Machine-Learned Information-Theoretic Optimal Data Compression for Outer Solar System Missions

Principal Investigator: Matthew Thill (332); Co-Investigators: Robert West (322), Hua Xie (332), Jeffrey Jewell (398), Dariush Divsalar (332), Philip Dumont (383), Matthew Tiscareno (SETI Institute), Paul Helfenstein (Cornell University)

Program: FY21 R&TD Topics **Strategic Focus Area: Data: Acquisition, Storage, and Analytics**

Objectives

The objective of this study was to produce advanced data compression algorithms tailored to outer solar system instrument concepts. We sought to augment standard "off-the-shelf" algorithms with machine learning to discover improvements in the source encoder/decoder's ability to optimally allocate compressed bits and minimize distortion. We aimed to characterize the impact of these algorithms on scientific retrievals.

We benchmarked our algorithms' performance using data from the Cassini Imaging Science Subsystem (ISS). In particular, we assessed their effects on 1) subtle features of haze density in images of Titan's atmosphere ("detached haze"), 2) barely-visible anomalies ("propeller structures") in images of Saturn's rings, and 3) the ability to match tie points in images of flyby encounters with Saturn's icy satellites.

Our long-term goal is to infuse our techniques into future missions, providing scientists a library of tunable compressors/decompressors and the means for optimizing them and selecting a satisfactory compression rate. This would change the paradigm by which instrument science teams currently perform compression, and encourage its application.

Background

Instruments such as spectrometers and imagers generate far more data than can be returned to Earth for processing and analysis. Generally, science teams are provided with a data compressor to be used at their discretion. Since retrieval processes are often sensitive to even subtle compression artifacts, scientists tend to be reluctant to use it.

A recent study [1] demonstrated that conventional techniques, such as entropy coding and wavelet transforms, could be used to compress Cassini data more aggressively than the Cassini team had, without compromising science goals. The study revealed that conventional methods could be tailored to particular data by identifying prominent information-bearing structure.

Machine learning allows us to address this problem in an automated fashion. Neural networks can be trained to optimize transformation and bit-allocation schemes for encoding data in compressed form, and for decompression and post-processing. Combining them with conventional algorithms known to perform well on specific data, we can achieve even more aggressive compression rates without sacrificing the scientific integrity of the data.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

Approach and Results

Titan Haze Retrievals:

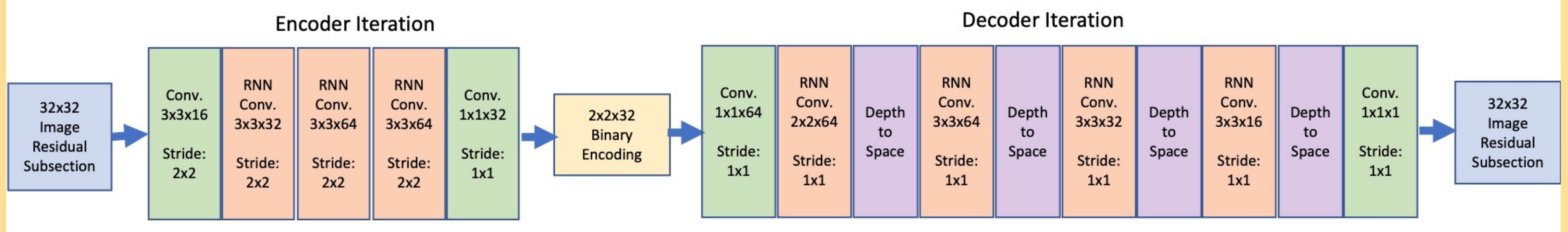
We found that a small number of Daubechies (db5) and ICER wavelet transform coefficients capture most of the information content of Titan images. Using a network design motivated by [2], shown in Fig. 1, we were able to encode the residuals of the images after removing a fixed set of wavelet coefficients. We pruned our training set to image subsections deemed likely to contain detached haze [3]. The network learned how to encode/decode image residuals to preserve the detached haze better than by simply keeping substantially more wavelet coefficients, and compressing at a higher overall factor. (Fig. 2) In some test cases, we were able to retrieve detached haze from only 1% of the db5 wavelet coefficients with full floating point precision, combined with neural network assistance, compressing by an overall factor of 9.8. By comparison, we were unable to recover the detached haze without the neural network, even after keeping ten times as many wavelet coefficients---compressing only by a factor of 2.5. (Fig. 3)

Identifying Regions-of-Interest in Wavelet domain for Compression of ISS-Rings:

We investigated methods to identify potential regions-of-interest based on auto-correlation in the transform domain. Focusing on ISS Saturn-rings images, we found that when the downlink bandwidth is highly constrained, subtle "propeller" structures [4] can be lost in a general compression framework due to their low amplitudes. By applying a wavelet domain auto-correlation, we were able to illuminate these features in multiple subbands. We could then prioritize the associated wavelet coefficients during compression, and preserve the regions-of-interest. (Fig. 4)

Tie Point Matching in Icy Satellites:

Analyzing flyby image sequences of Saturn's icy satellites, we found that by using earlier images to predict later ones, we can increase our overall compression rate for a given level of distortion. Using images of Tethys, we attempted to co-register successive images with translations described by few bits. Using earlier images to predict later ones, we applied the wavelet/neural network approach from our Titan haze retrievals to compress the residuals between each image and its prediction. Our results indicate that using image prediction, we can achieve lower distortion compared to compressing the original image even when storing only half as many db5 wavelet coefficients (0.5% compared to 1% of the total coefficients), increasing our compression factor from 9.8 to 12.1. (Fig. 5)



Encoder Iteration

32x32 Image Residual Subsection
Conv. 3x3x16 Stride: 2x2
RNN Conv. 3x3x32 Stride: 2x2
RNN Conv. 3x3x64 Stride: 2x2
RNN Conv. 3x3x64 Stride: 2x2
Conv. 1x1x32 Stride: 1x1
2x2x32 Binary Encoding

Decoder Iteration

Conv. 1x1x64 Stride: 1x1
RNN Conv. 2x2x64 Stride: 1x1
Depth to Space
RNN Conv. 3x3x64 Stride: 1x1
Depth to Space
RNN Conv. 3x3x32 Stride: 1x1
Depth to Space
RNN Conv. 3x3x16 Stride: 1x1
Depth to Space
Conv. 1x1x1 Stride: 1x1
32x32 Image Residual Subsection

Figure 1. A single iteration of the networks used to encode/decode image residuals after applying wavelet-based compression. The network applies a series of convolutional and recurrent convolutional layers to compress and decompress a 32x32 image subsection. Each iteration compresses the reconstruction error with respect to the previous one. By default, we apply this network for 8 iterations, adding a total of 1 bit per pixel.

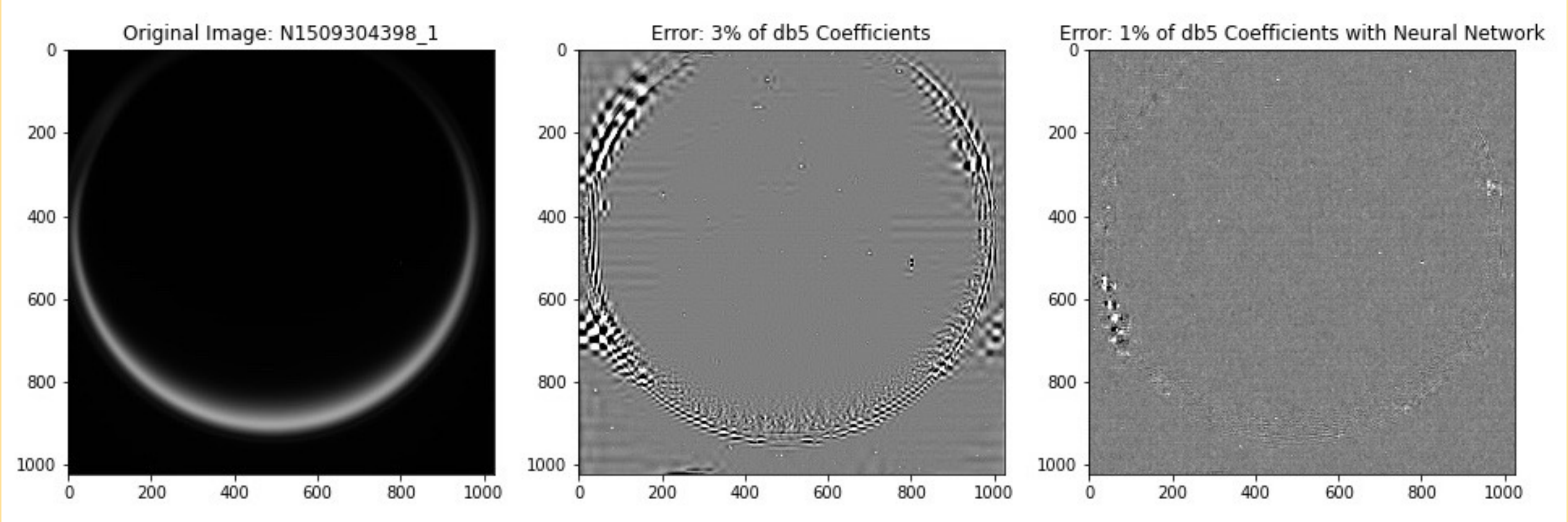


Figure 2. Results of compressing a Titan haze image. The leftmost figure is the original image. The center shows the error pattern in the reconstructed image from storing 3% of the coefficients of the db5 wavelet transform (compressing by a factor of 8.33). The right image is the error using only 1% of the db5 coefficients, with neural network assistance (a compression factor of 9.8). The latter method eliminates most compression artifacts while achieving a higher compression rate.

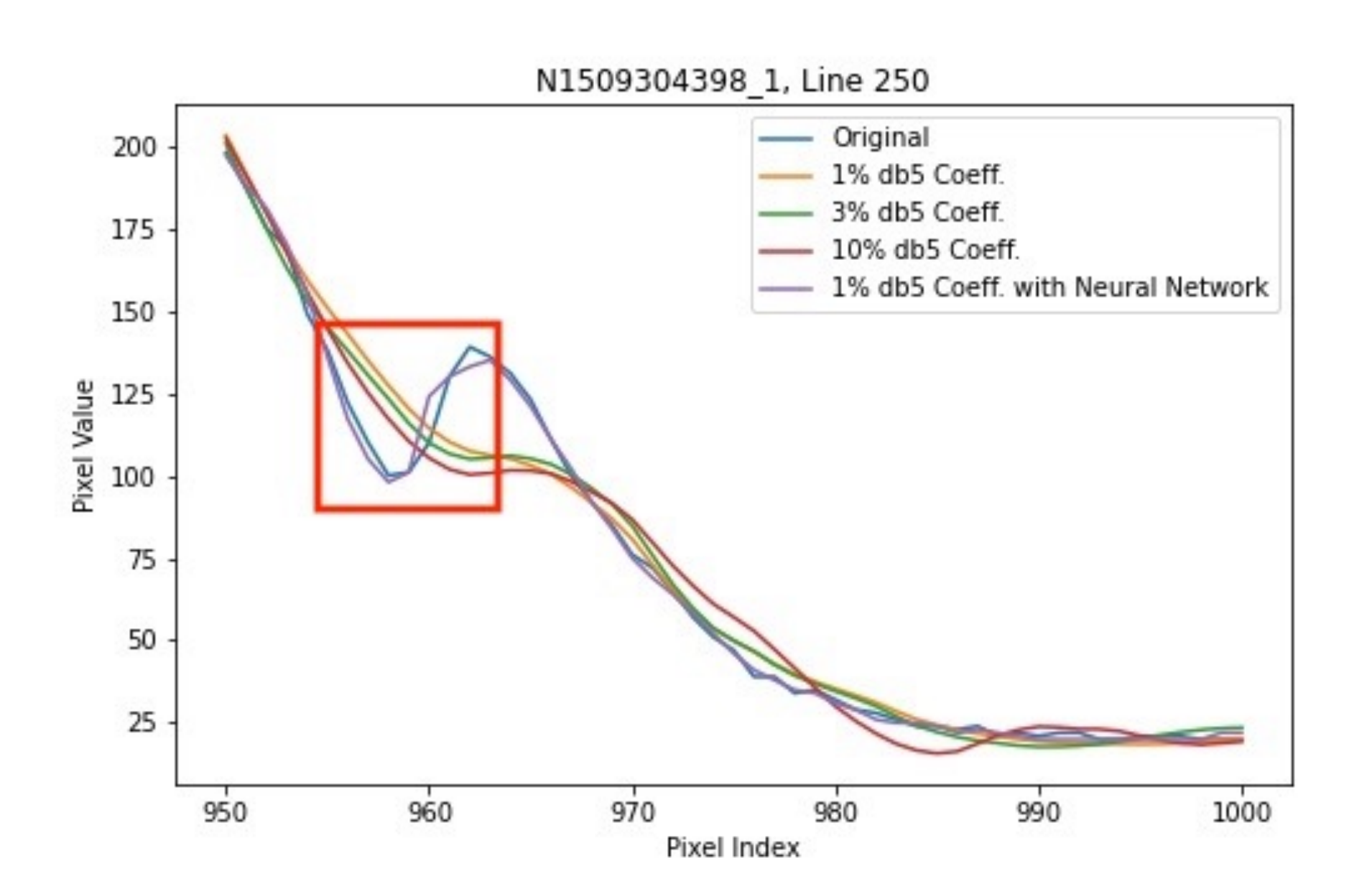


Figure 3. Pixel values of original and reconstructed images of Titan haze. Applying the network from Fig. 1 to the image residual after storing 1% of the db5 wavelet coefficients, we recover the signature "detached haze bump" outlined in the red box, compressing by an overall factor of 9.8. Without neural network assistance, we fail to recover the detached haze even when storing ten times as many coefficients, compressing only by a factor of 2.5.

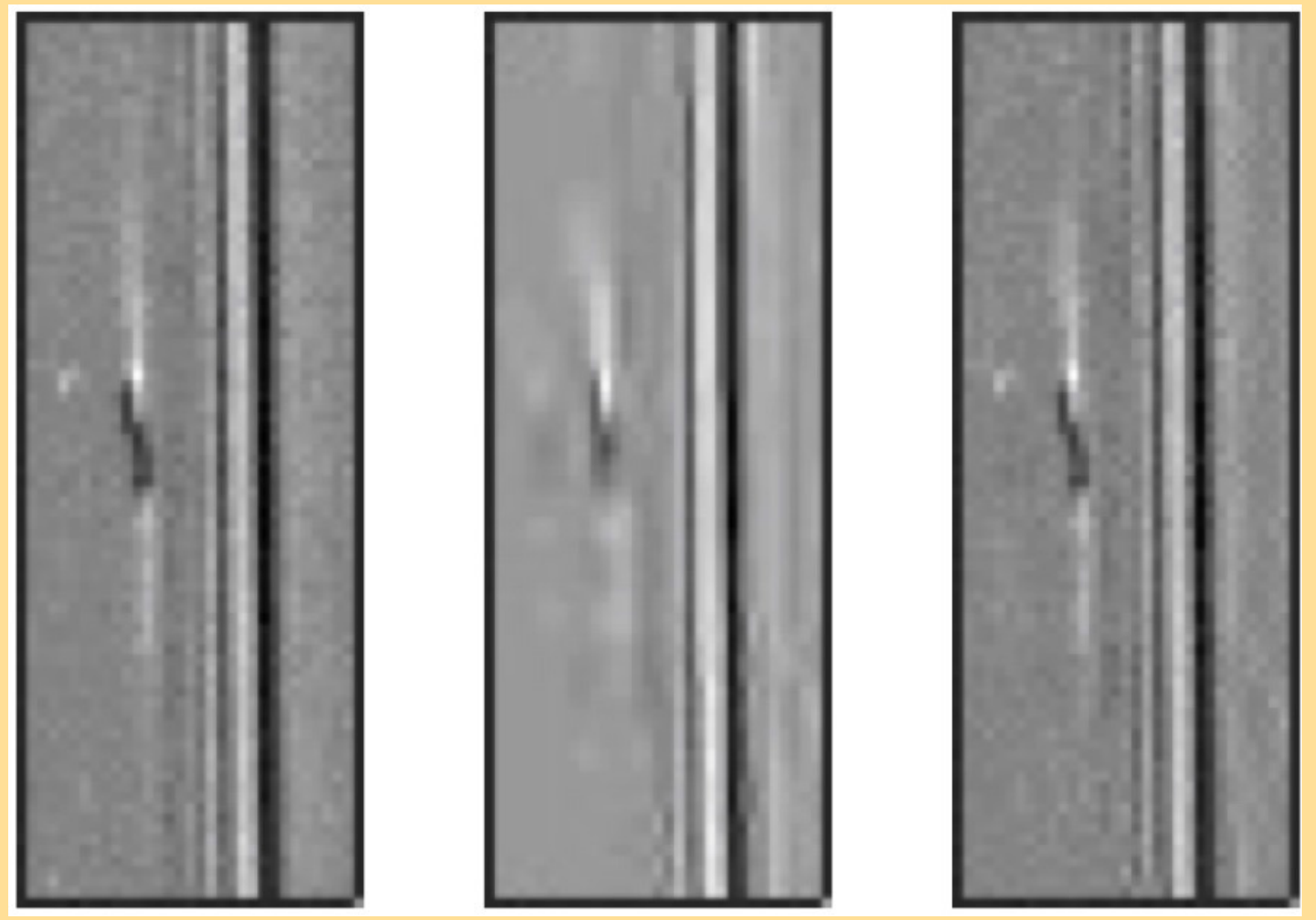


Figure 4: Left: Subsection of an original image of Saturn's rings which contains a propeller structure. Center: Reconstruction of image using original wavelet-transform-based compression, restoring image from 5% of wavelet coefficients. Right: Reconstruction of image using Region-of-Interest (ROI) transform compression, using an autocorrelation in the wavelet domain to identify wavelet coefficients likely associated to the ROI, and prioritizing them when selecting 5% of the coefficients from which to restore the image. ROI transform compression substantially reduces the smearing of the propeller.

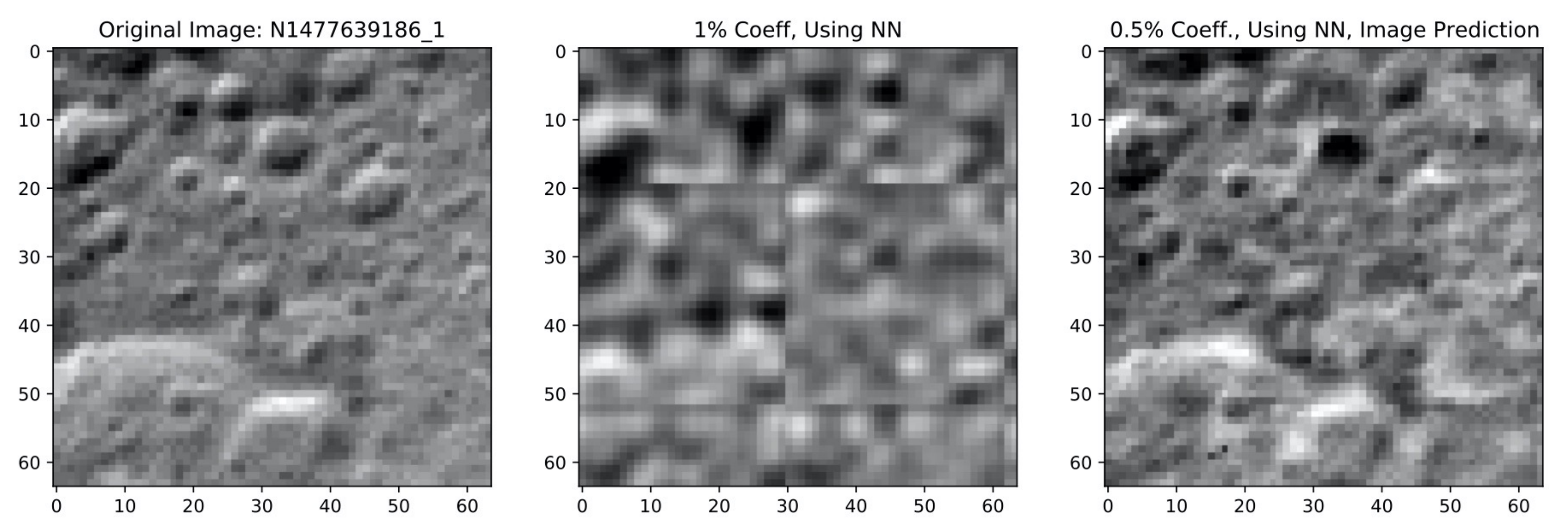


Figure 5: Left image: an original subsection of an image from a flyby sequence for Tethys. Center: the reconstructed version using 1% of the db5 wavelet coefficients and the neural network assistance, compressing by a factor of 9.8. Right: the reconstruction using only 0.5% of the wavelet coefficients, and applying the neural network compressor to the residual from predicting this image from earlier ones in the flyby. The overall compression factor rises to 12.1, and we are able to resolve some of the finer topographic features of the image, as well as reduce compression artifacts.

Significance/Benefits to JPL and NASA

Every bit of information from an Outer Solar System mission represents an engineering triumph over severe constraints on mass (limiting antenna size) and power (limiting signal strength). The potential for higher compression ratios promises a significant impact on future mission concepts by either returning significantly more data, allowing more instruments to make simultaneous observations, or enabling additional instruments to be carried on the same mission by trading antenna mass for science instrument mass. Currently, standard lossy compressors are inadequate for many scientific standards. Even JPEG compression---similar to the CCSDS standard for 2D image compression---was found to introduce noticeable artifacts into Cassini images of Saturn, and degrade high spatial frequency components.

This work will give researchers the tools and methodology to build custom compressors/decompressors catered to their specific demands. By using tunable algorithms, scientists will be able to select personalized combinations of conventional compression schemes and neural network assistants whose parameters can be trained on a sample of uncompressed or losslessly compressed data to optimize their particular retrievals. We envision a JPL-supplied spacecraft facility compressor hosting a library of compression schemes optimized for all high data volume instruments on future missions.

References

[1] Xie, H., West, R. A., Seignovert, B., Jewell, J., Kurth, W., Averkamp, T., "Compression Algorithms for High Data Volume Instruments on Planetary Missions: a Case Study for the Cassini Mission," J. Astron. Telesc. Instrum. Syst., 7(2), 028002, doi: 10.1117/1.JATIS.7.2.028002, 2121.

[2] Toderici, G., Vincent, D., Johnston, N., Jin Hwang, S., Minnen, D., Shor, J., Covell, M., "Full Resolution Image Compression with Recurrent Neural Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 5435-5443.

[3] West, R. A., Balloch, J., Dumont, P., Lavvas, P., Lorenz, R., Rannou, P., Ray, T., Turtle, E. P., "The evolution of Titan's detached haze layer near equinox in 2009," Geophysical Research Letters, 38(6), 2011.

[4] Tiscareno, M. S., Burns, J. A., Hedman, M. M., Porco, C. C., "The Population of Propellers in Saturn's A Ring," The Astronomical Journal, 135(3), 2008.

PI/Task Mgr Contact
Email: Matthew.D.Thill@jpl.nasa.gov

Clearance Number: URS303388
RPC/JPL Task Number: R21112