Instrument-To-Instrument translation:

Instrumental advances drive restoration of solar observation series via deep learning

R. Jarolim[1], A. M. Veronig[1,2], W. Pötzi[2], and T. Podladchikova[3]

[1] University of Graz, Institute of Physics, Universitätsplatz 5, 8010 Graz, Austria
[2] University of Graz, Kanzelhöhe Observatory for Solar and Environmental Research, Kanzelhöhe 19, 9521 Treffen am Ossiacher See, Austria
[3] Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, Moscow 121205, Russia
  e-mail: robert.jarolim@uni-graz.at

ABSTRACT

Context. The constant improvement of astronomical instrumentation provides the foundation for scientific discoveries. In general, these improvements have only implications forward in time, while previous observations do not profit from this trend. In solar physics, the study of long-term evolution typically exceeds the lifetime of single instruments and data driven approaches are strongly limited in terms of coherent long-term data samples.

Aims. We demonstrate that the available data sets can directly profit from the most recent instrumental improvements and provide a so far unused resource to foster novel research and accelerate data driven studies.

Methods. Here we provide a general method that translates between image domains of different instruments (Instrument-to-Instrument translation; ITI), in order to inter-calibrate data sets, enhance physically relevant features which are otherwise beyond the diffraction limit of the telescope, mitigate atmospheric degradation effects and can estimate observables that are not covered by the instrument.

Results. We demonstrate that our method can provide unified long-term data sets at the highest quality, by applying it to five different applications of ground- and space-based solar observations. We obtain 1) solar full-disk observations with unprecedented spatial resolution, 2) a homogeneous data series of 24 years of space-based observations of the solar EUV corona and magnetic field, 3) real-time mitigation of atmospheric degradations in ground-based observations, 4) a uniform series of ground-based Hα observations starting from 1973, that comprises solar observations recorded on photographic film and CCD, 5) magnetic field estimates from the solar far-side based on multi-band EUV imagery. The direct comparison to simultaneous high-quality observations shows that our method produces images that are perceptually similar and match the reference image distribution.

1. Introduction

With the rapid improvement of space-based and ground-based solar observations, unprecedented details of the solar surface and atmospheric layers have been obtained. As compared to the 11-year solar activity cycle the development of new instruments progresses over smaller time scales. This imposes additional challenges for the study of long-term variations and combined usage of different instruments. In this study, we address the question on how the information obtained from the most recent observations can be utilized to enhance observations of lower quality. This especially aims at homogenizing long time-series, mitigating atmospheric influences and to overcome instrumental limitations (e.g., resolution limitation, seeing).

The automated homogenization of data sets provides an integral component for long-term studies (e.g., solar cycle studies, historic sunspots, studies of rare events) and for studies that combine data from multiple instruments. Especially when dealing with large amounts of data, the automatic adjustment can be faster and more consistent than treating the data sets of the individual instruments separately (Hamada et al. 2020; Hamada et al. 2021). Data driven methods rely on the diversity and amount of data, and the inclusion of additional data sets can significantly increase the performance (Goodfellow et al. 2016). Enhancing old observation series (e.g., recorded on film) to the standard and quality of the primary modern data set can provide an easily accessible data source. Methods developed for specific instruments often depend on certain observables (e.g., magnetograms, filtergrams of specific wavelengths) that are only partially covered by other data sets. An approximation based on proxies can already provide a suitable basis for automated methods or gives additional information for solar monitoring (e.g., in the frame of space weather predictions). In the regular observation schedule the mitigation of atmospheric effects and quality enhancement is a frequently addressed problem (Ramos et al. 2018; Wöger et al. 2008; Rimmele & Marino 2011).

A principle problem of every enhancement method is the absence of a reference high-quality image. The inversion of the image degradation (e.g., lower spatial resolution, instrumental characteristics) is therefore inferred from artificial degradations (Schawinski et al. 2017; Rahman et al. 2020), from simulation data (Díaz Baso & Asensio Ramos 2018; Jia et al. 2019; Baso et al. 2019) or by estimating the degrading effects (Herbel et al. 2018; Asensio Ramos & Olspert 2021). We argue that a designed degradation can only represent a limited set of observations and can not account for the full diversity of real quality decreasing effects that occur in solar observations. Especially when dealing with atmospheric effects (i.e., clouds, seeing) and instrumental characteristics, the quality degradation is complex to model (Jarolim et al. 2020; Dos Santos et al. 2021). Even with the precise knowledge about the degrading function, every image
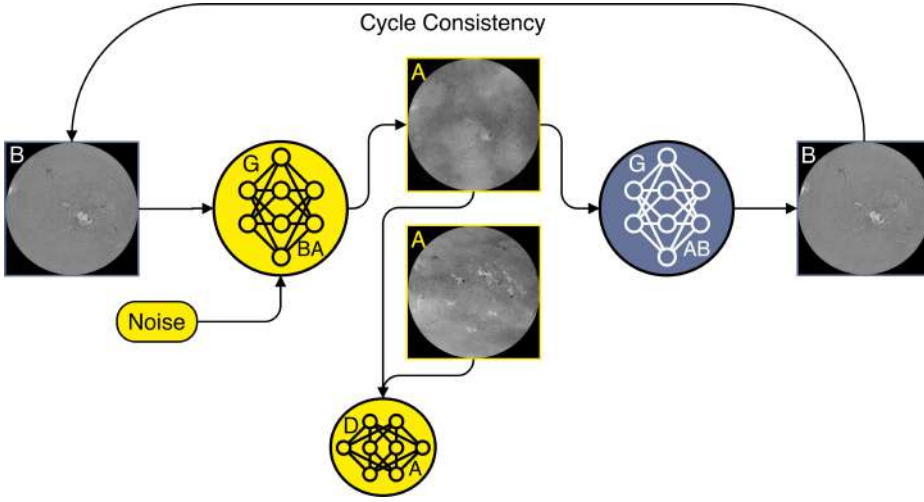
Fig. 1. Model training cycle for the synthesis of low-quality images. Images are transformed from the high-quality domain (B) to the low-quality domain (A) by generator BA (yellow). The synthetic images are translated by generator AB (blue) back to domain B. The mapping into domain A is enforced by discriminator A, which is trained to distinguish between real images of domain A (bottom) and generated images (top). Both generators are trained jointly to fulfill the cycle consistency between original and reconstructed image, as well as for the generation of synthetic images that correspond to domain A. The generation of multiple low-quality versions from a single high-quality image is accomplished with the additional noise term that is added to generator BA.

enhancement problem is ill-posed (Borman & Stevenson 1998; Yang et al. 2019b). We argue that we can reduce the number of possible high-quality solutions for a given low-quality image significantly by considering the distribution of real high-quality images and requiring that enhanced images correspond to the same image domain.

In this study, we propose an approach that uses real observations from state-of-the-art instruments in observational solar physics, to model the image quality distribution of high-quality images. We overcome the limitation of a high-quality reference image with the use of unpaired image-to-image translation (Zhu et al. 2017a). We provide a general method that translates from a given low-quality domain to a target high-quality domain (Instrument-to-Instrument translation; ITI). With this approach, we infer information from real observations to enhance physically relevant features which are otherwise beyond the diffraction limit of the telescope (e.g., super resolution), inter-calibrate data sets, mitigate atmospheric degradation effects and estimate observables that are not covered by the instrument.

Our primary model architecture consists of two neural networks, where the first generates synthetic low-quality images from a given high-quality image (*generator BA*). The second network is trained to invert the image degradation to reconstruct the original high-quality observation (*generator AB*). We enforce the generation of low-quality images with the use of competitive training between generator BA and a discriminator network. We include an additional noise factor for generator BA to model a variety of degrading effects, independent of the image content (Zhu et al. 2017b; Huang et al. 2018). With the synthesis of more realistic and diverse low-quality observations, the generator AB is capable to provide a similar reconstruction performance for real low-quality observations (Fig. 1). The artificial degradation leads inevitably to an information loss that needs to be compensated by the generator AB to reconstruct the original image. Analogously to the training cycle in Fig. 1 we employ a cycle translating low-quality observations to high-quality observations (A-B-A, Sect. 4). This enforces that images by generator AB correspond to the domain of high-quality images, restricting the possible enhanced solutions and gaining information from the high-quality image distribution.

2. Results

In this study, we address the validity of the enhanced images with the use of five data set pairs and the assessment of the sparse set of simultaneous observations. We use ITI to obtain 1) solar full-disk observations with unprecedented spatial resolution, 2) a homogeneous data series of 24 years of space-based observations of the solar EUV corona and magnetic field, 3) real-time mitigation of atmospheric degradations in ground-based observations, 4) a uniform series of ground-based Hα observations starting from 1973, that comprises solar observations recorded on photographic film and CCD, 5) magnetic field estimates from the solar far-side based on multi-band EUV imagery.

A description of the used instruments, data set preparation, preprocessing and training parameters can be found in the supplementary materials (App. A, B). Presented qualitative samples in this section were only selected based on the relevance of their content and are representative for the enhancement that we observe for the majority of samples of the considered data sets. Real high-quality observations that are used as reference are separated by a temporal split from the training data set (see App. B), to exclude a potential memorization of observations.

For all applications, we distinguish between high- and low-quality data sets. Hereby we consider data set pairs, where the high-quality data set contains observations with a better spatial resolution or less quality degradations (e.g., noise, atmospheric effects), as compared to the low-quality data set. As an example, the HMI instrument onboard SDO provides high resolution observations of the Sun, but is defined as low-quality data set as compared to high-resolution observations of Hinode/SOT.

2.1. Image super resolution with different field-of-view - SDO/HMI-to-Hinode/SOT

The Solar Optical Telescope onboard the Hinode satellite (Hinode/SOT; Tsuneta et al. 2008) provides a large data set of partial-Sun continuum images, similar to the full-disk continuum images of the Helioseismic and Mangetic Imager (HMI; Schou et al. 2012) onboard the Solar Dynamics Observatory (SDO; Pesnell et al. 2012). The observations by Hinode/SOT cover various regions of the Sun which makes them suitable as high-quality target for the enhancement of HMI continuum observations.

Using unpaired image translation we do not require a spatial or temporal overlap between the data sets, moreover the model training is performed with small patches of the full images (Sect. 4). This enables the use of instruments that can observe only a fraction of the Sun for the enhancement of full-disk observations.

Here, we resize Hinode observations to 0.15 arcsec pixels and use ITI to super resolve HMI observations by a factor of 4. Hinode/SOT provides a spatial sampling of up to 0.0541 arcsec pixels, for our application we found that a resolution increase by a factor of 4 is already at the limit of where we can properly assess enhanced details.

In Fig. 2 we show an example of the full-disk HMI and ITI enhanced HMI observations (panel a). For a direct comparison, we manually extract two subframes from the HMI full-disk images and the Hinode/SOT image (panel b). The blue box shows the umbra and penumbra of the observed sunspot. The direct comparison shows that the enhanced version is close to the actual observation and correctly separates fibrils that are only observed as blurred structure in the original observation. The yellow box shows a pore and the surrounding solar granulation pattern. Here, we can observe an enhancement of the shape that is close to the Hinode/SOT observation. The granulation pattern is similar to the Hinode/SOT observations for larger granule, but shows differences in terms of shapes and inter granular lanes. Comparing these results to the original HMI observation, the deviations result mostly from diffraction limited regions that are not related to more extended solar features (e.g., extended fibrils, coherent granulation pattern). The correct generation of fibrils in the penumbra, that are beyond the resolution limit of HMI observations, show that the neural network correctly learned to infer information from the high-quality image distribution. The clear structures in the granulation pattern can be interpreted as a result of the perceptual optimization (training cycle ABA).

Both instruments show a substantial overlap that allows for a quantitative assessment of our method. We select all observations that contain activity features in the months November and December, which were excluded from the training, and acquire the observations at the closest time instance from HMI (<10s difference; total of 62 observations). In most cases the specified coordinates require further adjustment to spatially align the images of the two instruments. We apply a matching by computing the cross-correlation between patches of HMI and Hinode/SOT observations and shifting the Hinode/SOT coordinate system to achieve a maximum cross-correlation. We apply this procedure over four resolution scales, with decreasing patch size. Quiet Sun and solar limb observations were excluded because of the difficulties in spatial alignment.

For the direct comparison we select subframes of $100'' \times 100''$ at the center of the Hinode/SOT observation. We linearly upsample the original HMI images to 0.15 arcsec pixels, to match the ITI images. For our evaluation we consider the distribution of pixel values and assume that the enhancement (e.g., contrast, scaling) leads to a distribution that matches the real Hinode/SOT observation. From each subframe we compute the histogram of the pixel distribution and normalize it for zero mean and unit variance, to exclude scaling biases between the instruments. We compute the area intersection of the Hinode/SOT histogram with the ITI and HMI histogram and average over the 62 matched observations. The HMI histograms have a mean intersection of 73.9% with the Hinode/SOT histograms. The enhanced ITI images provide an intersection of 90.8%. When considering the unnormalized histograms, the intersection between HMI and Hinode/SOT evaluates to 7.0%, due to the different scaling. The instrument-to-instrument translation provides a mean intersection of 64.1% for the unnormalized histograms, which shows that the translation additionally adjusts the HMI observations to the scale of Hinode/SOT observations. Samples of the aligned data set and the corresponding histograms can be
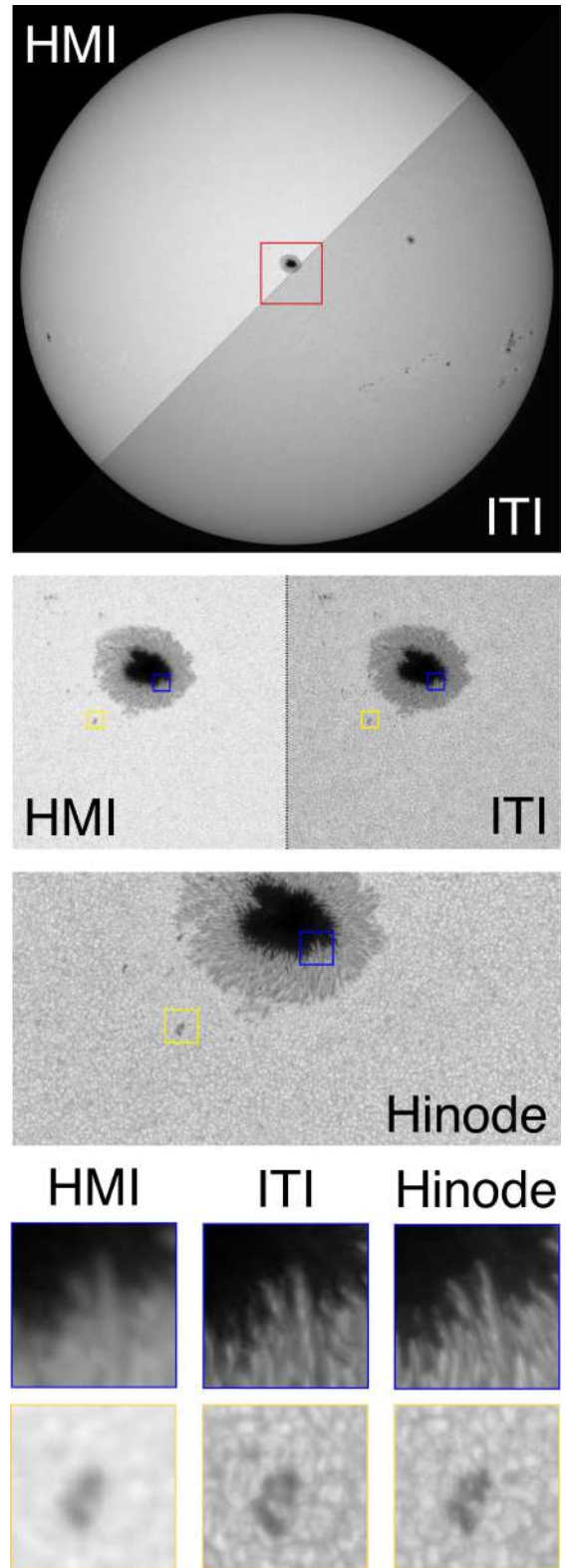

Fig. 2. Simultaneous observations of SDO/HMI and Hinode/SOT continuum from 2013-11-18 17:46:20 and 2013-11-18 17:46:28, respectively. a) Original HMI (top left) and ITI enhanced (bottom right) full-disk observation. b) Comparison of a $222'' \times 222''$ region. c) Corresponding Hinode observation with full field-of-view and resolution. d) Matched features of the original HMI, ITI enhanced and real Hinode observation with $13'' \times 13''$ spatial extend. The penumbral features of ITI match the real observation. The small scale separation of the individual fibrils is beyond the resolution limit of HMI, but are resolved in the ITI observation. The coarse shape of the granulation pattern and the solar pore match the Hinode observation, while smaller intergranular lanes show deviations.
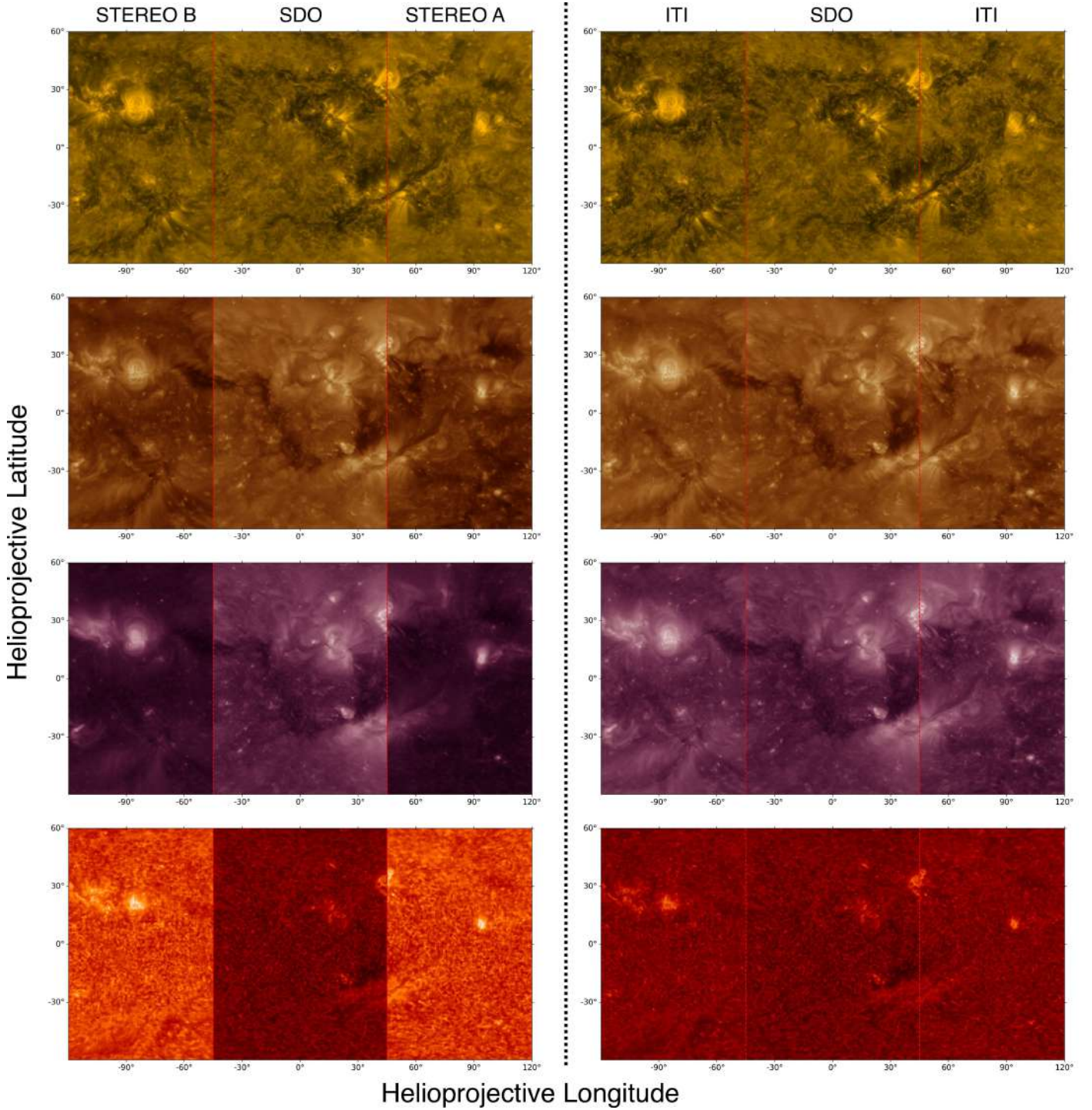
Fig. 3. Composite observations of STEREO-B/EUVI, SDO/AIA and STEREO-A/EUVI from 2010-12-12, where the telescopes are positioned at a separation angle of ±85 degree. From top to bottom we show the channels 131 Å, 195/193 Å, 284/211 Å and 304 Å. The left column shows the simple combination of the data from the three instruments. The right column shows the synchronic maps resulting from the ITI preprocessing. All maps are scaled by the minimum and maximum photon count of the combined map. The simple combination leads to a sharp transitions and shows the different calibration of the instruments. The ITI maps show a smooth transition with the original SDO/AIA observations.

found in App. D.1 and movies of full-cadence ITI enhanced observations are provided online (**Movie 1 and 2**).

2.2. Re-calibration of multi-instrument data - SOHO/EIT- and STEREO/EUVI-to-SDO/AIA

With the translation between image domains our method can account for both image enhancement and adjustment of the instru-

mental characteristics simultaneously (cf. Ignatov et al. 2018). We use ITI to enhance EUV images from the Solar and Heliospheric Observatory (SOHO; Domingo et al. 1995) and Solar Terrestrial Relations Observatory (STEREO; Kaiser et al. 2008) to SDO quality and calibrate the images into a unified series dating back to 1996.

With a spatial sampling of 0.6 arcsec pixels, we consider the filtergrams of the Atmospheric Imaging Assembly (AIA; Lemen

et al. 2011) and the magnetograms of HMI onboard SDO as high-quality reference. We use the EUV filtergrams from the Exteme Ulraviolet Imager (EUVI; Wülser et al. 2004) onboard STEREO as low-quality data set. For the translation of SOHO observations we use filtergrams of the Extreme-ultraviolet Imaging Telescope (EIT; Delaboudiniere et al. 1995) in combination with the LOS magnetograms of the Michelson Doppler Imager (MDI; Scherrer et al. 1995). The SDO images provide a 2.7 times higher pixel resolution than the STEREO/EUVI images. We reduce the resolution of STEREO observations to 1024×1024 pixels, before using ITI to increase the resolution by a factor 4, to the full 4096×4096 pixels resolution of SDO (STEREO/EUVI-to-SDO/AIA). For observations from SOHO/EIT+MDI we use half of the SDO resolution as reference (SOHO/EIT+MDI-to-SDO/AIA+HMI). All observations are normalized to a fixed scale of the solar radius, to avoid variations due to the ecliptic orbits (App. B). We consider all matching filters (131 Å, 195/193 Å, 284/211 Å, 304 Å, LOS magnetogram) and translate the combined set of channels with ITI, to benefit from the inter-channel relation.

The STEREO mission provides stereoscopic observations of the Sun, with the satellites being positioned at different vantage points. This excludes a direct comparison of the ITI enhanced observations with the SDO/AIA images. Here, we assess the usability of ITI to merge observations from multiple vantage points. In Fig. 3, we show heliographic maps of EUV filtergrams from STEREO-A, STEREO-B and SDO at a separation angle of ±85 deg. We use the central ±45 deg of each observation for the composite image and prefer the SDO observations in case of an overlap. All images are normalized for exposure time (DN/s) and are scaled by the maximum and minimum value of the composite image. The left panel of Fig. 3 shows the simple combination of the three filtergrams, where the different calibration of photon counts between the instruments can be clearly seen. The heliographic maps that use ITI enhanced STEREO filtergrams are shown in the right panel. The adjusted calibration levels can be best seen for 195/193 Å, 284/211 Å and 304 Å, where ITI provides a more consistent transition between the three data sources.

This calibration effect results from a feature dependent adjustment, where the observational characteristics of the high-quality instrument are reproduced. In Fig. 4 we show two examples of this effect. The ITI enhanced observations show an improvement of sharpness, that can be best seen from the coronal loops that are observed at the resolution limit of STEREO, and appear better resolved in the ITI image. The comparison of the 284/211 Å and 304 Å channels shows that additional structures are obtained by ITI that can not be seen in the original observation. We associate this enhancement with an inference of information from the other wavelength channels, by projecting the joined set of channels to the high-quality domain.

The SOHO mission provides observations that partially overlap with observations from SDO. In Fig. 5b, we provide a side-by-side comparison between the original SOHO, enhanced ITI and reference SDO observations. The full-disk images are a composite of the ITI enhanced and SDO observations, which show the same calibration effect as the STEREO heliographic maps (Fig. 5a). The sub-frames show the four EUV filtergrams and the LOS magnetogram of different regions of interest. We show examples of a filament, a quiet Sun region, the solar limb and an active region. For all samples we note a strong resolution increase and high similarity to the real reference observations. The filament in the top row, is difficult to observe in the original SOHO observation, but can be clearly identified in the ITI image.
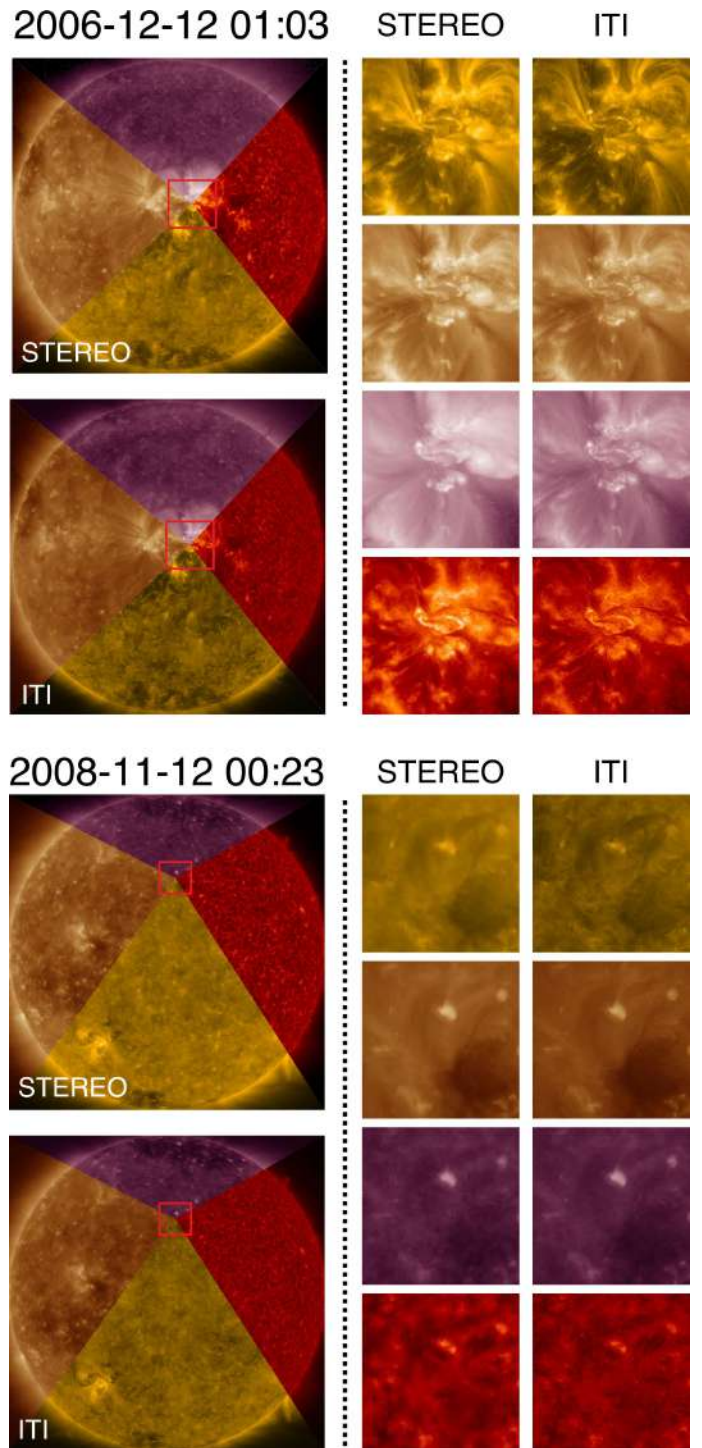

Fig. 4. Comparison of STEREO/EUVI and ITI enhanced EUV filtergrams. From top to bottom we show the 131 Å, 195/193 Å, 284/211 Å and 304 Å filtergrams. The sample at the top and bottom are taken during solar maximum and minimum, respectively. We note an overall improvement in sharpness by ITI, that can be best seen from the coronal loops of the active region (131 Å top). From the quiet Sun region (bottom) the enhancement of faint details can be seen, with the largest perceptual quality increase for the 304 Å channel.

For the frame at the solar-limb in the 284/211Å channel, we also note an inference from the multi-channel information that reconstructs the faint off-limb loop. From the quiet Sun region (second row) and the full-disk images, we can see that our method is consistent across the full solar disk. The observation of the

304 Å channel (active region; fourth row) shows the strongest improvement as compared to the original observation, but also shows that smaller features could not be fully reconstructed. We note that pixel errors can lead to wrong translations and can be accounted for prior to the translation (e.g., noise in SOHO/EIT 284 Å). The magnetic elements in the LOS magnetogram of ITI show a similar sharpness to the SDO/HMI magnetograms and are close to the actual observation. The magnetic flux elements at the lowest resolution level and the structure of the penumbra were not resolved by ITI. This shows the limit of the method, but also suggests that no artificial features are added by our model.

For the usage of long-term data sets the consistent calibration is of major importance (Boerner et al. 2014; Dos Santos et al. 2021). Datasets that comprise multi-instrument data require an adjustment into a uniform series (Hamada et al. 2020; Chatzistergos et al. 2019). We evaluate the model performance for long-term consistency over more than two solar cycles by computing the mean intensity per observation for each channel and instrument and comparing the resulting light-curves. For each instrument we sample 1000 equally distributed observations across the full mission lifetime (Fig. 5c). As can be seen from Fig. 5c, our model correctly scales the SOHO/EIT and STEREO/EUVI observations to the SDO intensity scale, with a good overlap that suggest also valid calibrations for pre-SDO times. The 171 Å channels show already a similar intensity scale. The 304 Å light curve shows the strongest deviations among the different instruments. The SOHO/EIT 304 Å observations shows a rapid change in calibration in 2011. This data was never considered for our model training and is therefore also not in agreement with the SDO/AIA light curve. Prior to this intensity jump, we note a good agreement between ITI and SDO/AIA 304 Å. The ITI enhanced STEREO/EUVI 304 Å show only a partial agreement. A possible reason for this could be the strong device degradation of the 304 Å channel (Boerner et al. 2014).

In the time between March 2010 and April 2011, magnetograms of both SOHO/MDI and SDO/HMI can be directly compared. For our evaluation we track the active region NOAA 11106 from 2010-09-13 to 2010-09-19, where we directly compare SOHO/MDI magnetograms and enhanced ITI magnetograms to the SDO/HMI magnetic field measure. For each observation we select a frame of 150×150 arcsec centered at the positive polarity of the active region, such that we can compare the small scale magnetic elements. At the top panel of Fig. 6a) we show the evolution of the mean unsigned magnetic flux of the selected regions (clipped at ± 1000 Gauss). All three curves show the same trend of magnetic flux evolution. The ITI magnetograms show less deviation from the reference SDO/HMI total magnetic flux evolution. The side-by-side comparison of magnetograms in Fig. 6b) shows that small elements that are not observed by SOHO/MDI are also not resolved by ITI, while details in extended elements are close to the SDO/HMI reference (e.g., negative polarities on 2010-09-14). In Fig. 6c) we use the SSIM and MAE to estimate the difference between SOHO/MDI and ITI magnetograms. The ITI enhancement leads throughout to a better magnetic field measure than the SOHO/MDI magnetograms, as compared to SDO/HMI. This supports the assumption that ITI can provide a physically valid enhancement. For the considered regions, we obtain an average deviation of 56 and 48 Gauss for SOHO/MDI and ITI, respectively. We note that this error partially also originates from the deviations in the spatial alignment, that affect both the original and enhanced images in the same way.

2.3. Mitigation of atmospheric effects - KSO Hα low-to-high quality

The mitigation of atmospheric effects in ground-based solar observations imposes two major challenges for enhancement methods. (1) Since observations are obtained from a single instrument, there exists no high-quality reference for a given degraded observation. (2) The diversity of degrading effects is large, which commonly leads to reconstruction methods that can only account for a subset of degrading effects.

We use ITI to overcome these challenges by translating between the image domains of clear observations (high-quality) and observations that suffer from atmospheric degradations (low-quality) of the same instrument. We use ground-based full-disk Hα filtergrams from Kanzelhöhe Observatory for Solar and Environmental Research (KSO; Pötzi et al. 2021) that we automatically separate into high- and low-quality observations (cf. Jarolim et al. 2020). For all observations we apply a center-to-limb variation (CLV) correction and resize the observations to 1024x1024 pixels (see App. B). In the resulting low-quality data set we include regular low-quality observations (e.g., clouds, seeing), while we exclude strong degradations (e.g., instrumental errors, strong cloud coverage).

In Fig. 7a we show two samples where a low-quality and high-quality observation were obtained within a 15 minutes time frame. We use these samples for a direct qualitative comparison of the ITI enhanced images (middle row Fig. 7a), which show a removal of clouds and are in agreement with the real high-quality observation.

We use the manually selected low-quality observation from Jarolim et al. (2020) to provide a quantitative evaluation of the quality improvement. For our evaluation we only consider the last two months of each year (total of 364 observations), to avoid a potential overlap with the training data. We adapt the image quality metric by Jarolim et al. (2020) for CLV corrected observations and only consider high-quality observations for model training, such that the quality metric estimates the deviation from the preprocessed synoptic KSO observations. In Fig. 7a we show the quality distribution of low-quality KSO observations and the ITI enhanced observations. The distributions show that ITI achieves a general quality increase, where the mean quality improves from 0.27 to 0.21, for the considered data set. In Fig. 7c we show three samples where the quality value intersect with the low-quality distribution after the ITI enhancement. The samples show that dense clouds and the contrail can be strongly reduced but not fully removed, which leads to the reduced image quality.

2.4. Restoration of long time-series - KSO Hα Film-to-CCD

Instrumental upgrades inevitably lead to significant difference in data series and require adjustment in order to use combined data series (Chatzistergos et al. 2019). We use ITI to restore Hα observations of scanned photographic film in the quality of the most recent CCD observations at KSO (Pötzi 2007). The film scans range from 1973 to 2000 and show degradations due to unequal illumination, atmospheric effects, low spatial resolution and digitalization. Figure 8a) shows samples across the full observing period. Similar to Sect. 2.3 we apply a center-to-limb correction, but use a resolution of 512×512 pixels. This is inline with our objective of homogenizing the data series, where we primarily adjust the global appearance, while we neglect small scale details. We apply ITI to correct degradations and match the observations with the high-quality KSO data series.
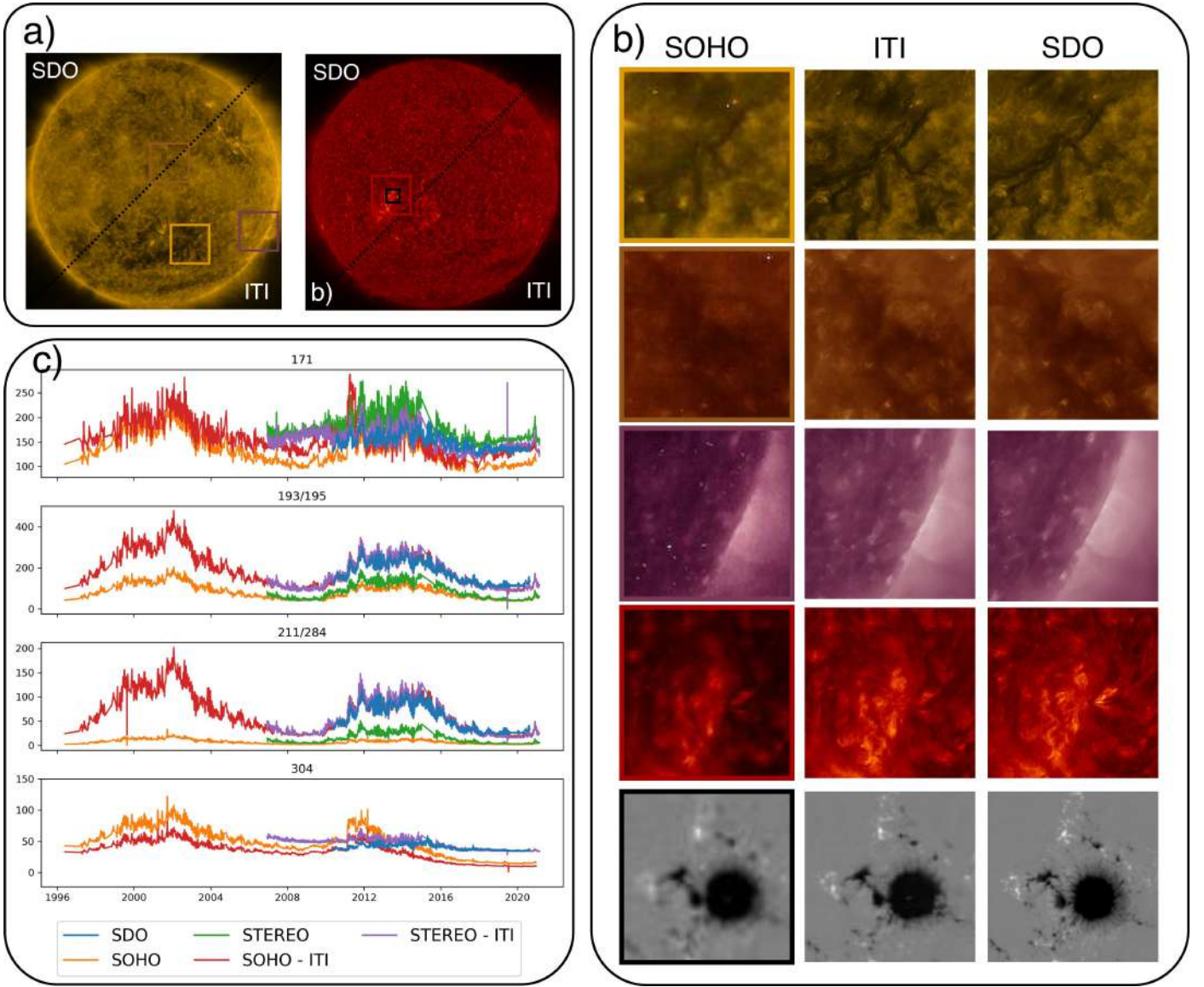
Fig. 5. ITI translation for the homogenization of SOHO with SDO observations and calibration of EUV data series. a) Two examples of SOHO-to-SDO translation from 2010-05-13 07:00 (left) and 2010-08-02 01:00 (right). The boxes indicate the cutouts in panel b) where we compare ITI images with aligned SDO/AIA filtergrams. We compare $300'' \times 300''$ and $100'' \times 100''$ cutouts of EUV filtergrams and LOS magnetogramsm, respectively. We directly compare observations taken at the same time from SOHO (left), ITI (center) and SDO (right). 131 Å: The observed feature is difficult to identify in the SOHO observation, while the ITI enhanced version resolves a clear filament structure that is consistent with the SDO observation. 195/193 Å: details in the quiet Sun region are completely blurred in the SOHO image. The obtained features by ITI are consistent on a global scale with SDO, but more deviations occur at the smallest resolution scales. 284/211 Å: ITI recovers faint of limb loops that are not resolved by SOHO. The pixel-noise at the bottom left is mitigated, but results in spurious features. 304 Å: The active region shows a valid reconstruction from the strongly pixelated observation. Magnetogram: Small magnetic elements are better resolved and appear deconvolved in the ITI image. Although a strong perceptual increase is noticeable, the full quality of the SDO/HMI magnetograms can not be reached by ITI. c) Comparison of the SOHO/EIT, STEREO/EUVI, SDO/AIA and ITI EUV light-curves. The mean intensities for each channel are plotted against time in the individual panels. The solar cycle trend can be seen for each light-curve (c.f. SILSO World Data Center 1998-2021). For the 171 Å, 193/195 Å and 211/284 Å channel, ITI adjusts the individual observations to a similar scale (DN/pix/s). The 304 Å channel shows deviations that originate from device degradation. The STEREO/EUVI 304 Å light-curve exceeds the scale.

In Fig. 8b) we show the refurbished observations and a KSO observation from 2013, taken with the recent instrument. The ITI images appear similar to the recent KSO observation, leading to a more homogeneous observation series. We note that ITI also accounts for atmospheric degradations (observation from 1980-04-29) and that enhanced images appear similar to the KSO observations in Fig. 7a. More detailed comparisons are given in the supplementary materials (App. D.3).

2.5. Approximation of observable - STEREO/EUVI-to-SDO/HMI magnetograms

In case of missing observables, a reasonable approach is to estimate the missing information based on proxy data. The STEREO mission observes the Sun in four EUV channels, but has no magnetic field imager onboard. We use ITI to complement the STEREO observations, by generating LOS magnetograms based on STEREO EUV filtergrams. Solar far-side magnetograms with
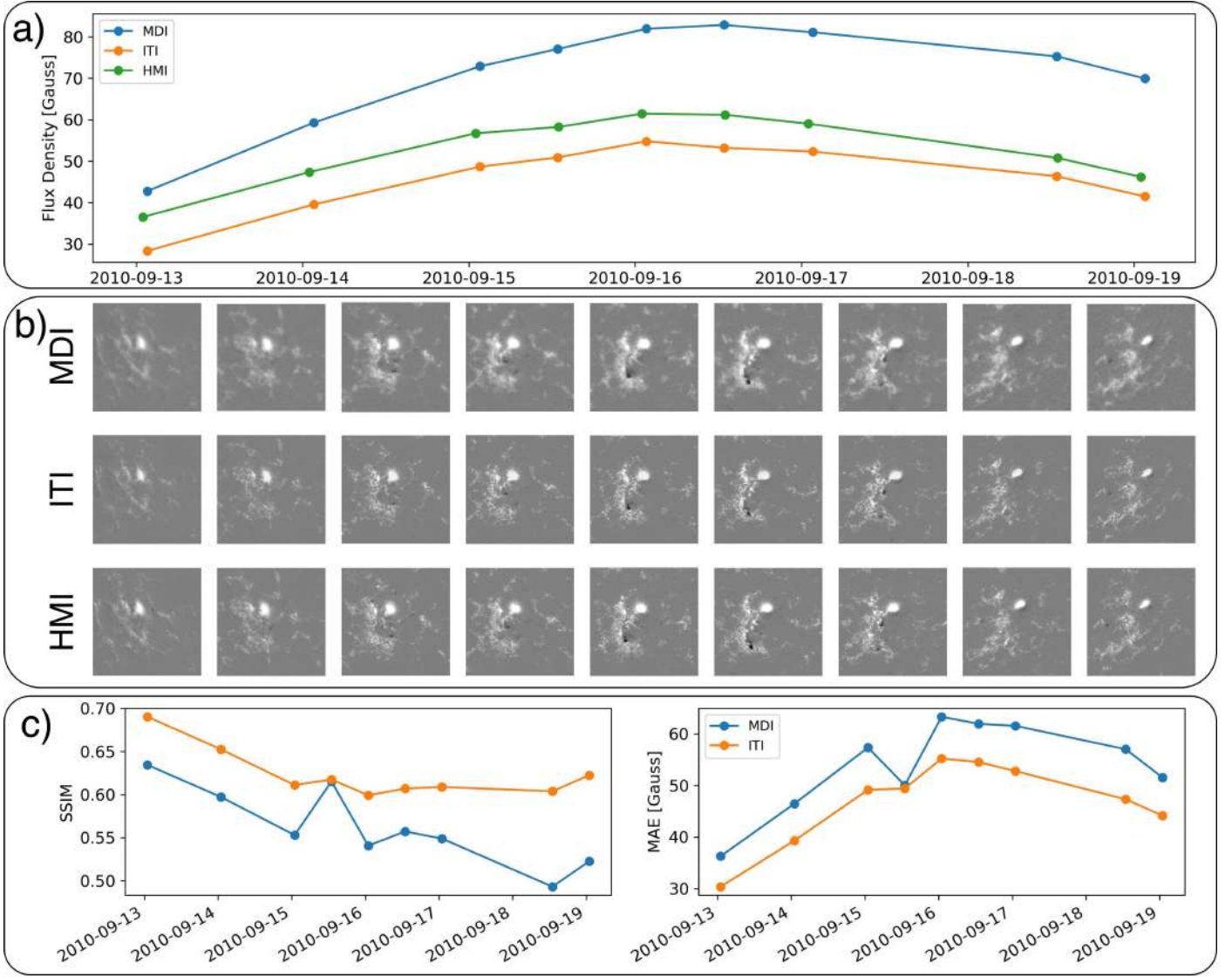
Fig. 6. Comparison of simultaneous LOS magnetograms from SOHO/MDI, SDO/HMI and ITI. The active region 11106 is tracked over seven consecutive days. a) Evolution of the mean unsigned magnetic flux. The ITI translation reduces the observed magnetic flux from SOHO/MDI, resulting in a better agreement with the high-resolution SDO/HMI observations. b) Direct comparison of the individual LOS magnetograms, scaled between [-1000, 1000] Gauss. ITI images show a similar appearance and sharpness to the SDO/HMI observations, while being also consistent with the low-resolution SOHO/MDI observations. c) SSIM and MAE of SOHO/MDI and ITI as compared to the SDO/HMI reference.

deep learning were first obtained by Kim et al. (2019), by training a neural network for SDO data and applying it to STEREO observations. In this study, we directly translate STEREO EUV filtergrams to the SDO domain. Similar to Sect. 2.2 we use the EUV filtergrams of STEREO and SDO as low- and high-quality domain respectively. In addition, we add the SDO LOS magnetograms to the high-quality domain. Thus, the generator AB translates sets of images with four channels to sets of images with five channels. For the magnetograms we use the unsigned magnetic flux, in order to prevent unphysical assumptions about the magnetic field polarity, which would require a global understanding of the solar magnetic field.

For our training setup we use a discriminator model for each channel and single discriminator for the combined set of channels, this provides three optimization steps that enable the estimation of realistic magnetograms. (1) The single channel discriminator for the magnetograms assures that the generated magnetograms correspond to the domain of real magnetograms, independent of their content. (2) The content of the image is bound by the multi channel discriminator, that verifies that the generated magnetograms are consistent with EUV filtergrams of the SDO domain. (3) From the identity transformation B-B, we truncate the magnetogram from the SDO input data and enforce the reconstruction of the same magnetogram (see Sect. 4.1).

The different vantage points of STEREO and SDO do not allow for a direct comparison of estimated ITI and real SDO/HMI magnetograms, but partially overlapping observations were obtained by SOHO/MDI at the beginning of the STEREO mission (November 2011). Here, we use these SOHO/MDI magnetograms as reference to assess the validity of the estimated ITI magnetograms. Figure 9 shows two examples of STEREO/EUVI 304 Å observations, the corresponding ITI magnetograms and the real SOHO/MDI reference magnetograms. We note a good agreement of the sunspot positions, that are not obvious from the EUV filtergrams. The magnetic elements appear similar in their distribution and magnitude, although we note a slight overestimation of magnetic field strength. The two examples show
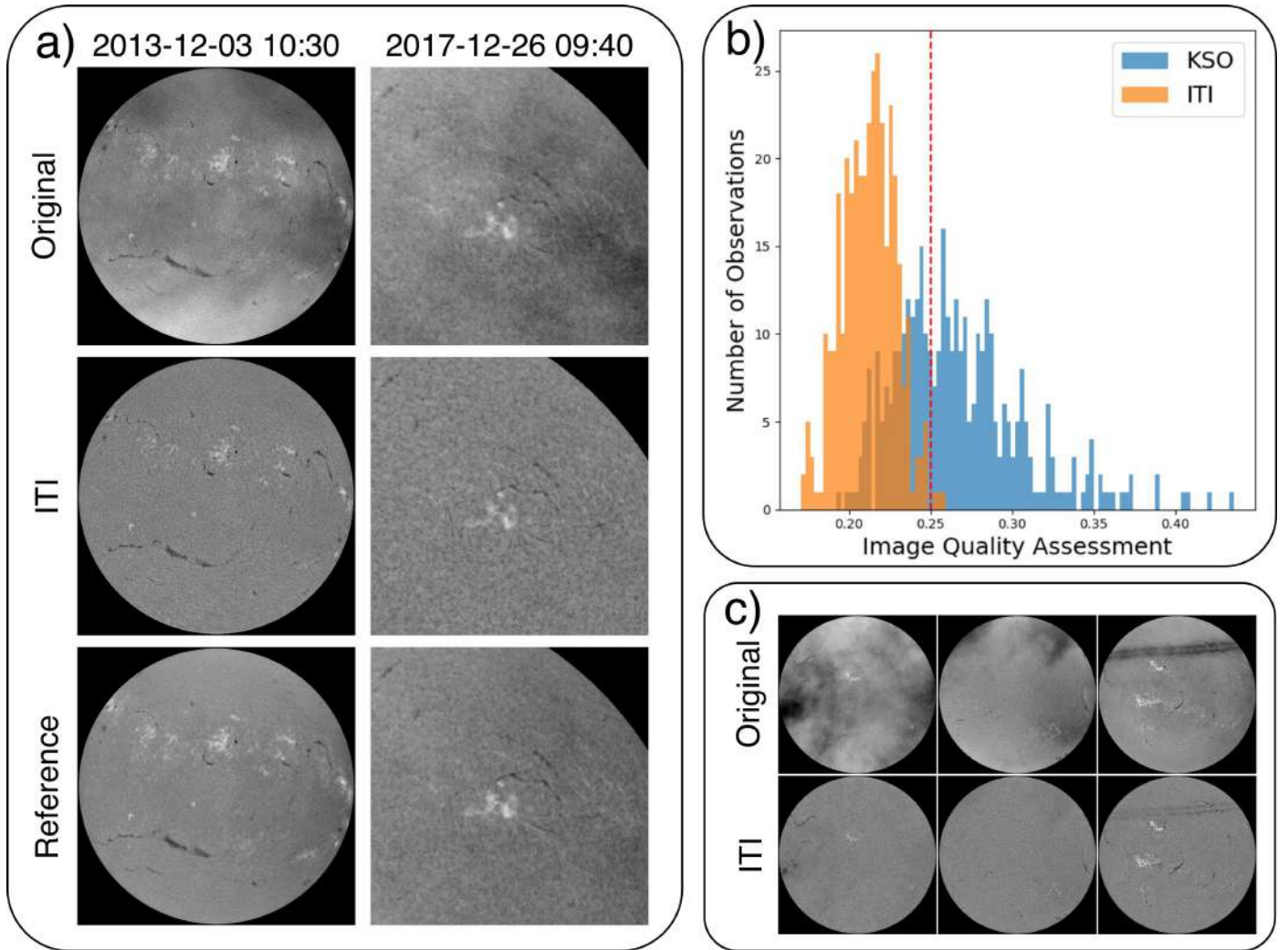
Fig. 7. ITI translation for the mitigation of atmospheric degradations. a) Comparison of the ITI mitigation of atmospheric effects during varying observing conditions. From top to bottom we show low-quality KSO observations, the ITI reconstructed observations and reference high-quality observations that were taken minutes after the low-quality observation. The left column shows a full-disk observation with clouds, that are strongly reduced by ITI. The right row shows a comparison of a cutout, where clouds are present in the low-quality KSO observation. The ITI reconstruction leads to a clear observation, where small chromospheric features remain unchanged and appear sharper. b) Estimated image quality distribution of the original low-quality KSO observations (blue) and the ITI enhanced observations (orange). The red dashed line indicates the 0.25 quality threshold. c) Three samples that exceed the quality threshold after the ITI enhancement.

the variations of the generated magnetograms, where the magnetogram in panel a) is close to the actual observation, while the magnetogram in panel b) shows more deviations. The biggest differences originate from the overestimation of magnetic field strengths, which especially affects regions with strong magnetic flux, where extended magnetic elements can be falsely contracted to a sunspot (i.e., spurious sunspot in panel b). An analysis of the temporal stability and the accompanying movie are given in the supplementary material (App. D.4; **Movie 3**)

A direct application of the far-side magnetograms are full-disk heliographic synchronic maps of magnetic flux. Analogously to Sect. 2.3 we combine the ITI magnetograms from STEREO A and B together with the real SDO/HMI magnetograms into heliographic maps. In panels a), b) and c) we show the magnetograms over 13 days, where the active regions (blue circle) are centered for STEREO B, SDO and STEREO A, respectively. In panels d), e) and f), we show the EUV filtergrams for comparison (c.f., Sect. 2.2). The overall distribution of magnetic flux and the position of the major sunspots is consistent with the SDO observation, although we note that again small

sunspots appear that are inconsistent with the SDO/HMI observation (panel a). Other magnetic features, such as the filament (green circle), can also be clearly identified in the ITI magnetograms. In all magnetograms the distinct separation between the two regions of strong magnetic flux is present, corresponding to the expected magnetic field topology.

3. Discussion

In this study we presented a method for the homogenization of data sets, stable enhancement of solar images and restoration of long-term image time series. We applied unpaired image-to-image translation to use observations of the most recent instruments to enhance observations of lower quality. Our method does not require any spatial or temporal overlaps and can account for the translation between instruments with different field-of-view, resolution and calibration, which makes our method applicable to many astrophysical imaging data sets. The approach is purely data driven, which avoids assumptions about the mapping between high- and low-quality samples and improves the appli-
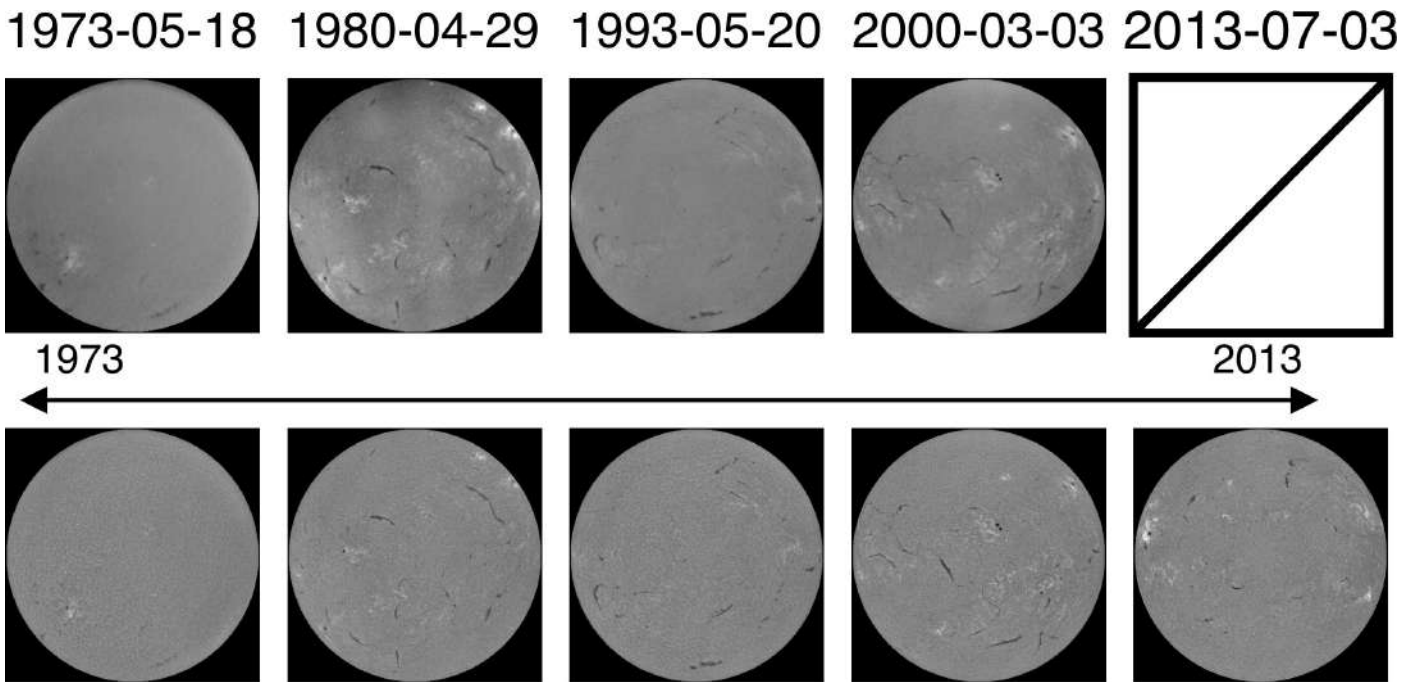
Fig. 8. Series of the observations on photographic film (top) and the restored ITI series (bottom). The observation on 2013-07-03 corresponds to a real high-quality KSO observation. The ITI images form a homogeneous series with the KSO archive past 2010.

cability to real observations. We demonstrated the flexibility of our method for five solar physics applications, that were selected to impose different challenges. All samples presented in Sect. 2, show a strong perceptual quality increase, reconstruction of faint details from a diverse set of degradations and are consistency across the full solar disk. Our method preserves measured intensity and flux values, such that the model results can be treated analogously to real observations.

We specifically addressed the question of the validity of the enhancement by comparing our results to real observations of the high-quality instruments. The enhanced images show throughout a strong similarity to the real observations and reveal a valid reconstruction of fibrils in the penumbra of sunspots (Fig. C.2), faint coronal loops (Fig. 5b) and blurred plage regions (Fig. 7a). This enhancement beyond the instrumental resolution, can be related to the inference of information from the high-quality image distribution. From the comparison of the enhanced fibrils (Sect. 2.1) we can see that the network takes the spatial context into account to resolve faint details. Deviations from real observations mostly occur at the smallest scales, where the enhancement is less bound by the low-quality observation.

The concept of translating between image domains has also implications for multi-channel environments. We showed that with the simultaneous translation of image channels, and requiring the consistency of the combined set of channels, information about the image content can be exchanged between related channels (SOHO-to-SDO; Fig. 5b) and missing channel information can be synthesized (far-side magnetograms; Sect. 2.5). The restoration of Hα film observations demonstrates that historic series can be homogenized with recent observation series, even if no overlapping observation periods exist (Sect. 2.4).

The quantitative evaluations demonstrate that our method achieves a stable performance across the full data sets. The analysis of pixel distributions, with a mean area intersection of 90.8% (SDO/HMI-to-Hinode continuum, Sect. 2.1), and quality assessment of ground-based observations (Sect. 2.3), suggest that ITI images are close to the high-quality image domain. The

analysis of the full-Sun EUV light-curves (Sect. 2.2) suggests that our method also provides a calibration of photon counts, leading to more consistent calibrated long-term series.

The demonstrated applications are of direct use in solar physics. 1) The results of the HMI-to-Hinode translation allow for a continuous monitoring of the Sun in a resolution of 0.15 arcsec pixels, producing full-disk images with unprecedented resolution. As can be seen from Fig. 2, the high-resolution observations of Hinode/SOT provide only a partial view on the objects of interest. The enhanced HMI images can provide useful context information by accompanying high-resolution observations. This applies to both the spatial extent as well as the extent of the time series by additional observations before and after the Hinode/SOT series. 2) The homogeneous series of space-based EUV observations enables the joined use of the three satellite missions. The data set provides a resource for the study of solar cycle variations (cf. Chatzistergos et al. 2019; Hamada et al. 2020), contributes additional samples for data-driven methods and enables the application of automated methods that were developed specifically for SDO/AIA data to the full EUV data series without further adjustments (Jarolim et al. 2021; Armstrong & Fletcher 2019). 3) The correction of atmospheric effects in ground-based observations can be operated in real-time (approximately 0.44 seconds per observation on a single GPU). This allows to obtain more consistent observations and assists methods that are sensitive to image degradations (e.g., flare detection; Veronig & Pötzi 2016). 4) With the reconstruction of the Hα film scans, we provide a homogeneous series starting in 1973, suitable for the study of solar cycle variations. 5) The generated STEREO far-side magnetograms give an estimate of the total magnetic flux distribution, which can provide a valuable input for space-weather awareness (Kim et al. 2019). Information about the magnetic polarities is required for the further application to global magnetic field extrapolations. From the patch-wise translation the inference of global magnetic field configurations (e.g., Hales law), would be arbitrary and was therefore omitted in this study. For all the considered
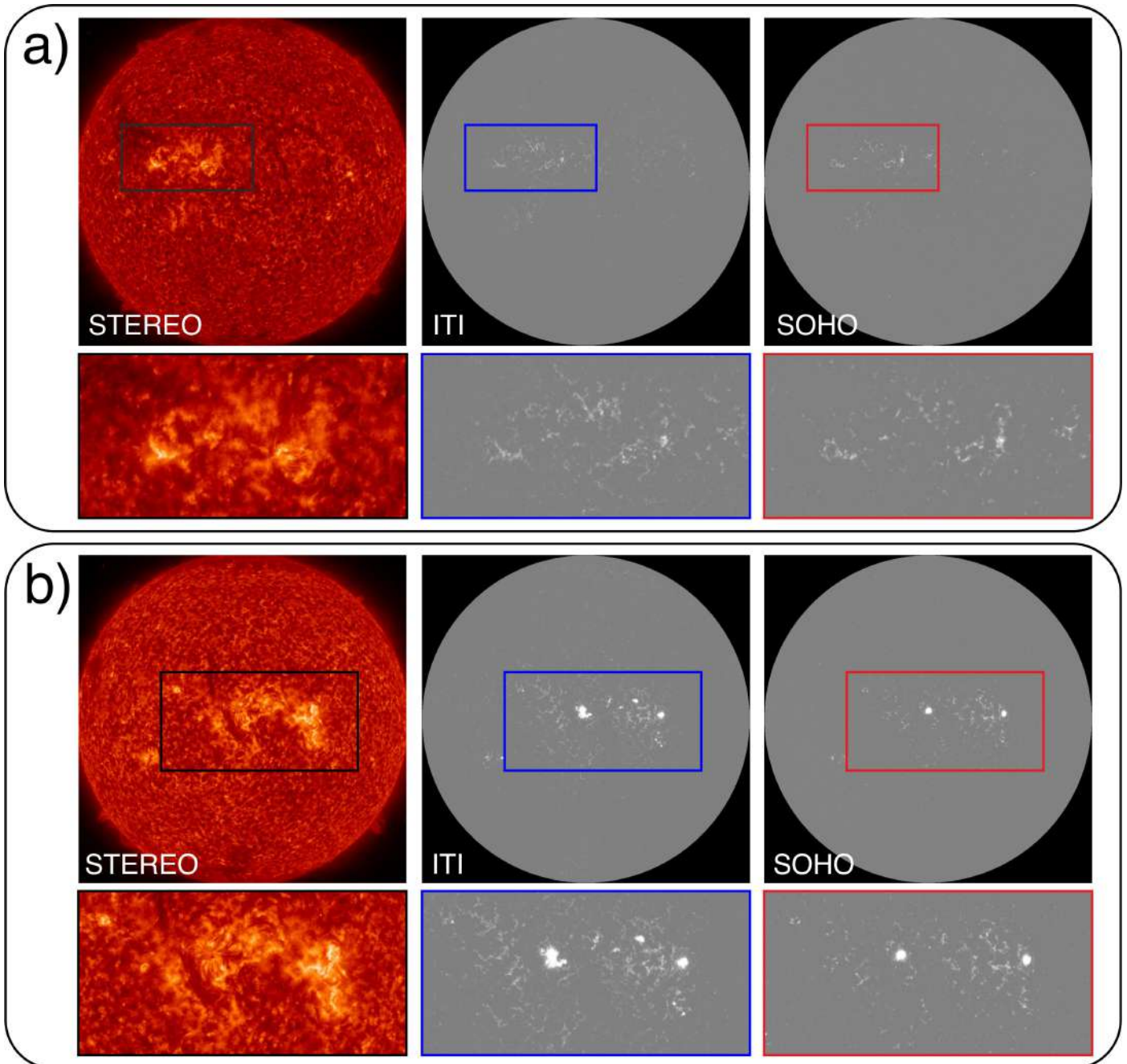
Fig. 9. Comparison between the synthetic ITI and overlapping SOHO/EIT magnetograms. In panel a and b we show observations from 2006-12-27 13:20 and 2007-01-07 19:20, respectively. STEREO/EUVI observations of the 304 Å channel are given for comparison and scaled by their maximum and minimum value. The LOS magnetograms show the absolute values of magnetic field strength and are scaled linearly between -1000 and 1000 Gauss. The flux distribution at the full-disk scale is in agreement between the ITI and SOHO/MDI observations. The comparison of the active region in a) shows that the ITI magnetogram have no major deviations and only deviate at the smallest scales. The active region in b) shows larger deviations that mainly originate from confined regions (i.e., sunspots) where ITI overestimates the magnetic field strength. An animated version is provided in the supplementary materials.

applications, our trained model processes images at higher rates than the cadence of the instruments, allowing the application in real-time and fast reconstruction of large data sets.

The extension to new data sets requires the acquisition of a few thousand images of low- and high-quality, where an alignment is not required, and the training of a new model. The translation is performed for observations of the same type (e.g., LOS magnetogram) or in the same (or similar) wavelength band with similar temperature response (e.g., EUV), but with different image quality, either reduced by atmospheric conditions or by in-

strumental characteristics, such as spatial resolution. The data sets need to provide similar observations in terms of features and regions, in order to avoid translation biases (Cohen et al. 2018). The estimation of magnetic field information based on EUV filtergrams illustrates an example where this condition is not strictly required. Here the image translation is constrained by the multi-channel context information and the learned high-quality image distribution.

We demonstrated that our neural network learns the characteristics of real high-quality observations, which provides an in-
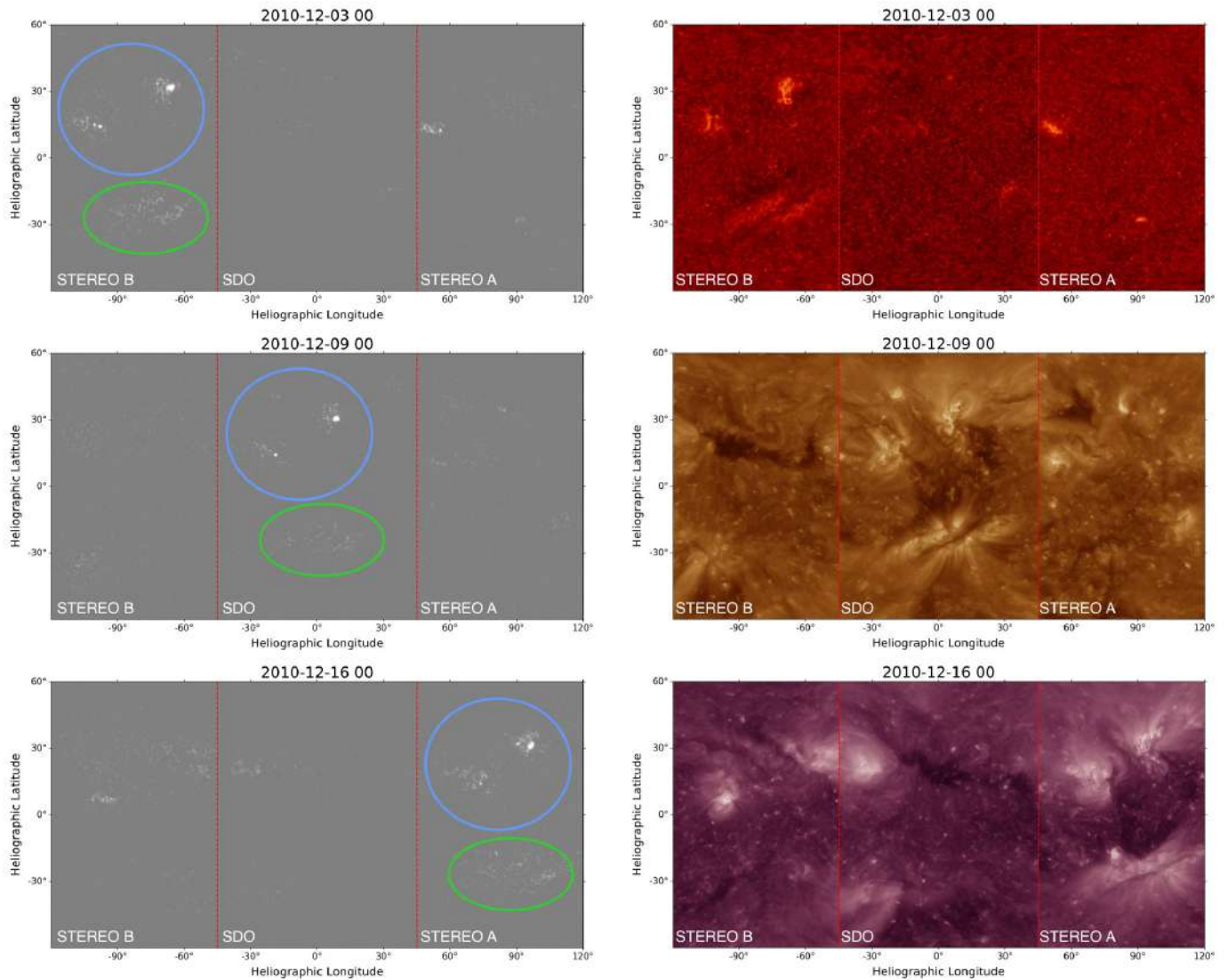
Fig. 10. Three consecutive observations of heliographic synchronic magnetograms (left) and reference EUV maps (right). We compare ITI magnetograms obtained from STEREO EUV filtergrams with real observations from SDO/HMI. We show the same active regions (blue circle) as observed by each instrument over 16 days. In the top row and bottom row the active regions are observed by STEREO B and STEREO A, respectively. The comparison to the real SDO/HMI observation (middle row) shows that ITI detects a similar magnetic flux distribution. Both ITI magnetograms show sunspot configurations with the characteristic tilt of preceding and following sunspot (Joy's law), but deviate from the SDO/HMI observation in terms of stronger and more confined magnetic fields (i.e., more sunspots). The green circle indicates the magnetic field configuration of the filament, that can be seen in the EUV observations. The ITI magnetograms show the expected topology of two opposing magnetic fields that sustain the filament channel and align with the SDO/HMI observation. An animated version of this figure is provided online (**Movie 4**).

formed image enhancement, where the considered high-quality data set implies a desired upper limit of quality increase.

4. Methods

Generative Adversarial Networks (GANs) have shown the ability to generate highly realistic synthetic images (Goodfellow et al. 2014; Radford et al. 2015; Karras et al. 2019). The training is performed in a competitive setup of two neural networks, the generator and discriminator. Given a set of images, the generator maps a random input from a prior distribution (latent space) into the considered image domain, while the discriminator is trained to distinguish between real images and synthetic images of the generator. The generator is trained to compete against the discriminator by producing images that are classified by the dis-

criminator as real images. The iterative step-wise optimization of both networks allows the generator to synthesize realistic images and the discriminator to identify deviations from the real image distribution (Goodfellow et al. 2014). By replacing the prior distribution with an input image, a conditional mapping between image domains can be achieved (Wang et al. 2018; Isola et al. 2017).

We use a GAN to generate highly realistic low-quality observations that show a large variation of degrading effects. We propose an informed image enhancement which uses domain specific knowledge to infer missing information. For this task, we employ a GAN to model the high-quality image distribution and constrain enhanced images to correspond to the same domain. With the use of a sufficiently large data set we expect that

we can learn to correctly model the true image distributions and find a mapping that is applicable for real observations.

4.1. Model training

The primary aim of our method is to transform images from a given low-quality domain to a target high-quality domain. We refer to the high-quality domain as B and to the low-quality domain as A. In order to achieve an image enhancement that can account for various image degradations we aim at synthesizing realistic images of domain A based on images of domain B. The pairs of high-quality and synthetic low-quality images are used to learn an image enhancement. Thus, the training process involves mappings from A to B (A-B), as well as mappings from B to A (B-A).

The model setup involves four neural networks, two generators and two discriminators. The generators learn a mapping between A-B and B-A. The discriminators are used to distinguish between synthetic and real images of domain A and B. The training cycle for image enhancement uses a high-quality image (B) as input, which is translated by the generator BA to domain A. The synthetic degraded image is then restored by the generator AB (Fig. 1). We optimize the generators to minimize the distance between the original and reconstructed image, as estimated by the reconstruction loss (cycle consistency). The simplest solution for this setting would be an identity mapping by both generators. We counteract this behavior with the use of discriminator A, which is trained to distinguish real images of domain A from synthetic images of generator BA. With this we constrain the generator BA to generate images that correspond to domain A and the generator AB to restore the original image from the degraded version. With the synthesis of more realistic low-quality images, we expect the generator AB to perform equally well for real low-quality images.

Similarly to Wang et al. (2018) each discriminator network is composed of three individual networks, where we decrease the image resolution by a factor 1, 2, 4 for the three networks, respectively. With this the perceptual quality optimization is performed at multiple resolution levels, thus estimating small-scale features as well as more global relations.

We follow the training setup of Zhu et al. (2017a) and use three additional training cycles. The second cycle enhances the perceptual quality of high-quality images with the use of discriminator B. The mapping is again performed under the cycle consistency criteria. Thus, we start with a low-quality image and perform an A-B mapping, followed by a B-A mapping (Fig. 11). The additional training with discriminator B ensures that images produced by generator AB correspond to domain B, which adds an additional constrain for image enhancement and improves the perceptual quality. The last two training cycles ensure a consistency under identity transformations. To this aim, we translate images of domain A with the generator BA and images of domain B with the generator AB, and then minimize the reconstruction loss between the original and transformed images (Fig. 12). For differences in resolution we use bilinear upsampling and average pooling for downsampling.

By only using an image as input to our generator, the results are deterministic (Almahairi et al. 2018). In the present case we are interested in modeling various degrading effects and explicitly want to avoid the generation of synthetic noise, based on image features (e.g., solar features, instrumental characteristics). This task is often addressed as multimodal translation (Zhu et al. 2017b; Huang et al. 2018) and also relates to style transfer between images (Huang & Belongie 2017; Karras et al. 2019; Choi
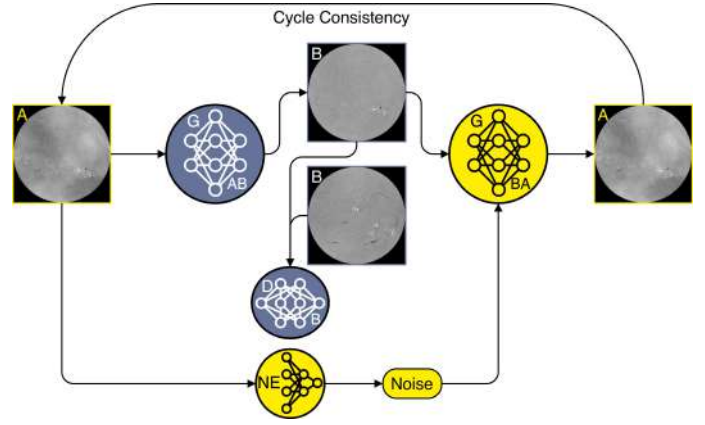

Fig. 11. Model training for image enhancement. Images of domain A are translated to domain B by generator AB (blue). The enhanced images are translated back to domain A by generator BA (yellow). The perceptual image quality is optimized with the use of discriminator B, which distinguishes between real images of domain B (bottom) and enhanced images by generator AB (top). The noise term for the reconstruction of the original images by generator BA is obtained by the noise-estimator (NE).
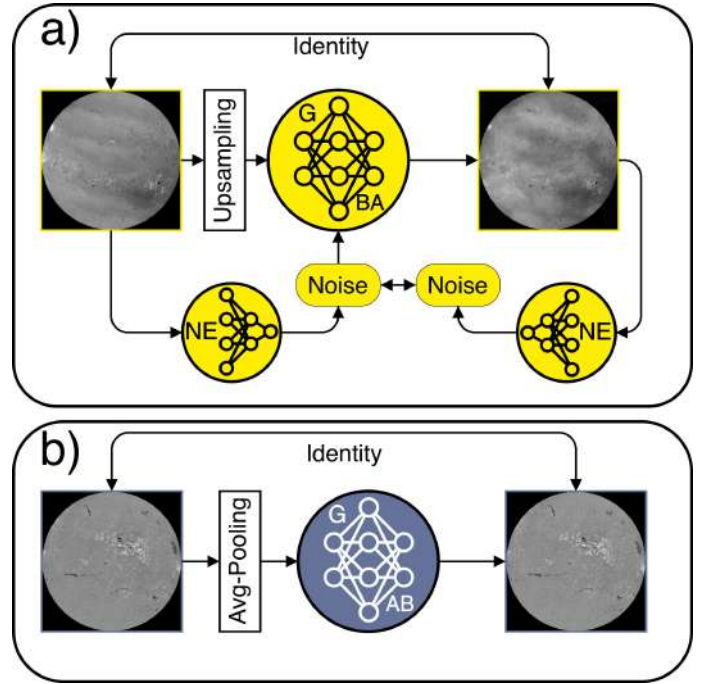

Fig. 12. Model training for the identity transformations. Images are optimized to remain unchanged for a mapping into the same domain. Top panel: mapping from domain A to domain A by generator BA (yellow). Bottom panel: mapping from domain B to domain B by generator AB (blue). The noise term for generator BA is obtained from the original image by the noise-estimator (NE). In addition the network is optimized to estimate the same noise before and after transformation (top panel). In case of different resolutions the low-quality images are first upsampled for A-A and downsampled for B-B.

et al. 2020). Here, we add an additional noise term to our generator BA, so that multiple degraded versions can be generated from a single high-quality image. For the generator AB we assume that there exists a unique high-quality solution for a given low-quality observation.

The cycle consistency of low-quality images is ensured by first estimating the noise of the original low-quality image. For

this task, we employ an additional neural network which we term noise-estimator. We use the noise-estimator for the A-B-A and A-A mapping and randomly sample a noise term for the B-A-B mapping from a uniform distribution [0, 1]. The advantage of this approach is two fold. 1) The mappings A-B-A and A-A are not ambiguous, which allows for a clear separation of noise from high-quality images. 2) The explicit encoding of low-quality features into the noise term representation benefits the relation between the generated low-quality features and the noise term, by enforcing the use of the noise term in the generator (Choi et al. 2020). For both the A-B-A and A-A mapping we minimize the distance between the estimated noise of the original image and the estimated noise of the reconstructed image (Fig. 12; Sect. 4.3). This approach relates to image style transfer (e.g. Johnson et al. 2016; Karras et al. 2019; Choi et al. 2020), where we consider the low-quality features as style that we transfer to the high-quality images.

Image quality can be addressed in terms of perceptual (e.g., mean-squared-error) and distortion (e.g., Wasserstein distance) quality metrics. The spanned space by both metrics is referred to as perception-distortion-plane, where the minimum on both metrics is not attainable (Blau & Michaeli 2018). In order to obtain the best quality, image enhancement algorithms should consider both metrics for optimization. This can be seen from image translation tasks, where the additional use of an adversarial loss achieved significant improvements in terms of image quality (Wang et al. 2018; Isola et al. 2017). We employ content loss, which shows a better correspondence to image features than pixel based metrics (e.g. MSE), as primary distortion metric and use adversarial loss of GANs for the perceptual optimization. From this setup we aim at achieving an optimal minimum in the perception-distortion plane.

Most full-disk observations exceed the computational limit when used in their full resolution. This can be overcome by reducing the resolution of the image or by training the neural network with image patches. In this study, we use image patches, in order to provide images with the highest resolution attainable. After model training, the neural network can be applied to full-resolution images, which is in most cases computationally feasible.

As described in Cohen et al. (2018), the use of data sets with an unequal distribution of features is likely to produce artifacts. For this reason we balance during training between image patches with solar features (e.g., active regions, filaments) and quiet Sun regions. During training we randomly sample patches from a biased distribution. For Hinode/SOT we additionally consider the solar limb, such that patches are equally sampled across the full solar disk.

4.2. Multi-channel architecture

In the case of multiple image channels (e.g., simultaneously recorded filtergrams at different wavelengths) we employ multiple discriminator networks (Fig. 13). We use an individual discriminator for each of the image channels (*single-channel*) and an additional discriminator that takes the combined channels as input (*multi-channel*). Here, each discriminator represents again a set of three single discriminators at three different resolution scales (cf., Sect. 4.1). From the usage of the single-channel discriminators we expect a better matching of the individual channel domains without influences of the other channels. The multi-channel discriminator is capable to assess the validity between the channels (e.g., appearance of features across the channels). This concept is especially important for the estimation of observ-

ables (Sect. 2.5), where the single-channel discriminator solely addresses the perceptual quality of the estimated observable, while the multi-channel discriminator restricts the estimated observable to be valid within the context of the other channels.

We note that the multi-channel architecture has no strict mapping requirements (e.g., same filters for each channel), such that the number of input and output channels is flexible. Thus, additional observables can be approximated (e.g., more output channels than input channels) and a conversion between similar channels can be achieved (e.g., SOHO/EIT 195 Å to SDO/AIA 193 Å).

4.3. Reconstruction loss

The cycle consistency serves as distortion metric for image enhancement. Pixel-based metrics (e.g., mean-absolute-error)

$$\mathcal{L}_{MAE,BAB} = \mathbb{E}[\|x_B - G_{AB}(G_{BA}(x_B, z))\|_1], \tag{1}$$

can prevent large divergences between the original image x_B and the reconstruction $G_{AB}(G_{BA}(x_B, z))$ but small shifts can cause a large increase of the reconstruction loss, which might provide a poor estimate of the image quality (Blau & Michaeli 2018). For this reason, we utilize in addition content loss, that compares feature similarity over pixel-wise similarity. Layer-wise activation of a neural network resemble extracted features and are therefore related to the content of the image. The content loss metric is computed by taking the mean-absolute-error (MAE) between the feature activation of the generated and original image. We define the content loss based on the discriminator, similar to Wang et al. (2018) as

$$\mathcal{L}_{Content,BAB,j} = \mathbb{E} \sum_{i=1}^{4} \frac{1}{N_i}[\|D_{B,j}^{(i)}(x_B) - D_{B,j}^{(i)}(G_{AB}(G_{BA}(x_B, z)))\|_1], \tag{2}$$

where x_B refers to a high-quality image; z to the noise term; $D_{B,j}^{(i)}$ to the activation layer i of network j from discriminator B; G_{AB} and G_{BA} to the generator AB and generator BA, respectively; and N_i to the total number of features of activation layer i. For each of our discriminators we use all intermediate activation layers (cf. Jarolim et al. 2020).

For consistency of the estimated noise terms, we minimize the MAE between the estimated noise term of the original low-quality image and its reconstructed version (A-B-A)

$$\mathcal{L}_{Noise,ABA} = \mathbb{E}[\|NE(x_A) - NE(G_{BA}(G_{AB}(x_A), NE(x_A)))\|_1], \tag{3}$$

and similarly for the identity mapping (A-A)

$$\mathcal{L}_{Noise,AA} = \mathbb{E}[\|NE(x_A) - NE(G_{BA}(x_A, NE(x_A)))\|_1], \tag{4}$$

where *NE* refers to the noise-estimator. To avoid the collapse of the noise term to a constant value, we minimize the MAE between the randomly sampled noise term and the estimated noise of the synthetic low-quality image (B-A-B)

$$\mathcal{L}_{Noise,BAB} = \mathbb{E}[\|z - NE(G_{BA}(x_B, z))\|_1]. \tag{5}$$

Equations 1 and 2 refer to the B-A-B cycle. The other training cycles are computed analogously, where we use the discriminator of the corresponding domain for the extraction of activation features (i.e., D_A for images of domain A). For the total
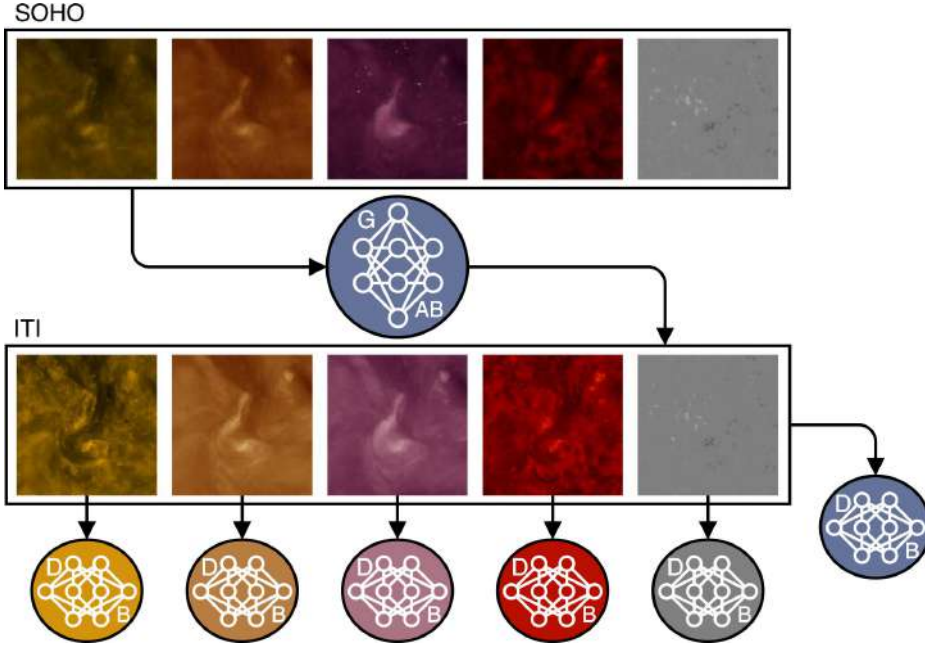
Fig. 13. Example of the multi-channel use of the discriminator. Each image channel of the source low-quality domain (top) is simultaneously translated to the target high-quality domain (bottom) by the generator AB (SOHO/EIT+MDI-to-SOHO/AIA+HMI). For each translated image channel a separate discriminator is used. The inter-channel consistency is assessed by a single discriminator that uses the combined set of channels as input.

reconstruction loss $\mathcal{L}_{Rec.}$ we combine the MAE loss and content loss

$$
\begin{aligned}
\mathcal{L}_{Rec.} = \; & \lambda_{mae} \cdot (\mathcal{L}_{MAE,BAB} + \mathcal{L}_{MAE,ABA}) + \\
& \lambda_{mae,id} \cdot (\mathcal{L}_{MAE,BB} + \mathcal{L}_{MAE,AA}) + \\
& \lambda_{content} \cdot \frac{1}{4M} \sum_{j=1}^{M} \left(\mathcal{L}_{Content,BAB,j} + \mathcal{L}_{Content,ABA,j} \right) + \\
& \lambda_{content,id} \cdot \frac{1}{4M} \sum_{j=1}^{M} \left(\mathcal{L}_{Content,BB,j} + \mathcal{L}_{Content,AA,j} \right) + \\
& \lambda_{noise} \cdot (\mathcal{L}_{Noise,ABA} + \mathcal{L}_{Noise,AA} + \mathcal{L}_{Noise,BAB}),
\end{aligned}
\tag{6}
$$

where M refers to the number of networks used for the discriminators and the λ parameters are used to scale the individual losses. The factor $4M$ is introduced to normalize the content loss for the number of discriminator layers per domain (cf. Eq. 2).

4.4. Adversarial loss

As originally proposed by Goodfellow et al. (2014), GANs are composed of a generating network (*generator*) that produces a synthetic image from a random input vector (*latent space*), and a discriminating network (*discriminator*) that distinguishes between generated and real images. The training is performed in a competitive setup between the generator and discriminator. We optimize the discriminator D and generator G for the objective proposed by Mao et al. (2017) (Least-Squares GAN). The discriminator A enforces the generation of low-quality images in the B-A-B cycle

$$
\mathcal{L}_{D_{A,j}} = \mathbb{E}\left[(D_{A,j}(x_A) - 1)^2 \right] + \mathbb{E}\left[D_{A,j}(G_{BA}(x_B, z))^2 \right],
\tag{7}
$$

where $D_{A,j}$ refers to the jth network of discriminator A, x_A to a real low-quality image, z to the noise term and $G_{BA}(x_B, z)$ to a generated low-quality image by generator BA. The generator is optimized to achieve a classification as real image by each discriminator network

$$
\mathcal{L}_{G,BA} = \mathbb{E}\left[\frac{1}{M} \sum_{j=1}^{M} (D_{A,j}(G_{BA}(x_B, z)) - 1)^2 \right].
\tag{8}
$$

The loss for discriminator B and generator AB is computed in the same way, where the discriminator enforces the generation of high-quality images in the A-B-A cycle.

4.5. Diversity loss

The variation of synthetic images is a frequently addressed topic for GANs (Gulrajani et al. 2017; Radford et al. 2015; Choi et al. 2020). For conditional image translation a variation of generated images can be obtained by inducing additional noise terms (Huang et al. 2018; Almahairi et al. 2018) or by including style information (Huang & Belongie 2017; Choi et al. 2020). Instance normalization achieved significant improvements in neural style transfer (Ulyanov et al. 2016). The normalization is applied after each convolution layer by normalizing each feature channel for zero mean and unit variance

$$
IN(x) = \gamma \frac{x - \mu(x)}{\sigma(x)} + \beta,
\tag{9}
$$

where $\mu(x)$ refers to the mean and $\sigma(x)$ to the variance across the spatial dimensions and γ and β to learnable parameters.

The effect of instance normalization can be interpreted as a normalization of the feature space in terms of style, where the affine parameters (γ and β) act as a projection to a different style (Huang & Belongie 2017). In the present case, the translation between different instruments can be understood as style transfer, where we transfer the high-quality style to images of low-quality. The affine parameters allow the network to learn a single style. For the training with image patches, we use running estimates of the normalization variables (μ, σ) with a momentum of 0.01 (statistical parameters are computed according to $x_{new} = (1 - 0.01) \cdot x_{old} + 0.01 \cdot x_{observed}$).

For the generator BA we enable the generation of various low-quality images from a single high-quality image by including a noise term, which we sample from a uniform distribution [0, 1]. A frequent problem of GANs is mode collapse, where the network generates the same image independent of the noise term (Radford et al. 2015; Gulrajani et al. 2017). We introduce an additional loss term to prevent mode collapse and increase the diversity of generated low-quality images (Yang et al. 2019a; Choi

Table 1. Data sets for training of instrument-to-instrument translation. The effective number of samples refers to the number of independent patches that can be extracted from the training set. Observations of Hinode/SOT have a varying field-of-views, the provided value gives the upper limit.

Instrument	Observables	Train	Effective Number of Samples	Test
SDO	171, 193, 211, 304, LOS magnetogram	3127	12508/50032	609
SOHO	171, 195, 284, 304, LOS magnetogram	5991	3.83e5	1379
STEREO	171, 195, 284, 304	3974	2.54e5	822
SDO/HMI	continuum	3003	30751	595
Hinode/SOT	continuum	3334	~6668	657
KSO synoptic	Hα	2679	42864	437
KSO LQ	Hα	1287	20592	364
KSO Film	Hα	4067	16268	772

et al. 2020; Mao et al. 2019). We sample two independent noise terms (z and $\hat{z}$) for the same high-quality image x_B and compute the content loss between the resulting two images

$$\mathcal{L}_{Content,Diversity,j}(x_B, z, \hat{z}) = \sum_{i=1}^{4} \frac{1}{N_i} [\|D_{A,j}^{(i)}(G_{BA}(x_B, z)) - D_{A,j}^{(i)}(G_{BA}(x_B, \hat{z}))\|_1], \quad (10)$$

where G_{BA} refers to the generator BA (cf. Eq. 2). We scale the difference in content loss by the distance of the noise terms and apply the logarithm to the result, which leads to an increased loss for nearly identical images and reduces divergences for large differences

$$\mathcal{L}_{Diversity} = \mathbb{E}\left[\log \frac{\frac{1}{4M}\sum_{j=1}^{M}\mathcal{L}_{Content,Diversity,j}(x_B,z,\hat{z})}{\|z-\hat{z}\|_1}\right]. \quad (11)$$

4.6. Combined loss

The training cycles for the generators and discriminators are performed end-to-end, where we alternate between generator and discriminator updates. Our full generator objective is given by combining Eq. 6, 8 and 11

$$\min_{G_{AB},G_{BA},NE}(\mathcal{L}_{Reconstruction} - \lambda_{divsersity}\cdot\mathcal{L}_{Diversity} + \lambda_{adversarial}\cdot(\mathcal{L}_{G,BA}+\mathcal{L}_{G,AB})), \quad (12)$$

where the λ parameters are used to scale the individual losses. The discriminator objective is obtained from Eq. 7:

$$\min_{D_A,D_B}\left(\frac{1}{M}\sum_{j=1}^{M}(\mathcal{L}_{D_{A,j}}+\mathcal{L}_{D_{B,j}})\right). \quad (13)$$

D_A and D_B refer to the combined discriminators for images of domain A and B, respectively ($D_A = \{D_{A,1}, ..., D_{A,M}\}$).

4.7. Data set

Table 1 summarizes the considered instruments, observing wavelengths, the amount of data samples and the number of independent patches that can be extracted. The images are normalized such that saturations are avoided, which we found beneficial for model training. Further details on data pre-processing, correction of device degradation and normalization is given in Appendix B. We apply a strict temporal separation of our train and test set, where we consider the last two months of each year only for model evaluations. In addition, we exclude the

SDO/AIA+HMI observations from 2010 that overlap with the SOHO/EIT+MDI observations.

The observations from SDO, STEREO and SOHO are taken at high cadence. For each data set we randomly sample observations from the full mission lifetime (SDO: 2010-2020, http://jsoc.stanford.edu; STEREO: 2006-2020, Solar Data Analysis Center via http://virtualsolar.org; SOHO: 1996-2011, STEREO Science Center via http://virtualsolar.org). For Hinode/SOT observations we use 2×2 binned red continuum observations at 6684 Å of the SOT Broadband Filter Imager (BFI) as high-quality reference. In the time between 2007 and 2016 we select a single observation from each observation series taken with BFI (https://darts.isas.jaxa.jp/solar/hinode). For the KSO Hα, we utilize the image quality assessment method by Jarolim et al. (2020), to automatically assemble a data set of ground-based KSO observations that suffer from atmospheric degradations (e.g., clouds, seeing), in the time between 2008 and 2019. As high-quality reference we use the synoptic archive of contrast enhanced KSO observations, that comprises manually selected observations of each observing day. In addition we perform a quality check and remove all observations of reduced image quality (http://cesar.kso.ac.at). The scanned KSO film Hα filtergrams include all observations taken in the time period from 1973 to 2000, independent of image quality (ftp://ftp.kso.ac.at/HaFilm/). We manually remove all observations that suffer from severe degradations (e.g., missing segments, strong overexposure, misalignment). Observations that suffer from minor degradations (e.g., spurious illumination, clouds, artifacts), are still considered for image enhancement.

With the use of unpaired data sets there is typically no limitation in data samples. For all our applications we select a few thousand images and extend them in case of diverging adversarial training.

4.8. Training Parameters

For our model training we use the Adam optimizer with a learning rate of 0.0001 and set $\beta_1 = 0.5$ and $\beta_2 = 0.9$ (Kingma & Ba 2014). As default value we set all λ parameters to 1. The content loss is scaled by 10 and the identity losses are scaled by 0.1 (s.t. $\lambda_{content} = 10$, $\lambda_{content,id} = 1$, $\lambda_{mae,id} = 0.1$). For all our models we track running statistics of the instance normalization layers (γ, β in Eq. 9) for the first 100 000 iterations and fix the values for the remaining iterations. For all space-based observations we set $\lambda_{diversity} = 0$, such that the main focus is the adaption of the instrumental characteristics and the generation of spurious artifacts is neglected (e.g., pixel errors). The training is performed

with batch size 1 and the size of image patches is chosen based on our computational capabilities.

In Table 2 we summarize the parameters of our model training. For each run we stop the training in case of convergence (no further improvement over 20 000 iterations). For the more complex low-quality features and limited number of samples, the generator BA can diverge to unrealistic low-quality images at later iterations. Here, we reduce the training iterations, where we note a stable training.

5. Code availability

The codes and trained models are publicly available: https://github.com/RobertJaro/InstrumentToInstrument

The software is designed as a general framework that can be used for automatic image translation and for training similar applications. Data preprocessing and download routines are provided.

6. Data availability

We publicly provide the enhanced and filtered KSO film observations with a resolution of 512×512 pixels. For the other presented data sets we provide lists of files used for training and evaluation. The provided models can be used to reproduce the data sets from our evaluation.

References

Almahairi, A., Rajeswar, S., Sordoni, A., Bachman, P., & Courville, A. 2018, arXiv preprint arXiv:1802.10151
Armstrong, J. A. & Fletcher, L. 2019, Sol. Phys., 294, 80
Asensio Ramos, A. & Olspert, N. 2021, A&A, 646, A100
Barnes, W. T., Bobra, M. G., Christe, S. D., et al. 2020a, The Astrophysical Journal, 890, 68
Barnes, W. T., Cheung, M. C. M., Padmanabhan, N., et al. 2020b, aiapy
Baso, C. D., de la Cruz Rodriguez, J., & Danilovic, S. 2019, Astronomy & Astrophysics, 629, A99
Blau, Y. & Michaeli, T. 2018, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6228–6237
Boerner, P., Testa, P., Warren, H., Weber, M., & Schrijver, C. 2014, Solar Physics, 289, 2377
Borman, S. & Stevenson, R. L. 1998, in 1998 Midwest symposium on circuits and systems (Cat. No. 98CB36268), IEEE, 374–378
Chatzistergos, T., Ermolli, I., Krivova, N. A., & Solanki, S. K. 2019, Astronomy & Astrophysics, 625, A69
Choi, Y., Uh, Y., Yoo, J., & Ha, J.-W. 2020, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8188–8197
Cohen, J. P., Luck, M., & Honari, S. 2018, in International conference on medical image computing and computer-assisted intervention, Springer, 529–536
Delaboudiniere, J.-P., Artzner, G., Brunaud, J., et al. 1995, in The SOHO Mission (Springer), 291–312
Díaz Baso, C. J. & Asensio Ramos, A. 2018, A&A, 614, A5
Diercke, A., Kuckein, C., Verma, M., & Denker, C. 2018, Astronomy & Astrophysics, 611, A64
Domingo, V., Fleck, B., & Poland, A. I. 1995, Solar Physics, 162, 1
Dos Santos, L. F. G., Bose, S., Salvatelli, V., et al. 2021, A&A, 648, A53
Galvez, R., Fouhey, D. F., Jin, M., et al. 2019, The Astrophysical Journal Supplement Series, 242, 7
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep learning (MIT press)
Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, in Advances in neural information processing systems, 2672–2680
Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. 2017, in Advances in neural information processing systems, 5767–5777
Hamada, A., Asikainen, T., & Mursula, K. 2020, Solar Physics, 295, 1
Hamada, A., Asikainen, T., & Mursula, K. 2021, Sol. Phys., 296, 40
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778
Herbel, J., Kacprzak, T., Amara, A., Refregier, A., & Lucchi, A. 2018, J. Cosmology Astropart. Phys., 2018, 054
Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. 2017, arXiv e-prints, arXiv:1706.08500
Howard, R. A., Moses, J., Vourlidas, A., et al. 2008, Space Science Reviews, 136, 67
Huang, X. & Belongie, S. 2017, in Proceedings of the IEEE International Conference on Computer Vision, 1501–1510
Huang, X., Liu, M.-Y., Belongie, S., & Kautz, J. 2018, in Proceedings of the European Conference on Computer Vision (ECCV), 172–189
Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., & Van Gool, L. 2018, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 691–700
Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. 2017, in Proceedings of the IEEE conference on computer vision and pattern recognition, 1125–1134
Jarolim, R., Veronig, A., Pötzi, W., & Podladchikova, T. 2020, Astronomy & Astrophysics, 643, A72
Jarolim, R., Veronig, A. M., Hofmeister, S., et al. 2021, A&A, 652, A13
Jia, P., Huang, Y., Cai, B., & Cai, D. 2019, ApJ, 881, L30
Johnson, J., Alahi, A., & Fei-Fei, L. 2016, in European conference on computer vision, Springer, 694–711
Kaiser, M. L., Kucera, T., Davila, J., et al. 2008, Space Science Reviews, 136, 5
Karras, T., Aila, T., Laine, S., & Lehtinen, J. 2017, arXiv preprint arXiv:1710.10196
Karras, T., Laine, S., Aittala, M., et al. 2019, arXiv e-prints, arXiv:1912.04958
Kim, T., Park, E., Lee, H., et al. 2019, Nature Astronomy, 3, 397
Kingma, D. P. & Ba, J. 2014, arXiv preprint arXiv:1412.6980
Lemen, J. R., Akin, D. J., Boerner, P. F., et al. 2011, in The solar dynamics observatory (Springer), 17–40
Mao, Q., Lee, H.-Y., Tseng, H.-Y., Ma, S., & Yang, M.-H. 2019, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1429–1437
Mao, X., Li, Q., Xie, H., et al. 2017, in Proceedings of the IEEE International Conference on Computer Vision, 2794–2802
Mumford, S. J., Christe, S., Freij, N., et al. 2020, SunPy
Otruba, W. & Pötzi, W. 2003, Hvar Observatory Bulletin, 27, 189
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, Solar Physics, 275, 3
Pötzi, W. 2007, Central European Astrophysical Bulletin, 31
Pötzi, W. 2008, Central European Astrophysical Bulletin, 32, 9
Pötzi, W., Veronig, A. M., Jarolim, R., et al. 2021, Solar Physics

Table 2. Overview of the Training parameters. For each instrument translation we train our model for the given number of iterations (Iter.). The upsampling parameter gives the number of upsampling operations that are applied to the low-quality images (Up). The diversity parameter is only considered for ground-based observations (λ_d). The patch size is given for the low-quality samples, such that the high-quality patch size corresponds to (Patch Size)·2^{Up}.

Instruments	Iter.	Up	λ_d	Patch Size [pixels]
SDO/HMI → Hinode continuum	500 000	2	0	160×160
SOHO/EIT+MDI → SDO/AIA+HMI	360 000	1	0	128×128
STEREO/EUVI → SDO/AIA	420 000	2	0	128×128
KSO LQ → KSO HQ Hα	360 000	0	1	256×256
KSO Film → KSO CCD Hα	240 000	0	1	256×256
STEREO/EUVI → SDO/AIA+HMI	360 000	0	0	256×256

Pötzi, W., Veronig, A. M., Riegler, G., et al. 2015, Solar physics, 290, 951

Radford, A., Metz, L., & Chintala, S. 2015, arXiv preprint arXiv:1511.06434

Rahman, S., Moon, Y.-J., Park, E., et al. 2020, The Astrophysical Journal Letters, 897, L32

Ramos, A. A., de la Cruz Rodríguez, J., & Yabar, A. P. 2018, Astronomy & Astrophysics, 620, A73

Rimmele, T. R. & Marino, J. 2011, Living Reviews in Solar Physics, 8, 2

Ronneberger, O., Fischer, P., & Brox, T. 2015, in International Conference on Medical image computing and computer-assisted intervention, Springer, 234–241

Schawinski, K., Zhang, C., Zhang, H., Fowler, L., & Santhanam, G. K. 2017, MNRAS, 467, L110

Scherrer, P. H., Bogart, R. S., Bush, R., et al. 1995, in The SOHO Mission (Springer), 129–188

Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, Solar Physics, 275, 229

Seitzer, M. 2020, pytorch-fid: FID Score for PyTorch, https://github.com/mseitzer/pytorch-fid, version 0.2.1

SILSO World Data Center. 1998-2021, International Sunspot Number Monthly Bulletin and online catalogue

Tsuneta, S., Ichimoto, K., Katsukawa, Y., et al. 2008, Solar Physics, 249, 167

Ulyanov, D., Vedaldi, A., & Lempitsky, V. 2016, arXiv preprint arXiv:1607.08022

Veronig, A. M. & Pötzi, W. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 504, Coimbra Solar Physics Meeting: Ground-based Solar Observations in the Space Instrumentation Era, ed. I. Dorotovic, C. E. Fischer, & M. Temmer, 247

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., et al. 2018, in Proceedings of the IEEE conference on computer vision and pattern recognition, 8798–8807

Wöger, F., von der Lühe, O., & Reardon, K. 2008, Astronomy & Astrophysics, 488, 375

Wülser, J.-P., Lemen, J. R., Tarbell, T. D., et al. 2004, in Telescopes and Instrumentation for Solar Astrophysics, Vol. 5171, International Society for Optics and Photonics, 111–122

Yang, D., Hong, S., Jang, Y., Zhao, T., & Lee, H. 2019a, arXiv preprint arXiv:1901.09024

Yang, W., Zhang, X., Tian, Y., et al. 2019b, IEEE Transactions on Multimedia, 21, 3106

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. 2017a, in Proceedings of the IEEE international conference on computer vision, 2223–2232

Zhu, J.-Y., Zhang, R., Pathak, D., et al. 2017b, in Advances in neural information processing systems, 465–476

Appendix A: Instruments and Data

This section describes the characteristics of the considered instruments for the evaluation of our method. The pair-wise correspondence of the data sets is discussed in Sect. 4.7.

The Solar Dynamics Observatory (SDO) is a space-based mission located in a circular geosynchronous orbit, that provides science data since 1 May 2010 (Pesnell et al. 2012). For our study we use EUV filtergrams from the Atmospheric Imaging Assembly (AIA; Lemen et al. 2011), and line-of-sight (LOS) magnetograms and continuum observations from the Helioseismic and Magnetic Imager (HMI; Schou et al. 2012). Both instruments provide solar full-disc images with a spatial sampling of 0.6 arcsec pixels and a spatial resolution of 1.5 arcsec. The data are recorded by 4096×4096 pixel CCDs. The AIA instrument operates at a cadence of 12 seconds and provides EUV filtgrams in seven band passes. In this study, we consider Fe IX (171 Å), Fe XII, XXIV (193 Å), Fe XIV (211 Å) and He II (304 Å) filtergrams. The emission lines are primarily formed in the solar corona and chromosphere and are associated with peak temperatures of ion formation at $6.3 \cdot 10^5$ K (Fe IX; quiet corona, upper transition region), $1.6 \cdot 10^6$ K, $2.0 \cdot 10^7$ K (Fe XII, XXIV; corona, hot flare plasma), $2.0 \cdot 10^6$ K (Fe XIV; active-region corona) and $5.0 \cdot 10^4$ K (He II; chromosphere, transition region) (Lemen et al. 2011). The HMI instrument provides maps of the photospheric magnetic field strength and orientation by observing polarizations of the Fe I absorption line (6173 Å). The continuum observations are calculated from six points in the Fe I line (Schou et al. 2012). For both the HMI magnetograms and continuum observations, we use the 720 seconds series.

The Solar and Heliospheric Observatory (SOHO) is a space-based mission, located at Lagrange point L1, that was launched in December 1995 (Domingo et al. 1995). In this study we use data from the Extreme-ultraviolet Imaging Telescope (EIT; Delaboudiniere et al. 1995) and LOS magnetograms from the Michelson Doppler Imager (MDI; Scherrer et al. 1995). Both instruments provide full-disc 1024×1024 pixels images of the Sun at a spatial sampling of 2.6 arcsec pixels and about 5 arcsec spatial resolution. From EIT we use Fe IX (171 Å), Fe XII (195 Å), Fe XV (284 Å) and He II (304 Å) filtergrams. The peak temperatures for ion formation are $1.3 \cdot 10^6$ K, $1.6 \cdot 10^6$ K, $2.0 \cdot 10^6$ K and $8.0 \cdot 10^4$ K, respectively (Delaboudiniere et al. 1995). The MDI instrument is the predecessor of SDO/HMI and derives the LOS magnetograms in the Ni I 6768 A absorption line (Scherrer et al. 1995).

The Solar Terrestrial Relations Observatory (STEREO) is a twin-satellite mission that operates two identical satellites on two orbits close to 1 AU. The orbits are selected to lead to a yearly separation of the spacecrafts by about 45 degree, as viewed from the Sun (Kaiser et al. 2008). The mission was launched in 2006 and provides stereoscopic observations of the Sun since then. In this study, we use the Exteme Ulraviolet Imager (EUVI; Wülser et al. 2004) of the Sun–Earth Connection Coronal and Heliospheric Investigation (SECCHI; Howard et al. 2008) instrument. The imager provides full-disc filtergrams with 2048×2048 pixels with a spatial sampling of about 1.6 arcsec pixels and 3.2 arcsec spatial resolution (Howard et al. 2008). Similarly to SOHO/EIT, filtergrams of Fe IX (171 Å), Fe XII (195 Å), Fe XV (284 Å) and He II (304 Å) are recorded. The associated peak temperatures are in the same range as for the SOHO/EIT filters (see above).

The Solar Optical Telescope (SOT) onboard the Hinode satellite (launched in 2006) is a 50 cm aperture telescope that provides high-resolution images of partial fields of the Sun (Tsuneta et al. 2008). The instrument comprises multiple broad- and narrow-band filters and provides spatial resolutions of up to 0.2 arcsec and a pixel scale of 0.054 arcsec pixels. In this study, we use continuum observations centered at a wavelength of 6684 Å of the Broadband Filter Imager (BFI). The BFI instrument provides a field-of-view of 218"×109", producing images with 4096×2048 pixels, and operates at a cadence <10s (Tsuneta et al. 2008). Observations are recorded on demand.

Kanzelhöhe Observatory for Solar and Environmental Research (KSO; https://kso.ac.at/) provides ground-based solar full-disk Hα filtergrams at a spatial resolution of 2 arcsec. The data are recorded by a 2048×2048 pixel CCD corresponding to a sampling of about 1 arcsec pixels. KSO regularly takes Hα images at a cadence of 6 seconds and provides a fully automated data reduction and data provision pipe line, which allows for data access in near real time (Otruba & Pötzi 2003; Pötzi et al. 2015; Pötzi et al. 2021). The current instrument setup is in operation since 2008. The Hα line is formed by absorption at 6563 Å in the solar chromosphere and by cooler plasma in the solar corona (filaments).

At KSO, regular monitoring of the Sun in the Hα spectral line is provided since 1973. The observations during 1973 and 2000 were recorded on photographic film at a cadence of about 4 minutes and were later digitized in 1024×1024 pixels format (Pötzi 2007, 2008).

Appendix B: Data pre-processing

Appendix B.1: Space-based observations

For each observation we center the Sun, normalize the field-of-view to 1.1 solar radii and crop the frame, such that the extent of the solar-disk is independent of yearly variations caused by the elliptic orbit (cf. Galvez et al. 2019). We correct for instrumental degradation and normalize the exposure time (Boerner et al. 2014). For STEREO/EUVI, SOHO/EIT, Hinode/SOT we use the IDL eit_prep, secchi_prep and fg_prep routines, respectively. For SDO/AIA we use the routine provided by aiapy (Barnes et al. 2020b). The maximum intensity value of each instrument channel is estimated from the maximum intensity values over the full data set by $\hat{I}_{max} = \text{mean}(I_{max}) + 0.5 \cdot \text{std}(I_{max})$, where I_{max} refers to the maximum intensity values of the images and std to the standard deviation. We clip negative values and values larger than the estimated maximum intensity and normalize to [0, 1]. We apply a asinh stretch for EUV filtergrams and scale the data afterwards to the interval [-1, 1], such that it corresponds to a tanh activation function

$$\hat{x} = \frac{\text{asinh}(x/a)}{\text{asinh}(1/a)}2 - 1, \qquad (B.1)$$

with $a = 0.005$. This stretching function provides for a logarithmic behavior for large values and a linear behavior for small values. For LOS magnetograms we use a linear normalization of values between [−1000, 1000] Gauss to [−1, 1] and set all off-limb values to zero.

For the continuum observations of SDO/HMI and Hinode we analogously estimate the maximum value and scale linearly between 0 and the maximum value.

Appendix B.2: Ground-based observations

For all KSO observations we apply the same preprocessing. We shift the Sun to the image center, crop the field-of-view to 1 so-
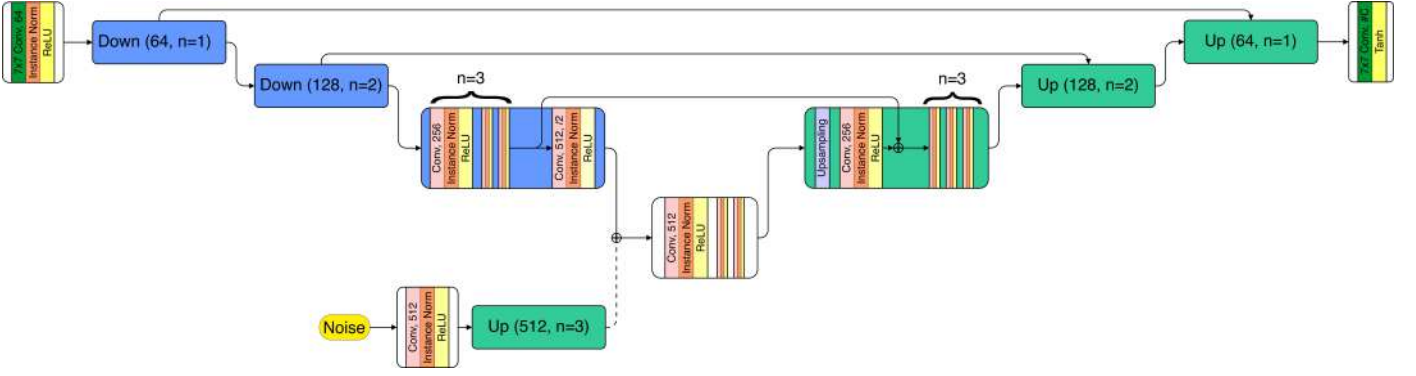
Fig. C.1. Overview of the generator architecture. Our generators are composed of an input block, three downsampling blocks, an intermediate block, three upsampling blocks and an output block. The input block transforms the image to 64 filter channels. Each downsampling block reduce the spatial resolution by 4 while increasing the number of filter channels by 2. The number of convolutional blocks is linearly increased. The upsampling blocks operate in reverse order. At the intermediate layer the noise term for generator BA is added by concatenation (dashed line). The noise term is matched to the dimensions of the intermediate layers by an initial convolution block and an upsampling block. Features from the downsampling blocks are concatenated to the upsampling block features at the same resolution. The output image is obtained by a convolution layer followed by a tanh activation.

lar radius and resize it to 512×512 pixels and 1024×1024 pixels for the Film-to-CCD and Low-to-High translation, respectively. The center-to-limb variation is corrected by plotting the theoretical correction $\mu = \cos\theta$ against the intensity values of the image, where θ refers to the heliocentric angle, and fitting a fourth order polynomial that gives the intensity correction $I_{corr}(\mu)$ (cf. Diercke et al. 2018). For KSO observations of the recent instrument (CCD), we scale values between 0.65 and 1.5 to $[-1, 1]$ with a asinh stretch ($a = 0.5$). For the KSO film observations we apply the same scaling but adjust the scaling to values between 0.39 and 1.94 to account for the variations of the degraded observations. We apply the preprocessing per frame and set all off-limb pixels to -1.

Appendix C: Model architecture

For image translation tasks the concept of multi-scale architectures has shown great success (Karras et al. 2017; Wang et al. 2018; Ronneberger et al. 2015). Skip connections have demonstrated the ability to enhance training performance and to preserve spatial details (He et al. 2016; Ronneberger et al. 2015).

For our generators we employ convolution blocks similar to Karras et al. (2017) and introduce skip connections at each resolution level, similar to Ronneberger et al. (2015). The full overview of our generator models is given in Fig. C.1. We use convolutional blocks composed of a convolutional layer with a kernel size of 3×3, followed by an instance normalization (Ulyanov et al. 2016) and a ReLU activation. In order to retain the image dimensions and to avoid boundary artifacts, we use reflection padding before the convolution layers (Wang et al. 2018). Our network is composed of three downsampling blocks, an intermediate block, followed by three upsampling blocks. A downsampling block consists of n convolutional blocks followed by a convolutional block with stride 2 (downsampling by 2). While downsampling we double the number of filter channels. For an upsampling block we first apply bilinear upsampling, reduce the number of filter channels by a factor of 2 with a convolutional block and apply n additional convolutional blocks. At each resolution level we use skip connections between the downsampling and upsampling blocks. Hereby the features prior to downsampling are concatenated with the features after the first convolutional block in the upsampling block. The input image is transformed by an initial convolutional block with a kernel size

of 7×7 and 64 filter channels. The three downsampling blocks are consecutively applied with $n = 1, 2, 3$. The intermediate layer consists of three convolution blocks with the same number of filters as the last downsampling block (512). The upsampling blocks are applied in the inverse order ($n = 3, 2, 1$). The output image is obtained by a convolution layer with kernel size 7×7, followed by a tanh activation function.

For the generator BA we include the noise term by concatenation with the features of the last downsampling block (prior to the intermediate block). The noise term is transformed by a convolutional block with 512 filter channels and matched to the dimensions of the last downsampling block by an upsampling block with $n = 3$ and 512 filter channels (dashed line in Fig. C.1).

Our networks are fully convolutional, therefore the training can be performed with image patches, while the evaluation is performed with full resolution images. For the noise term we use 16 channels and the spatial dimensions are adjusted to the considered image size (1/16 of the input image resolution). The instance normalization accounts for the global style transfer (see Sect. 4.5), while the noise term accounts for localized degradations in the image.

For the discriminators we use the architecture introduced in Wang et al. (2018), where each discriminator is composed of three individual networks that operate on different scales. The discriminators are composed of four consecutive stride 2 convolutions with instance normalization and ReLU activation. We start with 64 filter channels and consecutively increase them by a factor 2 for each layer. The discriminators output is obtained by a convolution layer with one filter channel. Therefore, each discriminator provides a grid of predictions instead of a single output (Isola et al. 2017). For the noise estimator we use the same architecture as for the discriminators, adjust the number of filters of the last layer to the noise dimensions (16) and apply a final sigmoid activation function.

For an image enhancement that involves a resolution increase, we extend the generator AB by additional upsampling blocks and the generator BA by additional downsampling blocks. In the case of images with multiple channels (e.g., multiple filtergrams, magnetograms), we translate all channels simultaneously. For each image channel we use an additional discriminator that ensures the correct representation of the generated channel, independent of the other image channels. The primary discriminator is adjusted to take all filter channels into account
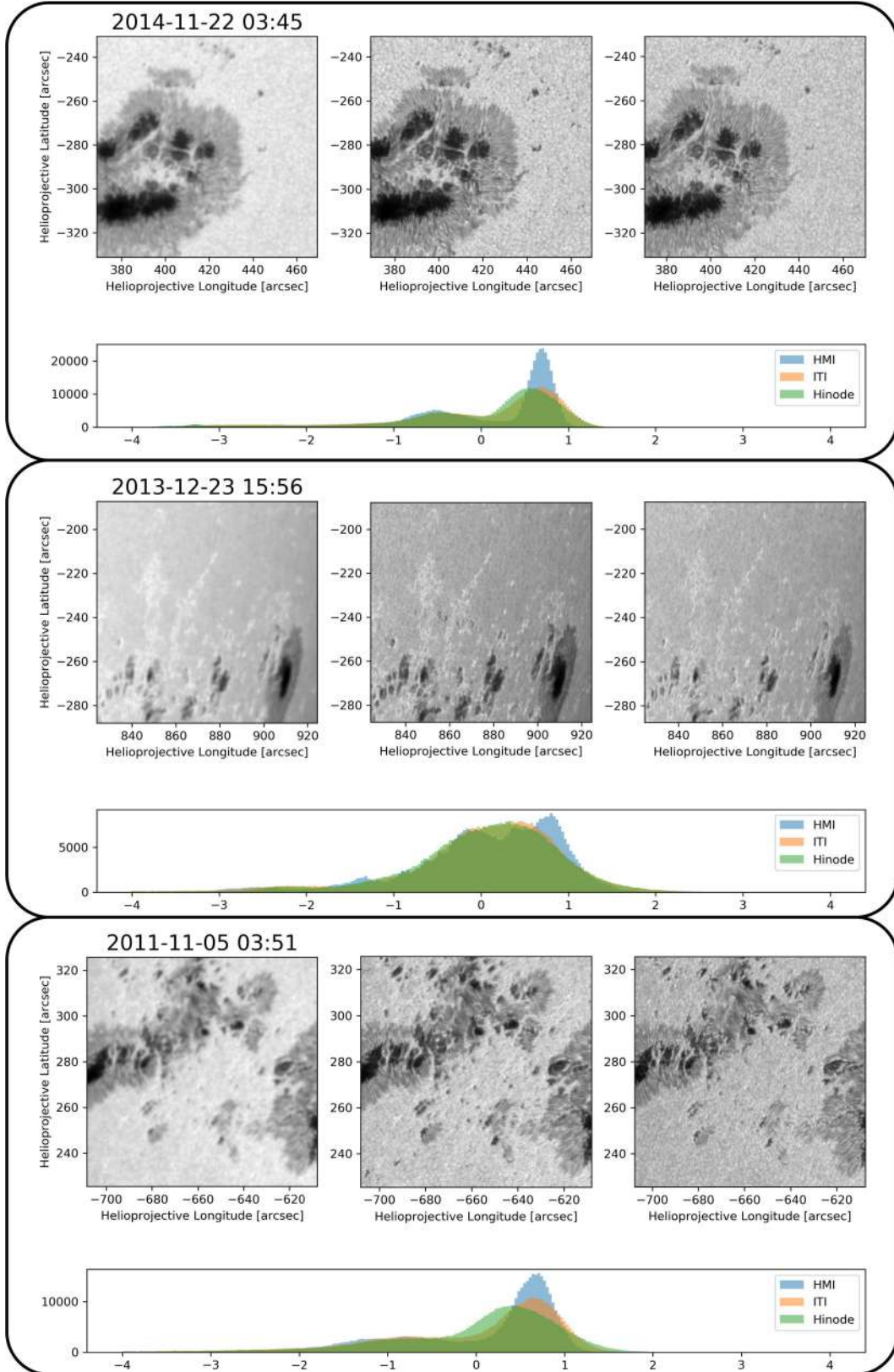
Fig. C.2. Three examples of the HMI-to-Hinode translation. We show the original SDO/HMI observation (left), the ITI translated image (middle), the overlapping Hinode/SOT observation (right) and the corresponding normalized histograms (bottom).

simultaneously to ensure the consistency between the channels.   For an unequal number of channels between domain A and B,
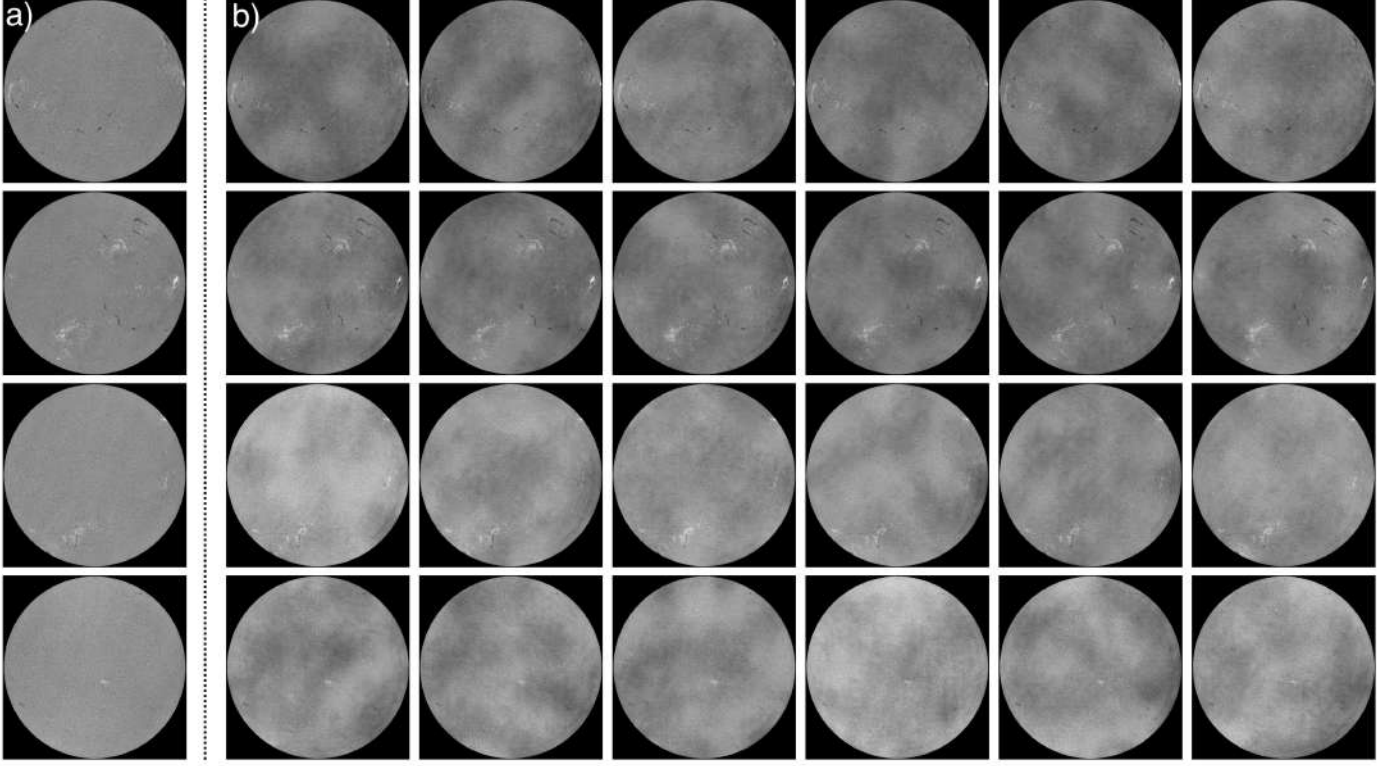
Fig. D.1. Diversity examples of the KSO low-to-high quality translation. a) real high-quality observations. b) artificially degraded observations. The images in one row are generated from the same high-quality input image. The artificial low-quality images show a variety of realistic atmospheric effects, while the solar features are mostly unchanged.

we adjust the identity cycles by truncating additional channels and extending missing channels with zeros.

Appendix D: Supplementary Evaluation

Appendix D.1: Aligned Hinode observations

In Fig. C.2 we give additional samples of the aligned SDO/HMI, ITI and Hinode/SOT continuum observations.

Appendix D.2: Variation of atmospheric degradations

By modeling a large variety of realistic degradation effects we expect that the reconstruction model can account for multiple atmospheric degradation effects simultaneously. From the included noise term in our generator BA, multiple low-quality observations can be synthesized from a single high-quality observation. We illustrate the diversity of the generated samples by randomly selecting a high-quality image and generating six corresponding low-quality images, where we use a different noise term for each example. Four examples are shown in Fig. D.1, where panel a) contains the high-quality input images and panel b) the synthetic low-quality images. The samples show a large spread of different cloud distributions, both in density and position. From a comparison with Fig. 7 it can be seen that the synthetic clouds appear similar to real low-quality observations. The solar features are consistent with the high-quality reference (e.g., active regions, filaments).

Appendix D.3: Detailed comparison of enhanced photographic film observations

A comparison between cutouts of the film scans and the ITI enhanced observations is shown in Fig. D.2. All samples show a clear improvement in image sharpness, that especially affects the plage regions and leads to more distinct representation of solar filaments. As can be seen from Fig. D.2b), the ITI translation also mitigates clouds, similar to the application in Sect. 2.3.

In Fig. D.3 we provide a comparison of a filtergram at three different scales. In Fig.D.3a) the calibration of the image intensity and correction of large scale inhomogeneities is visible. The cutouts show a comparison between the film observation and the restored image, where we note a perceptual quality improvement.

Appendix D.4: Temporal stability of generated ITI magnetograms

We further compare the temporal stability of our method by estimating the full-disk magnetograms from 2007-01-14 to 2007-01-18 and compare the sequence to the SOHO/MDI magnetograms (**Movie 3**). The ITI magnetograms are overall consistent and show no large artifacts (e.g., random active regions). The active region in the ITI magnetograms appears similar to the real observation and is consistent in its position on the solar disk. The largest inconsistencies originate from variations in the estimated sunspots and we note a quality decreases close to the solar limb.
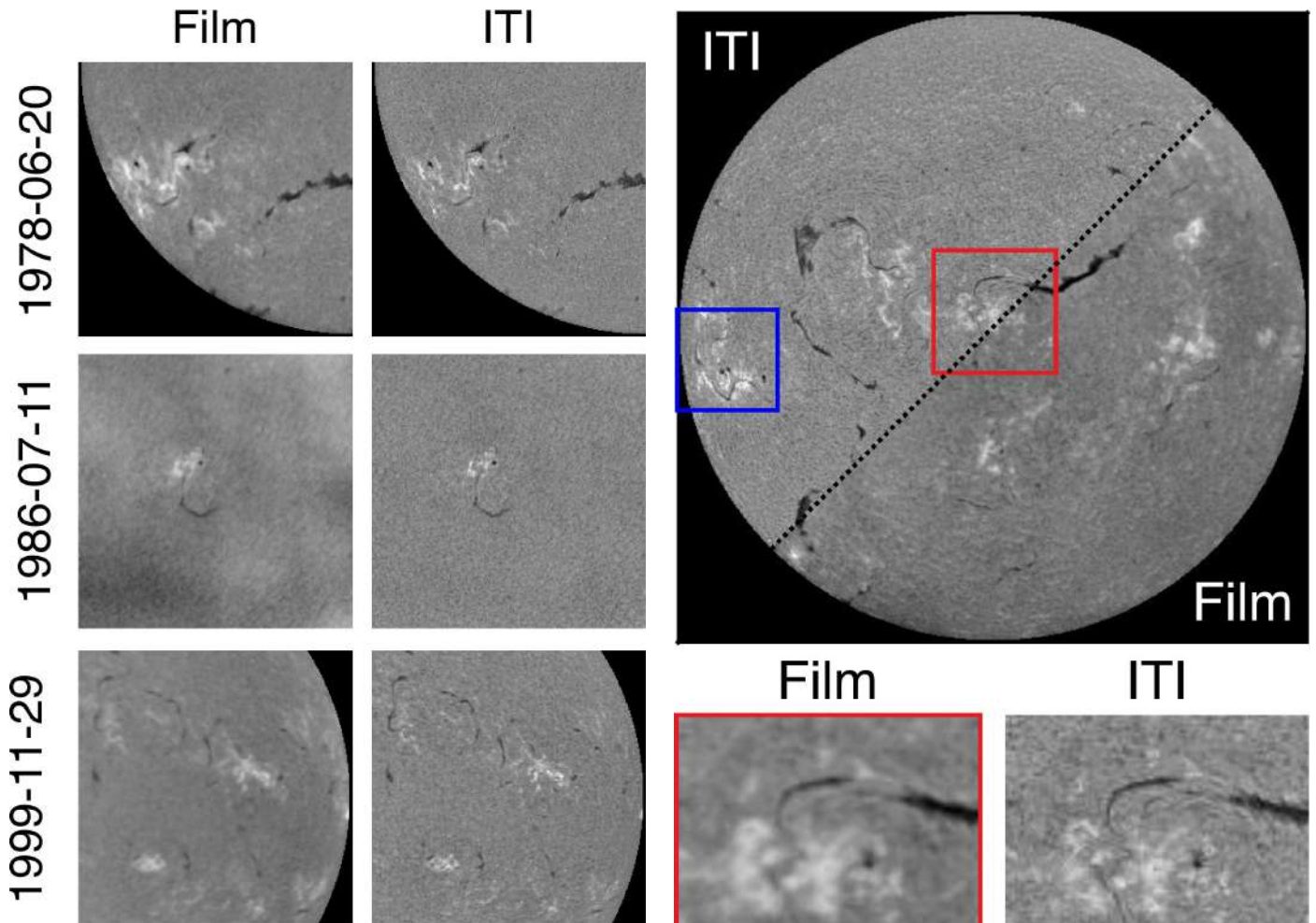
Fig. D.2. Example of KSO film-to-CCD translation. The ITI translation results in deblurring of plage regions, sharpening of filaments and enhances the structure of quiet Sun regions. Atmospheric effects are strongly mitigated in the enhanced version (middle).

Appendix D.5: Similarity of image distributions

The Fréchet inception distance (FID) is a commonly used metric for GANs to estimate the quality of synthesized images and is a measure for the distance between two image distributions (Heusel et al. 2017). We use the FID to compare the similarity of the high-quality data sets to the corresponding low-quality data sets and their ITI enhanced version.

We convert the test sets of each dataset to gray-scale images, where we scale the pre-processed data linearly between the minimum and maximum value. Differences in resolution between high- and low-quality data sets, are adjusted by bilinear upsampling. The reconstructed STEREO magnetograms are not included in this evaluation, since there exists no low-quality data set for comparison (Sect. 2.5).

We use an IncepctionV3 model with the weights of the tensorflow implementation for the FID. The FID is evaluated for each application and channel separately, where we use the full resolution images and do not normalize the inputs. The results in Table D.1 show that the ITI enhanced images are in every case closer to the high-quality image distribution than the low-quality distribution. This shows that our method is able to map images closer to the high-quality distribution, or in other words, leads overall to a perceptually better image quality. The interpretation of the FID in terms of distance to the high-quality image
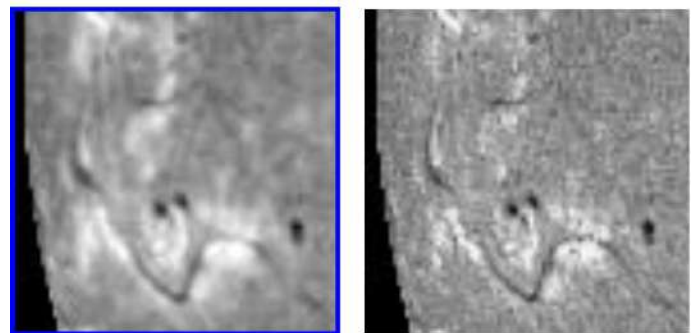

Fig. D.3. Detailed comparison of the KSO film-to-CCD translation. The ITI observation shows a more homogeneous appearance at the global scale (top). The comparison of two regions shows that active regions are well reconstructed (red). At smaller scales (blue), solar features can be better identified in the ITI observations, but the limits in resolution are visible.

distribution is not obvious for two reasons. (1) The used InceptionV3 network is trained on real world data, which can lead to unexpected behavior for scientific data and therefore variable FID scores. (2) In contrast to a GAN that synthesizes images from a random distribution, our translation model also takes the image content into account. The difference in content (e.g., solar

Table D.1. Evaluation of the FID for each application (lower better). We estimate the similarity of the ITI enhanced images and the original images to the reference high-quality observations.

Instruments	FID	
	LQ	ITI
SDO/HMI → Hinode continuum	36.6	**17.0**
SOHO/EIT → SDO/AIA 171 Å	35.7	**7.0**
SOHO/EIT → SDO/AIA 193 Å	27.3	**9.4**
SOHO/EIT → SDO/AIA 211 Å	17.6	**1.1**
SOHO/EIT → SDO/AIA 304 Å	32.3	**6.2**
SOHO/MDI → SDO/HMI magnetogram	25.2	**7.9**
STEREO/EUVI → SDO/AIA 171 Å	18.3	**10.8**
STEREO/EUVI → SDO/AIA 193 Å	39.0	**18.7**
STEREO/EUVI → SDO/AIA 211 Å	22.2	**13.9**
STEREO/EUVI → SDO/AIA 304 Å	13.7	**9.8**
KSO LQ → KSO HQ Hα	4.8	**2.0**
KSO Film → KSO CCD Hα	26.0	**8.1**

features, observed regions) can lead to an unavoidable bias in the FID.