Overview:

Our project involves an extension of the original Mamba model to handle a 2D visual sequence, inspired by work done in the paper entitled "VMamba: Visual State Space Model". Since Mamba originally dedicated its modeling to 1-dimensional time series data, it needed to adapt its modeling capability to be extended in 2-space dimensions without sacrificing efficiency in computation or flexibility. It is to be developed as a framework that copes with 2-dimensional visual data, such as images, and fully does so to a greater extent.

Contributions:

Implementation of VMamba for 2D Data:

New model pipeline based on the framework VMamba described in the associated paper.

The adapted and re-engineered components are the convolutional layers, blocks for 2D feature extraction, and task-specific heads.

Replaced the sequence processing steps with an architecture optimized for 2D spatial inputs.

Removal of S6, and simplification: First implemented the incorporation of the S6 block for enhanced representation learning; after limited performance benefits during evaluation, switched to the simplified model variant without S6.

The current implementation focuses on computational efficiency while maintaining competitive accuracy.

Custom Visualization:

Added custom visualization modules plotting model progress during epochs, loss curves, and sample outputs.

Model Training and Evaluation:

Optimized the training process using the CIFAR-10 dataset to make the pipeline best structured and most efficient.

The results shown here illustrate the potential of the model in applying the principles of Mamba for 2D input formats.

Outcomes

Achieved stable training behavior with loss curves and accuracy reflecting effective learning. Demonstrated the feasibility of adapting Mamba from 1D to 2D contexts while making trade-offs to balance performance and simplicity.