Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Rapid broad area search and detection of Chinese surface-to-air missile sites using deep convolutional neural networks

Richard A. Marcum
Curt H. Davis
Grant J. Scott
Tyler W. Nivin

SPIE.

Richard A. Marcum, Curt H. Davis, Grant J. Scott, Tyler W. Nivin, "Rapid broad area search and detection of Chinese surface-to-air missile sites using deep convolutional neural networks," *J. Appl. Remote Sens.* **11**(4), 042614 (2017), doi: 10.1117/1.JRS.11.042614.

Rapid broad area search and detection of Chinese surface-to-air missile sites using deep convolutional neural networks

Richard A. Marcum, Curt H. Davis,* Grant J. Scott, and Tyler W. Nivin
University of Missouri, Center for Geospatial Intelligence, Columbia, Missouri, United States

Abstract. We evaluated how deep convolutional neural networks (DCNN) could assist in the labor-intensive process of human visual searches for objects of interest in high-resolution imagery over large areas of the Earth's surface. Various DCNN were trained and tested using fewer than 100 positive training examples (China only) from a worldwide surface-to-air-missile (SAM) site dataset. A ResNet-101 DCNN achieved a 98.2% average accuracy for the China SAM site data. The ResNet-101 DCNN was used to process ~19.6 M image chips over a large study area in southeastern China. DCNN chip detections (~9300) were postprocessed with a spatial clustering algorithm to produce a ranked list of ~2100 candidate SAM site locations. The combination of DCNN processing and spatial clustering effectively reduced the search area by ~660*X* (0.15% of the DCNN-processed land area). An efficient web interface was used to facilitate a rapid serial human review of the candidate SAM sites in the China study area. Four novice imagery analysts with no prior imagery analysis experience were able to complete a DCNN-assisted SAM site search in an average time of ~42 min. This search was ~81*X* faster than a traditional visual search over an equivalent land area of ~88,640 km^2 while achieving nearly identical statistical accuracy (~90% F1). [DOI: 10.1117/1.JRS.11.042614]

Keywords: deep learning; convolutional neural networks; object detection; target recognition.

Paper 170417SS received May 15, 2017; accepted for publication Sep. 6, 2017; published online Oct. 6, 2017.

1 Introduction

Deep learning (DL) methods have shown through extensive experimental validation to deliver excellent performance for a wide variety of remote sensing data processing and analysis tasks. These include image preprocessing, pixel-based classification, object/target detection and recognition, land cover classification, and scene understanding. As a result, DL methods have been applied to many high-resolution remote sensing data types, including synthetic aperture radar (SAR), light detection and ranging, and electro-optical (EO) imaging sensors (multispectral, hyperspectral, etc.).[1] This array of sensing modalities combined with numerous satellite and airborne remote sensing platforms has generated a tremendous volume of high-resolution, remote sensing data that can be used for automated object/target recognition research.

A significant amount of automated target recognition (ATR) research has been performed using high-resolution airborne SAR imagery starting in the 1990s with the MSTAR dataset.[2] The MSTAR dataset consists of 15 small target classes like tanks and armored personnel carriers. In one recent example, Vasuki and Roomi[3] achieved a 96% accuracy for the MSTAR dataset using a three-stage framework consisting of image preprocessing, Gabor-wavelet feature extraction, and a multilayer perceptron for final target classification. Anglberger and Kempf[4] utilized a combination of signature simulation and change detection on TerraSAR-X satellite imagery for airplane detections that were then subjected to human review for confirmation.

*Address all correspondence to: Curt H. Davis, E-mail: DavisCH@missouri.edu

For more information about ATR systems, we recommend recent survey articles by Schacter[5] and Darymli et al.[6]

Historically, automated object/target recognition research has utilized hand-crafted feature extraction methods that are then used as input for a final classification stage. However, selection of the feature extraction method or methods that are best suited for a given application is not usually obvious. In addition, the heterogeneity of manmade objects and their surrounding environments in high-resolution remote sensing imagery adds additional complexity that limit the performance of classification methods that use hand-crafted features in real-world applications. The rise in popularity of convolutional neural networks (CNN), and more specifically deep convolutional neural networks (DCNN), can be partially attributed to the ability of DCNN to eliminate altogether the need for hand-crafted features. Given a sufficient amount of labeled training data, a DCNN can learn optimal image processing filters, and ultimately extracted features, that can then be utilized in the final classification stage for a given application domain.

Chen et al.[7] created a custom DCNN to extract variable-scale features for vehicle detection and demonstrated an increase in vehicle detection rates over standard DCNN models. Recently, Zhang et al.[8] used a DCNN for airport detection in high-resolution EO satellite imagery and obtained a classification accuracy of 84%. Regions with long straight parallel lines were detected as possible airport runways and then a DCNN was used to determine if the detected regions were in fact airports. Also, Cao et al.[9] used a region-based CNN to identify the number of aircraft in high-resolution EO satellite imagery and achieved a recall rate of 79.6%.

The UC-merced (UCM) land use dataset[10] is a benchmark high-resolution EO image dataset that has been widely studied using a variety of DL methods. The UCM dataset has 21 different labeled classes each with 100 examples per class. Luus et al.[11] created a custom DCNN and achieved a classification accuracy of 93.5% for the UCM dataset. Cheng et al.[12] used GoogLeNet[13] as a feature extractor and a set of one-versus-all SVM for classification and obtained an overall accuracy of 98.3%. Scott et al.[14] introduced enhanced data augmentation techniques designed specifically for remote sensing imagery and combined this with transfer learning from ImageNet[15] to the ResNet-50 DCNN architecture and obtained a 98.5% classification accuracy on the UCM dataset. Scott et al.[16] then utilized an information fusion framework with the Sugeno fuzzy integral to fuse three different DCNN outputs to obtain a classification accuracy of 99.3% on the UCM dataset.

While the UCM dataset is widely characterized as a land use/land cover dataset, the 21 classes are a mixture of traditional land cover classes like forest, beach, chaparral, etc., along with smaller-scale object classes like airplane, storage tank, baseball diamond, tennis court, etc. The average DCNN classification accuracy for the UCM object classes is presently 99% (96% minimum).[14] This, therefore, suggests that DCNN can perform automated object detection at a very high level, at least using standard statistical assessments of testing data that are withheld from DCNN training.

The primary goal in this effort is to evaluate how automated DCNN object detection can be applied to assist in the traditional and labor-intensive process of human visual searches for objects/targets of interest in high-resolution EO imagery over very large areas of the Earth's surface. We use the term "broad area search" (BAS) to describe this process and note that BAS has been a standard practice in airborne and satellite photo reconnaissance applications for many decades. Given the recent promising results of DCNN for object recognition in high-resolution EO remote sensing imagery, we believe that there is great potential for DCNN to be effectively applied in BAS applications.

The remainder of this paper is organized as follows. In Sec. 2, we describe an enterprise data curation framework that is used to generate a comprehensive worldwide dataset of surface-to-air missile (SAM) site image chips derived from high-resolution EO satellite imagery. In Sec. 3, we evaluate the statistical performance of four DCNN for automated SAM site detection using the curated worldwide SAM site dataset. In Sec. 4, we establish a baseline set of BAS results for SAM sites over a large study area in southeastern China using the traditional approach of human visual search. In Sec. 5, we develop and then apply a suitably trained DCNN for large-scale processing and detection of SAM sites in the China study area. The DCNN detection results

are then spatially postprocessed to produce ranked lists of candidate SAM site locations. The candidate SAM sites are then used to assist human analysts in performing a DCNN-assisted BAS. The results are then compared with the baseline visual BAS results to assess and quantify the utility of the DCNN-assisted BAS approach. Finally, in Sec. 6, we summarize our major findings and discuss avenues for future research.

2 Source Data Ingestion and Curation

2.1 *Data Ingestion and Curation Framework*

DL methods require large amounts of labeled training data to build robust image classifiers. The success of DL on datasets, such as ImageNet,[15] is made possible by the sheer volume of labeled training data developed and made available for computer vision and machine learning research. For example, the ImageNet dataset currently has one thousand class labels and over one million annotated ground photos. Conversely, comparable labeled datasets for training object/target recognition algorithms are not widely available for remote sensing imagery.

The lack of labeled remote-sensing image data is a problem for many object/target classes of interest, though we note that labeled image training sets have been created for airplanes, buildings, and a few other object classes. Nevertheless, within the last 10 years, there has been an explosion in the amount of open-source geospatial information that can be exploited to construct labeled remote sensing imagery datasets for DL applications. Toward that end, we have developed an efficient enterprise framework to ingest open-source geospatial vector datasets and combine those with high-resolution remote sensing imagery accessed from web services to produce curated sets of labeled image data for DL research. This framework consists of: (1) geospatial vector data ingestion, (2) image labeling and curation, and (3) quality assurance review (QAR). Figure 1 provides an overview of the enterprise framework, the flow of data, and the various processes involved.

The ingestion step processes geospatial vector data and then stores it into a PostGIS geospatial database. This step is designed to accommodate a variety of different geospatial file format parsers. As such, once the framework has been extended to ingest a given file type,

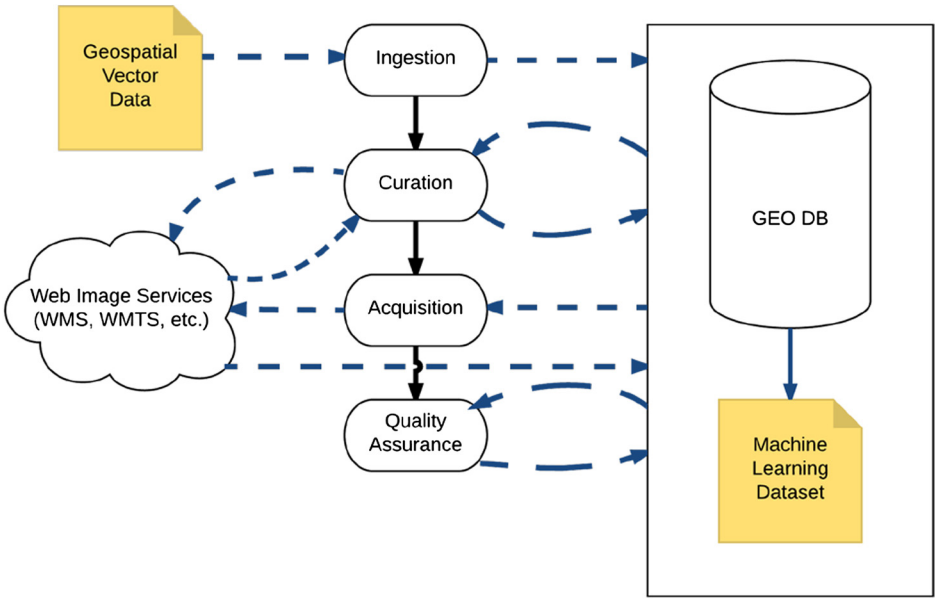


Fig. 1 Overview of enterprise framework for ingesting open-source geospatial vector datasets and combining those with high-resolution remote sensing imagery to produce curated sets of labeled imagery data for DL research.

little to no work needs to be done to ingest more geospatial data in the same file format. The framework currently supports both KML and CSV file formats and has ingested, to date, more than 80 open-source geospatial feature data (GFD) sources containing more than 5.7 million labeled feature points.

Data parsed from GFD sources are normalized during ingestion as they are loaded into the PostGIS geospatial database. First, all input coordinates are converted to a standard spatial reference system (WGS84—SRID 4326). In addition, various input geometries (line string, polygon, etc.) are converted to single point features. For example, line strings are decomposed to their constituent points and polygons are converted to a point at their center of mass. This is done for both data management and efficiency reasons. Once loaded into the database, point features can be selected for curation based on a number of metadescriptors, such as location, expected feature type, and data source. This allows subsequent feature data curation to be prioritized based on a wide range of metadata, including feature type, feature attributes, geography (country, state, etc.), and data source. This flexibility is important for a variety of reasons. For example, some GFD are more likely to be accurate and/or reliable than others for certain feature types. In addition, utilization of feature attributes, when present, can help target the curation work to the most desirable features and/or create stratified samples from very large and heterogeneous feature datasets. This, in turn, supports the creation of training datasets that adequately represent the variability present in a given feature class.

After GFD ingestion, point features are selected from the database and then curated in a custom web application that facilitates rapid human verification/correction of geographic location and semantic labels for all assigned point features. Curation begins when a high-resolution remote sensing image of the geographic area surrounding each feature point is autoloaded in the web browser along with a placemark overlay showing the geographic location of the feature of interest. Currently, the framework utilizes DigitalGlobe's Maps API Premium Imagery,[17] where access was provided for this effort through a partnership with the DigitalGlobe Foundation.[18] In addition, the framework can also access and utilize remote sensing imagery from any web mapping service via WMS and WMTS. DigitalGlobe's Maps API provides a high-resolution orthoimage basemap of the entire globe, which ensures on-demand, online access to quality imagery for each feature location. After loading the curation image, the web application allows the user to confirm or correct the geographic position of the feature of interest and also confirm or correct the feature's pre-assigned semantic label that was derived from the GFD source. The feature information is automatically updated in the geospatial database when any changes to geographic position or semantic label are made during curation.

After image curation, the system downloads high-resolution images for each feature utilizing various web image services (WMS, WMTS, etc.) and stores these in the geospatial database. The web service requests can be made at various image resolutions depending on the most desirable ground sample distance (GSD) for a given feature type. After image acquisition, each feature undergoes a QAR. The QAR involves loading the acquired imagery stored in the geospatial database into another custom web application along with its semantic label. Then, the feature's semantic label, geographic location, and/or tertiary metadata (e.g., quality labels) can be revised during the QAR and the updated/corrected information will then be stored in the geospatial database. Typically, no revision is needed during the QAR due to the earlier curation step, but this can be highly dependent on the experience level of the curation imagery analyst (IA). Thus, the two step curation + QAR workflow allows for less experienced analysts to perform curation while ensuring high quality/accuracy by allowing a more experienced analyst to perform the QAR. After curation and QAR, labeled image datasets containing semantically identical features can be pulled from the geospatial database and used for a wide variety of machine learning research and development activities.

2.2 *Surface-to-Air Missile Background*

A SAM is a ground-launched missile intended to intercept and destroy approaching enemy aircraft and/or missiles. Operational SAM systems were first deployed in the 1950s and have

Fig. 2 Transporter-erector-launcher (TEL) vehicle for U.S. Patriot SAM system with four rectangular missile launch canisters. The TEL is shown in the erect position ready for missile launch.

replaced most other forms of antiaircraft weapons (e.g., artillery) in modern armed forces throughout the world. SAM systems are an integral part of multilayered air defense networks and can have a maximum range up to many hundreds of kilometers. Modern SAM systems utilize transporter-erector-launcher (TEL) vehicles to carry, elevate to a firing position, and then launch missiles from multiple launch tubes. Figure 2 shows an example of a U.S. Patriot SAM TEL.

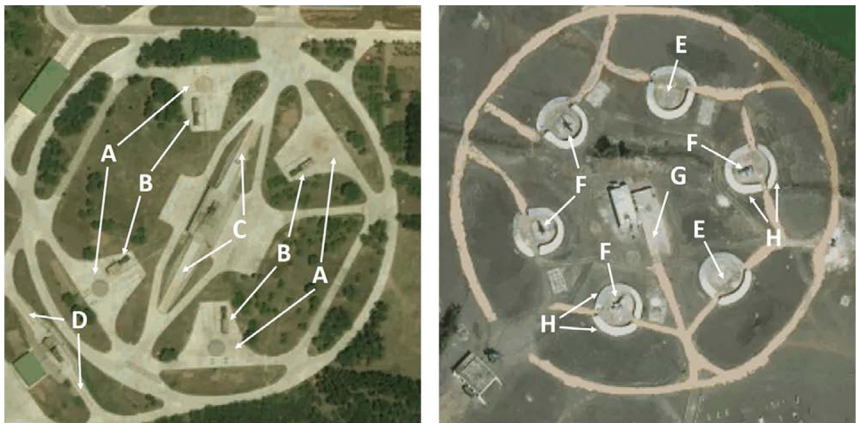
SAM sites typically contain four to six launch surfaces or pads placed in a circular or semicircular ring around a center command and control position. These configurations allow for missiles to be launched in a variety of directions. The launch pads may be formed simply as a flat area of bare earth or more formally constructed using concrete. The launch pads are often connected by an outer circular road and are sometimes interconnected by short road segments between the pads. This connectivity allows the TELs/missiles to be quickly moved and/or repositioned as needed. The launch pads come in a variety of shapes (circular, square, trapezoid, etc.) but are most often circular. In addition, the launch pads are often, but not always, surrounded by revetments, usually made of earth, to provide both protection and concealment of the launch area from its surrounding environment.

Modern SAM sites are usually accompanied by fire control radar systems that can also be vehicle mounted for mobility. The mobile fire control radars are often placed on elevated ramp platforms to enhance the radar's signal propagation relative to clutter in the surrounding local area. The radar ramp platforms are sometimes situated in the center of the ring of launch pads but they can also be placed in nearby locations adjacent to the site. Figure 3 provides two examples of typical SAM-site configurations.

2.3 *SAM Site Training Data Curation*

For this study, we applied the framework described in Sec. 2.1 to an open-source compilation of worldwide SAM sites and related features.[19] This compilation included class labels for historical SAM sites, i.e., not in use for a long period of time, inactive SAM sites, active SAM sites, and other related features, such as early warning radar sites, missile training and test sites, military garrisons, etc.

The geospatial point dataset of SAM sites and related features was first ingested into the geospatial database. Next, the point data were examined sequentially in the image curation web application to evaluate high-resolution commercial EO satellite imagery provided by DigitalGlobe's Maps API Premium Imagery web service. The curation web application was used to (1) visually evaluate the online source imagery quality (e.g., cloud free, image resolution,



(a) (b)

Fig. 3 Two examples of typical circular/symmetrical SAM-site configurations. (a) SAM site with four trapezoidal launch pads (A); each with a mobile missile TEL (B) (see also Fig. 2) on the launch pad. A center double-entry ramp platform (C) and another double-entry ramp platform (D) in the SW corner likely have mobile fire-control radar systems on top of each platform ramp. (b) SAM site with six circular launch pads (E) with missiles (F) present on four of the launch pads. A center single-entry ramp platform (G) is present but the platform is empty. Each of the launch pads is surrounded by two semicircular concrete revetments (H).

etc.), (2) determine if the source imagery contained the location of a valid SAM site (either inactive or active), and (3) properly adjust the point's geographic location to the approximate center of the SAM site.

Step 2 was necessarily subjective because evaluation of a single high-resolution EO source image cannot be reasonably used to discriminate between active versus inactive versus historical SAM sites in many circumstances. This is because the lack of TELs and/or missiles cannot, by itself, be used to determine a SAM site's status simply because modern SAM systems are designed to have mobile missiles that are routinely moved between sites, moved for military exercises, or concealed/stored in nearby structures for short periods of time when a site is temporarily not an active duty.

Consequently, TELs and/or missiles will not always be present or observable at a given SAM site location at any one time. Thus, the evaluation of whether or not a potential SAM site location was included in the final SAM site dataset was based on a visual assessment of whether the site contained multiple missile launch areas and other related features (e.g., radar platforms, military vehicles, etc.) in some recognizable pattern that could be reasonably interpreted, with high confidence, as a SAM site, either active or inactive.

After passing the initial curation stage, 512 m $\times$ 512 m image chips with a nominal 1-m resolution GSD were downloaded for the valid SAM site locations and stored in the geospatial database. The downloaded image chips then underwent a QAR (Sec. 2.1) to verify that the image chips stored in the database were valid. A small percentage of the image chips were rejected during QAR because a 1-m GSD image chips were not correctly downloaded from the web mapping service or because the initial evaluation of the SAM site's validity was deemed questionable or in error by a more experienced IA. A total of 2570 valid SAM sites were stored in the geospatial database after QAR.

Finally, the 2570 SAM site image chips that passed the QAR stage were visually evaluated to assess if a characteristic pattern of missile launch pads were present within a smaller ~256 m $\times$ 256 m window (256 $\times$ 256 pixels at 1-m GSD) centered on each 512 m $\times$ 512 m image chip. Valid SAM sites were excluded from the final curated SAM site dataset used in this study if the missile launch pads were outside this truncated chip size. This was necessary because the DCNN used in this study have a maximum input window size of 227 $\times$ 227 pixels. As a result, only SAM sites that were less than ~250 m in diameter were used for this study to ensure that a majority of a SAM site's salient features (e.g., launch areas) were present in the input processing window used by the DCNN. As a result, 361 valid SAM sites stored in the geospatial database

Table 1 Regional and country details for the number of valid SAM sites present in the curated worldwide dataset.

Africa	Asia	Europe and Russia	Middle East	North and South America	Total
455	469	779	410	96	2209
Top three countries					
Egypt	China	Russia	Syria	U.S.	
357	108	230	170	76	
Libya	Kazakhstan	Ukraine	Iraq	Cuba	
42	82	113	83	12	
Algeria	North Korea	Germany	Iran	Peru	
18	54	113	60	8	

were excluded from the final curated SAM site dataset used in this study because of this size constraint.

2.4 *Curated SAM Site Dataset Summary*

The curated worldwide dataset of labeled SAM site image chips resulting from the process described in Sec. 2.3 contains SAM sites from Africa, Asia, Europe/Russia, Middle East, and North/South America. Table 1 provides a breakdown of the number of SAM sites for each geographic region along with the number of SAM sites for the top three countries in each region. The countries containing the most SAM sites for these regions were Egypt (357), China (108), Russia (230), Syria (170), and the U.S. (76). Figure 4 shows representative examples of the curated SAM site image chips for various countries.

In total, the curated worldwide SAM site dataset used in this study consists of 2209 labeled image chips at 1-m GSD. The vast majority (>99%) of the labeled SAM site image chips are RGB images. However, 14 black and white (B/W) image chips were included in the final curated dataset due to the unavailability of RGB imagery from DigitalGlobe's Maps API image service at a few SAM site locations.

In Sec. 5, DCNN object detection is used to assist in BAS for SAM sites over a very large study area in southeastern China of ~94,000 km^2. This requires processing large volumes of imagery with a binary DCNN classifier to automatically detect image chips with candidate SAM site locations while rejecting all other image chips in the study area. This binary classification approach requires counter or negative examples for training the DCNN. Furthermore, it is desirable for the negative training examples to have similar land cover as the image chips that contain valid SAM sites. Thus, four counter image chips were downloaded for each curated SAM site location using a 5 km offset from the SAM site center point in the four cardinal directions (N/S/E/W). As a result, $4 \times 2209 = 8836$ counter or negative training image chips were added to the valid SAM site training dataset. Thus, the total size of the DCNN training dataset containing both positive and negative SAM site training examples was 11,045 image chips.

While the description provided in Sec. 2.2 and the examples shown in Figs. 3 and 4 are typical of most SAM sites, there is a great deal of variability in the morphology, construction materials, land cover, concealment, and many other factors related to the overall visual presentation of SAM sites. Figure 5 provides examples showing the variability of SAM sites present in the curated worldwide dataset used in this study. This heterogeneity presents a significant challenge for automated object/target recognition using traditional hand-crafted feature extraction approaches for machine learning. Consequently, DL approaches are particularly attractive for automated object/target detection because robust features for object recognition can be learned given a sufficiently large training dataset that adequately represents the heterogeneity for a given object/target class of interest.

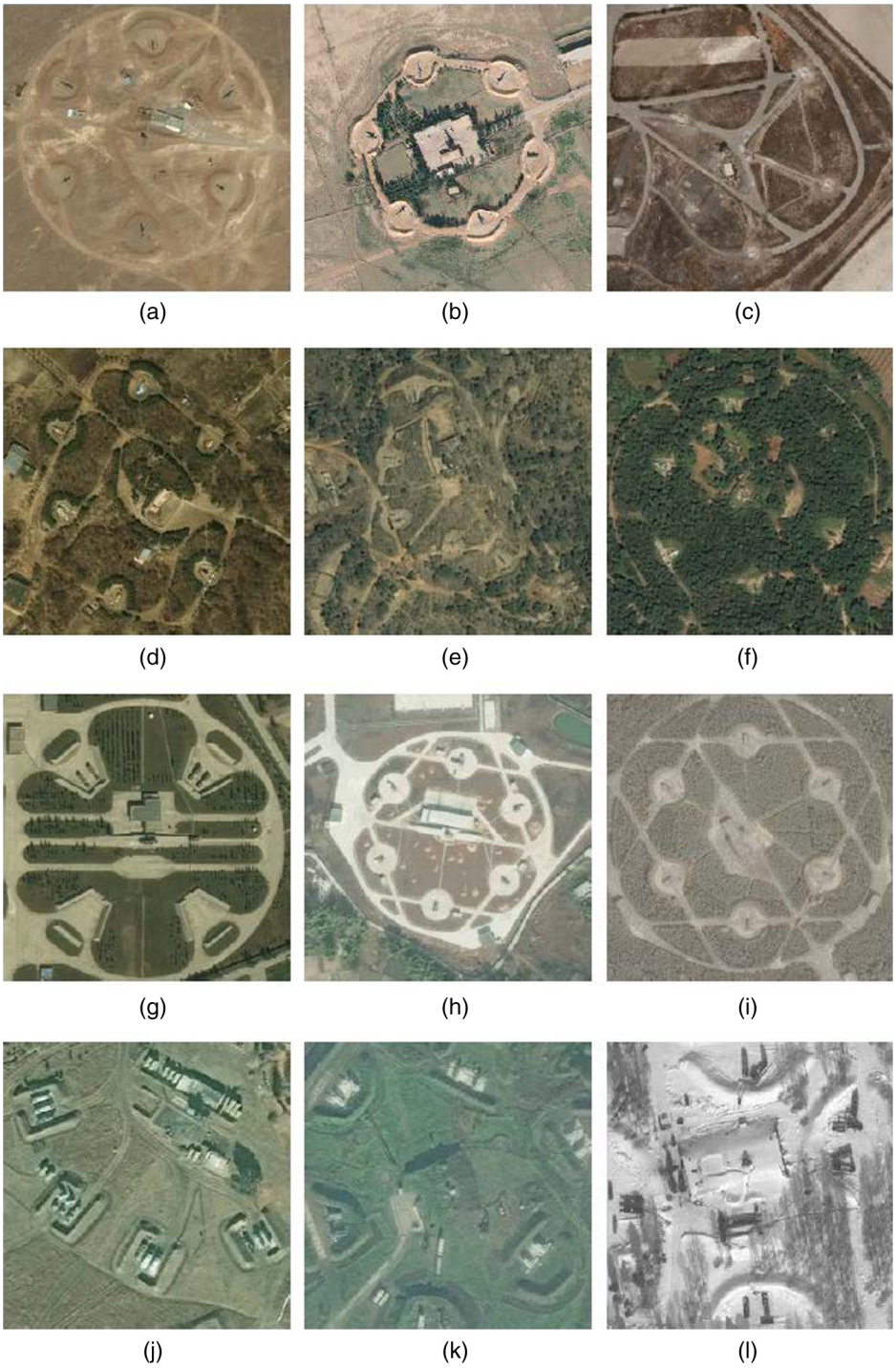


Fig. 4 Curated SAM site image chip examples for various countries. (a–c) Iran. (d–f) North Korea. (g–i) China. (j–l) Russia. Each curated SAM site image chip is 256×256 pixels at 1-m GSD.

3 SAM Site Detection Using Deep Convolutional Neural Networks

3.1 *DCNN Architectures*

Table 2 provides a summary of the DCNN utilized in this study. Specifically, the CaffeNet,[20] GoogLeNet,[13] ResNet-50, and ResNet-101[21] DCNN were all experimentally evaluated for SAM site detection using the curated worldwide dataset described in Sec. 2. The CaffeNet DCNN is a derivative of the AlexNet[15] architecture implemented in the Caffe DL framework.[20] The

Fig. 5 Examples of atypical SAM sites relative to the more typical SAM site configurations shown in Figs. 3 and 4. The examples shown here are just a small sample illustrating the large variability in morphology, land cover, construction materials, etc., present in the curated worldwide SAM site dataset used in this study.

Table 2 Overview of the DCNN architectures used in this study.

DCNN	Type	Convolutional layer depth	# convolutional weights
CaffeNet	Basic CNN	5	2.3M
GoogLeNet	CNN + inception layers	21	6M
ResNet-50	CNN + residual connections	49	23.5M
ResNet-101	CNN + residual connections	100	48M

CaffeNet architecture has five convolutional layers for the feature extraction phase followed by three fully connected layers for the classification phase. The three fully connected layers consist of two inner product layers followed by a "softmax" output layer. The softmax layer produces a probability distribution across the DCNN output classes.

The GoogLeNet[13] architecture has an overall depth of 21 convolutional layers. The novel feature of this DCNN is the so-called inception layer. An inception layer takes the output of the previous layer and performs 3×3 max pooling, 1×1, 3×3, and 5×5 convolutions. All of these outputs are stacked into one feature vector and then passed on to the next layer.

The variable sizes of the convolutional kernels allow features at various scales to be extracted and utilized simultaneously. In addition, three output classification layers are used at various stages of the DCNN to improve error backpropagation to the lower levels because of the overall depth of the network.

The ResNet[21] architecture is a derivative of VGG Nets,[22] which primarily uses 3×3 convolutional filters. The following rules govern the architectural design of the ResNet family of DCNN. First, if the output of the feature map is the same, then the same number of convolutional filters will be used. Second, if the output of the feature map is halved, then twice as many convolutional filters will be used. Finally, the ResNet architecture employs residual connections that bypass two or more convolutional layers at a time. These shortcuts allow errors to propagate backward through the network easier, which then allows for even deeper DCNN than CaffeNet and GoogleNet. Consequently, ResNet DCNN can have up to 200 convolutional layers in their neural network architecture.

3.2 *DCNN Training and Validation*

We employed techniques from Scott et al.[14] to develop robust DCNN classifiers using the curated worldwide SAM site dataset. Specifically, we utilized a combination of data augmentation and transfer learning from pretrained DCNN models. We adopted this approach based on the experimental results from Ref. 14 that demonstrated a 3% to 4% increase in overall classification accuracy for the benchmark UCM dataset (see Sec. 1) compared to using only pretrained DCNN models with no data augmentation for three different DCNN architectures (CaffeNet, GoogLeNet, and ResNet).

Data augmentation was used to greatly enhance the amount of labeled image data needed for training the DL networks. Each source image chip was first mirrored about its y axis to start the data augmentation process. Then, the original and mirrored image chips were each rotated by θ_i, where $\theta_i = 5 * i$ for $i = 0,1,2,\ldots,71$. This resulted in a $2 \times 72 = 144X$ increase in the size of the original, unaugmented training dataset. The resulting total size of the worldwide SAM site training dataset was $11{,}045 \times 144 = 1{,}590{,}480$ image chips (positive and negative examples) after augmentation.

The transfer learning used pretrained DCNN feature extraction weights that were originally learned from the ImageNet dataset[15] that consists of 1.2 million labeled ground photos. The pretrained feature extraction weights from the CaffeNet, GoogLeNet, ResNet-50, and ResNet-101 DCNN were fine-tuned during the SAM site training using a learning rate of 0.001 and a momentum of 0.9.

The DCNN were evaluated for SAM site detection using a fivefold cross validation technique. Fivefold cross validation partitions the data into five disjoint sets for both DCNN training and testing. For each fold, four of the partitions are used for DCNN training and the remaining single partition is used for DCNN testing/validation. Statistical results from each testing/validation fold are then averaged across all fivefolds to characterize overall DCNN performance. Cross-validation testing is useful for assessing the learned model's ability to generalize to an independent dataset.

We created two different sets of DCNN based on the type of image chips used for DCNN training. The first set of DCNN was trained for each of the fivefolds using only the augmented RGB image chips. The second set of DCNN was trained for each of the fivefolds using augmented RGB and augmented B/W image chips. The motivation for creating the RGB + B/W trained DCNN was to increase the robustness of the DCNN classifier when only B/W source imagery is available (see Sec. 5.1).

The augmented B/W image chips were derived from their corresponding augmented RGB image chips by converting the RGB pixel values in each image chip to their relative luminance using $0.21 * \mathrm{R} + 0.71 * \mathrm{G} + 0.07 * \mathrm{B}$. As a result, the size of the augmented RGB and B/W DCNN training dataset was double that of the ~1.6 M augmented RGB image chip dataset.

After training, DCNN testing was done separately for each of the fivefolds using only the original 2209 image chips (RGB or B/W), i.e., the augmented image chips were not used for DCNN testing/validation. Statistical results from each fold were computed and then averaged

Table 3 Fivefold cross validation statistics for RGB-only trained DCNN applied to the curated worldwide SAM site dataset.

Testing data	RGB				B/W			
DCNN	TPR (%)	TNR (%)	ACC (%)	AUC (%)	TPR (%)	TNR (%)	ACC (%)	AUC (%)
CaffeNet	90.3 ± 2.1	98.6 ± 0.3	94.4 ± 1.2	98.4	89.6 ± 3.5	97.1 ± 1.0	93.3 ± 2.3	97.9
GoogLeNet	93.4 ± 1.8	98.5 ± 0.6	96.0 ± 1.2	98.5	95.9 ± 1.7	96.6 ± 1.1	95.9 ± 1.4	98.5
ResNet-50	94.0 ± 1.0	98.6 ± 0.2	96.3 ± 0.6	99.0	95.1 ± 1.0	96.8 ± 0.7	95.9 ± 0.9	99.0
ResNet-101	94.4 ± 0.7	98.8 ± 0.3	96.6 ± 0.7	99.5	95.1 ± 1.1	97.1 ± 0.8	96.1 ± 0.9	99.3

Table 4 Fivefold cross validation statistics for RGB + B/W trained DCNN applied to the curated worldwide SAM site dataset.

Testing data	RGB				B/W			
DCNN	TPR (%)	TNR (%)	ACC (%)	AUC (%)	TPR (%)	TNR (%)	ACC (%)	AUC (%)
CaffeNet	91.2 ± 2.2	98.5 ± 0.3	94.9 ± 1.2	98.5	89.4 ± 1.8	98.7 ± 0.4	94.0 ± 1.1	98.3
GoogLeNet	93.5 ± 1.0	98.7 ± 0.3	96.1 ± 0.6	98.8	93.2 ± 1.4	98.5 ± 0.4	95.9 ± 0.9	98.6
ResNet-50	94.2 ± 0.4	98.5 ± 0.3	96.4 ± 0.4	99.3	94.0 ± 0.8	98.3 ± 0.3	96.2 ± 0.6	99.3
ResNet-101	94.1 ± 1.4	98.8 ± 0.2	96.4 ± 0.8	99.4	93.9 ± 1.2	98.6 ± 0.4	96.2 ± 0.8	99.5

across all five testing folds for TPR = true positive rate, TNR = true negative rate, accuracy = ACC = $(\mathrm{TPR} + \mathrm{TNR})/2$, and AUC, which is the area under the ROC curve. Tables 3 and 4 summarize the fivefold cross validation statistical results for these DCNN experiments. CaffeNet had the worst statistical performance for all the DCNN, and this is very likely due to the fact that it is the most shallow of the DCNN utilized in these experiments (see Table 2).

For the RGB-trained DCNN (Table 3), the ACC average and standard deviation across all fivefolds of the RGB testing data were: CaffeNet with $94.4 \pm 1.2\%$, GoogLeNet with $96.0 \pm 1.2\%$, ResNet-50 with $96.3 \pm 0.6\%$, and ResNet-101 with $96.6 \pm 0.7\%$. We also evaluated the RGB-trained DCCN using only the B/W testing data. The results in Table 3 show that the average ACC and AUC dropped by <1.2% for each DCNN relative to the results from the RGB testing data and by <0.5% for the three deepest DCNN (i.e., excluding CaffeNet). This suggests that color information in the RGB image data does not add much discriminating power for DCNN SAM site detection. This seems intuitively reasonable given that morphology and/or the spatial relationships between constituent objects within a SAM site should dominate the DCNN feature selection.

For the RGB and B/W-trained DCNN (Table 4), the ACC average and standard deviation across all fivefolds for the RGB testing data were: CaffeNet with $94.9 \pm 1.2\%$, GoogLeNet with $96.1 \pm 0.6\%$, ResNet-50 with $96.4 \pm 0.4\%$, and ResNet-101 with $96.4 \pm 0.8\%$. As before, we evaluated the RGB and B/W-trained DCNN using only the B/W testing data and found that the average ACC and AUC dropped by <1% relative to the RGB testing results for all four DCNN and <0.5% for the three deepest DCNN (i.e., excluding CaffeNet).

The statistical results in Tables 3 and 4 and summarized above are all based on a 50% confidence threshold for the final DCNN output classification. Figure 6 provides a more comprehensive view of the overall performance of the RGB and B/W-trained DCNN at various confidence thresholds, where the average ACC is plotted as a function of the confidence threshold. From the plot, we see that GoogLeNet had a higher average ACC than the other DCNN for confidence thresholds >70% but that ResNet-101 had the best performance for confidence thresholds <70%. In addition, CaffeNet had the worst performance for all confidence values,

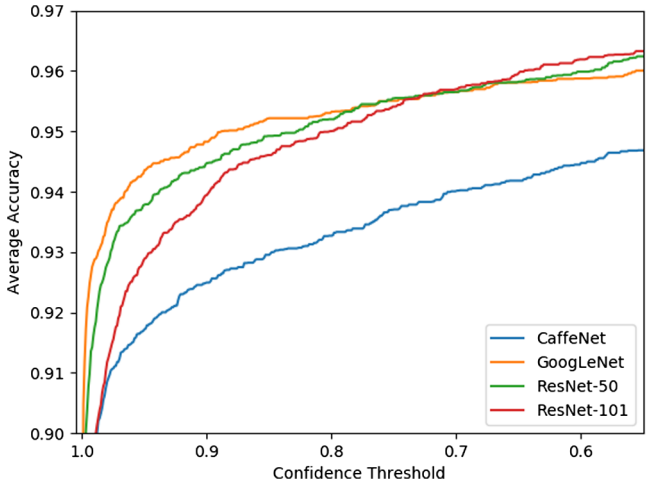


Fig. 6 Average DCNN accuracy from fivefold cross-validation results for the curated worldwide SAM site dataset. GoogLeNet has the best accuracy for confidence values >70%, whereas ResNet-101 has the best accuracy for lower confidence values. CaffeNet's performance is significantly worse than the other three DCNN, and this is very likely due to CaffeNet's much shallower network architecture compared to the other DCNN.

and this is very likely due to the fact that it is the most shallow of the DCNN utilized in these experiments (see Table 2).

Given the large training dataset size of ~1.6 M samples, DCNN differences for any of the performance metrics (e.g., ACC, AUC) in Tables 3 or 4 that are 0.1% or larger are statistically significant at the 99% confidence level ($p < 0.01$). In looking at the trends in ACC and AUC from Tables 3 and 4 across the four DCNN, we consistently see that the DCNN performance increases as the network depth increases (see Table 2). This suggests that the deeper networks are better able to leverage low-level features by aggregating successive levels to develop high-level visual concepts that facilitate more refined image understanding and therefore improve the overall performance of the final classification.

Overall, the ResNet-101 DCNN had the best ACC and AUC for both the RGB and B/W testing data even though GoogLeNet had better ACC values for high-confidence thresholds. Finally, the excellent overall ResNet-101 performance with ACC > 96% and AUC > 99% on the curated worldwide SAM site dataset demonstrates the great potential of DCNN for automated and/or semiautomated search and discovery of objects/targets of interest in high-resolution remote sensing imagery.

4 Visual Broad Area Search

As stated in Sec. 1, the primary goal in this study is to evaluate how automated DCNN object detection can be used to assist in the labor-intensive process of human visual searches for objects/targets of interest in high-resolution EO imagery over very large search areas. We use the term BAS to describe this process. In order to quantitatively compare the performance and assess the potential benefits of DCNN for assisting in the BAS process, we require a baseline set of experimental results, where human BAS performance is measured. In this section, we describe a baseline BAS for SAM sites over a large study area in China. Summary data are then presented to quantify human BAS performance in terms of statistical accuracy relative to ground truth SAM site data as well as time spent (human labor hours) performing the BAS.

4.1 *China Study Area*

Over the last 20 years, China has developed a modern multilayered air defense network that includes both short and long range mobile SAM systems. In addition, significant tensions exist between China and Taiwan and between China and other countries surrounding the

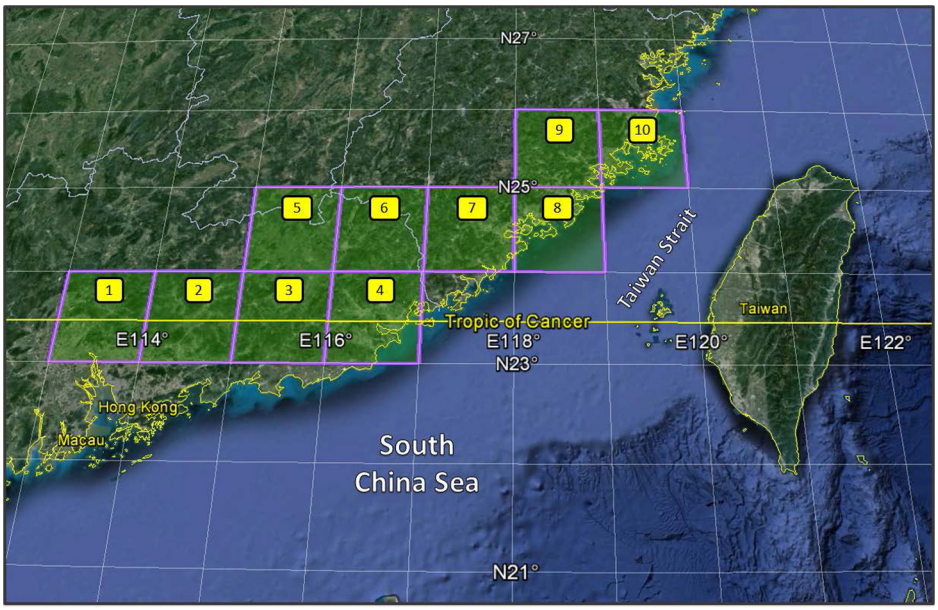


Fig. 7 Overview of the study area along the southeast coast of China. The study area is divided into ten 1 deg × 1 deg AoIs. See Table 5 for more details.

South China Sea due to ongoing territorial water disputes. Considering these factors, we selected a study area along the southeast coast of China immediately west of Taiwan and north of the South China Sea. Figure 7 provides an overview of the study area, which is divided into ten 1 deg × 1 deg areas of interest (AoIs).

Table 5 provides additional summary information for the ten AoIs. In total, the 10 AoIs have a land area of ~94,000 km^2. In addition, there were a total of 14 SAM sites in the study area that were known *a priori* since they were part of the open-source compilation of SAM sites[19] used to produce the curated worldwide SAM site dataset (see Sec. 2). By design, we selected three AoIs (2, 3, and 6) that had no known SAM sites from the open-source dataset in Ref. 19. For the remaining AoIs, four AoIs (4, 5, 8, and 9) had only one known SAM site and three AoIs

Table 5 Summary information for the BAS study area along the coast of southeastern China. The study area contains 10 1 deg × 1 deg AoI, where the alphanumeric designation (column 2) is the lower left latitude and longitude for each AoI. See also Fig. 7.

AoI #	AoI	Area (km^2)	Water (%)	Land area (km^2)	# SAM sites
1	23°N113°E	11,041	0%	11,041	4
2	23°N114°E	11,041	0%	11,041	0
3	23°N115°E	11,041	0%	11,041	0
4	23°N116°E	11,041	18%	9,036	1
5	24°N115°E	10,956	0%	10,956	1
6	24°N116°E	10,956	0%	10,956	0
7	24°N117°E	10,956	3%	10,578	3
8	24°N118°E	10,956	63%	4,033	1
9	25°N118°E	10,867	2%	10,629	1
10	25°N119°E	10,867	59%	4,443	3
	Total	109,724	14.6%	93,756	14

Table 6 Geographic coordinates for SAM sites located in the BAS study area (see also Table 5 and Fig. 7). All 14 SAM sites were known *a priori* to be in the study area because they were part of the open-source dataset[19] used to produce the curated worldwide SAM site dataset (Sec. 2).

AoI #	Lat (deg)	Long (deg)
1	23.1425	113.0553
1	23.1615	113.3923
1	23.1517	113.3989
1	23.5123	113.2768
4	23.4537	116.7189
5	24.1624	115.7718
7	24.2160	117.9384
7	24.5684	117.6604
7	24.5012	117.8766
8	24.6687	118.2830
9	25.0331	118.8075
10	25.3462	119.0679
10	25.5762	119.4534
10	25.5927	119.4493

(1, 7, and 10) had three or four known SAM sites. Table 6 provides the geographic coordinates for all 14 of the *a priori* SAM sites.

4.2 *Visual Search*

We use the term IA in this section and throughout the rest of this paper to convey the functional role of someone who, in the context of this study, performs a visual search of high-resolution EO imagery for SAM sites. The term is not intended to convey or represent that these individuals were professionals whose primary work role was to perform imagery analysis.

Three IAs (IA1, IA2, IA3) were selected to conduct a visual BAS over the study area. These IAs were all undergraduate students enrolled at the University of Missouri, who had significant, but part-time, prior work experience analyzing high-resolution EO imagery. Specifically, they had between 500 and 900 h (average of ~700 h) of recent experience analyzing high-resolution EO imagery to identify, extract, and attribute a wide variety of cartographic features within a GIS production environment. Furthermore, they were selected from a group of undergraduate students with similar experience because their work performance relative to their peers was considered by their immediate supervisor to be well above average. However, they did not have any prior experience or knowledge related to SAM sites or the identification of SAM sites in high-resolution EO imagery. With respect to the two groups of IAs that performed the DCNN-assisted BAS presented in Sec. 5, we characterize the experience level of IAs IA1, IA2, and IA3 as intermediate.

The three selected IAs were given a 1-h overview briefing on how to identify SAM sites in high-resolution EO imagery. They were also provided: (1) a hardcopy set of ~15 image chips of SAM sites to serve as a visual reference, as needed, during the visual BAS of the study area, and (2) a one page high-level bullet summary of various factors to consider when trying to identify or invalidate a potential SAM site during their visual BAS. The SAM site image chips provided in item 1 were all from China but did not, by design, include any of the *a priori* known SAM sites (see Table 6) within the study area.

The visual BAS for each AoI was subdivided into a 4×4 search grid (0.25 deg $\times$ 0.25 deg) of 16 total sub-AoI regions. The three IAs performed their visual BAS using commercial desktop GIS software that could access and display multiple high-resolution EO basemap imagery layers over the study area. The IAs recorded vector points in the GIS software at each location they visually identified to be a SAM site. The IAs also recorded the amount of time spent conducting the visual search separately for each of the 16 sub-AoI regions. IAs were instructed not to discuss any aspect of their visual BAS with anyone else. Finally, no additional instruction or guidance was provided to the three IAs during their visual search across the entire study area.

4.3 *Visual Search Results*

After the visual search was completed for the entire study area, each of the three IAs produced a single master shapefile identifying their detected SAM site locations and an excel sheet with their recorded search times for all 16 subregions in each of the 10 AoIs. Each point location within the IA's master shapefile was evaluated against the known SAM site locations (Table 6) to determine if the SAM site was a true positive (TP) based on the *a priori* SAM site ground truth. Each additional point in the master shapefile that was not a TP based on the *a priori* SAM site information was then evaluated to determine if it correctly represented a new SAM site (NSS) location. If the additional point was positively identified as a NSS, then it was added to the total TP count, otherwise it was counted as a false positive (FP). Finally, any *a priori* SAM site locations not found within the IAs master shapefile was counted as a false negative (FN).

Table 7 provides the performance data for the visual BAS and detection of SAM sites by the three IAs for each AoI in the study area. All three IAs correctly identified two new SAM site locations, one in AoI 2 and the other in AoI 7. The two NSS along with their geographic locations are shown in Fig. 8. The semicircle layout of the launch pads in the new SAM site in AoI 7 is unique compared to all the other SAM sites in the curated China SAM site dataset (>100; see Table 1). The eastward orientation of the new SAM site in AoI 7 site directly faces Taiwan and indicates that this site is very likely a long-range SAM system designed to defend the airspace over the Taiwan strait (see Fig. 7).

Adding the two NSS with the 14 *a priori* SAM sites from Table 6 brings the total number of ground truth positive (P) SAM site points in the study area to be P = 16. Across all three IAs,

Table 7 IA performance data for traditional/visual BAS and detection of SAM sites for the southeastern China study area. The three IAs (IA1, IA2, IA3) had an intermediate level of experience. See Sec. 4.3 for description of the reported values.

	Analyst	IA1				IA2				IA3			
AoI #	AoI	TP	FN	FP	NSS	TP	FN	FP	NSS	TP	FN	FP	NSS
1	23°N113°E	3	1	0		2	2	1		2	2	0	
2	23°N114°E	1	0	0	1	1	0	0	1	1	0	0	1
3	23°N115°E	0	0	0		0	0	0		0	0	0	
4	23°N116°E	1	0	0		1	0	0		1	0	0	
5	24°N115°E	1	0	0		1	0	0		1	0	0	
6	24°N116°E	0	0	0		0	0	0		0	0	0	
7	24°N117°E	4	0	0	1	4	0	0	1	4	0	0	1
8	24°N118°E	1	0	0		1	0	1		1	0	0	
9	25°N118°E	1	0	0		1	0	0		1	0	0	
10	25°N119°E	3	0	0		1	2	0		3	0	0	
	Total	15	1	0	2	12	4	2	2	14	2	0	2

Fig. 8 Two new SAM site locations that were correctly identified by three different IAs during their visual BAS of the China study area. The semicircular layout of the SAM site launch pads in AoI 7 is unique compared to all other SAM sites (>100; see Table 1) in the curated China SAM site dataset. The eastward orientation of the new SAM site in AoI 7 directly faces Taiwan and indicates that site is very likely a long-range SAM system designed to defend the airspace over the Taiwan Strait (see Fig. 7).

the total number of TP, FP, and FN detections varied from 12 to 15, 1 to 2, and 0 to 2, respectively. Table 8 provides the IA search time (labor hours) for each of the 10 AoIs and the total search time for the study area. The total search time for all 10 AoIs varied between 40 and 80 h for the three IAs with an average total search time of ~60 h.

We evaluated the visual BAS performance of the IAs using three statistical measures: true positive rate $(\mathrm{TPR}) = \mathrm{TP}/\mathrm{P}$, positive predictive value $(\mathrm{PPV}) = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$, and $\mathrm{F1} = 2 * \mathrm{TP}/(2 * \mathrm{TP} + \mathrm{FP} + \mathrm{FN})$. Table 9 provides these summary statistics for each IA's visual BAS performance over the complete study area. IA1 had the top performance among the three IAs

Table 8 AoI search times (h) for three IAs (IA1, IA2, IA3) that performed a traditional/visual BAS over the southeastern China study area.

		Visual BAS time (h)			
AoI #	AoI	IA1	IA2	IA3	AVG
1	23°N113°E	12.3	11.3	10.4	11.3
2	23°N114°E	8.7	8.8	4.8	7.4
3	23°N115°E	9.4	5.6	4.6	6.5
4	23°N116°E	7.1	6.6	3.4	5.7
5	24°N115°E	9.4	6.8	3.4	6.5
6	24°N116°E	7.3	5.6	2.5	5.1
7	24°N117°E	7.9	5.9	3.9	5.9
8	24°N118°E	3.4	3.2	2.2	2.9
9	25°N118°E	6.1	4.9	3.2	4.7
10	25°N119°E	5.3	2.9	2.2	3.4
	Total	76.8	61.5	41.5	59.9

Table 9 IAs overall statistical performance for traditional/visual BAS and detection of SAM sites in the China study area.

Analyst	TPR (%)	PPV (%)	F1 (%)
IA1	93.8	100	96.8
IA2	75.0	85.7	80.0
IA3	87.5	100	93.3
AVG	85.4	95.2	90.0

with a TPR of 93.8%, PPV of 100%, and F1 of 96.8%. We note that IA1 also spent the most total time (77 h; Table 8) of the three IAs to complete the visual BAS for the study area. Since the F1 score is the harmonic mean of TPR and PPV, and includes positive detections (TP) and both error types (FP and FN) in its calculation, we use this as the single best measure of IA performance for the visual BAS across the entire study area. The aggregate F1 score for the three IAs was 90.0% for the study area.

5 DCNN-Assisted Broad Area Search

In this section, we first describe the technical approach used to develop and apply a suitably trained DCNN for large-scale processing and detection of SAM sites in the southeastern China study area. Following this, we describe strategies for spatial postprocessing of the DCNN detection results and the subsequent integration of the postprocessed DCNN results into a simple user interface to assist human analysts in a BAS application. Finally, summary data are presented to quantify the DCNN-assisted human BAS performance, and those results are then compared with the baseline visual BAS results given in the previous section.

5.1 *China DCNN Training*

The trained DCNN models developed and evaluated in Sec. 3 from the curated worldwide SAM site dataset were not utilized for processing source imagery for the study area AoIs (Sec. 5.2). One reason for this is because the *a priori* SAM sites from the China study area (Table 6) were included in various DCNN training partitions for the curated worldwide SAM site data and experiments presented in Sec. 3. Thus, including the *a priori* SAM sites in the DCNN training for the AoI, processing could potentially bias the China AoI experimental results by artificially increasing the detection rate (TPR) of the *a priori* SAM sites.

Furthermore, we speculated that using a modified version of the curated worldwide dataset (i.e., excluding the study area *a priori* SAM sites) for DCNN training may not be the best approach for the China AoI processing given the significant amount of heterogeneity observed in the worldwide SAM site dataset (e.g., Fig. 5). Finally, we were also interested in understanding and evaluating the DCNN AoI performance using a much more limited set of DCNN training data since obtaining large labeled training datasets is currently a significant impediment for the implementation of DL methods for remote sensing applications. This is especially true when the object/target class of interest is an extremely rare and/or difficult to locate feature, as is the case for the SAM sites selected for this study.

Based on the above considerations, we selected a DCNN training dataset consisting of all SAM sites from the curated worldwide dataset that were in the country of China but were not part of the study area, i.e., the *a priori* SAM sites from Table 6 were excluded from the China-only SAM site training dataset. The resulting number of original, unaugmented RGB SAM site image chips used for the China AoI DCNN training was 94 positive training examples.

Since ~3.2% (2980 km^2) of the high-resolution EO source imagery provided by the DigitalGlobe Maps API was B/W imagery over the China study area (Fig. 7), we included B/W image chips ($2X$) in the data augmentation scheme by converting RGB values to relative luminance (Sec. 3.2). We then jittered (shifted) each RGB and B/W image chip by 10, 20, 30, and

Table 10 Fivefold cross validation statistics for RGB and B/W-trained China AoI DCNN with RGB China AoI SAM site testing data.

DCNN	TPR (%)	TNR (%)	ACC (%)	AUC (%)
CaffeNet	91.1 ± 8.4	98.7 ± 1.4	94.9 ± 4.9	98.0
GoogLeNet	93.3 ± 7.2	99.7 ± 0.6	96.5 ± 3.9	98.9
ResNet-50	96.7 ± 5.0	99.5 ± 1.2	98.1 ± 3.1	99.4
ResNet-101	96.7 ± 5.0	99.7 ± 0.6	98.2 ± 2.8	99.9

40 pixels ($4X$) in the N/E/S/W+NE/SE/NW/SW ($8X$) directions. This results in a $1X + 4X \times 8X = 33X$ increase in the size of the RGB and B/W training dataset.

The jitter augmentation was implemented to account for the fact that there is no guarantee that a SAM site will be positioned near the center of image chips extracted from the AoI source imagery for DCNN processing. Thus, the jitter will cause well-centered SAM site training examples to be slightly shifted, often resulting in partial or complete loss of some of the salient features (e.g., launch pads). This should help mitigate false rejection (FN) of true SAM sites (TP) when a jitter-trained DCNN is used to process AoI image chips that contain a partial SAM site.

We then applied the augmentation scheme from Sec. 3, where each RGB and B/W image chip (original and jittered) was flipped about its y axis ($2X$) and then fully rotated in 5 deg steps ($72X$) through 0 deg to 359 deg. The resulting increase of the positive SAM site training dataset relative to the original, unaugmented 94 RGB SAM site image chips is therefore $2X \times 33X \times 2X \times 72X = 9504X$ or $94 \times 9504 = 893{,}376$ positive SAM site training chips.

Four additional counter (negative) training image chips were obtained for each SAM site location using 5 km N/S/E/W offsets (Sec. 2.4). Applying the same data augmentation process to the original $4 \times 94 = 376$ RGB negative training examples generates a total of 3,573,504 negative training examples. The resulting size of the overall DCNN training dataset for the China AoI is therefore 4,466,880 total (positive and negative) image chips.

After constructing the China AoI DCNN training dataset, fivefold cross validation experiments (see Sec. 3.2) were performed for the RGB and B/W-trained DCNN (CaffeNet, GoogLeNet, ResNet-50, and ResNet-101). Table 10 summarizes the statistical results, at a 50% confidence threshold, for each of the China AoI DCNN using RGB testing data only (i.e., B/W testing data are not shown but were generated). As with the results presented in Sec. 3.2 for the curated worldwide SAM site experiments, CaffeNet had the worst overall performance (ACC = 94.9%, AUC = 98.0%) and ResNet-101 had the best overall performance (ACC = 98.2%, AUC = 99.9%).

In addition, the overall trend of better statistical performance in Table 4 (Sec. 3.2) for increasing depth of the DCNN (Table 2) is also present in Table 10. It is worth noting that the ACC and AUC results from Table 10 are better than the corresponding ACC and AUC results (RGB testing) in Table 4 for GoogLeNet, ResNet-50, and ResNet-101. Notably, both ResNet-50 and ResNet-101 had a >1.5% increase in ACC for the China AoI DCNN compared to the worldwide SAM site DCNN. This provides strong evidence that constructing China AoI-trained DCNN versus using the DCNN from the curated worldwide dataset was the best approach.

Moreover, it is reasonable to assume that the improved statistical performance for the China AoI DCNN (Table 10) versus the worldwide SAM site DCNN (Table 4) is due largely to the fact that the China AoI SAM site training/testing data have significantly less heterogeneity compared to the worldwide SAM site dataset. This suggests that constructing region and/or country-specific DCNN for object/target recognition may be an effective strategy when there are significant geographic variations in the visual presentation of various object/target classes.

Finally, it is important to note that the excellent ResNet-101 China AoI DCNN results (ACC = 98.2%, AUC = 99.9%) in Table 10 were obtained using <100 original (unaugmented) RGB SAM site training examples. This was achieved by applying the combination of data augmentation and transfer learning first demonstrated by Scott et al.,[14] where the pretrained DCNN weights were learned from the ImageNet dataset.[15] Scott et al.[14] obtained a 99% average ACC

across all object-like classes (e.g., airplane, tennis court, etc.) in the UCM dataset,[10] which had only 100 training examples per class. Thus, the results presented here provide further evidence that robust DCNN object/target recognition results can be achieved with DCNN for modest training dataset sizes by using data augmentation and transfer learning in lieu of much larger amounts of labeled training data normally associated with DCNN. Moreover, the China AoI DCNN obtained much better statistical performance than the worldwide DCNN for SAM site detection even though it used $>20X$ fewer training examples. Thus, having large labeled training datasets does not necessarily guarantee better performance when using DCNN in specific application scenarios.

5.2 *China DCNN AoI Processing*

All source imagery for the China study area was downloaded using DigitalGlobe's Maps API Premium Imagery,[17] where access was provided for this effort through a partnership with the DigitalGlobe Foundation.[18] Specifically, we downloaded ~66,000 1280 m × 1280 m source image tiles at a nominal resolution of 1-m GSD, where adjacent tiles had 100-m overlap in both the N/S and E/W directions. The 100-m N/W and E/W tile overlap was done to mitigate the impact of possible splitting of ~200 to 250 m diameter SAM sites between adjacent tiles if no tile overlap were used. The 1-m GSD is the same as the SAM site image chips used to train the China AoI DCNN in the previous section.

As mentioned in the previous section, B/W source imagery covered about 2980 km^2 (3.2%) of the study area, and this is why the China AoI DCNN were trained with both RGB and B/W imagery. In addition, about 5120 km^2 (5.5%) of the downloaded source image tiles, located in mostly rural, low population density areas, were at a nominal GSD >= 2 m. Since this source imagery was not at the nominal 1-m GSD of the China AoI DCNN training data, it was excluded from the China AoI DCNN processing. It should be noted that the excluded source imagery did not contain any of the *a priori* SAM site locations (Table 6) or the two new additional SAM sites (*a posteriori*) discovered during the visual BAS (Fig. 8, Sec. 4.3). Thus, all *a priori* and *a posteriori* SAM site locations were present in the final source image tiles used for the China AoI DCNN processing.

The primary criterion used to select the final DCNN for processing the China AoI source imagery tiles was the average ACC from the fivefold cross validation results presented in the previous section. The average ACC was used since we wanted to consider both the TPR and FPR (i.e., 1 − TNR) performance of the DCNN. The ACC results reported in Table 10 for the various DCNN are for a confidence threshold of 50%. However, adopting a confidence threshold this low would generate an excessive number of FP, which would, in our opinion, be highly impractical for a BAS application.

Figure 9 shows the average ACC from the fivefold cross validation results for the China AoI SAM site dataset. Since the ResNet-101 DCNN had the best ACC for confidence thresholds <= 99.75%, it was selected as the final DCNN used for processing all the China AoI source imagery. Furthermore, a 99% confidence threshold was selected for the ResNet-101 DCNN China AoI processing. The corresponding ACC, TPR, and FPR for the ResNet-101 DCNN at 99% confidence threshold were 96.5%, 94.5%, and 1.6%, respectively.

While the primary focus of this study was to evaluate the human performance using DCNN to assist in the BAS process, here, we briefly summarize the computational aspects of the China AoI DCNN training and processing. Creation of the fully augmented DCNN training dataset from the initial RGB positive and negative training example image chips took ~4.8 h. Creation of the lightning memory-mapped data base (LMDB) for each of the cross-validation folds took 3.5 h. Training the ResNet-101 DCNN with the China AoI dataset took 15 h. In total, the processing pipeline to go from the unaugmented China training dataset to a fully-trained DCNN was ~23 h on a server with dual Intel Xeon X5670 (2 CPU, 12 cores/CPU, 2.93 GHz), 96 GB of RAM, 2 Geforce 1080, 2 Geforce Titan Z, and a 1 Gb connection to network attached storage (NAS).

Next, the DCNN inference processing for the China AoIs used a sliding window of 256 × 256 pixels with a 64 pixel stride size representing a 75% overlap between image chips. Figure 10 illustrates the DCNN image chip processing grid with overlapping processing

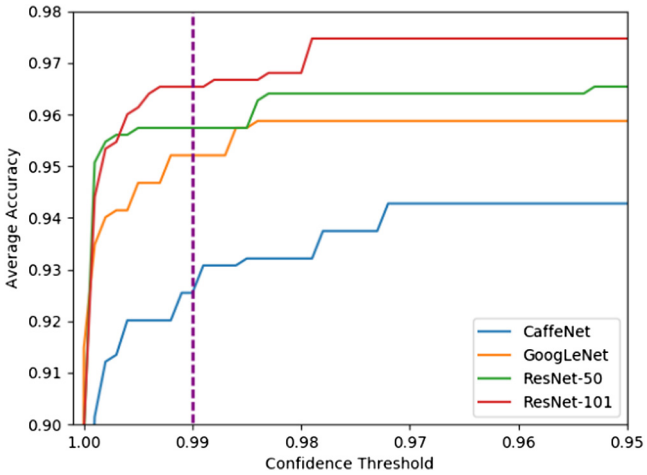


Fig. 9 Average DCNN accuracy from fivefold cross validation results for the China AoI SAM site training dataset (Sec. 5.1). The ResNet-101 DCNN had the best accuracy for confidence thresholds <= 99.75%. Consequently, the ResNet-101 DCNN at a confidence threshold of 99% (~96.5% average accuracy) was selected for China AoI processing.

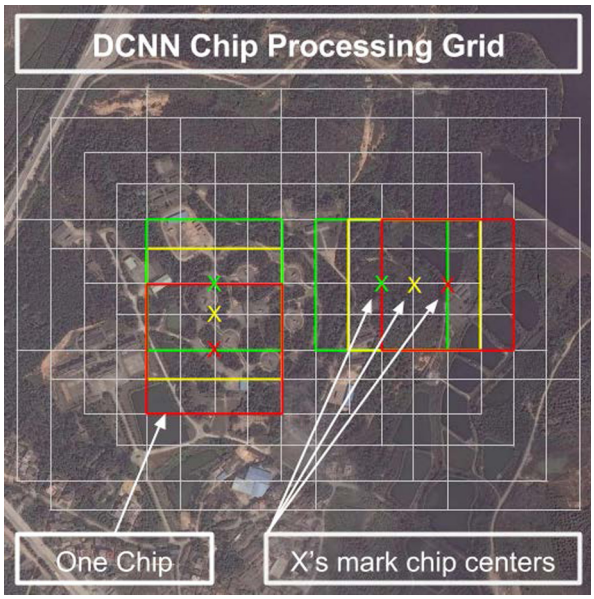


Fig. 10 Illustration showing how 1280 m × 1280 m tiles in each AoI were partitioned into 256 m × 256 m image chips with 75% overlap in both the N/S and E/W directions. Each image chip was then input into the ResNet-101 DCNN for processing to detect candidate SAM site locations.

windows. The sliding window with 75% overlap generates 289 image chips for each 1280×1280 pixel tile in the AoI. It should be noted that the Caffe framework center crops each 256×256 pixel chip to a 227×227 pixel chip for input into the ResNet-101 DCNN. As a result, the effective overlap for the sliding processing window is ~71%, i.e., slightly less than the 75% for the uncropped image chip overlap.

China AoIs with no water area have 8179 tiles and would generate a maximum number of 2,363,442 image chips (75% overlap) for DCNN processing. To facilitate faster I/O within the Caffe framework, an LMDB was generated for each AoI by first chipping the tiles (6 h) and then loading the LMDB (2 h) using a dual Intel Xeon X5667 server with 96 GB of RAM and a 1 Gb NAS. Next, the ResNet-101 DCNN inference processing took a maximum of 6 GPU hours for an AoI. However, we processed the ten AoIs in parallel using four separate GPUs, so the maximum DCNN GPU processing time for the 10 AoIs was $(10 \times 6)/4 = 15$ h.

Table 11 Summary information for ResNet-101 DCNN processing of the 10 AoIs in the southeastern China study area.

AoI #	Area processed (km^2)	Tiles[a] processed	Image chips[b] processed	DCNN chip detections[c]	Spatial clusters	Chip reduction (%)
1	11,041	8179	2,363,442	714	163	77.2
2	11,041	8179	2,363,442	899	219	75.6
3	11,041	8179	2,363,442	453	101	77.7
4	9036	6694	1,991,210	2332	494	78.8
5	10,956	8116	2,363,442	715	161	77.5
6	8562	6343	1,846,999	1,059	239	77.4
7	9600	7111	2,097,851	668	163	75.6
8	4033	2988	1,078,837	1,146	251	78.1
9	8890	6586	1,921,850	592	135	77.2
10	4443	3291	1,300,789	700	153	78.1
Total	88,643	65,666	19,691,304	9278	2079	77.6

[a]1280 × 1280 pixel image tiles at 1-m GSD with 100 pixel overlap in N/S and E/W directions.
[b]256 × 256 pixel image chips at 1-m GSD with 75% overlap in N/S and E/W directions.
[c]ResNet-101 DCNN with 99% confidence threshold cutoff.

Also, the GPU processing time for a given AoI could be significantly less than the 6 GPU h/AoI because the actual number of processed image chips for a given AoI (see Table 11; Sec. 5.3) could be less due to water coverage (Table 5) and the rejection of source image tiles that had a nominal GSD >= 2 m, as described above. Consequently, the actual processing time for the 10 AoIs in the China study area covering ~88,600 km^2 of land area was ~13 h.

5.3 *Spatial Clustering and AoI Ranking*

As detailed in the previous section, 10 AoIs covering ~88,600 km^2 of land area were processed by the China ResNet-101 DCNN. The resulting DCNN processing outputs potentially have up to 2,363,442 output points for a given AoI. The actual number of points (image chips with 75% overlap) processed for each AoI is provided in Table 11 and is variable due to water area (Table 5) and rejection of source imagery with GSD >= 2 m (Sec. 5.2). After DCNN processing, an alpha-cut is applied to eliminate all DCNN output points with a confidence <99% (Sec. 5.2, Fig. 9). As shown in Fig. 11(a), the resulting surface after the alpha-cut is a sparse spatial field of positive detections over the AoI, where spatial clustering can be observed. Thus, it is desirable to find the center of the various detection clusters since these are the most likely candidates for positive identification of SAM sites in a given AoI.

As can be seen in Fig. 12, the DCNN detection surface is often flat (1.00 confidence) over the area occupied by a SAM site. To aid in the elicitation of an unknown number of clusters in a given AoI, we transformed the sparse DCNN detection surface into an amplified spatial field forming clusters as localized density maxima peaks. The amplified spatial density surface was generated for each point by aggregating the mass of its neighbors onto a distance–decay structuring function. This is an additive function-to-function morphology, with a structuring function centered over chip c as $s(p) = \{e(-d/D) \text{ if } [d = dist(c, p)] < max\, D; \text{else } 0\}$. In this effort, $\text{dist}(c, p)$ is the haversine distance from chip center c to neighbor point p, and $max\, D$ was set to 150 m. By defining the neighborhood of a response as $N(r)\ \forall\ r \in R$, the cluster centers were located according to Algorithm 1. Figure 11(b) shows the resulting amplified spatial density over AoI 1, where the detection surface height is visualized as a circular radius and the number of contributing image chip detections is the circle's darkness.

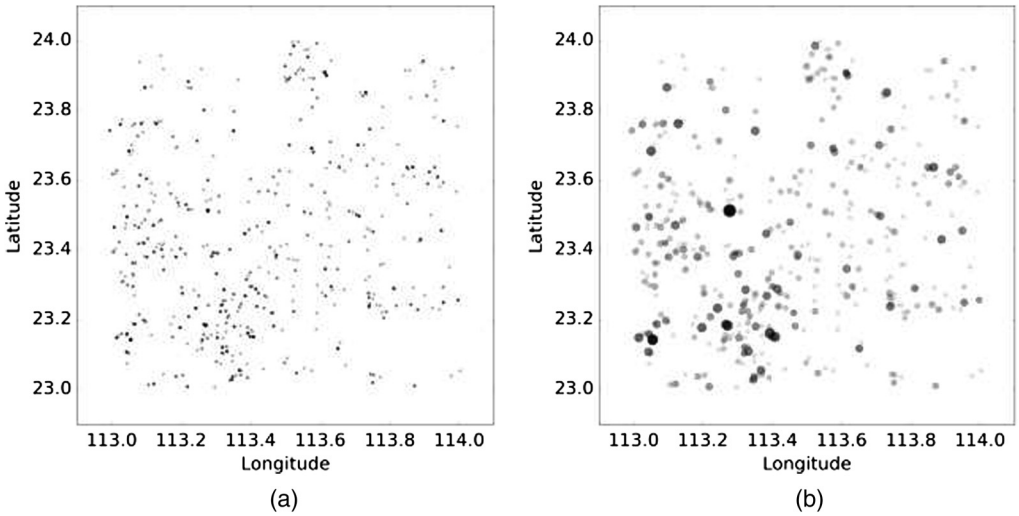


Fig. 11 DCNN detection response for AoI 1. (a) Possible SAM site locations (total of 714) after alpha-cut (<99%) on the ResNet-101 DCNN output confidence; (b) candidate SAM site locations (total of 163) from spatial clustering described in Sec. 5.3. The detection surface height is visualized as a circular radius with the number of contributing image chip detections determining the circle's darkness. In this AoI, the spatial clustering reduced the number of possible SAM site locations by 77.5% (see also Table 11).

Mode-seeking clustering algorithms are designed to discover the number of clusters without relying on the specification of expected partitions ahead of time, which is a common drawback of *k*-means variants. In the case of the amplified spatial densities, the modes within the field lie within the local maxima found as the center of mass of spatially connected densities. A straightforward method to discover modes is the mean-shift algorithm, which defines a spatial aperture of nearest neighbors to be evaluated at each iteration for each point. Each point is then moved to the center of mass of its spatial local neighbor set.

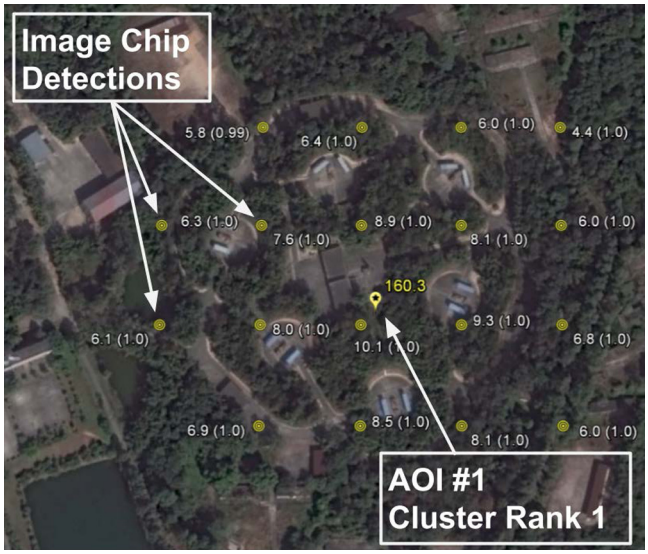


Fig. 12 DCNN detection density surface generated for a candidate SAM site in AoI 1. The image chip centers (bullseyes) are labeled with the spatially amplified density value followed by the DCNN output confidence in parentheses. The cluster center (yellow placemark) is measured as the sum of the densities (160.3) that aggregate into this cluster's overall score. The overall cluster score is then used to rank order (high to low) the candidate SAM sites for subsequent human review.

Algorithm 1 Spatial clustering of DCNN response surface R.

1. $R' = \delta[s(p), N(r)]$: function-to-function dilation of R by $s(p)$
2. While (movement < epsilon)
   a. $R' = weightedmeanshift[R', s(p)]$
   b. Movement = haversine(R'_{old}, R'_{new})
3. Clusters = co-locations of R'
4. Rank clusters in descending mass

In this study, we used 150 m as the aperture (i.e., neighborhood). The spatial decay function, defined above, was used to weigh the computation of the local field mean density. In the case of Algorithm 1, we used a mean-shift subalgorithm that is weighted by $s(p)$ about each drifting response $r' \in R'$. Points were continually evaluated and shifted until the total Earth surface movement, haversine in our case, of all the points was less than epsilon = 1 m. At this stage, the points that had converged under the modes were then mapped into their respective clusters and labeled. The cluster's score was computed as the area under its amplified spatial density. The cluster score was then used to rank (highest to lowest score) all clusters in an AoI to determine the order for subsequent human review of candidate SAM site locations (Sec. 5.4). In this study, single-detection clusters (i.e., spatial outliers) were filtered out from the results presented for subsequent human review. Therefore, DCNN detections that presented as an isolated spike were not presented for the human review described in Sec. 5.4. This operator characteristic could be adjusted along with other parameters related to DCNN confidence sensitivity, but this was not done as part of this study.

Table 11 provides a detailed summary of the DCNN processing and subsequent spatial clustering results by AoI. Approximately 19.7M image chips (75% overlap) from ~66 K tiles were processed by the China ResNet-101 DCNN covering an area of 88,600 km^2 (94.5% of the study land area of ~94,000 km^2; see Table 5). Out of these 19.7 M image chips, the DCNN processing produced ~9300 chip detections at a 99% confidence threshold. The spatial clustering of the DCNN chip detections generated ~2100 ranked clusters, which is a ~78% reduction relative to the ~9300 input chip detections. Assuming that each cluster location represents a local area of roughly one image chip (256 m × 256 m), then the DCNN processing followed by the spatial clustering effectively reduced the human visual BAS space to a total area of only 136 km^2. Thus, the DCNN-assisted search space is only 0.15% of the total processed area, representing a ~660X reduction in the amount of imagery submitted for subsequent human review.

5.4 *DCNN-Assisted Rapid Human Review of Candidate SAM Sites*

As discussed in Sec. 1, visual BAS using large high-resolution image archives is a labor-intensive process. Consequently, a number of applications have been developed to facilitate rapid BAS for a variety of domains. These systems often produce a ranked set of image chips and/or localized image areas for subsequent human visual inspection. In addition, research, such as Ref. 23, has demonstrated that rapid serial presentation of image chips can achieve drastic acceleration in image review tasks.

For example, GeoIRIS[24] produced a ranked list of visually similar image chips from a large corpus of high-resolution EO imagery for serial human review. GeoIRIS also allowed a user to customize the relative importance of hand-crafted feature spaces to improve the relevance of the retrieved image chips. In another example, GeoCDX[25] processed temporal image pairs for automated change detection and then presented change-intensity ranked image chips also for serial human review. Additionally, GeoCDX augmented the change detection with unsupervised machine learning techniques[26] to present tiers of visually similar change clusters to the user for rapid identification of relevant change types based on application domain requirements. Griparis et al.[27] proposed a method for visual information mining of imagery by projecting image chips into a 3-D reduced feature space to facilitate visual cluster analysis. Also, the

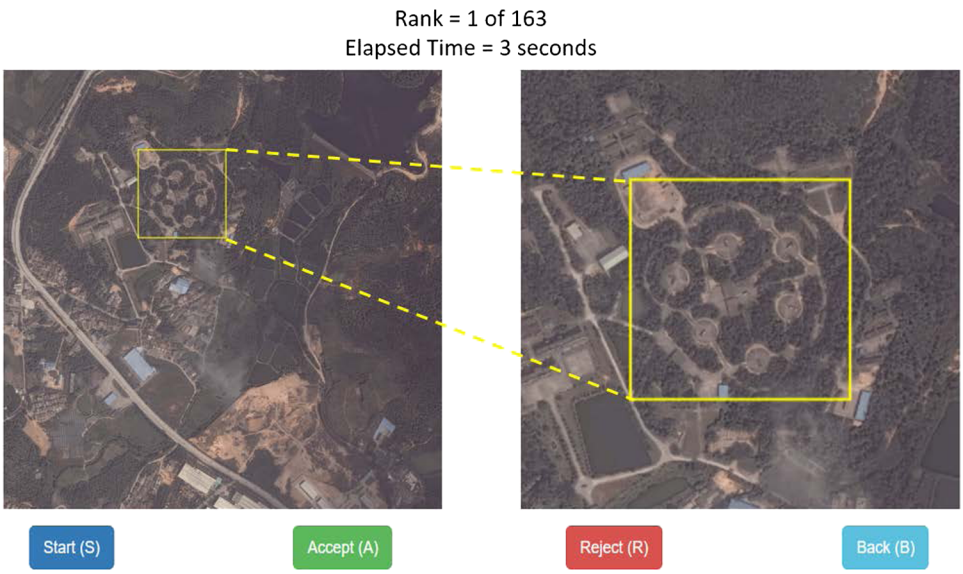


Fig. 13 WUI used for rapid serial presentation and human review of ranked candidate SAM sites generated from spatial clustering of DCNN image chip detection results. The left hand side of the WUI presents the contextual view of the candidate SAM site and surrounding area while the right hand side of the WUI presents a zoomed in view of the candidate SAM site for more detailed human inspection of small-scale objects and features. The actual result shown here is for the top-ranked candidate SAM site for AoI 1, as shown in Fig. 12. See Sec. 5.4 for more detailed discussion of the WUI functions.

semantics-enabled spatial image information mining in Ref. 28 is a contemporary work that facilitates a number of image chip retrieval methods from a large image corpus.

Based on this prior work, we implemented a simple but efficient web user interface (WUI) to facilitate rapid serial human review of the candidate SAM sites[29] in each AoI. Candidate SAM sites for a given AoI were generated by spatial clustering of the China ResNet-101 DCNN output results and subsequent ranking of each spatial cluster, as discussed in Sec. 5.3. Figure 13 shows the WUI presentation for the top-ranked candidate SAM site (cluster rank 1; see Fig. 12) from AoI 1. The left-hand side of the WUI presents a 1280 m × 1280 m AoI tile containing the candidate SAM site area that is highlighted by a 256 m × 256 m bounding box. The right-hand side of the WUI presents a 512 m × 512 m view of the local area surrounding the cluster center of the candidate SAM site. In addition, a hot-key function allows the right-hand side presentation to toggle back and forth between the default 512 m × 512 m view and a zoomed in 256 m × 256 m view to allow closer visual inspection of small-scale objects and features. In the context of this application, the toggle zoom feature was helpful for inspecting candidate SAM site launch pads for the presence of missiles and/or TELs.

After starting the application for a given AoI, the human operator is allowed to accept/reject a candidate SAM site using hot keys (A and R) and/or mouse clicks of WUI buttons. After making an accept/reject decision, the WUI will autoload the next candidate SAM site in the ranked list until all ranked items have been reviewed. In addition, the operator is allowed to traverse backward in the ranked list of candidate SAM sites using a hot key (B) or mouse click of a WUI button. This allows the human operator to quickly and easily correct obvious mistakes. This functionality is important because during rapid serial review, it is very common for a human operator to be predisposed to repetitive keystrokes and/or mouse clicks that can easily introduce errors even for expert users.

Finally, a notification area at the top of the WUI displays the current rank out of the total number of ranked candidate SAM sites in the AoI along with the elapsed time in seconds spent on the human review after the application is started. All accept/reject decisions for each candidate SAM site are recorded to a log file along with user identification and total elapsed time for the human review for all candidate SAM sites in the AoI.

Two groups of IAs were selected to perform the DCNN-assisted BAS over the China study area. As mentioned in Sec. 4.2, we use the term IA to convey the functional role of someon, who performs a visual search of high-resolution EO imagery for SAM sites. The term is not intended to convey or represent that these individuals were professionals whose primary work role was to perform imagery analysis.

The first IA group consisted of four individuals (IA4, IA5, IA6, IA7), who had no prior work experience whatsoever analyzing high-resolution EO imagery. Three were undergraduate students enrolled at the University of Missouri and the fourth was a recent graduate, who had taken an IT staff position with the university. With respect to the IAs that performed the visual BAS presented in Sec. 4, we characterize the experience level of IA4, IA5, IA6, and IA7 as novice.

The second IA group consisted of two individuals (IA8, IA9), who had extensive experience analyzing high-resolution EO imagery. IA8 was the second author of this paper who, over a 15+ year period, had accumulated at least 1000 h of academic R&D type experience working with high-resolution EO imagery but without specific formal training. IA9 was an undergraduate student enrolled at the University of Missouri, who had recent active duty military service, where he had at least 1000 h of high-resolution EO imagery analysis as part of his military service. However, IA9 also did not have any formal training and had no prior experience analyzing air defense related features like SAM sites. With respect to the IAs that performed the visual BAS presented in Sec. 4, we characterize the experience level of IA8 and IA9 as expert.

The four novice IAs and IA9 were given the same overview briefing and hardcopy reference materials that were given to the intermediate IAs (Sec. 4.2) to assist them in identification of SAM sites in high-resolution EO imagery. All these materials were prepared by IA8 based on open-source research and study of the open-source SAM site compilation used as the source for producing the curated worldwide SAM site dataset (Sec. 2). Finally, each IA was given a short demonstration of the WUI and of small amount of time (15 min or less) to become familiar with the WUI operation described here for the DCNN-assisted BAS. The WUI familiarization was done using a 1 deg $\times$1 deg test AoI from China that was outside of the study area shown in Fig. 7.

The DCNN-assisted BAS was performed independently by the novice and expert IAs. The WUI did not provide any feedback/indications as to the accuracy of the IAs accept/reject decisions during performance of the DCNN-assisted BAS over the 10 AoIs. No additional human instruction or guidance was provided to the IAs during their work, and the IAs did not discuss any aspect of their BAS with anyone else during their performance of the search. Finally, the IAs did not have to record their time spent performing the BAS since that was automatically logged by the WUI.

5.5 *DCNN-Assisted Broad Area Search Results*

After the DCNN-assisted BAS was completed for the entire study area, the WUI log files were compared with a master file containing the 14 *a priori* (Table 6) and 2 *a posteriori* (Fig. 8) SAM site locations to determine the number of TP, FP, and FN detections for each IA's DCNN-assisted human review of the candidate SAM sites. Table 12 provides the performance data for the DCNN-assisted BAS and detection of SAM sites by the four novice and two expert IAs for each AoI in the study area. Five of the six IAs correctly identified the two new SAM sites (*a posteriori*) in AoI 2 and AoI 7 (Fig. 8). No additional new SAM site locations were detected by the IAs in the DCNN-assisted BAS. Thus, we are extremely confident that the total number of SAM sites in the study area is $P = 16$ given the consistency in the total number of observed SAM sites between the visual and DCNN-assisted BAS. This is especially true given that both the visual and DCNN-assisted BAS were performed by multiples IAs (high redundancy).

It should be noted that all novice and expert IAs had the same FN in AoI 9 because this particular SAM site was not presented by the WUI as a candidate SAM site, i.e., this FN was not due to human error. Investigation for the cause of this determined that the China ResNet-101 DCNN did properly detect this SAM site since its detection confidence of 99.999% was well above the 99% confidence threshold cutoff. However, there were no other

Table 12 IA performance data for DCNN-assisted BAS and detection of SAM sites for the southeastern China study. See Sec. 4.3 for description of the reported values.

	Analyst	IA4				IA5				IA6				IA7				IA8				IA9			
AoI #	AoI	TP	FN	FP	NSS	TP	FN	FP	NSS	TP	FN	FP	NSS	TP	FN	FP	NSS	TP	FN	FP	NSS	TP	FN	FP	NSS
1	23°N113°E	4	0	0		4	0	0		4	0	0		4	0	0		4	0	0		4	0	0	
2	23°N114°E	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
3	23°N115°E	0	0	1		0	0	1		0	0	1		0	0	0		0	0	0		0	0	0	
4	23°N116°E	1	0	0		1	0	0		1	0	0		1	0	0		1	0	0		1	0	0	
5	24°N115°E	1	0	0		1	0	0		1	0	0		1	0	0		1	0	0		1	0	0	
6	24°N116°E	0	0	0		0	0	0		0	0	0		0	0	0		0	0	0		0	0	0	
7	24°N117°E	4	0	0	1	4	0	0	1	2	2	0	0	3	1	0	1	4	0	0	1	4	0	0	1
8	24°N118°E	1	0	0		1	0	0		1	0	0		1	0	0		1	0	0		1	0	0	
9	25°N118°E	0	1	0		0	1	0		0	1	0		0	1	0		0	1	0		0	1	0	
10	25°N119°E	3	0	0		3	0	0		1	2	0		3	0	0		3	0	0		3	0	0	
	Total	15	1	1	2	15	1	1	2	11	5	1	1	14	2	0	2	15	1	0	2	15	1	0	2

Table 13 AoI search times (min) for novice IAs (IA4, IA5, IA6, IA7) and expert IAs (IA8, IA9) that performed DCNN-assisted BAS over the southeastern China study area.

		DCNN-assisted BAS time (min)							
AoI #	AoI	IA4	IA5	IA6	IA7	AVG	IA8	IA9	AVG
1	23°N113°E	3.4	5.6	6.8	5.3	5.3	2.1	4.4	3.3
2	23°N114°E	3.2	8.5	6.0	3.8	5.4	2.5	6.8	4.7
3	23°N115°E	3.6	4.0	3.8	3.5	3.7	1.3	5.4	3.4
4	23°N116°E	9.1	10.6	10.0	8.9	9.7	5.9	17.9	11.9
5	24°N115°E	2.5	3.4	3.1	2.7	2.9	1.7	3.9	2.8
6	24°N116°E	2.8	3.0	2.5	3.5	3.0	2.3	5.4	3.9
7	24°N117°E	2.6	4.2	4.1	3.0	3.5	3.1	5.5	4.3
8	24°N118°E	3.2	4.3	3.3	4.7	3.9	3.7	3.6	3.7
9	25°N118°E	2.0	3.3	1.6	2.0	2.2	1.9	2.7	2.3
10	25°N119°E	2.5	2.6	3.8	2.5	2.9	2.0	3.8	2.9
	Total	34.9	49.5	45.0	39.9	42.3	26.5	59.4	43.0

DCNN chip detections adjacent to this correct detection, and, as a result, the spatial clustering algorithm described in Sec. 5.3 treated this as spatial outlier and removed this point from the list of candidate SAM sites presented for human review.

Table 13 provides the IA search time (labor minutes) for each of the 10 AoIs and the total search time for the study area. The total search time for all 10 AoIs varied between 25 to 60 min for the IAs with an average total search time of ~43 min. We evaluated the IAs DCNN-assisted BAS performance using the same three statistical measures (TPR, PPV, F1) used in Sec. 4.3 to evaluate the IAs visual BAS performance. Table 14 provides these summary statistics for each IA's DCNN-assisted BAS performance over the entire study area. The summary statistics are partitioned into two IA groups (novice and expert) due to the vast difference in experience levels. Unsurprisingly, the expert IAs (IA8, IA9) had the best overall statistical performance with TPR of 93.8%, PPV of 100%, and F1 96.8%. In addition, the 93.8% TPR result is the best human performance that could have been achieved given that the single FN from AoI 9 was due to removal of this SAM site by the spatial clustering algorithm.

Table 14 Overall statistical performance for novice IAs (IA4, IA5, IA6, IA7) and expert IAs (IA8, IA9) that performed DCNN-assisted BAS over the southeastern China study area.

Analyst	TPR (%)	PPV (%)	F1 (%)
IA4	93.8	93.8	93.8
IA5	93.8	93.8	93.8
IA6	68.8	91.7	78.6
IA7	87.5	100	93.3
Avg	85.9	94.8	89.9
IA8	93.8	100	96.8
IA9	93.8	100	96.8
AVG	93.8	100	96.8

The combined statistical performance of the four novice IAs (IA4, IA5, IA6, IA7) was TPR of 85.9%, PPV of 94.8%, and F1 of 89.9%, and this is 5% to 8% lower than the expert IA performance across the three measures. However, comparing the overall performance of the novice IAs from the DCNN-assisted BAS to that of the intermediate IAs that performed the visual BAS (Table 9, Sec. 4.3), we see nearly identical statistical performance, where all three statistical measures differed by <0.5% between the novice and intermediate IA groups. Notably, the F1 score was ~90% for both IA groups with only a 0.1% difference between the aggregate F1 scores of the two IA groups. Recall from Sec. 4.3 that the intermediate IAs had an average of ~700 h of prior experience analyzing high-resolution EO imagery for GIS cartographic feature extraction, analysis, and attribution. Thus, the novice IAs performing the DCNN-assisted BAS were able to achieve the same overall statistical performance as the intermediate IAs performing the visual BAS, even though they had no prior experience analyzing high-resolution EO imagery. This is a very important finding from this study.

The average visual BAS time for the intermediate IAs was 59.9 h (Table 8) covering a land area of ~93,760 km^2 (Table 5). The average DCNN-assisted BAS time for the novice IAs was 42.3 min (Table 13) for a DCNN-processed land area of ~88,640 km^2 (Table 11). The difference of ~5120 km^2 between the visual BAS area and the DCNN-assisted BAS land area coverage was due to the exclusion of downloaded source image tiles, located in mostly rural, low population density areas, where the nominal was GSD >= 2 m (Sec. 5.2). Therefore, we estimated the amount of extra time the intermediate IAs spent performing the visual BAS over the additional land area by overlaying the excluded DCNN source image tiles on the 4×4 search grid (0.25 deg×0.25 deg) used by the intermediate IAs for recording their visual search times (Sec. 4.2). Then, an area-weighted search time differential was computed over the affected AoIs (6, 7, 9) and found to be 2.7 h. Subtracting this from the average visual BAS time (59.9 h) yields an adjusted visual BAS time of 57.2 hours. Thus, the DCNN-assisted BAS time was $(57.2 \times 60)/42.3 = 81X$ faster than the adjusted visual BAS time for an equivalent land area of ~88,640 km^2.

6 Summary and Future Work

In this study, a comprehensive worldwide dataset of high-resolution EO image chips was produced for ~2200 SAM sites using an enterprise geospatial data curation framework. Four DCNN were trained using the worldwide SAM site dataset using a combination of data augmentation ($144X$) and transfer learning. The statistical performance of the four DCNN was then evaluated for automated SAM site detection using the curated worldwide dataset. A ResNet-101 DCNN was found to have the best overall performance with a 96.8% average accuracy from a fivefold cross validation.

Next, a traditional human visual BAS was performed to locate SAM sites over a study area of ~94,000 km^2 along the coast of southeastern China. The human performance was measured statistically using ground truth SAM site locations and also by time spent (labor hours) performing the visual search. The results show that three IAs with an intermediate level of experience were able to complete the visual BAS in an average time of ~60 h with an average F1 score of 90%.

DCNN were then trained and tested using a China-only subset of the worldwide SAM site dataset that contained less than 100 positive training examples. An enhanced data augmentation strategy that included jitter/shift of the original image chips increased the China training and testing data by ~$9500X$. A ResNet-101 DCNN was again found to have the best overall performance for the China-only SAM site dataset with a 98.2% average accuracy from a fivefold cross validation.

The China DCNN result demonstrates that a robust DCNN object/target classifier can be achieved for modest training dataset sizes (<100) using data augmentation and transfer learning in lieu of much larger amounts of labeled training data normally associated with DCNN. Moreover, the China DCNN obtained much better statistical performance than the worldwide DCNN for SAM site detection even though it used >$20X$ fewer training examples. Thus, using large labeled training datasets does not necessarily guarantee better performance when using DCNN in specific applications.

We attribute the better statistical performance of the China DCNN to the fact that the China SAM site training/testing data had significantly less heterogeneity compared to the worldwide SAM site dataset. This suggests that constructing region and/or country-specific DCNN for object/target recognition may be an effective strategy when there are significant geographic variations in the visual presentation of various object/target classes.

The China ResNet-101 DCNN was then used to process ~19.6 M image chips (256 m × 256 m) over the China study area, which produced ~9300 chip detections (99% confidence threshold cutoff) of possible SAM site locations. The DCNN chip detections were then postprocessed using a spatial clustering algorithm to produce a ranked list of ~2100 candidate SAM site locations. The combination of DCNN processing and spatial clustering effectively reduced the BAS space by ~660*X* to 0.15% of the DCNN-processed land area.

An efficient web interface was developed to facilitate rapid serial human review of the candidate SAM sites in the China study area. Two groups of IAs (novice and expert) then performed a DCNN-assisted BAS over the China study area. Two IAs with an expert level of experience were able to complete the DCNN-assisted BAS in an average time of ~43 min with an average F1 score of 96.8%. Four IAs with only a novice level of experience were able to complete the DCNN-assisted BAS in an average time of ~42 min with an average F1 score of 89.9%.

The novice IAs who had no prior experience analyzing high-resolution EO imagery were able to achieve the same overall statistical performance (~90% F1) for the DCNN-assisted BAS as the intermediate IAs that performed the visual BAS. More importantly, the DCNN-assisted BAS time was ~81*X* faster than the visual BAS time for an equivalent land area of ~88,640 km^2.

Thus, the overall results from this study clearly demonstrate that: (1) robust DCNN object/target classifiers can be developed using only modest amounts of training data (<100 samples) and (2) that DCNN processing can effectively assist humans in visual searches for objects/targets of interest in high-resolution EO imagery over very large areas of the Earth's surface to achieve dramatic labor/time savings at comparable statistical accuracy.

Currently, we are actively working to create a variety of curated image training datasets for high-resolution satellite EO imagery using the enterprise curation framework described in this paper. For example, in the near future, we will produce several curated datasets having 50 to 100 object/target classes with 100 to 1000 training samples per class. This will allow us to evaluate the suitability of extending our DCNN detection approach to a diverse set of object/target classes. Nevertheless, this will present additional challenges in the post-DCNN inference processing. For example, algorithms for spatial aggregation of DCNN detection responses would need to be extended to operate on spatially clustered vectors.

Another important research direction will be to develop active learning mechanisms, where both TP and FP detections presented in the high-ranked response clusters can be fed back into the DCNN object recognition system. This would allow the DCNN object detection model to be continually fine-tuned, similar to the original transfer learning approach, to improve the discrimination between TPs and TNs (i.e., SAM site and not-SAM site) for example.

Acknowledgments

The authors wish to express our "deep" appreciation to the DigitalGlobe Foundation for providing access to the DigitalGlobe Maps API platform. This research study simply would not have been possible without access to the worldwide high-resolution EO image basemap provided by the DigitalGlobe Maps API. In addition, we are pleased to acknowledge the contributions of M. Blackwood, J. Bongard, J. Carmichael, L. Darrough, J.P. Davis, J.O. Davis, A. Kemp, M. Munir, and M. Thunhorst who all made important contributions to this study.

References

1. L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: a technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.* **4**(2), 22–40 (2016).
2. U.S.A.F., "MSTAR overview," http://tinyurl.com/pc8nh3s (March 2017).

3. P. Vasuki and S. M. M. Roomi, "Automatic target classification of manmade objects in synthetic aperture radar images using Gabor wavelet and neural network," *J. Appl. Remote Sens.* **7**(1), 073592 (2013).
4. H. Anglberger and T. Kempf, "A simulation-based approach towards automatic target recognition of high resolution space borne radar signatures," *Proc. SPIE* **10004**, 1000413 (2016).
5. B. J. Schachter, "Target classification strategies," *Proc. SPIE* **9476**, 947602 (2015).
6. K. El-Darymli et al., "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access* **4**, 6014–6058 (2016).
7. X. Chen et al., "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.* **11**(10), 1797–1801 (2014).
8. P. Zhang et al., "Airport detection from remote sensing images using transferable convolutional neural networks," in *Int. Joint Conf. on Neural Networks (IJCNN)*, pp. 2590–2595, IEEE (2016).
9. Y. Cao, X. Niu, and Y. Dou, "Region-based convolutional neural networks for object detection in very high resolution remote sensing images," in *12th Int. Conf. on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 548–554, IEEE (2016).
10. Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. of the 18th SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems*, pp. 270–279, ACM (2010).
11. F. P. Luus et al., "Multiview deep learning for land-use classification," *IEEE Geosci. Remote Sens. Lett.* **12**(12), 2448–2452 (2015).
12. G. Cheng et al., "Scene classification of high resolution remote sensing images using convolutional neural networks," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, pp. 767–770, IEEE (2016).
13. C. Szegedy et al., "Going deeper with convolutions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, 1–9 (2015).
14. G. J. Scott et al., "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.* **14**(4), 549–553 (2017).
15. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. of the 25th Int. Conf. on Neural Information Processing Systems (NIPS)*, pp. 1097–1105, Curran Associates, Inc. (2012).
16. G. J. Scott et al., "Fusion of deep convolutional neural networks for land cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.* **14**, 1638–1642 (2017).
17. DigitalGlobe, "DigitalGlobe maps API," https://platform.digitalglobe.com/maps-api/ (March 2017).
18. DigitalGlobe Foundation, "DigitalGlobe Foundation homepage," http://www.digitalglobefoundation.org/ (March 2017).
19. S. O'Connor, "IMINT and analysis blog," http://geimint.blogspot.com/ (8 August 2015).
20. Y. Jia et al., "Caffe: convolutional architecture for fast feature embedding," in *Proc. of the 22nd ACM Int. Conf. on Multimedia*, ACM, 675–678 (2014).
21. K. He et al., "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, 770–778 (2016).
22. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
23. N. Bigdely-Shamlo et al., "Brain activity-based image classification from rapid serial visual presentation," *IEEE Trans. Neural Syst. Rehabil. Eng.* **16**(5), 432–441 (2008).
24. C.-R. Shyu et al., "GeoIRIS: geospatial information retrieval and indexing system-content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.* **45**(4), 839–852 (2007).
25. M. N. Klaric et al., "GeoCDX: an automated change detection and exploitation system for high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.* **51**(4), 2067–2086 (2013).
26. O. Sjahputera et al., "Clustering of detected changes in high-resolution satellite imagery using a stabilized competitive agglomeration algorithm," *IEEE Trans. Geosci. Remote Sens.* **49**(12), 4687–4703 (2011).

27. A. Griparis, D. Faur, and M. Datcu, "Dimensionality reduction for visual data mining of earth observation archives," *IEEE Geosci. Remote Sens. Lett.* **13**, 1701–1705 (2016).
28. K. R. Kurte et al., "Semantics-enabled framework for spatial image information mining of linked earth observation data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**, 29–44 (2017).
29. Center for Geospatial Intelligence, "China SAM site search and detection challenge," https://ssc.cgi.missouri.edu/ (2017).

Richard A. Marcum received his BS degree in applied mathematics from the University of Illinois Urbana-Champaign in 2010. He received his MS degree in mathematics from Northern Illinois University in 2013. He will obtain his MS degree in computer engineering at the University of Missouri Columbia in December 2017. His research interests are in the areas of image and signal processing, machine learning, and pattern recognition applied to remote sensing.

Curt H. Davis received his BS and PhD degrees in electrical engineering from the University of Kansas in 1988 and 1992, respectively. He is currently the Naka endowed professor of electrical engineering and computer science and the director of the Center for Geospatial Intelligence at the University of Missouri. His current research interests are focused on the development of automated methods (change detection, object recognition, etc.) for processing and exploitation of high-resolution imagery for a variety of application domains. In 2008, he was elected IEEE Fellow for his contributions to satellite remote sensing.

Grant J. Scott received his PhD degree in computer engineering and computer science from the University of Missouri in 2008. He is currently the director of the University of Missouri Data Science and Analytics Program, as well as an assistant research professor at the Center for Geospatial Intelligence at the University of Missouri. His current research interests include computer vision, high performance data intensive computing, and large-scale pattern databases. He also serves as the MU program manager for the National Geospatial-Intelligence Agency Program of Study in Data Science.

Tyler W. Nivin received his BS degree in computer science from the University of Missouri in 2016. He is currently pursuing his MS degree in computer science at the University of Missouri. His research interests include machine learning and high performance computing.