A machine-learning approach to correcting atmospheric seeing in solar flare observations

John A. Armstrong [1]★ and Lyndsay Fletcher[1,2]

[1]*SUPA School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK*
[2]*Rosseland Centre for Solar Physics, University of Oslo, PO Box 1029 Blindem, NO-0315 Oslo, Norway*

ABSTRACT

Current post-processing techniques for the correction of atmospheric seeing in solar observations – such as Speckle interferometry and Phase Diversity methods – have limitations when it comes to their reconstructive capabilities of solar flare observations. This, combined with the sporadic nature of flares meaning observers cannot wait until seeing conditions are optimal before taking measurements, means that many ground-based solar flare observations are marred with bad seeing. To combat this, we propose a method for dedicated flare seeing correction based on training a deep neural network to learn to correct artificial seeing from flare observations taken during good seeing conditions. This model uses transfer learning, a novel technique in solar physics, to help learn these corrections. Transfer learning is when another network already trained on similar data is used to influence the learning of the new network. Once trained, the model has been applied to two flare data sets: one from AR12157 on 2014 September 6 and one from AR12673 on 2017 September 6. The results show good corrections to images with bad seeing with a relative error assigned to the estimate based on the performance of the model. Further discussion takes place of improvements to the robustness of the error on these estimates.

Key words: atmospheric effects – techniques: image processing – Sun: flares.

1 INTRODUCTION

Atmospheric scintillation is ubiquitous in ground-based astronomy. This poses a problem for all observers, particularly those studying highly variable phenomena. State-of-the-art observing facilities utilize adaptive optics (AO) systems in their optical path to correct for the wavefront deviations introduced by the atmosphere. However, when the seeing conditions are particularly bad, the object being observed evolves faster than the speed of the AO system, or the field of view is much larger than the isoplanatic patch, post-processing techniques must be introduced to correct for seeing. The two most common post-processing techniques used in solar physics are Speckle interferometry and Phase Diversity (PD) methods.

Speckle interferometry is the process that involves dividing the field of view into subfields that are smaller than the turbulence coherence length and correcting each individually. As a result, this depends on accurate statistics of the atmospheric turbulence and, consequently, requires hundreds of frames to estimate a diffraction-limited reconstruction (von der Lühe & Dunn 1987; von der Lühe 1993). For the most dynamic of processes (e.g. solar flares), the evolution time will be shorter than the cumulative length of the exposures required to obtain the frames necessary for the reconstruction. Therefore, the atmospheric parameters will need to be estimated from a number of consecutive frames, which is much less than optimal for the algorithm. This will lead to greater uncertainty in the atmospheric parameters and a poorer restoration, as a result.

PD methods jointly estimate the restored image and the distortions responsible for the aberrated image in a maximum likelihood estimation. The state-of-the-art PD method in solar physics is multi-object multiframe blind deconvolution (MOMFBD; Van Noort, Der Voort & L'ofdahl 2005). MOMFBD implements a simple model of the optics and detectors used in the observations, eliminating the need to rely on the atmospheric statistics as in Speckle reconstruction. As explained in Van Noort et al. (2005), PD methods work best when contrast is high, noise is low, and exposure time is short. This is difficult to achieve in narrow-band solar observations and, as a result, wide-band data collected simultaneously must be used to aid in the MOMFBD restoration. It is this that poses the biggest problem for the restoration of flare data. Chromospheric energy deposition in a flare is mostly seen through the enhancement of optical and near-infrared spectral lines and not necessarily strong continuum enhancements (Fletcher et al. 2011). This can lead to the objects being studied looking very different in the wide-band and narrow-band observations. Given that the wide-band is used to help the optimization of the restoration, in cases where there is no continuum enhancement in a flare, it can actually be a hindrance to the restoration.

Furthermore, both Speckle and PD methods have a limit to their restoration capabilities (as all methods will). This is detrimental to flare observations due to their sporadic nature meaning observers cannot wait for optimal seeing conditions to observe. For these reasons, we propose a dedicated flare seeing-correction tool based on training a deep neural network (DNN) on diffraction-limited narrow-band flare data synthesized with artificial seeing.

Deep learning is the science of using DNNs to learn previously not easily programmable tasks, to learn computationally expensive tasks

★ E-mail: j.armstrong.2@research.gla.ac.uk

to increase efficiency, and to gain insight into data via the use of data-driven models. The use of deep learning in solar physics has already been applied to solar image restoration for increasing the speed at which MOMFBD restores images (Asensio Ramos, Rodriguez & Yabar 2018) and for the estimation of the point spread function (PSF) to do multiframe blind deconvolution (Asensio Ramos 2020). Furthermore, the correction of noisy narrow-band polarization data has been achieved using deep learning (Díaz Baso, de la Cruz Rodríguez & Danilovic 2019). Here, we present a tool for correcting for seeing in observations of Stokes I in narrow-band observations of solar flares, utilizing a trained DNN to learn the mapping from seeing-plagued data to diffraction-limited data without the constraints or Speckle of PD methods.

The structure of the paper is as follows: Section 2 introduces the model used to generate synthetic seeing and a description of the data used to train the DNN; Section 3 highlights the architecture and training of the DNN with some validation data shown; Section 5 is the application of the model to observations from the data set described in Section 2.1 with natural bad seeing; Section 4 shows how well the model performs on data from the X9.3 flare SOL2017-09-06T12:03; and Section 6 presents the conclusions.

2 SEEING MODEL

The following outlines the synthetic seeing model that will be used to generate the training set for the neural network algorithm. This synthetic seeing model will be applied to images that we assume are corrected completely by the AO and data reduction pipelines, and are not affected by residual seeing (i.e. the images are diffraction-limited).

Atmospheric seeing is the refraction of light as it travels through the Earth's atmosphere due to the turbulent nature of the atmosphere's refractive index. This turbulence is caused by random variations in the density and temperature structure of the atmosphere. This greatly impacts the resolution of observations as images are subject to the following:

$$O = I * P_{\mathrm{atmos}} + G, \tag{1}$$

where $*$ denotes the convolution, I is the diffraction-limited image, P_{atmos} is the PSF of the seeing, G is random Gaussian noise (for a discussion on the use of Gaussian noise, see Van Noort et al. 2005), and O is the observed image. The effect on astronomical imaging is that regardless of telescope aperture size, images appear as though observed through a telescope with effective aperture size equal to the Fried parameter, r_0 (equation 6). The construction of the atmospheric PSF will provide a basis for applying synthetic seeing to diffraction-limited images for the network to learn from. The general form for the PSF of the atmosphere is given by the Hankel transform of the modulation transfer function

$$P_{\mathrm{atmos}}(\rho) = \int_0^\infty J_0(\rho\nu) \exp\{-0.5\, D_{\mathrm{S}}(\nu)\}\nu\,\mathrm{d}\nu, \tag{2}$$

where $D_{\mathrm{S}}(\rho)$ is the two-point correlation (structure) function between the phase of two wavefronts in the telescope focal plane, $J_0(\rho\nu)$ is the zeroth-order Bessel function, and ν is the spatial frequency.

To find a form for this structure function, the assumption is made that the Earth's atmosphere can be modelled as a medium with smoothly varying turbulence (Tatarski 2016). Then, the structure function is written as

$$D_{\mathrm{S}}(\rho) = 2.91 k^2 \rho^{5/3} \int_{\vec{\ell}} C_n^2(\vec{r})\, \mathrm{d}\vec{r}, \tag{3}$$

where $k = 2\pi/\lambda$ is the wavenumber of the light observed, ρ is the Euclidean distance in the sky, and C_n^2 is the profile describing the

structure of the atmosphere at a point $\vec{r}$ (i.e. this encompasses the turbulent nature of the refractive index of the atmosphere). $\vec{\ell}$ is the path taken by a photon through the atmosphere. For a photon incident on the 'top' of the Earth's atmosphere (the point where the medium becomes turbulent) at an angle θ to the normal of the atmosphere, the C_n^2 profile can be written as

$$\int_{\vec{\ell}} C_n^2(\vec{r})\, \mathrm{d}\vec{r} = \sec\theta \int_0^\ell C_n^2(z)\, \mathrm{d}z, \tag{4}$$

where the z-direction represents the direction of the normal to the atmosphere and the limit $z = 0$ corresponds to the top of the atmosphere, and $z = \ell$ corresponds to the total distance travelled by the photon. Equation (3) can then be rewritten as

$$D_{\mathrm{S}}(\rho) = 2.91 k^2 \rho^{5/3} \sec\theta \int_0^\ell C_n^2(z)\, \mathrm{d}z. \tag{5}$$

The Fried parameters are then written explicitly as (Fried 1966)

$$r_0 = \left(0.423 k^2 \sec\theta \int_0^\ell C_n^2(z)\, \mathrm{d}z \right)^{-3/5}. \tag{6}$$

This can then be used to simplify equation (5) to

$$D_{\mathrm{S}}(\rho) = 6.88 \left(\frac{\rho}{r_0} \right)^{5/3} = 6.88 \left(\frac{\lambda\nu}{2\pi r_0} \right)^{5/3}, \tag{7}$$

where λ represents the air wavelength of the light observed and the spatial frequency ν is expressed in units of radians of phase per radian field of view (Racine 1996).

The form of the structure function given by equation (7) is used as it eliminates the need to choose a model for the C_n^2 profile and instead, the Fried parameter becomes a free parameter in the model with a variety of different values explored.

As the image is degraded in quality to equivalent to one taken with an aperture of diameter r_0, the angular size of the PSF in the sky can be found using

$$\alpha = 2.021 \times 10^5 \times \frac{\lambda}{r_0}, \tag{8}$$

where α is measured in arcseconds. The size of the PSF in detector pixels (n_{pix}) can then be calculated by dividing by the angular size of a single pixel (α_{pix})

$$n_{\mathrm{pix}} = \frac{\alpha}{\alpha_{\mathrm{pix}}}. \tag{9}$$

n_{pix} is then the size of the PSF array to be convolved with the image (under the diffraction-limited assumption). The PSF is then populated using equations (2) and (7).

2.1 Generating training data

A range of Fried parameters $r_0 = \{1, 2.5, 5, 7.5, 10, 12.5, 15\}$ cm is used to generate many different PSFs to convolve with the good seeing images following equation (1). This creates a diverse training data set for the neural network to learn from.

The data used are taken with the Swedish 1-m Solar Telescope's CRisp Imaging SpectroPolarimeter (SST/CRISP) instrument (Scharmer et al. 2003; Scharmer 2006; Scharmer et al. 2008). CRISP is a dual Fabry–Pérot interferometer capable of narrow-band imaging spectropolarimetry and wide-band imaging. Imaging spectroscopy data are used in the training of the network with observations in two spectral lines: H α and Ca II $\lambda8542$. The observations are of three flares: the M1.1 two-ribbon solar flare SOL20140906T17:09 that took place in NOAA AR 12157 with heliocentric coordinates
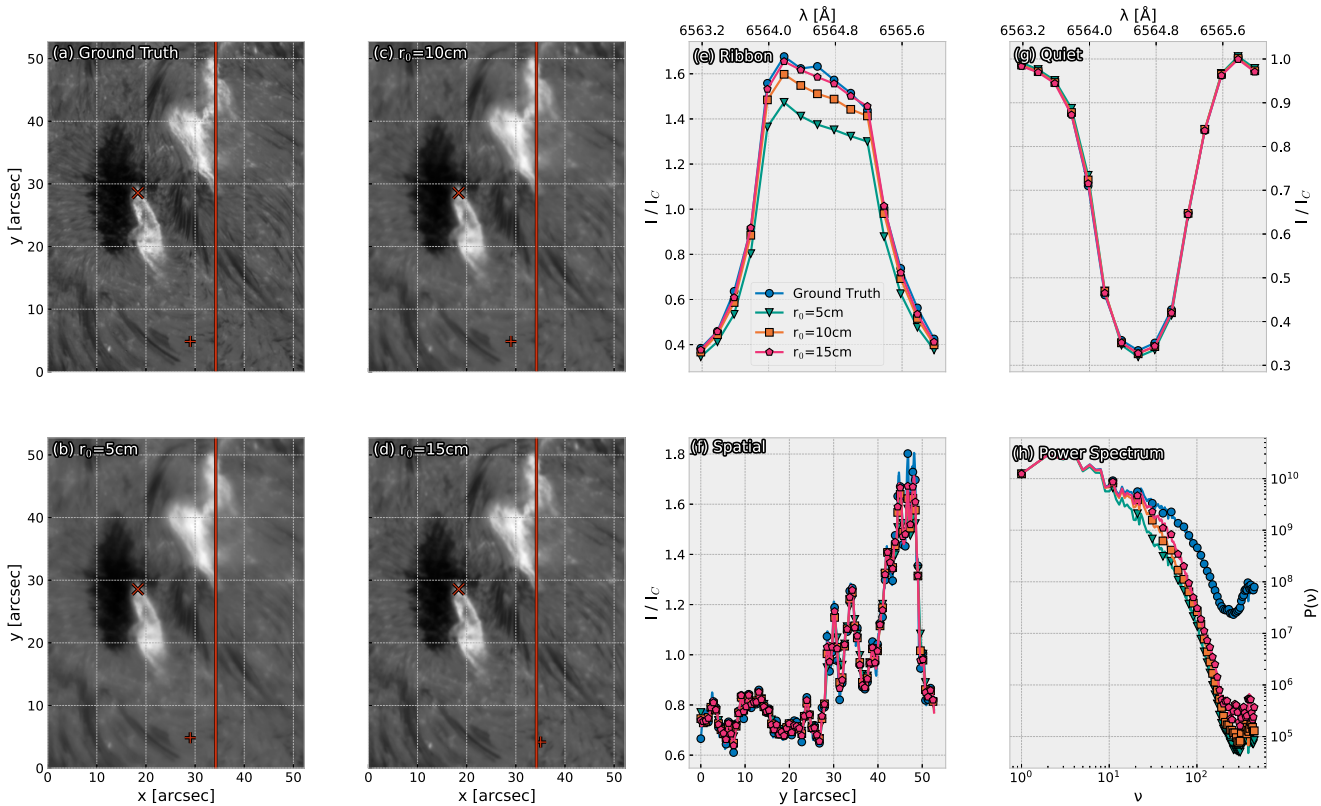
Figure 1. Example of the seeing model described in Section 2 applied to one of the good seeing observations from the data set described in Section 2.1; particularly, this observation is from 17:11:33UTC approximately 2 min after the flare soft X-ray peak, shown in panel (a). The image used here for demonstration is of the H α blue wing at $\Delta\lambda = -600$ mÅ. The seeing model is applied for three different Fried parameters as can be seen in (b) $r_0 = 5$ cm, (c) $r_0 = 10$ cm, and (d) $r_0 = 15$ cm. (e), (f), and (g) show the change in the spectral line on the flare ribbon, a spatial slice, and the spectral line in a quieter part of the atmosphere, respectively. The flare ribbon line is indicated by the cross in panels (a)–(d) with the quiet point being the plus sign and the slice shown by the vertical line. (h) shows the azimuthally averaged power spectrum across the images. In panels (e), (f), (g), and (h), the circles correspond to the ground truth, the triangles to $r_0 = 5$ cm, the squares to $r_0 = 10$ cm, and the pentagons to $r_0 = 15$ cm.

(-732 arcsec, -302 arcsec); an X2.2 event SOL20170906T09:10 that took place in NOAA AR 12673 with heliocentric coordinates (537 arcsec, -222 arcsec); and an X9.3 event SOL201709-06T12:02 taking place in the same AR as SOL20170906T09:10. For SOL2014-09-06T17:10, the H α data are sampled at 15 wavelength points in intervals of 200 mÅ from the line core, and the Ca II data consist of 25 wavelength points sampled at 25 wavelengths in intervals of 100 mÅ from the line core. These data are made publicly available via the F-CHROMA solar flare data base (Cauzzi et al. 2014).[1] For SOL20170906T09:10 and SOLT20170906T12:02, the H α data are taken at 13 wavelength points with the line core more densely sampled than the wings. The Ca II data are similarly sampled but for 11 wavelength points. All data have been pre-processed using the CRISPRED data reduction pipeline (de la Cruz Rodríguez et al. 2015) that includes all alignment, instrument calibration, and image restoration using MOMFBD. Therefore, the ground truth to be recovered makes the assumption that images without bad seeing conditions are completely corrected for seeing and other aberrations by the CRISPRED pipeline.

An example of the seeing model for three different values of the Fried parameter ($r_0 = 5$, 10, and 15 cm) is shown in Fig. 1. Fig. 1(a) shows the observation from 17:11:33UTC from the data set

described above (approximately 2 min after the flare soft X-ray peak) taken in the H α blue wing at $\Delta\lambda = -600$ mÅ. Figs 1(b)–(d) then show seeing corresponding to $r_0 = 5$, 10, and 15 cm, respectively, applied to the observation. Two points are then selected: one on the eastern flare ribbon indicated by the cross in Figs 1(a)–(d) and one in the quiet atmosphere represented by the plus in the same figures. The spectra of these points in each of the four cases are plotted in Fig. 1(e) for the ribbon and Fig. 1(g) for the quiet atmosphere. This indicates that worsening seeing will result in a reduction in the intensity of bright features due to the spatial smearing of the intensity. Fig. 1(f) shows the intensity variation with y for a fixed x indicated by the vertical line in Figs 1(a)–(d). Fig. 1(f) solidifies the previous points, showing that peaks and troughs of intensity are lost as seeing worsens. Also, small-scale features are lost due to bad seeing, which is apparent as curves in Fig. 1(f) become smoother for worse seeing. The power spectrum is also calculated for each image shown in Fig. 1(h), which further conveys the loss of small-scale features as the power in the higher frequencies is substantially reduced as seeing worsens.

3 NEURAL NETWORK APPROACH

The following section outlines the DNN architecture trained and used to correct for atmospheric seeing, and a description of how the network is trained.
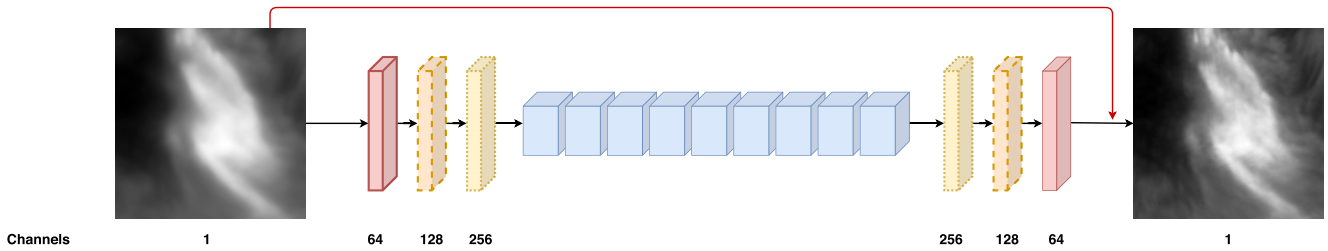
[1] https://star.pst.qub.ac.uk/wiki/doku.php/public/solarflares/start

Figure 2. Schematic of the neural network used to learn the seeing correction. This network consists of six convolutional layers and nine residual layers. The picture on the left of the network shows the input that is an image from the data set generated in Section 2.1 of an image imbued with synthetic seeing. The picture on the right is the ground truth the network is trying to recover. The first block (with the solid lines) is a convolutional layer using a 7×7 kernel and generating 64 feature maps. The block with the dashed lines downsamples the feature maps produced by the first block by a factor of 2 using a strided convolution of 3×3 kernel and produces 128 feature maps. The dotted line block downsamples the feature maps by a further factor of 2 using a strided convolution of 3×3 kernel and produces 256 feature maps. The shorter blocks in the middle are the residual layers that all consist of 3×3 kernel convolutions with 256 feature maps. The inner structure of the residual layers is shown in Fig. 3. The next dotted line block upsamples the feature maps by a factor of 2 using nearest neighbour interpolation and reduces the number of feature maps to 128. The second dashed line block then upsamples by a further factor of 2 using the same method while reducing the number of feature maps to 64. The last block in the network is a convolutional block that reduces the number of feature maps to the number of output channels using a 7×7 kernel convolution before passing the output through a hyperbolic tangent (tanh) function. This is then combined with the input to the network (red arrow) to produce the output of the network. In each of the convolutional and residual layers, the normalization is batch normalization and the activation is ReLU.

3.1 Architecture

The DNN architecture used is illustrated in Fig. 2 and inspired by the generator network used in Kupyn et al. (2017). The network follows an encoder–decoder framework wherein the input data – in this case, the image with bad seeing – are downsampled to a lower dimensional, abstract representation of itself that can be reconstructed without the bad seeing by the learned network by upsampling the representation at the other end of the network. This is accomplished using a combination of convolutional layers (consisting of convolution, normalization, and activation) and residual layers (Fig. 3; He et al. 2015). The normalization used in the convolutional and residual layers is batch normalization (Ioffe & Szegedy 2015), and the activation used is rectified linear unit (ReLU; Nair & Hinton 2010). The convolutional layers are used for the downsampling and upsampling of the data while the residual layers learn the complexities of the abstract representation. Nine residual layers were found to be the optimal number to learn to correct for the bad seeing, with three convolutional layers on either side of the residual layers to perform the down/upsampling.

The first convolutional layer (shown with emboldened vertices in Fig. 2) convolves the image with a 7×7 kernel and transforms the input to 64 feature maps. The dashed line layer convolves these feature maps with a 3×3 kernel, downsampling their dimension by a factor of 2 and doubling the number of feature maps to 128. The dotted line layer convolves the 128 feature maps in the same way as the previous layer, downsampling the feature maps by a factor of 2 and doubling the number of feature maps to 256.

After this, these feature maps are passed to the nine residual layers, shown as the shorter blocks in Fig. 2. Each of these layers has the structure shown in Fig. 3. The convolution kernel sizes are all 3×3 with each residual layer keeping the number of feature maps at 256.

Subsequently, the feature maps are given to the second dotted line layer that upsamples the feature maps by a factor of 2 using nearest neighbour interpolation and reduces the number of feature maps by a factor of 2–128 using a convolution with a 3×3 kernel. Then, the second dashed line layer follows the same process resulting in there being 64 feature maps a factor of 2 larger than those before being passed to the final layer. The final layer transforms the feature maps to the number of output channels (in this case, 1) using a convolution with a 7×7 kernel. The output of this layer is then operated on by a hyperbolic tangent (tanh) function before being combined with

the input to the network (shown by the red arrow in Fig. 2). Being combined with the input is what Kupyn et al. (2017) coined as a 'ResOut' connection. The philosophy behind this is that the network is trying to learn some function f that maps an input with bad seeing to an output with good seeing. Given that f depends on the input, there exists some residual function, H, such that

$$H(x) = f(x) - x, \tag{10}$$

where the input to the network has been denoted by x. Therefore, adding the input at the end of the network allows it to learn only the residual H that may be easier to learn than the function f.[2]

3.2 Training

The network described in Section 3.1 is then trained using the data in Section 2.1 with the images synthesized with seeing as the input to the network and the corrected images as the output.

Aside from generation of a good training data set, the key to training a network is to use the correct loss function to track how well the network is doing. In this case, the loss function takes the form of two individual loss functions in a linear combination: perceptual loss and mean square error (MSE) loss

$$\mathcal{L} = \mathcal{L}_{\mathrm{P}} + \mathcal{L}_{\mathrm{MSE}}. \tag{11}$$

Perceptual loss (introduced by Johnson, Alahi & Fei-Fei 2016) is a measure of similarity between two images based on how they are perceived by a different neural network from the one being trained. This is an example of transfer learning: the process of using a previously trained neural network to influence the learning of a new network. The network from Armstrong & Fletcher (2019) (henceforth, referred to as Slic) is used here due to it being trained to classify features in the solar atmosphere. The argument is that a network trained sufficiently well on recognizing features should produce the same feature maps for two identical images. Therefore, using a measure of the difference of these features maps produced deep within the Slic network will give a measure of the similarity in the two images. This works by taking the network generated image I_{G}

[2]This is also the basis of how residual layers work and why networks can be much deeper using them.
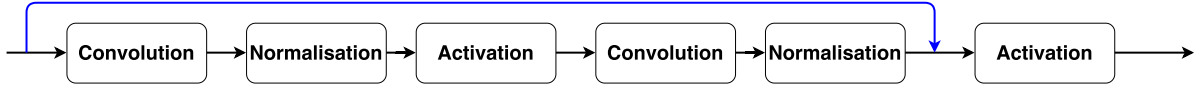
Figure 3. Inside of a residual block. This consists of two convolution layers applied to the input like traditional convolutional neural networks but with a skip connection (blue arrow) adding the input of the layer to the output before the second activation. This allows residual networks to be deeper than traditional networks as it prolongs the onset of the vanishing gradient problem.
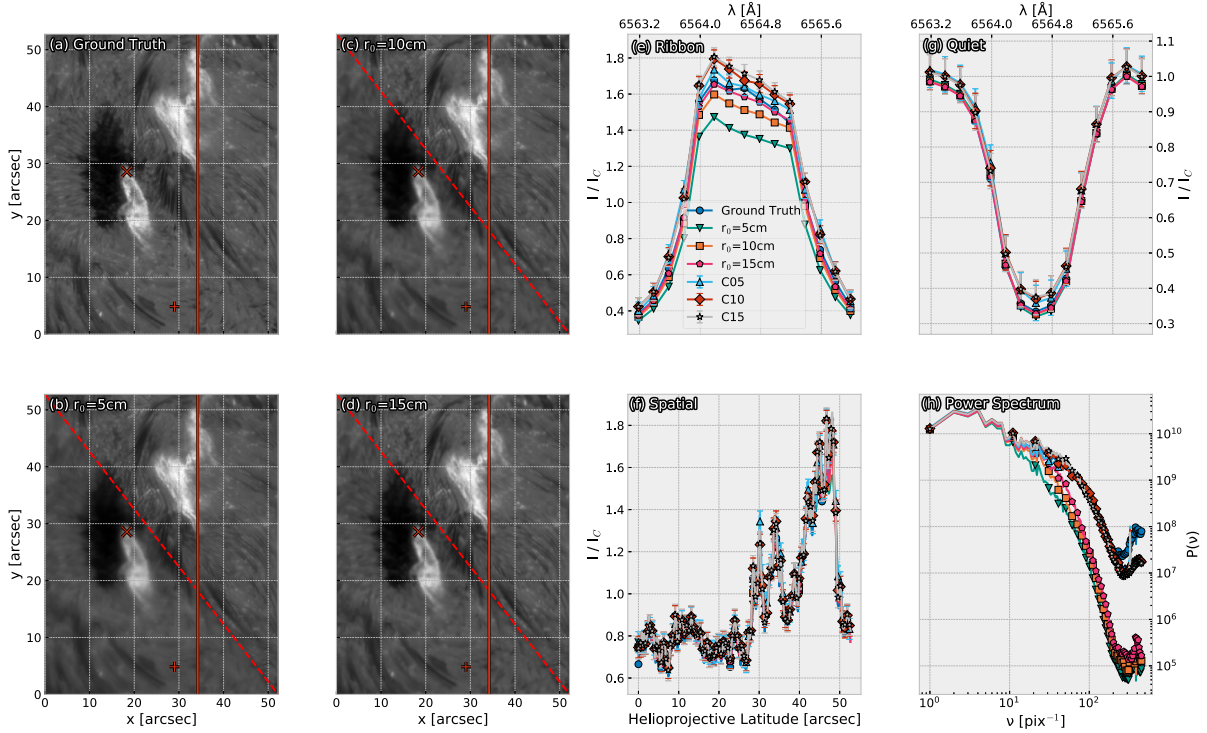

Figure 4. The results for applying the trained model to the data presented in Fig. 1. The layout of this figure is equivalent to the layout of Fig. 1 with a few differences. Panels (b)–(d) show both the ground truth data contaminated with the artificial seeing (data below the dashed line) and the reconstructed data from the network (data above the dashed line). This demonstrates the reconstructive power of the trained network for images with artificial seeing as the small-scale features that can be reconstructed are visible by eye. Note that the reconstruction is worse for worsening seeing conditions, which is to be expected from any algorithm of this kind. Panels (e)–(h) show the same as in Fig. 1 with the profiles for each of these cases for the reconstructions also plotted and indicated in the legend as e.g. 'C05' refers to correction of the data with bad seeing characterized by $r_0 = 5$ cm, etc. The reconstructed profiles are also plotted with their error bars, showing that the reconstruction is at least within the error bar of the ground truth.

and the ground truth image I_S and applying Slic to them. The output is then cut after the eighth layer and the feature maps compared using an MSE metric. This can be written as

$$\mathcal{L}_P = \frac{1}{W_j H_j} \sum_{x=0}^{W} \sum_{y=0}^{H} \left(\phi_j \left(I_S \right)_{x,y} - \phi_j \left(I_G \right)_{x,y} \right)^2, \tag{12}$$

where W_j are H_j are the width and height of the j-th output layer of Slic, respectively. ϕ_j is the function resulting from feeding the images through Slic and taking the output after the eighth layer.

The MSE loss is the N-dimensional Euclidean distance function squared where the data are N-dimensional.

$$\mathcal{L}_{MSE} = ||I_S - I_G||^2 \tag{13}$$

The MSE ensures that the magnitude of the reconstructions matches similarly to the ground truth images.

The perceptual and MSE losses are then minimized simultaneously using the Adam optimizer (Kingma & Ba 2014) using minibatching

and a variable learning rate following cosine annealing. Minibatching consists of not using the entirety of the training and validation data sets while training the network. Instead, 10 per cent of the training and validation data are used randomly per epoch for training. This increases the speed of the epoch that can speed up the convergence of the network (diversity across the data will lead to better generation as the network does not see the same data every epoch and having more but quicker epochs leads to more parameter updates and thus faster learning). In training, a batch size of 12 is used with 100 minibatches per epoch for the training data and 10 minibatches for the validation data. Cosine annealing (Loshchilov & Hutter 2016) is a method for dynamically changing the learning rate of the system every epoch following

$$\eta_t = \eta_{min} + \frac{1}{2} \left(\eta_{max} - \eta_{min} \right) \left(1 + \cos \left(\frac{T_{cur}}{T_{max}} \pi \right) \right), \tag{14}$$

where η_t is the current learning rate, η_{min} is the minimum learning rate, η_{max} is the starting learning rate, T_{cur} is the current epoch
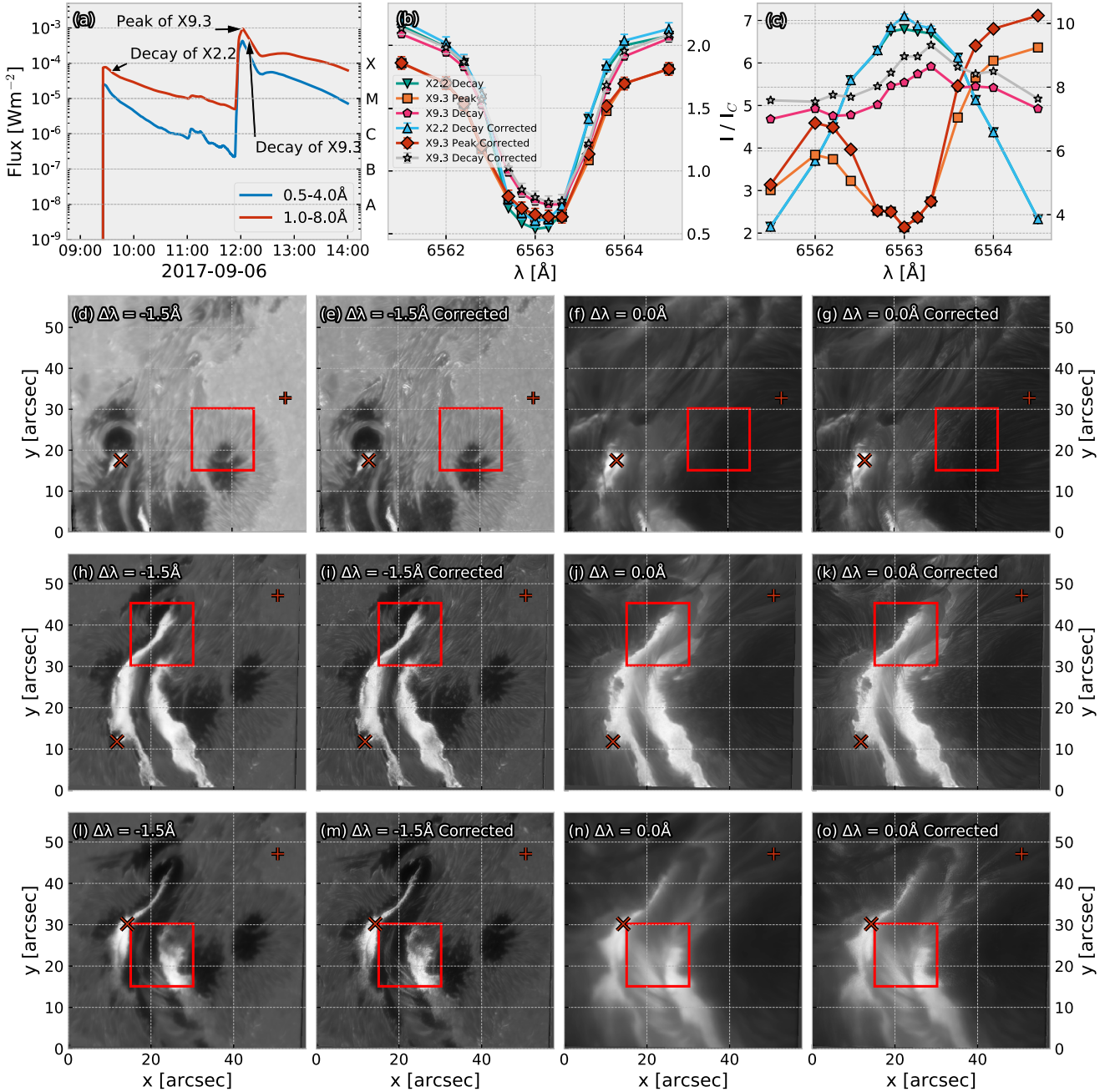
Figure 5. Panel (a) shows the GOES soft X-ray curve for the AR 12673 data, annotated to show where each observation corresponds to. (b) and (c) show spectra off and on the flare ribbon from the raw frames, respectively, and the corrected spectra for each of the cases. Trained model applied to observations from AR 12673 for the decay phase of the X2.2 flare SOL20170906T09:10 is shown in panels (d)–(g); the soft X-ray peak of the X9.3 flare SOL20170906T12:02 is shown in panels (h)–(k); and the decay phase of SOL20170906T12:02 is shown in panels (l)–(o). In each row, the first panel is the observation before correction taken in the far blue wing of Hα ($\Delta\lambda = -1.5$ Å); the second panel is the correction to the blue wing image; the third panel is the image in the line core before correction; and the last panel is the correction to the line core. The spectra shown are indicated in (d)–(o) using '+' and 'x' for (b) and (c), respectively. The boxes in panels (d)–(o) represent the subfields shown in Fig. 6.

number, and T_{max} is the number of epochs to get to the minimum learning rate. This method allows the exploration of local minima by decreasing the learning rate from η_{max} to η_{min} over T_{max} epochs while allowing the network to escape from incorrect local minima as the learning rate is reset to η_{max} after T_{max} epochs. The network here has $\eta_{max} = 5 \times 10^{-3}$, $\eta_{min} = 1 \times 10^{-6}$, and $T_{max} = 100$. The network is trained on an NVIDIA Titan Xp for 1900 epochs. The results are shown in Section 3.3 below.

3.3 Training results

To test the trained model, the data generated and shown in Fig. 1 are evaluated by the trained model with a spectral, spatial and power spectrum comparison as in Fig. 1. Given that the neural network model is an approximate fitting (in line with the Universal Function Approximation Theorem; Cybenko 1989; Lu et al. 2017), a formulation of an error on the estimate by the network is important. As such, an ad hoc error is calculated by evaluating the whole training

Table 1. The percentage changes in the total number of DNs between the non-corrected and corrected images. Negative percentage implies a lower number of DNs in the reconstruction and vice versa.

$\Delta\lambda$ (Å)	X2.2 decay (per cent)	X9.3 peak (per cent)	X9.3 decay (per cent)
−1.50	−2.50	−3.77	−3.62
−1.00	−2.77	−3.57	−3.72
−0.80	−3.12	−3.55	−3.42
−0.60	−3.53	−3.10	−3.29
−0.30	−4.60	−3.83	−3.38
−0.15	−4.68	−3.75	−3.82
0.00	−5.15	−3.88	−4.02
0.15	−5.22	−3.71	−3.65
0.30	−5.24	−3.39	−3.44
0.60	−3.96	−3.31	−3.36
0.80	−3.73	−3.43	−3.67
1.00	−3.18	−3.52	−3.39
1.50	−4.00	−3.80	−3.41

set by the trained model and taking the average error obtained from equation (11). Other methods of error estimation for neural networks are discussed in Section 6 but not explored in this work. The results of the reconstruction are shown in Fig. 4.

Figs 4(b)–(d) show both the ground truth data (Fig. 4a) contaminated with artificial seeing (the data below the dashed line) and the degraded data that have been reconstructed by the trained neural network model (the data above the dashed line). This shows the reconstructive power of the trained model, with small-scale structure recovery visible by eye. As with all algorithms of this kind, the reconstructions are of better quality for better seeing conditions.

For more quantitative measures of the reconstruction, three profiles are compared: a spectral line profile on the flare ribbon; a spectral line profile in a quieter part of the atmosphere; and a spatial line profile that is a slice of constant y shown by the cross, plus, and vertical line in Figs 4(a)–(d), respectively. The on-ribbon spectral line profile is plotted in Fig. 4(e). The ground truth is indicated by the circular markers with the degraded data indicated by downward-facing triangles for $r_0 = 5$ cm, squares for $r_0 = 10$ cm, and pentagons for $r_0 = 15$ cm. The reconstructed profiles are also plotted with their error bars and are indicated by upward-facing triangles for the reconstruction of $r_0 = 5$ cm, diamonds for $r_0 = 10$ cm, and stars for $r_0 = 15$ cm. Despite the large error bars, each case is reconstructed well by the model with the ground truth falling within the error bars and a noticeable return of the unusual shape near the peak of the line. Fig. 4(g) shows the results for the spectral line from the quieter part of the atmosphere. This follows the same convention as Fig. 4(e). In this case, the reconstruction is somewhat worse regardless of the seeing conditions as the wings of the line still have a discrepancy compared with the ground truth intensity values. This may be due to the focus of the trained model being subverted by the bright features with the lower contrast features not being as crucial in the reconstruction. Fig. 4(f) shows the slice of constant y. This illustrates on the whole that the brighter and darker features are reconstructed well by the model as there is not much discrepancy between reconstruction and ground truth along the slice.

Fig. 4(h) shows the power spectrum for the ground truth, each of the degraded images and the reconstructed images following the same descriptions as Figs 4(e)–(g). The reconstructions show that the large- to medium-scale structure (up to $\nu = 10$ pix^{-1}) within the field of view is almost perfectly reconstructed regardless of seeing

conditions. The rest of the spectrum for each reconstruction shows a tendency to reconstruct smaller features but not with the power they are represented by in the ground truth image. This is noticeable towards the highest frequencies where length-scales approach a single pixel; however, when the features are on scales of tens of pixels their power is still restored well for all seeing conditions. For example, when $r_0 = 15$ cm, there is still a good reconstruction up to approximately $\nu = 45$ pix^{-1}. The shapes of the reconstructed power spectra are correct compared with the ground truth, which suggests that learning for a better convergence of the L2 loss may result in the restoration of the lost power.

This has demonstrated the flexibility and accuracy of the restoration when applied to diffraction-limited images contaminated with artificial seeing. Next, the soundness of the model applied to images with no ground truth will be explored. This is done for two separate flares: Section 5 explores the M1.1 flare that the network is trained on but uses images that naturally could not be corrected by the CRISPRED data pipeline fully and Section 4 performs reconstructions on data from an X9.3 solar flare SOL2017-09-06T11:53.

4 CORRECTING BAD SEEING ON THE MOST EXPLOSIVE ACTIVE REGION OF SOLAR CYCLE 24

The most energetic flare of Solar Cycle 24, SOL20170906T12:02, occurred in NOAA active region AR12673 with GOES class X9.3 (Fig. 5a). This flare was observed by CRISP as described in Section 2.1. Unfortunately, seeing conditions were so poor that most of the data are affected by residual seeing. The true test for the neural network model is to see if it can reconstruct these data with accurate photometry across the field of view, perceptual similarities to their high-resolution counterparts, and sensible spectral line reconstruction.

In Fig. 5, we show three examples of corrections made to Hα observations with the neural network. Fig. 5(a) shows the GOES soft X-ray light curves indicating the de facto flare classification. This is annotated to show the three different times the examples are from: in the decay phase of SOL20170906T09:10 at 09:34:26UTC; at the peak of SOL20170906T12:02 at 12:02:26UTC; and in the decay phase of SOL20170906T12:02 at 12:09:11UTC. Figs 5(d)–(g) show the SOL20170906T09:10 decay phase observation in the Hα blue wing $\Delta\lambda = -1.5$ Å, the corrected observation in the blue wing, the observation in the Hα line core, and the corrected observation in the line core, respectively. Similarly, Figs 5(h)–(k) show the peak of SOL20170906T12:02 and Figs 5(l)–(o) show the decay phase of SOL20170906T12:02.

Each of Figs 5(d)–(o) is annotated with a '+' and an 'x'. The '+' indicates a point in a quieter part of the atmosphere with 'x' indicating a point on the flare ribbons. Correspondingly, the spectra from these points are shown in Figs 5(b) and (c). In these panels, the downward-facing triangles represent the spectral line before correction for the decay of SOL20170906T09:10, with the upward-facing triangles representing the spectrum following correction; the square points show the line profile before correction for the peak of SOL20170906T12:02, with the diamonds showing the line profile post correction; and the pentagons correspond to the profile before correction for the decay phase of SOL20170906T12:02, with the stars showing the profile post correction. The line profiles in Fig. 5(b) retain their shape when corrected with the intensity values in the wings (and, to a lesser extent, the core) increasing, which we would expect as seeing will effectively 'smear' light over many pixels causing a reduction in intensity in one pixel. This correction also
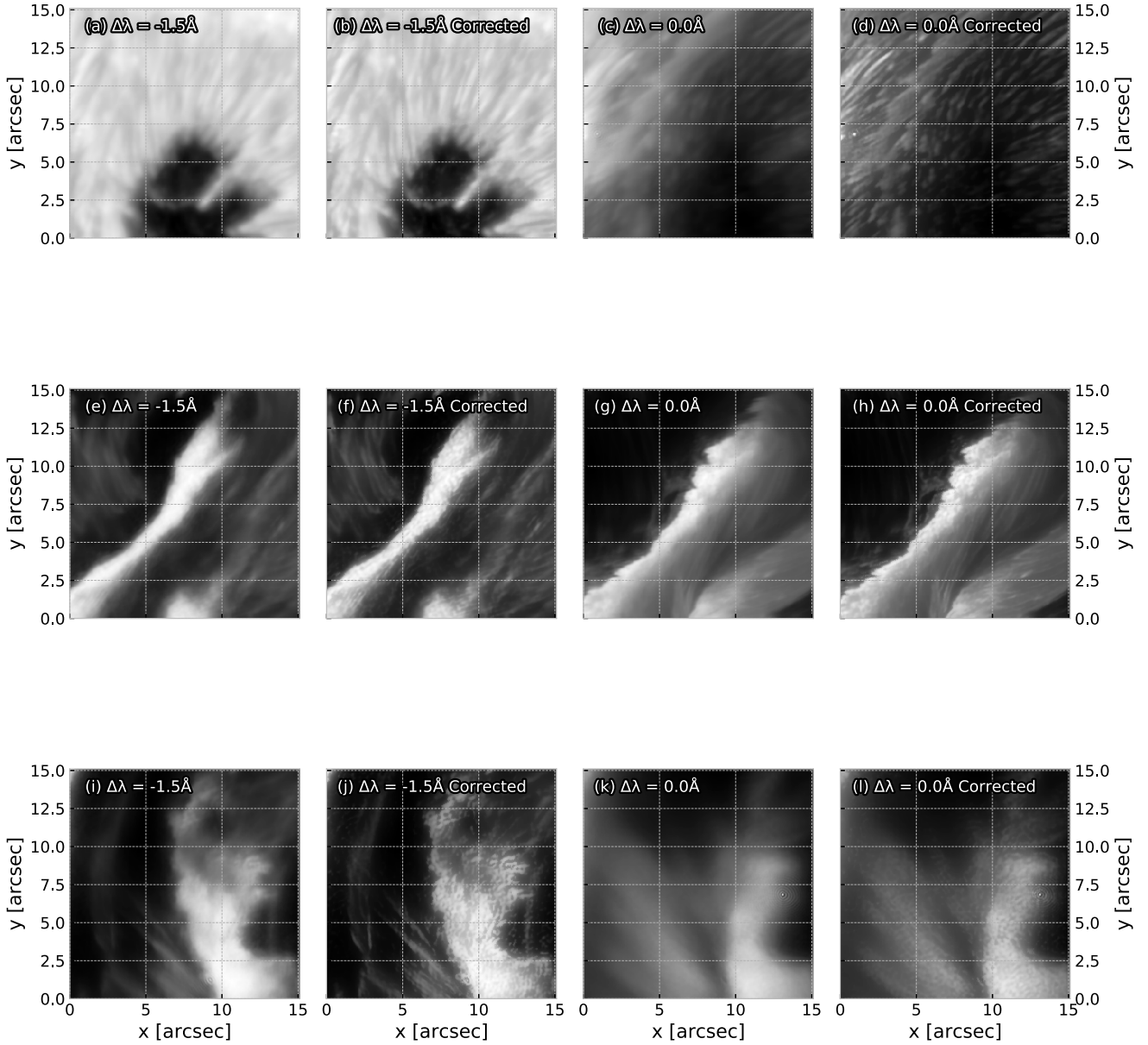
Figure 6. The subfields indicated by the boxes in Fig. 5. This shows the correction on the small-scale features in the images for the three different observations of AR 12673 indicated in Fig. 5(a). Panels (a)–(d) show the model applied to a sunspot umbra/penumbra region in the decay phase of SOL20170906T09:10 in both the Hα blue wing – (a) and (b) – and Hα line core – (c) and (d). Panels (e)–(h) show the application to the eastern flare ribbon at the peak of the SOL20170906T12:02 event following the same convention as the previous row. Similarly, panels (i)–(l) show the application to the western flare ribbon during the decay phase of SOL2017:0906T12:02.

preserves asymmetry in the line profile and Doppler shifts that can be seen clearly due to the differences in wing intensities between the blue and red wings and the intensity-averaged line core not being equal to the emitted wavelength, respectively. The line profiles in Fig. 5(c) show three very different line profiles on the flare ribbon depending on the time at which it is observed. For the decay phase of SOL20170906T09:10, the line profile has small changes in the wings after correction but a larger change towards the line core. The peak of SOL20170906T12:02 spectral line before correction appears as the characteristic twin-peaked Hα profile (with a very broad red wing) with the correction implying that the blue wing should be stronger than that in the raw observations. The decay phase of SOL20170906T12:02 spectral line before and after correction

maintains a similar shape with the intensities of the corrected profile being larger at every wavelength point. The increases in intensity of each of these line profiles are to be expected by the same spatial 'smearing' effect mentioned earlier but we would not expect the intensity to increase in every pixel (otherwise, we would be introducing phantom photons to our observations). To examine this further, another metric that we look to quantify the reconstruction is how well photon counts are preserved across both uncorrected and corrected images. Since the data are level 2, i.e. corrected for the effects of instrumental and detector noise as well as alignment between channels and other data in the data set and application of a seeing correction method, the unscaled intensity will have units of data numbers (DNs). This means that to check conservation of
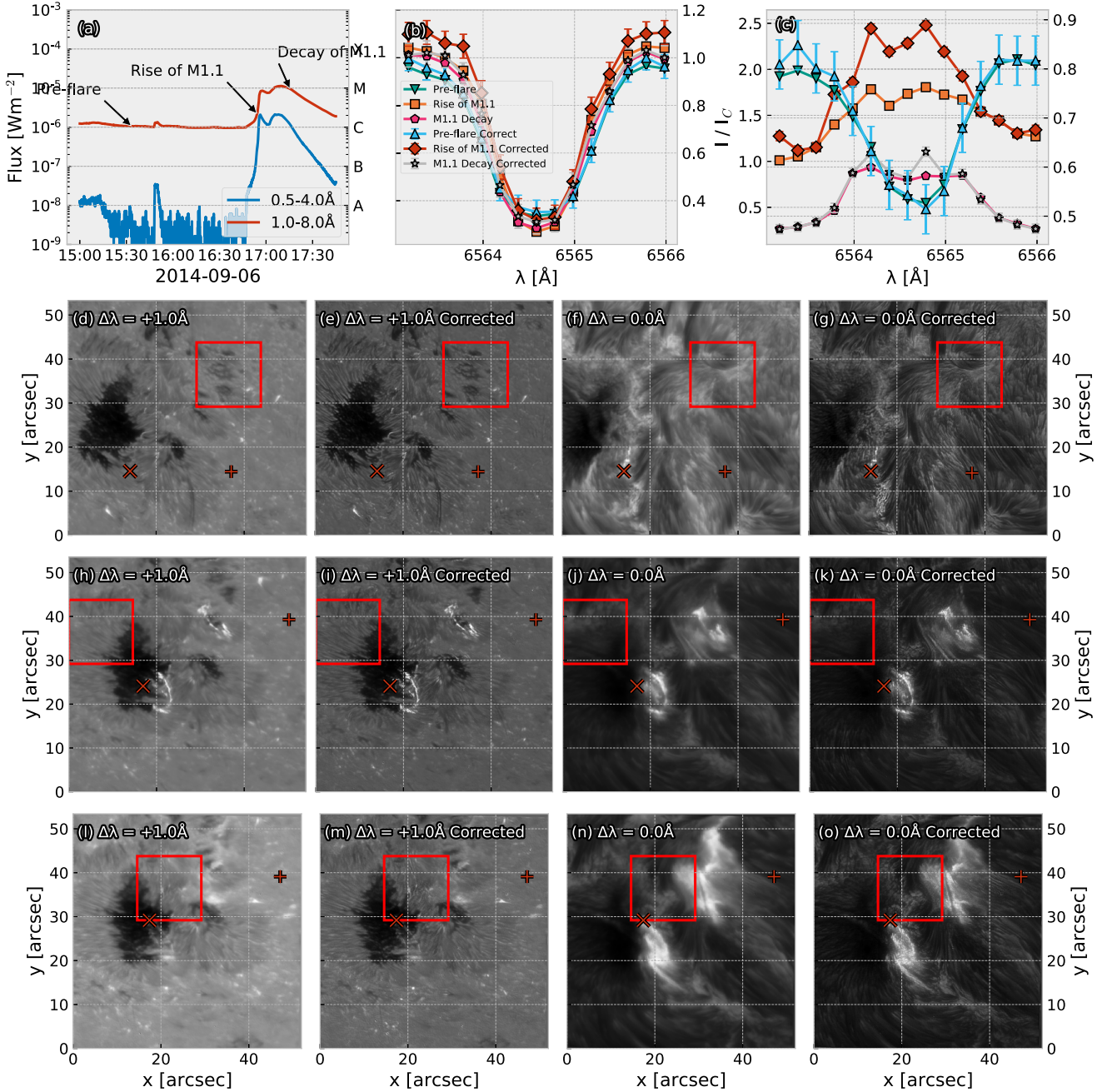
Figure 7. Panel (a) shows the GOES soft X-ray curve for the AR12157 data, annotated to show the time of each observation. (b) and (c) show absorption and emission spectra, respectively, from the raw frames, and the corrected versions for each of the cases. Trained model applied to observations from AR12157 for the pre-flare of SOL20140906T17:10 is shown in (d)–(g); the rise of the soft X-ray peak of SOL20140906T17:10 is shown in (h)–(k); and the decay of SOL20140906T17:10 is shown in (l)–(o). In each row, the first panel is the observation before correction, taken in the red wing of H α ($\Delta \lambda = +1.0$ Å); the second is the correction to the red wing image; the third panel is the image in the line core before correction; and the last panel is the corrected line-core image. The spectra shown are indicated in (d)–(o) using the '+' and 'x' for (b) and (c), respectively. The boxes in panels (d)–(o) represent the subfields shown in Fig. 8.

photons, a sum over the field of view will suffice. The results of this are shown in Table 1. In each frame, regardless of the event, the total DN decreases by a few per cent. This could mean that our neural network is undercompensating when correcting the flare ribbons or is overcompensating when correcting absorption features. This is discussed more in Section 6.

The boxes in Figs 5(d)–(o) reference the subfields shown in Fig. 6. This is to illustrate how well our model recovers small-scale features

in the flare ribbons and quieter parts of the Sun. Figs 6(a)–(d) show part of the umbra/penumbra of AR 12673 for the decay phase of SOL20170906T09:10 both in the far blue wing – panels (a) and (b) – and line core – panels (c) and (d) – of H α. Figs 6(e)–(h) show the eastern flare ribbon in the prior format for the peak of SOL20170906T12:02 and Figs 6(i)–(l) show the western flare ribbon in the same format for the decay phase of SOL20170906T12:02. This figure is for illustrative purposes and shows the good recovery
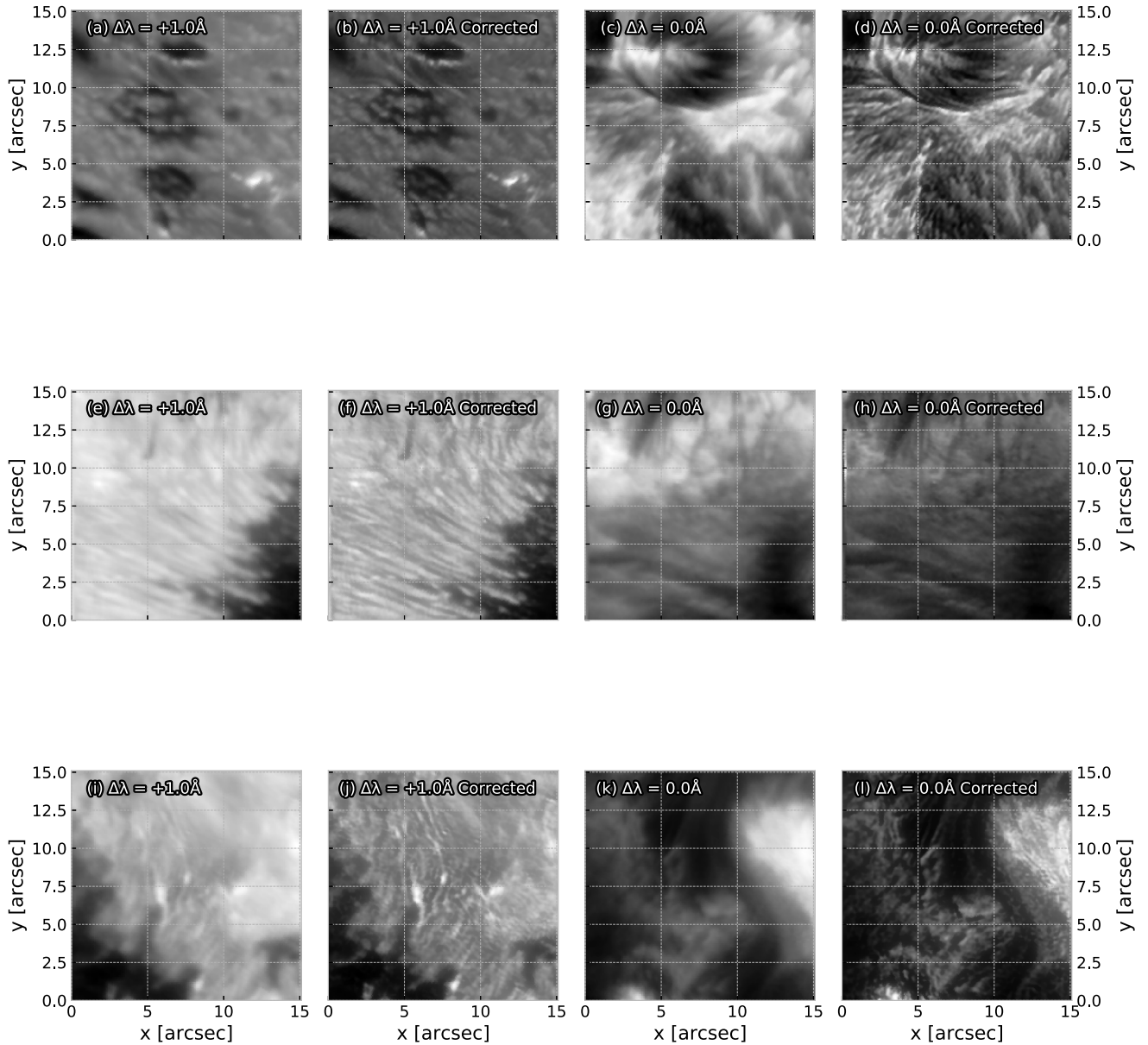
Figure 8. The subfields indicated by the boxes in Fig. 7. These show the correction on the small-scale features in the image for the three different observations of AR12157 indicated in Fig. 7(a). Panels (a)–(d) show the model applied to part of AR12157 north-west of the main sunspot during the pre-flare of SOL20140906T17:10 in both the Hα red wing – (a) and (b) – and Hα line core – (c) and (d). Panels (e)–(h) show the application to part of the sunspot umbra/penumbra during the rise of SOL20140906T17:10 following the same layout as the previous row. Likewise, panels (i)–(l) show the application to a region containing some sunspot penumbra and some of the northern flare ribbon during the decay of SOL20140906T17:10.

of small-scale features even when the seeing is particularly bad, as is most prominently seen in the observation of the decay phase of SOL20170906T12:02.

5 CORRECTING BAD SEEING ON THE M1.1 FLARE

As with the AR12673 data, the AR12157 data described in Section 2.1 contain both well-corrected data and data with seeing still present after the CRISPRED reconstruction. The data set can be used for both training and testing of the neural network model that provides some semblance of what the ground truth should be even when there is not a ground truth.

There are three examples in this section where the data contain bad seeing: one from the pre-flare phase, one from the rise of the soft X-ray peak, and one in the decay phase. The results of the neural network training and testing are shown in Figs 7 and 8.

Fig. 7(a) shows the GOES soft X-ray light curves indicating the flare classification and is annotated to show the three different times. The examples are: the pre-flare of SOL20140906T17:10 at 15:33:14UTC; during the rise of the soft X-ray peak of SOL20140906T17:10 at 16:54:13UTC; and in the decay of SOL20140906T17:10 at 17:15:24UTC. Figs 7(d)–(g) show the pre-flare observation in the Hα red wing Δλ = +1.0 Å, the corrected red wing observation, the Hα line core observation, and the corrected line core observation, respectively. Figs 7(h)–(k) and

Table 2. The percentage changes in the total number of DNs between the non-corrected and corrected images using the same conventions as Table 1.

$\Delta\lambda$ (Å)	Pre-flare (per cent)	Rise of M1.1 (per cent)	M1.1 decay (per cent)
−1.40	−4.38	−4.48	−4.27
−1.20	−4.53	−4.66	−4.46
−1.00	−4.54	−4.55	−4.63
−0.80	−5.16	−5.13	−4.95
−0.60	−5.65	−6.39	−5.91
−0.40	−8.39	−8.72	−7.21
−0.20	−10.39	−11.87	−9.64
0.00	−11.88	−10.84	−10.87
0.20	−10.78	−10.86	−11.56
0.40	−8.85	−8.90	−8.42
0.60	−6.29	−6.56	−6.17
0.80	−5.30	−5.23	−5.57
1.00	−4.77	−4.62	−4.36
1.20	−4.59	−4.37	−4.27
1.40	−4.09	−4.60	−4.42

(l)–(o) follow the same layout for the rise of the soft X-ray peak of SOL20140906T17:10 and the decay of SOL20140906T17:10, accordingly.

Each of Figs 7(d)–(o) is annotated with a '+' and an 'x'. The '+' indicates the spectra shown in Fig. 7(b) and the 'x' the spectra shown in Fig. 7(c). In these panels, the downward-facing triangles represent the spectral line before correction for the pre-flare observation and the upward-facing triangles represent the post-correction spectrum; the square points show the line profile before correction for the rise of the soft X-ray peak of SOL20140906T17:10, with the diamonds showing the line profile post correction; and the pentagons correspond to the profile before correction of the decay observation, with the stars showing the profile post correction. As with Fig. 5(b), the line profiles in Fig. 7(b) retain their shape and have enhanced intensities across the lines. The pre-flare-corrected spectrum is the one that has changed the most with a noticeable increase in intensity towards the line core. This is a result of the 'smearing' of light mentioned in Section 4. Again, the Doppler shifts and intensity-averaged wavelengths are approximately conserved. The line profiles from the rise-phase observations in Fig. 7(c) show a different story. The shape of the line profile is not too dissimilar from the line profile before correction but the intensity values are estimated at around $2\times$ higher. This could be in part due to not only the 'smearing' of the flare ribbon emission before correction but also due to overestimation of the bright features, which is discussed more in Section 6. This correction is outside of the range of the error bars of the estimate and a more robust approach to error calculation for this model may be needed and is discussed further in Section 6. The profile in the decay phase is corrected in a similar manner to the profile in the peak of SOL20170906T12:02 (Section 4), in that the shape develops the typical two-horn profile of H α in flare ribbons.

The sum of the DNs across the field of view at each wavelength for each of the cases before and after correction is also calculated as in Section 4. The results are shown in Table 2. These corrections show the same behaviour as for the AR12673 data with a decrease in the number of DNs by a few per cent. This implies that a further DN conservation metric could be implemented in equation (11) to offset this underestimate.

The boxes in Figs 7(d)–(o) reference the subfields shown in Fig. 8. Again, this is to illustrate how well our model recovers small-scale features across the varied field of view. Figs 8(a)–(d) show the north-easterly part of AR12157 for the pre-flare both in the red wing – panels (a) and (b) – and line core – panels (c) and (d) – of H α. Figs 8(e)–

(h) show part of the sunspot umbra/penumbra in the previous format during the rise of the soft X-ray curve of the flare, and Figs 8(i)–(l) part of the sunspot penumbra during the decay of the flare in the same format. This figure, again, is for illustrative purposes and shows the good recovery of small-scale feature particularly in the line core.

6 DISCUSSION AND CONCLUSIONS

We have presented a new method for seeing correction of intensity (Stokes I) images in ground-based solar flare observations. This method can be adapted to other problems after generation of the training set. In this method, a neural network is trained to learn to correct for synthetic seeing, generated by a mathematical model, which is applied to data observed in good seeing. This network is then applied to real data taken in bad seeing. We found that the network performs best when the effects of seeing are minimal, as expected. When seeing is worse, the network is still good at recovering large-scale features in the images (see the bottom row in Fig. 5) but, on small scales, the reconstruction is perceivably less accurate (see Figs 6j and l). Moreover, when the seeing is worse, the network seems to overcompensate on the small scales introducing features that are not necessarily physical (again, see Figs 6j and l and 8l). On the other hand, the overcompensation may not be due to the bad seeing entirely as a small instrumental blemish can be seen in Figs 6(k) and (l) just below the $y = 7.5$ arcsec line at around $x = 12.5$ arcsec. This takes the form of a Moiré pattern that may be introduced during the observation or the calibration of the data. Further examples of this pattern appearing can be seen on larger scales in Figs 7(h) and (l). These patterns may cause inaccuracies in the reconstruction by the network.

An estimate of the error of the network was made by taking the final trained model and applying it to the training and validation sets combined and calculating the mean of the calculated losses by equation (11). This is a rather ad hoc error that we wish to improve in the future using the method proposed in Lowe & Zapart (1999) of training a network with an additional input that is the variance of the estimate that the network generates. This will add a robustness to our error calculation and deliver a network capable of providing corrections and their confidence intervals. This may help us to solve the problem of the underestimation of DNs as discussed in Sections 4 and 5 as the discrepancy in DNs may be encapsulated by the confidence interval.

Furthermore, the change in DNs was considered for the training data set that yielded interesting findings: The discrepancy in the number of DNs for the training set is always small (<1 per cent) and is both positive and negative (meaning that there is sometimes an overestimation and sometimes an underestimation). This contrasts the examples we study in Sections 4 and 5. This means that the discrepancy in photons may just be apparent in the examples that we have chosen but it could also be systematic of the model we have trained. In case of the latter, we propose, in the future, to have an added term to equation (11) in the optimization of our system that minimizes the difference in the total number of DNs between the ground truth and the corrected frames.

All in all, the model that has been trained produces nice corrections on spectroscopic images that would otherwise be plagued with bad seeing. This allows us to study these flare events at higher time resolution more confidently as the geometry of the ribbons and their intensities have been corrected for bad seeing. We note that our model only performs seeing correction for Stokes I and in the case of having full spectropolarimetric imaging it is hypothesized that the seeing in Stokes Q, U, and V can be corrected for using the method in Díaz Baso et al. (2019).

DATA AVAILABILITY

The code used here is available at https://github.com/rhero12/Shaun. The code also includes a model trained for corrections to Ca II 8542 Å data where seeing on average will be worse due to its longer wavelength.

REFERENCES

Armstrong J. A., Fletcher L., 2019, Sol. Phys., 294, 80
Asensio Ramos A., 2020, preprint (astro-ph/2006.01438)
Asensio Ramos A., Rodriguez J. d. l. C., Yabar A. P., 2018, A&A, 620, A73
Cauzzi G., Fletcher L., Mathioudakis M., Carlsson M., Heinzel P., Berlicki A., Zuccarello F., 2014, Am. Astron. Soc. Meeting Abstr., #224
Cybenko G., 1989, Math. Control Signals Syst., 2, 303
de la Cruz Rodríguez J., L'ofdahl M. G., Sütterlin P., Hillberg T., Rouppe van der Voort L., 2015, A&A, 573, A40
Díaz Baso C. J., de la Cruz Rodríguez J., Danilovic S., 2019, A&A, 629, A99
Fletcher L. et al., 2011, Space Sci. Rev., 159, 19
Fried D. L., 1966, J. Opt. Soc. Am., 56, 1372
He K., Zhang X., Ren S., Sun J., 2015, preprint (arXiv:1512.03385)
Ioffe S., Szegedy C., 2015, preprint (arXiv:1502.03167)
Johnson J., Alahi A., Fei-Fei L., 2016, preprint (arXiv:1603.08155)
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kupyn O., Budzan V., Mykhailych M., Mishkin D., Matas J., 2017, preprint (arXiv:1711.07064)
Loshchilov I., Hutter F., 2016, preprint (arXiv:1608.03983)
Lowe D., Zapart K., 1999, Neural Comput. Appl., 8, 77
Lu Z., Pu H., Wang F., Hu Z., Wang L., 2017, preprint (arXiv:1709.02540)
Nair V., Hinton G. E., 2010, in FÜrnkranz J., Joachims T., eds, Proc. 27th Int. Conf. Mach. Learn., ICML'10, Omnipress, Madison, WI, USA. p. 807
Racine R., 1996, PASP, 108, 699
Scharmer G. B., 2006, A&A, 447, 1111
Scharmer G. B., Bjelksjo K., Korhonen T. K., Lindberg B., Petterson B., 2003, in Keil S. L., Avakyan S. V., eds, Proc. SPIE Conf. Ser. Vol. 4853, Innovative Telescopes and Instrumentation for Solar Astrophysics. Waikoloa, Hawai'i, USA, p. 341
Scharmer G. B. et al., 2008, ApJ, 689, L69
Tatarski V. I., 2016, Wave Propagation in a Turbulent Medium, reissue edn. Dover Publ. Inc., Mineola, New York
Van Noort M., Der Voort L. R. V., L'ofdahl M. G., 2005, Sol. Phys., 228, 191
von der Lühe O., 1993, A&A, 268, 374
von der Lühe O., Dunn R., 1987, A&A, 177, 265

This paper has been typeset from a TEX/LATEX file prepared by the author.