Improvements to the Loss function of TSA-Net

1 Introduction

The loss functions were changed to be more sensitive to the nature of spectral image reconstruction. It was found that the model with content-aware loss alone converged faster to its optimum, with a low PSNR. Combined with the fact that the original model's loss became quite small near the optimum, and might be leading to slower optimization, the intuition was that the combination of these losses might lead to faster training as well as more accuracy.

Multi-scale Structure loss was used to further increase the robustness of the model.

2 Loss Functions

The loss functions used in the TSA-Net model are defined as follows:

2.1 Original Model Loss

The loss function for the original model is given by:

$$\text{Loss}_{\text{original}} = \text{loss}_{\text{RMSE}}(\hat{y}, y) + \text{loss}_{\text{spec}}(\hat{y}, y)$$

where: - $\hat{y}$: predicted output, - y: ground truth output, - $\text{loss}_{\text{RMSE}}$: Root Mean Square Error loss, - $\text{loss}_{\text{spec}}$: Spectral loss.

2.2 Improved Model Loss (Mix Loss)

The improved model loss function includes additional terms for content-aware spectral loss and multi-scale structure loss:

$$\text{Loss}_{\text{improved}} = \frac{\text{loss}_{\text{CAS}}(\hat{y}, y)}{100} + \text{loss}_{\text{multi-scale}}(\hat{y}, y) + \text{loss}_{\text{RMSE}}(\hat{y}, y) + \text{loss}_{\text{spec}}(\hat{y}, y)$$

where: - loss_{CAS}: Content-Aware Spectral loss, - $\text{loss}_{\text{multi-scale}}$: Multi-Scale Structure loss.

2.3 Loss Function Details

1. Root Mean Square Error (RMSE) Loss:

$$\text{loss}_{\text{RMSE}}(\hat{y}, y) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

2. Spectral Loss:

$$\text{loss}_{\text{spec}}(\hat{y}, y) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\nabla y_i - \nabla \hat{y}_i\right)^2}$$

3. Multi-Scale Structure Loss:

$$\text{loss}_{\text{multi-scale}}(\hat{y}, y) = \sum_{i=0}^{4}\left(0.5^i \cdot (1 - \text{SSIM}(\text{downsample}_i(\hat{y}), \text{downsample}_i(y)))\right)$$

4. Content-Aware Spectral Loss:

$$\text{loss}_{\text{CAS}}(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\text{Var}(\text{FFT}(y))}{\epsilon + \text{Var}(\text{FFT}(y))} \cdot \left(\text{FFT}(\hat{y}) - \text{FFT}(y)\right)^2\right)$$

where ϵ is a small constant to prevent division by zero.

3 Results for TSA-Net

At epoch 100 of training with the improved method, following are the results on test data.

Image	**PSNR, SSIM**
Scene 1	28.16, 0.8892
Scene 2	24.00, 0.7854
Scene 3	22.32, 0.6185
Scene 4	33.69, 0.9320
Scene 5	24.41, 0.8572
Scene 6	28.83, 0.9492
Scene 7	20.05, 0.7087
Scene 8	27.75, 0.9368
Scene 9	25.74, 0.8109
Scene 10	25.45, 0.8301
Average	26.04, 0.8318

These results are comparable to those obtained with the original method, with the same data, at epoch 200. This indicates that the new model converges faster.

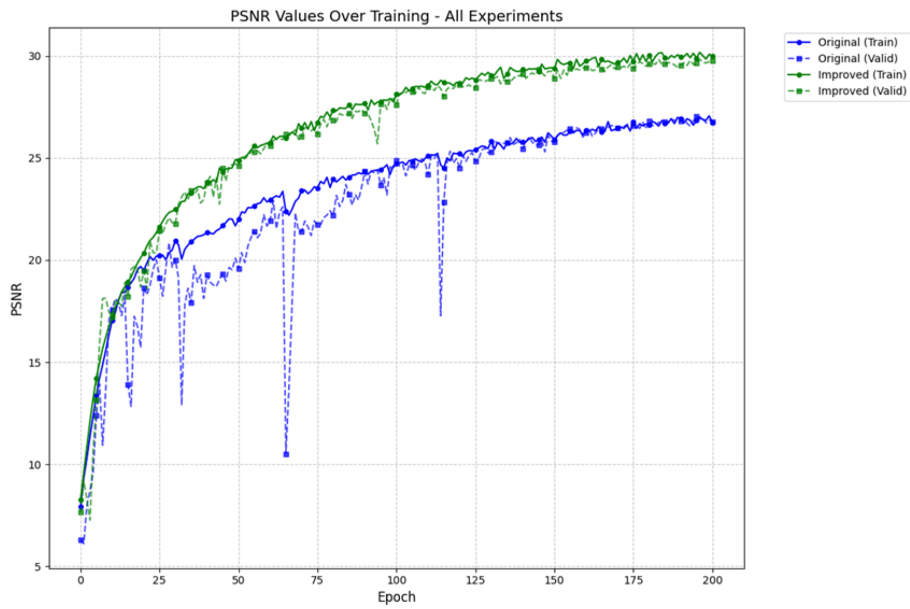

Figure 1: Plot of training and validation loss with epochs, for original and improved models

4 Observations

- Each epoch took the same time as it did while training the previous model.
- Mix loss leads to lesser deviations of validation accuracy from training accuracy during the initial training phase.
- The validation PSNR of the original method after 200 epochs was achieved by the mix loss within just 75 epochs.
- Given 200 epochs, Mix loss achieved a validation PSNR of 30, representing an 11% improvement over the original loss which yielded a PSNR of 26.
- As mentioned earlier, a test PSNR of 26, which took 200 epochs in the original method, was achieved within 100 epochs using the new Mix loss. This indicates a 50% reduction in training time.
- However, test accuracy with Mix loss remained at 26 at epoch 200 as well, despite good validation performance (PSNR of 30). This could indicate a higher dependence on other characteristics of the data beyond those captured by the current training setup.
- Further improvements in test PSNR may be possible with additional and varied data for training.