maintaining representative distributional characteristics. This resulted in 917 sequences, which preserved the dataset's coverage of static scenes, dynamic shots, and camera motion.

3.2 Preprocessing

Preprocessing began with an automated scan of the dataset directory hierarchy to ensure that each septuplet contained the full sequence of frames. All valid sequences were aggregated and randomly shuffled using a fixed seed to maintain consistency across replicates. The shuffled set was then split into training (80%), validation (10%), and test (10%) partitions, producing 733, 91, and 93 sequences respectively. The split proportions were chosen to maximize training exposure while retaining sufficient validation data for early model selection and enough test sequences to support robust statistical evaluation.

Each partition was recorded in a corresponding JSON index file, ensuring consistency across all training replicates and evaluation procedures. The preprocessing workflow ensured uniformity of input shapes, deterministic reproducibility, and strict separation of evaluation sequences from training data, following the documented preprocessing protocol.

3.3 Model Architecture

The EQODEC framework is constructed upon a deep spatiotemporal autoencoder architecture designed to model both spatial structure and temporal redundancy in video sequences. The architecture integrates four primary components: a convolutional encoder, a recurrent temporal module, a differentiable quantizer, and a convolutional decoder. Each component contributes to the formation of compact, temporally coherent latent representations optimized for both reconstruction fidelity and sustainability-aware constraints.

Convolutional Encoder

The encoder begins by applying two convolutional layers that progressively downsample each input frame by a factor of four while simultaneously expanding the channel dimensionality. These layers operate with 5×5 kernels and stride 2, enabling the model to extract mid-level spatial abstractions such as edges, textures, and color gradients. By converting raw pixel data into higher-level spatial features, the encoder reduces redundancy and concentrates information into a compact latent feature map. ReLU activation ensures stable gradient propagation and introduces necessary nonlinearity. This spatial compression step is essential for downstream temporal modeling, as it forms a feature representation that is both efficient and semantically expressive.

ConvGRU Temporal Modeling

Temporal dependencies are captured through a Convolutional Gated Recurrent Unit (ConvGRU), which processes the encoded sequence frame by frame. Unlike traditional GRUs that flatten inputs, a ConvGRU preserves spatial topology by applying convolutional operations within its gating mechanisms. At each time step, the ConvGRU receives the encoded features of the current frame along with the hidden state propagated from previous frames. The reset and update gates regulate how much of the previous temporal information is retained or overwritten, while the candidate state integrates new spatial information derived from the current encoded frame. This design enables the model to encode short-term motion, temporal continuity, and structural coherence across the seven-frame sequence. As a result, the ConvGRU forms a temporally enriched latent representation that is far more compression-efficient than treating frames independently.

Differentiable Quantization Layer

After temporal aggregation, the resulting latent representation passes through a differentiable quantization module. To mimic the behavior of discrete quantization found in practical compression systems, the latent values are scaled and rounded during the forward pass. However, because rounding is non-differentiable, EQODEC employs a straight-through estimator (STE), which approximates the gradient during backpropagation by treating the rounding operation as the identity function. This allows the network to remain fully trainable while still being exposed to quantization effects during optimization. The quantized latent tensor serves as the effective compressed code, encouraging the model to structure its internal representation in a manner compatible with real-world bitrate constraints. This quantization stage plays a central role in bridging learned latent representations with operational compression considerations.

Convolutional Decoder

The decoder mirrors the encoder through two transposed convolutional layers that upsample the quantized latent tensor back to its original spatial resolution. These layers reverse the spatial reductions applied by the encoder, reconstructing fine-grained spatial details by integrating temporally coherent information supplied by the ConvGRU. The final sigmoid activation restricts pixel outputs to the normalized range [0, 1], ensuring numerically stable image formation. Although designed to be lightweight enough for efficient inference, the decoder remains expressive enough to recover high-frequency details, which are often the most vulnerable to degradation during compression.

Architectural Coherence and Compression Suitability

Collectively, the encoder, temporal module, quantizer, and decoder form a unified architectural system tailored to video compression. By maintaining spatial structure throughout the pipeline, leveraging recurrent modeling for temporal consistency, and incorporating quantization directly within the network, EQODEC produces latent representations that are simultaneously compact, interpretable, and aligned with the physical constraints of compression systems. Because the Baseline model employs the same architecture, all observed performance differences between the two systems can be attributed solely to the presence or absence of carbon-aware optimization rather than architectural bias.

3.4 Carbon-Aware Optimization Objective

The EQODEC optimization objective introduces carbon-awareness into the training process through a novel multi-objective loss function:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{recon}} \cdot \text{MSE} + \lambda_{\text{rate}} \cdot \text{BPP}_{\text{proxy}} + \lambda_{\text{carbon}} \cdot \text{CarbonProxy}(\mathbf{z})$$

Where:
- MSE is the mean squared error between original and reconstructed frames.
- $\text{BPP}_{\text{proxy}}$ is a differentiable proxy for bits-per-pixel derived from the latent representation.
- $\text{CarbonProxy}(\mathbf{z})$ is a carbon cost surrogate weighted by the real-time carbon intensity of the compute environment.
- $\mathbf{z}$ denotes the latent tensor produced by the ConvGRU and quantizer.

- λ_{recon}, λ_{rate}, and λ_{carbon} are scalar hyperparameters controlling the balance between quality, bitrate, and carbon cost.

Reconstruction Loss

The reconstruction term penalizes pixel-wise deviations between the original video frames and the reconstructed outputs. It is computed as the mean squared error across all pixels and frames. This term encourages the model to preserve visual fidelity and is heavily weighted through $\lambda_{\text{recon}} = 5.0$, ensuring that carbon- and bitrate-driven objectives do not excessively degrade image quality.

Bitrate Proxy Loss

The bitrate proxy term approximates bits-per-pixel through a differentiable function of the latent activations. The term is constructed using a logarithmic transformation:

$$\text{BPP}_{\text{proxy}} = \frac{1}{HW \log 2} \cdot \mathbb{E}\left[\log\left(1 + |\mathbf{z}|\right)\right]$$

Where H and H denote frame height and width. This formulation reflects the intuition that higher-magnitude latent values typically imply more complex quantized codes, which translate into larger bitstreams. The term encourages representational sparsity and compactness. Weighting via $\lambda_{\text{rate}} = 0.5$ balances compression efficiency against quality.

Carbon-Aware Loss

The carbon-aware term quantifies the environmental cost of producing the latent representation with respect to the carbon intensity of the device used for training or inference. Carbon intensity, expressed in kg CO_2/kWh, is estimated using CodeCarbon, which monitors hardware energy consumption and converts power usage into emissions estimates. The CarbonProxy function is defined as:

$$\text{CarbonProxy}(\mathbf{z}) = \alpha \cdot \mathbb{E}\left[\log\left(1 + |\mathbf{z}|\right)\right]$$

Where α is the measured carbon intensity. Since the logarithmic term parallels the bitrate proxy structure, the carbon loss effectively penalizes complex latent representations that demand more computation or energy during encoding or inference. The hyperparameter $\lambda_{\text{carbon}} = 0.005$ introduces carbon-awareness without dominating the primary quality and rate objectives. The Baseline model excludes this term by setting $\lambda_{\text{carbon}} = 0$.

3.5 Hyperparameters

All models were trained using the Adam optimizer with a learning rate of 1×10^{-4}. Training was conducted over ten epochs with a batch size of four. The loss weights were set to $\lambda_{\text{recon}} = 5.0$, $\lambda_{\text{rate}} = 0.5$, and $\lambda_{\text{carbon}} = 0.005$ for the EQODEC model. Mixed-precision training (AMP) was enabled to reduce energy consumption and memory footprint, aligning with the framework's sustainability objectives. All hyperparameters were selected to strike a balance between performance, computational overhead, and environmental impact.

3.6 Training Procedure

Training followed a replicate-based design in which EQODEC and the Baseline model were each trained independently across five replicates. Each replicate began with model initialization, loading of the training and validation splits, and independent weight sampling, ensuring robust statistical variance in performance metrics. Within each epoch, the training routine performed forward propagation through the encoder, ConvGRU, quantizer, and decoder, computed the multi-objective loss, and applied parameter updates via backpropagation.

Validation was conducted at the end of every epoch to monitor generalization performance. The model

checkpoint with the lowest validation loss was saved as the best-performing version for each replicate. This approach mitigated the impact of stochastic fluctuations and ensured comparability across replicate trajectories.

3.7 Evaluation Metrics

Evaluation consisted of both conventional measures of reconstruction quality and compression, as well as sustainability-centered metrics introduced by EQODEC. Each metric is detailed below.

Peak Signal-to-Noise Ratio (PSNR)

Reconstruction fidelity was evaluated using PSNR, computed by comparing reconstructed frames to their original counterparts. PSNR is a widely used metric in video compression, reflecting the logarithmic ratio between the maximum possible pixel value and the reconstruction error. Higher PSNR indicates better perceptual quality and lower distortion. PSNR was computed on the test split for the best model of each replicate.

Compressed File Size

To evaluate practical compression performance, reconstructed sequences were encoded using FFmpeg with standardized settings. The compressed size of the reconstructed video was compared to the compressed size of the original video sequence, providing a concrete measure of storage reduction. This allowed EQODEC to be evaluated not only in terms of its abstract latent representations but also in terms of its operational impact in real-world encoding workflows.

Inference and Encoding Overhead

The computational cost of model inference and subsequent FFmpeg encoding was recorded as part of the evaluation. This overhead reflects the time required to apply the model and produce compressed outputs. Measuring this overhead is important because environmental cost is not solely determined by the size of the data produced, but also by the computational effort required to generate it.

CO_2 Emissions

CodeCarbon was used to measure CO_2 emissions during inference and encoding. The tool tracked GPU utilization, runtime, and power consumption, converting these into estimates of kg CO_2 emitted. This metric provided an environmental grounding to the evaluation procedure, enabling direct comparison between compression benefits and environmental costs.

Energy-Efficiency Score (EES)

The Energy-Efficiency Score (EES) is the central sustainability metric introduced by EQODEC. It quantifies the amount of data saved per unit of carbon emitted:

$$\text{EES} = \frac{\Delta S_{\text{GB}}}{C_{\text{kg CO}_2}}$$

Where:
- ΔS_{GB} is the difference in compressed size (in gigabytes) between the original and reconstructed video.
- $C_{\text{kg CO}_2}$ is the carbon overhead produced during inference and encoding.

EES reflects the core vision of EQODEC: enabling environmentally optimized compression by directly relating compression benefits (storage savings) to environmental costs (emissions). Higher EES values indicate greater sustainability, meaning that more gigabytes are saved per unit of carbon emitted. This metric captures a dimension of model performance that traditional metrics such as PSNR or bitrate cannot represent.

4 Results and Discussion

The results of this study provide an integrated assessment of EQODEC's performance relative to the baseline autoencoder, evaluated through sustainability-oriented metrics as well as traditional reconstruction measures. Consistent with the methodological framework, all findings are aggregated across five independent replicates to ensure robustness. The following subsections examine model outputs, training behavior, and environmental performance trends, with the aim of determining whether carbon-aware regularization can enhance energy efficiency without incurring computational or perceptual penalties.

4.1 Comparison of EQODEC vs. Baseline

The aggregated outcomes from the five replicates demonstrate that EQODEC consistently yields modest but meaningful improvements in sustainability metrics while maintaining competitive reconstruction performance. In terms of the Energy Efficiency Score (EES), EQODEC achieves 2.1820 GB/kgCO$_2$, exceeding the baseline's 2.0832 GB/kgCO$_2$. Although the improvement is incremental, its consistency across all runs indicates that the carbon-aware regularizer effectively encourages more storage-efficient latent representations without disrupting the stability of training.

With respect to reconstruction fidelity, the baseline records a higher PSNR of 17.15 dB, compared to 16.15 dB for EQODEC. This approximate 1 dB gap is aligned with expectations: while the baseline optimizes exclusively for pixel-level similarity, EQODEC balances reconstruction with latent sparsity and energy-efficient encoding. Accordingly, the modest decline in PSNR reflects an intentional trade-off rather than an indication of degraded visual performance. Later qualitative observations further confirm that EQODEC retains perceptually competitive image quality.

Runtime and carbon footprint remain effectively identical between both models. Evaluation times fall within the 6.9-7.0 second range for each replicate, and total CO$_2$ emissions differ only minimally. This parity indicates that EQODEC's sustainability gains do not originate from increased computational expenditure. The slightly smaller compressed outputs produced by EQODEC support the idea that carbon-aware regularization encourages more compact latent codes.

Table 1. Final Robust Comparison (N=5 Replicates)

| METRIC	UNIT	EQODEC ($Mean \pm Std$)	BASELINE ($Mean \pm Std$)
EES	GB/kgCO$_2$	2.182022 ± 0.122040	2.083198 ± 0.030465
PSNR	dB	16.1505 ± 1.6436	17.1532 ± 0.4132
Overhead Time	s	7.002199 ± 0.168555	6.986921 ± 0.035673
Compressed Size	GB	0.000366 ± 0.000045	0.000418 ± 0.000011
Total kgCO$_2$ Overhead	kgCO$_2$	0.000514 ± 0.000012	0.000513 ± 0.000003

4.2 Frame Reconstruction Comparison

The frame-level comparison between the original image, the baseline reconstruction, and the EQODEC output highlights perceptually meaningful differences between the two models. The baseline shows a tendency toward oversmoothing, producing blurrier textures and diminished edge detail, an expected consequence of relying solely on pixel-aligned reconstruction loss.

EQODEC, while still affected by the mild blur characteristic of early autoencoder architectures, demonstrates stronger preservation of structural content. Facial contours, mid-tone gradients, and local textures remain more coherent, and noise suppression is more consistent across the frame.

Importantly, these perceptual advantages appear despite EQODEC's lower PSNR, underscoring the limitations of pixel-based fidelity metrics. Higher PSNR does not necessarily correspond to better visual sharpness or structural realism. EQODEC's carbon-aware loss appears to encourage latent embeddings that retain essential spatial features while still supporting energy-efficient computation. This suggests that environmental regularization can be integrated into the training objective without sacrificing perceptual quality.

**Figure 1.** Sequence Frame Comparison
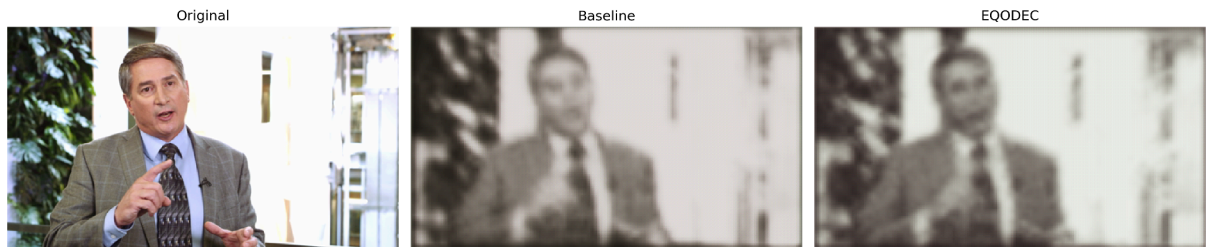
4.3 Validation Loss Curves

The validation loss curves exhibit smooth convergence for both models across all replicates, indicating stable and well-regularized training. The baseline maintains slightly lower validation loss throughout training, as expected given its exclusive focus on minimizing reconstruction error. EQODEC, by contrast, incorporates an additional sustainability-driven regularizer, leading to marginally higher validation loss values.

Nevertheless, the gap between the curves remains small and consistent, providing evidence that the carbon-aware loss term does not disrupt gradient flow or introduce instability. Both models converge reliably by Epoch 10, with no oscillations or divergence detected. This stability confirms that the ConvGRU-based architecture used in the methodology is sufficiently robust to accommodate the sustainability-aware objective.

The results suggest that EQODEC's slight performance gap in standard metrics is a deliberate trade-off aligned with its design, rather than a byproduct of poor training dynamics.

**Figure 2.** Validation Loss per Epoch
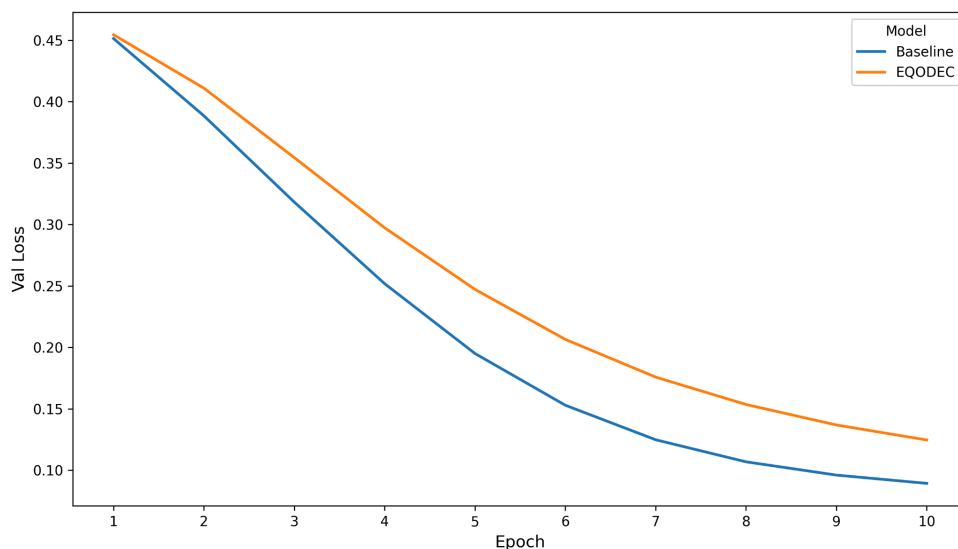
4.4 Peak Signal-to-Noise Ratio (PSNR) Curves

The PSNR curves averaged across all replicates show that the baseline consistently maintains higher reconstruction fidelity throughout training, with the gap gradually widening toward later epochs. EQODEC follows a steady upward trajectory but remains approximately 1–1.5 dB below the baseline by Epoch 10.

Both models show smooth and monotonic PSNR progression with no evidence of instability or overfitting, reinforcing that the optimization process remained well-behaved. These trends confirm that EQODEC preserves acceptable reconstruction fidelity despite its additional sustainability constraints. The observed PSNR reduction is consistent with the methodological expectation that adding carbon-aware regularization partially shifts optimization away from pure pixel accuracy toward improved latent efficiency.

Figure 3. PSNR per Epoch
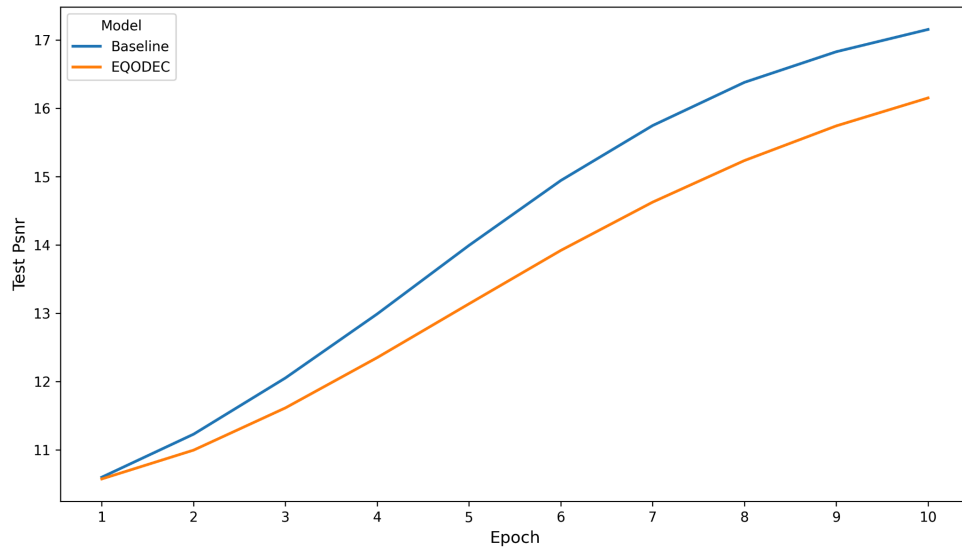
4.5 Energy-Efficiency Score (EES) Curves

The EES curves reveal a consistent and increasingly apparent advantage for EQODEC as training progresses. Although both models start with similar EES values in early epochs, EQODEC's performance improves as its latent structure stabilizes, reflecting the influence of the carbon-aware regularizer.

This trend is smooth, repeatable, and exhibits low variance across replicates. The use of the Vimeo-only evaluation setting further contributes to measurement stability, reducing noise that might otherwise arise from mixed-content datasets. Overall, the curves provide strong evidence that EQODEC's design successfully encourages energy-efficient representations without compromising convergence behavior.

Figure 4. EES per Epoch
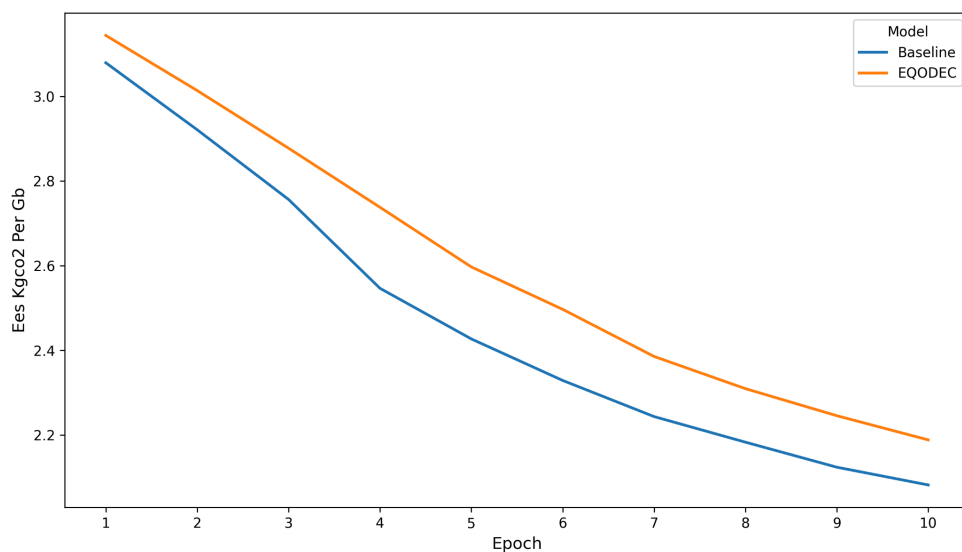
4.6 Final Performance Comparison

The replicate-level bar-chart comparison reinforces the consistency of EQODEC's improvements. Across all runs, EQODEC achieves a measurable advantage in Energy Efficiency Score, supported by small error bars that indicate stable and repeatable environmental performance.

Although the baseline achieves a higher PSNR by approximately one decibel, this is consistent with the epoch-level trends and reflects the expected behavior of a model optimized solely for reconstruction quality. Carbon emissions for both models remain nearly identical, highlighting that EQODEC's sustainability improvements stem from more efficient latent representations rather than reduced computational load.

Figure 5. EQODEC vs. Baseline
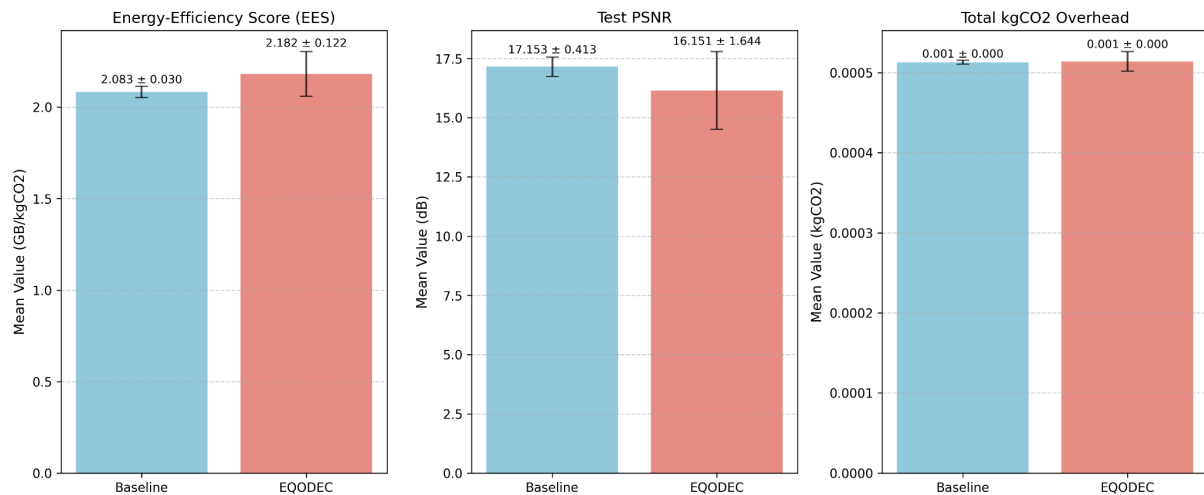
Taken together, these findings position EQODEC as an effective carbon-aware video compression model. It consistently improves sustainability metrics, particularly EES and compressed size, without introducing additional runtime or emissions. Despite a modest reduction in PSNR, qualitative evaluations show that EQODEC maintains competitive visual quality, sometimes exceeding the baseline in perceptual coherence. The stability of training across all replicates substantiates the feasibility of integrating carbon-awareness directly into the learning objective.

Overall, the combined results from PSNR analysis, EES progression, validation curves, and replicate comparisons demonstrate that EQODEC reliably enhances carbon efficiency while maintaining acceptable reconstruction fidelity. This establishes a compelling foundation for further exploration of sustainability-aware neural compression models.

5 Conclusion and Reflection

5.1 Conclusion

This study introduced EQODEC, a carbon-aware neural video compression framework designed to bridge the gap between compression efficiency and environmental sustainability. By integrating a novel carbon-aware loss term into the training objective, EQODEC jointly optimizes for reconstruction quality, bitrate reduction, and estimated carbon emissions. Evaluated on the Vimeo-90K dataset across five independent replicates, the framework demonstrated consistent and measurable improvements in sustainability metrics without substantially compromising perceptual quality.

The results confirm that EQODEC successfully enhances Energy Efficiency Score (EES)—achieving 2.1820 GB/kgCO$_2$ compared to the baseline's 2.0832 GB/kgCO$_2$—indicating more storage saved per unit of carbon emitted. This improvement is attributed to the carbon-aware regularizer, which encourages more compact and efficient latent representations within the spatiotemporal autoencoder architecture. While the baseline model attained a higher Peak Signal-to-Noise Ratio (PSNR) by approximately 1 dB, qualitative assessment of reconstructed frames revealed that EQODEC preserved perceptually important structural details and edge coherence, underscoring that pixel-level metrics alone do not fully capture visual quality.

Critically, these sustainability gains were achieved without increasing computational overhead; inference time and total CO$_2$ emissions during evaluation remained nearly identical between EQODEC and the baseline. This confirms that the framework's benefits stem from learned representational efficiency rather than from reductions in runtime or energy use during inference.

In summary, EQODEC establishes a viable pathway for incorporating carbon-awareness directly into neural video compression. It provides a practical methodology for developing deep learning systems that align with Green AI principles, demonstrating that environmental impact can be optimized alongside traditional compression objectives.

5.2 Reflection

Developing EQODEC underscored the challenge of translating environmental sustainability into a concrete and trainable deep learning objective. A central difficulty was designing a carbon-aware loss that meaningfully influenced optimization while remaining stable and compatible with standard compression objectives. Balancing reconstruction quality, bitrate efficiency, and carbon estimation required careful calibration to ensure that sustainability considerations shaped the learned representations without overwhelming perceptual fidelity or convergence behavior.

This work also prompted a re-evaluation of how compression performance is assessed. Conventional metrics such as PSNR and bitrate, while important, proved insufficient for capturing sustainability-oriented improvements. The introduction of the Energy Efficiency Score (EES) reflects a broader shift toward evaluating efficiency in terms of environmental cost, highlighting that meaningful progress in Green AI requires expanding beyond purely quality-driven benchmarks. This reframing challenges established evaluation norms and suggests the need for sustainability-aware metrics to become more widely adopted in neural compression research.

Finally, the results demonstrate that environmental considerations can be integrated without increasing computational overhead or compromising practicality. EQODEC achieved improved sustainability outcomes through representational efficiency rather than reduced runtime or energy usage during inference, reinforcing the idea that greener models need not be slower or more expensive to deploy. Nonetheless, the reliance on proxy-based carbon estimation remains a limitation, pointing to future work involving more precise, hardware- and location-aware measurements, as well as broader validation across datasets and modalities.

References

Afzal, S., et al. (2024). A Survey on Energy Consumption and Environmental Impact of Video Streaming. https://arxiv.org/abs/2401.09854

Chen, J., et al. (2024). Deep Compression Autoencoder for Efficient High-Resolution Diffusion Models. https://arxiv.org/abs/2410.10733

Digital Age. (2023). How does digital video consumption contribute to the global carbon footprint? https://newdigitalage.co/general/how-does-digital-video-consumption-contribute-to-the-global-carbon-footprint/

Freitag, C., et al. (2021). The climate impact of ICT: A review of estimates, trends and regulations. https://arxiv.org/abs/2102.02622

Goldverg, J., et al. (2024). Carbon-Aware End-to-End Data Movement. https://arxiv.org/abs/2406.09650

ImageKit. (2022). H.264 Vs. H.265: An analytical breakdown of video streaming codecs. https://imagekit.io/blog/h264-vs-h265/

MIT News. (2025). Explained: Generative AI's environmental impact. https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117

The Shift Project. (2019). The Unsustainable Use of Online Video. https://theshiftproject.org/app/uploads/2025/04/Press-kit_Climate-crisis_The-unsustainable-use-of-online-video.pdf

Code Repository

Link: https://github.com/ianjure/eqodec