Datasets

Note: References here do not match reference numbers in the paper. The references are provided at the end.

Online content provided with paper "A Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community" by Ball, Anderson and Chan

Dataset	Description	References
15 Scenes	This dataset is an extension of 13 scene categories data set provided by Fei-Fei and Perona [1] and Oliva and Torralba [2]. This data set contains coast, forest, mountain, open country, highway, inside city, tall building, street, bedroom, kitchen, living room, office, suburb, industrial, and store. The average resolution is 300×250. Website: http://vision.stanford.edu/resources_links.html	[1],[2],[3]
19-Class	High resolution satellite dataset. 19 categories and each of them has 50 images, with a size of 600 × 600 pixels. 19-Class is composed of 19 classes of scenes, including airport, beach, bridge, commercial area, desert, farm- land, football field, forest, industrial area, meadow, mountain, park, parking, pond, port, etc. Website: http://dsp.whu.edu.cn/cn/staff/yw/HRSscene.html	[4],[5]
19-class satellite	Image datasets with semantic categories spatial envelope model. The database contains about 8100 pictures of environ- mental scenes so as to cover a large variety of outdoor places. Images were 256×256 pixels in size, in 256 gray levels. They come from the Corel stock photo library, pictures taken from a digital camera and images downloaded from the web. Website: https://stock.adobe.com	[2]
Atlantic Deep Sea	50 cold-water reefs with annotated images. Five coral and four non-coral classes.	[6]
Banja-Luka LU Public	606 RGB aerial images of size 128 by 128 pixels. This database was constructed from a part of Banja-Luka city, Bosnia and Herzegovina. Classes are houses, cemetery, industry, field, river, and trees.	[7]
Bern	This dataset contains two SAR images over Bern, Switzerland, in April and May 1999. Between these dates, the Aare river flooded parts of the cities of Thun and Bern, and flooded the Bern airport entirely.	[8]
BigBIRD	BigBIRD has per object: (1) 600 Kinect-style RGB-D images, (2) 600 high-resolution images, (3) accurate calibration information for every image, (4) segmented objects per image, and (5) full-object meshes. Website: http://rll.eecs.berkeley.edu/bigbird	[9]
BIWI	The BIWI RGBD-ID Dataset1 consists of video sequences of 50 different subjects, performing a certain routine of motions and walks in front of a Kinect. Website: http://robotics.dei.unipd.it/reid	[10]

Dataset	Description	References
Botswana	The dataset was collected by Hyperion sensor on EO-1 over the Okavango Delta, Botswana. This dataset contains 1496 × 256 pixels with 30-m spatial resolution, and 242 bands covering the 400–2500 nm portion of the spectrum in 10 nm windows. The data consists of observations from 14 classes. Website: http://aviris.jpl.nasa.gov/data/free_data.html	[11]
Brazilian Coffee Scenes	SPOT sensor in 2005 over four counties in the State of Minas Gerais, Brazil: Arceburgo, Guaran´esia, Guaxup´e, and Monte Santo. Challenging dataset. R-G-NIR. Website: www.patreo.dcc.ufmg.br/downloads/2ultisens-coffee-dataset/	[12]
Caltech	The training data has six training sets, each with 6-13 one-minute long sequence files, along with all annotation information (see the paper for details). The testing data consists of five sets. Website: https://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/	[13]
Caltech Pedestrian	A large-scale, challenging dataset with about 10 hours of 640 × 480, 30fps video, acquired from a vehicle driving through regular traffic in an urban environment under good weather conditions. There are 350,000 pedestrian bounding boxes (BB). Occlusions and temporal correspondences are also annotated. Website: https://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/	[14]
Caltech1999	The Caltech1999 dataset contains 126 images of cars from the rear. Approximate scale normalization. JPEG format (896 X 592). These images were taken in the Caltech parking lots. Website: http://www.vision.caltech.edu/archive.html	[15]
Caltrans Performance Measurement System database	Highway information using 39,000+ sensors along California freeways. Website: http://pems.dot.ca.gov/	[16]
CelebA	CelebA labels images selected from two challenging face datasets, Celeb-Faces (reference [26] in [17]) and LFW(reference [12] in [17]). CelebA contains ten thousand identities, each of which has twenty images. There are 200,000 images, each annotated with forty face attributes and five key points by a professional labeling company. Website: http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html	[17]
CIFAR-10	The CIFAR-10 dataset consists of 60,000 32 X 32 color images (50,000 for training and 10,000 for testing) in 10 classes of generic objects, with 6000 images per class. The images were extracted from the world wide web for images using WordNet database nouns. Website: https://www.cs.toronto.edu/~kriz/cifar.html	[18]
COIL20	The Columbia Object Image Library (COIL20) is a multi-class data set with 1440 gray-scale images of 20 objects. Each pattern is a 32×32 gray	[19]

Dataset	Description	References
	scale image of one object taken from a specific view. The COIL20(B) data set is a binary classification task obtained from COIL20. Website: http://www.cs.uchicago.edu/vikass/research.html	
Copernicus SENTINEL	MSI dataset. Images have a radiometric resolution of 12 bits/pixel. Instrument has 4 bands at 10m spatial resolution, 6 at 20m spatial resolution, and 3 at 60m spatial resolution. Website: https://scihub.copernicus.eu/dhus	[20]
Copperas Cove HYDICE	This data set was captured by the hyperspectral digital imagery collection experiment (HYDICE) sensor in October 1995. The image area was located at Copperas Cove, near Fort Hood, Texas, USA. The urban scene is 307 pixels × 307 pixels, with a spatial resolution of 2 m per pixel, with 210 bands and a spectral resolution of 10 nm. Website: http://www.agc.army.mil/	[21]
CUHK	The CUHK occlusion pedestrian dataset mainly includes images with occluded pedestrians. This dataset contains 1063 images. Each image contains at least one occluded pedestrian. Each pedestrian is labeled with a bounding box and a tag indicating whether the pedestrian is occluded or not. Website: http://www.ee.cuhk.edu.hk/~xgwang/CUHK_pedestrian.html	[22]
Cuprite	A hyperspectral dataset over Cuprite, NV. Many different minerals are in the scene. Website: http://aviris.jpl.nasa.gov/data/free_data.html	[23]
Daimler	Pedestrian detection dataset. Website: http://www.gavrila.net/Datasets/Daimler_Pedestrian_Benchmark_D/3altech_pedestrian_benchmark_d.html	[24]
DC Mall	The Washington DC image was collected by the HYDICE sensor over a mall in Washington DC. It has 1280 × 307 pixels with 210 (191 usable) spectral bands in the range of 0.4–2.4 μm. The spatial resolution is 2 m/pixel. Website: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html	[25]
ETH	Stereo vision pedestrian detection dataset. Website: http://www.vision.ee.ethz.ch/~aess/dataset/	[26],[27]
FC1	The Flightline C1 data is a 12-band multispectral image taken over Tippecanoe County, Indiana by the M7 scanner. The image is 949 × 220 pixels and contains ten agricultural classes. A ground survey of 70,847 is provided. Website: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html	[28]
Fish Recognition Ground-Truth dataset	This underwater live fish dataset is acquired from a live video dataset captured from the open sea. There are totally 27,370 verified fish images of 23 clusters and each cluster is presented by a representative species. The fish species are manually labeled by following instructions from marine biologists. Website: http://groups.inf.ed.ac.uk/f4k/	[29]

Dataset	Description	References
MNIST	The MNIST database of images of handwritten digits (0-9) is a standard benchmark data set used in the machine learning community. It has a training set of 60,000 examples (approximately 6000 examples per digit), and a test set of 10,000 examples. The dimensionality of images is 28x28. Website: http://yann.lecun.com/exdb/mnist/	[30]
G50C	The G50C is a binary classification data set of which each class is generated by a 50-dimensional multivariate Gaussian distribution. This classification problem is explicitly designed so that the true Bayes error is 5%. Website: http://vikas.sindhwani.org/datasets/ssl/	[19]
Graz	Images containing multiple objects. Website: www.emt.tugraz.at/~pinz/data/	[31]
GTSRB	Traffic signs images. More than 50,000 images. Has effects from distance, illumination, weather conditions, partial occlusions, rotations. 43 classes. Website: http://benchmark.ini.rub.de/?section=gtsrb&subsection=news	[32]
HiRISE	This dataset contains imagery from the High Resolution Imaging Science Experiment (HiRISE) camera on board the Mars Reconnaissance Orbiter (MRO).	[33]
IAS-LAB	The IAS-Lab RGBD-ID dataset consists of 33 sequences of 11 people acquired using the OpenNI SDK and the NST tracker. For every subject, the Training and Testing sequences were collected in different rooms, with strong illumination changes caused by the different auto-exposure level of the Kinect in the two rooms Website: http://www.lorisbazzani.info/code-datasets/caviar4reid	[34]
iCoseg	The iCoseg dataset contains 38 groups with 17 images/group on average (total 643 images) and pixel-wise hand-annotated ground truth. Website: http://amp.ece.cornell.edu/projects/touch-coseg/	[35]
IEEE GRSS 2008 Data Fusion Contest	The data set consisted of airborne data from the reflective optics system imaging spectrometer (ROSIS-03) optical sensor. The flight over the city of Pavia, Italy has 102 usable bands with spectral coverage 0.43 to 0.86 µm and spatial resolution is 1.3 m. Classes are buildings, roads, shadows, vegetation, and water. Website: http://tlclab.unipv.it/dftc/home.do	[36]
IEEE GRSS 2013 Data Fusion Contest	National Science Foundation Center for Airborne Laser Mapping collected a dataset in the summer of 2012 over the University of Houston and the neighboring urban area. The dataset has a spatial resolution of 2.5m. The imagery has144 spectral bands ranging from 380 to 1050 nm. The ground truth is provided by 2013 IEEE GRSS Data Fusion Contest. Website: http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/	[37]

Dataset	Description	References
IEEE GRSS 2015 Data Fusion Contest	The 2015 Contest was focused on multiresolution and 5 multisensory fusion at extremely high spatial resolution. A 5-cm resolution color RGB orthophoto and a LiDAR dataset, for which both the raw 3D point cloud with a density of 65 pts/m² and a digital surface model with a point spacing of 10 cm, were distributed to the community. These data were collected using an airborne platform over the harbor and urban area of Zeebruges, Belgium. Website: http://www.grss-ieee.org/community/technical-committees/data-fusion/2015-ieee-grss-data-fusion-contest/	[38]
IEEE GRSS 2016 Data Fusion Contest	The imaging data were acquired on March, 31, and May, 30, 2015, over Vancouver, Canada from the DEIMOS-2 satellite. DEIMOS-2 operates from a Sun-synchronous orbit at a mean altitude of 620km. The spacecraft design is based on an agile platform for fast and precise off-nadir imaging (up to +/-30° over nominal scenarios and up to +/-45° in emergency cases), and carries a push-broom very high resolution camera with 5 spectral channels (1 panchromatic, 4 multispectral with red, green, blue and NIR bands). For each date, four images are provided: panchromatic images at 1 m resolution and multispectral product (R, G, B, NIR) at 4 m resolution, both at levels 1B (a calibrated and radiometrically corrected product, not resampled; with the geometric information contained in a RPC separated file) and 1C (a calibrated and radiometrically corrected product, manually orthorectified and resampled to a map grid; the geometric information is contained in the GeoTIFF tags.) Level 1C images cover exactly the same ground area for both dates. The full color, UHD video was acquired over Vancouver on July, 2nd, 2015. The High-Resolution camera, Iris, is installed on the Zvezda module of the International Space Station (ISS). Iris uses a CMOS detector to capture RGB videos with a Ground Sample Distance as fine as 1-meter, at 3 frames per second. Iris videos use image frames that have been fully orthorectified and resampled to 1-meter. Frame format is 3840×2160 pixels and cover approximately 3.8km × 2.1km. Website: http://www.grss-ieee.org/community/ technical-committees/data-fusion/	[39]
IIT PAVIS	This dataset is composed by four different groups of data. The first “Collaborative” group has been obtained by recording 79 people with a frontal view, walking slowly, avoiding occlusions and with stretched arms. This happened in an indoor scenario, where the people were at least 2 meters away from the camera. This scenario represents a collaborative setting, the only one that we considered in these experiments. The second (“Walking”) and third (“Walking2”) groups of data are composed by frontal recordings of the same 79 people walking normally while entering the lab where they normally work. The fourth group (“Back- wards”) is a back view recording of the people walking away from the lab. Since all the acquisitions have been performed in different days, there is no guarantee that visual aspects like clothing or accessories will be kept constant.	[40]

Dataset	Description	References
	Website: https://www.iit.it/research/lines/pattern-analysis-and-computer-vision/pavis-datasets/534-rgb-d-person-re-identification-dataset	
ILSVRC 2013 ImageNet	This is the 2013 edition of the ILSVRC dataset. There are 12,125 images for training (9877 of them contain people, for a total of 17,728 instances), 20,121 images for validation (5,756 of them contain people, for a total of 12823 instances) and 40,152 images for testing. There is significant variability in pose and appearance, in part due to interaction with a variety of objects. In the validation set, people appear in the same image with 196 of the other labeled object categories. There are 200 classes described here: http://www.image-net.org/challenges/LSVRC/2013/browse-det-synsets. Website: http://image-net.org/challenges/LSVRC/2013/	[41]
ImageNet	Large scale training dataset for Deep Learning. Website: http://www.image-net.org/	[42]
Indian Pines	AVIRIS data over Indian Pines area. 145 X 145 pixels, 220 spectral bands. Website: https://purr.purdue.edu/publications/1947/1	[43]
ISPRS	ISPRS Semantic labeling dataset. Two state-of-the-art airborne image datasets, consisting of very high resolution true orthophoto (TOP) tiles and corresponding digital surface models (DSMs) derived from dense image matching techniques. Both areas cover urban scenes. While Vaihingen is a relatively small village with many detached buildings and small multi story buildings, Potsdam shows a typical historic city with large building blocks, narrow streets and dense settlement structure. Website: http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html	[44],[45]
JHUIT-50	The JHUIT-50 dataset contains 50 industrial objects and hand tools frequently used in mechanical operations. Objects are segmented from the background following the same procedures in the BigBIRD dataset. Fine-grained visual cues are often employed to distinguish these types of objects. Website:https://cirl.lcsr.jhu.edu/research/human-machine-collaborative-systems/visual-perception/jhu-visual-perception-datasets/	[46]
Kennedy Space Center	The NASA AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) instrument acquired data over the Kennedy Space Center (KSC), Florida, on March 23, 1996. AVIRIS acquires data in 224 bands of 10 nm width with center wavelengths from 400 – 2500 nm. Spatial resolution of 18 m and 176 bands. 13 classes representing the various land cover types that occur in this environment were defined for the site. Website: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Kennedy_Space_Center_.28KSC.29	[47]

Dataset	Description	References
LifeCLEF 2015 Plant Task Dataset	The dataset is composed of 113,205 pictures belonging to 41,794 observations of 1000 species of trees, herbs and ferns living in Western European regions. This data was collected by 8,960 distinct contributors of the Tela Botanica social network in the context of the Pl@ntNet project [30]. Each picture belongs to one and only one of the 7 types of views reported in the meta-data (entire plant, fruit, leaf, flower, stem, branch, leaf scan) and is associated with a single plant observation identifier allowing to link it with the other pictures of the same individual plant (observed the same day by the same person). Website: http://www.imageclef.org/lifeclef/2015/plant	[48]
LineMOD	The LineMOD dataset is a large dataset of 15 registered video sequences of 15 texture-less 3D objects. Each object was attached to the center of a planar board with markers attached to it, for model and image acquisition. The markers on the board provided the corresponding ground truth poses. Each object was reconstructed first using a set of images and the corresponding poses using a simple voxel based approach. After reconstruction, close range and far range 2D and 3D clutter was added to the scene and took the evaluation sequences. Each sequence contains more than 1,100 real images from different viewpoints. Our sequences provide uniformly distributed views from 0-360 degrees, 0-90 degree tilt rotation, 65 cm-115 cm scaling and ±45 degree in-plane rotation. Website: http://campar.in.tum.de/twiki/pub/Main/StefanHinterstoisser	[49]
LULC	The data set is composed of the following 21 Land Use/Land Classification (LULC) classes: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium-density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. Each class consists of 100 images measuring 256×256 pixels, with a pixel resolution of 30 cm in the red–green–blue color space. Website: http://vision.ucmerced.edu/datasets	[50]
Madonna	Madonna dataset was collected on the site of Villelongue, France, by the HYSPEX sensors. The data consists in 32224 pixels, with 160 spectral bands (from 400 to 1000 nm), and a spatial resolution of 50 cm per pixel. Twelve woody species are included. Some estimated mixed pixels are included.	[51]
Mass. Building, Mass. Roads	The Massachusetts Buildings Dataset consists of 151 aerial images of the Boston area, with each of the images being 1500 × 1500 pixels for an area of 2.25 square kilometers. The entire dataset covers roughly 340 square kilometers. The data is randomly split the data into a training set of 137 images, a test set of 10 images and a validation set of 4 images. The target maps were obtained by rasterizing building footprints obtained from the OpenStreetMap project. The dataset covers mostly urban and suburban areas and buildings of all sizes, including individual	[52]

Dataset	Description	References
	houses and garages, are included in the labels. Figures 6.1(a) and 6.1(b) show two representative regions from the Massachusetts Buildings dataset. The Massachusetts Roads Dataset consists of 1171 aerial images of the state of Massachusetts. As with the building data, each image is 1500×1500 pixels in size, covering an area of 2.25 square kilometers. We randomly split the data into a training set of 1108 images, a validation set of 14 images and a test set of 49 images. The dataset covers a wide variety of urban, suburban, and rural regions and covers an area of over 2600 square kilometers. With the test set alone covering over 110 square kilometers, this is by far the largest and most challenging aerial image labeling dataset. Website: https://www.cs.toronto.edu/~vmnih/data/	
Moorea	University of San Diego Labeled Coral dataset. About 2,000 images from three different years. Five coral classes and four non-coral classes. The Moorea Coral Reef Long Term Ecological Research project has been collecting image data from the island of Moorea (French Polynesia) since 2005. This project monitors six sites around the island, and four habitats at each site. Moorea Labeled Corals: The MLC dataset is a subset of the MCR LTER packaged for computer vision research. It contains 2055 images from three habitats: fringing reef, outer 10m and outer 17m, from 2008, 2009 and 2010. It also contains random point annotation (row, col, label) for the nine most abundant labels, four non coral labels: (1) Crustose Coralline Algae (CCA), (2) Turf algae, (3) Macroalgae and (4) Sand, and five coral genera: (5) Acropora, (6) Pavona, (7) Montipora, (8) Pocillopora, and (9) Porites. These nine classes account for 96% of the annotations and total to almost 400,000 points. There is a large variation in the number of samples from each class. Website: http://vision.ucsd.edu/data	[53]
MSRC	The MSRC dataset consists 240 manually segmented and annotated photographs which depict different objects in completely general positions, lighting conditions and viewpoints. The objects belong to the nine classes: building, grass, tree, cow, sky, airplane, face, car and bicycle. Website: http://research.microsoft.com/vision/cambridge/recognition/	[54]
MSTAR	The MSTAR dataset was collected in September of 1995 at the Redstone Arsenal, Huntsville, AL by the Sandia National Laboratory SAR sensor platform. The collection is part of the Moving and Stationary Target Acquisition and Recognition (MSTAR) program. SNL used an X-band SAR sensor in one foot resolution spotlight mode. Strip map mode was used to collect the clutter data. Various military targets are included in the dataset. Website: https://www.sdms.afrl.af.mil/index.php?collection=mstar&page=targets	[55]

Dataset	Description	References
Multi-View RGB-D	A large-scale, hierarchical multi-view object dataset collected using RGB-D cameras. The RGB-D Object Dataset contains visual and depth images of 300 physically distinct objects taken from multiple views. The chosen objects are commonly found in home and office environments, where personal robots are expected to operate. Objects are organized into a hierarchy taken from WordNet hypernym/hyponym relations and is a subset of the categories in ImageNet. Website: http://www.cs.washington.edu/rgbd-dataset	[56]
Naples 99	This data set consists of images from ERS2 synthetic aperture radar (SAR) and Landsat TM sensors acquired in 1999 over Naples Italy. The problem is binary classification: detection of urban versus nonurban areas. The available features were the seven LandSAT bands, two SAR backscattering intensities (0–35 days), and the SAR interferometric coherence.	[57]
NUS-WIDE	This dataset was randomly generated by crawling more than 300,000 images together with their tags from the image sharing site Flickr.com through its public API. The images whose sizes are too small or with inappropriate length-width ratios are removed. The remaining set contains 269,648 images with a total of 425,059 unique tags. Figure 1 illustrates the distribution of the frequencies of tags in the dataset. Website: http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm	[58]
NWPU VHR-10	This data set contains a total of 800 VHR optical remote sensing images, where 715 color images were acquired from Google Earth with the spatial resolution ranging from 0.5 to 2m, and 85 pansharpened color infrared images were acquired from Vaihingen data with a spatial resolution of 0.08 m. There are two image sets in this data set: a positive image set with 650 images with each image containing at least one target to be detected, and a negative image set with 150 images. From the positive image set, 757 airplanes, 302 ships, 655 storage tanks, 390 baseball diamonds, 524 tennis courts, 159 basketball courts, 163 ground track fields, 224 harbors, 124 bridges, and 477 vehicles were manually annotated with bounding boxes used for ground truth. Website: http://pan.baidu.com/s/1hqwzXeG	[59]
NYU-v2	The NYU depth dataset – version 2 – of Silberman and Fergus, composed of 407,024 couples of RGB images and depth images. The data was collected with a Kinect. Among these images, 1449 frames have been labeled. The object labels cover 894 categories. The dataset is provided with the original raw depth data that contain missing values, with inpainted depth images. Website: http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html	[60]
Ottawa	This dataset is a fusion of one airborne scanner and four car-mounted scanners, which provides higher density along the streets but lower density far from the streets (e.g., parking lots). This is a large dataset with about 100 million points and 1000 objects of interest. The data was collected by Neptec with one airborne scanner and four car-mounted TITAN scanners, facing left, right, forward- up, and forward-down. A	[61]

Dataset	Description	References
	single merged point cloud has 954 million points, each with a position, intensity, and color. The reported error in alignments between airborne and car-mounted scans is 0.05 meters, and the reported vertical accuracy is 0.04 meters.	
PASCAL VOC2012	Imagery with categories airplane, bicycle, bird, boat, etc. Training, testing and validation images are included. Website: http://host.robots.ox.ac.uk/pascal/VOC/voc2012/	[62]
Pavia Center Pavia University	ROSIS-3, 610 X 340 pixels. The number of spectral bands is 102 for Pavia Centre and 103 for Pavia University. Pavia Centre and Pavia University have 9 classes. Website: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Pavia_Centre_and_University	[47]
Prague Texture Segmentation Benchmark	This dataset consists of a set of synthetic mosaics of remote sensing images captured using the Advanced Land Imager (ALI). Each image is 512 x 512, has ten bands, a spatial resolution of 30 m. The dataset provides a common test bed on which different segmentation methods can be compared. In addition, it provides the ground-truth data which are needed for supervised segmentation techniques. Textures in these images are natural and boundaries are straight. Website: http://mosaic.utia.cas.cz	[63]
Princeton ModelNet	The dataset contains 151,128 3D CAD models belonging to 660 unique object categories. ModelNet was constructed by downloaded 3D CAD models from 3D Warehouse, and Yobi3D search engine indexing 261 CAD model websites. Next, common object categories from the SUN database were queried that contain no less than 20 object instances per category, removing those with too few search results, resulting in a total of 660 categories. After downloading, we remove miscategorized models using Amazon Mechanical Turk. Turkers are shown a sequence of thumbnails of the models and answer "Yes" or "No" as to whether the category label matches the model. The authors then manually checked each 3D model and removed irrelevant objects from each CAD model. Website: http://3DShapeNets.cs.princeton.edu	[64]
RS19	The RS19 dataset contains 1,005 high-spatial resolution images with 600 × 600 pixels divided into 19 classes, with approximately 50 images per class. Exported from Google Earth, which provides high-resolution satellite images up to half a meter, this dataset has samples collected from different regions all around the world, which increases its diversity but creates challenges due to the changes in resolution, scale, orientation and illuminations of the images. The classes include airport, beach, bridge, river, forest, meadow, pond, parking, port, viaduct, residential area, industrial area, commercial area, desert, farmland, football field, mountain, park and railway station.	[65],[4]
RSSCN7	The RSSCN7 data set contains 2,800 remote sensing scene images, which are from seven typical scene categories: grassland, forest,	[66], [67]

Dataset	Description	References
	farmland, parking lot, residential region, industrial region, and river and lake. For each category, there are 400 images collected from the Google Earth, which are sampled on four different scales with 100 images per scale. Each image is 400 × 400 pixels. This data set is rather challenging due to the wide diversity of the scene images that are captured under changing seasons and varying weathers and sampled on different scales. Website: https://sites.google.com/site/qinzoucn/documents	
Russian Traffic Signs Dataset (RTSD)	RTSD consists of 9508 images with signs and 71050 background images. It contains 14,360 sign bounding boxes, 6387 of which are also labeled with a physical sign id. There are 863 labeled physical signs, thus each physical sign is encountered on average 7.3 times. The dataset is divided into training and test part. There are 4,754 training images with signs and 44817 background images. The remaining images are test images. Website: ftp://anonymous@kiviuq.gml-team.ru/AnonymousFTP/RTSD/	[68]
Salinas	This Salinas scene was collected by the 224-band AVIRIS sensor over Salinas Valley, California, and is characterized by high spatial resolution (3.7-meter pixels). The area covered comprises 512 lines by 217 samples. As with Indian Pines scene, we discarded the 20 water absorption bands, in this case bands 108-112, 154-167, 224. This image was available only as at-sensor radiance data. Salinas's ground truth contains 16 classes. Website: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Salinas	[69]
SAT-4	SAT-4 has 500,000 image patches, each sized 28x28, with 4 bands (R,G,B,NIR) covering four broad land cover classes: barren land, trees, grassland and a class that consists of all land cover classes other than the above three. 400,000 patches (comprising of four-fifths of the total dataset) were chosen for training and the remaining 100,000 (one-fifths) were chosen as the testing dataset. We ensured that the training and test datasets belong to disjoint set of image tiles. Each image patch is size normalized to 28x28 pixels. Once generated, both the training and testing datasets were randomized using a pseudo-random number generator. Website: http://csc.lsu.edu/~saikat/deepsat/	[70]
SAT-6	SAT-6 consists 405,000 image patches (28x28) contain 4 bands (R,G,B,NIR) covering 6 land cover classes: barren land, trees, grassland, roads, buildings and water bodies. 324,000 images (comprising of four-fifths of the total dataset) were chosen as the training dataset and 81,000 (one fifths) were chosen as the testing dataset. Similar to SAT-4, the training and test sets were selected from disjoint NAIP tiles. Website: http://csc.lsu.edu/~saikat/deepsat/	[70]
SOCAT	The Surface Ocean CO2 Atlas (SOCAT) project was initiated by the international marine carbon science community in 2007 with the aim of	[71]

Dataset	Description	References
	providing a comprehensive, publicly available, regularly updated, global data set of marine surface CO2, which had been subject to quality control (QC). Website: http://www.socat.info/	
Sports-8	This dataset contains eight sports event categories collected from the Internet: bocce, croquet, polo, rowing, snowboarding, badminton, sailing, and rock climbing. This event dataset is a very challenging one. Some of the difficulties are: (1) The background of each image is highly cluttered and diverse; (2) Object classes are diverse; (3) Within the same category, sizes of instances from the same object are very different; (4) The pose of the objects can be very different in each image; (5) Number of instances of the same object category change diversely even within the same event category; and (6) Some of the foreground objects are too small to be detected. Website: http://vision.stanford.edu/resources_links.html	[72]
SUN	SUN stands for Scene Understanding dataset. This dataset contains 908 categories and 131,072 images. It is continually growing as scripts extract more images over time. Website: http://vision.princeton.edu/projects/2010/SUN/hierarchy/	[73]
SUN-397	"SUN" stands for Scene Understanding. This is a large scale object recognition database imagery, containing 130,519 images from 899 categories. The dataset represents scenes from WordNet [74] with images available on the tiny images database [75]. The final dataset reaches 899 categories and 130,519 images. We refer to this dataset as database. Website (WordNet): https://wordnet.princeton.edu/ Website (tiny images): http://people.csail.mit.edu/torralba/tinyimages Website: http://groups.csail.mit.edu/vision/SUN/	[74],[75],[76]
Sydney	Sydney dataset contains 7 different scene categories (residential, airport, meadow, rivers, ocean, industrial, and runway) and 613 HSR images. The spatial resolution of this image was about 0.5 m.	[77]
TudBrussels	Pedestrian detection dataset taken using camera in a driving car. Website: http://www.d2.mpi-inf.mpg.de/tud-brussels	[78],[79]
UC Merced Land Use dataset	Aerial optical images with low- level characteristics similar to those of ImageNet. Website: http://vision.ucmerced.edu/datasets/landuse.html	[80]
URBAN1, URBAN2	Road detection datasets. The URBAN1 dataset consists of over 500 square kilometers of training data and 48 square kilometers of test data at a resolution of 1.2m per pixel. This dataset contains both urban and suburban areas of a large city and has relatively few registration problems, but does contain omission errors. The URBAN2 dataset consists of a 28 square kilometer subset of a different city than the one covered by URBAN1 and has significant registration problems in addition to containing omission errors.	[81]
USPST	The USPST data set is a subset (the testing set) of the well-known handwritten digit recognition data set USPS. The USPST(B) data set is a	[19]

Dataset	Description	References
	binary classification task obtained from USPST by grouping the first 5 digits as Class 1 and the last 5 digits as Class 2. Website: https://www.otexts.org/1577	
VL-CMU-CD	The VL-CMU-CD extracts extracted 152 RGB and depth image sequences for change detection from the VL-CMU dataset. Each sequence contains on average 9 pairs of corresponding images taken from different time instances. There are 1, 362 registered image pairs, each with a manually annotated ground truth change and sky masks. Website: http://www.saistent.com/proj/RSS2016.html The VL-CMU (Visual Localization) dataset consists of 16 sequences captured over the period of one year in the city of Pittsburgh, PA, USA. The sequences are recorded at 15Hz by a pair of 1024×768 pixel Point Grey Flea 2 vehicle-mounted cameras at 45 degrees left and right from the forwards direction and zero overlap between the pair. In each of the sequences the vehicle traversed approximately the same 8km route. The dataset also includes measurements from an inertial sensor and a GPS. Website: http://3dvis.ri.cmu.edu/data-sets/localization/	[82]
Washington DC Mall	The DC mall dataset was published by Purdue. The sensor is Hyperspectral digital imagery collection experiment (HYDICE) and has 191 spectral bands. Website: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html	[83]
WHU-RS	50 satellite images with a size of 600×600 for each of the 19 classes, collected from Google Earth. Classes include airport, beach, bridge, commercial, desert, farmland, football field, etc.	[84], [85]
Yellow River	SAR images acquired by Radarsat-2 at the region of Yellow River Estuary in China in June 2008 and June 2009. Website: http://www.asc-csa.gc.ca/eng/satellites/radarsat2/default.asp	[86]

References

[1] L. Fei-Fei and P. Pietro, "A bayesian hierarchical model for learning natural scene categories," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 2, pp. 524–531.

[2] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, 2006, vol. 2, pp. 2169–2178.

[4] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maitre, "Structural high-resolution satellite image indexing," in *ISPRS TC VII Symposium-100 Years ISPRS*, 2010, vol. 38, pp. 298–303.

[5] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image

representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 173–176, 2011.
[6] J. Roberts and S. Party, "Changing Oceans Expedition 2012 RRS James Cook 073 Cruise Report," 2013.
[7] V. Risojević, S. Momić, and Z. Babić, "Gabor descriptors for aerial image classification," in *International Conference on Adaptive and Natural Computing Algorithms*, 2011, pp. 51–60.
[8] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, 2005.
[9] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "Bigbird: A large-scale 3d database of object instances," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 509–516.
[10] M. Munaro, A. Basso, A. Fossati, L. Van Gool, and E. Menegatti, "3D reconstruction of freely moving persons for re-identification with a depth sensor," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 4512–4519.
[11] "Botswana." [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Botswana. [Accessed: 10-Feb-2017].
[12] A. B. Penatti, K. Nogueira, and J. A. Santos, "Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains ?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 44–51.
[13] "Caltech Pedestrian Dataset." [Online]. Available: https://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/. [Accessed: 18-Feb-2017].
[14] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, 2012.
[15] "Caltech 1999." [Online]. Available: http://www.vision.caltech.edu/archive.html. [Accessed: 04-Mar-2017].
[16] "Caltrans, Performance Measurement System (PeMS)," 2014. [Online]. Available: http://pems.dot.ca.gov.
[17] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3730–3738.
[18] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Citeseer, 2009.
[19] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: from transductive to semi-supervised learning," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 824–831.
[20] "Copernicus Sentinel Data." [Online]. Available: https://scihub.copernicus.eu/dhus.
[21] "Copperas Cove HYDICE data set." [Online]. Available: http://www.agc.army.mil/. [Accessed: 10-Feb-2017].
[22] W. Ouyang and X. Wang, "A discriminative deep model for pedestrian detection with occlusion handling," in *2012 IEEE Conf. on Computer Vision and Pattern Recog. (CVPR)*, 2012, pp. 3258–3265.
[23] "Cuprite Dataset." [Online]. Available: http://aviris.jpl.nasa.gov/data/free_data.html. [Accessed: 11-Feb-2017].
[24] "Daimler Pedestrian Dataset." [Online]. Available: http://www.gavrila.net/Datasets/Daimler_Pedestrian_Benchmark_D/daimler_pedestrian_benchmark_d.html. [Accessed: 18-Feb-2017].
[25] D. A. Landgrebe, *Signal theory methods in multispectral remote sensing*, vol. 29. John Wiley & Sons, 2005.

[26] “ETH Pedestrian Dataset.” [Online]. Available: http://www.vision.ee.ethz.ch/~aess/dataset/. [Accessed: 18-Feb-2017].
[27] A. Ess, B. Leibe, K. Schindler, and and L. van Gool, “A Mobile Vision System for Robust Multi-Person Tracking,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’08)*, 2008.
[28] Q. Jackson and D. A. Landgrebe, “An adaptive classifier design for high-dimensional data analysis with a limited training data set,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 12, pp. 2664–2679, 2001.
[29] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher, “Supporting ground-truth annotation of image datasets using clustering,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 1542–1545.
[30] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[31] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer, “Weak hypotheses and boosting for generic object detection and recognition,” in *European conference on computer vision*, 2004, pp. 71–84.
[32] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “The German traffic sign recognition benchmark: a multi-class classification competition,” in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, 2011, pp. 1453–1460.
[33] C. W. Hamilton, S. A. Fagents, and T. Thordarson, “Lava--ground ice interactions in Elysium Planitia, Mars: geomorphological and geospatial analysis of the Tartarus Colles cone groups,” *J. Geophys. Res. Planets*, vol. 116, no. E3, 2011.
[34] M. Munaro, S. Ghidoni, D. T. Dizmen, and E. Menegatti, “A feature-based approach to people re-identification using skeleton keypoints,” in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 5644–5651.
[35] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, “icoseg: Interactive co-segmentation with intelligent scribble guidance,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 3169–3176.
[36] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot, and others, “Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, 2009.
[37] “2013 IEEE GRSS Data Fusion Contest.” [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/.
[38] “2015 IEEE GRSS Data Fusion Contest.” [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/.
[39] “2016 IEEE GRSS Data Fusion Contest,” 2016. [Online]. Available: http://www.grss-ieee.org/community/ technical-committees/data-fusion/.
[40] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, “Re-identification with rgb-d sensors,” in *European Conference on Computer Vision*, 2012, pp. 433–442.
[41] R. Timofte, V. De Smet, and L. Van Gool, “Anchored neighborhood regression for fast example-based super-resolution,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1920–1927.
[42] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and others, “Imagenet large scale visual recognition challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[43] “Indian Pines Dataset.” [Online]. Available: http://dynamo.ecn.purdue.edu/biehl/MultiSpec.
[44] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf, “The ISPRS benchmark on urban object classification and 3D building reconstruction,” *ISPRS Ann.*

Photogramm. Remote Sens. Spat. Inf. Sci, vol. 1, pp. 293–298, 2012.
[45] M. Cramer, “The DGPF-test on digital airborne camera evaluation-overview and test design,” *Photogrammetrie-Fernerkundung-Geoinformation*, vol. 2010, no. 2, pp. 73–82, 2010.
[46] C. Li, A. Reiter, and G. D. Hager, “Beyond Spatial Pooling : Fine-Grained Representation Learning in Multiple Domains,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, no. 1, pp. 4913–4922, 2015.
[47] “Pavia Dataset.” [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Pavia_Centre_and_University. [Accessed: 29-Nov-2016].
[48] A. Joly, H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.-P. Vellinga, J. Champ, R. Planqué, S. Palazzo, and H. Müller, “LifeCLEF 2016: multimedia life species identification challenges,” in *International Conference of the Cross-Language Evaluation Forum for European Languages*, 2016, pp. 286–310.
[49] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, “Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes,” in *Asian conference on computer vision*, 2012, pp. 548–562.
[50] Y. Yang and S. Newsam, “Spatial pyramid co-occurrence for image classification,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1465–1472.
[51] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, “Nonlinear unmixing of hyperspectral images: Models and algorithms,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, 2014.
[52] V. Mnih, “Machine Learning for Aerial Image Labeling,” University of Toronto, 2013.
[53] O. Beijbom, P. J. Edmunds, D. I. Kline, B. G. Mitchell, and D. Kriegman, “Automated annotation of coral reef survey images,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 1170–1177.
[54] J. Winn, A. Criminisi, and T. Minka, “Object categorization by learned universal visual dictionary,” in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, vol. 2, pp. 1800–1807.
[55] “MSTAR SAR Dataset.” [Online]. Available: https://www.sdms.afrl.af.mil/index.php?collection=mstar&page=targets. [Accessed: 18-Feb-2017].
[56] K. Lai, L. Bo, X. Ren, and D. Fox, “A large-scale hierarchical multi-view rgb-d object dataset,” in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 1817–1824.
[57] L. Gomez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila, and G. Camps-Valls, “Urban monitoring using multi-temporal SAR and multi-spectral data,” *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 234–243, 2006.
[58] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, “NUS-WIDE: a real-world web image database from National University of Singapore,” in *Proceedings of the ACM international conference on image and video retrieval*, 2009, p. 48.
[59] G. Cheng, J. Han, P. Zhou, and L. Guo, “Multi-class geospatial object detection and geographic image classification based on collection of part detectors,” *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 119–132, 2014.
[60] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgbd images,” in *European Conference on Computer Vision*, 2012, pp. 746–760.
[61] A. Golovinskiy, V. G. Kim, and T. Funkhouser, “Shape-based recognition of 3D point clouds in urban environments,” in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 2154–2161.
[62] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The PASCAL Visual

Object Classes Challenge 2012 (VOC2012) Results.” .
[63] S. Mikeš, M. Haindl, and G. Scarpa, “Remote sensing segmentation benchmark,” in *Pattern Recognition in Remote Sensing (PRRS), 2012 IAPR Workshop on*, 2012, pp. 1–4.
[64] Z. Wu and S. Song, “3D ShapeNets : A Deep Representation for Volumetric Shapes,” pp. 1912–1920, 2015.
[65] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, “Towards better exploiting convolutional neural networks for remote sensing scene classification,” *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.
[66] “RSSCN7 Remote Sensing Dataset.” [Online]. Available: https://sites.google.com/site/qinzoucn/documents. [Accessed: 19-Feb-2017].
[67] B. Zhao, Y. Zhong, and L. Zhang, “A spectral-structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery,” *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 73–85, 2016.
[68] “Russian Traffic Signs Dataset,” *ftp://anonymous@kiviuq.gml- team.ru/AnonymousFTP/RTSD/*. .
[69] “Salinas Dataset.” [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Salinas. [Accessed: 09-Feb-2017].
[70] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, “Deepsat: a learning framework for satellite imagery,” in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015, p. 37.
[71] C. L. Sabine, S. Hankin, H. Koyuk, D. C. E. Bakker, B. Pfeil, A. Olsen, N. Metzl, A. Kozyr, A. Fassbender, A. Manke, and others, “Surface Ocean CO2 Atlas (SOCAT) gridded data products,” *Earth Syst. Sci. Data Discuss.*, vol. 5, no. 2, pp. 781–804, 2012.
[72] L.-J. Li and L. Fei-Fei, “What, where and who? classifying events by scene and object recognition,” in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1–8.
[73] J. Xiao, K. A. Ehinger, J. Hays, A. Torralba, and A. Oliva, “Sun database: Exploring a large collection of scene categories,” *Int. J. Comput. Vis.*, vol. 119, no. 1, pp. 3–22, 2016.
[74] A. Kilgarriff and C. Fellbaum, “WordNet: An Electronic Lexical Database.” JSTOR, 2000.
[75] A. Torralba, R. Fergus, and W. T. Freeman, “80 million tiny images: A large data set for nonparametric object and scene recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1958–1970, 2008.
[76] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, “Sun database: Large-scale scene recognition from abbey to zoo,” in *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, 2010, pp. 3485–3492.
[77] F. Zhang, B. Du, and L. Zhang, “Saliency-guided unsupervised feature learning for scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, 2015.
[78] “TUD-Brussels Dataset.” [Online]. Available: http://www.d2.mpi-inf.mpg.de/tud-brussels. [Accessed: 18-Feb-2017].
[79] C. Wojek, S. Walk, and B. Schiele, “Multi-cue onboard pedestrian detection,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 794–801.
[80] Y. Yang and S. Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.
[81] V. Mnih and G. E. Hinton, “Learning to detect roads in high-resolution aerial images,” in *European Conference on Computer Vision*, 2010, pp. 210–223.
[82] P. F. Alcantarilla, S. Stent, G. Ros, R. Arroyo, and R. Gherardi, “Streetview change detection with deconvolutional networks,” in *Robotics: Science and Systems Conference (RSS)*, 2016.
[83] “Washington DC Mall.” [Online]. Available:

https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html.
[84] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv Prepr. arXiv1405.3531*, 2014.
[85] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.
[86] "Yellow River Radarsat 2." [Online]. Available: http://www.asc-csa.gc.ca/eng/satellites/radarsat2/default.asp. [Accessed: 04-Mar-2017].