NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

DanceCam: atmospheric turbulence mitigation in wide-field astronomical images with short-exposure video streams

Bialek, Spencer; Bertin, Emmanuel; Fabbro, Sébastien; Bouy, Hervé; Rivet, Jean-Pierre; Lai, Olivier; Cuillandre, Jean-Charles

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1093/mnras/stae1018
Monthly Notices of the Royal Astronomical Society, 531, 1, pp. 403-421, 2024-04-19

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=1927c950-8b25-4ea6-9fc8-19cbf661bbe4
https://publications-cnrc.canada.ca/fra/voir/objet/?id=1927c950-8b25-4ea6-9fc8-19cbf661bbe4

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **531,** 403–421 (2024)
Advance Access publication 2024 April 19

https://doi.org/10.1093/mnras/stae1018

DanceCam: atmospheric turbulence mitigation in wide-field astronomical images with short-exposure video streams

Spencer Bialek [1,2★] Emmanuel Bertin [2,3★] Sébastien Fabbro [1,4] Hervé Bouy [5,6] Jean-Pierre Rivet [7] Olivier Lai [7] and Jean-Charles Cuillandre [3]

[1]*Department of Physics and Astronomy, University of Victoria, Victoria, BC V8W 3P2, Canada*
[2]*Canada–France–Hawaii Telescope, Kamuela, HI 96743, USA*
[3]*AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Cité, F-91191 Gif-sur-Yvette, France*
[4]*National Research Council Herzberg Astronomy and Astrophysics, Victoria, BC V9E 2E7, Canada*
[5]*Laboratoire d'Astrophysique de Bordeaux, CNRS and Université de Bordeaux, Allée Geoffroy St. Hilaire, F-33165 Pessac, France*
[6]*Institut Universitaire de France*
[7]*Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire J.–L. Lagrange, F-06304 Nice Cedex 4, France*

Accepted 2024 April 5. in original form 2024 March 16

ABSTRACT

We introduce a novel technique to mitigate the adverse effects of atmospheric turbulence on astronomical imaging. Utilizing a video-to-image neural network trained on simulated data, our method processes a sliding sequence of short-exposure (∼0.2 s) stellar field images to reconstruct an image devoid of both turbulence and noise. We demonstrate the method with simulated and observed stellar fields, and show that the brief exposure sequence allows the network to accurately associate speckles to their originating stars and effectively disentangle light from adjacent sources across a range of seeing conditions, all while preserving flux to a lower signal-to-noise ratio than an average stack. This approach results in a marked improvement in angular resolution without compromising the astrometric stability of the final image.

Key words: methods: data analysis – methods: observational – techniques: image processing – software: simulations – turbulence – atmospheric effects.

1 INTRODUCTION

Modern astronomy demands data of unparalleled precision and resolution to further its insights and discoveries. In their pursuit of this goal, astronomers often wrestle with the challenges imposed by the Earth's atmosphere. Due to random fluctuations in pressure, temperature, and wind speed, turbulence arises in the atmosphere along with a stratified refractive index distribution. The rapidly varying nature of these fluctuations causes the wavefront of incoming light to be dynamically distorted, leading to the twinkle of starlight. Short exposure images reveal the starlight to be composed of speckles with randomly varying shapes and positions around a central point. As a result, long exposures lead to the degradation of image resolution and the overall effect can be framed as a blurring operation from a bell-like point spread function (PSF). Atmospheric turbulence thus causes a significant loss in the spatial resolution of astronomical images and is a major impediment to obtaining high-quality imaging data (Roddier 1981).

The only way for a ground-based telescope to combat turbulence is by recording its effects in real-time and using the information from these snapshots to move the speckles back to a central point. Adaptive optics (AO) is one such method that has emerged as a revolutionary tool to counteract these disturbances, enabling telescopes to achieve near diffraction-limited observations (Beckers 1993; Hardy 1998). AO systems include a wavefront sensor to measure the wavefront distortions, from a laser or natural guide star, caused by the atmosphere, and then compensate for these distortions, typically with a deformable mirror, on millisecond timescales. The performance of AO has seen remarkable improvements over time, with modern systems capable of delivering images with resolutions that rival those from space telescopes (Davies & Kasper 2012).

However, AO is not without its limitations. Traditional AO systems often have a limited field of view, correcting only a small region around the guide star. Their effectiveness is also contingent on the presence of a *bright* guide star or the use of artificial laser guide stars, limiting where and when they can be used. To address these limitations, modern AO systems have evolved to incorporate advanced techniques such as multi-conjugate adaptive optics (MCAO) and ground-layer adaptive optics (GLAO). MCAO uses multiple deformable mirrors conjugated to specific layers of the atmosphere to correct over a wider field of view, while GLAO aims to provide a uniform correction over a wide field by primarily correcting the turbulence close to the ground (Johnston & Welsh 1994; Rigaut, Ellerbroek & Flicker 2000; Tokovinin 2004). While these systems have achieved impressive corrections over wider fields (e.g. Massari et al. 2016; Abdurrahman et al. 2018), the field of view still remains significantly smaller compared to what is achievable with uncor-

★ E-mail: spencer.bialek@gmail.com (SB); emmanuel.bertin@universite-paris-saclay.fr (EB)

© 2024 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

rected modern imaging fields, and there are additional challenges related to the cost and complexity of implementing and maintaining advanced AO systems. Time overheads during observations are also often an issue.

An alternative method that has gained traction in the astronomical community is high cadence imaging, in particular popularized with the 'lucky imaging' technique. Lucky imaging capitalizes on the intermittency of turbulence and the brief moments of atmospheric stability. By taking a rapid series of short-exposure images, only the sharpest frames – those taken during moments of optimal seeing conditions – are selected. These 'lucky' frames are then aligned and combined to produce a single high-resolution image. While lucky imaging can achieve impressive resolutions over wide fields (Mackay et al. 2018), especially for brighter targets, its effectiveness is inherently tied to the whims of the atmosphere. In poor seeing conditions, the probability of capturing lucky frames decreases, making the technique less effective (Faedi et al. 2013).

Another popular seeing mitigation technique using high cadence imaging is fast guiding, which focuses on compensating the image jittering caused by atmospheric turbulence. This is only effective in the regime where the tilt component of the wavefront distortions outweighs the sum of all other contributions to the total phase error budget, that is, when the PSF features a dominant speckle which can be used to track and compensate for random image motions on the focal plane. In practice, the correction is most effective for telescopes with apertures ≈3–4 times the Fried parameter (Fried 1966; Young 1974). The compensation may be done using, e.g. a tip-tilt mirror (e.g. McClure et al. 1989), 2D synchronous charge transfers (Tonry, Burke & Schechter 1997), or shift-and-add stacking of short exposures (Bates & Cady 1980). In the regime where the PSF features several speckles, and where a sufficiently bright star is available inside the region of interest, one may also consider holographic image reconstruction (Schödel et al. 2013).

In all cases of fast guiding, the correction remains only effective over a limited solid angle around the reference star(s) (isokinetic or isoplanatic disc, depending on the type of correction). The diameter of the disc depends on the altitude of the turbulent layer and ranges from a degree or more for the ground layer down to a few arcseconds for the highest layers. Hence in practice this type of method is only applicable to small patches individually, and correcting apparent motions over a wide field-of-view requires adjusting a 'rubber' focal plane model (Kaiser, Tonry & Luppino 2000), which consists of a distorted virtual pixel grid whose deformations are continuously controlled by a number of guide stars over its surface.

As the field of astronomical instrumentation continues to push the boundaries of observational capabilities, there is a pressing need for more versatile and cost-effective solutions that can operate under a broader range of conditions. With the rapidly evolving landscape of digital technology, there is a growing opportunity for machine learning (ML) methods to meet these demands and help fill in some of the gaps of AO and lucky imaging systems.

ML has already been successfully employed to mitigate atmospheric turbulence effects in long-range imaging applications (Nieuwenhuizen & Schutte 2019; Vint et al. 2020; Hoffmire et al. 2021; Zhang et al. 2024). These methodologies primarily rely on training ML algorithms on a collection of artificially distorted images, enabling them to predict the known ground truth and rectify turbulence-induced aberrations, though there are other methods which attempt to accomplish this in an unsupervised way (e.g. Li et al. 2021). In essence, ML can be used as a digital counterpart to lucky imaging, where instead of relying on elusive high quality exposures, algorithms *infer* the turbulence-free image from a video sequence of short exposures. This 'DanceCam' approach, as we call it, offers the potential to consistently produce sharp, high-resolution images, even in less-than-ideal seeing conditions and with a more efficient use of telescope time.

Despite the success in long-range imaging, the adaptation and application of ML techniques to turbulence in astronomical images remains largely uncharted, presenting a promising frontier for future research. The uniqueness of astronomical imaging poses specific challenges and requirements that differentiate it from long-range imaging. Characteristics like the extreme range of object brightness, object sizes and scales, and the requirement for ultra-high resolution and precision measurements, make astronomical imaging a domain where a distinct approach is required (Tyson & Frazier 2022).

In this paper, we present a novel method to combat the deleterious effects of atmospheric turbulence in astronomical images using ML trained on simulations of turbulent and noisy video-streams of stellar fields. Section 2 outlines the method used to simulate atmospheric turbulence and the ML methods used in this study. Section 3 describes both how the simulated data sets were created and how real test data were collected from the C2PU telescope. Section 4 is an overview of the main results from evaluating the proposed method on simulated and real data. Section 5 summarizes the strengths and limitations of the proposed method, and concluding remarks are in Section 6.

2 METHODS

2.1 Simulating atmospheric turbulence

We implement a variation of the 'split-step' method (e.g. Chatterjee & Mohamed 2014) for simulating the propagation of a wavefront through the atmosphere, in which the atmosphere is decomposed into several distinct layers which perturb the wavefront as it passes through. This section describes the theoretical framework used for the split-step simulations.

The Kolmogorov theory of turbulence (Kolmogorov 1941a,b) was one of the first models used for describing the statistical properties of atmospheric turbulence (Frisch & Kolmogorov 1995). It assumes that the turbulence is isotropic, homogeneous, and fully developed, which means that the turbulence has reached a statistically steady state. The Kolmogorov model is characterized by a power-law scaling of the spatial frequency spectrum of the turbulence, and the refractive-index power spectral density is given by:

$$\Phi_n = 0.033\, C_n^2\, k^{-11/3} \qquad \text{for} \qquad \frac{1}{L_0} \ll k \ll \frac{1}{l_0},$$

where L_0 is the outer scale, i.e. the average size of the largest eddies, l_0 is the inner scale, i.e. the average size of the smallest eddies, C_n^2 is the refractive index structure constant, which is a measure of the strength of the turbulence, and k is the angular spatial frequency. The spatial frequency spectrum in the Kolmogorov model has a power-law behaviour with an exponent of $-11/3$, which means that high spatial frequencies are strongly attenuated by atmospheric turbulence.

There exist more sophisticated models which include inner-scale and outer-scale factors to improve fits between theory and experiment. In the modified von Kármán model, for example, the refractive-index power spectral density is given by:

$$\Phi_n = 0.033\, C_n^2\, \frac{\exp\left(-k^2/k_m^2\right)}{\left(k^2 + k_0^2\right)^{11/6}},$$

where $k_m = 5.92/l_0$ and $k_0 = 2\pi/L_0$. The modified von Kármán model is used in the atmospheric turbulence simulations in this study

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

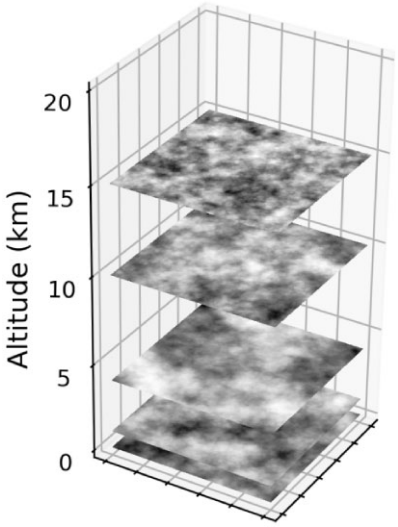

Figure 1. An example of the phase screens used in the simulation pipeline. For every simulated video sequence, each layer is initialized with a different r_0 and wind speed to account for varying seeing conditions.

because it provides a more realistic description of the statistical properties of the turbulence than the simpler Kolmogorov model, at the expense of more free parameters.

Atmospheric turbulence is usually modelled using phase screens. A phase screen is a realization of the 2D random phase distortion field introduced by turbulence at a particular altitude. In the modified von Kármán model, the statistical properties of the phase power spectral density are given by[1]:

$$\Phi_\phi = 0.023\, r_0^{-5/3}\, (f + 1/L_0^2)^{-11/6} \exp(-1.126\, l_0^2\, f), \tag{1}$$

where f is the frequency in cycles/m, and r_0 is the Fried parameter, which represents the size of the region over which the wavefront distortion due to atmospheric turbulence is roughly constant. r_0, L_0, and l_0 may differ from layer to layer. Fig. 1 shows an example of the phase screens computed in our simulation software.

To calculate the impact of the phase screens on the wavefront of the light, we leverage the mathematical framework of Fourier optics in the Fresnel regime to describe the propagation of light through a series of planes on the line of sight, starting from the highest altitude layer.

In the case of atmospheric turbulence, the ith layer carries a phase screen represented by the 2D phase distribution $\phi_i(x, y)$, where x and y are the spatial coordinates on the plane. This phase screen interacts with a monochromatic incident wavefront, characterized by a complex amplitude $W_i(x, y)$, to generate the amplitude of the emergent wavefront

$$\tilde{W}_i(x, y) = W_i(x, y) \exp j\phi_i(x, y). \tag{2}$$

In the Fresnel (paraxial) approximation, the distorted wavefront is propagated to the next layer at distance $d = |z_{i+1} - z_i|$ by convolving with the impulse response of free-space propagation:

$$W_{i+1}(x, y) = h_i(x, y, |z_i - z_{i+1}|) * \tilde{W}_i(x, y), \tag{3}$$

where

$$h(x, y, d) = \frac{e^{2j\pi\frac{d}{\lambda}}}{j\lambda d} \exp j\pi \frac{x^2 + y^2}{\lambda d},$$

[1]See e.g. Schmidt (2010) for the full derivation.

z_i the altitude of the ith layer and λ the wavelength. (3) is more conveniently computed in its 'angular spectrum' form in Fourier space (Schmidt 2010):

$$\mathcal{F}W_{i+1}(f_x, f_y) = H_i(f_x, f_y)\mathcal{F}\tilde{W}_i(f_x, f_y), \tag{4}$$

with

$$H(f_x, f_y, d) = e^{2j\pi\frac{d}{\lambda}} e^{-j\pi\lambda d(f_x^2 + f_y^2)},$$

where f_x and f_y are the spatial frequencies on the plane.

To mitigate aliasing effects in Fresnel propagation, we apodize the incident planewave from the distant point source using a cosine-tapered (Tukey) radial window (e.g. Harris 1978), with inner and outer diameters $1.5D$ and $2D$, respectively, where D is the diameter of the telescope entrance pupil.

After cascading through n atmospheric layers, the simulated wavefront reaches the telescope where the formalism remains the same except that pupil functions replace $\exp j\phi_i(x, y)$ in (2), first with a circular obstruction with diameter D' by the secondary mirror or prime focus:

$$P_{n+1}(x, y) = \begin{cases} 1, & \text{if } x^2 + y^2 \geq \left(\frac{D'}{2}\right)^2 \\ 0, & \text{otherwise} \end{cases},$$

and finally by the circular windowing from the primary mirror.

$$P_{n+2}(x, y) = \begin{cases} 1, & \text{if } x^2 + y^2 \leq \left(\frac{D}{2}\right)^2 \\ 0, & \text{otherwise} \end{cases}$$

Assuming perfect optics, the (incoherent) optical PSF on the focal plane is obtained by taking the square modulus of the inverse Fourier transform of the complex wavefront amplitude on the entrance pupil:

$$\mathrm{PSF}(x, y) = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{W}_{n+2}(x', y') e^{2j\pi\left(\frac{xx' + yy'}{\lambda F}\right)} \mathrm{d}x' \mathrm{d}y' \right|^2,$$

where F is the effective focal length of the instrument. The PSF describes the response of the telescope and atmosphere to a point source object, and can be used to estimate the resolution of the imaging system. Finally, the PSF is convolved with the intra-pixel response function, which we assume to be a perfect 2D boxcar the size of the camera pixel. Fig. 2 shows what the PSF of a star looks like when imaged with a 1 m aperture telescope through an atmosphere with different strengths of turbulence (characterized by r_0 in equation 1).

The images are sampled at the camera pixel resolution, and initially generated noiseless. A spatially constant sky background is added and a constant gain factor is applied. Finally, realizations of noisy images are generated following a Poisson–Gaussian mixture representing the shot noise of the collecting instrument plus stationary white noise from the readout electronics.

The entire simulation pipeline is written with PYTORCH so that GPUs could be maximally utilized with Fast Fourier Transforms (Brigham & Morrow 1967). This results in the capability to render ~150 000 PSFs per second, which is a couple orders of magnitude faster than other similar implementations (e.g. Hardie et al. 2017). Maximizing efficiency in the simulation pipeline is required because the data sets we generate contain hundreds of thousands of frames.

2.2 Deep learning inference of turbulence-free images

2.2.1 Model architecture

The cornerstone of our proposed method is the application of the Residual U-Net, a variant of the traditional U-Net architecture known for its proficiency in semantic segmentation and image reconstruction

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

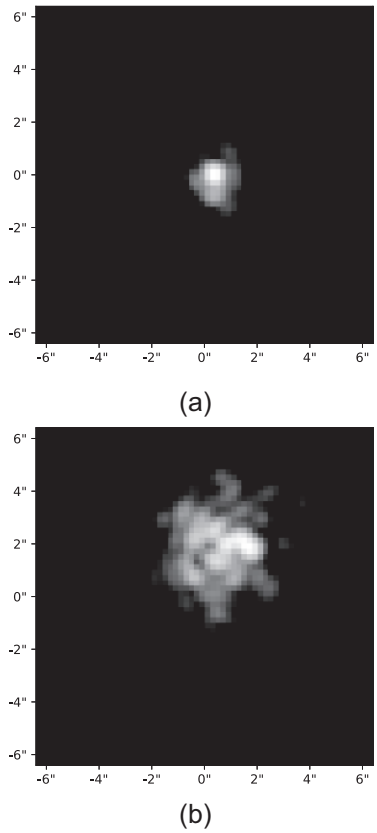

Figure 2. The PSFs of the same star imaged with a 1 m telescope in (a) mild turbulence ($D/r_0 \approx 3$) and (b) strong turbulence ($D/r_0 \approx 7$). In the case of mild turbulence, the PSF is mostly concentrated in the centre with a clear Airy pattern around it, whereas with the strong turbulence, several Fried parameter length scales can fit within the area of the telescope's aperture and so the light is distributed away from the centre in multiple 'speckles'.

tasks (Ronneberger, Fischer & Brox 2015; Çiçek et al. 2016; Yao et al. 2018; Zhang, Liu & Wang 2018; Mizusawa et al. 2021). The model used in this study, developed by the Medical Imaging Computing Group at the German Cancer Research Center (MIC-DKFZ)[2], is part of a widely-used Python package used for deep learning-based biomedical image segmentation (Isensee et al. 2021). It was selected based on its unique architecture and inherent properties – detailed below – that make it suitable for our image reconstruction task.

A U-Net can be defined as a combination of a contracting path (encoder) and an expansive path (decoder), bridged by a bottleneck which helps us to reduce the computational complexity of the model. The encoder performs consecutive convolutions and downsampling to output *feature maps*, learning the contextual information while decreasing the spatial dimension of the input. The expansive path, on the other hand, performs transposed convolutions on the feature maps and then concatenates them with the corresponding feature maps from the encoder (skip connections), allowing the decoder to merge high-level features with preserved local information.

The U-Net is composed of stages, each containing a certain number of *blocks*. Each block includes convolutional layers with activation and normalization functions – in our case, the leaky variant of the Rectified Linear Unit (`LeakyReLU`) activation and instance normalization (`InstanceNorm`) were used, noting that `InstanceNorm` was chosen for its ability to perform well on smaller batch sizes (Kolarik, Burget & Riha 2020), which we were limited to due to GPU memory constraints. The encoder block reduces the spatial dimension while increasing the feature channels progressively, and the decoder block does the inverse operation. In our U-Net, we used two blocks per stage.

The residual U-Net uses *residual* blocks in the encoder to help alleviate the vanishing gradient problem (He et al. 2016; Borawar & Kaur 2023) and allow for deeper networks. For an input feature map X, each residual block in the architecture performs two convolutional operations F_1 and F_2, separated by a `LeakyReLU` activation function. The output of these operations is added to the input feature map, and another `LeakyReLU` function is applied. This can be mathematically represented as:

$$Y = g_2(F_2(g_1(F_1(X))) + W(X)), \tag{5}$$

where F_1 and F_2 denote 2D convolutional operations, W is a transformation operation that adjusts the input feature map's dimensions and/or number of channels to match $F_2(g_1(F_1(X)))$, and g_1 and g_2 represent the `LeakyReLU` functions. Each layer consists of two such residual blocks along with down- or upsampling operations.

Following the principle of *deep supervision* (Wang et al. 2015; Li et al. 2022), which enables the backpropagation of gradients from the deeper layers of the network to the earlier layers more effectively, our approach compares the output from each decoder block with a correspondingly downsampled version of the ground truth. The loss is calculated for each comparison, and the overall loss is determined by a weighted sum of these individual losses; the weight assigned to each downsampled image is set to be half of that of the preceding image in the sequence. The entire network is shown schematically in Fig. 3.

2.2.2 *Training configuration*

During training, as a batch of data is loaded, the background is first subtracted by iteratively determining the median and excluding points that deviate more than three times the mean absolute deviation (MAD) from the median. Such an approach effectively minimizes the impact of outliers, such as stars, in the estimation of a statistically robust background level, and ensures the varied background levels of any test data will not negatively influence predictions.

When working with image reconstruction or denoising, it is often beneficial to transform the input data such that the noise level is approximately constant across the image. This can help algorithms, like those used in U-Net architectures, to perform more consistently. We performed experiments with cross-validation on various scaling transformations. After evaluating the results, in particular the flux conservation, we converged on the Anscombe transformation for both the input and target images. Not only does it help stabilizing the variance, making the noise homoscedastic in the Poisson regime, but it also amplifies signals in dim regions, enhancing the visibility of faint features. The transformation is given by:

$$f(x) = 2\sqrt{x + \frac{3}{8}}$$

For its inversion, we adopt the closed-form approximation of the exact unbiased inverse as defined in Makitalo & Foi (2011):

$$f^{-1}(x) = \frac{1}{4}x^2 + \frac{1}{4}\sqrt{\frac{3}{2}}x^{-1} - - - \frac{11}{8}x^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}x^{-3} - \frac{1}{8}$$

The application of the Anscombe transform in the context of deep supervision, particularly during downsampling of the target frames,

[2]https://github.com/MIC-DKFZ/dynamic-network-architectures

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

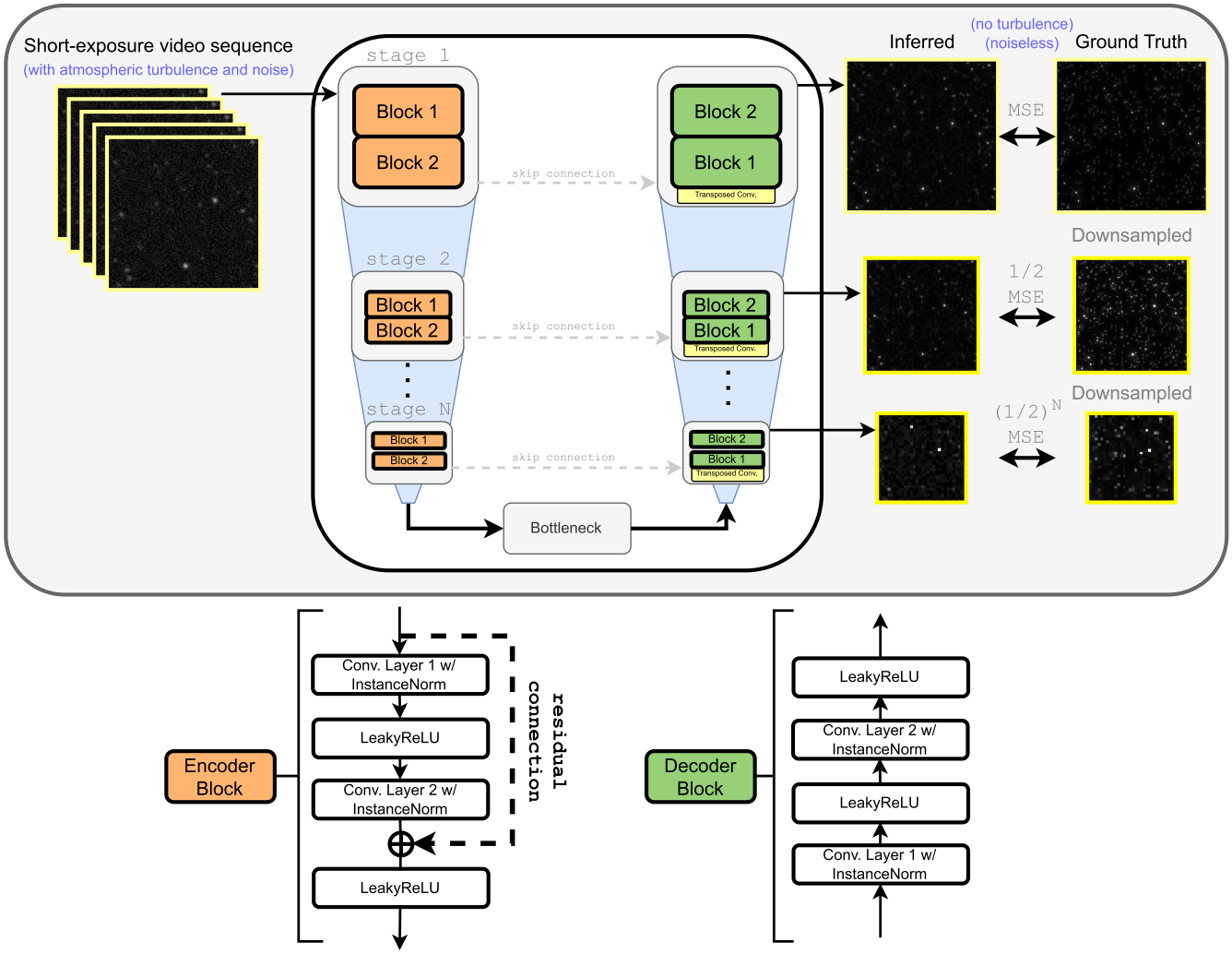

Figure 3. The DanceCam Residual U-Net architecture. A set of simulated short-exposure video streams of stellar fields – with turbulence and noise – along with their corresponding ground truth frames – with no turbulence or noise – is used to train the model. Instead of a single output, the model additionally has outputs from each stage in the decoder which are compared to downsampled versions of the ground truth using a weighted mean-squared error (MSE) loss function. Once trained, either a simulated or real video stream can be used as input and only a single (not downsampled) inferred image is retrieved.

presents a methodological concern. Downsampling is typically more effective on linear, untransformed data, yet our approach involves downsampling on Anscombe-transformed data. This raises questions about the potential need for an alternative method, such as downsampling untransformed data and then comparing it to de-transformed outputs from the U-Net, while also considering the implications for the U-Net decoder's handling of dynamic range transformations at each resolution. Alternative strategies will be considered in future iterations of the proposed method; as it stands, the current implementation was deemed sufficient.

For the optimization process, we employed the adaptive moment estimation (Adam) optimizer, a popular first-order stochastic gradient descent algorithm using estimates of the first and second moments of the gradients (Kingma & Ba 2014). In our implementation, the default values were used such that the initial learning rate was set to 0.001, and beta values of 0.9 and 0.999 were used for the exponential decay rates of the gradient and squared gradient, respectively.

To expedite the training process, we utilized multiple GPUs – specifically 32GB NVIDIA Tesla V100s – made available by the Digital Research Alliance of Canada on their cluster. On average, each epoch took approximately 14 min to complete (noting that a forward pass of a single example took fractions of a second). The mean-squared error (MSE) loss was used, and a learning rate scheduler was incorporated which adjusted the learning rate when the training MSE loss plateaued.

Since we were particularly concerned with the image reconstructions conserving the flux of stars, `SExtractor` (Bertin & Arnouts 1996) was run on the inferred and ground truth images in the validation set after every epoch to track the magnitude estimates of detected stars. The final model was chosen such that the mean absolute error between the magnitudes of the inferred and ground truth stars was minimized. We note that astrometric precision was not a criterion during training, but including it in future iterations of the method may help improve overall astrometric performance.

3 DATA

3.1 Synthetic data generation

Each stellar field of a video sequence in a training data set was created with one of two methods:

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

(i) *Homogeneous field*: The number of stars to be simulated is uniformly sampled from [3, 2000] and placed on a given uniform randomly generated x and y coordinates. An exponential law is used to generate magnitudes for each star: $p(m) \propto 10^{\alpha(m-m_{\max})}$, where $m_{\max}$ was 21 in this study and α (the slope of the differential source counts, i.e. dlog N / dmag) was 0.4. The inclusion of homogeneous fields helps the model generalize better by learning to handle a diverse set of scenarios with uniform sampling of dense and sparse fields.

(ii) *Realistic field*: To include more realistic priors in our simulation pipeline, we generate the fields according to the distributions of real star clusters of the Milky Way. The catalogue from Kharchenko et al. (2013) includes the sky coordinates of thousands of clusters, which we use to query the Gaia DR3 data base (Prusti et al. 2016; Vallenari, Brown & Prusti 2022) to obtain the RA, Dec, and `phot_g_mean_mag` values of each star. A random offset in both RA and Dec is uniformly sampled from $[-FOV/2, FOV/2]$, where FOV is the field-of-view of the telescope, and applied to get a shifted image, i.e. a more diverse training set.

The data set therefore contains both randomly generated and more realistic stellar fields, noting that the inclusion of extended objects (like galaxies) will be included in future iterations of the method (see Section 5.1).

We generated a training data set containing 40 000 12 s video sequences; 12 s was chosen as a compromise between GPU memory constraints and collecting enough information about the turbulence and faint stars. Each frame is 256×256 pixels and, to match the properties of the C2PU Telescope (see Section 3.2), we used a 1-m diameter telescope aperture, a central wavelength of 650 nm, a pixel scale of 0.235 arcsec pixel^{-1}, a readout noise standard deviation of $1e^-$, and a sampling rate of 5.25s^{-1} (i.e. sampled every ∼200 ms) – noting that while this sampling rate is generally insufficient at capturing the quickly evolving turbulence, it does offer an advantage of having higher signal-to-noise in each frame, and in any case we plan on increasing the sampling rate in future iterations of the method. In each video sequence, the Fried parameter and wind speeds for each layer of the atmosphere are sampled from a normal distribution. Along with each video sequence, we generated the corresponding ground truth frame in which we disabled contributions from the atmosphere and any sources of noise in our simulation pipeline. The data sets were split 90 per cent/10 per cent into training/validation sets.

3.2 C2PU telescope

The Centre Pédagogique Planète et Univers (C2PU) facility (Bendjoya et al. 2012) is located on the Plateau de Caussols at an elevation of 1260 m, approximately 50 km from Nice in southern France. The site benefits from good seeing in summer time, with a median of 1.06 arcsec, down to 0.8 arcmin at the end of the nights (Aristidi et al. 2020).

On the night of 2022 May 27, we collected around 60 s of short-exposure (200 ms) images of the globular cluster M92 in the SDSS *r* bandpass from the wide-field camera installed at the prime focus of the C2PU Omicron 1.04 m telescope. No guiding was used as the tracking accuracy of the telescope allows for unguided exposures up to a few minutes without image degradation. The wide field camera provides a 37.6 arcmin × 25.2 arcmin field of view with excellent image quality, through a three-lens Wynne coma corrector, at a resulting F/3.17 focal ratio. One important feature of this optical setup for our project is its relatively low obstruction (30 per cent linear) for a wide-field instrument, which preserves more than 80 per cent of the central peak intensity of the telescopic point spread function, compared to an unobstructed aperture. The camera at the C2PU Omicron prime focus is a QHY600Pro equipped with a Sony IMX 455M sCMOS sensor. Despite its modest quantum efficiency at redder wavelengths, this generation of sensors has proven to be competitive for quantitative astronomy (Alarcon et al. 2023; Betoule et al. 2023). The sequence images have a size of 1024×1024 pixels and a pixel scale of 0.235 arcsec.

4 EXPERIMENTAL EVALUATION

Visually, the proposed method does an excellent job at taking in a short sequence of turbulent images and producing a clear, sharp, noise-, and turbulence-mitigated image. For example, Fig. 4 shows the massive improvement – compared to temporally averaging the sequence – in image quality of frames inferred from sequences with good (0.7 arcmin) and bad (1.4 arcmin) seeing conditions. Some stars which are barely visible (or not visible at all) in the averaged frame appear quite clearly in the inferred frame.

Validating the method *quantitatively* in addition to qualitatively, however, required analysing the performance metrics of a large number of simulated observations, as detailed in the following section.

4.1 Quality assurance

To quantitatively assess the performance of the Residual U-Net, a series of quality assurance tests were implemented. These tests were designed to evaluate the improvements in image quality, flux conservation, and astrometric stability in comparison to simply stacking the images.

A 30-s video sequence of a stellar field was simulated following the *random field* process described in Section 3.1, including the random sampling of Fried parameters and wind speeds for each atmospheric layer. We iteratively step along the sequence one frame at a time and partition a 12-s subset to obtain an inferred U-Net frame until a full sequence of inferred frames was collected. The inferred sequence was stacked by taking the mean along the temporal dimension, resulting in a single frame representing a 30-s inferred observation. The same 30-s sequence was used to obtain a single averaged frame. We additionally saved the ground truth frame (simulated with no turbulence or noise) for each sequence. The entire procedure was repeated 500 times, resulting in a 'quality assurance' data set of 500 inferred frames, 500 averaged frames, and 500 ground truth frames.

`SExtractor` was then used on all of the frames in the data set to identify the stars[3] and calculate for each one the magnitude (by using an aperture diameter of 10 pixels to measure the total flux), the angular diameter (in terms of the D50, which measures the diameter at which 50 per cent of the light is contained within an aperture), and the x-position, for the purpose of showing flux conservation, improved image quality, and astrometric stability, respectively. Fig. 5 shows the residuals of these estimates, as a function of magnitude, on the inferred and averaged frames compared to the ground truth frames.

The magnitude estimates of stars from the inferred frames can be seen to have overall a smaller dispersion throughout the magnitude range than stars in the averaged stack [which is evidenced by the

[3]For the sake of consistency across the vastly different regimes of source density and image quality, in this work we used fixed, conservative source extraction settings for `SExtractor` except where otherwise noted.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

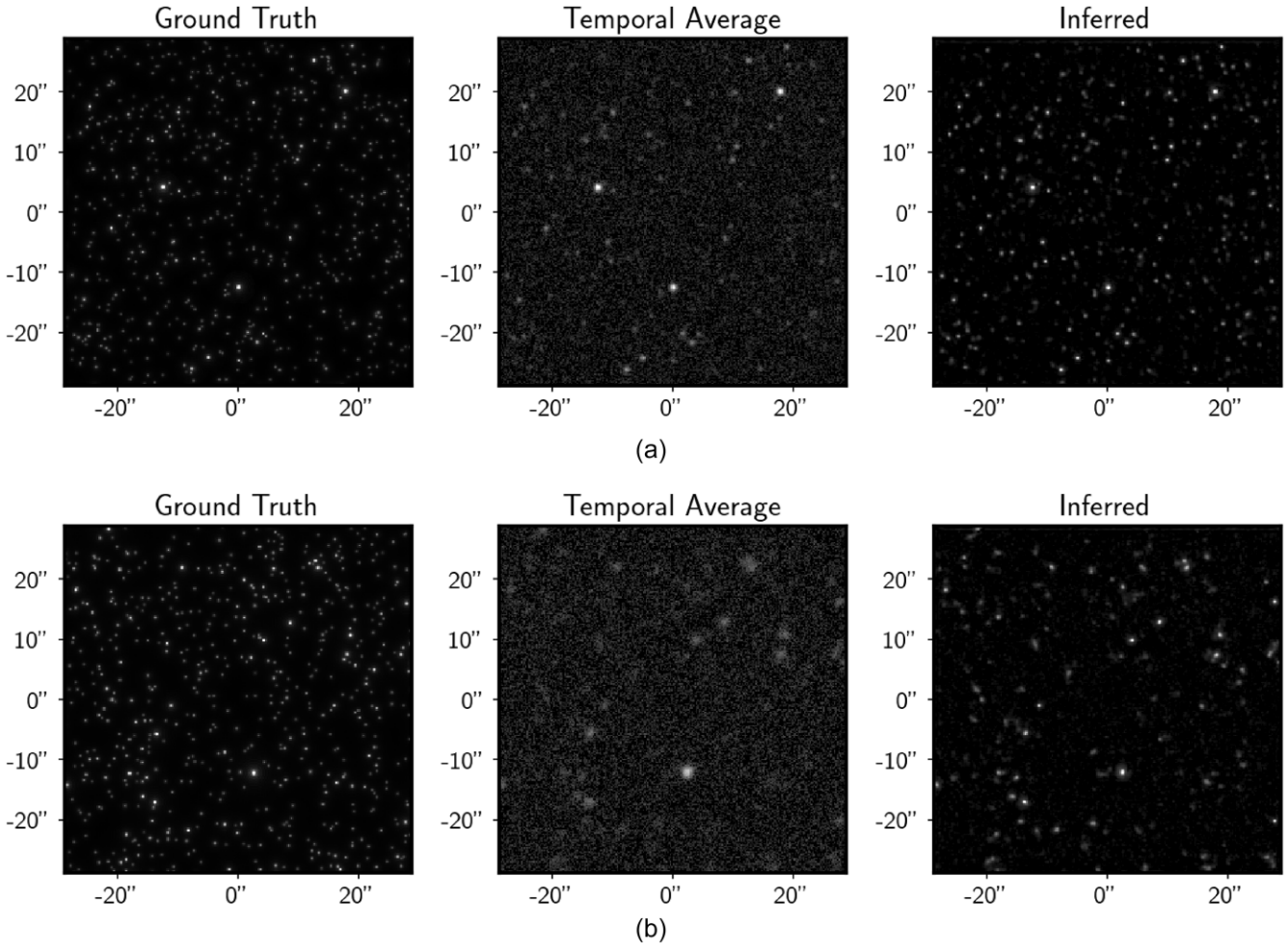

Figure 4. Two examples highlighting the ability of the proposed method to remove the effects of atmospheric turbulence and produce a sharp, clear image. 6-s sequences of random stellar fields were simulated with (a) 0.7 arcmin seeing and (b) 1.4 arcmin seeing, and the ground truth, temporally averaged sequence, and inferred frames are shown here.

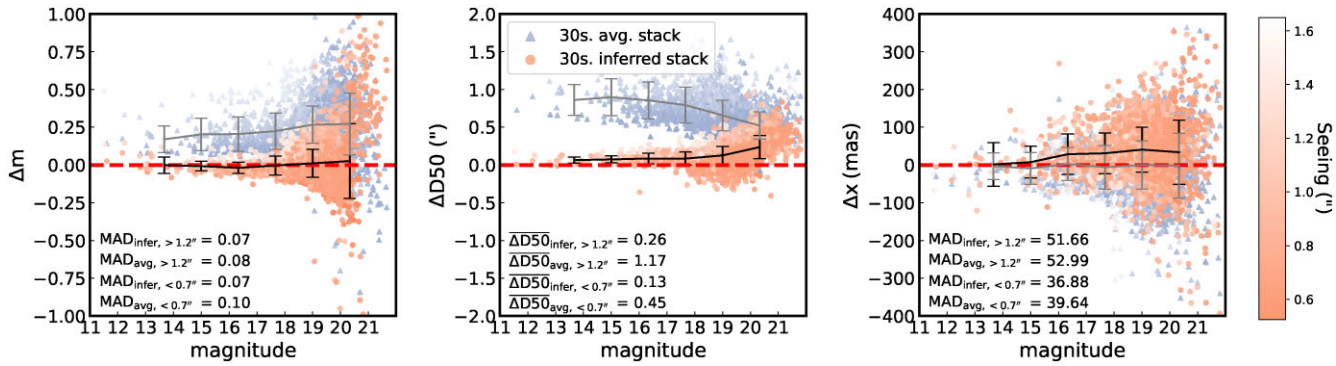

Figure 5. A series of quality assurance tests were made to validate the image reconstructions made by the U-Net. Hundreds of 30-s simulated observations of random stellar fields, with varying seeing conditions, were created and two images were made for each example: a stack made from the U-Net inferred images and a simple averaged stack of the raw frames. `SExtractor` was run on each frame, along with their corresponding ground truth frame, to collect information about each detected star's (*left panel*) magnitude – using a 10 pixel diameter aperture – to test for flux conservation, (*middle panel*) source sizes – defined as the diameter of the aperture within which 50 per cent of the light from a star is contained (D50) – to test for improvements in image quality, and (*right panel*) centroid coordinates, to test for astrometric stability. Shown here are the residuals of those metrics for the inferred stack (red circles) and simple averaged stack (blue triangles) when compared to the matching stars in the ground truth frames as a function of magnitude, along with their binned means and standard deviations (shown as error bars). Also shown are the computed means for 'bad seeing' and 'good seeing' subsets of the data (> 1.2 arcmin and < 0.7 arcsec, respectively). The 'fainter fatter' effect can be seen for the inferred stacks in the D50 figure, wherein the U-Net tends to smooth the fainter stars (see Section 5.1 for a further discussion).

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

median absolute deviation (MAD) being 0.01 arcsec–0.03 arcmin lower across the seeing range]. It is also the case that the magnitude residuals are closer to ground truth for the inferred frames, but in practice this could easily be resolved by using an adaptive aperture size on the averaged frames; we emphasize the importance of the magnitude residual plot is showing that flux is conserved in the inferred frames with lower dispersion than for the averaged frames. The lowered dispersion is also shown in the astrometric stability plot, where the MAD is 1–3 mas less for the inferred frames than the averaged frames, a very modest reduction but importantly there is no apparent degradation in astrometric stability. The strength of our proposed method is showcased in the angular diameter plot, where the average D50 residual is a factor of 3 –4x lower in the inferred frames than the averaged frames, and for bright stars can even be a factor of about 8× less.

In all cases, the performance on the inferred frames gets worse towards fainter magnitudes: at a magnitude of 16, the standard deviations for magnitude residuals and D50 residuals are 0.01 and 0.07 arcsec, respectively, whereas at a magnitude of 19, the residuals increase to 0.08 and 0.1 arcsec. It appears the U-Net tends to 'smooth' fainter stars, leading to a 'fainter fatter' effect (as opposed to the more common 'brighter fatter' effect for CCDs), further discussed in Section 5.1.

4.2 Test case: M92

After ensuring that the trained model performed well in conserving flux, increasing image quality, and stabilizing the astrometry, it was time to task it with a more realistic scenario: inferring the turbulence-free image of the globular cluster M92. With the high stellar densities in the core of the cluster, this would be a true test of our method's ability to disentangle the light from nearby stars and improve the spatial resolution. Testing was performed on both simulated data and real data obtained with the C2PU telescope.

4.2.1 Simulated data

We began the test case of M92 with a simulated observation of it (30 seconds total @ 5.25 frames/second for a total of 160 frames) to obtain a baseline to compare to. A wide field of view (∼4 arcmin, 1024 × 1024 pixels) was used to showcase the ability of our method to perform on larger images than it was trained on. As described in Section 3.1, *Gaia* coordinates and magnitudes were used for each star. We set a maximum magnitude of $G = 21$, consistent with the model's training parameters, leading to a total of ∼12 000 stars being simulated. A relatively poor seeing of 1.36 arcmin was chosen to test the method's proficiency under suboptimal conditions.

Since the model was trained on frames of size 256 × 256 pixels, the full 1024 × 1024 frames of M92 could not be used as input. Instead, each frame was split into tiles of size 256 × 256 with 50 per cent overlap, resulting in the full 1024 × 1024 ×160 video sequence being split into 49 overlapping sequences of size 256 × 256 ×160. As described in Section 4.1, the U-Net was used on each of these sequences to produce a series of inferred images along the temporal axis which were then averaged together to produce a single 256 × 256 inferred image, thereby reducing the 49 sequences of size 256 × 256 ×160 into overlapping inferred tiles of size 256 × 256. In the final reconstruction, the overlapping regions of these tiles were blended smoothly: masks were generated for each tile, consisting of a central region with full contribution (value 1), and edge regions with a linear gradient from 0 to 1, reflecting the degree of overlap. The final image B at each pixel was reconstructed by calculating the normalized weighted sum of the n overlapping tiles, expressed as $B(x, y) = \frac{\sum_{i=1}^{n} T_i(x,y) \cdot M_i(x,y)}{\sum_{i=1}^{n} M_i(x,y)}$, where $T_i(x, y)$ and $M_i(x, y)$ are the value and mask at position (x, y) in the ith tile, respectively. This ensured a seamless integration of the tiles, with the centre of each tile retaining its original value and the overlapping edges merging smoothly, resulting in the final 1024 × 1024 inferred image.

Fig. 6 shows the ground truth frame, the inferred frame, and the temporally averaged frame, along with zoomed in images of the central region of M92 which show, with red circles, stars identified with `SExtractor`. It is visually clear that the inferred frame is far sharper and had significantly more stars identified than the averaged frame, so a more quantitative analysis was conducted to confirm this.

Fig. 7 shows the relative number of stars identified by `SExtractor` in both the inferred and averaged frames in bins of magnitude, where true positives were defined as being no more than 2 pixels away from the corresponding star in the ground truth frame. We note that the detection threshold for `SExtractor` was optimized on each frame individually to maximize both the precision and recovery rate of its star identification. For the averaged frame, the precision was 97.1 per cent (i.e. 2.9 per cent false positive rate), and 63.1 per cent of stars were successfully recovered. In contrast, for the inferred frame, the detection precision was 97.8 per cent (i.e. 2.2 per cent false positive rate), and 86.2 per cent of stars were successfully recovered, verifying that substantially more stars were identified in the inferred frame.

To test for the improvements in image quality, the D50 measurements from `SExtractor` were used and, because it is a more common diagnostic metric in AO analyses, the Strehl ratio, S, was calculated for each star according to Roberts Jr et al. (2004):

$$S = \frac{I(x = 0)}{\sum I} \frac{\sum P}{P(x = 0)},$$

where x is the position vector, $I(x = 0)$ is the maximum of the measured PSF, $P(x = 0)$ is the maximum of the diffraction limited PSF, and $\sum I$ and $\sum P$ are computed over squares of size 10 × 10 pixels (centred on the coordinates extracted by `SExtractor`) and used to normalize the ground truth PSF to have the same total intensity as the observed PSF. This Strehl ratio estimator tends to be noisy due to the statistical noise and crowding, but we are more interested in *relative* improvements between the average and inferred stacks. Fig. 8 shows how these values change over the magnitude range. Below a *Gaia* G magnitude of 17.5, the inferred stars have a 3× average reduction in D50 and an average of ∼6× improvement in Strehl ratio. Beyond a magnitude of 17.5, the performance on inferred stars drops yet still maintains a 1.5–3× improvement on D50 measurements, and 2–4× improvement on Strehl ratios; again we find that the model struggles with fainter stars, tending to smooth their PSFs.

4.2.2 C2PU telescope data

All of the steps in 4.2.1 were repeated for the real M92 data collected by the C2PU telescope. Fig. 9 shows a comparison between the ground truth (simulated with *Gaia* coordinates and magnitudes), inferred, and temporally averaged images of a 30 s observation of M92 containing 160 total frames of size 1024×1024. The zoomed in areas of the central regions of M92 show that the brighter stars in the inferred image appear substantially sharper, though there are clearly some spurious effects in the fainter parts; in the video stream of M92, there were indications that an intermittently very turbulent ground

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

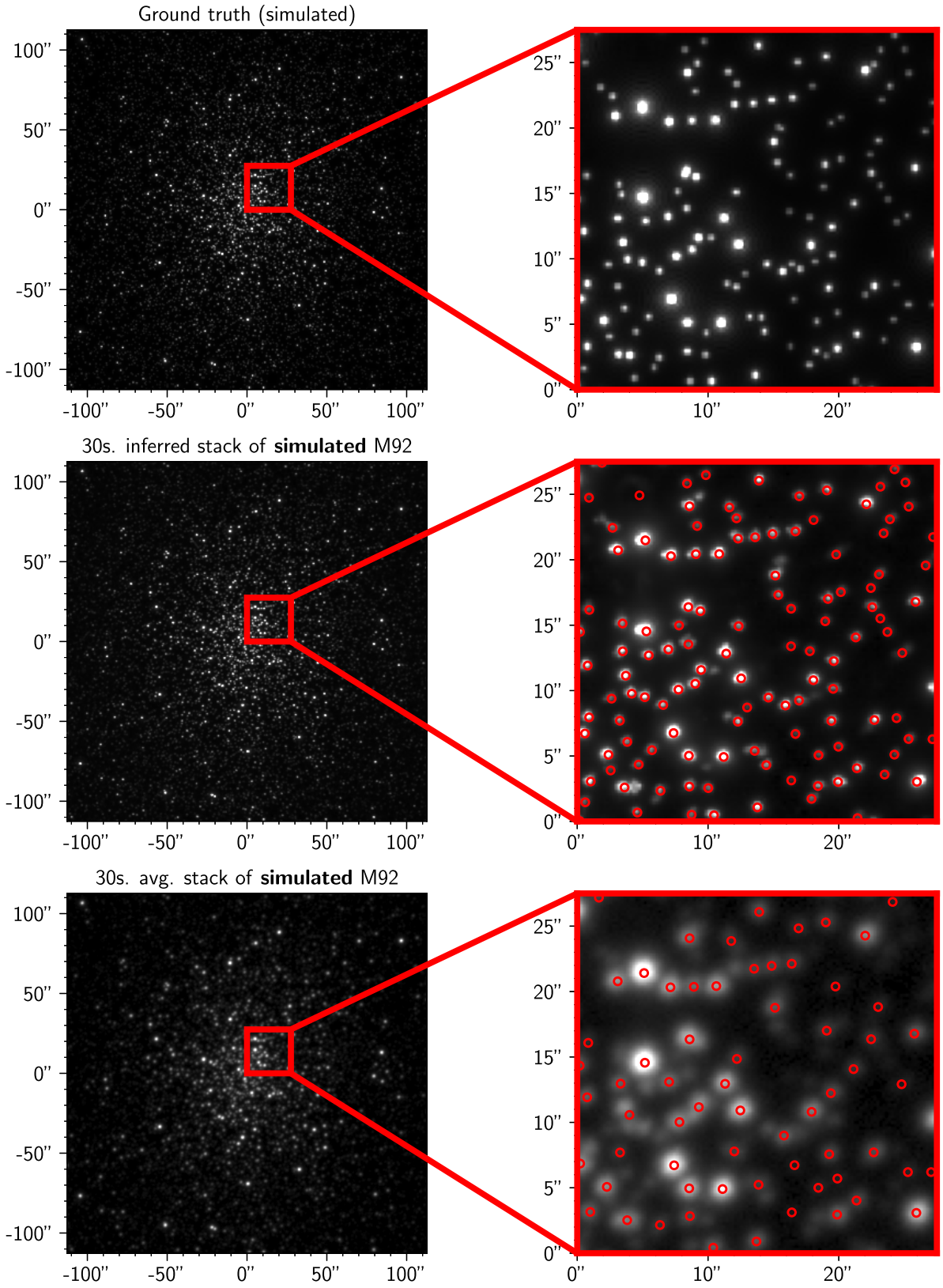

Figure 6. Our simulation pipeline was used to create a 30 s video stream – 160 frames in total @ 5.25 frames s^{-1} – of the globular cluster M92 using *Gaia* positions and G magnitudes and a total seeing of 1.36 arcsec. Shown here is a comparison of the simulated ground truth, a stack made from inferred images, and a simple averaged stack of the frames. The circles indicate stars that were detected by `SExtractor`, using conservative detection settings.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

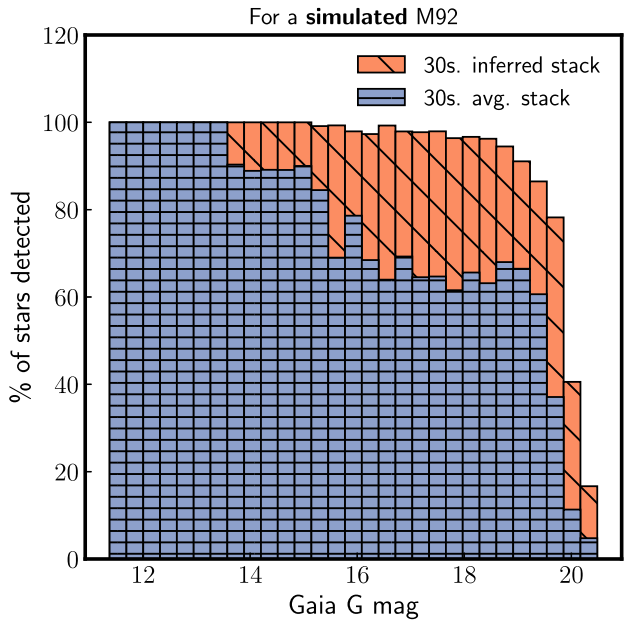

Figure 7. `SExtractor` was run on the ground truth frame, inferred stack, and simple averaged stack of the simulated M92 shown in Fig. 6. The stars that were identified in the inferred stack and averaged stack were matched to the stars in the ground truth frame (by ensuring their measured positions differed by less than 2 pixels), and shown here is the completeness of the detected stars as a function of *Gaia* G magnitude. Note that the false positive rate for both frames was ∼2 per cent, i.e. of all the stars `SExtractor` identified in the inferred and averaged frames, 2 per cent were not matched to those in the ground truth frame.

layer was causing all the stars in the field to move in lockstep by up to a couple arcseconds, an effect which is not currently accounted for in the simulations. This and other limitations, as discussed in Section 5.1, contributed to a decreased visual performance on real data.

As for the number of stars correctly identified, Fig. 10 shows that slightly more stars – throughout the entire magnitude range – were found in the inferred frame than the averaged frame (∼9.6 per cent versus ∼8.7 per cent of the total stars in the ground truth image), however the total number of stars recovered in either frame, as well as the precision of the detections, depends on the detection threshold used in `SExtractor`. Fig. 11 shows the relationship between the precision and total number of stars correctly identified as a function of detection threshold, where it can be seen that better performance (a maximum of a few per cent more stars recovered for a given precision) is achieved for the inferred frame. We also investigated *where* the stars were being more readily identified in both images. Fig. 12 shows that the inferred image works particularly well at deblending the crowded central regions of M92, recovering about 3× more stars within 25 arcsec from the centre than the averaged image.

Finally, Fig. 13 shows quantitatively how the D50 and Strehl ratio measurements on the inferred and averaged images change across the magnitude range. Again we find that, for the inferred image, the performance on brighter stars (G < 17) is enhanced, with an average of 2.5× reduction in D50 measurements and 5× increase in Strehl ratio measurements. For fainter stars, the performance drops in a similar manner as in the case of a simulated M92, leading to only slightly decreased average PSF width than the brighter stars.

The results from the analysis on the synthetic and observed M92 are summarized in Table 1, including the results from using a smaller temporal input context for training and testing the U-Net (described in Section B). Additionally, results from training on single temporally averaged frames are discussed in Section A.

5 DISCUSSION

The proposed method introduces a novel approach to handling the challenges posed by atmospheric turbulence in astronomical imaging, especially when dealing with wide fields. By segmenting the observed field into smaller overlapping tiles, producing an inferred frame for each tile, and merging the inferred tiles, the method allows for an effective mitigation of atmospheric turbulence effects in arbitrarily large images without compromising the quality of turbulence correction.

An important feature of the DanceCam approach is its independence from the traditional reliance on guide stars, instead using the information from the entire field to understand the turbulence properties. While AO systems are tethered to either natural or artificial guide stars for measuring and correcting atmospheric distortions, our method's flexibility broadens its applicability, offering the potential to correct turbulence even in regions where suitable guide stars might be elusive. Moreover, the ability to analyse distortions in captured images paves the way for reverse-engineering the atmospheric layers and their respective turbulence properties. Such insights can be beneficial for our understanding of atmospheric conditions and refining observation strategies.

Modern technological advancements further strengthen the efficacy of the method. With the requirement of short exposures to capture the near real-time turbulence, most CCDs would produce images with too much readout noise, effectively drowning out the faint targets. The integration of sCMOS cameras, on the other hand – renowned for their rapid readout speeds, reduced noise, and enhanced quantum efficiency (Guidash et al. 2016; Zhang et al. 2020; Zhu et al. 2022) – complements the turbulence mitigation technique. Modern GPUs, additionally, allow for an impressive processing speed; the time required to produce a single inferred frame from a fully trained model is mere fractions of a second, allowing for the correction of a video stream comprising hundreds of short-exposure wide-field images in near real-time.

The proposed method is also appealing due to its minimalistic hardware requirements. Eschewing the intricate setups of traditional AO systems, which involve expensive deformable mirrors and wavefront sensors, as well as significant overheads, our approach is predominantly software-centric. This orientation not only simplifies its implementation but also offers significant cost savings, making it an attractive solution for budget-conscious observatories and researchers, and even amateur astronomers.

5.1 Limitations and known issues

Acknowledging the current limitations and exploring potential avenues for improvement of the proposed method is essential for its efficacy and ensuring its applicability in future astronomical observations:

(i) **Intermittent turbulence:** One of the primary constraints is the absence of intermittent turbulence in the current implementation. Intermittent turbulence can introduce sporadic and unpredictable distortions, which the method might not handle effectively in its present form. Interestingly, it could even enhance the method's

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

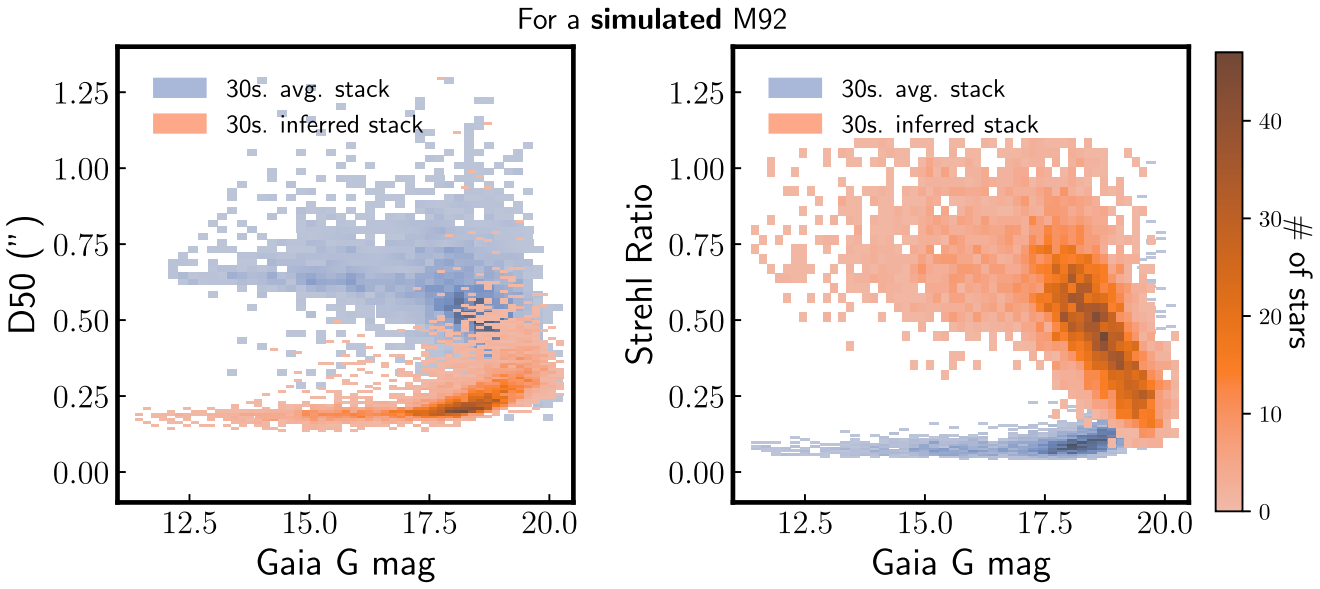

Figure 8. `SExtractor` was run on the inferred stack and simple averaged stack of the simulated M92 shown in Fig. 6. Shown here are the distributions of the angular diameter of the aperture inside of which 50 per cent of the flux is contained (D50, left plot) and Strehl ratio estimates (right plot) of the detected stars as a function of *Gaia* G magnitude.

effectiveness by occasionally providing 'lucky frames' in a sequence. A straightforward solution could involve incorporating dynamic changes to r_0 for each atmospheric layer in the simulation pipeline.

(ii) **Motion blurring:** Another limitation is the lack of motion blurring implementation. In real-world scenarios, the shutter collects light over a duration (e.g. 0.2 s), whereas the simulation pipeline currently takes instantaneous snapshots at regular intervals. Addressing this could involve splitting each frame into N sub-frames, simulating turbulence at every sub-time-step, and averaging the sub-frames together. This approach would effectively simulate an open shutter, offering a more realistic representation. Alternatively, using shorter exposures could also address this limitation.

(iii) **Monochromatic light:** The method's current reliance on simulating monochromatic light simplifies the computational process but overlooks chromatic aberrations introduced by atmospheric turbulence. Enhancing the simulation pipeline to cover a broader band-pass of light could improve the method's accuracy.

(iv) **Readout noise:** In the current scheme, we assume that readout noise follows a Gaussian distribution. However in sCMOS sensors each pixel has its own amplifier circuit, and therefore its own noise characteristics; some pixels are markedly noisier than others, and the overall histogram of a real uniformly illuminated sCMOS exposure is significantly non-Gaussian.

(v) **Rolling shutter:** Today's astronomical sCMOS sensors are usually operating in rolling shutter mode: lines of pixels are read and reset in sequence. The efficiency of photon collection is thereby optimized, but this has the inconvenience that two successive sensor lines are recorded at slightly different times. The maximum time difference amounts to only a few milliseconds inside the current 256 × 256 analysis window. However this delay would reach ∼100 ms if distant parts of the full sensor were to be analysed jointly, in which case it could not be neglected anymore.

(vi) **GPU constraints:** Temporal context is currently limited due to GPU memory constraints. Optimizing GPU usage, such as parallellizing the pipeline to distribute computational load across multiple GPUs, could address this limitation. Additionally, as GPUs continue to technologically advance, the method's capabilities will naturally expand.

(vii) **Only simulating stars:** A major limitation of the method is that it is currently tailored to simulate stars only, excluding other celestial objects like nebulae, cirri, galaxies, or planets. Integrating diverse light profiles and spatial structures associated with a variety of astronomical objects would enhance its scope and versatility.

(viii) **Transient objects:** The current system does not handle transient objects that appear temporarily and unpredictably, e.g. satellites or other moving objects, or objects that change in brightness. These types of objects could be included in the simulation pipeline. Alternatively, or additionally, owing to the high cadence video streams, frames containing transients of interest could be detected and saved for later processing using automated methods (e.g. Cabrera-Vives et al. 2017; Gieseke et al. 2017). Deleterious transients (specifically telecommunications satellites) could be removed before processing. This is indeed a major benefit of the video stream approach to observations, since with long exposures the best you can do is mask satellite streaks (e.g. Paillassa, Bertin & Bouy 2020). See Beskin et al. (2023) for a more in-depth look at the benefit of wide-field, high cadence imaging for transient detection and analysis.

(ix) **High-volume data streams:** Given our focus on wide fields and short exposures, the method generates a significant volume of data, which can be challenging to process and store efficiently. To manage these massive data streams, it will likely be crucial to process and analyse the collected data in quasi-real-time using one or several GPUs, helping to filter and prioritize data for storage and detailed examination.

(x) **Circular stationarity of phase screens:** In the current scheme, we initialize a static phase screen and 'roll' the values along for each time-step, where the values at the edge roll over to the opposite edge. This approach, while efficient, is not a problem when using short temporal contexts for training. However, it could become increasingly problematic when using longer video streams, as it introduces artificial periodicity and may not accurately represent

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

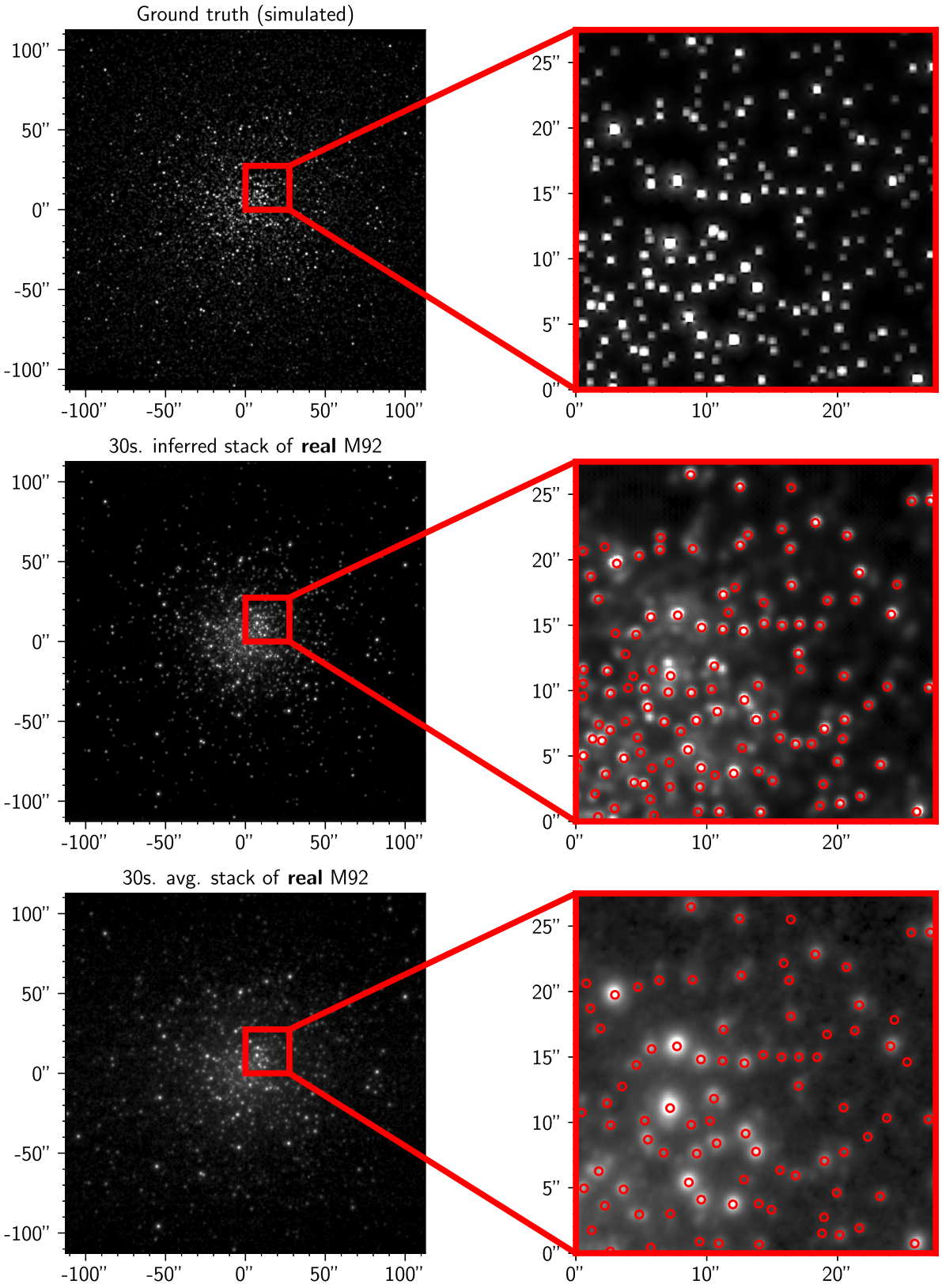

Figure 9. The C2PU telescope was used to obtain a 30 s video stream – 160 frames in total @ 5.25 frames s^{-1} – of the globular cluster M92. Shown here is a comparison of the simulated ground truth using *Gaia* positions and magnitudes, a stack made from inferred images, and a simple averaged stack of the frames. The circles indicate stars that were detected by `SExtractor` using conservative detection settings. Note that some relatively bright stars are missing from the *Gaia* catalogue in these crowded areas.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

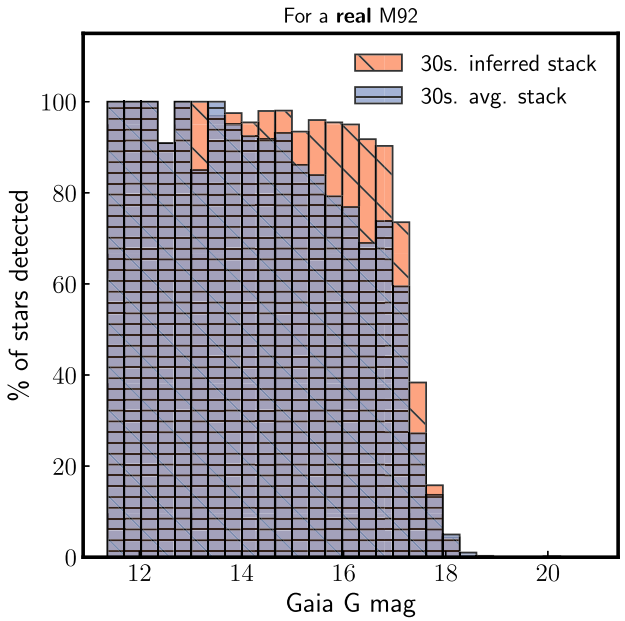

Figure 10. `SExtractor` was run on the ground truth frame, inferred stack, and simple averaged stack of the real M92 shown in Fig. 9. The stars that were identified in the inferred stack and averaged stack were matched to the stars in the ground truth frame, and shown here is the completeness of the detected stars as a function of *Gaia* G magnitude. Note that most of the stars in M92 are fainter than G ∼ 18, so the majority of stars not detected had G > 18.

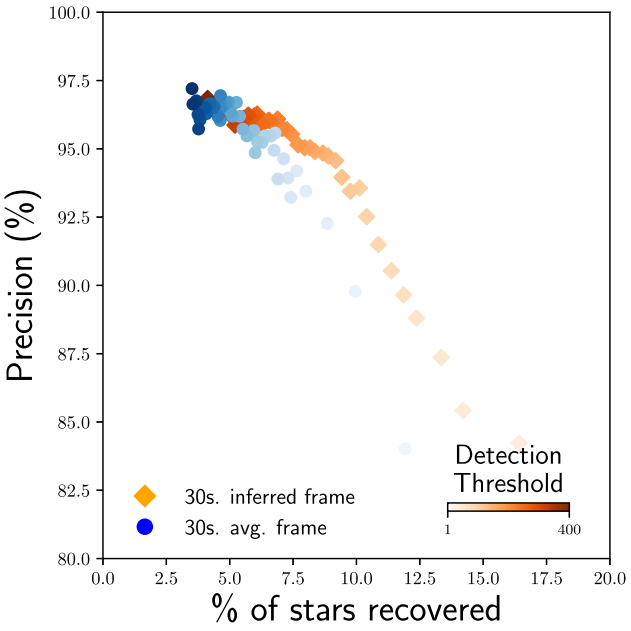

Figure 11. The chosen detection threshold of `SExtractor` affects the precision of detected sources and total percentage of recovered stars. Shown here are the results of changing the detection threshold on a 30 s averaged frame and a 30 s inferred frame of M92.

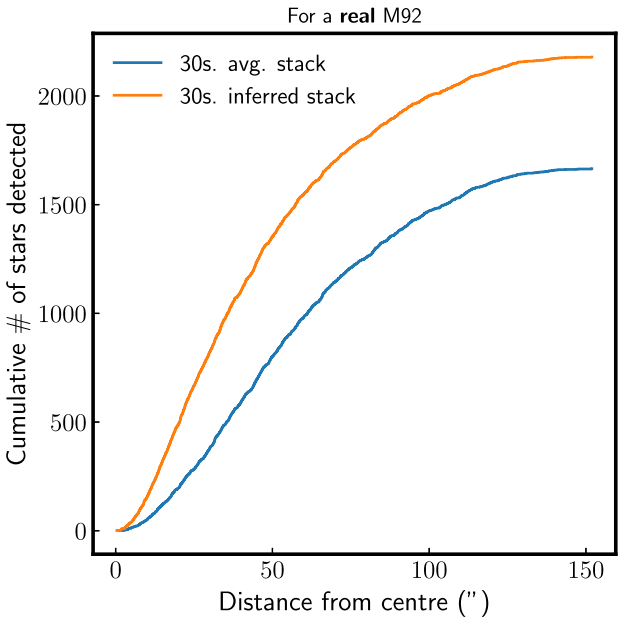

Figure 12. `SExtractor` was run on the inferred and simple averaged stack of the real M92 shown in Fig. 9. Shown here is the cumulative number of stars found as a function of distance from the centre of the cluster for each stack, highlighting the ability of the proposed method to work in the crowded central regions of a stellar cluster.

the evolving nature of atmospheric turbulence. To address this, a more dynamic model of phase screens could be developed, where new turbulence patterns are continuously generated rather than recycled. This would better mimic the natural, non-repetitive behaviour of atmospheric turbulence over extended periods. Implementing algorithms that can generate realistic, time-evolving turbulence patterns without significantly increasing computational load would be key.

In addition to these limitations, there are other effects not accounted for in the simulations, including 16-bit quantization, dark currents, quantum efficiency, malfunctioning pixels, and filter variations. These effects can all be added in future iterations of the method, either in simulations or by using a semisupervised approach to jointly train on real observations.

Two notable issues that have emerged are the 'fainter fatter' effect, where fainter stars are excessively smoothed, and the 'hallucination' effect, where the U-Net erroneously interprets noise as non-existent stars (as evidenced by imperfect precision of detected sources). These phenomena have multiple potential causes:

(i) Limited temporal context: Atmospheric turbulence disperses starlight, reducing the signal-to-noise ratio per pixel, sometimes burying the signal entirely. With limited information, the U-Net's tendency to misinterpret noise (either as part of a faint star's profile or as a new star) increases, especially for faint stars with inherently weaker signals.

(ii) U-Net's inherent design: The U-Net's output, when trained with a square-error cost, converges to the mean of the posterior distribution. Indeed, the output is a probability-weighted average of inferred images compatible with the data, where the solution's likelihood is much more 'peaky' for bright sources than for faint sources.

(iii) Difficulty tracking faint, isolated stars: The U-Net might struggle to consistently 'track' extremely faint, isolated stars. Ex-

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

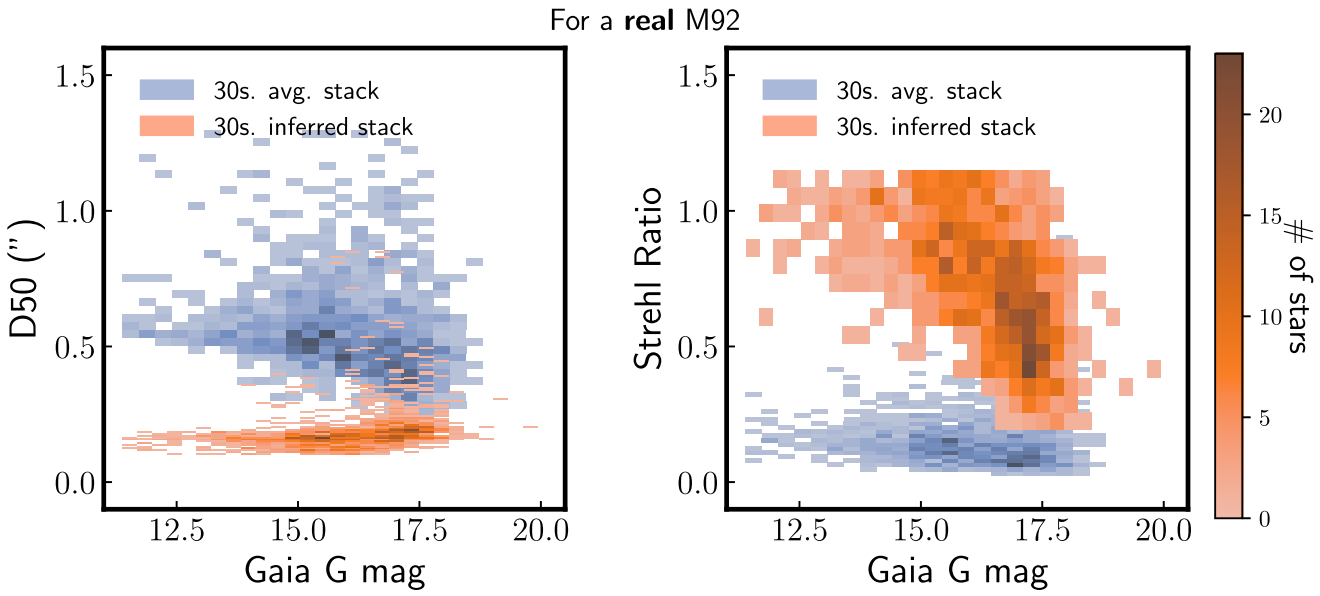

Figure 13. `SExtractor` was run on the inferred stack and simple averaged stack of M92 shown in Fig. 6. Shown here are the distributions of D50 (left plot) and Strehl ratio (right plot) estimates of the detected stars as a function of *Gaia* G magnitude.

Table 1. Comparison of the precision, percentage of recovered stars, average angular diameter ($\overline{\text{D50}}$), and average Strehl ratio ($\overline{\text{S}}$) for the various image processing methods used for simulated and real M92 observations. We note that the chosen detection threshold in `SExtractor` can affect the precision and recall substantially (see Fig. 11). In the case of real M92 observations the detection threshold for the averaged stack could be chosen such that the precision was fairly close to 100 per cent but with a very small total number of stars detected, whereas the inferred stacks had a definite maximum precision. For a more direct comparison, the thresholds for the averaged stack were chosen such that their resulting precision was approximately equal to the maximum precision for the inferred stack.

Method	Simulated M92 results				Real M92 results			
	Precision (per cent)	Recovered stars (per cent)	$\overline{\text{D50}}$ (arcmin)	$\overline{\text{S}}$	Precision (per cent)	Recovered stars (per cent)	$\overline{\text{D50}}$ (arcmin)	$\overline{\text{S}}$
30s. Avg. Stack	97.1	63.1	0.63	0.13	97.1	8.73	0.50	0.20
30s. U-Net Inferred Stack (6s. input context)	96.4	85.8	0.32	0.51	97.2	10.04	0.24	0.52
30s. U-Net Inferred Stack (12s. input context)	97.8	86.2	0.23	0.60	97.2	9.57	0.19	0.74

tending the temporal context may offer limited improvement for this specific issue.

These observed limitations highlight the complex interplay between atmospheric conditions, U-Net architecture, and the challenges inherent in detecting faint objects.

Addressing these limitations could involve expanding the temporal context of the observations, as shown in Table 1 and discussed in Section B, where it is confirmed that an increase in temporal context correlates with improvements in precision, Strehl ratio, and D50 measurements. This suggests that a broader temporal scope allows for a more accurate signal reconstruction, particularly for fainter stars.

Future enhancements to the proposed method could benefit from exploring diverse approaches to uncertainty quantification. For example, the method outlined by Angelopoulos & Bates (2021) offers a framework for creating statistically rigorous, distribution-free uncertainty sets for any pre-trained model's predictions, without relying on distributional or model assumptions. This technique ensures that the generated sets contain the ground truth with a specified probability. In contrast, probabilistic or generative models, such as Bayesian neural networks, generative adversarial networks (GANs), or diffusion-based models, provide an alternative means of assessing uncertainty. These approaches could generate probabilistic distributions for each pixel, offering a detailed view of the possible inferences and their associated uncertainties. These methods would allow for a more granular understanding of uncertainty, presenting the variability of reconstructions in a quantitatively rich manner, thus offering a distinct advantage in scenarios where understanding the distribution of inferences is crucial.

Another important limitation of our current method is its inability to significantly improve astrometric precision despite sharper images. This shortcoming likely stems from the residual jitter caused by atmospheric turbulence, as the U-Net, though capable of tracking the turbulent PSF across exposures, cannot determine the true position of stars. Averaging the centroids of these wandering PSFs over the 32 or 64 frames is not significantly more beneficial than averaging the light distribution, leaving a gap in astrometric accuracy. The addition of priors coming, for example, from the *Gaia* catalogue could help us to remove this limitation.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

While the method has its limitations, many of these challenges offer exciting avenues for future research and development. Addressing these constraints can evolve the method into a more robust and comprehensive solution for astronomical imaging in the presence of atmospheric turbulence. Indeed, we would like to emphasize that the work presented in this paper represents only the first iteration of the proposed method. There are numerous optimizations which can be implemented to enhance the performance by, e.g. incorporating more realism into the simulations, broadening the temporal and spatial contexts over which the U-Net makes its inferences, using shorter duration images to capture more information about the turbulence, and implementing more advanced machine learning methodologies and architectures.

6 CONCLUSIONS

In this study, we introduced a novel machine learning-based approach to counteract the challenges posed by atmospheric turbulence in astronomical imaging. By utilizing a U-Net architecture, we have demonstrated the potential to significantly enhance the sharpness of astronomical images. Our method, trained on simulated observations, is adept at inferring a turbulence- and noise-free image from a sequence of short-exposure observations of a stellar field, effectively associating speckles with their source star and disentangling light from proximate sources, while conserving flux.

Visually, the method showcased an enhancement in image clarity, especially under suboptimal seeing conditions. Quantitatively, our results have been compelling: when tested on the simulated M92 data set, the inferred frames exhibited an average reduction in D50 measurements by a factor of 3 for stars brighter than a *Gaia* G magnitude of 17.5, and an average 6× improvement in the Strehl ratio. This performance, however, tapered for fainter stars, indicating areas for further refinement. Furthermore, our quantitative analysis using `SExtractor` revealed that, when using the inferred frame, up to 36 per cent more stars were identified relative to the averaged frame, with a precision rate of ∼98 per cent. This is a testament to the model's ability to enhance image quality and resolution, even in densely populated stellar fields.

In real-world tests, using a 30 s video sequence of the globular cluster M92 as a case study, our method demonstrated its ability to de-blend crowded regions. Specifically, the inferred image recovered about 3× more stars within 25 arcsec from the centre of M92 compared to the temporally averaged image, with an average reduction in D50 measurements by a factor of 2.5 and an average 3.5× improvement in Strehl ratio. However, the performance metrics, particularly for fainter stars, indicated a more rapid decline in the real data scenario than in the simulated one. This highlights the challenges posed by real-world conditions, such as the intermittent turbulent ground layer observed in the M92 video stream and the non-Gaussian behaviour of pixel noise in sCMOS devices. The model's tendency to 'smooth' and/or possibly 'hallucinate' fainter stars is an expected limitation that warrants further investigation. The performance on real data, while promising, highlights areas of improvement, particularly in handling effects not currently accounted for in simulations. Indeed, given the suite of current limitations in the simulations mentioned in Section 5.1, it was not granted that meaningful inferences could be made on real data.

In conclusion, this first 'DanceCam' study presents a machine learning-based approach to addressing the challenge of atmospheric turbulence in astronomical imaging. The results obtained from both simulated and real data demonstrate the capabilities and potential of this method. However, it is important to acknowledge that further development and refinement are necessary for the DanceCam approach, particularly in enhancing the reconstruction of fainter stars and improving astrometric precision. The method, in its current form, shows promise for contributing to the field of astronomical research. Approaches like ours, while still in their nascent stages, could play a role in improving the clarity and accuracy of wide-field ground-based observations. Nevertheless, we recognize the need for cautious optimism regarding the method's scientific potential, and we encourage continued exploration and testing to fully ascertain its efficacy in diverse observational scenarios.

ACKNOWLEDGEMENTS

We acknowledge and respect the lək̓ʷəŋən peoples on whose traditional territory the University of Victoria stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 682903, P.I. H. Bouy), from the French State in the framework of the 'Investments for the future' Program, IdEx Bordeaux, reference ANR-10-IDEX-03-02, and from the French 'Projets nationaux' PNCG, PNHE and PNP.

We thank the National Sciences and Engineering Research Council of Canada for funding through the Discovery Grants and CREATE programs, and the Canada-France-Hawaii Telescope – where a large part of the work was done – for support both financially and from the excellent staff. We also thank Prof. Kim A. Venn for helpful comments and support on this project. This research was enabled in part by support provided by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

DATA AVAILABILITY

Videos of the simulations and the observations, together with additional information about the DanceCam project are available at https://dancecam.info. All other data underlying this article will be shared on reasonable request to the corresponding author. *Code availability:* A public repository of the code which produced the figures in this paper is available at https://github.com/DanceCam/dancelibpublic. The rest of the code will be shared on reasonable request to the corresponding author.

REFERENCES

Abdurrahman F. et al., 2018, AJ, 156, 100
Alarcon M. R., Licandro J., Serra-Ricart M., Joven E., Gaitan V., de Sousa R., 2023, PASP, 135, 055001
Angelopoulos A. N., Bates S., 2021, preprint (arXiv:2107.07511)
Aristidi E., Fantéï-Caujolle Y., Ziad A., Chabé J., Giordano C., Renaud C., Rafalimanana A., 2020, Adaptive Optics Systems VII. Vol. 11448, SPIE, p. 1488
Bates R. H. T., Cady F. M., 1980, Opt. Commun., 32, 365
Beckers J. M., 1993, Annu. Rev. Astron. Astrophys., 31, 13
Bendjoya P., Abe L., Rivet J., Suárez O., Vernet D., Mékarnia D., 2012, proc. of the SF2A-2012, p. 643
Bertin E., Arnouts S., 1996, Astron. Astrophys. Suppl. Ser., 117, 393
Beskin G. et al., 2023, MDPI, 10, 1352
Betoule M. et al., 2023, A&A, 670, A119
Borawar L., Kaur R., 2023, in Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022. Springer, p. 235
Brigham E. O., Morrow R., 1967, IEEE Spectrum, 4, 63

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

Cabrera-Vives G., Reyes I., Förster F., Estévez P. A., Maureira J.-C., 2017, ApJ, 836, 97
Chatterjee M. R., Mohamed F. H., 2014, Opt. Eng., 53, 126107
Çiçek Ö., Abdulkadir A., Lienkamp S. S., Brox T., Ronneberger O., 2016, in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19. Springer, p. 424
Davies R., Kasper M., 2012, ARA&A, 50, 305
Faedi F. et al., 2013, MNRAS, 433, 2097
Fried D. L., 1966, J. Opt. Soc. Am., 56, 1372
Frisch U., Kolmogorov A. N., 1995, Turbulence: the Legacy of AN Kolmogorov. Cambridge Univ. Press, Cambridge
Gieseke F. et al., 2017, MNRAS, 472, 3101
Guidash M., Ma J., Vogelsang T., Endsley J., 2016, Sensors, 16, 517
Hardie R. C., Power J. D., LeMaster D. A., Droege D. R., Gladysz S., Bose-Pillai S., 2017, Opt. Eng., 56, 071502
Hardy J. W., 1998, Adaptive Optics for Astronomical Telescopes, Vol. 16. Oxford Univ. Press, Oxford
Harris F., 1978, Proc. IEEE, 66, 51
He K., Zhang X., Ren S., Sun J., 2016, in Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. p. 770
Hoffmire M. A., Hardie R. C., Rucci M. A., Van Hook R., Karch B. K., 2021, Opt. Eng., 60, 033103
Isensee F., Jaeger P. F., Kohl S. A., Petersen J., Maier-Hein K. H., 2021, Nat. Methods, 18, 203
Johnston D. C., Welsh B. M., 1994, J. Opt. Soc. Am. A, 11, 394
Kaiser N., Tonry J. L., Luppino G. A., 2000, PASP, 112, 768
Kharchenko N., Piskunov A., Schilbach E., Röser S., Scholz R.-D., 2013, A&A, 558, A53
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kolarik M., Burget R., Riha K., 2020, in 2020 43rd international conference on telecommunications and signal processing (TSP). IEEE, p. 677
Kolmogorov A. N., 1941a, in Dokl. Akad. Nauk SSSR. p. 299
Kolmogorov A. N., 1941b, Cr Acad. Sci. URSS, 30, 301
Li N., Thapa S., Whyte C., Reed A. W., Jayasuriya S., Ye J., 2021, in Proceedings of the IEEE/CVF International Conference on Computer Vision. p. 2522
Li R. et al., 2022, preprint (arXiv:2207.02376)
Mackay C. et al., 2018, Publ. Astron. Soc. Aust., 35, e047
Makitalo M., Foi A., 2011, IEEE Trans. Image Process., 20, 2697
Massari D. et al., 2016, A&A, 595, L2
McClure R. D. et al., 1989, PASP, 101, 1156
Mizusawa S., Sei Y., Orihara R., Ohsuga A., 2021, Comput. Med. Imag. Graph., 90, 101920
Nieuwenhuizen R., Schutte K., 2019, in Dijk J., ed., Proc. SPIE Conf. Ser. Vol. 11870, Artificial Intelligence and Machine Learning in Defense Applications III. SPIE, Bellingham, p. 153
Paillassa M., Bertin E., Bouy H., 2020, A&A, 634, A48
Prusti T. et al., 2016, A&A, 595, A1
Rigaut F. J., Ellerbroek B. L., Flicker R., 2000, in Adaptive Optical Systems Technology. SPIE, p. 1022
Roberts Jr L. C. et al., 2004, in Advancements in Adaptive Optics. SPIE, p. 504
Roddier F., 1981, Progress in Optics, vol. 19. Elsevier, p. 281
Ronneberger O., Fischer P., Brox T., 2015, in International Conference on Medical image computing and computer-assisted intervention. Springer, p. 234
Schmidt J. D., 2010. SPIE.
Schödel R., Yelda S., Ghez A., Girard J. H., Labadie L., Rebolo R., Pérez-Garrido A., Morris M. R., 2013, MNRAS, 429, 1367
Tokovinin A., 2004, PASP, 116, 941
Tonry J., Burke B. E., Schechter P. L., 1997, PASP, 109, 1154
Tyson R. K., Frazier B. W., 2022, Principles of Adaptive Optics. CRC Press, Boca Raton
Vallenari A., Brown A., Prusti T., 2022, A&A, 674, A1
Vint D., Di Caterina G., Soraghan J., Lamb R., Humphreys D., 2020, in Dijk J., ed., Proc. SPIE Conf. Ser. Vol. 11543, Artificial Intelligence and Machine Learning in Defense Applications II. SPIE, Bellingham, p. 1154303
Wang L., Lee C.-Y., Tu Z., Lazebnik S., 2015, preprint (arXiv:1505.02496)
Yao W., Zeng Z., Lian C., Tang H., 2018, Neurocomputing, 312, 364
Young A. T., 1974, ApJ, 189, 587
Zhang Z., Liu Q., Wang Y., 2018, IEEE Geosci. Remote Sens. Lett., 15, 749
Zhang Y.-h. et al., 2020, in Holland A. D., Beletic J., eds, Proc. SPIE Conf. Ser. Vol. 11454, X-Ray, Optical, and Infrared Detectors for Astronomy IX. SPIE, Bellingham, p. 19
Zhang X., Mao Z., Chimitt N., Chan S. H., 2024, IEEE Transactions on Computational Imaging, IEEE
Zhu J. et al., 2022, in Holland A. D., Beletic J., eds, Proc. SPIE Conf. Ser. Vol. 12191, X-Ray, Optical, and Infrared Detectors for Astronomy X. SPIE, Bellingham, p. 743

APPENDIX A: COMPARISON TO INFERENCE ON TEMPORALLY AVERAGED FRAME

Our primary methodology for mitigating atmospheric turbulence in astronomical images utilizes short video sequences to train a Residual U-Net. An alternative approach involves training the U-Net with a single, long exposure image. This section reiterates the rationale behind our video-based method and the potential limitations associated with a long exposure input.

Atmospheric turbulence introduces dynamic, time-varying distortions in stellar light. A video stream, comprised of numerous short-exposure frames, effectively captures these temporal variations within the turbulence. By providing the U-Net with this sequence of frames, we enable it to analyse the patterns of turbulence and discriminate between photons belonging to different stars, even in cases where blurring leads to image overlap in single frames.

When instead presented with a long exposure image, the U-Net receives a composite view where the effects of turbulence are accumulated over time. This obscures the finer temporal patterns of turbulence, making it more difficult for the model to disentangle light from closely spaced astronomical sources.

For comparative testing, we simulate long exposure images by temporally averaging the frames within a video stream. This averaging process mimics the effect of a traditional long exposure observation. Fig. A1 shows a *qualitative* comparison of sources detected by `SExtractor` – in particularly crowded regions of randomly generated stellar fields – between output images inferred from either a video stream or a single temporally averaged frame. For a more *quantitative* analysis, we test on several hundred simulated stellar fields to collect statistics of the detected sources in crowded sub-regions. As seen in Fig. A2, inferring on video frames allows for recovery of $\sim 2\times$ more stars in the fainter end of the magnitude distribution ($m > 19$). These simple tests help us to confirm that the video streams enable the U-Net to better disentangle the light from neighbouring stars and consistently recover more sources in crowded regions.

While the long exposure input approach may provide practical benefits, such as a reduced computational cost and a simplified data acquisition process, our research demonstrates that the video-based method holds superior potential for reconstructing turbulence-free astronomical images. By exploiting the temporal dynamics of atmospheric distortions, our approach enables the Residual U-Net to more effectively reconstruct stellar images, ultimately enhancing the resolution of astronomical images.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

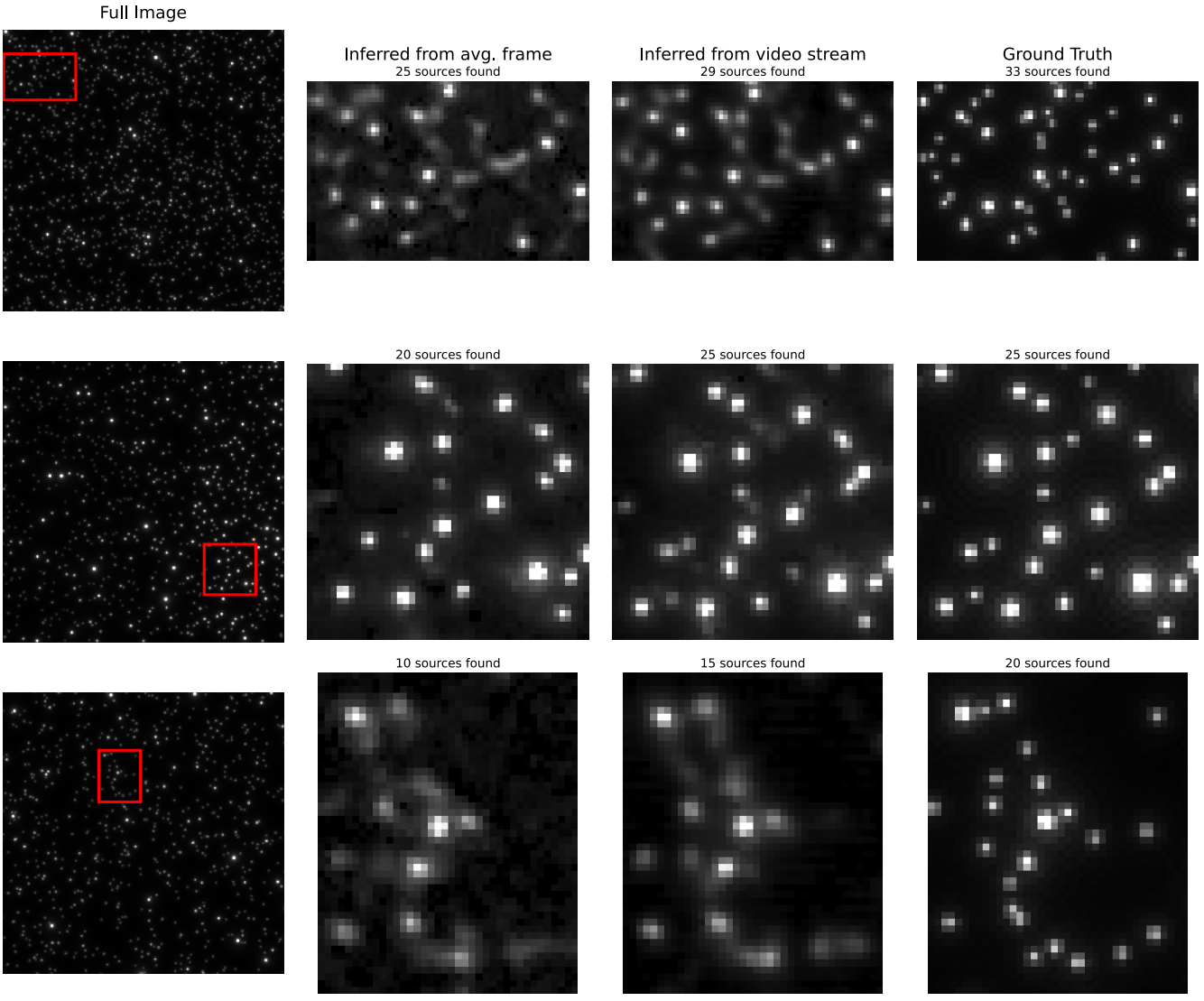

Figure A1. Qualitative performance comparison of source detection in crowded regions of simulated stellar fields. `SExtractor` is applied to the U-Net's inferred output when trained on either a video stream or its temporally averaged equivalent. The video-based approach demonstrates superior ability to discriminate light from nearby sources, as seen in both the source counts and the zoomed in densely-crowded sub-regions (indicated by rectangles).

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

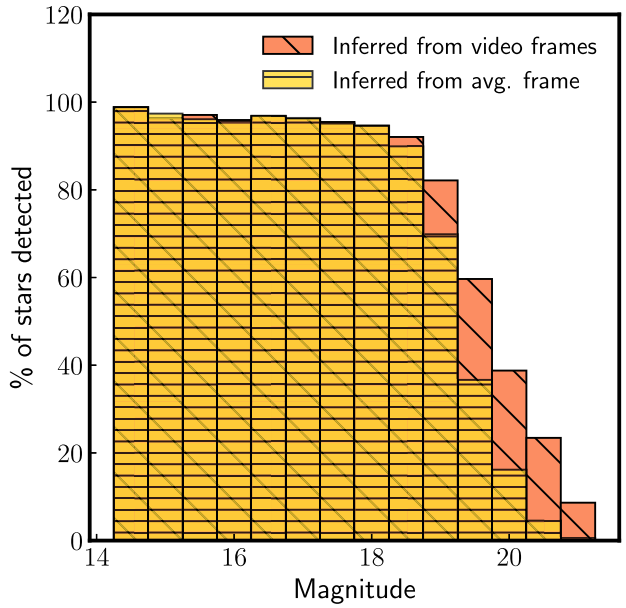

Figure A2. Quantitative performance comparison of source detection in crowded regions of simulated stellar fields. The same process as described in Fig. A1 was performed for several hundred simulated stellar fields and the number of detected stars compared to the ground truth was binned according to magnitude. The video-based approach again demonstrates a consistently better ability to discriminate light from neighbouring, especially faint sources.

APPENDIX B: RESULTS WITH A SMALLER TEMPORAL INPUT CONTEXT

To investigate the impact of the input temporal context on U-Net training and testing, we conducted experiments with a reduced context window of 6 s (32 frames @ 5.25 frames per second). This appendix presents the results of these experiments, offering a comparison to the performance achieved with a 12-s context. We repeated the M92 tests outlined in Section 4.2, with results presented below.

For simulated M92 observations, Fig. B1 shows the proportion of stars recovered across different magnitudes. While a 6-s context still achieves a respectable 85.8 per cent recovery rate and 96.4 per cent precision, the 12-s context (Fig. 7) demonstrates a noticeable advantage (86.2 per cent recovery rate and 97.8 per cent precision), especially for fainter stars.

Fig. B2 highlights performance differences for D50 and Strehl ratio measurements when compared to 8. The 12-s context yields superior results (0.23 arcsec D50, 0.60 average Strehl ratio) compared to the 6-s context (0.32 arcsec D50, 0.51 Strehl ratio). Again, these benefits are most pronounced for fainter sources.

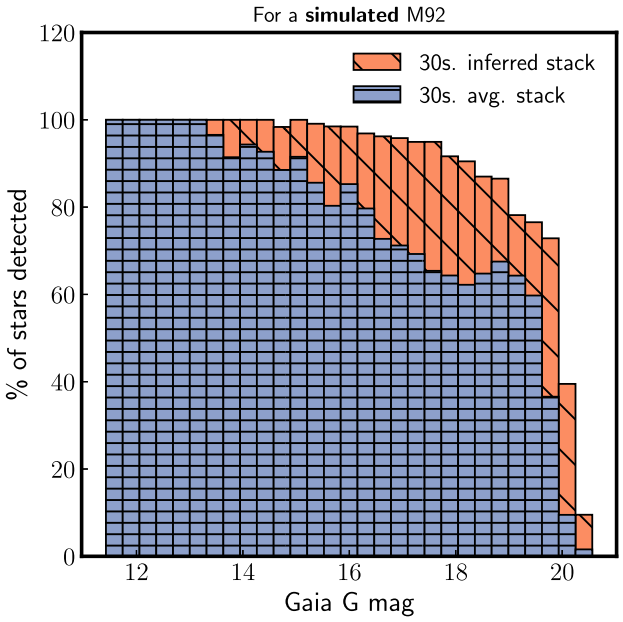

Figure B1. Percentage of stars recovered by magnitude range for simulated M92 observations given a 6-s temporal input to the U-Net (only relevant to the inferred stack). Compare to Fig. 7 for the 12-s temporal context.

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

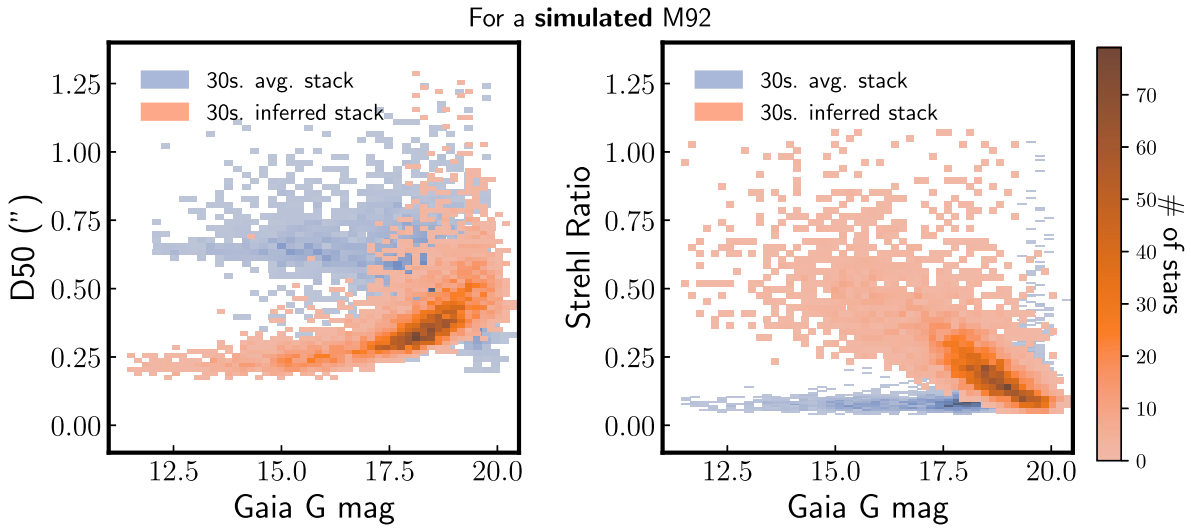

Figure B2. D50 and Strehl ratio performance versus magnitude for simulated M92 observations given a 6-s temporal input to the U-Net (only relevant to the inferred stack). Compare to Fig. 8 for the 12-s temporal context.

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

Downloaded from https://academic.oup.com/mnras/article/531/1/403/7654005 by guest on 12 September 2024

© 2024 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.