Transformer-Based Neural Video Compression on Solar Imagery

Atefeh Khoshkhahtinat, Ali Zafari, Piyush M. Mehta, Nasser M. Nasrabadi, Barbara J. Thompson, Daniel da Silva, Michael S. F. Kirk

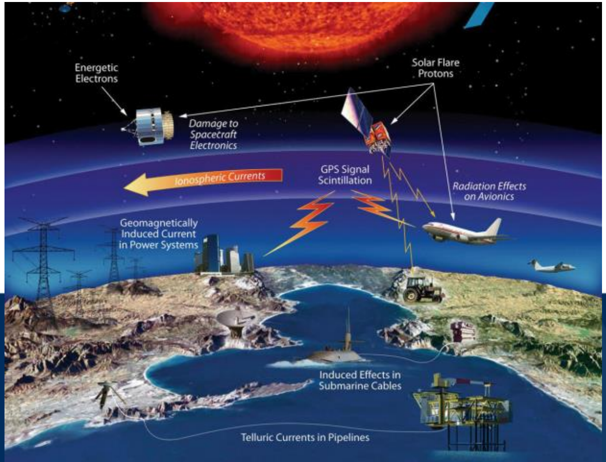
American Meteorological Society (AMS 2024)

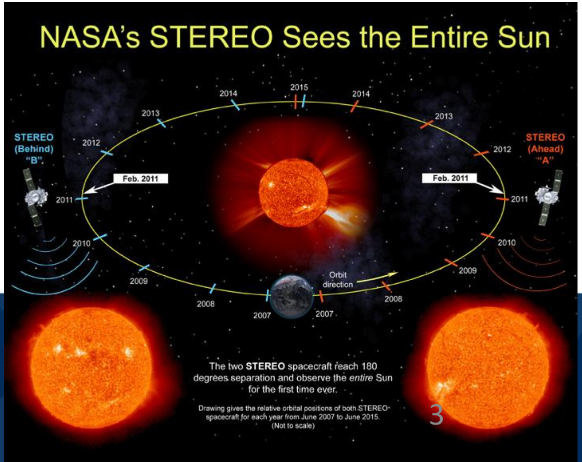
Overview

- Motivation
- Solar Dynamics Observatory (SDO)
- Neural Network-based Image Compression Approaches
- Neural Network-based Video Compression Approaches
- Our Contribution: Adaptation of Transformer-based Video Compression Scheme to SDO Image Dataset
- Results and Comparison
- Conclusions and future work

Motivation

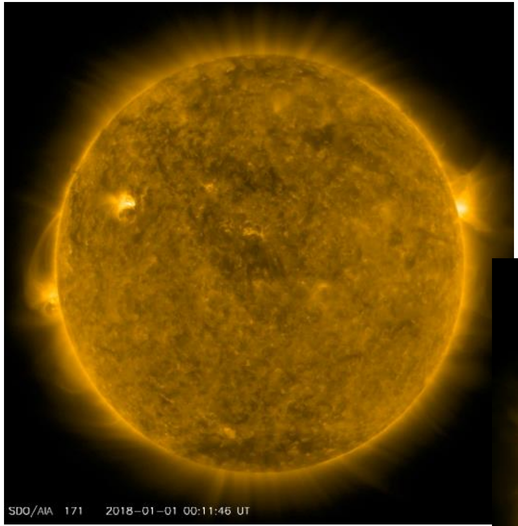
- The sun is essential for life in the solar system but can also wreak havoc on life and technology infrastructure in space and on ground (impacts of space weather)
- There is growing impetus on studying the sun for advanced scientific knowledge towards improved predictive capability
  - Multiple active and planned (deep-space) missions imaging the sun from different locations
    - SDO
    - SOHO
    - STEREO A and B
    - Solar Orbiter
    - Parker Solar Probe
    - Vigil (planned by ESA)
  - These missions (will) require effective and efficient data compression schemes to increase science and information return for operations

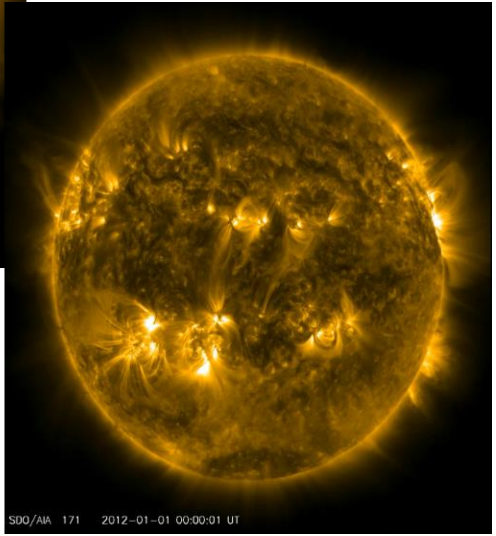




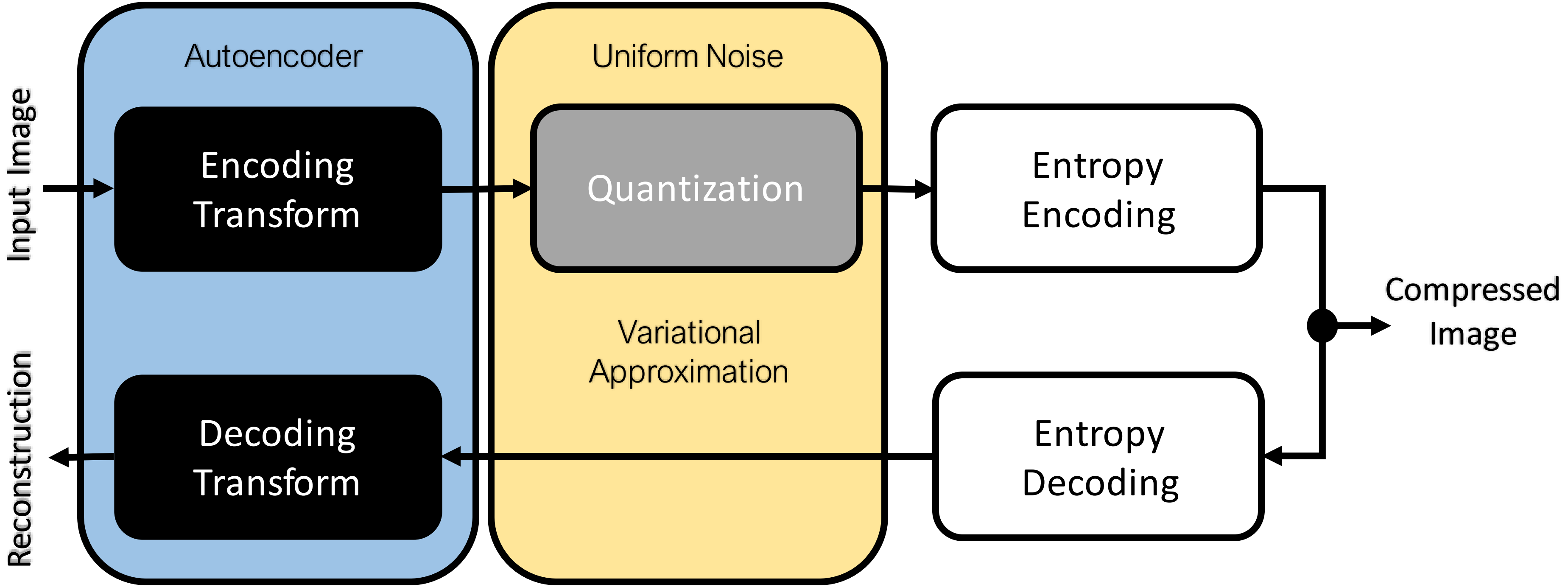
Solar Dynamics Observatory

- SDO is one-of-a-kind mission dedicated to continuously observe the sun from GEO
  - Launched in 2010, it has a dedicated ground station for data transmission
  - It captures a stack of 4K solar images across 9 different spectral wavelengths every 12 seconds
  - This results in 1.4TB of data each day, totaling multiple PBs thus far.
- The data collected has provided a wealth of scientific information about our star
- For this demonstration purpose, we use the curated ML-ready dataset (arXiv:1903.04538)



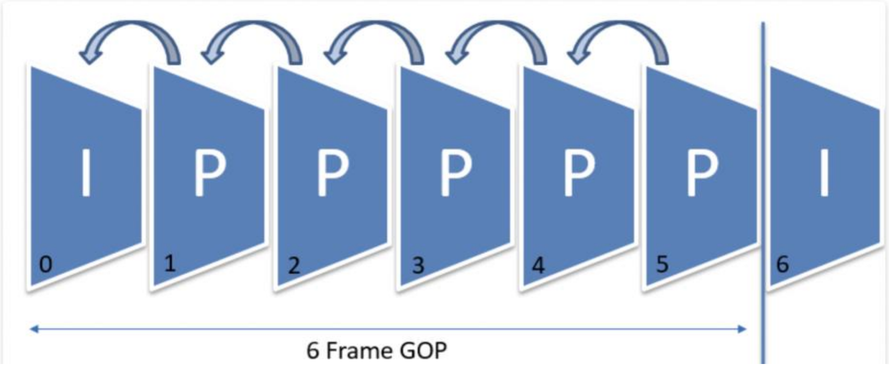


Neural-based Image Compression

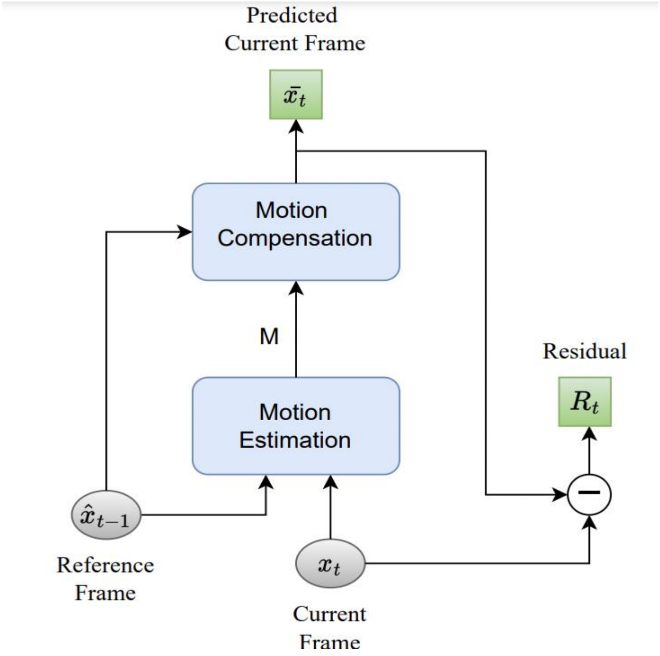


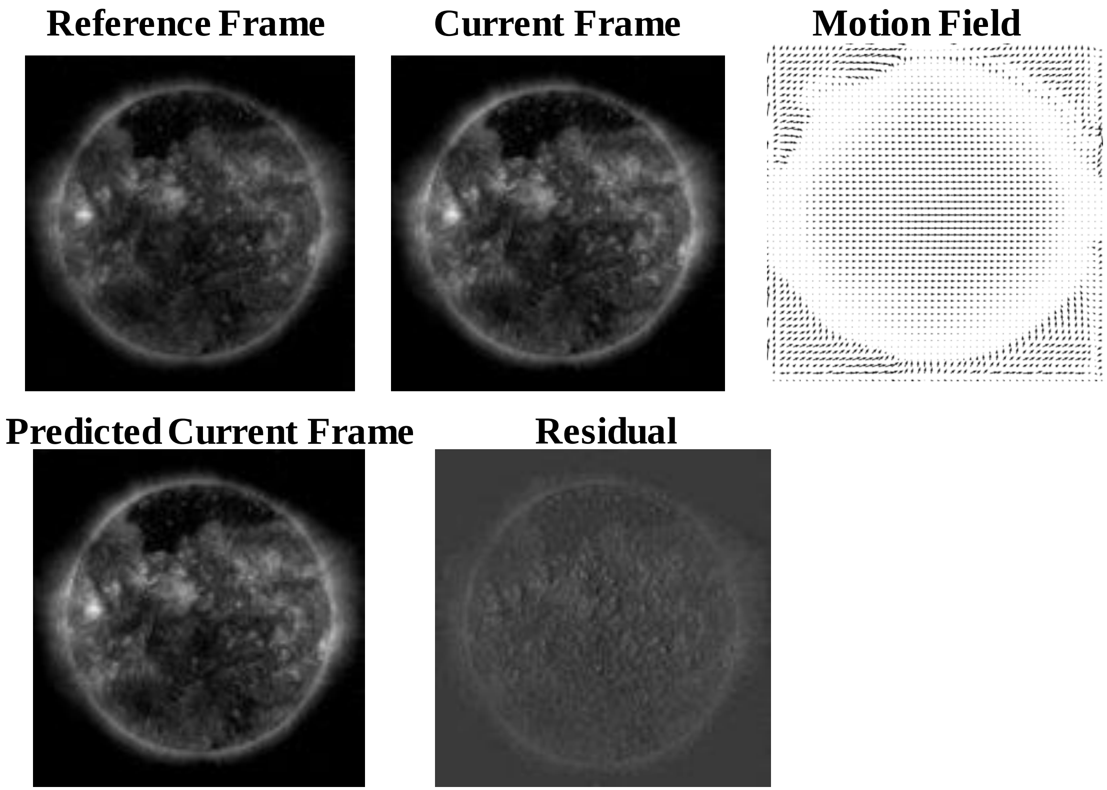
Video Compression

- Video Compression: Remove **temporal redundancy** + **spatial redundancy**
- There are two types of coding: Intraframe coding, Interframe coding
- **Intraframe coding** : Image compression **Interframe coding**: Using reference frame + motion information

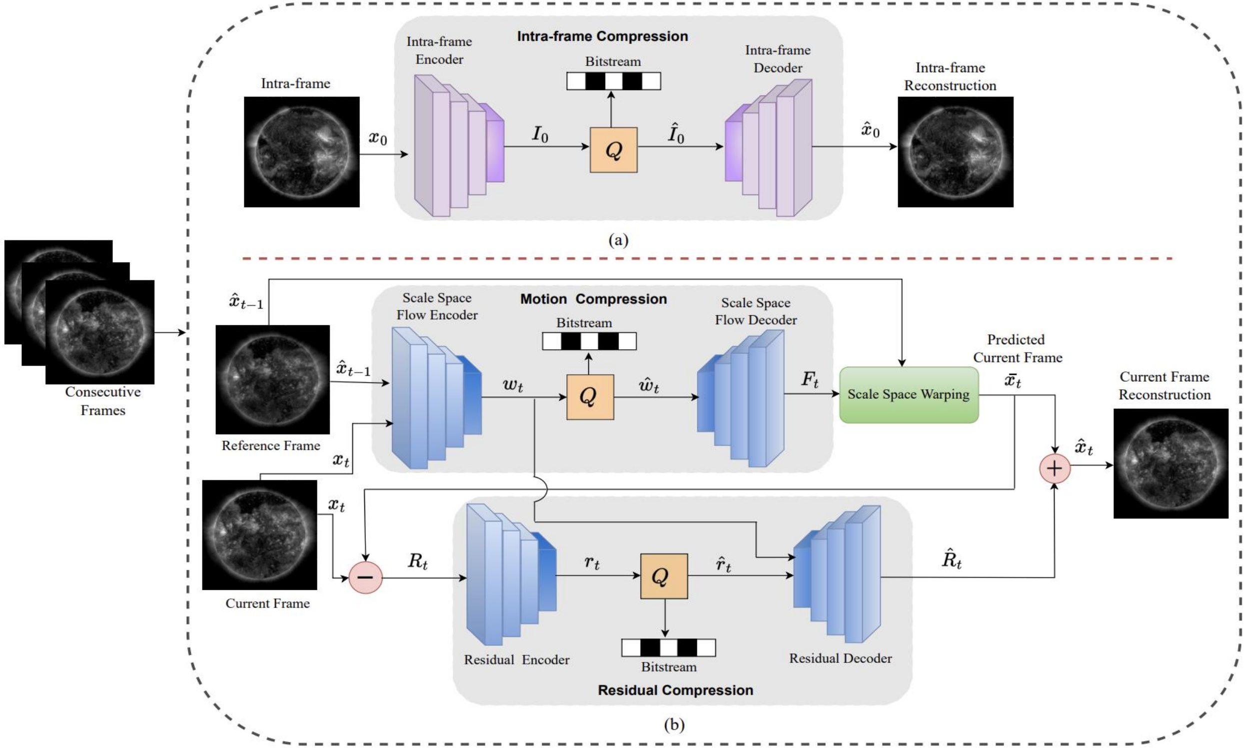


Inter-frame Pipeline





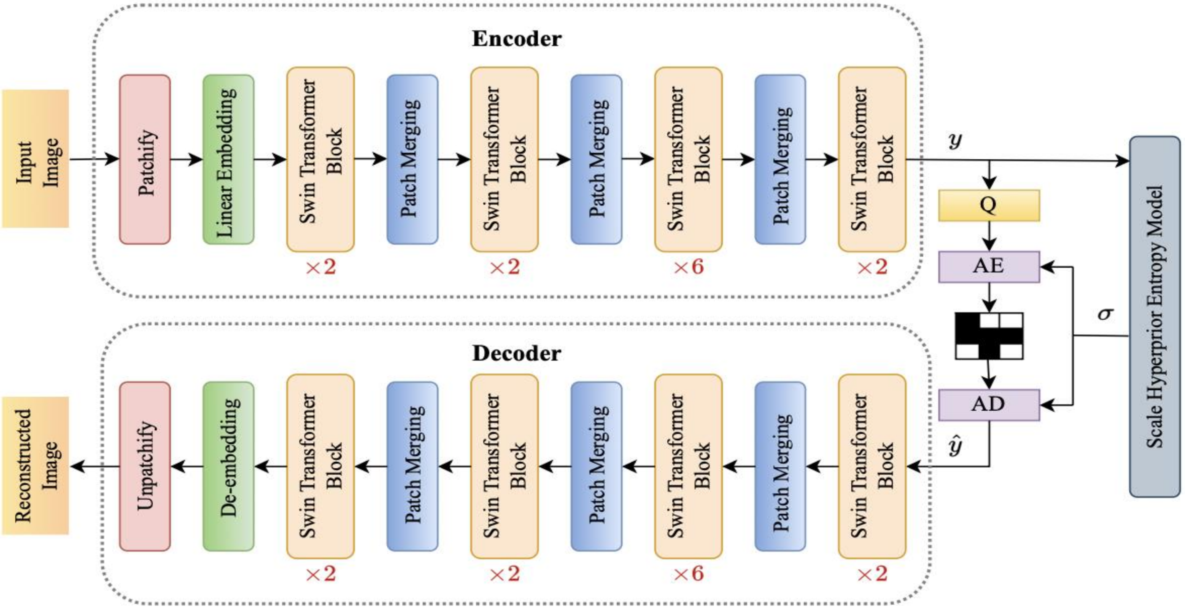
SSF: Neural-based Video Compression



Total rate–distortion loss over Sequence with N frames:

$$D+\lambda R=\sum_{t=0}^{T-1} d(x_t,\hat{x}_t)+\lambda\Big[R(I_0)+\sum_{t=1}^{T-1} R(w_t)+R(r_t)\Big]$$

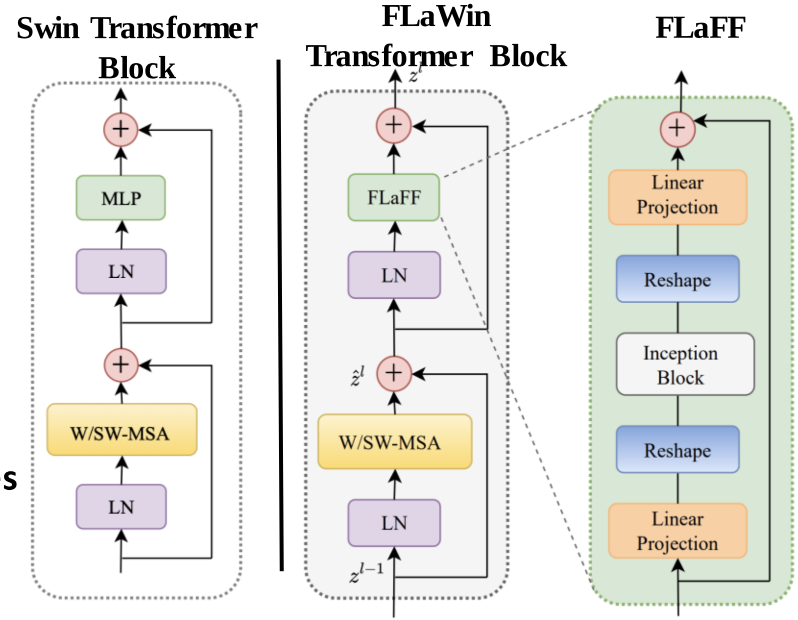
Swin Transformer-based Compression



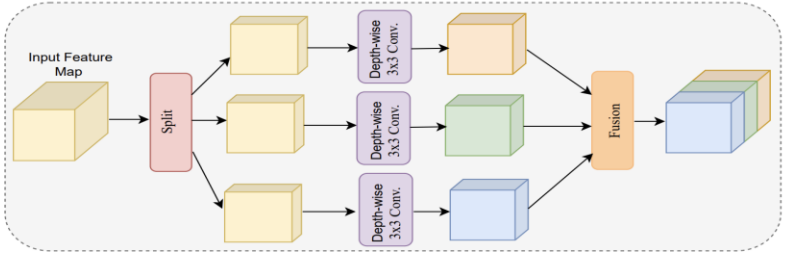
(a)

Transformer Blocks

- Transformer block is the main part of the Transformer architecture.
- The Transformer block consists of multiple layers of self-attention, feed-forward neural network, layer normalization and residual connection.
- Self-attention helps model to capture long-range dependencies.
- FLaWin is constructed by replacing the feed-forward network in the Swin Transformer block with the fused local-aware feed-forward network(FLaFF) :Extracting **local dependencies + global dependencies**



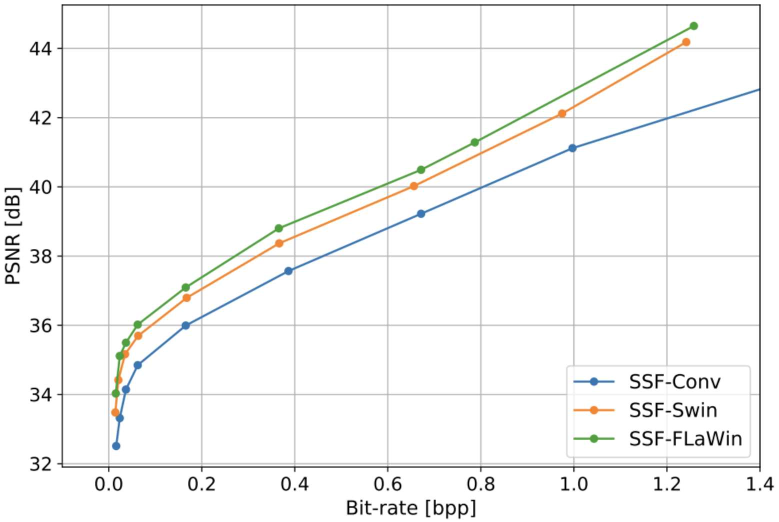
Inception Block

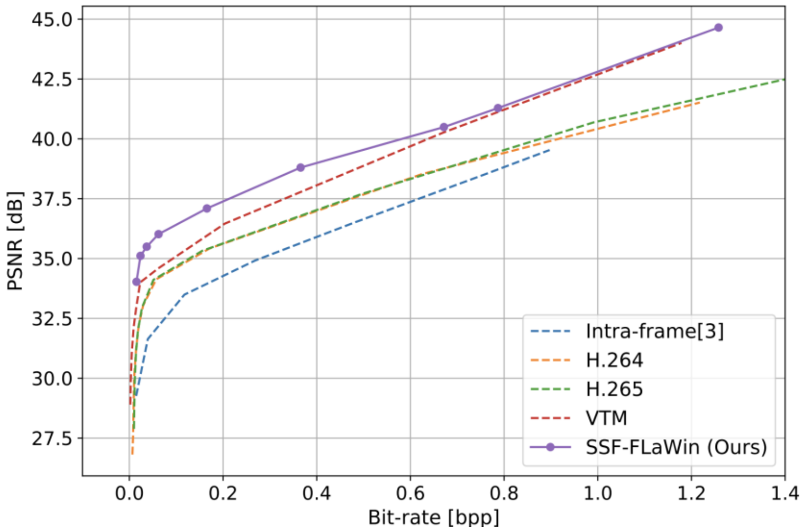


Inception block: dividing the input along the channel dimension and directing these split components into three separate branches.

Results and Comparison

- Comparison Metric for Distortion:
  PSNR: Signal-to-noise ratio





Conclusion and Future Work

- Applying the video compression approach on the dataset boost the compression ratio because of the high temporal correlation between the images.
- Utilizing Transformers-based autoencoder, we improve the network capability to capture correlations more robustly, leading to a more decorrelated latent code.
- We demonstrate that our Transformer-based compression scheme outperform traditional codes for the SDO data.
- Next step involves adaptation of the proposed compression scheme to Level 0 (no post-processing) data.
- The more accurate entropy model is needed to estimate the distribution of the latent representations.
- Other types of Transformer networks can be designed to improve the compression efficiency.

THANK YOU!

Contact:
ak00043@mix.wvu.edu

This work is supported by NASA award 80NSSC21M032