

# Toward Generalized Change Detection on Planetary Surfaces With Convolutional Autoencoders and Transfer Learning

Hannah Rae Kerner<sup>1</sup>, Kiri L. Wagstaff<sup>2</sup>, Brian D. Bue<sup>2</sup>, Patrick C. Gray<sup>2</sup>, James F. Bell III<sup>3</sup>, and Heni Ben Amor<sup>4</sup>

**Abstract**—Ongoing planetary exploration missions are returning large volumes of image data. Identifying surface changes in these images, e.g., new impact craters, is critical for investigating many scientific hypotheses. Traditional approaches to change detection rely on image differencing and manual feature engineering. These methods can be sensitive to irrelevant variations in illumination or image quality and typically require before and after images to be coregistered, which itself is a major challenge. Additionally, most prior change detection studies have been limited to remote sensing images of earth. We propose a new deep learning approach for binary patch-level change detection involving transfer learning and nonlinear dimensionality reduction using convolutional autoencoders. Our experiments on diverse remote sensing datasets of Mars, the moon, and earth show that our methods can detect meaningful changes with high accuracy using a relatively small training dataset despite significant differences in illumination, image quality, imaging sensors, coregistration, and surface properties. We show that the latent representations learned by a convolutional autoencoder yield the most general representations for detecting change across surface feature types, scales, sensors, and planetary bodies.

**Index Terms**—Change detection algorithms, earth, machine learning, mars, moon, neural networks, remote sensing, supervised learning, unsupervised learning.

## I. INTRODUCTION

ONGOING planetary imaging investigations such as the High Resolution Imaging Science Experiment (HiRISE) [1] and ConTeXt Camera (CTX) [2] on the Mars Reconnaissance Orbiter are returning large volumes of image data that continue to

grow faster than scientists can analyze and categorize. There is a need for systems that can rapidly and intelligently analyze these data and prioritize observations of interest to scientists. Change detection, the process of automatically identifying changes in surface features between two images collected over the same location at different points in time, is a critical tool for analyzing these data. For example, recurring slope lineae (RSL) are narrow, low-albedo features observed on Mars that extend from bedrock to incrementally lengthen down steep slopes and are thought to be formed by shallow subsurface liquid water flows [3]. RSL appear to lengthen on timescales of several months and appear/disappear on timescales close to one year [3]. Because of their implications for the past and present history of water on Mars, RSL are key features that scientists are actively monitoring for changes as they develop and evaluate theories on RSL formation and growth mechanisms [e.g., Fig. 1(a)]. A system that automatically detects RSL in images from HiRISE or other high-resolution imaging systems could help scientists better understand where RSL occurs, how they evolve, and how or why they form in the first place.

New meteorite impacts are another key surface feature that scientists monitor for change. Daubar *et al.* [4] reported the discovery of 248 new impact sites that formed on Mars within the last few decades [e.g., Fig. 1(b)]. The landscape of Mars and other planets continues to be altered by impact events. Speyerer *et al.* reported the discovery of 222 new impact sites on the moon, which was 30% more impacts than was expected based on current estimates of the current cratering rate on the moon [5] [e.g., Fig. 1(c)]. Documenting when and where new impacts occur helps scientists to refine estimates of the past and present cratering rates in the solar system, which in turn enables improved age estimates of important events in the solar system's history [4], [5].

High spatial and temporal resolution imaging of the earth by government and commercial satellites enables observation of countless surface features that change due to natural or human-induced processes. Detecting changes in surface features on the earth—e.g., new construction, fires, or volcanic eruptions [e.g., Fig. 1(d)]—is important for our scientific understanding of the earth as well as for many commercial, humanitarian, and defense applications.

Popular approaches to change detection are difference-based methods, which compare differences in pixel intensities between

Manuscript received April 1, 2019; revised June 18, 2019; accepted August 7, 2019. Date of publication September 8, 2019; date of current version November 22, 2019. This work was supported by the Jet Propulsion Laboratory, California Institute of Technology, Internal Strategic University Research Partnerships (SURP) program under a contract with the National Aeronautics and Space Administration. (Corresponding author: Hannah Rae Kerner.)

H. R. Kerner is with the Department of Geographical Sciences, University of Maryland, College Park, MD 20742 USA (e-mail: hkerner@umd.edu).

K. L. Wagstaff and B. D. Bue are with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA (e-mail: kiri.l.wagstaff@jpl.nasa.gov; bbue@jpl.nasa.gov).

P. C. Gray is with the Nicholas School of the Environment, Duke University, Durham, NC 27710 USA (e-mail: patrick.c.gray@duke.edu).

J. F. Bell III is with the School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85282 USA (e-mail: Jim.Bell@asu.edu).

H. Ben Amor is with the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85282 USA (e-mail: hbenamor@asu.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2936771

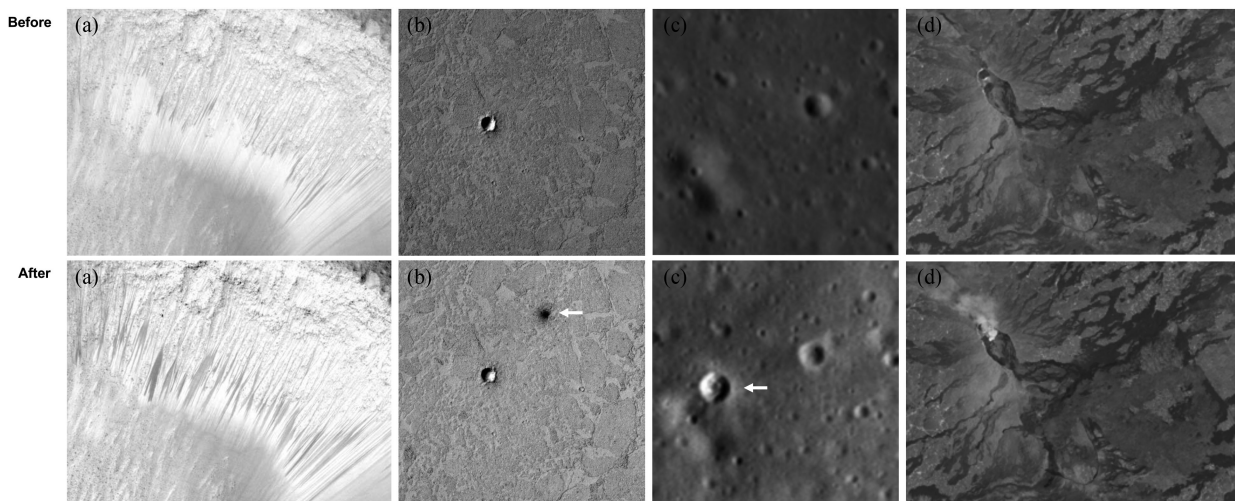


Fig. 1. Example before and after images from the (a) HiRISE RSL (Mars), (b) CTX impacts (Mars), (c) LROC impacts (Moon), and (d) PlanetScope (Earth) datasets. Image IDs from Table II ESP\_030769\_1685, ESP\_031059\_1685, P06\_003451\_2035\_XN\_23N171W, P13\_006286\_2073\_XN\_27N171W, M181330922L, and M1139065512L, ertaale-20170116-before, and ertaale-20170123-after.

images acquired at two different times, and postclassification comparison methods, which compare object or land-cover classifier predictions at the image or pixel level. Many machine learning approaches exist for detecting change between remote sensing images, including kernel-based methods such as support vector machines (SVMs) as well as deep learning methods. In general, prior approaches to change detection have at least one of the following limitations: they require large amounts of training data to pretrain classifiers, they require classes for each type of surface feature being monitored for change to be defined, they require images to be precisely coregistered, or they are sensitive to irrelevant variations caused by illumination and processing artifacts. Furthermore, the majority of prior work for change detection in remote sensing images has been limited to images of earth and have not been evaluated for images of terrestrial planets beyond earth (e.g., Mars and the Moon).

The goal of our approach is to learn general representations of bitemporal image pairs (i.e., pairs of images acquired on two different dates) that are useful for identifying when changes in surface features have occurred on a planetary body. Using diverse datasets of remote sensing images of Mars, the moon, and earth, we find that existing difference-based approaches may be limited in their ability to generalize to new datasets. We propose a new deep learning approach to change detection that uses representations of bitemporal images and transfer learning to detect surface feature changes in diverse remote sensing datasets using a relatively small number of labeled training examples. We performed several experiments to characterize the capacity for the proposed representations to enable generalization to new surface feature types, imaging sensors, resolutions, level of coregistration, and planetary bodies unseen during training.

## II. RELATED WORK

Change detection is an active area of research in remote sensing, and there is a large body of research on its methods and applications [6]–[9]. Tewkesbury *et al.* [9] proposed organizing change detection literature in terms of the *comparison method*,

describing the method used to determine if a change in surface features has occurred between two images, and the *unit of analysis*, describing the image representation that will be analyzed by the comparison methods. Raw pixel intensities, obtained from images that have had little to no preprocessing applied, are the most common units of analysis and have been used widely since the beginning of remote sensing change detection research (e.g., [9]–[14]). Difference images, in which each pixel represents the difference between corresponding pixel intensities in a pair of images, are also a common unit of analysis [9]. Thresholding is a common change detection method in approaches where the unit of analysis is the difference in pixel values (e.g., [15]). If the difference between two images for one pixel location is greater than a threshold  $t$ , which pixel is classified as a *change* pixel. If the difference is less than or equal to  $t$ , the pixel is classified as a *no-change* pixel. This threshold can be derived empirically as one (or more) standard deviations from the mean pixel difference in a distribution of pixel differences across each band of an image [6]. When comparing pixel values between before and after images directly, it is common to use a threshold on the ratio between the corresponding pixel values to classify pixels that constitute change in surface features as in [5]. While these approaches are computationally fast, they are sensitive to changes caused by misregistration, illumination, and image artifacts that are not relevant for assessing surface feature change [16], which we demonstrate using diverse datasets of remote sensing images of multiple planetary bodies and surface feature types.

As the spatial resolution of satellite cameras improved and pixels sampled smaller surface areas, researchers began developing higher level representations from pixels that were more suitable for detecting land-cover classes, made comparison methods less sensitive to misregistration, and reduced noise that contributed to false detections [17]. Land-cover classifiers have been used widely to produce labels that can be compared between before and after images, e.g., [18]–[22]. Comparing the land-cover classes predicted for the before and after image by a classifier is known as postclassification comparison

and is widely used in the remote sensing change detection literature. Postclassification approaches that produce pixelwise labels suffer from misregistration issues in postclassification comparison because the labels are at the same resolution of the input image. The unique advantage of this approach is that the semantics of the change are provided with the detection. The primary disadvantage of postclassification comparison methods is that performance depends on the accuracy of the land-cover classifier since the classification error is compounded when the predictions are compared for change classification [6], [9]. Additionally, it is difficult and computationally expensive to learn a new classifier to predict land-cover classes for new applications and image datasets. For planets beyond Earth, there are some surface features such as RSL (which we discuss in this article) or dust devils on Mars for which relatively few images exist for training a classifier, which can preclude the use of postclassification comparison methods.

Object-based methods are another category of change detection approaches driven by the availability of high-resolution satellite images [8], [9]. In object-based approaches, image objects are identified before applying a comparison method to image pairs or sequences. Castilla and Hay [17] presented an analysis of how image objects derived from pixels in remote sensing images relate to geographic objects in ground truth. Walter [23] used objects already present in a GIS database to compare land-use classes in multitemporal images. Klaric *et al.* [24] proposed a framework that extracts high-level statistical features from images that could be collectively compared and clustered to detect and describe change in large volumes of remote sensing images.

Many successful machine learning approaches have been proposed for change detection in remote sensing images. In [25]–[27], kernel methods were used to model the temporal relationships between spatial features in images obtained from multiple sensors. In [28] and [29], the authors proposed nonlinear change classifiers using SVMs and kernel functions. Im and Jensen [30] combined neighborhood correlation images with decision trees to detect change based on correlations between local spatial and temporal patterns. Bovolo and Brozzone [31] presented a theoretical analysis of change vectors and their use for unsupervised change detection. Though not an automated method, Heyer *et al.* [32] presented a web-based tool for scientists to visually analyze overlapping orbital images of Mars and identify potential changed regions.

Deep neural networks are machine learning models that learn hierarchical representations of input data by composing layers of nonlinear transformations, the parameters of which are optimized during model training. Methods that employ deep neural networks are often referred to as “deep learning” methods. In contrast to traditional machine learning methods, deep learning methods learn features (e.g., for classification or regression) from the input data automatically, rather than using manually engineered features [33]. Many supervised deep learning approaches have been proposed for change detection in remote sensing images (e.g., [34]–[38]). In [39]–[41], supervised and unsupervised deep learning methods are combined to learn intermediate representations of before and after images that are useful for identifying important changes between the images.

Zhang [42] used a combined feed-forward neural networks and  $K$ -means clustering for unsupervised difference representation learning between before and after images. Saha [43] used a pretrained deep neural network to extract high-level feature vectors from before and after images that were used for change vector analysis. In [44], an autoencoder neural network, which learns a mapping from input images to reduced-dimension latent representations by minimizing the difference between the input and autoencoder reconstruction, was used to learn the correspondence between features in before and after images in an unsupervised manner; in contrast, we use a *convolutional* autoencoder (CAE) in this article to learn the salient features of before and after images independently as a preprocessing step for a separate change classifier. Siamese neural networks have also been recently proposed for change detection in remote sensing images [45]–[47]. Siamese networks consist of two identical networks that extract features from two inputs that are combined into one network using a similarity or distance layer. The conjoining layer minimizes differences between similar examples to ensure examples that are similar in the input space are also close in the embedded space [48]. Another important direction for deep learning approaches to change detection makes use of recurrent neural networks. Recurrent neural networks maintain information about prior states of network activations, rather than relying on the output from the previous layer alone as is the case for traditional neural networks [49]. Prior work proposed using recurrent neural networks to model temporal dependencies between surface features in multitemporal remote sensing images [40], [50], [51]. Rather than using a deep neural network to identify classes of surface features in remote sensing images, we propose using a deep neural network to identify *change* in a variety of surface features without specifying the feature type. We propose several input representations of bitemporal image pairs that enable change classification and compare how each representation enables the classifier to generalize beyond its limited training dataset.

### III. APPROACH

#### A. Fine-Tuned Neural Network for Change Detection

We approached change detection as a binary classification problem. Given an input containing information about a bitemporal pair of images, a binary classifier should predict 1 if there is a change in surface features, and predict 0 if there are no changes in surface features. Surface changes are changes to morphologies on the surface that are present in the image, e.g., RSL or impact craters. There might be other differences between a pair of images, e.g., illumination or resolution differences, that are not considered surface changes and should not result in change detection.

Deep neural networks are very effective at building hierarchical representations of complex data. Previous work has shown that the features learned by a neural network when trained on a large dataset of natural images in an object dataset, such as the ImageNet database [52], are generic and can be useful for classification tasks that may be very different from the original task (e.g., [53] and [54]). Zhu *et al.* [35] surveyed the use of pretrained neural networks for remote sensing tasks. Audebert *et al.* [55] and Volpi and Tuia [56] presented segmentation



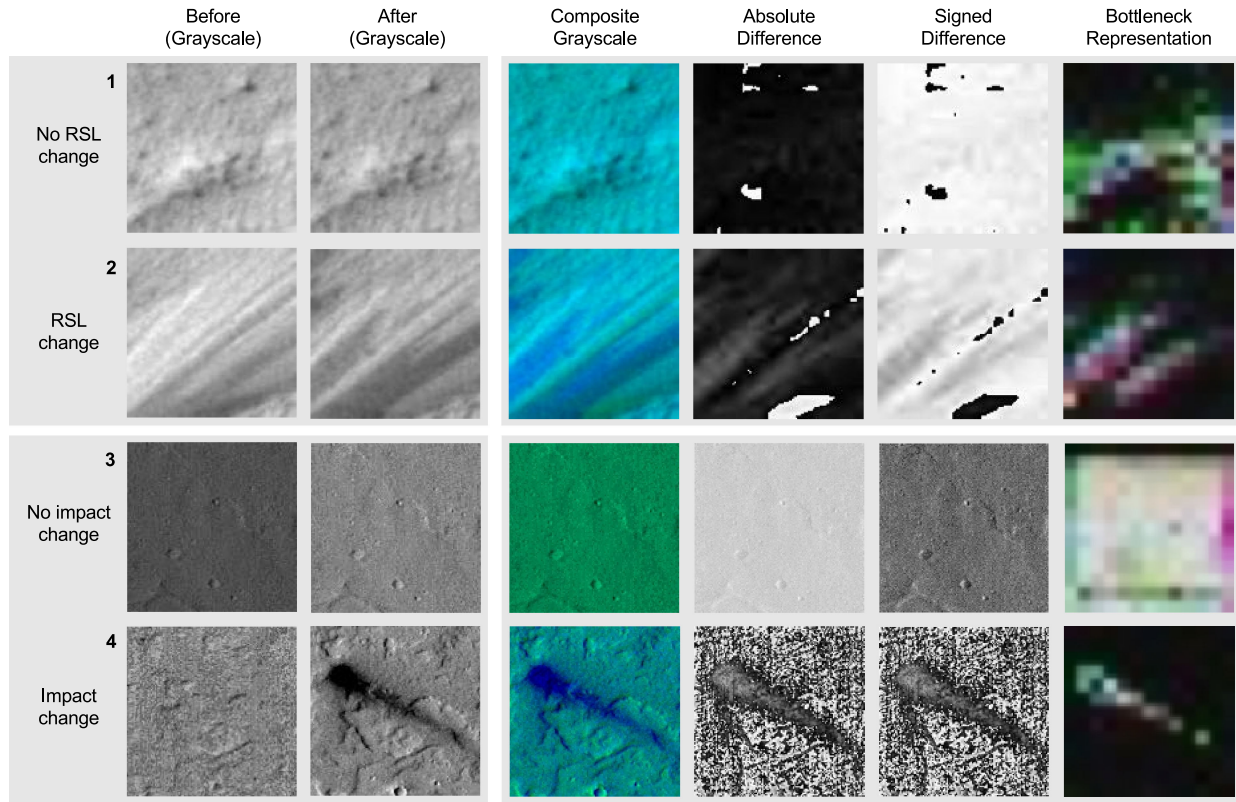


Fig. 2. Top section: four representations of two example image pairs from the HiRISE RSL dataset, one *without* RSL change (row 1) and one *with* RSL change (row 2). Bottom section: four representations of two example image pairs from the CTX Fresh impacts dataset, one *without* impact change (row 3) and one *with* impact change (row 4).

networks pretrained on remote sensing images; however, a formal study of the transferability of features learned by networks pretrained on remote sensing images has not been conducted as has been done for object datasets such as ImageNet (e.g., [54]). The process of refining the weights in a pretrained neural network for a new task by updating some or all of the network weights via training on the smaller, task-specific dataset is referred to as “fine-tuning.” Penatti *et al.* [57] showed that the features learned by architectures pretrained on object datasets are also useful for aerial and remote sensing images, but were outperformed by low-level color descriptors for remote sensing images. Fine-tuned networks were successfully employed for land-cover classification in remote sensing images in [58]–[62]. In contrast to prior work, we propose using fine-tuned networks to classify general *change* in surface features rather than a specific surface feature or set of features. We initialized the Inception-v3 network [63] with weights learned from pretraining on the ImageNet database and replaced the final softmax layer of Inception-v3 with a new softmax layer with two outputs (for *change* and *no-change*). We fine-tuned the Inception-v3 network for our change detection task by optimizing weights for this new softmax layer using our change detection training and validation datasets.

### B. Input Image Representations

The Inception-v3 network requires 3-channel input images, which are typically RGB color images. Our goal is to represent

a pair of images as a single three-channel image that can be classified directly. We considered four representations of before/after image pairs as the input to Inception-v3: composite grayscale, absolute difference, signed difference, and autoencoder bottleneck maps (see Fig. 2). These representations are analogous to the “units of analysis” discussed in Section II and constitute various levels of abstraction from the original image pair. The image filters learned by ImageNet pretraining have been shown to be useful for application to remote sensing and other unrelated image datasets (e.g., [54], [57]–[62], [64]–[66]). In Section V, we report on an evaluation of which representation is most suitable for classifying change between the image pair.

1) *Composite Grayscale*: The datasets described in Section IV contain  $100 \times 100$ -pixel grayscale images. We can use this representation directly. Since the input to Inception-v3 is a single 3-channel image, we created a composite of the before and after grayscale images in which the blue channel contains the before image, the green channel contains the after image, and the red channel contains all zeros.

2) *Absolute Difference*: In this approach, we computed the absolute value of the difference between each pixel in the grayscale before and after images. The resulting single-channel image is replicated in each of the three input channels.

3) *Signed Difference*: As in the previous approach, we computed the difference between each pixel in the grayscale before and after images. Because image formats do not allow negative values, we rescaled the difference values (which nominally



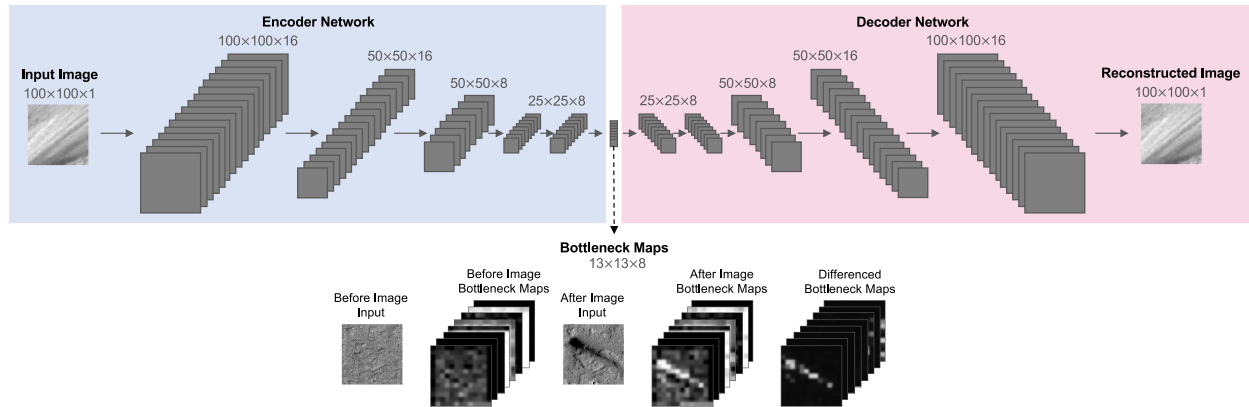


Fig. 3. Convolutional autoencoder architecture.

range from  $-255$  to  $+255$ ) into the range  $[0,255]$ . The resulting single-channel image is replicated in each of the three input channels.

4) *Autoencoder Bottleneck Maps*: A CAE is a type of self-supervised neural network that learns a low-dimensional representation or “code” capturing the most salient features in a dataset by optimizing its reconstruction of the original input from the learned encoding, using convolutional layers for feature extraction [67]. Autoencoders are sometimes called “encoder–decoders” because they consist of an encoder network that projects the input into the low-dimensional “bottleneck” representation and a decoder network that projects from the bottleneck representation back up to the input space (see Fig. 3). Once an autoencoder is trained to reconstruct all examples in a dataset well, the encoder network can be used for dimensionality reduction. This is similar to dimensionality reduction through projection into the eigenspace in principal component analysis [68]. By training an autoencoder to produce representations at the bottleneck layer that capture the salient features in the image, we can take advantage of the large dataset of *un-labeled* patches in the HiRISE RSL dataset to produce a potentially more refined representation for the downstream change detection classifier.

We trained a CAE with images indicated in Column 4 of Table II. The encoder part of the network consists of three sequences of  $3 \times 3$  convolution, batch normalization [69], and  $2 \times 2$  max pooling. In the decoder part of the network, we use the same three sequences but instead of max pooling we upsample using nearest-neighbor interpolation. We generated bottleneck map representations by applying only the encoder function of the trained network. Table I lists the size of image representations following each layer in the autoencoder. The dimension at the bottleneck layer is  $13 \times 13 \times 8$ , which is  $\sim 7.4$  times smaller than the input dimension. We selected the three (of eight) bottleneck maps that were most discriminative for change in surface features to populate the three input channels to Inception-v3. To determine which maps were most discriminative, for each of the eight bottleneck maps, we computed two distributions: one of the mean squared error between the before and after image maps for the *no-change* examples in the validation dataset, and one for the *change* examples in the validation dataset. We computed the Kullback–Leibler (KL) divergence between these

TABLE I  
OUTPUT SIZE OF EACH LAYER IN CAE

Layer Name	Layer Output Size
Input	$100 \times 100 \times 1$
$3 \times 3$ Convolution (1)	$100 \times 100 \times 16$
Batch Normalization (1)	$100 \times 100 \times 16$
Max Pool (1)	$50 \times 50 \times 16$
$3 \times 3$ Convolution (2)	$50 \times 50 \times 8$
Batch Normalization (2)	$50 \times 50 \times 8$
Max Pool (2)	$25 \times 25 \times 8$
$3 \times 3$ Convolution (3)	$25 \times 25 \times 8$
Batch Normalization (3)	$25 \times 25 \times 8$
Max Pool (3)	$13 \times 13 \times 8^*$
$3 \times 3$ Convolution (4)	$13 \times 13 \times 8$
Batch Normalization (4)	$13 \times 13 \times 8$
Up Sample (1)	$25 \times 25 \times 8$
$3 \times 3$ Convolution (5)	$25 \times 25 \times 8$
Batch Normalization (5)	$25 \times 25 \times 8$
Up Sample (2)	$50 \times 50 \times 8$
$3 \times 3$ Convolution (6)	$50 \times 50 \times 16$
Batch Normalization (6)	$50 \times 50 \times 16$
Up Sample (3)	$100 \times 100 \times 8$
$3 \times 3$ Convolution (7)	$100 \times 100 \times 1$

Asterisk indicates bottleneck representation.

two distributions for each of the eight maps and selected the three maps with the largest KL divergence. We up-sampled each of these three feature maps to  $100 \times 100$  pixels and populated the three input channels with these three maps.

### C. Baseline Change Detection

Approaches based on image differencing are the most widely used change detection approaches [6]. Since there is no widely accepted baseline method for change detection, we designed two baselines based on common methods in the change detection literature.

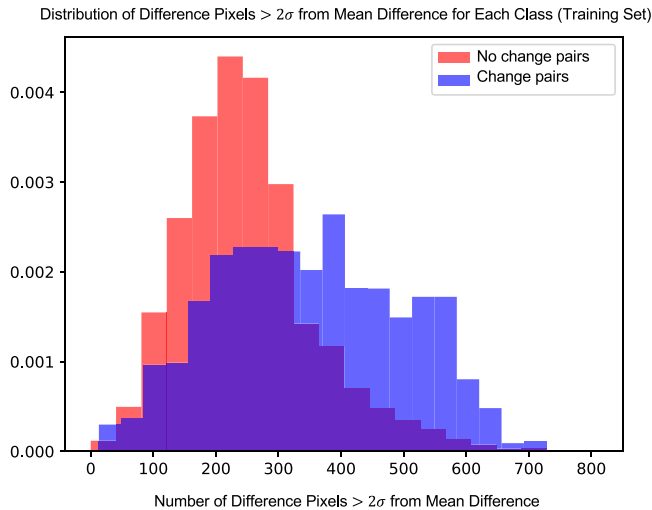


Fig. 4. Distribution of difference pixels greater than two standard deviations from the mean in each image pair for *change* and *no-change* examples (see Experiment 1).

1) *Naive Bayes With High-Change Pixel Values*: The underlying assumption for difference-based change detection methods is that the difference in pixel intensity at locations where surface change occurred will be significantly higher than at locations where change did not occur (or only occurred due to illumination or other irrelevant changes). We designed the following procedure for change detection based on this principle.

- 1) Apply local contrast normalization to before and after images using the Contrast Limited Adaptive Histogram Equalization Algorithm [70].
- 2) Compute pixelwise difference between before and after images in training set.
- 3) Compute mean and standard deviation of distribution of pixelwise differences for each image pair.
- 4) Compute number of pixels  $n$  in which the pixelwise difference is greater than two standard deviations from the mean pixel difference for each image pair.
- 5) Fit conditional Gaussian probability distributions  $p(x = n|y = \text{change})$  and  $p(x = n|y = \text{no-change})$  (see Fig. 4).
- 6) Use Naive Bayes [71] to predict class label for test examples using posterior probabilities:  

$$p(y = \{\text{change}, \text{no-change}\} | x = n).$$

2) *SVM With Difference Image*: SVMs have been utilized in prior work for detecting changes at the pixel and image level using differenced or filtered before/after images as input (e.g., [28] and [29]). In a classification setting, SVMs use kernel functions to transform inputs to a higher dimensional space in which the classifier finds a linear decision boundary. SVMs typically generalize well to unseen examples because they find a decision boundary with maximal margin between the training data and decision boundary [72]. We used an SVM classifier to make image-level classifications of *change* or *no-change* using the absolute difference between pixels in the before and after images, represented as a  $1 \times d$  feature vector where  $d$  is the number of pixels in the difference image. We used the radial basis function kernel and determined the hyperparameters for

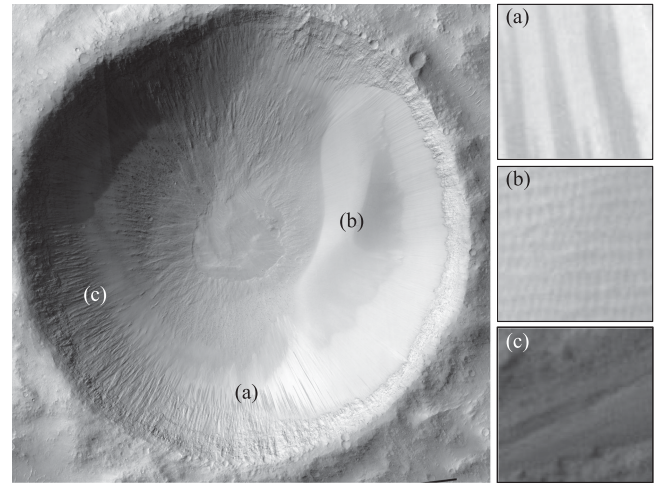


Fig. 5. Example of  $100 \times 100$ -pixel patches sampled from HiRISE image of Garni crater (ESP\_027802\_1685) where (a) RSL occur, (b) where RSL do not occur, and (c) where it is difficult to distinguish RSL from other shadowed topography.

the SVM using grid search and threefold cross validation (see Appendix B for details).

We performed each experiment in Section V using these baseline methods *with* local contrast normalization and *without* local contrast normalization (“No LCN”).

#### IV. DATASETS

##### A. HiRISE: Recurring Slope Lineae (RSL), Mars

RSL are dark, narrow features (typically 0.5–5 m) that incrementally lengthen down steep slopes and fade/recur throughout the year. They are thought to be formed by shallow subsurface liquid water flows [3]. Scientists are actively monitoring RSL for changes as they develop and evaluate theories on RSL formation and growth mechanisms [e.g., Fig. 1(a)]. We created a dataset for change detection of RSL using repeat observations of a well-studied site with known RSL activity called Garni Crater in Valles Marineris, Mars [73]. These observations were made by the HiRISE camera onboard the Mars Reconnaissance Orbiter. HiRISE has a spatial resolution of  $\sim 30$  cm/pixel and three color channels: red (550–850 nm), blue-green (400–600 nm), and near infrared (800–1000 nm) [1]. We used the HiRISE red channel orthorectified products in Table II corresponding to 12 different acquisition dates. We used the red band because it has the highest spectral coverage and signal-to-noise ratio [1]. We cropped each image to the  $10\,000 \times 10\,000$ -pixel boundary of Garni crater and converted to grayscale. Since RSL are relatively small, we subsampled  $100 \times 100$ -pixel patches from the cropped images of Garni crater for our experiments (e.g., Fig. 5). We subsampled using a sliding window with a stride size of 50 pixels. This resulted in a dataset of 39 601 patches for each of the 12 images (475 212 total). The process of subsampling large remote sensing images for change detection has also been employed in prior work, e.g., [66], [74]–[76]. These image patches were used to train the CAE and change detection classification models, as we describe in more detail in Section IV. In Table II, Column 4 indicates which (unlabeled) images were used to train the

TABLE II  
IMAGE PRODUCTS USED IN THIS ARTICLE

Instrument	Image ID	Date Acquired	Training		Validation	Testing
			Autoencoder	Classifiers		
HiRISE	ESP_027802_1685	07/01/2012		•		
HiRISE	ESP_028501_1685	08/25/2012		•		
HiRISE	ESP_029213_1685	10/19/2012	•			•
HiRISE	ESP_029780_1685	12/02/2012				•
HiRISE	ESP_030347_1685	01/16/2013			•	
HiRISE	ESP_030769_1685	02/18/2013			•	
HiRISE	ESP_031059_1685	03/12/2013	•			
HiRISE	ESP_031771_1685	05/07/2013	•			
HiRISE	ESP_032048_1685	05/28/2013	•			
HiRISE	ESP_032615_1685	07/11/2013	•			
HiRISE	ESP_034184_1685	11/11/2013	•			

Bullets indicate which part of the article the image was used for.

TABLE III  
NUMBER OF EXAMPLES LABELED IN EACH CLASS FOR THREE IMAGE  
PAIRS FROM HiRISE RSL DATASET

Before Image	After Image	Change	No-Change
ESP_027802_1685	ESP_028501_1685	203	3,333
ESP_030347_1685	ESP_030769_1685	73	613
ESP_029213_1685	ESP_029780_1685	75	179

autoencoder and which (labeled) images were used to train the classification models. Columns 5 and 6 indicate which (labeled) images were used for validation and testing with respect to the classification models.

We selected three bitemporal pairs that exhibited changes in RSL for labeling (see Table III). We selected bounding boxes surrounding the changed regions and labeled bitemporal patches as having *change* or *no-change* within these bounds by inspecting animated GIFs transitioning between before and after tiles. We labeled 203 *change* and 3333 *no-change* patches for training, 73 *change* and 613 *no-change* patches for validation, and 75 *change* and 179 *no-change* patches for testing (see Table III).

### B. CTX: Meteorite Impacts, Mars

The landscape of Mars and other planets is continually altered by meteorite impacts [see Fig. 1(b)]. Scientists use data about when and where these impacts occur to refine estimates of the current cratering rate in our solar system and constrain the impact production function over time [4]. In one study of contemporary impact cratering, Daubar *et al.* [4] reported 248 new impact sites that formed on Mars within the last few decades discovered by comparing images taken over the same location at different times from multiple instrument datasets. We selected eight image pairs (16 images) from Daubar *et al.*'s study in which both the before and after images were taken by the CTX onboard the Mars Reconnaissance Orbiter (see Table VII). CTX has a spatial resolution of  $\sim 6$  m/pixel and

a single channel (500–700 nm) [2]. We map projected and coregistered each image pair using the PIXL Visual Precision Targeting algorithm [77]. While this algorithm produces good coregistration for most regions in the image, some regions are not perfectly coregistered (e.g., Fig. 6). For each image pair, we cropped one  $150 \times 150$ -pixel patch around the approximate center of the impact and six  $150 \times 150$ -pixel patches at random locations in the nonimpact region of the image, then resized the patches to  $100 \times 100$  pixels. We augmented the eight *change* patches with horizontally flipped, vertically flipped, and  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  rotated versions of each tile. We did not augment the *no-change* tiles. This resulted in 48 positive (*change*) and 48 negative (*no-change*)  $100 \times 100$ -pixel image pairs in this dataset. We used this dataset to assess generalization to surface features, image sensors, and resolution that were different than in the training set.

### C. LROC: Impacts, Moon

Speyerer *et al.* [5] found that the current lunar cratering rate is significantly higher than previously thought. These new estimates, informed by updated images of the lunar surface and automatic classification methods, improved models of current cratering rates and surface regolith turnover. These models are used to constrain the ages of surface features on the moon and other planetary bodies. To test the transfer performance of the proposed change detection approaches, we created a dataset of lunar impacts captured using the Lunar Reconnaissance Orbiter Camera's (LROC) Narrow Angle Camera (NAC) [78]. LROC has two identical NACs (for stereo imaging) that collect  $\sim 0.5$  m/pixel panchromatic images. Since the LROC mission began, the NACs have collected a repository of over 1.7 million 500-megapixel images.<sup>1</sup> A subset of these images were captured over the same region under similar illumination conditions at different points in time. We selected five of these image pairs in which an impact occurred due to a meteorite or the Chang'e

<sup>1</sup>[Online]. Available: <http://wms.lroc.asu.edu/lroc/thumbnails>



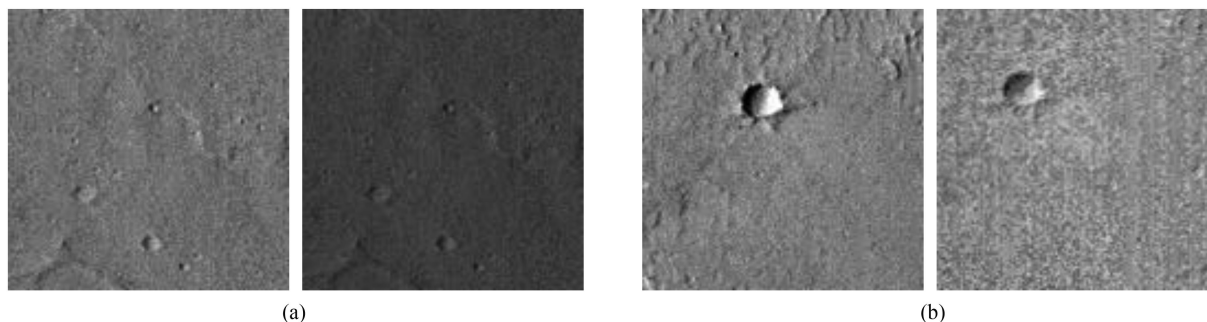


Fig. 6. Example CTX patches where before and after images are (a) coregistered perfectly and (b) not coregistered perfectly. In (b), the large crater in the first image is shifted up and to the left in the second image, and many small details in the first image do not appear in the second image). Both patches were cropped from the same two images: P09\_004477\_1906\_XN\_10N100W and P13\_006178\_1907\_XN\_10N100W (see Table VII).

lander [79] as reported in Speyerer *et al.* [5] [e.g., Fig. 1(c)]. In Table VII, we give the NAC ID for each image as well as the name used by the LROC mission for each pair in parentheses. These images were map-projected, radiometrically calibrated, and coregistered to within 20 m ( $\sim 40$  pixels). For each pair of images, we cropped one  $100 \times 100$ -pixel patch around the approximate center of the impact and one  $100 \times 100$ -pixel patch at a random location in the nonimpact region of the image. This resulted in five *change* and five *no-change*  $100 \times 100$ -pixel image pairs in this dataset. We used these images to assess generalization to a different planetary body (see Experiment 3).

#### D. PlanetScope: Miscellaneous Processes, Earth

The earth’s surface is continually undergoing change from natural geologic processes as well as human activity. Planet Inc. is a commercial remote sensing company that operates a constellation of small satellites (“cubesats”) that acquire daily color images of the earth’s surface using their PlanetScope camera [80]. These images have  $\sim 3$ -m spatial resolution and four bands: red, green, blue, and near-infrared. The sun-synchronous orbit of the satellites enables a consistent equatorial overpass time between 9:30 and 11:30 A.M. local time [81], so images acquired on different dates typically have similar illumination conditions. We manually identified before and after images of four different locations taken at different points in time that showed changes in surface features on the earth [e.g., Fig. 1(d)]; we give the names of these images in Table VII.<sup>2</sup> We chose desert regions where the landscape would be most similar to the moon and Mars (Nevada, Saudi Arabia, Ethiopia, and the Gobi desert). These images exhibit surface feature change due to fire, oil production, volcanic eruptions, and solar array construction. We cropped one  $200 \times 200$ -pixel patch around a *change* feature in each before/after pair and one  $200 \times 200$ -pixel patch in a region of the image pair without changes in surface features, then resized the patches to  $100 \times 100$  pixels.<sup>3</sup> We converted these RGB images to grayscale. This resulted in five *change* and five

*no-change*  $100 \times 100$ -pixel image pairs, which we used to assess generalization to a different planetary body (see Experiment 3).

The source images for these four datasets are publicly available and we provided instructions for accessing them in Appendix A. The datasets are publicly available online.

## V. EXPERIMENTS

We performed three experiments to evaluate the performance of Inception-v3 for binary change detection after fine-tuning with each of the four proposed input image representations. We used TensorFlow [82] for fine-tuning Inception-v3 and Keras<sup>4</sup> for the CAE. We describe the details of training, including number of training steps, and hyperparameter settings, in Appendix B. Area under the curve (AUC) for receiver operating characteristics (ROC) curves gives a better indication of model performance than accuracy, since accuracy depends on a chosen threshold on the posterior probability. Because we want to detect as many patches that contain true surface changes as possible, even at the expense of more false positives, a low false negative rate (FNR) is more important than a low false positive rate (FPR), or even an overall high classification accuracy or AUC. For this reason, we also reported the FPR at 5% FNR to assess which approach results in the fewest false positives given a 5% maximum FNR.

#### A. Experiment 1: Generalization From Train to Test Set

In this experiment, we evaluated the ability of each of four input representations to Inception-v3 to generalize from detecting changes in one type of surface feature in a training dataset to detecting changes in that same feature in a held-out test set. We used the labeled image pairs from the HiRISE RSL dataset described in Section IV. We augmented the positive-labeled (*change*) image pairs from the training and validation sets with horizontal flips, vertical flips, and  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  rotations, resulting in 1218 positive training examples and 438 positive validation examples. We used the labeled bitemporal image pairs ESP\_027802\_1685 and ESP\_028501\_1685 for training, ESP\_030347\_1685 and ESP\_030769\_1685 for validation, and ESP\_029213\_1685 and ESP\_029780\_1685 for testing (see Table II). We chose to partition the training, validation, and

<sup>2</sup>We selected images using the PlanetScope Gallery tool. [Online]. Available: <https://planet.com/gallery>

<sup>3</sup>We cropped two patch pairs with and without change from the Erta Ale, Ethiopia image pair; hence, we refer to these as “Erta Ale 1” and “Erta Ale 2.” [Online]. Available: <https://doi.org/10.5281/zenodo.2373798>

<sup>4</sup>F. Chollet, “Keras,” 2015. [Online]. Available: <https://keras.io>

TABLE IV  
PERFORMANCE METRICS FOR EXPERIMENT 1: TRAIN AND TEST ON HiRISE RSL (MARS) DATASET

Classification Method	Representation	FPR at 5% FNR	Accuracy	AUC
<b>Inception-v3</b>	<b>Absolute Difference</b>	<b>7.8%</b>	<b>94.5%</b>	<b>0.984</b>
Inception-v3	Bottleneck Representation	93.3%	70.1%	0.464
Inception-v3	Composite Grayscale	40.8%	89.8%	0.898
Inception-v3	Signed Difference	10.6%	91.7%	0.974
SVM	Absolute Difference	40.8%	87.8%	0.941
SVM	Absolute Difference - No LCN	97.2%	75.6%	0.687
<b>Naive Bayes</b>	<b>Pixel Difference <math>&gt; 2\sigma</math></b>	<b>2.2%</b>	<b>94.5%</b>	<b>0.991</b>
Naive Bayes	Pixel Difference $> 2\sigma$ - No LCN	29.6%	92.9%	0.958

Shaded rows and bold values indicate methods with best performance.

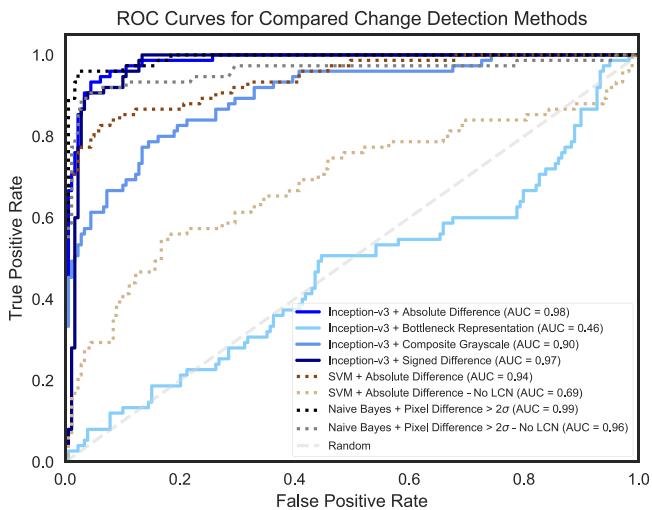


Fig. 7. ROC curves for Experiment 1: train and test on HiRISE RSL (Mars) dataset.

test sets by image pair rather than a random sample from all available images for two reasons. First, the RSL-forming region is located on the southern wall of Garni crater in the training pair but on the northeastern wall of Garni crater in the validation and test pairs. By sampling the training set and validation/test sets from different geographic locations within Garni crater, we avoid overestimating the classifier’s performance with spatial overlap between the training and validation/test sets. Second, separating the datasets by acquisition date represents how our change detection approach would be used in practice during a science mission. Table IV lists performance metrics and Fig. 7 shows the ROC curves for each representation.

### B. Experiment 2: Generalization to New Surface Feature Type, Instrument, and Misregistration

In this experiment, we evaluated the ability of each representation to generalize to new geographic locations on Mars (outside of Garni crater), a new type of surface feature change (meteorite impacts), a new instrument (CTX), and realistic misregistration compared to the training dataset. We used the same models that were trained for Experiment 1. We used the CTX Meteorite

impacts dataset of 96 images (48 *change*, 48 *no-change*) as the test set. Table V gives performance metrics and Fig. 8 shows the ROC curves for each classification approach in this experiment.

### C. Experiment 3: Generalization to New Body

In this experiment, we evaluated how well each representation enabled generalization to a new planetary body. Specifically, we wanted to test how general the representations learned by the autoencoder were for change detection. We used the same training set as in Experiments 1 and 2 of HiRISE RSL image pairs. We tested images of surface feature changes on the moon (LROC dataset) and on the earth (PlanetScope dataset) as described in Section IV. For each dataset, we evaluated five image pairs *with* surface changes and five image pairs without. In the LROC dataset, surface changes are the result of meteorite impacts and a spacecraft landing. These images are map-projected but not coregistered, so matching features are misregistered by as many as 40 pixels. In the PlanetScope dataset, surface changes are the result of natural geologic processes (e.g., lava flows) as well as some anthropogenic processes (e.g., solar array construction). These image pairs are precisely coregistered and have very similar lighting conditions between before and after images. Figs. 9 and 10 show the images we evaluated in this experiment for the LROC and PlanetScope datasets, respectively. The third row of each section in these figures shows the difference between the autoencoder bottleneck representation of the before image and the after image for each tested pair. The outline of each image indicates if it was correctly classified (green) or misclassified (red).

## VI. DISCUSSION

### A. Summary of Findings

We found in Experiment 1 that the Inception-v3 network fine-tuned using absolute difference image representations of HiRISE RSL images and the Naive Bayes baseline method exhibited the best change detection performance on the test set of HiRISE RSL images from a spatially distinct region of Garni crater. The Naive Bayes baseline method achieved slightly higher AUC and lower FPR at 5% FNR values than Inception-v3 with absolute difference representations, and these scores were

TABLE V  
PERFORMANCE METRICS FOR EXPERIMENT 2: TRAIN ON HiRISE RSL (MARS) DATASET, TEST ON CTX METEORITE IMPACTS (MARS) DATASET

Classification Method	Representation	FPR at 5% FNR	Accuracy	AUC
Inception-v3	Absolute Difference	50.0%	71.9%	0.807
<b>Inception-v3</b>	<b>Bottleneck Representation</b>	<b>14.6%</b>	<b>75.0%</b>	<b>0.942</b>
Inception-v3	Composite Grayscale	97.9%	55.2%	0.545
Inception-v3	Signed Difference	85.4%	54.2%	0.691
SVM	Absolute Difference	39.6%	67.7%	0.853
SVM	Absolute Difference - No LCN	54.2%	54.2%	0.720
<b>Naive Bayes</b>	<b>Pixel Difference <math>&gt; 2\sigma</math></b>	<b>33.3%</b>	<b>52.08%</b>	<b>0.952</b>
Naive Bayes	Pixel Difference $> 2\sigma$ - No LCN	50.0%	55.21%	0.931

Shaded rows and bold values indicate methods with best performance.

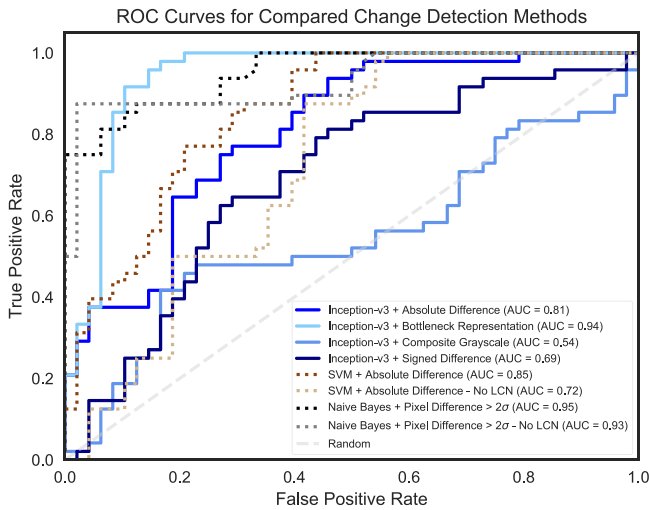


Fig. 8. ROC curves for Experiment 2: train on HiRISE RSL (Mars) dataset, test on CTX Meteorite impacts (Mars) dataset.

closely followed by fine-tuned Inception-v3 using signed difference image representations (see Fig. 7 and Table IV). However, when local contrast normalization (LCN in Fig. 7 and Table IV) is *not* applied prior to processing with Naive Bayes, our methods outperformed this baseline by a significant margin. Thus, our deep learning methods could be used successfully without the need for preprocessing in change detection pipelines.

The purpose of Experiment 2 was to test how sensitive the input image representations and the features learned by each classification method were to the type of surface feature undergoing change, the sensor that collected the images, and realistic misregistration. Using a test set where these properties were different from the training set, we found that the autoencoder bottleneck representations gave the best change detection performance overall in this scenario. While the Naive Bayes baseline method achieved a slightly higher AUC score (by 0.01), the bottleneck approach yielded the lowest FPR at 5% FNR (14.6%) as well as the highest accuracy of the tested approaches (see Fig. 8 and Table V). This was a surprising outcome given that the bottleneck representations resulted in the worst performance in Experiment 1. We discuss and explore potential explanations for this in Section VI-B.

In Experiment 3, we evaluated how general the latent representations learned by the autoencoder were for different types of features, sensors, scales, and planetary bodies. The surface features in both the LROC impacts and PlanetScope earth datasets were different than those in training, especially for the earth examples, as were the imaging sensor and level of misregistration. We found that the Inception-v3 network fine-tuned with differenced latent representations of HiRISE RSL image pairs correctly classified all PlanetScope earth images and three out of five LROC impacts images tested (see Figs. 9 and 10).

### B. Autoencoder Bottleneck Representations

It is interesting that the bottleneck representations yielded the lowest performance when training and testing on similar examples (see Experiment 1), but the highest performance when the train and test examples were significantly different (see Experiments 2 and 3). Surface changes in the CTX impacts, LROC impacts, and PlanetScope earth datasets mostly follow a pattern in which the feature is completely absent from one image in the pair and then appears in the other image of the pair (see Figs. 2, 9, and 10). In the HiRISE RSL dataset, changes tend to be more gradual and often manifest as *growing* or *receding* rather than appearing on a blank slate (see Figs. 1 and 2). Thus, one hypothesis is that the autoencoder is learning similar representations for both the before and after images in these cases such that the *differenced* bottleneck maps are not as suitable for change detection as they are in the other datasets. In Fig. 11, we show six images from the *change* class in the test set in Experiment 1 that our Inception-v3 with bottleneck representations approach assigned the *highest* probability of *change* (left) and the *lowest* probability of *change* (right). In Fig. 12, we show six images from the *no-change* class that our approach correctly (left) and incorrectly (right) classified. While this “gradual change” hypothesis could explain some of the misclassifications, it does not explain all or even most of them. In some of the misclassified *change* examples, the RSL change is on the edge and nearly out of the frame, whereas features in the other three datasets are (by design) located in the center of the frame. This would suggest that the autoencoder encoding function is less effective in representing features that are only partially in the frame.



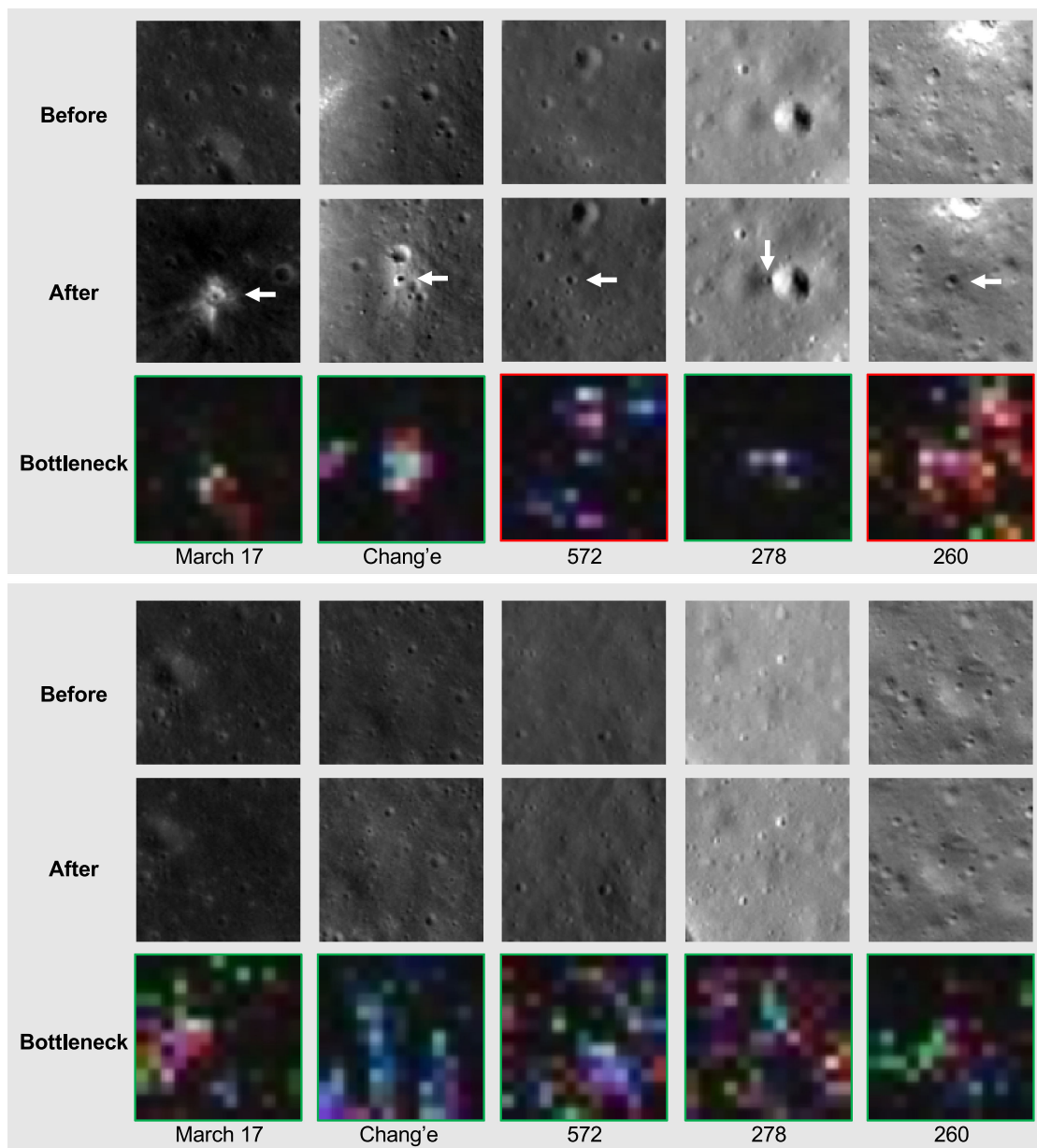


Fig. 9. *Change* (top section) and *no-change* (bottom section) image pairs from the LROC impacts (Moon) dataset used for Experiment 3. White arrows point to the new feature in the after images. The third row of each section shows the difference between bottleneck representations of before and after images. Correctly classified images are outlined in green and misclassified images in red.

### C. Generalization to Different Planetary Bodies

In Experiment 3, we tested how sensitive the change detection approaches were to the planetary body being studied with test images of the lunar surface and the earth's surface instead of the Martian surface seen during training. The lunar dataset also had the additional challenges from Experiment 2 in that the feature-type undergoing change (meteorite and spacecraft impacts), sensor-type (LROC), and level of misregistration (up to 40 pixels) were different than in the training images. The images in the earth dataset were precisely coregistered as in the training dataset, but the feature-type and sensor-type differed significantly. Comparing the bottleneck representations of the

misclassified LROC images (see Fig. 9, row 3) with the bottleneck representations of the other *change* image representations in Figs. 9 and 10, the misclassified representations do not appear as localized to the surface feature change. In the case of image 572, the representation seems to have picked up on the actual new impact crater as well as an impact crater that is present in the before image but not the after image due to misregistration. The representation for image 260 also appears to have several false detections in addition to the actual new impact crater, perhaps due to features that appear in one image but not in the other due to misregistration. This suggests a limit to the level of misregistration this approach can successfully tolerate.

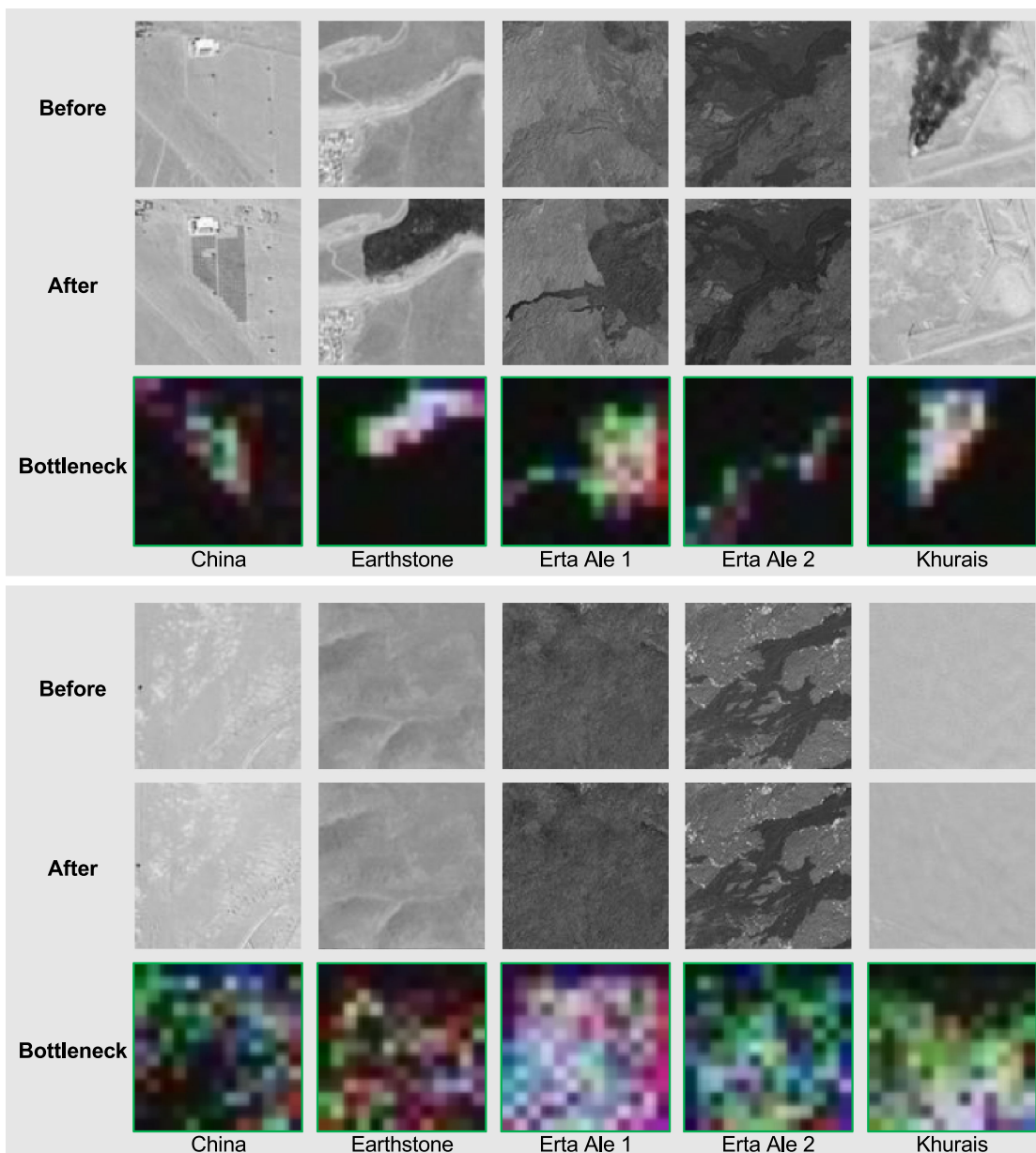


Fig. 10. *Change* (top section) and *no-change* (bottom section) image pairs from the PlanetScope (Earth) dataset used for Experiment 3. The third row of each section shows the difference between bottleneck representations of before and after images. The green outlines indicate that all images were correctly classified.

Comparing the bottleneck representations in Fig. 10 (row 3) to the corresponding before and after images in Fig. 10 (rows 1 and 2), we see that despite never having seen images of earth or the types of features in the PlanetScope dataset during training, the autoencoder is still able to encode useful representations of the surface features for change detection. This result further supports our hypothesis that the autoencoder bottleneck representations are the most general of those we studied and might be the most suitable for general purpose change detection for surface features on planetary bodies. These representations can also enable a more interpretable approach than the other compared methods. The representations in Fig. 10 suggest that the difference between the encodings of before and after images corresponds to the regions where features changed within the

image. In the future work, we will explore the use of these encodings to classify change at the region or pixel level given only patch-level labels.

#### D. Baseline Performance

We found that the Naive Bayes baseline method exhibited good performance for the HiRISE RSL, CTX impacts, and PlanetScope earth test examples, but not for the LROC impacts examples. In these three datasets, the surface feature exhibiting change occupies a large portion of the  $100 \times 100$  image (see Figs. 2 and 10). The scale of the impacts in the LROC impacts dataset is much smaller compared to the features in the other three datasets. This method depends directly on the number  $n$  of

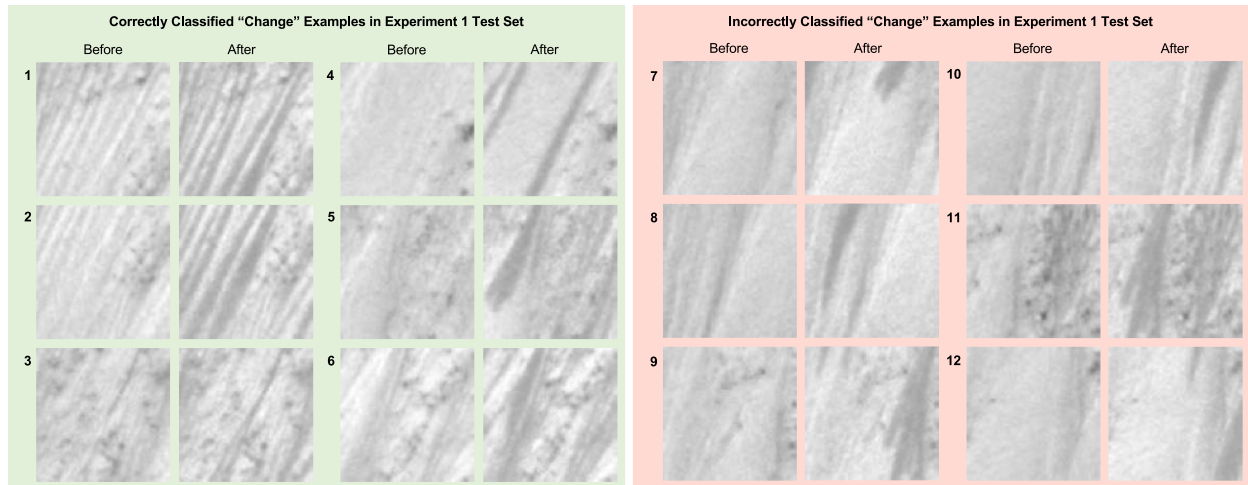


Fig. 11. Examples from Experiment 1 test set that the Inception-v3 with bottleneck representations method assigned the *highest* probability of *change* (left) and *lowest* probability of *change* (right). The true label for all examples shown is *change*; thus, examples on the right were misclassified.

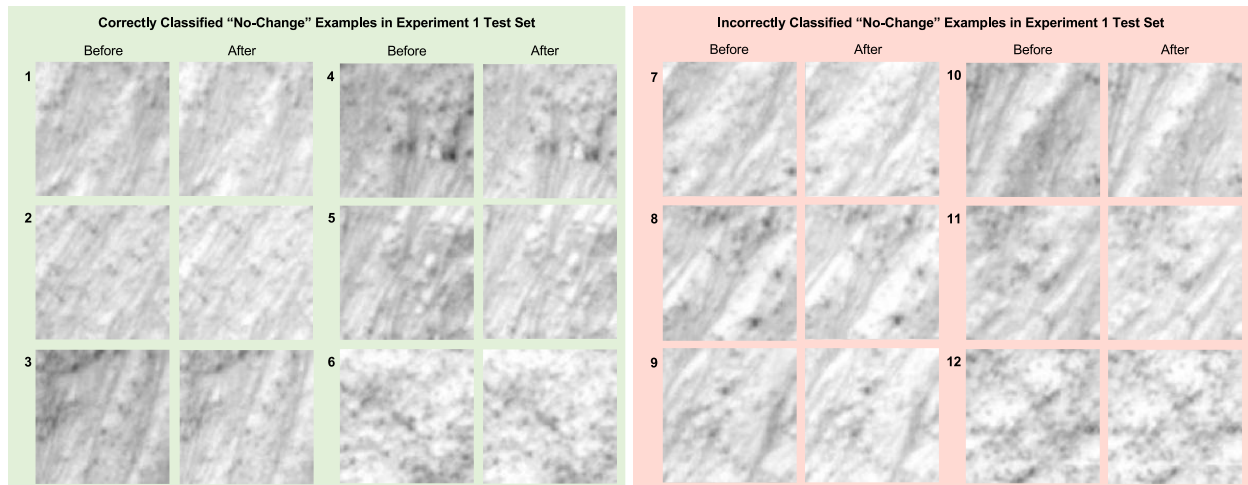


Fig. 12. Examples from Experiment 1 test set that the Inception-v3 with bottleneck representations method assigned the *lowest* probability of *change* (left) and *highest* probability of *change* (right). The true label for all examples shown is *no-change*; thus, examples on the right were misclassified.

differenced pixels in an image pair that are beyond two standard deviations from the mean difference in pixels. If  $n$  represented only the pixels where surface feature changes occurred, we would expect this number to be larger for changes in *larger* features (as in the HiRISE RSL, CTX impacts, and PlanetScope earth images) and smaller for changes in *smaller* features (as in the LROC impacts images). In Table VI, we show the mean and standard deviation of  $n$  as a percentage of the total pixels in each image ( $100 \times 100 = 10,000$ ) for all datasets used for training or testing. Given the distributions of  $n$  in the test datasets shown in this table, we would expect Naive Bayes to perform well for the LROC impacts dataset, since its distribution of  $n$  is similar to the training set. Furthermore, we would expect Naive Bayes to have lower performance on the CTX impacts dataset, since its distribution of  $n$  is the most different from the training set. Yet, this was not the case. This implies that pixels where a surface feature change occurred are not always represented by pixels with the highest difference in intensity, and that difference-based

TABLE VI  
MEAN AND STANDARD DEVIATION OF THE NUMBER OF DIFFERENCE PIXELS  $n > 2\sigma$  (NAIVE BAYES INPUT) AS A FRACTION OF THE TOTAL ( $100 \times 100$ ) PIXELS FOR EACH DATASET USED FOR TRAINING OR TESTING

Dataset	Mean/Std of $n$ (Percent of Total Pixels)
HiRISE RSL train	$0.028\% \pm 0.013$
HiRISE RSL test (Exp. 1)	$0.032\% \pm 0.019$
CTX Impacts test (Exp. 2)	$0.079\% \pm 0.020$
LROC Impacts test (Exp. 3)	$0.026\% \pm 0.007$
PlanetScope Earth test (Exp. 3)	$0.035\% \pm 0.019$

approaches to change detection will not reliably detect changes in such examples.

We found that the SVM baseline method exhibited good performance for all four datasets, but was not competitive with



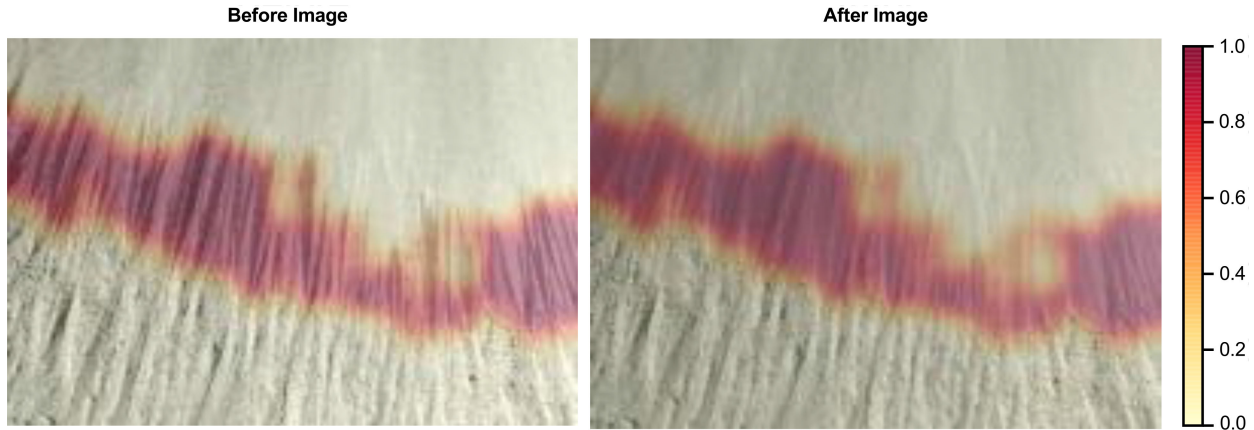


Fig. 13. Change map visualization, for example, HiRISE RSL image showing fading RSL, computed by convolving the Inception-v3 with absolute difference representations change classifier over image. Scale bar represents the average likelihood of change estimated in each pixel.

our Inception-v3 with absolute difference and bottleneck representation methods overall. We found that in all experiments, the performance was better when local contrast normalization was *not* applied for this method; this is in contrast to the Naive Bayes baseline, which depends on the *distribution* of difference pixels rather than the spatial pattern of difference pixels as in the SVM baseline. We also found worse performance for Inception-v3 when local contrast normalization was applied to the absolute difference image representation. This suggests that local contrast normalization might be more useful for methods such as the Naive Bayes baseline that depends on statistical properties rather than spatial patterns in the image.

## VII. CHANGE MAP VISUALIZATION

We trained our classifiers on  $100 \times 100$ -pixel patches sampled from full-resolution images. In practice, temporal image pairs that scientists wish to assess for surface feature change might be in the patch form, e.g., if they are intermediate products in a map patch server. Most often, image pairs will be two map-projected, coregistered images that are much larger than the patch size. In this case, it is useful to produce a change map across the image pairs using our change detection classifier. To produce change maps, we convolved the classifiers over the full-resolution image pair, stored the prediction made for each tile, and averaged the predictions that were computed over each pixel to produce a likelihood estimate for each pixel. The stride size controls the resolution of the change map since pixels will be visited more frequently (and, thus, more predictions will be averaged for each pixel) for smaller stride sizes. Fig. 13 shows the change map for a region in the southern wall of Garni crater from the HiRISE RSL dataset in which RSL that were present in the before image have faded away in the after image (before: ESP\_028501\_1685, after: ESP\_029213\_1685). We used the absolute difference representation and fine-tuned Inception-v3 classifier since this was the best performing approach for this dataset and a stride size of two pixels.

## VIII. CONCLUSION

We presented a new deep learning approach that leverages transfer learning and CAEs for patch-level change detection using a relatively small number of training examples. We created new, diverse datasets that include changes in surface features such as RSL, meteorite impact craters, and human-made structures on Mars, the moon, and earth to evaluate this approach for remote sensing images of multiple planetary bodies. Our experiments showed that our change detection methods outperformed the difference-based baseline methods with equal preprocessing for the HiRISE RSL dataset, and regardless of preprocessing on all other datasets. These experiments revealed a key insight that changed surface features may not always be represented by pixels with the highest difference in intensity between before and after images, which limits their generalization ability. We showed that latent (“bottleneck”) representations learned by a CAE provide the most general representation for surface feature change detection in our article, and that Inception-v3 fine-tuned with bottleneck representations could detect surface feature changes even when the feature type, imaging sensor, level of misregistration, feature scale, and planetary body are different than in the training dataset.

*Future Work:* In this article, our experimental approach was to isolate our change detection datasets by feature type, instrument, planet, and other properties in order to reveal the strengths and weaknesses of each input representation. The success of our change detection methods, in particular the Inception-v3 network using autoencoder bottleneck representations as input, on examples that deviated significantly from the training set suggests that variants of these methods hold promise for general-purpose change detection for surface features on planetary bodies that share similar overall characteristics. We plan to explore this hypothesis in future work. In this article, our goal was to predict patch-level labels of *change* or *no-change*. In future work, we will investigate how to leverage bottleneck representations or class activation maps to predict pixel-level or region-level labels of *change* or *no-change* given only patch-level labels.

TABLE VII  
IMAGE PRODUCTS USED FOR TEST DATASETS

Instrument	Image ID	Date Acquired
CTX	P02_001790_1871_XN_07N182W	12/13/2006
CTX	P03_002169_1937_XI_13N091W	01/12/2007
CTX	P06_003451_2035_XN_23N171W	04/22/2007
CTX	P09_004477_1906_XN_10N100W	07/11/2007
CTX	P13_005954_1927_XI_12N105W	11/03/2007
CTX	P13_006178_1907_XN_10N100W	11/20/2007
CTX	P13_006286_2073_XN_27N171W	11/29/2007
CTX	P14_006560_1936_XN_13N091W	12/20/2007
CTX	B01_009923_1790_XN_01S113W	09/07/2008
CTX	B01_010213_1790_XN_01S113W	09/30/2008
CTX	B02_010424_1849_XI_04N113W	10/16/2008
CTX	B06_012006_1912_XI_11N104W	02/17/2009
CTX	B06_012022_1845_XI_04N182W	02/18/2009
CTX	B09_012981_1844_XI_04N084W	05/03/2009
CTX	G14_023886_1819_XN_01N083W	08/31/2011
CTX	G17_024942_1813_XI_01N112W	11/22/2011
LROC	M104318871L (572)	08/07/2009
LROC	M142531289R (260)	10/24/2010
LROC	M161489808R (278)	05/31/2011
LROC	M176811566R (278)	11/24/2011
LROC	M181330922L (March 17)	01/16/2012
LROC	M1111614190R (572)	12/31/2012
LROC	M1119134763 (260)	03/28/2013
LROC	M1129602407L (Chang'e)	07/27/2013
LROC	M1139065512L (March 17)	11/14/2013
LROC	M1144922100L (Chang'e)	01/21/2014
PlanetScope	china-solar-20151217-before	12/17/2015
PlanetScope	khurais-20151218-after	12/18/2015
PlanetScope	khurais-20160212-before	02/12/2016
PlanetScope	china-solar-20160227-after	02/27/2016
PlanetScope	ertaale-20170116-before	01/16/2017
PlanetScope	ertaale-20170123-after	01/23/2017
PlanetScope	earthstone-20170702-before	07/02/2017
PlanetScope	earthstone-20170704-after	07/04/2017

#### APPENDIX A ADDITIONAL DATASET DETAILS

In Table II, we provided identifiers of images used for the HiRISE RSL Dataset and indicated how each image was used in our experiments. Three more datasets—CTX Meteorite impacts, LROC impacts, and PlanetScope earth—were used for testing generalization of the change detection classifiers in Section V. Table VII gives the identifiers for the images used in these datasets. We provide instructions for accessing these images in the following sections.

#### A. HiRISE RSL Dataset

The HiRISE images we used for this article can be accessed<sup>5</sup> <Image ID>, where <Image ID> is the Image ID from Table II. We used the JP2 black and white (red channel) map-projected products. To crop the images to the 10 000 × 10 000-pixel region of Garni crater, we used the following ImageMagick command: `convert image.jp2 -crop 10000x10000+3300+12000 cropped_image.jp2`.

#### B. CTX Meteorite Impacts Dataset

The CTX images we used for this dataset can be found<sup>6</sup> using the search tool on the planetary data system (PDS) imaging node. On the filter menu, select “Mars Reconnaissance Orbiter” under Mission, “ctx” under instrument, and type the Image ID from Table VII in the search bar.

#### C. LROC Impacts Dataset

The LROC images used for this article can be found on the Arizona State University School of Earth and Space Exploration’s LROC portal available online.<sup>7</sup>

#### D. PlanetScope Earth Dataset

The images used for this dataset are browse products from the Gallery on the Planet website. They can be accessed online.<sup>8</sup>

### APPENDIX B

#### DETAILS OF EXPERIMENTAL SETUP

To help reproduce the experiments in this article, we provide details for implementing and training all models in the following sections.

<sup>5</sup>[Online]. Available: <https://www.uahirise.org/>

<sup>6</sup>[Online]. Available: <https://pds-imaging.jpl.nasa.gov/search>

<sup>7</sup>[Online]. Available: [http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/17%20March%202013%20Event/feature\\_highlights/481](http://lroc.sese.asu.edu/featured_sites/lroc_features/17%20March%202013%20Event/feature_highlights/481)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/17%20March%202013%20Event/feature\\_highlights/490](http://lroc.sese.asu.edu/featured_sites/lroc_features/17%20March%202013%20Event/feature_highlights/490)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/New%20Crater%20572/feature\\_highlights/504](http://lroc.sese.asu.edu/featured_sites/lroc_features/New%20Crater%20572/feature_highlights/504)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/New%20Crater%20572/feature\\_highlights/509](http://lroc.sese.asu.edu/featured_sites/lroc_features/New%20Crater%20572/feature_highlights/509)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/New%20Crater%20260/feature\\_highlights/491](http://lroc.sese.asu.edu/featured_sites/lroc_features/New%20Crater%20260/feature_highlights/491)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/New%20Crater%20260/feature\\_highlights/492](http://lroc.sese.asu.edu/featured_sites/lroc_features/New%20Crater%20260/feature_highlights/492)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/New%20Crater%20278/feature\\_highlights/496](http://lroc.sese.asu.edu/featured_sites/lroc_features/New%20Crater%20278/feature_highlights/496)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/New%20Crater%20278/feature\\_highlights/497](http://lroc.sese.asu.edu/featured_sites/lroc_features/New%20Crater%20278/feature_highlights/497)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/Chang'e%203%20Landing%20Site/feature\\_highlights/553](http://lroc.sese.asu.edu/featured_sites/lroc_features/Chang'e%203%20Landing%20Site/feature_highlights/553)  
[http://lroc.sese.asu.edu/featured\\_sites/lroc\\_features/Chang'e%203%20Landing%20Site/feature\\_highlights/558](http://lroc.sese.asu.edu/featured_sites/lroc_features/Chang'e%203%20Landing%20Site/feature_highlights/558)

<sup>8</sup>[Online]. Available: <https://www.planet.com/gallery/china-solar-20170130/>

<https://www.planet.com/gallery/khurais/>

<https://www.planet.com/gallery/ertaale-20170123/>

<https://www.planet.com/gallery/earthstone-fire-20170705/>

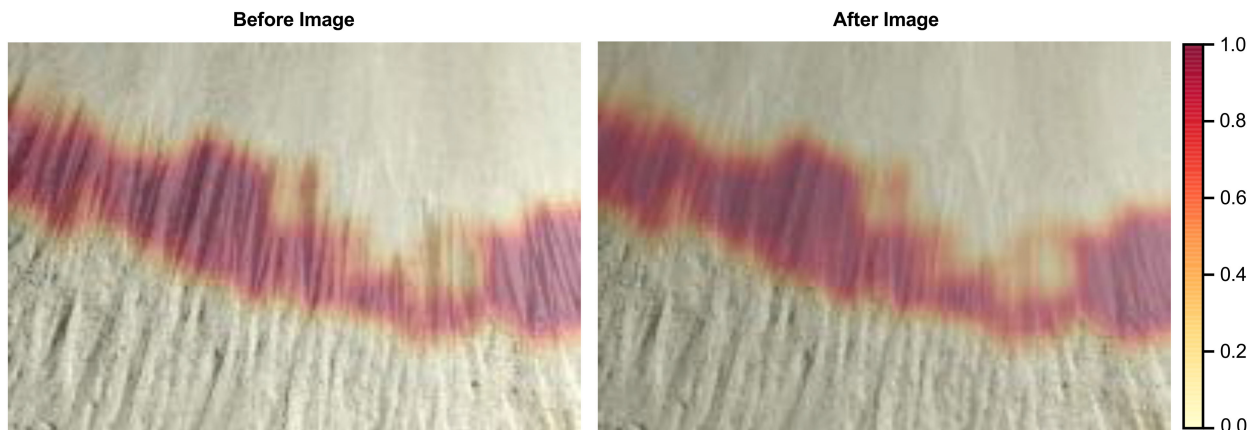


Fig. 14. Change map visualization, for example, HiRISE RSL image pair showing fading RSL, computed by convolving the Inception-v3 with signed difference representations change classifier over the image pair. Scale bar represents the average likelihood of change estimated in each pixel.

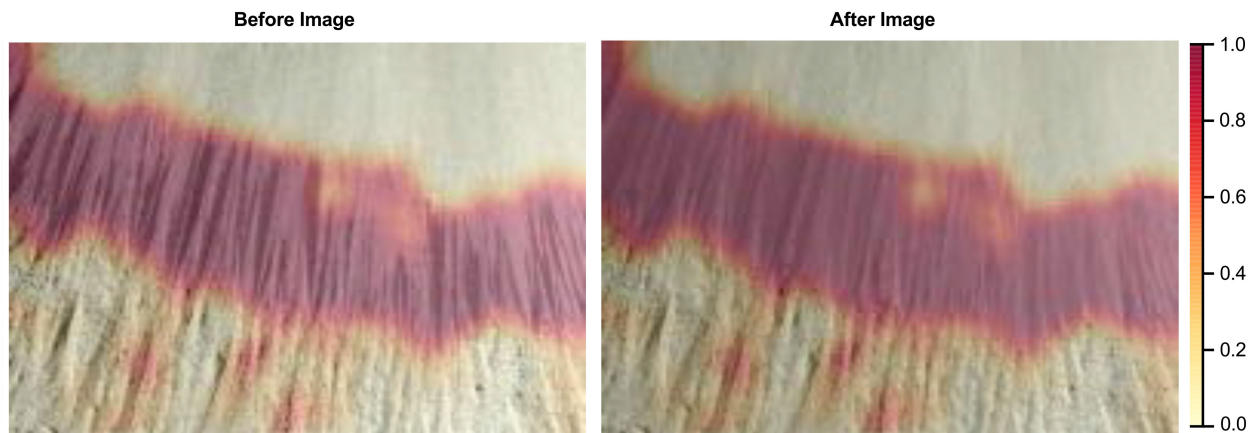


Fig. 15. Change map visualization, for example, HiRISE RSL image pair showing fading RSL, computed by convolving the Inception-v3 with composite grayscale representations change classifier over the image pair. Scale bar represents the average likelihood of change estimated in each pixel.

#### A. Fine-Tuned Inception-v3

We implemented the Inception-v3 approaches using TensorFlow. Details of how to fine-tune Inception-v3 for new categories can be found on the TensorFlow website.<sup>9</sup> We used stochastic gradient descent optimization and the sparse softmax cross-entropy loss function. We used a training batch size of 100 and learning rate of 0.001. We trained a different model for each image representation until validation loss was minimized, which resulted in different training times for each representation. We fine-tuned the absolute difference model for 2220 steps ( $\sim 48$  epochs), the autoencoder bottleneck model for 3160 steps ( $\sim 69$  epochs), the composite grayscale model for 790 steps ( $\sim 17$  epochs), and the signed difference model for 3700 steps ( $\sim 81$  epochs).

#### B. Convolutional Autoencoder

We implemented the CAE using Keras. The size (and, thus, number of feature maps) is given in Table I. We used  $3 \times 3$ -pixel kernels for convolution and  $2 \times 2$ -pixel kernels for max

pooling, both using a stride size of 1 pixel. We used the Adam optimizer with  $\beta_1$  0.9,  $\beta_2$  0.999,  $\epsilon$   $1e-7$ , and decay 0.0. We used a batch size of 100 and learning rate of 0.001. We used the binary cross-entropy loss function. We trained the autoencoder for 50 epochs.

#### C. Baseline Methods

We implemented both the Naive Bayes and SVM baseline methods using Scikit-learn<sup>10</sup> in python. The hyperparameter settings determined using grid search with threefold cross validation for the SVM baseline were  $C = 10$ ,  $\gamma = 0.01$  with local contrast normalization and  $C = 100$ ,  $\gamma = 0.1$  without local contrast normalization.

### APPENDIX C ADDITIONAL CHANGE MAPS

In Section VII, we showed an example change map for a region of Garni crater computed by convolving the Inception-v3 model fine-tuned with absolute difference image representations

<sup>9</sup>[Online]. Available: [https://www.tensorflow.org/hub/tutorials/image\\_retraining](https://www.tensorflow.org/hub/tutorials/image_retraining)

<sup>10</sup>[Online]. Available: <https://scikit-learn.org>



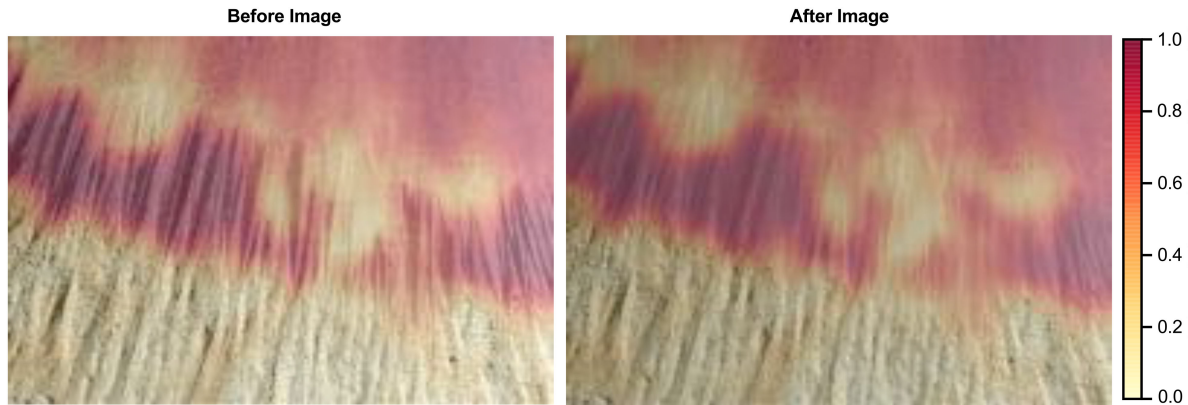


Fig. 16. Change map visualization, for example, HiRISE RSL image pair showing fading RSL, computed by convolving the Inception-v3 with autoencoder bottleneck representations change classifier over the image pair. Scale bar represents the average likelihood of change estimated in each pixel.

over the entire image with a stride size of 2 pixels (see Fig. 13). In Figs. 14–16, we show the change maps computed by convolving additional methods tested over the same image pair as Fig. 13.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. D. Stillman of the Southwest Research Institute (SwRI) and Dr. I. Daubar of the Jet Propulsion Laboratory for their expert knowledge on RSL and meteorite impacts, the PDS for supporting the development of this article, and Dr. G. Doran of the Jet Propulsion Laboratory for his assistance with processing CTX image pairs. This research was carried out (in part) at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

#### REFERENCES

- [1] A. S. McEwen *et al.*, “Mars Reconnaissance Orbiter’s High Resolution Imaging Science Experiment (HiRISE),” *J. Geophys. Res.*, vol. 112, no. E5, May 2007, Art. no. E05S02.
- [2] M. C. Malin *et al.*, “Context camera investigation on board the Mars Reconnaissance Orbiter,” *J. Geophys. Res.*, vol. 112, no. E5, May 2007, Art. no. E05S04.
- [3] D. E. Stillman, T. I. Michaels, R. E. Grimm, and K. P. Harrison, “New observations of Martian southern mid-latitude recurring slope lineae (RSL) imply formation by freshwater subsurface flows,” *Icarus*, vol. 233, pp. 328–341, May 2014.
- [4] I. Daubar, A. McEwen, S. Byrne, M. Kennedy, and B. Ivanov, “The current Martian cratering rate,” *Icarus*, vol. 225, no. 1, pp. 506–516, Jul. 2013.
- [5] E. J. Speyerer, R. Z. Povilaitis, M. S. Robinson, P. C. Thomas, and R. V. Wagner, “Quantifying crater production and regolith overturn on the moon with temporal imaging,” *Nature*, vol. 538, no. 7624, pp. 215–218, Oct. 2016.
- [6] A. Singh, “Review article: Digital change detection techniques using remotely-sensed data,” *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, Jun. 1989.
- [7] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, “Review article digital change detection methods in ecosystem monitoring: A review,” *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, May 2004.
- [8] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, “Change detection from remotely sensed images: From pixel-based to object-based approaches,” *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, Jun. 2013.
- [9] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, “A critical synthesis of remotely sensed optical image change detection techniques,” *Remote Sens. Environ.*, vol. 160, pp. 1–14, Apr. 2015.
- [10] T. Hame, I. Heiler, and J. San Miguel-Ayanz, “An unsupervised change detection and recognition system for forestry,” *Int. J. Remote Sens.*, vol. 19, no. 6, pp. 1079–1099, Jan. 1998.
- [11] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, “PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data,” *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008.
- [12] M. A. Torres-Vera, R. M. Prol-Ledesma, and D. Garcia-Lopez, “Three decades of land use variations in Mexico City,” *Int. J. Remote Sens.*, vol. 30, no. 1, pp. 117–138, Jan. 2009.
- [13] O. Abd El-Kawy, J. Rød, H. Ismail, and A. Suliman, “Land use and land cover change detection in the western Nile delta of Egypt using remote sensing data,” *Appl. Geogr.*, vol. 31, no. 2, pp. 483–494, Apr. 2011.
- [14] R. Peiman, “Pre-classification and post-classification change-detection techniques to monitor land-cover and land-use change using multi-temporal landsat imagery: A case study on Pisa province in Italy,” *Int. J. Remote Sens.*, vol. 32, no. 15, pp. 4365–4381, Aug. 2011.
- [15] A. A. Alesheikh, A. Ghorbanali, and N. Nouri, “Coastline change detection using remote sensing,” *Int. J. Environmental Sci. Technol.*, vol. 4, no. 1, pp. 61–66, Dec. 2007.
- [16] X. Dai and S. Khorram, “The effects of image misregistration on the accuracy of remotely sensed change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1566–1577, Sep. 1998.
- [17] G. Castilla and G. J. Hay, “Image objects and geographic objects,” in *Object-Based Image Analysis*. Berlin, Germany: Springer, 2008, pp. 91–110.
- [18] C. Munyati, “Wetland change detection on the Kafue flats, Zambia, by classification of a multitemporal remote sensing image dataset,” *Int. J. Remote Sens.*, vol. 21, no. 9, pp. 1787–1806, Jan. 2000.
- [19] J. Nichol and M. S. Wong, “Satellite remote sensing for detailed landslide inventories using change detection and image fusion,” *Int. J. Remote Sens.*, vol. 26, no. 9, pp. 1913–1926, May 2005.
- [20] F. Yuan, K. E. Sawaya, B. C. Loeffelholz, and M. E. Bauer, “Land cover classification and change analysis of the twin cities (Minnesota) metropolitan area by multitemporal landsat remote sensing,” *Remote Sens. Environ.*, vol. 98, no. 2-3, pp. 317–328, Oct. 2005.
- [21] A. Shalaby and R. Tateishi, “Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the northwestern coastal zone of Egypt,” *Appl. Geogr.*, vol. 27, no. 1, pp. 28–41, Jan. 2007.
- [22] P. Gray *et al.*, “Integrating drone imagery into high resolution satellite remote sensing assessments of estuarine environments,” *Remote Sens.*, vol. 10, no. 8, Aug. 2018, Art. no. 1257.
- [23] V. Walter, “Object-based classification of remote sensing data for change detection,” *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3-4, pp. 225–238, Jan. 2004.
- [24] M. N. Klaric *et al.*, “GeoCDX: An automated change detection and exploitation system for high-resolution satellite imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2067–2086, Apr. 2013.
- [25] L. Bruzzone and D. Prieto, “An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images,” *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 452–466, Apr. 2002.
- [26] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Rojo-Alvarez, and M. Martinez-Ramon, “Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.

- [27] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. Hoboken, NJ, USA: Wiley, 2009.
- [28] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 20, pp. 77–85, Feb. 2013.
- [29] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
- [30] J. Im and J. R. Jensen, "A change detection model based on neighborhood correlation image analysis and decision tree classification," *Remote Sens. Environ.*, vol. 99, no. 3, pp. 326–340, Nov. 2005.
- [31] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
- [32] T. Heyer *et al.*, "The multi-temporal database of planetary image data (MUTED): A web-based tool for studying dynamic Mars," *Planet. Space Sci.*, vol. 159, pp. 56–65, Sep. 2018.
- [33] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [34] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 125–138, Jan. 2016.
- [35] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [36] G. Cao, B. Wang, H.-C. Xavier, D. Yang, and J. Southworth, "A new difference image creation method based on deep neural networks for change detection in remote-sensing images," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 7161–7175, Dec. 2017.
- [37] H. Lyu and H. Lu, "A deep information based transfer learning method to detect annual urban dynamics of Beijing and New York from 1984–2016," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 1958–1961.
- [38] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 1, no. 57, pp. 3–13, Jan. 2019.
- [39] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, Jun. 2016.
- [40] H. Lyu, H. Lu, and L. Mou, "Learning a transferable change rule from a recurrent neural network for land cover change detection," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 506.
- [41] M. Gong, T. Zhan, P. Zhang, and Q. Miao, "Superpixel-based difference representation learning for change detection in multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2658–2673, May 2017.
- [42] P. Zhang, M. Gong, H. Zhang, J. Liu, and Y. Ban, "Unsupervised difference representation learning for detecting multiple types of changes in multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2277–2289, Apr. 2019.
- [43] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.
- [44] Y. Xu, S. Xiang, C. Huo, and C. Pan, "Change detection based on auto-encoder model for VHR images," in *Proc. SPIE Pattern Recognit. Comput. Vis.*, Oct. 2013, vol. 8919, p. 891902.
- [45] S. Lefevre, D. Tuia, J. D. Wegner, T. Produit, and A. S. Nassar, "Toward seamless multiview scene analysis from satellite to street level," *Proc. IEEE*, vol. 105, no. 10, pp. 1884–1899, Oct. 2017.
- [46] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.
- [47] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018.
- [48] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 539–546.
- [49] S. Hochreiter and J. J. Urgan Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [50] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 2, pp. 924–935, Feb. 2019.
- [51] A. Song *et al.*, "Change detection in hyperspectral images using recurrent 3D fully convolutional networks," *Remote Sens.*, vol. 10, no. 11, Nov. 2018, Art. no. 1827.
- [52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [53] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.
- [54] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
- [55] N. Audebert, B. Le Saux, and S. Lefevre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS J. Photogramm. Remote Sens.*, 2018, pp. 20–32.
- [56] M. Volpi and D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, Feb. 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/7725499/>
- [57] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 44–51.
- [58] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," Aug. 2015, *arXiv*.
- [59] F. Hu *et al.*, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14 680–14 707, Nov. 2015.
- [60] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
- [61] N. Liu, L. Wan, Y. Zhang, T. Zhou, H. Huo, and T. Fang, "Exploiting convolutional neural networks with deeply local description for remote sensing image classification," *IEEE Access*, vol. 6, pp. 11 215–11 228, 2018.
- [62] Y. Ge, S. Jiang, Q. Xu, C. Jiang, and F. Ye, "Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 17 489–17 515, Jul. 2018.
- [63] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
- [64] J. Donahue *et al.*, "Decaf: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.
- [65] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7426826/>
- [66] H. R. Kerner, D. F. Wellington, K. L. Wagstaff, J. F. Bell, and H. Ben Amor, "Novelty detection for multispectral images with application to planetary exploration," in *Proc. 31st Annu. Conf. Innovative Appl. Artif. Intell.*, 2019, pp. 9484–9491.
- [67] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.*, 2011, pp. 52–59.
- [68] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [69] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [70] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. San Mateo, CA, USA: Morgan Kaufmann, 1994, pp. 474–485.
- [71] H. Zhang, "The optimality of naive Bayes," in *Proc. Florida Artif. Intell. Res. Soc. Conf.*, 2004.
- [72] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

- [73] D. E. Stillman, "Unraveling the mysteries of recurring slope lineae," in *Dynamic Mars: Recent and Current Landscape Evolution of the Red Planet*, R. J. Soare, S. J. Conway, and S. M. Clifford, Eds. Amsterdam, The Netherlands: Elsevier, 2018, ch. 2, pp. 51–85.
- [74] M. Molinier, J. Laaksonen, and T. Hame, "Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 861–874, Apr. 2007.
- [75] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, Jun. 2007.
- [76] H. R. Kerner, J. Bell III, and H. Ben Amor, "Context-dependent image quality assessment of JPEG-compressed Mars science laboratory mastcam images using convolutional neural networks," *Comput. Geosci.*, vol. 118, pp. 109–121, 2018.
- [77] G. Doran, D. R. Thompson, and T. Estlin, "Precision instrument targeting via image registration for the Mars 2020 rover," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 3352–3358.
- [78] M. S. Robinson *et al.*, "Lunar Reconnaissance Orbiter camera (LROC) instrument overview," *Space Sci. Rev.*, vol. 150, no. 1–4, pp. 81–124, Jan. 2010.
- [79] W.-H. Ip, J. Yan, C.-L. Li, and Z.-Y. Ouyang, "Preface: The Chang'e-3 lander and rover mission to the moon," *Res. Astron. Astrophys.*, vol. 14, no. 12, pp. 1511–1513, Dec. 2014.
- [80] Planet Team, "Planet Application Program Interface: In Space for Life on Earth," 2018.
- [81] R. Houborg and M. F. McCabe, "High-resolution NDVI from Planet's constellation of earth observing nano-satellites: A new data source for precision agriculture," *Remote Sens.*, vol. 8, no. 9, 2016, Art. no. 768.
- [82] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Operating Syst. Des. Implementation*, 2016, pp. 265–283.



**Hannah Rae Kerner** received the B.S. degree in computer science from the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, in 2014 and the Ph.D. degree in exploration systems design from Arizona State University, Tempe, AZ, USA, in 2019.

She is an Assistant Research Professor with the University of Maryland, College Park, MD, USA. Her research interests include new machine learning methods for remote sensing applications.



**Kiri L. Wagstaff** received the B.S. degree in computer science from the University of Utah, Salt Lake City, UT, USA, in 1997, the M.S. and Ph.D. degrees in computer science from Cornell University, Ithaca, NY, USA, in 2000 and 2002, respectively, the M.S. degree in geological sciences from the University of Southern California, Los Angeles, CA, USA, in 2008, and the M.L.I.S. degree in library and information science from San Jose State University, San Jose, CA, USA, in 2017.

She is a Principal Research Technologist in artificial intelligence and machine learning with the Jet Propulsion Laboratory, Pasadena, CA, USA. Her research interests include new machine learning and data analysis methods for spacecraft.



**Brian D. Bue** received the B.S. degree in computer science and the B.A. degree in mathematics from Augsburg College, Minneapolis, MN, USA, in 2003; the M.S. degree in computer science from Purdue University, West Lafayette, IN, USA, in 2007; and the Ph.D. degree in electrical and computer engineering from Rice University, Houston, TX, USA, in 2013.

He is a Research Data Scientist with the Machine Learning and Instrument Autonomy Group, NASA Jet Propulsion Laboratory, Pasadena, CA, USA. His research interests include machine learning techniques for ground-based and onboard applications in planetary science, earth science, and astronomy.



**Patrick C. Gray** received the B.A. degree in computer science from the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, in 2014. He is currently working toward the Ph.D. degree from Duke University, Durham, NC, USA.

His doctoral research in the Duke Marine Robotics and Remote Sensing Lab focuses on machine learning approaches to remote sensing analysis, coordinating satellite, drone, and *in-situ* environmental monitoring, and understanding how autonomous systems will benefit field scientists—all with a focus on coastal and polar environments.



**James (Jim) F. Bell III** received the B.S. degree in geological and planetary sciences from the California Institute of Technology, Pasadena, CA, USA, in 1987, and the M.S. and Ph.D. degrees in geology and geophysics from the University of Hawaii, Honolulu, HI, USA, in 1989 and 1992, respectively.

He is a Professor of geology and planetary science with the School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA, an Adjunct Professor with the Department of Astronomy, Cornell University, Ithaca, NY, USA, and a Distinguished Visiting Scientist with the Jet Propulsion Laboratory, Pasadena, CA, USA. His research focuses on the geology, mineralogy, and composition of the surfaces of planets, moons, asteroids, and comets.



**Heni Ben Amor** received the Dipl.-Inf. degree in computer science from the University of Koblenz-Landau, Mainz, Germany, in 2005 and the Ph.D. degree in robotics and computer science from the Technical University Bergakademie Freiberg, Freiberg, Germany, in 2010.

He is an Assistant Professor for robotics with Arizona State University, Tempe, AZ, USA, where he directs the ASU Interactive Robotics Laboratory. Prior to joining ASU, he was a Research Scientist with Georgia Tech, a Postdoctoral Researcher with the Technical University Darmstadt (Germany), and a Visiting Research Scientist with the Intelligent Robotics Lab, University of Osaka (Japan). His primary research interests include artificial intelligence, machine learning, robotics, and human–robot interaction.

Dr. Amor was the recipient of the NSF CAREER Award in 2018, the Fulton Outstanding Assistant Professor Award in 2018, and the Daimler-and-Benz Fellowship in 2012.