Automatic detection of impact craters on Al foils from the Stardust interstellar dust collector using convolutional neural networks

Logan JAEGER1, Anna L. BUTTERWORTH1, Zack GAINSFORTH1, Robert LETTIERI1, Dan ZEVIN1, Augusto ARDIZZONE2, Michael CAPRARO3, Mark BURCHELL3, Penny WOZNIAKIEWICZ3, Ryan C. OGLIORE3, Bradley T. DE GREGORIO5, Rhonda M. STROUD5, and Andrew J. WESTPHAL1

1Space Sciences Laboratory, University of California at Berkeley, Berkeley, California 94720, USA
2Stardust Citizen Science Project
3Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK
4Washington University, St. Louis, Missouri 63130, USA
5Materials Science and Technology Division, Code 6366, US Naval Research Laboratory, Washington, District of Columbia 20375-5320, USA

*Corresponding author. E-mail: loganjaeger@berkeley.edu

(Received 17 March 2021; revision accepted 26 August 2021)

Abstract—NASA’s Stardust mission utilized a sample collector composed of aerogel and aluminum foil to return cometary and interstellar particles to Earth. Analysis of the aluminum foil begins with locating craters produced by hypervelocity impacts of cometary and interstellar dust. Interstellar dust craters are typically less than one micrometer in size and are sparsely distributed, making them difficult to find. In this paper, we describe a convolutional neural network based on the VGG16 architecture that achieves high specificity and sensitivity in locating impact craters in the Stardust interstellar collector foils. We evaluate its implications for current and future analyses of Stardust samples.

INTRODUCTION

NASA’s Stardust mission successfully returned cometary and interstellar dust collectors to Earth in 2006, after an encounter with the comet 81P/Wild 2 in 2004, and exposure to the interstellar dust stream in 2000 and 2002. The cometary and interstellar dust particles it collected have been extensively studied, leading to scientific insights on the origins of the solar system (Brownlee, 2014; Brownlee et al., 2006), the nature of comets (Brownlee et al., 2004; Flynn et al., 2006; Gainsforth et al., 2019; Joswiak et al., 2012; McKeegan et al., 2006; OglioRE, Huss, et al., 2012; Simon et al., 2008; Zolensky et al., 2006), the presence of cometary amino acids (Elsila et al., 2009; Sandford et al., 2006), and the properties of interstellar dust (Bechtel et al., 2014; Brenker et al., 2014; Butterworth et al., 2014; Flynn et al., 2014; Gainsforth et al., 2014; Stroud et al., 2014; Westphal, Bechtel, et al., 2014; Westphal, Stroud, et al., 2014). Stardust carried one collector to capture cometary particles from 81P/Wild 2 and another to capture interstellar dust particles. Each collector consisted of aerogel tiles and aluminum foils. The foils were intended to facilitate aerogel removal, but also provided an additional capture medium (Tsou et al., 2003). When particles hit the aluminum foils, they created small craters (Hörz et al., 2006; Kearsley, Borg, et al., 2008; Price et al., 2010; Wozniakiewicz et al., 2012). These craters are sparsely distributed on the foils and the majority are only detectable by scanning electron microscope (SEM) imaging. Even then they can easily be confused with imperfections in the foil (see Figs. 1 and 2), making the task of locating them nontrivial.

A large citizen science project, Stardust@home (hereafter SAH), engaged human volunteers to locate craters on the Stardust foils (Westphal et al., 2005). This project has attracted more than 30,000 volunteers and has been particularly successful with respect to the identification of particles in aerogel (Westphal, Anderson, et al., 2014). However, even with this massive participation, human views of the foils are still sparse. Nearly 20% of the available SAH foil images have not
been viewed more than once. Furthermore, the SAH site only contains images from 4 of the roughly 200 individual foils present in the Stardust collector. More importantly, the majority of the foils in the Stardust collector have never been scanned by an SEM. Stroud et al. (2014) report that approximately 4.8% of the foils were examined by an SEM during the preliminary examination. Since then, just over half of the foils have been removed from the collector. An efficient and effective automated procedure could provide an impetus to scan the rest of these foils, as the examination of images would be much easier.

Previous successful attempts at automatic crater identification include a normalized cross-correlation and template matching algorithm by Ogliore, Floss, et al. (2012) and an algorithm based on a circular Hough transform and Canny edge detection developed by De Gregorio et al. (2018). Both of these algorithms have revealed multiple new craters. Convolutional neural networks (CNNs), which have been established as effective tools in image classification (He et al., 2015; Krizhevsky et al., 2012), have the potential to complement or outperform the previously developed machine learning techniques.

By way of comparison, the Hough transform algorithm detects circular craters, and template matching detects craters matching a set of known craters. A CNN can be trained on craters in combination with other features, such as scratches and image exposure, and can integrate those observations into the final metric. Thus, while both the Hough transform algorithm and the cross-correlation template matching algorithm only seek to find crater-like features, a CNN can both identify crater-like features and address the environment of a given image (see Fig. 3). Because the fraction of SEM images that contain craters is very small, Bayes’s theorem implies that the majority of images classified as containing craters by our algorithm will actually be false positives, even with a generous estimation of the neural net accuracy, for example, if we suppose that our CNN achieves a value of 0.999 for both sensitivity and specificity, and we take the estimate that there are 100 total craters in the sample of approximately $10^7$ images.

![Fig. 1. Some examples of images that contain real craters from the Stardust interstellar collector. Notice the differences in size, shape, and image exposure. (a) and (c) come from foil I1020W, while (b) comes from foil I1126N. (Color figure can be viewed at wileyonlinelibrary.com.)(a) and (c) come from foil I1020W, while (b) comes from foil I1126N. (Color figure can be viewed at wileyonlinelibrary.com.)](image1)

![Fig. 2. Examples of images that do not contain craters, but do contain some misleading imperfections in the foil. These features are either impurities introduced during the manufacturing process or are impacts produced by secondary ejecta from the Stardust spacecraft (Burchell et al., 2012). c) Contains an iron oxide “volcano,” which occurs when there is a iron impurity near the surface of the foil that “erupts” out. a and b) From foil I1020W. c) From foil I1126N. (Color figure can be viewed at wileyonlinelibrary.com.)](image2)
required to search the entire collector (Landgraf et al., 1999). Even so, Bayes’s theorem says there is still less than a 1% chance that an image classified to contain a crater actually contains a crater. Where $P(c)$ represents the probability that a crater is actually present in an image, $P(\neg c)$ represents the probability that a crater is not present in an image, and $P(\pm)$ represents the probability that our CNN predicts a crater is present in an image:

$$P(c|\pm) = \frac{P(\pm|c)P(c)}{P(\pm|c)P(c) + P(\pm|\neg c)P(\neg c)} = \frac{0.999 \times (100/10^7)}{0.001 \times (1 - (100/10^7)) + 0.999 \times (100/10^7)} \approx 1\%$$

So a significant amount of human interaction will still be necessary to ultimately find and identify craters, even if we reach the bounds of practical CNN performance. The focus of the project, then, is not to make a perfect crater identification algorithm that eliminates the need for human interaction with these images. Instead, we intend to make an algorithm that is good enough to correctly identify the large majority of images that do not contain craters, thereby reducing the amount of time and effort humans must expend looking through images. That is, we are biasing our CNN toward being highly specific, rather than being highly sensitive.

Although all of the data used in this paper come from the interstellar dust collector, the results could be applied to the cometary collector. Interstellar craters are typically $<1$ μm in diameter. The cometary foils contain many impact craters that are much larger, from 5 to 200 μm in diameter (Kearsley, Borg, et al., 2008). However, sub-μm cometary craters outnumber larger craters (Price et al., 2010), and there is interest in examining the size, shape, and mineral composition of these smaller cometary particles as well (Leitner et al., 2008; Leroux et al., 2008; Stroud et al., 2010).

Fig. 3. Heat maps created by overlaying a scaled up activation map from one of the filters on the final convolutional layer over the input image. These reveal an advantage of a CNN over previous attempts at crater identification. Earlier methods only “trigger” on features that look like craters without accounting for the surrounding conditions. A CNN contains many filters in each convolutional layer, with each filter having a different application. Some filters are focused on identifying craters, like that in (b), while others can see features like the scratch identified by the filter in (d). In this way, a CNN can account for the relationship between a crater-like object and its environment to better determine whether or not it truly is a crater. Both images are from the foil I1020W. (Color figure can be viewed at wileyonlinelibrary.com.)
EXPERIMENTAL

Sample Preparation

The top foil surface was separated from the collector apparatus using a twin rotary cutter in the Stardust Lab at Johnson Space Center. This resulted in thin Al strips that were stored and shipped in a custom foil mount (Kearsley, Borg, et al., 2008). In preparation for SEM imaging, we clamped the foils mechanically at the ends and stretched them to ensure flatness. This method of foil mounting was also described in Stroud et al. (2014) and has been shown to ensure stability and meet the standards necessary for automatic image acquisition. We imaged the foils in SEMs using voltage between 5 and 15 keV, current between 0.2 and 5 nA, and image resolution between 40 and 60 nm px$^{-1}$ (Stroud et al., 2014). Care was taken to ensure that imaging times were sufficiently short, and that the vacuum chambers within the SEMs used were sufficiently free of carbon, in order to minimize deposition that could compromise later investigation. For example, Auger spectroscopy can be used for elemental mapping of craters in the Stardust foils but is extremely sensitive to carbon deposition and can be obscured even by a 4 nm layer of carbon (Stadermann et al., 2009).

Due to the small number of actual interstellar impact craters that have thus far been observed in the SEM images of the Stardust collector, this project would be impossible without analog craters produced by alternative means, namely light gas gun (LGG) shots of aluminum foils (Burchell et al., 1999). Forty-four such craters had originally been imaged for use by Stadermann et al. (2008), and had previously been used by Ogliore, Floss, et al. (2012). We produced a further 53 with the two-stage LGG in the Centre for Astrophysics and Planetary Sciences at the University of Kent (see Burchell et al., 1999), using the procedure described in Wozniakiewicz et al. (2012). The method involved firing projectiles at an analog Stardust aluminum foil using a two-stage LGG. The projectiles were a mixed sample, comprising 0.5 and 1.0 μm diameter SiO$_2$ spheres, mixed with larger grains of hand-ground pyroxene, olivine, and diopside. We measured the impact speed to be 6.02 ± 0.05 km s$^{-1}$, comparable to the Stardust cometary dust impact speed. It was reported by Westphal, Stroud, et al. (2014) that the impact speed of interstellar particles is also <10 km s$^{-1}$. The craters from the smaller SiO$_2$ grains were identifiable by their regular circular shape and size. Based on Price et al. (2010), it is predicted that, for silica beads similar to those here impacting Al foil at 6 km s$^{-1}$, crater size = 1.54 times projectile diameter. Hence, crater sizes of 0.77 and 1.54 μm are expected from the SiO$_2$ spheres, as well as larger craters from the mineral impacts. The larger craters from the various minerals would have distinct differences in residue composition inside the craters. We imaged the foil using a Tescan Vega SEM at the Space Sciences Laboratory in Berkeley.

Data Preprocessing

Adequate CNN training requires a set of thousands of images that contain craters and thousands that do not. While we have many images without craters, images with craters in the interstellar collector are rare. We resolved this issue by augmenting a set of blank images with analog craters that sample the range of crater morphologies, as described in the section above. The process is shown in Fig. 4. A crater is first imaged (Fig. 4a), isolated from the surrounding foil (Fig. 4b), and then combined with a blank image (Figs. 4c and 4d). In order to isolate the crater from the surrounding foil, we hand-edited each image to delete the surrounding foil using a transparency channel. We did this using the GIMP digital image processing package (The GIMP Development Team, 2019, https://www.gimp.org) with brushes of varying sizes and hardnesses between 25 and 100 pixels in order to reduce the effect of a hard boundary at the edge of the crater or sensitivity to Fourier components harmonic with the brush size. We then performed a process of augmentation of these craters through rotating, rescaling, and altering the aspect ratio to produce about 30,000 artificial crater images (van der Walt et al., 2014). The aspect ratio is sampled from a normal distribution centered at 1 with a standard deviation of 0.1, and the rescaling is such that a crater is no smaller than five pixels in diameter and no larger than about 25, matching the observed scale. Such data augmentation techniques, known collectively as data warping, are well established in the field of deep learning and computer vision (Shorten & Khoshgoftaar, 2019; Wong et al., 2016; Yu et al., 2017). We then pasted these augmented craters into foil images that do not contain craters. We visually checked that these synthetic images looked reasonable (Fig. 4d). Before training, we normalized the images such that their pixel intensities were between 0 and 1. Finally, we split the synthetic images into training, validation, and test sets of 10,000 each, with a similar set of blank images for a total database of 60,000 images.

Network Overview

Our network structure is similar to the VGG16 network (Simonyan & Zisserman, 2014) frequently used
in the machine vision community, but is tailored to our specific application. The most significant difference is that our network is substantially smaller than VGG16, having $<250,000$ free parameters, while VGG16 has $>10^9$ free parameters. Figure 5 shows an overview of our network structure. We used a set of three coarse feature finding units, each with two convolutional layers and a max pooling layer (Giusti et al., 2013). These are followed by two more convolutional layers each with a max pooling layer, before feeding into a fully connected head. The coarse feature finding units identify features up to about 24 pixels wide, which is the size of a typical crater in our images. The additional two convolutional units provide context up to 48 pixels across. During hyperparameter optimization, we found that features larger than $\sim50$ pixels across contributed little to determining the presence of a feature that looks like a crater. Thus, we do not include convolutional layers for larger feature sizes. The head combines the features found by the convolutional filters to determine whether a crater is present or not.

**Detailed Architecture**

We programmed the CNN in Keras (Chollet et al., 2015) with a Tensorflow backend (Abadi et al., 2015) using Python. We utilized a number of other packages in image preprocessing and evaluation, including matplotlib (Hunter, 2007), scikit-learn (Pedregosa et al., 2011), scikit-image (Van der Walt et al., 2014), hdf5...

---

**Fig. 4.** The process of augmenting images. A crater is imaged by hand with an SEM (a), then isolated from the background using a soft brush (b). An SEM image of a Stardust foil that does not contain a crater (c) is the background into which the isolated crater is pasted (d), after having been augmented. Here, the SEM image of the stardust foil is 150 × 150 pixels, of the type used in the first step of CNN training. The stigmation in the crater image is slightly different than the stigmation in the background image, which will create a slight bias when we train our network, as a true image that contains a crater would have the exact same stigmation for both the crater and the background. In future versions of the network, we plan to adjust the stigmation of the crater to match that of the background.
A tabular representation of the network architecture can be seen in Table 1. For brevity, we have only included the main layers in Table 1. The CNN had a total of 216,465 trainable parameters, 8 two-dimensional convolutional layers, and 7 densely connected layers. All hidden layers are immediately followed by the ReLU activation function (Krizhevsky et al., 2012), while the output layer is equipped with a sigmoid activation function. Every convolutional layer uses a $3 \times 3$ kernel size, with a stride of 1 (stride is the number of pixels between the centers of adjacent kernel operations), and does not use padding. All max pooling layers use a $2 \times 2$ pool size and a stride of two pixels. We used several methods to prevent overfitting: Following every convolutional layer, there is a two-dimensional spatial dropout layer (Tompson et al., 2015) with a dropout rate of 0.25. Following every densely connected layer, there is a simple dropout layer (Srivastava et al., 2014) with a dropout rate of 0.3. Every convolutional layer is equipped with a ridge regression L2-loss kernel regularizer with regularization parameter $\lambda = 0.0001$ (Krogh & Hertz, 1992). As a bridge between the feature-finding and densely connected decision-making sections of the network, we used a global max pooling layer, as opposed to a flatten layer. A global max pooling layer not only allows a variable-size input but also helps prevent overfitting by minimizing the amount of noise passed to the decision-making section (Lin et al., 2013).

Network Training

The raw data consist of $384 \times 512$ pixel grayscale JPG images. It is computationally costly to train on images of that size. To alleviate this cost, we allowed our network to accept a variable-size grayscale input, and trained in multiple steps. In the first step, we trained on up to 20,000 synthetic $150 \times 150$ pixel grayscale images (10,000 each for positive and negative images). In the second step, we retrained the network, starting from the previously obtained weights, on up to 20,000 grayscale images sized $384 \times 512$ pixels. In both steps, we used the testing and validation sets of up to 20,000 images each.

For our optimizer, we used Adam with Nesterov momentum (Dozat, 2016), with a constant learning rate of 0.002. We used binary cross-entropy for our loss function. In order to bias the network against false positives, we adjusted the class weights such that a false positive was penalized 10 times more than a false negative.

Biasing the Training Sets

Performance can often be increased by ensemble learning (e.g., Caruana et al., 2004; Dietterich, 2000; Huang et al., 2017). Our current prediction algorithm is in fact an ensemble of three networks each trained with slightly different training data. They mostly identify the same images. However, some edge cases (ambiguous features that may be craters) are found by only one or two of the networks in the ensemble, and therefore, we capture more interesting craters by using the union of the predictions.

We produced different versions by varying the training sets as shown in Table 2. There are two main differences that separate the versions: the number of images used in training and the content of training images. The impact of using more images is clear:
Larger training sets resulted in more accurate networks after training. However, while V3 and V4 are nominally trained using the same number of images, V4 was trained using only synthetic images produced with our analog craters using the process described above, whereas V2 and V3 were trained with images from the SAH site that had been biased by human volunteer responses.

The negative images for the V2/V3 training sets were chosen from images that had been consistently marked negatively by the SAH users (fewer than 33% of the views had indicated a crater). Positive images were chosen from SAH “calibration” images, which are used to measure volunteer performance and had been previously manufactured at the launch of the SAH site in a similar manner described by Westphal, Anderson, et al. (2014), that had been consistently marked positively by SAH users (greater than 50% had indicated a crater). Note that all of these calibration images contain synthetic craters, and the volunteer bias was used so that we could be sure that the craters were clear and realistic. This approach was used to ensure that our network could mirror the activity of SAH volunteers, who have been established as largely reliable in the aggregate (Westphal, Anderson, et al., 2014).

Table 1. Simplified tabular representation of the network. See the Network Architecture section for a more detailed description. The “units” column here represents the final dimension of the output space of the layer. That is, it is equal to the number of filters in a convolutional layer, and the number of output nodes in a densely connected layer.

| Layer         | Units |
|---------------|-------|
| Conv2D        | 64    |
| Conv2D        | 64    |
| MaxPool       | –     |
| Conv2D        | 64    |
| Conv2D        | 64    |
| MaxPool       | –     |
| Conv2D        | 64    |
| Conv2D        | 64    |
| MaxPool       | –     |
| Conv2D        | 32    |
| MaxPool       | –     |
| Conv2D        | 32    |
| GlobalMaxPool | –     |
| Dense         | 32    |
| Dense         | 32    |
| Dense         | 16    |
| Dense         | 16    |
| Dense         | 16    |
| Dense         | 16    |
| Dense         | 1     |

Table 2. Comparison of different versions of our network. Specificity and synthetic crater sensitivity were determined from a testing set of 500 images that reflect the way in which the network was trained, as explained in the Network Training section, calculated with a threshold of 0.5.

|                   | V2    | V3    | V4    |
|-------------------|-------|-------|-------|
| Number of 150 × 150 training images | 6000  | 20,000| 20,000|
| Number of full-size training images | 2000  | 20,000| 20,000|
| Small FOV epochs  | 557   | 115   | 115   |
| Large FOV epochs  | 162   | 120   | 118   |
| Specificity       | 98.8% | 99.2% | 99.8% |
| Sensitivity (synthetic craters) | 96.6% | 98%   | 99.8% |
| Sensitivity (true craters) | 6/27  | 6/27  | 18/27 |

These “calibration” craters were manufactured using only a few analog craters that were all largely circular, and the use of more analog craters with a larger variety of structures helped V4 achieve higher sensitivity and specificity than V2 or V3.

RESULTS

An overview of characteristics and performance of each CNN version is shown in Table 2. The specificity was measured on a set of 500 images that do not contain craters, chosen randomly from the SAH database. Two different measures of sensitivity are presented so as to evaluate both the CNN and our method of synthetic crater generation. The sensitivity with synthetic craters was measured on a set of 500 images that had synthetic craters, reflective of the way in which the CNNs were trained. That is, the data set used for V2 and V3 contained the SAH calibration images, and the data set used for V4 contained synthetic craters created in the manner described in the Data Preprocessing section.

When an image from the SAH database is examined by our team and is determined to be a crater candidate, that corresponding section of the foil is rescanned to obtain a higher resolution image. Not all of these crater candidates that are rescanned are determined to be of interstellar origin (some are crater-looking foil defects, others are craters created by spacecraft ejecta). Our CNN should nevertheless identify these features as craters because, without a higher resolution image or elemental analysis, distinguishing them from interstellar craters is not possible. The sensitivity with true craters is given as the proportion of 27 images verified by any and all means over the years including higher resolution SEM imaging, and focused ion beam transmission electron...
microscopy. Each was verified by further study to be a real impact crater. It is apparent that the V4 version of the network, which incorporates the most advanced image augmentation techniques, performed the best at finding features that were later verified to be valid craters.

The normalized cross-correlation algorithm of Ogliore, Floss, et al. (2012; https://presolar.physics.wustl.edu/Laboratory_for_Space_Sciences/software/cratersource.zip) identified 8 of 25 true craters, indicating that its sensitivity is higher than V3 but lower than V4 of the network. However, the specificity is significantly lower, with one to three false positives for most images. Each image took a comparatively long time to search: about 4 minutes per processor core. Clearly, the network provides a significant improvement in sensitivity, specificity, and speed over a conventional normalized cross-correlation approach.

The output of the CNN is not binary, but is a continuous variable, known as the “prediction,” which is related to the probability that a crater is present. In Fig. 6, we plotted the sensitivity and specificity of each version of the network as a function of the minimum CNN prediction. Because much of the change in specificity and sensitivity occurs near a prediction of zero, we define a new parameter $\xi$ as

$$\xi = \log_{10}(p + 10^{-6})$$

where $p$ is the prediction of a given version of the network. A small constant, $10^{-6}$, has been added to the prediction inside the logarithm so that a prediction of zero does not diverge. This new parameter allows us to “zoom in” on prediction values near zero. The sensitivity in Fig. 6 is calculated on 129 images that have been selected as especially likely to contain craters by two of us, AA and MC. Notice that the specificity is high for any threshold greater than zero, and climbs relatively slowly after that, while the sensitivity continues to drop steadily. The equations used to calculate the sensitivity and specificity are shown below, where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively.

sensitivity = $\frac{TP}{TP + FN}$  
specificity = $\frac{TN}{TN + FP}$

Because SAH volunteers have been established as largely reliable, we can use their responses as a proxy measurement for the success of the network. Figure 7 shows that there is a clear correlation between SAH volunteer responses and the predictions of the network for all three versions, with V4 being the most successful. Only about 15,000 (~4%) of the 424,000 images in our data set meet a threshold of a prediction 0.01 from any of V2, V3, or V4. Thus, the network accomplishes our goal of eliminating the vast majority of images. (See Fig. 8 for detail on the distribution of predictions for all three networks.)

**Craters Found by Our Ensemble**

Approximately 40 promising crater candidates that had previously gone unnoticed were identified by our CNNs. Four of these can be seen in Fig. 9. It will take further examination in order to conclusively determine whether or not these are created by interstellar dust impacts or simply imperfections in the foil.

Our CNN took about an hour running on an Nvidia GeForce RTX 2060 GPU to generate predictions for the entire data set. This speed is important because it contrasts with the pace of SAH. Over the years 2015–2018, the average number of images viewed by volunteers per day was approximately 1200. With 424,000 images in the data set, it took nearly a year ($\frac{424,000}{1200}$ ≈ 353 days) for the SAH volunteers to work through that many images.

Some of the images that contain crater candidates have only been seen once or twice, like Fig. 9b, while others have been seen many times and have failed to earn a positive score. For example, Fig. 9c had been disapproved all four times it had been viewed. We would likely never be alerted to images like these, which are prone to being overlooked by volunteers, without an automated approach.

**DISCUSSION**

**A Replacement or a Complement to Stardust@Home?**

Deep learning is successful at classification tasks, but has inherent limitations. Whereas a human mind can leverage common sense and relatively few examples to perform well on simple tasks, neural networks lack general knowledge and real understanding, and require a large number of examples to reach proficiency with the same tasks.

Taking the statistical and practical limitations into account, we opted for a hybrid deployment strategy. After running the network on all of the images in our data set, we eliminated all images that the network negatively classified from the SAH website, and left in the ones with predictions greater than 0.01 so that volunteers would have to sort through significantly fewer images without craters. As we improve the quality of our network, it is possible to gradually reduce our...
reliance on SAH volunteers, but some form of human engagement will likely always be necessary.

**Future Improvement**

As is evident from Table 2, there is a rift between our network’s true sensitivity and its synthetic sensitivity, even with the augmented data set used to train V4, which points to an important shortcoming in our algorithm. The fact that the synthetic sensitivity is so high means that our network is training effectively: It gets very good at identifying the craters that we feed it. However, that success does not appear to transfer smoothly to the images of true craters in our data set. There are three main possibilities.

The first is that our augmentation techniques are inadequate, and we are changing the crater analogs to such an extent that they no longer resemble true craters. This is unlikely because we are using established data warping techniques (Shorten & Khoshgoftaar, 2019) with reasonable augmentation parameters. Furthermore, we can visually confirm that the augmented craters still resemble genuine craters (Fig. 4d).

A second explanation is that the morphologies of the analog craters do not adequately replicate the true crater morphologies. The shape of the impact craters depends on the density and internal structure of the dust grains (Kearsley, Graham, et al., 2008), and interstellar dust is composed of porous, composite particles (Mathis & Whiffen, 1989). In the future, we could use training sets based on analog craters from other compound impactors, for example, those used in Kearsley, Graham, et al. (2008).

The third is that the number of analog craters we use to construct our training set is too small. This is more likely. Data augmentation, though effective,
cannot replace the simple act of collecting more raw data (Lee et al., 2015; Wong et al., 2016). The only way to reduce the interdependence of training images created through data augmentation is to collect more raw data. Therefore, we could create more analog craters with the LGG, or re-examine the LGG foils which have already been imaged for new morphologies. In this way, we can increase the number of analog craters in our data set by at least a factor of two. The second way is to let the network find actual Stardust craters. Even though the true sensitivity is not perfect, it is still good enough to provide us with new images of true craters and, just as importantly, false positives. As we gain access to these new data, we can use this feedback to improve the performance of the network: The network finds some images that it thinks are craters and these images are evaluated both by experts and by volunteers on the SAH website. These images are then incorporated correspondingly into an improved training set to retrain the network, and this process is iterated with the new version of the CNN.

Data warping–based augmentation is an established technique for improving neural network performance when limited data are available, but there are other approaches to data generation that may be worth exploring. An entirely different way to improve our training data in the absence of new raw images is to deploy generative adversarial networks (GANs), which have seen success in recent years as an approach to data augmentation (Frid-Adar et al., 2018; Shorten & Khoshgoftaar, 2019). Perez and Wang (2017) reports GAN-based augmentation alone improves CNN performance about as much as traditional augmentation techniques, but recommends a combination of augmentation techniques that would likely outperform any single technique.

**Outlook for Further Efforts on Stardust Samples**

Aerogel composes 85% of Stardust’s collector area (Tsou et al., 2003) and preserves particles in a way that aluminum foil does not. Thus, it would be useful to have an effective computer vision model for identifying particles in the aerogel section of the collector. Here, we outline the challenges associated with this application.

The most obvious difference is the dimensionality of the data: When a particle embeds itself in aerogel, it creates a track that is imaged in three dimensions. On the Stardust@home website, aerogel tracks are viewed as “movies,” in which users view 30–40 frames in succession, each successive frame with its focus slightly lower than the last. Many tracks cannot be identified by a single frame. Much of the literature on volumetric computer vision comes from classification of volumetric

![Fig. 7. The average \( \xi \) score versus crater identification fraction. The crater identification fraction for a given image is the number of times an SAH volunteer has indicated a crater is likely present in the image, divided by the total number of times that the image has been seen by SAH volunteers. (Color figure can be viewed at wileyonlinelibrary.com.)](image)
medical images, like magnetic resonance imaging (MRI) and computed tomography (CT) scans. A number of different approaches to classifying this type of data have been attempted, many of which attempt to adapt 2-D networks to this problem: Prasoon et al. (2013) used three separate 2-D CNNs on three orthogonal planes of the volumetric data; Roth et al. (2014) adapted the approach used by Prasoon et al. (2013) and mapped the three orthogonal planes to the RGB channels of a 2-D CNN; Grewal et al. (2018) ran a 2-D CNN on successive slices in a single plane, but added a long short-term memory layer to their network, so previous slices could influence the predictions of later slices. Of course, others have attempted to apply 3-D CNNs to this type of classification problem, with early attempts by Dou et al. (2016) and Payan and Montana (2015), and a more recent attempt by Ker et al. (2019) that demonstrated higher performance than previous attempts at reasonable computational cost. Although MRI and CT scans produce data of a different form than optical images of aerogel, these results provide optimism that a CNN might be successfully applied to the aerogel section of Stardust’s collector.

**Application to Other Missions**

Our success with computer vision on the Stardust samples lends credence to the hope that similar CNN methods could be successfully applied to other sample return missions, especially those that collect particulate rather than bulk samples.

The Tanpopo mission, managed by the Japanese Aerospace Exploration Agency (JAXA) and designed to investigate the interplanetary migration of microbes...
Yamagishi et al., 2009), used a sample collection method very similar to the one used by the Stardust mission: aerogel held in place with aluminum foil (Tabata et al., 2014). Although it appears little analysis has been done on samples collected by the aluminum foil part of the Tanpopo collector, our research shows that the barrier to beginning this analysis is low, as applying a CNN similar to ours on SEM images of the Tanpopo foils would likely expose many craters, which could then be examined. Furthermore, our research lends optimism to an accelerated search for particles trapped in Tanpopo aerogel, as it does for those trapped in Stardust aerogel.

Stardust’s sister mission, Genesis, included collectors that are similar in many respects to Stardust’s foils, namely, flat surfaces with rare features of interest and many “contamination” features. The main goal of Genesis was to gather data on solar wind (Lo et al., 1998), but there has been at least some interest in the analysis of interplanetary dust particles coincidentally caught by the Genesis collectors (Graham et al., 2004). Also, NASA’s Long Duration Exposure Facility (LDEF) performed an experiment to measure the flux of small interplanetary dust particles (Singer et al., 1985). A re-examination of these surfaces with our CNN could provide an updated calculation.

CONCLUSION

This article presented the implementation of a CNN to the task of finding craters within SEM images of the aluminum foil portion of the NASA Stardust mission collector. The CNN was effective and demonstrated immediate utility. We were able to identify crater candidates in the Stardust at Home (SAH) database that had not been noticed before, and we were able to effectively eliminate 96% of the over 400,000 foil images in the SAH database, leaving SAH volunteers with a much more manageable set of images to look through. Even with these successes, our CNN presents clear paths for improvement. Particularly, the collection of more analog and true craters would increase the size of our training set and improve the effectiveness of our network. Another way to address the small amount of craters we currently have is to use techniques like GAN-based augmentation. This CNN might be a practical and convenient tool for future analyses of Stardust foils, and our CNN could be applied to sample

Fig. 9. Some promising images that may contain craters, as identified by our CNN. (a) and (d) are from the foil 11008W, (b) is from the foil 11126N, and (c) is from the foil 11020W. Prior to the implementation of the network, (b) had one approval and one disapproval, while (c) had no approvals and four disapprovals. (a) and (d) are from a data set not present on the SAH website. An “approval” means a volunteer stated that a crater exists in the image, and a “disapproval” means the volunteer stated no crater was present (Color figure can be viewed at wileyonlinelibrary.com.)
return missions with similar collectors, such as Genesis and Tanpopo. The success of deep learning methods on Stardust foils further gives optimism that deep learning methods can be applied to Stardust’s aerogel collector.

Acknowledgments—Computational work at the Advanced Light Source and Molecular Foundry at Lawrence Berkeley National Laboratory was supported by the U.S. Department of Energy under contract no. DE-AC02-05CH11231. This work was supported by NASA grant NNX16AI15G. We specifically thank David Shapiro for providing access to computing resources, allowing us to train the CNN.

Data Availability Statement—Our code base can be found at https://github.com/ljaeg/Foils-ConvNet

Editorial Handling—Dr. Donald Brownlee

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z. & Citro, C. et al. (2015) TensorFlow: large-scale machine learning on heterogeneous systems. tensorow.org.

Bechtel, H.A., Flynn, G.J., Allen, C., Anderson, D., Ansari, A., Bajt, S. et al. (2014) Stardust Interstellar Preliminary Examination III: infrared spectroscopic analysis of interstellar dust candidates. Meteoritics & Planetary Science, 49, 1548–1561.

Brenker, F.E., Westphal, A.J., Vincze, L., Burghammer, M., Schmitz, S., Schoonjans, T. et al. (2014) Stardust Interstellar Preliminary Examination V: XRF analyses of interstellar dust candidates at ESRF ID 13. Meteoritics & Planetary Science, 49, 1594–1611.

Brownlee, D. (2014) The Stardust mission: analyzing samples from the edge of the solar system. Annual Review of Earth and Planetary Sciences, 42, 179–205.

Brownlee, D.E., Horz, F., Newburn, R.L., Zolensky, M., Duxbury, T.C., Sandford, S. et al. (2004) Surface of young Jupiter family comet 81P/Wild 2: view from the Stardust spacecraft. Science, 304, 1764–1769.

Brownlee, D., Tsou, P., Aleon, J., Alexander, C.M.O., Araki, T., Bajt, S. et al. (2006) Comet 81P/Wild 2 under a microscope. Science, 314, 1711–1716.

Burchell, M.J., Cole, M.J., McDonnell, J. & Zarnecki, J.C. (1999) Hypervelocity impact studies using the 2 mv Van de Graaff accelerator and two-stage light gas gun of the University of Kent at Canterbury. Measurement Science and Technology, 10, 41.

Burchell, M.J., Cole, M.J., Price, M.C. & Kearsley, A.T. (2012) Experimental investigation of impacts by solar cell secondary ejecta on silica aerogel and aluminum foil: implications for the Stardust Interstellar Dust Collector. Meteoritics & Planetary Science, 47, 671–683.

Butterworth, A.L., Westphal, A.J., Tyliszczak, T., Gainsforth, Z., Stodolna, J., Frank, D.R. et al. (2014) Stardust Interstellar Preliminary Examination IV: scanning transmission X-ray microscopy analyses of impact features in the Stardust Interstellar Dust Collector. Meteoritics & Planetary Science, 49, 1562–1593.

Caruana, R., Niculescu-Mizil, A., Crew, G. & Ksikes, A. (2004) Ensemble selection from libraries of models. In Proceedings of the Twenty-First International Conference on Machine Learning, p. 18.

Chollet, F. et al. (2015) Keras. https://keras.io/getting_started/faq/#how-should-i-cite-keras

De Gregorio, B., Brintlinger, T. & Stroud, R. (2018) Fast automated identification of dust impact craters in aluminum foils from the Stardust collection (abstract #1255). 49th Lunar and Planetary Science Conference. CD-ROM.

Dietterich, T.G. (2000) Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems. Cagliari, Italy: Springer, pp. 1–15.

Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D. et al. (2016) Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Transactions on Medical Imaging, 35, 1182–1195.

Dozat, T. (2016) Incorporating nesterov momentum into Adam. In: Proceedings of 4th International Conference on Learning Representations, Workshop Track.

Elsila, J.E., Glavin, D.P. & Dworkin, J.P. (2009) Cometary glycine detected in samples returned by Stardust. Meteoritics & Planetary Science, 44, 1323–1330.

Flynn, G.J., Bleuet, P., Borg, J., Bradley, J.P., Brenker, F.E., Brennan, S. et al. (2006) Elemental compositions of comet 81P/Wild 2 samples collected by Stardust. Science, 314, 1731–1735.

Flynn, G.J., Sutton, S.R., Lai, B., Wirick, S., Allen, C., Anderson, D. et al. (2014) Stardust Interstellar Preliminary Examination VII: synchrotron X-ray fluorescence analysis of six Stardust interstellar candidates measured with the Advanced Photon Source 2-ID-D microprobe. Meteoritics & Planetary Science, 49, 1626–1644.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J. & Greenspan, H. (2018) Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321, 321–331.

Gainsforth, Z., Brenker, F.E., Simionovici, A.S., Schmitz, S., Burghammer, M., Butterworth, A.L. et al. (2014) Stardust Interstellar Preliminary Examination VIII: identification of crystalline material in two interstellar candidates. Meteoritics & Planetary Science, 49, 1645–1665.

Gainsforth, Z., Westphal, A.J., Butterworth, A.L., Jilly-Rehak, C.E., Brownlee, D.E., Joswiak, D. et al. (2019) Fine-grained material associated with a large sulfide returned from comet 81P/Wild 2. Meteoritics & Planetary Science, 54, 1069–1091.

Giusti, A., Cireşan, D.C., Maschi, J., Gambardella, L.M. & Schmidhuber, J. (2013) Fast image scanning with deep max-pooling convolutional neural networks. In: 2013 IEEE International Conference on Image Processing. Melbourne, Australia: IEEE, pp. 4034–4038.

Graham, G.A., Kearsley, A.T., Butterworth, A., Bland, P.A., Burchell, M.J., McPhail, D.S. et al. (2004) Extraction and microanalysis of cosmic dust captured during sample return missions: laboratory simulations. Advances in Space Research, 34, 2292–2298.

Grewal, M., Srivastava, M.M., Kumar, P. & Varadarajan, S. (2018) Radnet: radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Washington, DC: IEEE, pp. 281–284.
Ker, J., Singh, S.P., Bai, Y., Rao, J., Lim, T. & Wang, L.
Kearsley, A.T., Graham, G.A., Burchell, M.J., Cole, M.J.,
Joswiak, D.J., Brownlee, D.E., et al. (2006) Impact features on Stardust:
implications for comet 81P/Wild 2 dust. Science, 314, 1716–1719.
Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E. &
Weinberger, K.Q. (2017) SnapShot ensembles: train 1, get
m for free. arXiv preprint arXiv:1704.00109
Hunter, J.D. (2007) Matplotlib: a 2D graphics environment.
Computing in Science & Engineering, 9, 90–95.
Joswiak, D.J., Brownlee, D.E., Matrajt, G., Westphal, A.J.,
Snead, C.J. & Gainsforth, Z. (2012) Comprehensive
examination of large mineral and rock fragments in
Stardust tracks: mineralogy, analogous extraterrestrial
materials, and source regions. Meteoritics & Planetary
Science, 47, 471–524.
Kearsley, A.T., Borg, J., Graham, G.A., Burchell, M.J., Cole, M.J.,
Leroux, H. et al. (2008) Dust from comet Wild 2:
interpreting particle size, shape, structure, and composition
from impact features on the Stardust aluminum foils.
Meteoritics & Planetary Science, 43, 41–73.
Kearsley, A.T., Graham, G.A., Burchell, M.J., Cole, M.J.,
Wozniakiewicz, P., Teslich, N. et al. (2008) Micro-craters
in aluminum foils: implications for dust particles from
comet Wild 2 on NASA’s Stardust spacecraft.
International Journal of Impact Engineering, 35, 1616–1624.
Ker, J., Singh, S.P., Bai, Y., Rao, J., Lim, T. & Wang, L.
(2019) Image thresholding improves 3-dimensional
convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans.
Sensors, 19, 2167.
Kluvyer, T., Ragan-Kelley, B., Pérez, F., Granger, B.E.,
Bussonnier, M., Frederic, J. et al. (2016) Jupyter
notebooks—a publishing format for reproducible
computational workflows. In: Loizides, F. & Schmidt, B.
(Eds.) Power in Academic Publishing: Players, Agents and
Agendas: Proceedings of the 20th International Conference
on Electronic Publishing. Amsterdam, Netherlands: IOP
Press, pp. 87–90.
Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012) ImageNet
classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L. & Weinberger, K.Q.
(Eds.) Advances in neural information processing systems 25.
New York, NY: Curran Associates Inc., pp. 1097–1105.
Krogh, A. & Hertz, J.A. (1992) A simple weight decay can
improve generalization. Advances in Neural Information
Processing Systems 4, 950–957.
Landgraf, M., Müller, M. & Grün, E. (1999) Prediction of the
in-situ dust measurements of the Stardust mission to
comet 81P/Wild 2. Planetary and Space Science, 47, 1029–
1050.
Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z. & Tu, Z. (2015)
Deeply-supervised nets. Artificial Intelligence and Statistics,
562–570.
Leitner, J., Stephan, T., Kearsley, A.T., Hörz, F., Flynn, G.J.
& Sandford, S.A. (2008) TOF-SIMS analysis of crater residues from Wild 2 cometary particles on Stardust
aluminum foil. Meteoritics & Planetary Science, 43, 161–
185.
LeNail, A. (2019) Nn-svg: publication-ready neural network
architecture schematics. Journal of Open Source Software, 4, 747.
Leroux, H., Stroud, R.M., Dai, Z.R., Graham, G.A., Troadec, D.,
Bradley, J.P. et al. (2008) Transmission electron microscopy of cometary residues from micron-sized craters
in the Stardust Al foils. Meteoritics & Planetary Science,
43, 143–160.
Lin, M., Chen, Q. & Yan, S. (2013) Network in network. arXiv preprint arXiv:1312.4400.
Lo, M., Williams, B., Bollman, W., Han, D., Hahn, Y., Bell, J. et al. (1998) Genesis mission design. In: AIAA/AAS
astrodynamics specialist conference and exhibit. Boston,
MA: AIAA/AAS, pp. 4468–4718.
Mathis, J.S. & Whiffen, G. (1989) Composite interstellar
grains. The Astrophysical Journal, 341, 808–822.
McKeegan, K.D., Aleon, J., Bradley, J., Brownlee, D.,
Müller, M., Hörz, F., Bajt, S., Baratta, G.A. et al. (2008) TOF-SIMS analysis of crater residues from micron-sized craters
in the Stardust Al foils. Meteoritics & Planetary Science, 4718.
Ogliore, R.C., Floss, C., Stadermann, F.J., Kearsley, A.,
Leitner, J., Stroud, R.M. et al. (2012) Automated
searching of Stardust interstellar foils. Meteoritics & Planetary Science, 47, 729–736.
Ogliore, R., Huss, G., Nagashima, K., Butterworth, A.,
Gainsforth, Z., Stodolna, J. et al. (2012) Incorporation of
a late-forming chondrule into comet Wild 2. The
Astrophysical Journal Letters, 745, L19.
Payan, A. & Montana, G. (2015) Predicting Alzheimer’s
disease: a neuroimaging study with 3D convolutional
neural networks. arXiv preprint arXiv:1502.02506.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V.,
Thirion, B., Grisel, O. et al. (2011) Scikit-learn: machine
learning in Python. Journal of Machine Learning Research,
12, 2825–2830.
Perez, L. & Wang, J. (2017) The effectiveness of data
augmentation in image classification using deep learning.
arXiv preprint arXiv:1712.04621.
Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E. &
Nielson, M. (2013) Deep feature learning for knee cartilage
segmentation using a triplanar convolutional neural
network. In: International Conference on Medical Image
Computing and Computer-Assisted Intervention. Berlin:
Springer, pp. 246–253.
Price, M., Kearsley, A., Burchell, M., Hörz, F., Borg, J.,
Bridges, J. (2010) Distribution of finer (sub-10 micron)
dust collected by the Stardust spacecraft. Meteoritics & Planetary Science, 45, 1409–1428.
Roth, H.R., Lu, L., Seif, A., Cherry, K.M., Hoffman, J.,
Wang, S. et al. (2014) A new 2.5 D representation for
lymph node detection using random sets of deep convolutional neural network observations. In: Mori, K.,
Sakura, I., Yashinobu, S., Barillot, C. & Navab, N. (Eds.),
International Conference on Medical Image Computing and Computer-Assisted Intervention. Nagoya, Japan: Springer,
pp. 520–527.
Sandford, S.A., Aleon, J., Alexander, C.M.O’D., Araki, T.,
Bajt, S., Baratta, G.A. et al. (2006) Organics captured from
comet 81P/Wild 2 by the Stardust spacecraft. Science, 314, 1720–1724.
Shorten, C. & Khoshgoftaar, T.M. (2019) A survey on image
data augmentation for deep learning. Journal of Big Data, 6, 60.
Simon, S.B., Joswiak, D.J., Ishii, H.A., Bradley, J.P., Chi, M., Grossman, L. et al. (2008) A refractory inclusion returned by Stardust from comet 81P/Wild 2. *Meteoritics & Planetary Science*, 43, 1861–1877.

Simonyan, K. & Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singer, S., Stanley, J. & Kassel, P. (1985) The LDEF interplanetary dust experiment. In: Giese, R. & Lamy, P. (Eds.) *International Astronomical Union Colloquium*. Cambridge: Cambridge University Press, vol. 85, pp. 117–120.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.

Stadermann, F.J., Floss, C., Bose, M. & Lea, A.S. (2009) The use of Auger spectroscopy for the in situ elemental characterization of sub-micrometer presolar grains. *Meteoritics & Planetary Science*, 44, 1033–1049.

Stadermann, F.J., Hoppe, P., Floss, C., Heck, P.R., Hörz, F., Huth, J. et al. (2008) Stardust in Stardust: the C, N, and O isotopic compositions of Wild 2 cometary matter in Al foil impacts. *Meteoritics & Planetary Science*, 43, 299–313.

Stroud, R.M., Allen, C., Ansari, A., Anderson, D., Baji, S., Bassim, N. et al. (2014) Stardust Interstellar Preliminary Examination XI: identification and elemental analysis of impact craters on Al foils from the Stardust Interstellar Dust Collector. *Meteoritics & Planetary Science*, 49, 1698–1719.

Stroud, R., Koch, I., Bassim, N., Piccard, Y. & Nittler, L. (2010) Structure and composition of comet Wild 2 residues in sub-micron to micron-sized craters (abstract #1792). 41st Lunar and Planetary Science Conference. CD-ROM.

Tabata, M., Imai, E., Yano, H., Hashimoto, H., Kawai, H. & Kawaguchi, Y. et al. (2014) Design of a silica-aerogel-based cosmic dust collector for the Tanpopo mission aboard the International Space Station. *Transactions of the Japan Society for Aeronautical and Space Sciences*, Aerospace Technology Japan 12Pk 29-Pk 34.

The GIMP Development Team (2019) Gimp. https://www.gimp.org

The HDF Group (1997) Hierarchical data format, version 5. http://www.hdfgroup.org/HDF5/

Tompson, J., Goroshin, R., Jain, A., LeCun, Y. & Bregler, C. (2015) Efficient object localization using convolutional networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 648–656.

Tsou, P., Brownlee, D., Sandford, S., Hörz, F. & Zolensky, M. (2003) Wild 2 and interstellar sample collection and Earth return. *Journal of Geophysical Research: Planets*, 108 (E10), 1–21.

Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N. et al. (2014) scikit-image: image processing in Python. *PeerJ*, 2, e453.

Westphal, A.J., Anderson, D., Butterworth, A.L., Frank, D.R., Lettieri, R., Marchant, W. et al. (2014) Stardust Interstellar Preliminary Examination I: identification of tracks in aerogel. *Meteoritics & Planetary Science*, 49, 1509–1521.

Westphal, A.J., Bechtel, H.A., Brenker, F.E., Butterworth, A.L., Flynn, G., Frank, D.R. et al. (2014) Final reports of the Stardust Interstellar Preliminary Examination. *Meteoritics & Planetary Science*, 49, 1720–1733.

Westphal, A.J., Butterworth, A.L., Snead, C.J., Craig, N., Anderson, D. & Jones, S.M. et al. (2005) Stardust@home: a massively distributed public search for interstellar dust in the Stardust Interstellar Dust Collector (abstract #1908). 36th Lunar and Planetary Science Conference. CD-ROM.

Westphal, A.J., Stroud, R.M., Bechtel, H.A., Brenker, F.E., Butterworth, A.L., Flynn, G.J. et al. (2014) Evidence for interstellar origin of seven dust particles collected by the Stardust spacecraft. *Science* 345:786–791.

Wong, S.C., Gatt, A., Stamatescu, V. & McDonnell, M.D. (2016) Understanding data augmentation for classification: when to warp? In 2016 International Conference on Digital Image Computing: Techniques and applications (DICTA), IEEE, pp. 1–6.

Wozniakiewicz, P.J., Ishii, H.A., Kearsley, A.T., Burchell, M.J., Bradley, J.P., Price, M.C. et al. (2012) Stardust impact analogs: resolving pre-and postimpact mineralogy in Stardust Al foils. *Meteoritics & Planetary Science*, 47, 708–728.

Yamagishi, A., Yano, H., Okudaira, K., Kobayashi, K., Yokobori, S.-I. & Tabata, M. et al. (2009) Tanpopo: astrobiology exposure and micrometeoroid capture experiments. *Transactions of the Japan Society for Aeronautical and Space Sciences*, 7, Tk49-Tk-55.

Yu, X., Wu, X., Luo, C. & Ren, P. (2017) Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience & Remote Sensing*, 54, 741–758.

Zolensky, M.E., Zega, T.J., Yano, H., Wirick, S., Westphal, A.J., Weisberg, M.K. et al. (2006) Mineralogy and petrology of comet 81P/Wild 2 nucleus samples. *Science*, 314, 1735–1739.