The Astrophysical Journal Supplement Series, 268:33 (8pp), 2023 September https://doi.org/10.3847/1538-4365/ace9d7

Homogenizing SOHO/EIT and SDO/AIA 171 Å Images: A Deep-learning Approach

Subhamoy Chatterjee[1], Andrés Muñoz-Jaramillo[1], Maher A. Dayeh[2,3], Hazel M. Bain[4,5], and Kimberly Moreland[2,3,4,5]
[1] Southwest Research Institute, Boulder, CO, USA; subhamoy.chatterjee@swri.org
[2] Southwest Research Institute, San Antonio, TX, USA
[3] The University of Texas at San Antonio, San Antonio, TX, USA
[4] Cooperative Institute for Research in Environmental Sciences, University of Boulder, CO, USA
[5] Space Weather Prediction Center, NOAA, Boulder, CO, USA
Received 2023 May 30; revised 2023 July 18; accepted 2023 July 22; published 2023 September 11

Abstract

Extreme-ultraviolet (EUV) images of the Sun are becoming an integral part of space weather prediction tasks. However, having different surveys requires the development of instrument-specific prediction algorithms. As an alternative, it is possible to combine multiple surveys to create a homogeneous data set. In this study, we utilize the temporal overlap of Solar and Heliospheric Observatory Extreme ultraviolet Imaging Telescope and Solar Dynamics Observatory Atmospheric Imaging Assembly 171 Å surveys to train an ensemble of deep-learning models for creating a single homogeneous survey of EUV images for two solar cycles. Prior applications of deep learning have focused on validating the homogeneity of the output while overlooking the systematic estimation of uncertainty. We use an approach called "approximate Bayesian ensembling" to generate an ensemble of models whose uncertainty mimics that of a fully Bayesian neural network at a fraction of the cost. We find that ensemble uncertainty goes down as the training set size increases. Additionally, we show that the model ensemble adds immense value to the prediction by showing higher uncertainty in test data that are not well represented in the training data.

Unified Astronomy Thesaurus concepts: Solar corona (1483); Convolutional neural networks (1938); Calibration (2179); Solar atmosphere (1477); Solar extreme ultraviolet emission (1493)

1. Introduction

Solar observations span across several past decades being recorded by multiple ground-based and space-based observatories. Due to changes in instrumentation, the data sets differ in resolution (spatial and temporal), field of view, dynamic range, and noise characteristics. While dealing with long-term solar data, the detection of solar features/events requires building custom algorithms for each instrument. Cross-calibrating those surveys and creating a single homogeneous data set for the entire span of observation enables the scientific community to make long-term studies and discover underlying patterns without spending effort to deal with instrument differences.

There have been several efforts to homogenize solar images from different surveys using traditional approaches such as oversampling of pixels using interpolation, intensity rescaling, etc. Recent studies find a significant improvement over those baseline approaches in homogenization tasks such as super-resolving solar magnetograms (Jungbluth et al. 2019) utilizing state-of-the-art machine-learning (ML) approaches such as convolutional neural networks (CNNs).

Deep-learning (DL) approaches applied in scientific domains suffer from the shortcoming of limited availability of data. Limited data create barriers for ML models to learn the diversity of natural phenomena. Thus, in such situations trusting a point estimate by an ML model on unseen data can result in underestimation and overestimation. Creating model ensembles can help mitigate this problem by generating prediction uncertainty. There are different approaches to creating DL model ensembles that range from multiple selections of data to random disconnections in neural networks (Gal & Ghahramani 2015) and approximating Bayesian neural networks by adding several anchors that drive the regularization term in the model loss (Pearce et al. 2018). Studies find the application of usage model ensembles to be useful in overcoming the demerits of data scarcity in scientific domains such as heliophysics (Chatterjee et al. 2022).

In this study, we use CNNs to homogenize and estimate the uncertainty of full-disk solar extreme-ultraviolet (EUV) images in 171 Å from the Extreme-ultraviolet Imaging Telescope (EIT; Delaboudinière et al. 1995) on board Solar and Heliospheric Observatory (SOHO) and the Atmospheric Imaging Assembly (AIA; Lemen et al. 2012) on board Solar Dynamics Observatory (SDO) using their overlapping period.

2. Observational Data

We use level-1 SOHO/EIT and SDO/AIA 171 Å full-disk images (Figure 1) with a cadence of 1 day and over the period 2010–2020 for our study. We make sure that the EIT–AIA pairs are observed at the same time.

3. Data Alignment

3.1. Reprojection

The EIT–Sun–AIA angle, focal length, and detector differences necessitate reprojecting EIT–AIA images to a common point of view and field of view. We use the "reproject" module of Python to map AIA images to EIT detector with 4 times the original resolution of EIT (Figure 2). The reprojection of the corona is nontrivial. So, we make an assumption that coronal structures lie on a sphere centered on Earth, with a radius of the Sun–Earth distance.

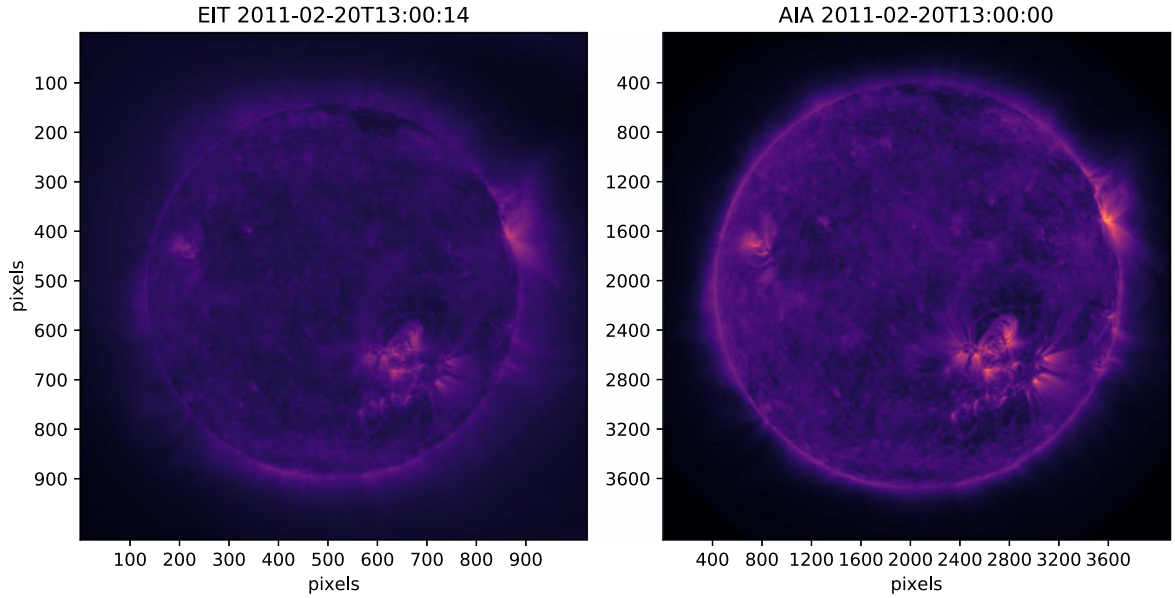

Figure 1. Level-1 representative images from SOHO/EIT and SDO/AIA from the same date. Left panel shows an EIT 171 Å image at the native resolution and field of view. Right panel shows an AIA 171 Å image at native resolution and field of view. A comparison of the images clearly depicts the higher field of view of EIT.

3.2. Template Matching

Reprojection is insufficient when we want to perform one-to-one mapping of EIT to AIA patches. Several other factors such as resolution and optical aberration result in pixel shifts and may affect our homogenization objective. To make an accurate alignment, we perform template matching (Brunelli 2009) using the following steps:

1. Divide the EIT full-disk images into contiguous non-overlapping 64×64 patches.
2. Blow up each of those patches by 4 times and map that location to the reprojected AIA.
3. Correlate that blown-up EIT patch within the corresponding AIA location having a margin of (1/8)th patch size on all four sides.
4. Find the 256×256 region from the AIA window providing the best correlation.
5. Follow the above steps after generating another set of nonoverlapping patches bounded by the center pixels of the original patches.

The steps above produce 365 patch pairs [64×64 EIT, 256×256 AIA] per full-disk EIT–AIA image pair (Figure 3).

3.3. Data Preparation: Training, Validation, and Test Set

We sample decade-long data (2010 May–2020 December) within the training, validation, and test set by adding the first 7 months of each year to the training set, August and September to the validation set, and the rest to the test set. We also put the entire 2015 data aside in the test set and tune the model hyperparameters based on the validation set performance.

4. DL-based Image Transformation

4.1. DL Model

We prepare the DL model to transform 64×64 EIT patches to 256×256 AIA patches. The model is based on the superresolution CNN presented in Deudon et al. (2020). However, our network is simple as we do not have multiple views and avoid shift-net by prealigning the EIT and AIA patches. The model consists of an encoder and a decoder structure. The encoder consists of 2D convolution layers and residual blocks. The decoder consists of $4\times$ upsampling and 2D convolution layers. In each residual block, two convolution layers are used and the convolution layer outcome is added to the input layer to generate the output. We use a kernel size of 5×5, 32 feature layers and a parametric ReLU activation in all the convolution layers except for the final layer that has only one feature layer and a kernel size of 1×1. We use reflection padding the deal bring edge continuity in the adjacent superresolved tiles.

4.2. Loss Function Definition

To minimize the difference between the target (T) and DL predicted (P) image, we consider the following as elements of loss function:

1. Mean squared error (MSE): $\mathrm{MSE} = \frac{1}{N}\sum_i (T_i - P_i)^2$ where N stands for the number pixels in output patches.
2. Histogram (Wang et al. 2018) difference: $H = \frac{1}{K}\sum_i (HT_i - HP_i)^2$ where K stands for the number of bins in the histograms.
3. Negative of structural similarity index metric (Zhou et al. 2004): $S = -\frac{(2\mu_T\mu_P + C_1)(2\sigma_{T,P} + C_2)}{(\mu_T^2 + \mu_P^2 + C_1)(\sigma_T^2 + \sigma_P^2 + C_2)}$ with $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, L being dynamic range and $K_1, K_2 \ll 1$. μ_X, σ_X, and $\sigma_{X,Y}$ respectively represent mean of X, standard deviation of X, and covariance of X, Y.
4. Gradient difference: $G = \frac{1}{N}\sum_i (\nabla_x T_i - \nabla_x P_i)^2 + \frac{1}{N}\sum_i (\nabla_y T_i - \nabla_y P_i)^2$.

Our objective is to perform the best-performing weighted combination of MSE and one of the other losses. To identify the weights, we combine MSE with each of the remaining components and calculate the best weight through a set of diagnostics as described in the following subsection.

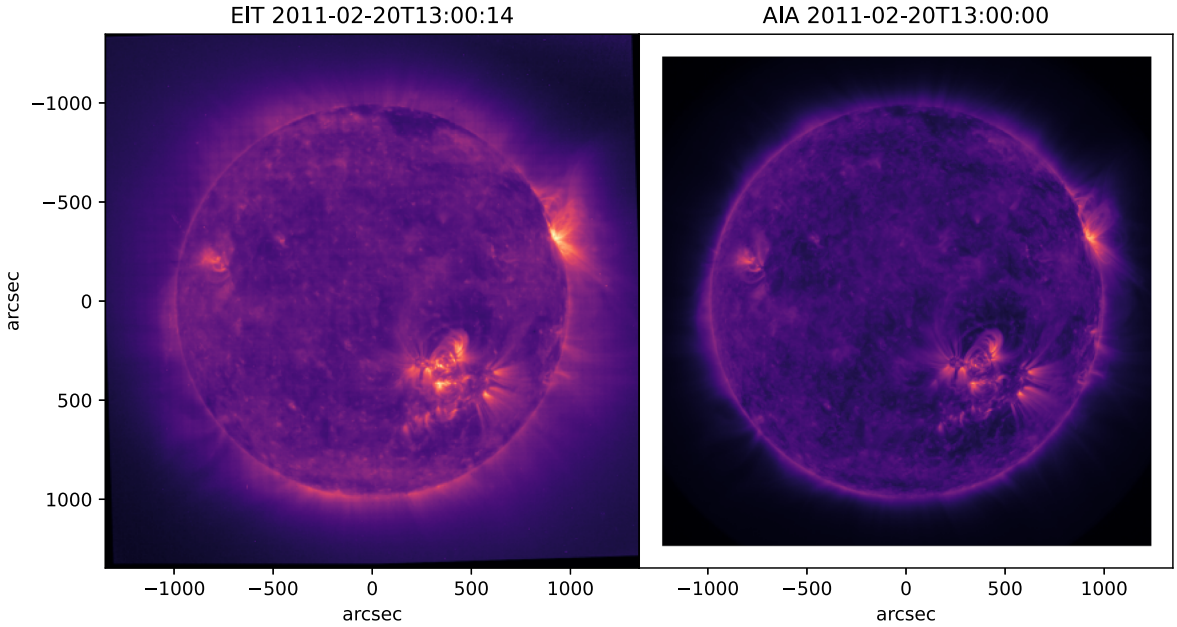

Figure 2. Outcome of reprojection. Left panel shows an EIT 171 Å image; right panel shows a reprojected AIA image to match the field of view and point of view seen by the EIT detector.

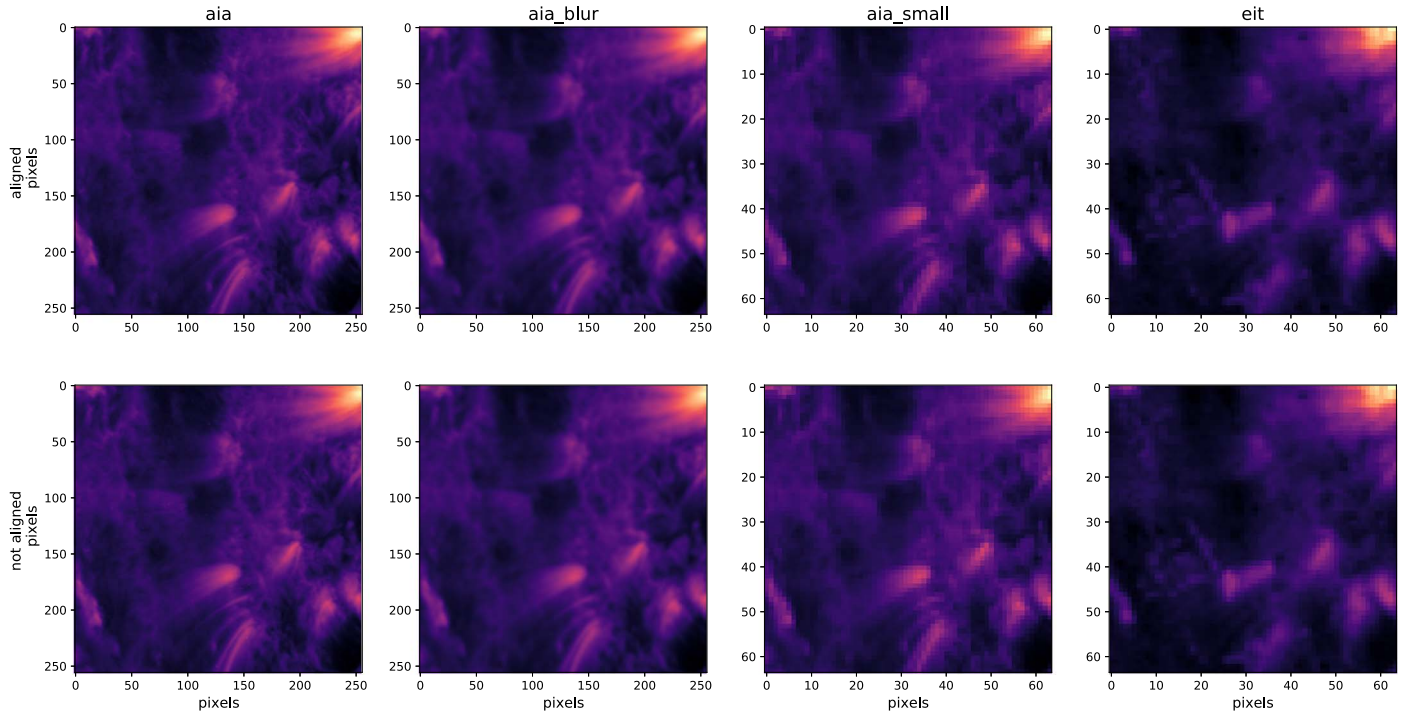

Figure 3. Benefit of patch alignment while creating input-target pair for EIT-to-AIA transformation. Top row shows the effect of patch alignment while the bottom row shows unaligned patches. While comparing the unaligned AIA patches with the aligned ones, the vertical and horizontal shift correction becomes apparent by looking at the bright region on the top right. Different columns for the AIA patch represent steps toward undersampling for visual comparison with EIT at native resolution.

4.3. Approximate Bayesian Ensembling

Having a small amount of training data it is important that we estimate the uncertainty in model inference. We use a state-of-the-art technique called approximately Bayesian ensembling (ABE) that draws a set of anchor weights from a prior distribution and factor them through an additional regularization term to the loss function defined by

$$\text{Loss}_n = \frac{1}{N}\,\|\boldsymbol{I} - \hat{\boldsymbol{I}}_n\|_2^2 + \frac{1}{N}\,\|\boldsymbol{\Gamma}^{1/2}.(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{\text{anchor},n})\|_2^2, \quad (1)$$

where $diag\,(\boldsymbol{\Gamma})_i = \frac{\sigma_\epsilon^2}{\sigma_{\text{prior},i}^2}$ and N is the number of data points.

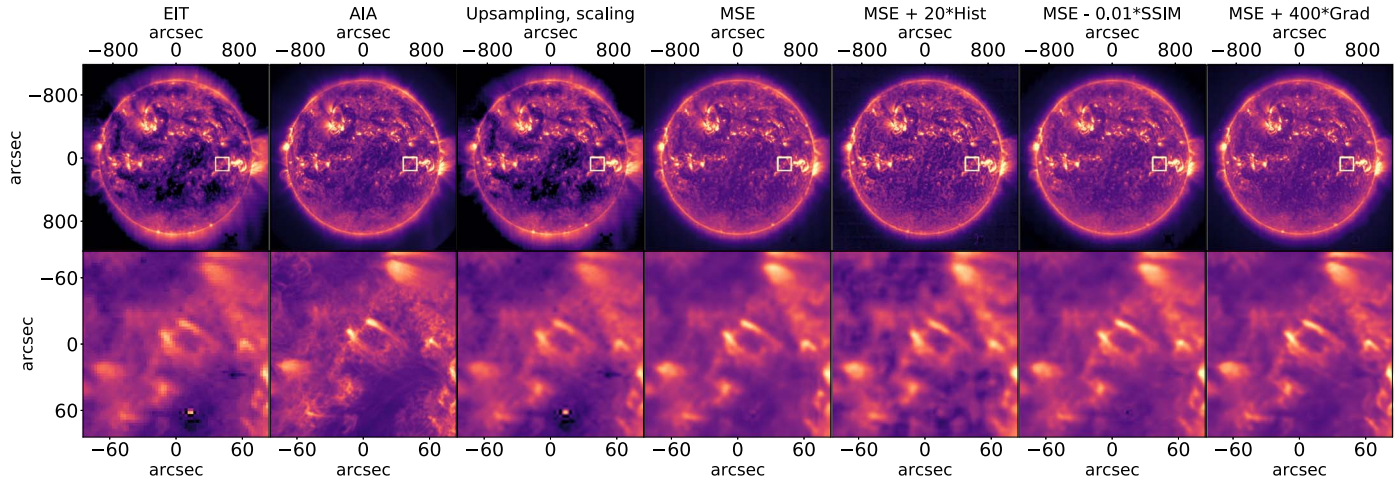

Figure 4. Effect of loss function variations of inferred high-resolution image. The baseline outcome is depicted by the third column while starting from the fourth column all the outcomes toward the right are from a DL model with loss function shown at the titles. The effect of loss function variation is highly apparent the model outcome. For example, MSE + 20*Hist does not show as many microdetails as depicted by the MSE + 400*Grad. It can also be seen that the DL outcomes get rid of the apparent artifact (bright dot on the patch bottom half) on EIT.

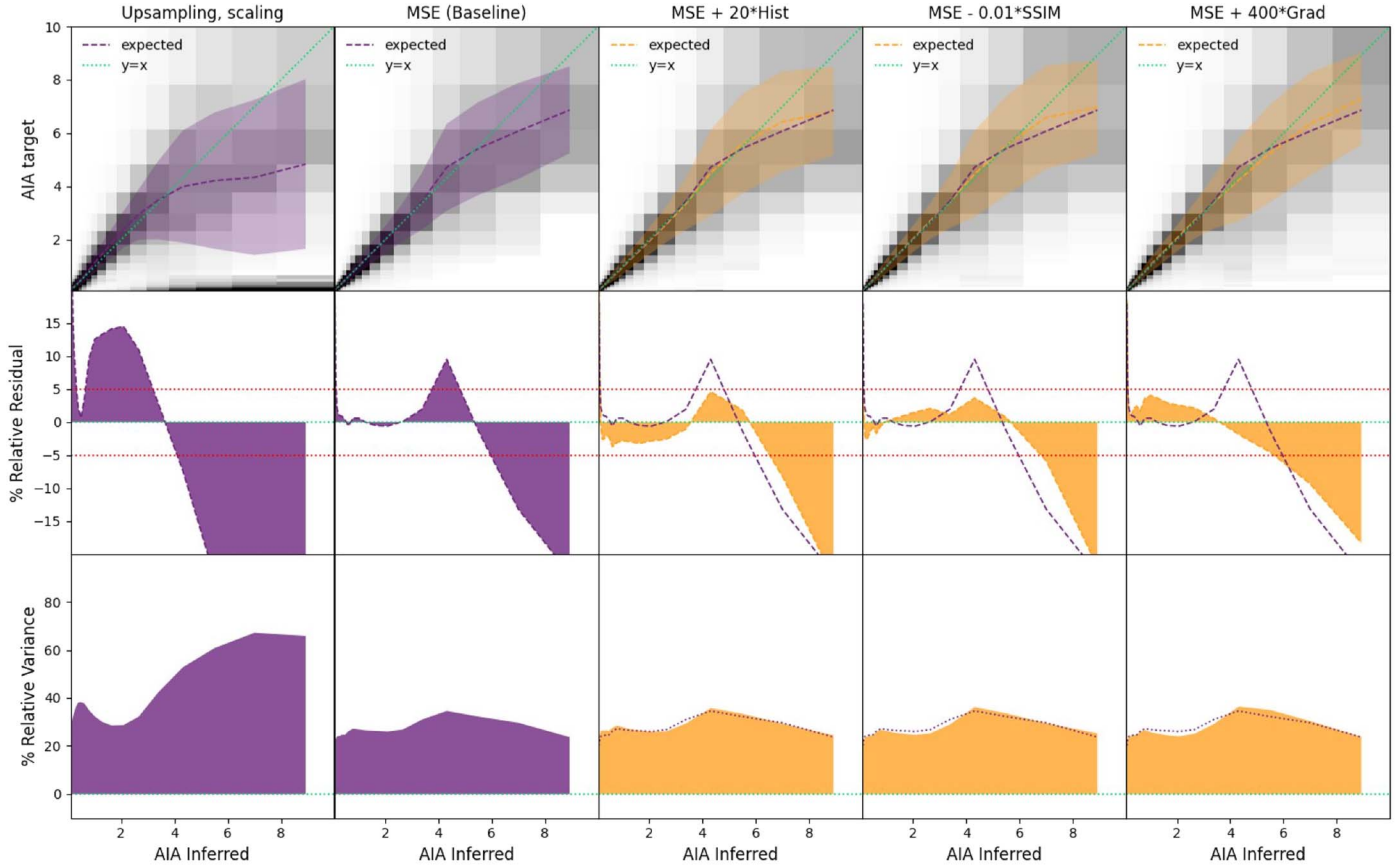

Figure 5. Change in metrics defining the quality of inferred images with change in the loss function. The leftmost column shows the outcome of upsampling and scaling. Columns (2)–(5), starting from left, show the outcomes of the DL model acquired through minimizing a loss printed as the title of each column. The top row shows a 2D histogram between the target and inferred AIA intensities across all the test patches. The dashed lines represent the median line calculated through each input inferred intensity bin and the shaded region represents the interquartile range. The green diagonal line represents "$y = x$." The middle row depicts the relative residual for meaning percentage deviation of the median line from "$y = x$" line. The horizontal dotted lines mark $+/-5\%$ relative residual. The bottom row represents relative variance meaning the percentage uncertainty of outcome with respect to the median. The outcome for MSE loss is overplotted as a baseline on all other loss function outcomes.

During the training phase, the weights ($\boldsymbol{\theta}_n$) are optimized but the anchors ($\boldsymbol{\theta}_{\text{anchor},n}$) are kept unchanged. We randomly select a set of four anchors and optimize a model for each giving rise to the ensemble of models. Here the index n runs over the ensemble members. We set σ_ϵ^2 to be the mean of the AIA patch histogram. We use anchored regularization in the last layer of

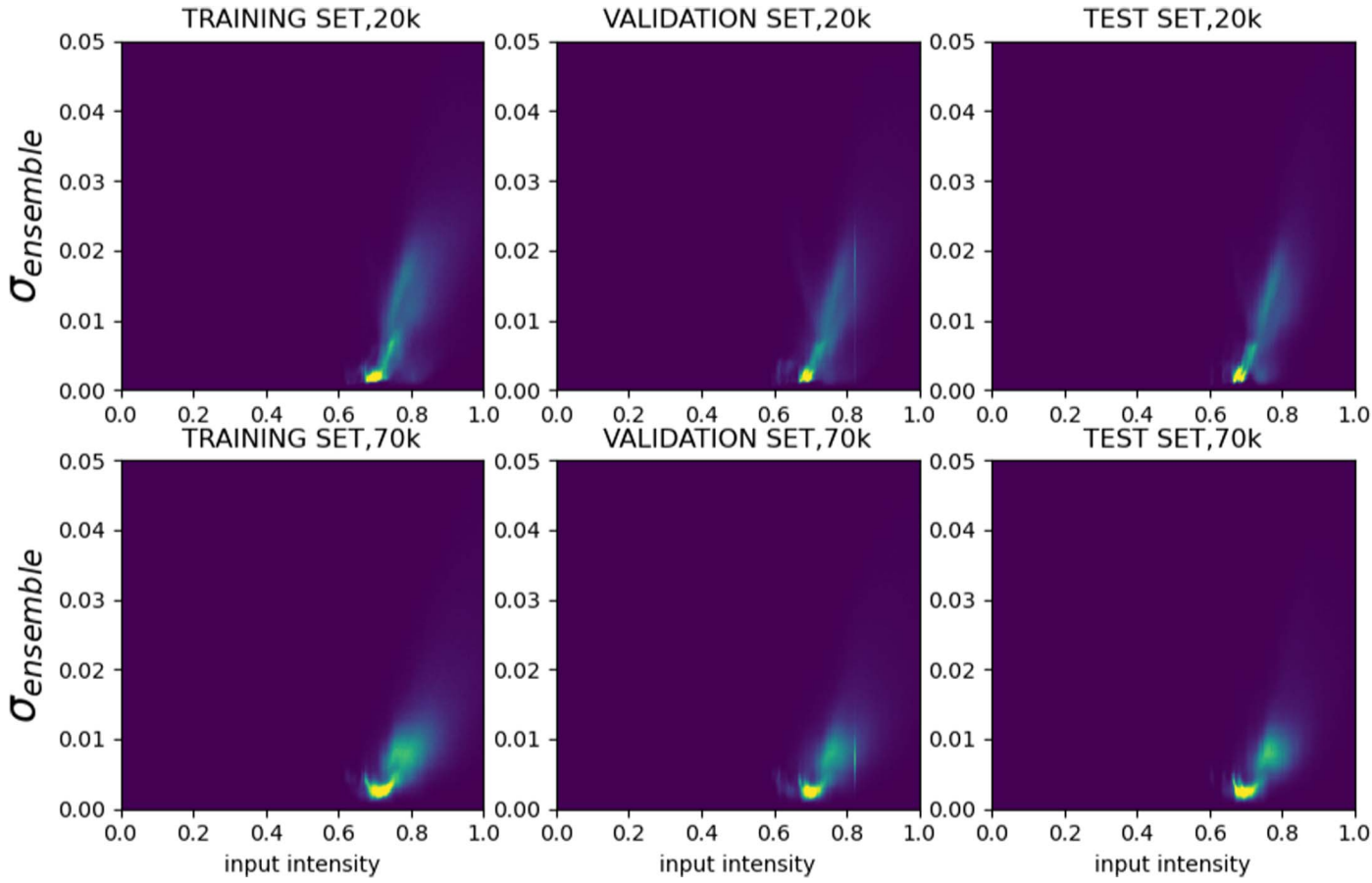

Figure 6. Change in inference uncertainty with training set size. Each panel represents a 2D histogram between input pixel intensity and standard deviation of ensemble inference (σ_{ensemble}) across four superresolved pixels. The top row represents the 2D histograms made from training, validation, and test patches when the models were trained with 20,000 patch pairs. The bottom row represents the 2D histograms made from training, validation, and test patches when the models were trained with 70,000 patch pairs. A clear reduction in ensemble SD can be observed especially for higher-intensity pixels when the training set gets bigger.

our model to select both priors and anchors from a Glorot uniform (Glorot & Bengio 2010) distribution with $\sigma^2_{\text{prior},i} = \frac{6}{n_l + n_{l+1}}$, where n_l and n_{l+1} are the numbers of units in two consecutive layers. We rewrite $diag(\mathbf{\Gamma})_i$ as $\frac{k_1(n_l + n_{l+1})}{6}$. We use k_1 (with a default value of $\sigma^2_\epsilon \approx 0.5$) to change the regularization weight and another parameter k_2 (with a default value of 1) to scale the standard deviation of the anchor weight distribution. We examine the sensitivity of the results to the anchored regularization with different values of k_1 (later referred to as "LAMBDA") and k_2 (later referred to as "SD").

5. Results

5.1. Model Prediction, Diagnostics, and Comparison with Baseline Approach

We develop a set of diagnostics to compare the model inference between the baseline and outcomes for different loss function combinations. For this purpose, we collect all the validation patch pairs and model inference on those. We generate 2D histograms through logarithmic binning (uniform bins in log space) from the tuple of pixel intensity from model inference and target. We calculate the median and standard deviation for each inferred intensity bin. This generates a curve representing the expected AIA target and uncertainty for each inferred value. From those, we estimate the deviation from the expected rule, i.e., "target = expected." We name that metric as "% relative residual." We also add a metric called "% relative variance" representing the standard deviation and median ratio.

From the visual inspection and metrics (Figures 4 and 5), we find that all the DL outcomes give rise to superior results as compared to those corresponding to simple Upsampling and scaling. As depicted in Figure 4, the DL outcomes get rid of the bright point-like artifact present in the EIT patch. Among the DL outcomes, we use MSE loss as a baseline and examine the improvement caused by additional terms in the loss function. We find that unlike MSE additional loss terms drive the "% relative residual" to be within $\pm 5\%$ for an inferred intensity range of $[0, 6]$ (Figure 5). We also observe that MSE + 400*Grad performs the best for higher intensities (>6). Also, by visual inspection, one can discern finer details in MSE +400*Grad more easily as compared to other DL outcomes.

The results show that (although better than the baseline) the DL model cannot reach the same level of detail as the target AIA 171 Å images (Figure 4). This could be because of the ill-posed nature of the problem with AIA having higher diffraction-limited resolution and better pixel pitch than EIT. It could also be the limited training data that are not able to constrain the outcome enough or that the small-scale information cannot be uniquely recovered from the large scale. More experiments should be performed to understand the optimal number of loss function terms beyond which performance cannot be improved given the size of the training data. Recent advancement in conditional generative adversarial networks

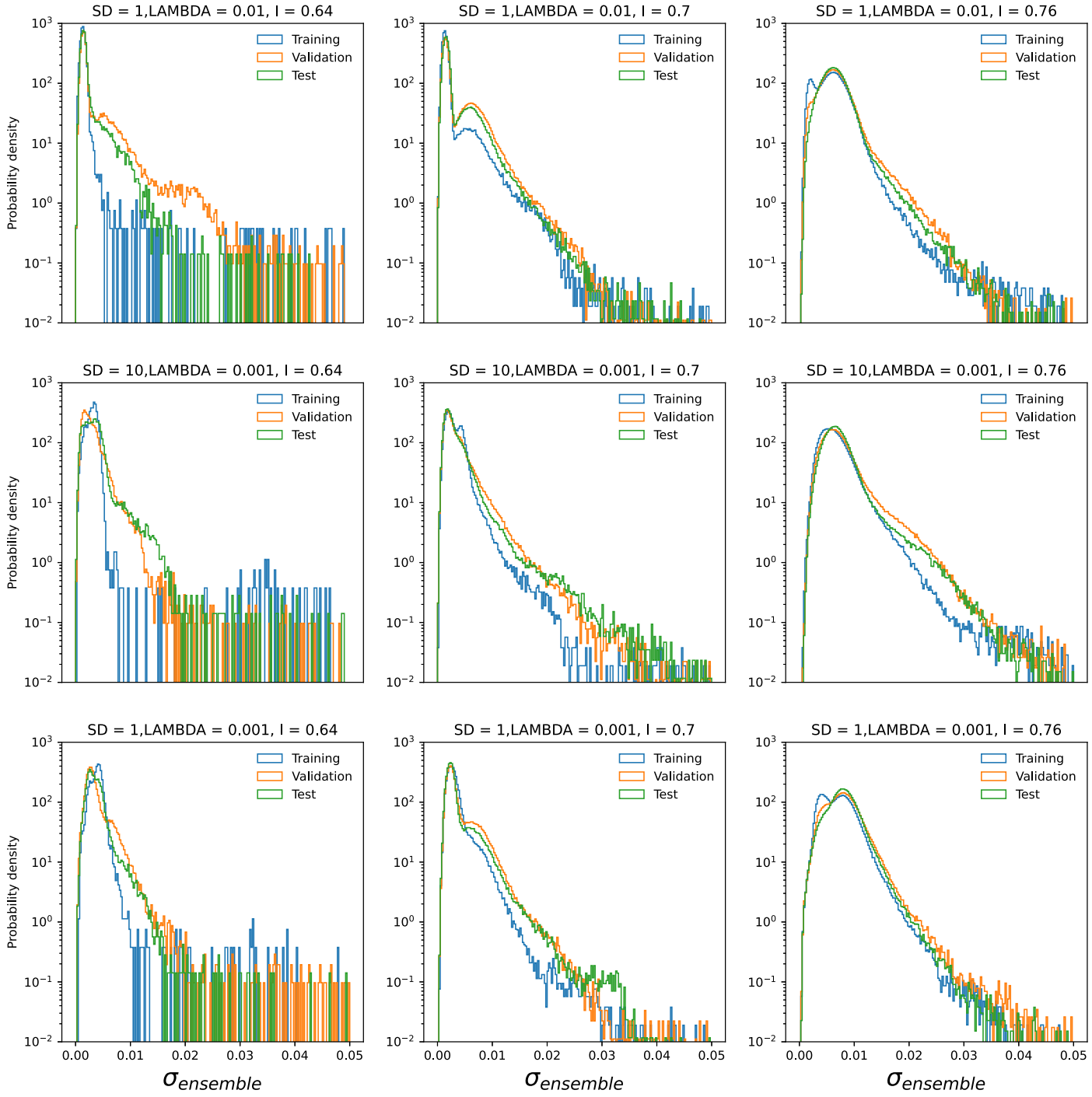

Figure 7. Change in ensemble uncertainty (σ_{ensemble}) with the intensity of input pixels. Each panel shows uncertainty histograms on training, validation, and test sets for a particular input intensity and setting of the ensemble run. Each row represents the effect of a particular standard deviation scaling factor (SD) of the regularization term's anchor distribution and weightage (LAMBDA). Each column represents the outcome for a particular input intensity (I). The stronger difference in histogram tails can be observed for the lowest input intensity, i.e., $I = 0.64$ with validation and test set outcomes being more extended toward higher uncertainty values.

also inspires further experiments to bring back surface texture information but physical validation is needed for them to be fit for scientific applications.

5.2. Sensitivity of Uncertainty to Training Set Size

We select four anchors and train a model with each of those without changing the training and validation set. Initially, we train the model with 20,000 EIT–AIA patch pair and then increase the number of training patches to 70,000. To record the change in performance, we generate 2D histograms between ensemble standard deviation versus input EIT intensity over all the patch pixels for training, validation, and test sets. The ensemble standard deviation (σ_{ensemble}) is calculated over 64 pixels (4×4 for 4 models) corresponding to each EIT pixel. From 2D histograms, we find a clear impact of training set size on the ensemble uncertainty. We observe that for each EIT intensity, the distribution of ensemble uncertainty when trained with 70,000 patches shifts toward

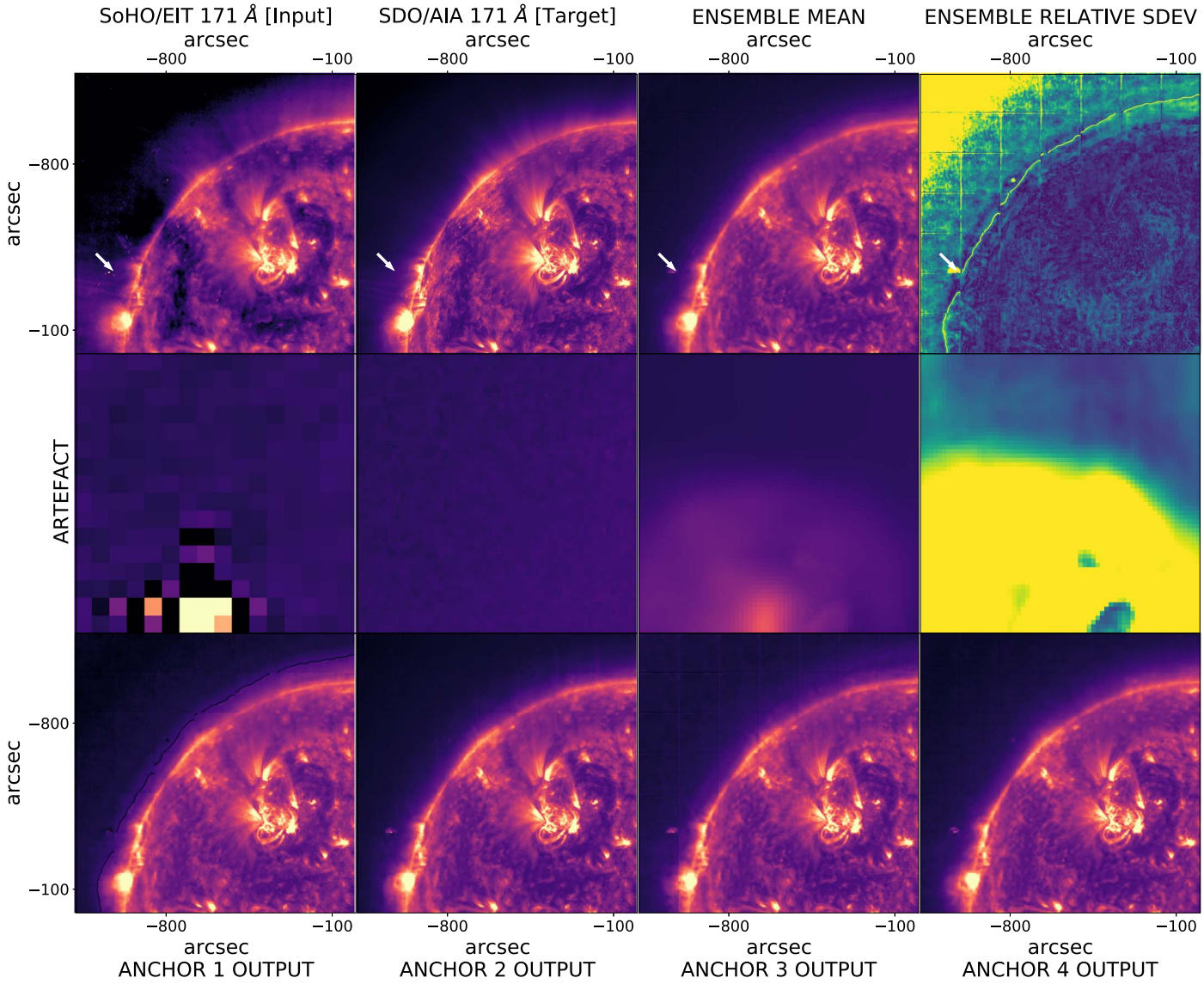

Figure 8. Ensemble inference on a single image quadrant of a representative EUV 171 Å image. The top row starting from left depicts the model input, target, ensemble inference mean, and ensemble relative standard deviation (σ/μ). The bottom row shows the outcomes of different ensemble members. It can be observed those bright point-like artifacts (marked by white arrows in the top row and zoomed in the middle row of images) on the input images show high uncertainty and the across ensemble members. Also, the image for relative ensemble standard deviation depicts high uncertainty for the coronal region in the top left corner.

smaller values as compared to when trained with 20,000 patches (Figure 6).

5.3. Sensitivity of Uncertainty to Ensembling Parameters

We also record changes in the distribution of ensemble standard deviation (σ_{ensemble}) for values of input intensity and parameters defining anchor distribution (k_2 or "SD") and weightage of the regularization term (k_1 or "LAMBDA") in the loss function over training, validation, and test set. It appears that for all the input intensities and anchor parameters, the σ_{ensemble} distributions have a longer tail in the validation and test set as compared to the training set. The difference between validation/test and training set distribution diminishes for higher input intensities. However, the distributions get wider for higher input intensities. We do not find as pronounced an effect by changing the anchor parameters. However, a close inspection reveals that the distribution gets wider with a reduction in regularization weightage for $I = 0.64$. Also, for $I = 0.76$ the second peak in the distribution gets less pronounced for lower weightage (Figure 7).

5.4. Sensitivity of Uncertainty to Observational Artifacts

On a representative solar image quadrant, we find that regions that are not well represented in the training data (e.g., coronal regions) produce high disagreement among the ensembles. Also, structures of nonsolar origin generate high uncertainty (Figure 8).

It is unknown at this moment how many ensemble members are optimal beyond which the uncertainty in reconstruction converges. Further investigations are needed for that.

6. Discussion and Conclusion

We thus trained an ensemble of CNNs to superresolve SOHO/EIT 171 Å images to SDO/AIA 171 Å images using the overlapping time period. The salient features of this work are listed below:

1. We prepared ML-ready data consisting of aligned EIT 171 Å 64 × 64 patches and AIA 171 Å 256 × 256 patches as input and output images.
2. We trained a CNN based on residual blocks and upsampling layers to translate an EIT patch to an AIA patch.
3. We found the DL outcome is always superior to a simple upsampling based on intensity scaling and bicubic interpolation.
4. We added different terms in the loss function and found having SSIM or image gradient as additional terms helps improve the quality of the superresolved images both visually and quantitatively as compared to MSE baseline.
5. We also estimated the uncertainty of the prediction by training an ensemble of models through ABE.
6. We found that uncertainty improves while making the size of the training set larger.
7. We also found that the model ensemble generated high uncertainty for regions that are not well represented in the training set such as coronal regions and regions of nonsolar origin such as artifacts.

We would like to highlight that ABE has been demonstrated to behave like a Bayesian neural network for toy applications for which there is ground truth. However, there is no Bayesian neural network ground truth for homogenization and super-resolution. Our results are encouraging in that ABE uncertainty has desirable and verifiable properties such as sensitivity to input data, robustness to changes in ensembling parameters, and sensible behavior to unexpected features.

As a future work, we would like to investigate the point of convergence of ensemble uncertainty. Also, translation between unpaired image domains (Zhu et al. 2017; Jarolim 2022) can help make use of the large database beyond the overlapping time period and better constrain high-resolution textural information. We also plan to use our homogenized EUV data for scientific investigations in heliophysics toward independent verification of the data quality (or its scientific readiness) indicated by the metrics.

Acknowledgments

This work is mainly supported by SwRI internal research grant R6134 and NASA HGIO grant 80NSSC23K0416. The authors also acknowledge partial support from NASA SWO2R grant 80NSSC20K0290.

ORCID iDs

Subhamoy Chatterjee https://orcid.org/0000-0002-5014-7022
Andrés Muñoz-Jaramillo https://orcid.org/0000-0002-4716-0840
Maher A. Dayeh https://orcid.org/0000-0001-9323-1200
Hazel M. Bain https://orcid.org/0000-0003-2595-3185
Kimberly Moreland https://orcid.org/0000-0002-6202-8565

References

Brunelli, R. 2009, Template Matching Techniques in Computer Vision: Theory and Practice (New York: Wiley)
Chatterjee, S., Muñoz-Jaramillo, A., & Lamb, D. A. 2022, NatAs, 6, 796
Delaboudinière, J. P., Artzner, G. E., Brunaud, J., et al. 1995, SoPh, 162, 291
Deudon, M., Kalaitzis, A., Goytom, I., et al. 2020, arXiv:2002.06460
Gal, Y., & Ghahramani, Z. 2015, arXiv:1506.02142
Glorot, X., & Bengio, Y. 2010, Proc. of Machine Learning Research 9, Proc. of the 13th Int. Conf. on Artificial Intelligence and Statistics, ed. Y. W. Teh & M. Titterington, (PMLR), 249, https://proceedings.mlr.press/v9/glorot10a.html
Jarolim, R. 2022, Research Square preprint: 10.21203/rs.3.rs-1021940/v1
Jungbluth, A., Gitiaux, X., Maloney, S. A., et al. 2019, arXiv:1911.01490
Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17
Pearce, T., Leibfried, F., Brintrup, A., Zaki, M., & Neely, A. 2018, arXiv:1810.05546
Wang, Z., Li, H., Ouyang, W., & Wang, X. 2018, CoRR, arXiv:1804.09398
Zhou, W., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. 2004, ITIP, 13, 600
Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. 2017, in 2017 IEEE Int. Conf. on Computer Vision (ICCV), ed. K. Ikeuchi (Los Alamitos, CA: IEEE Computer Society), 2242