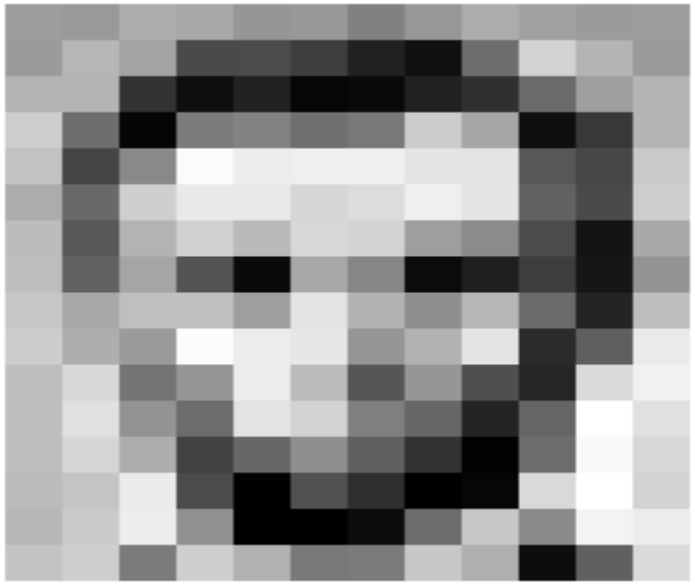
Deep Learning for Remote Sensing

AI60002

28th Feb 2022

Deep Learning for Images

- Provided: images (2D or 3D matrices)
- X(i,j,k): a pixel on the i-th row, j-th column and k-th channel
- Information: neighboring pixels likely to have similar values (spatial autocorrelation)
- Further: neighboring pixels likely to belong to same object



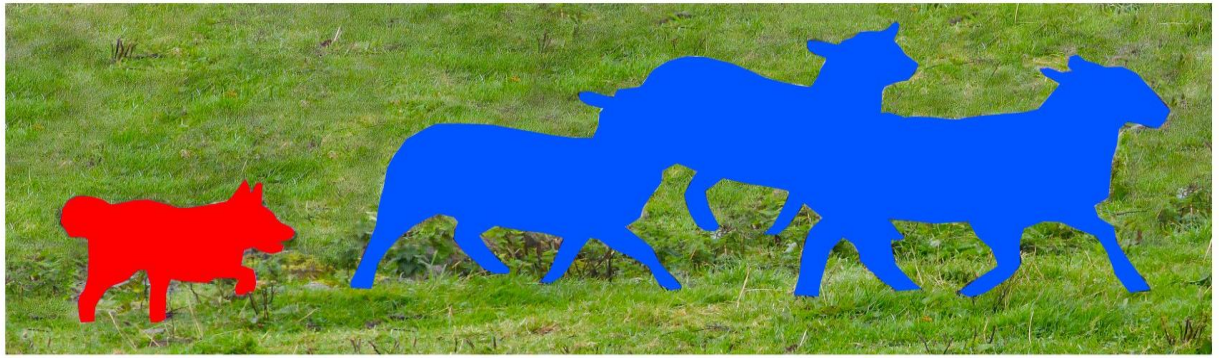
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

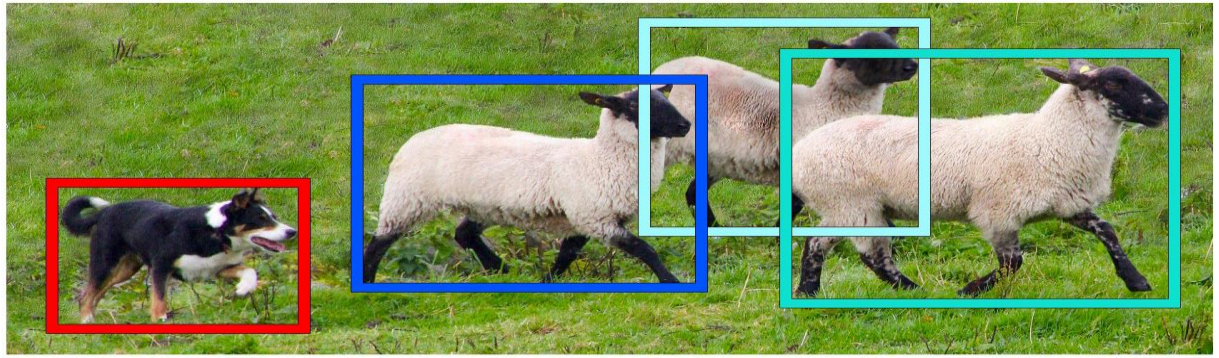
Deep Learning for Images

- Object Recognition
- Object Detection (for a particular object/set of objects)
- Image segmentation

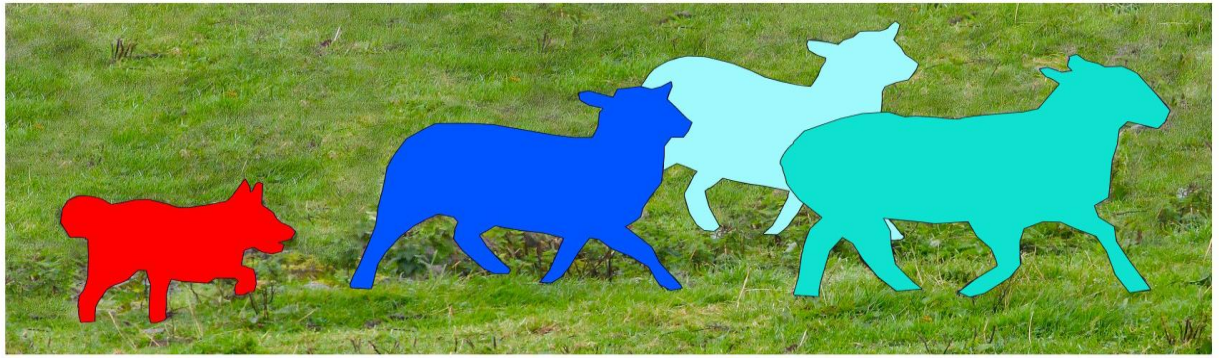
Image Recognition



Semantic Segmentation



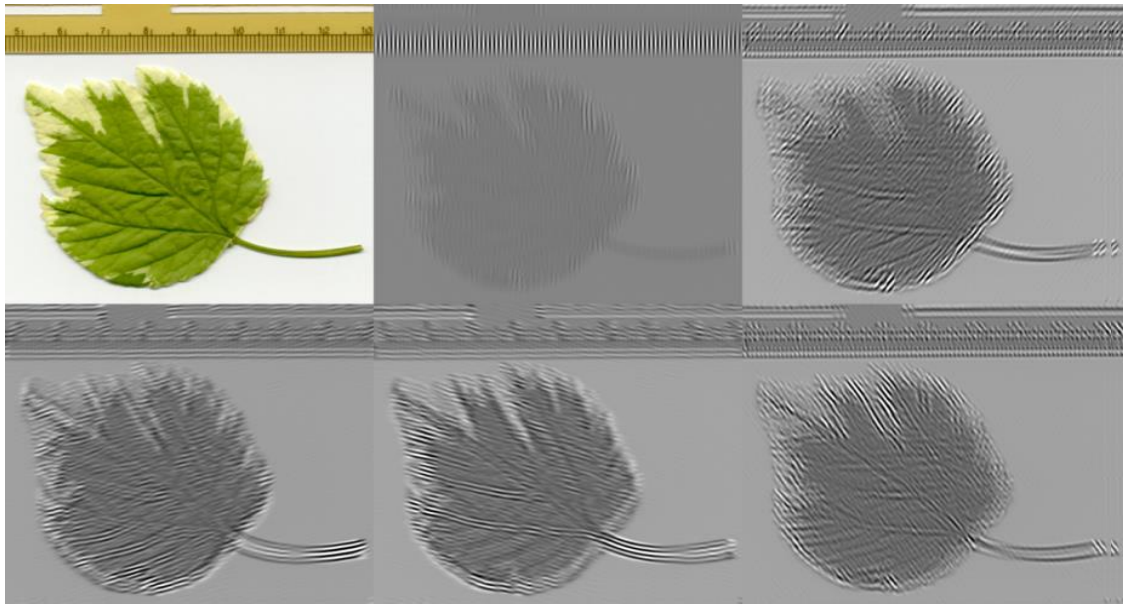
Object Detection

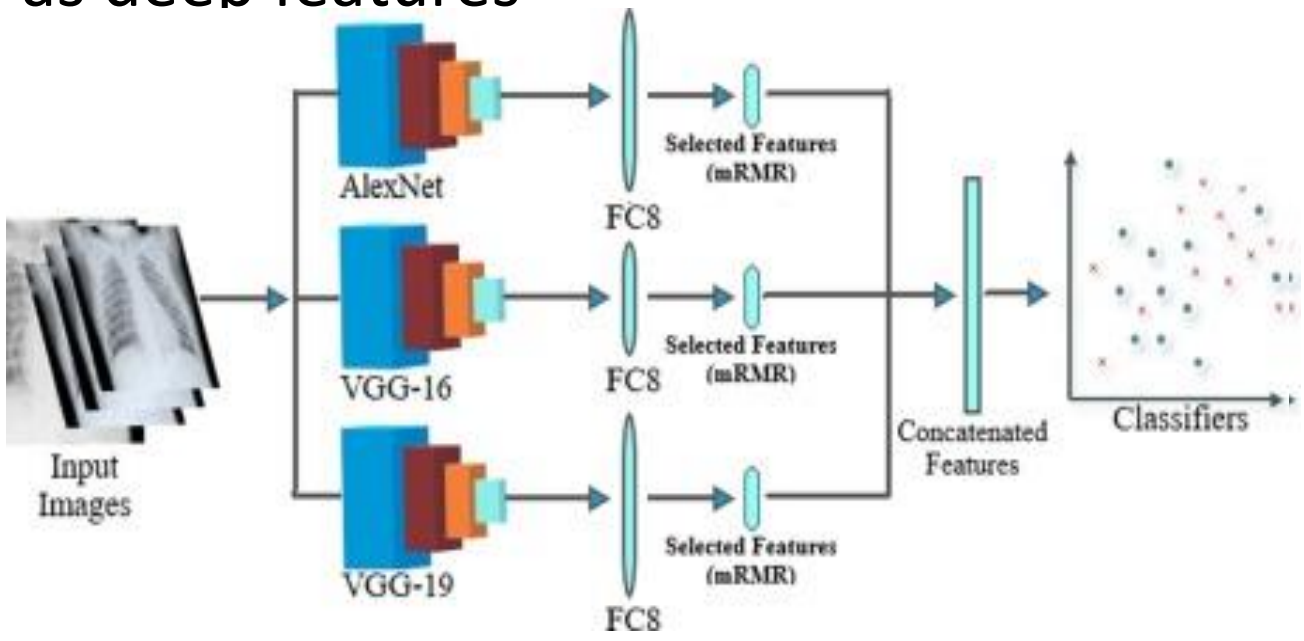


Instance Segmentation

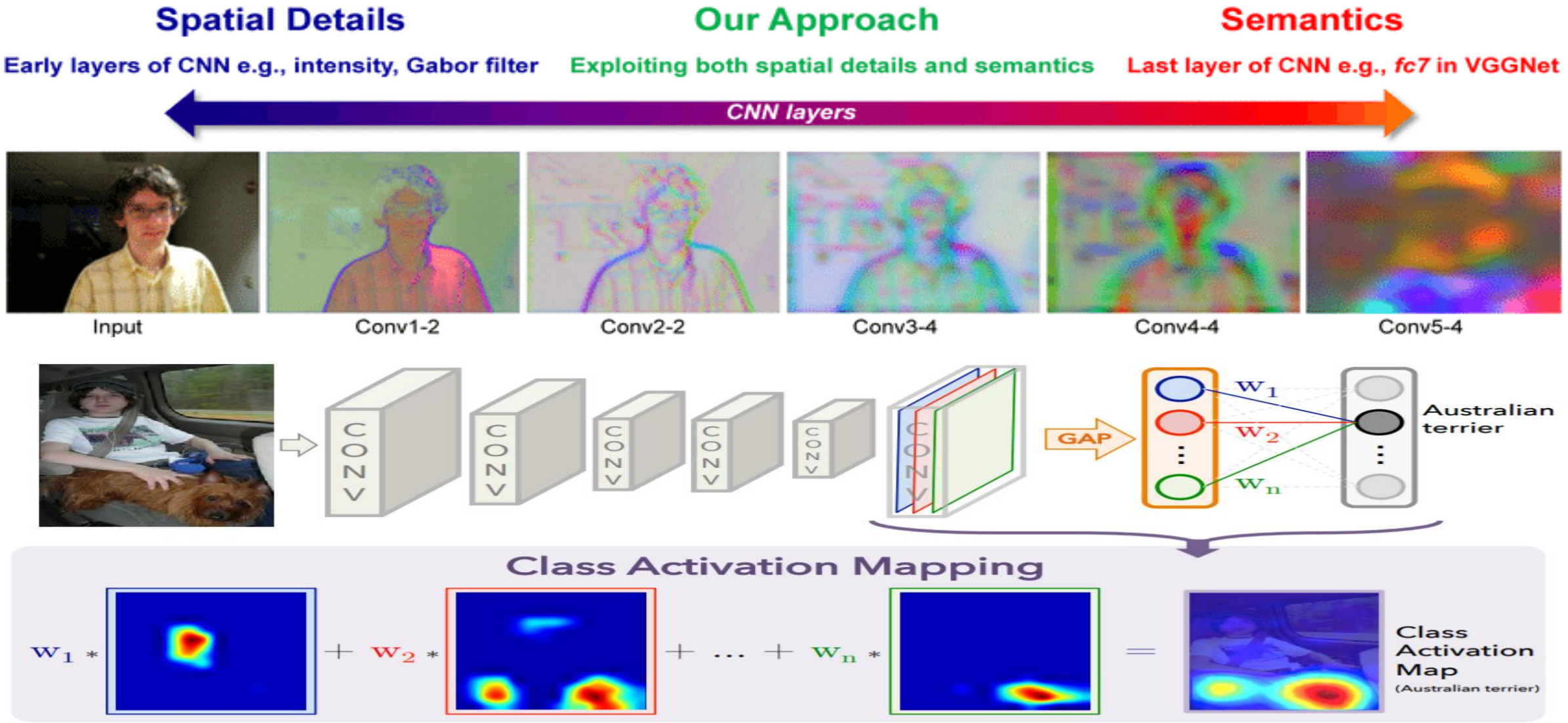
Deep Learning for Images

- Image representation – each image must be represented, as a whole and in parts
- Representation may be using raw pixel values, filter outputs
- Deep features: an image/sub-image provided as input to a neural network
- Each layer of the neural network creates a new representation of the image
- These representations can be used as deep features





Deep Learning for Images



Max-Pooling and Convolution Operations

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

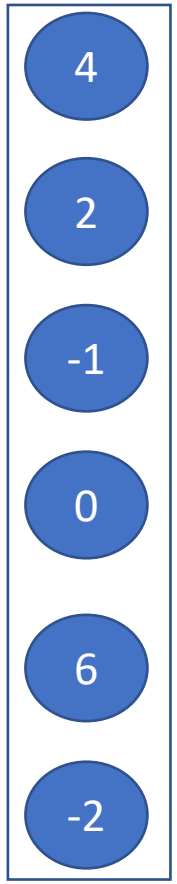
Block size:2
Stride: 2

6	8
3	4

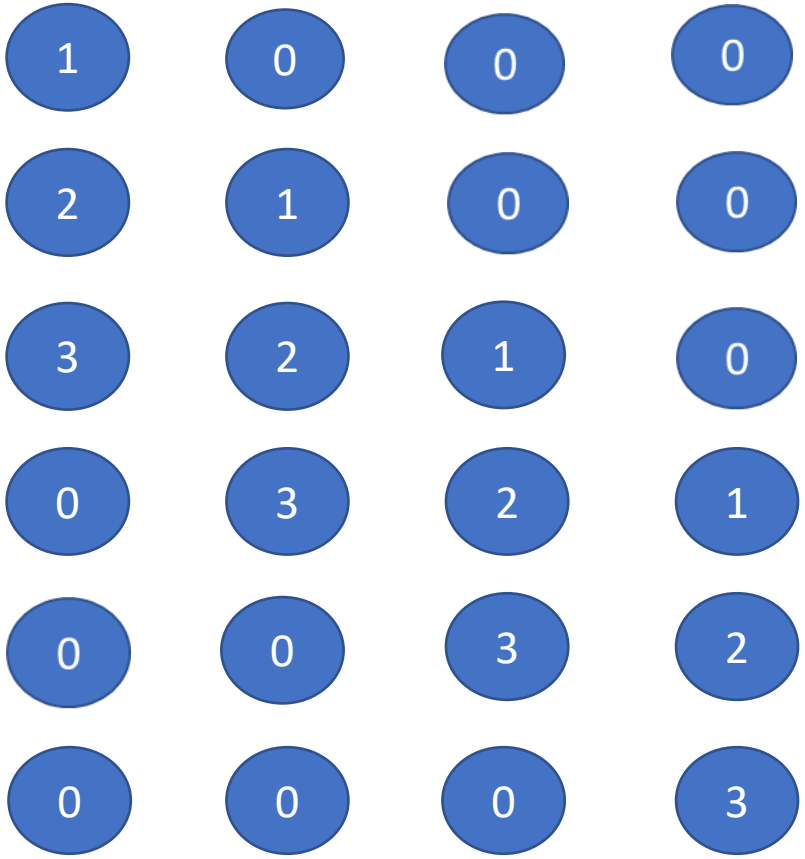
1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

Block size:3
Stride: 1

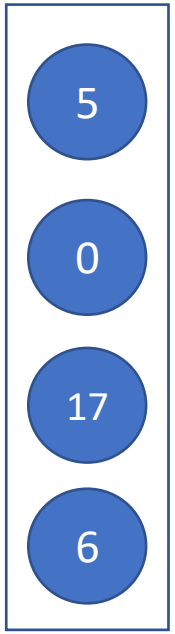
7	8
7	8



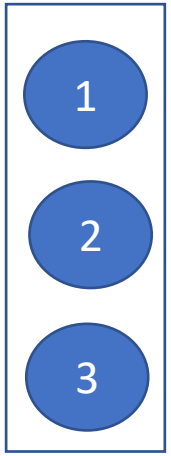
INPUT



WEIGHT MATRIX



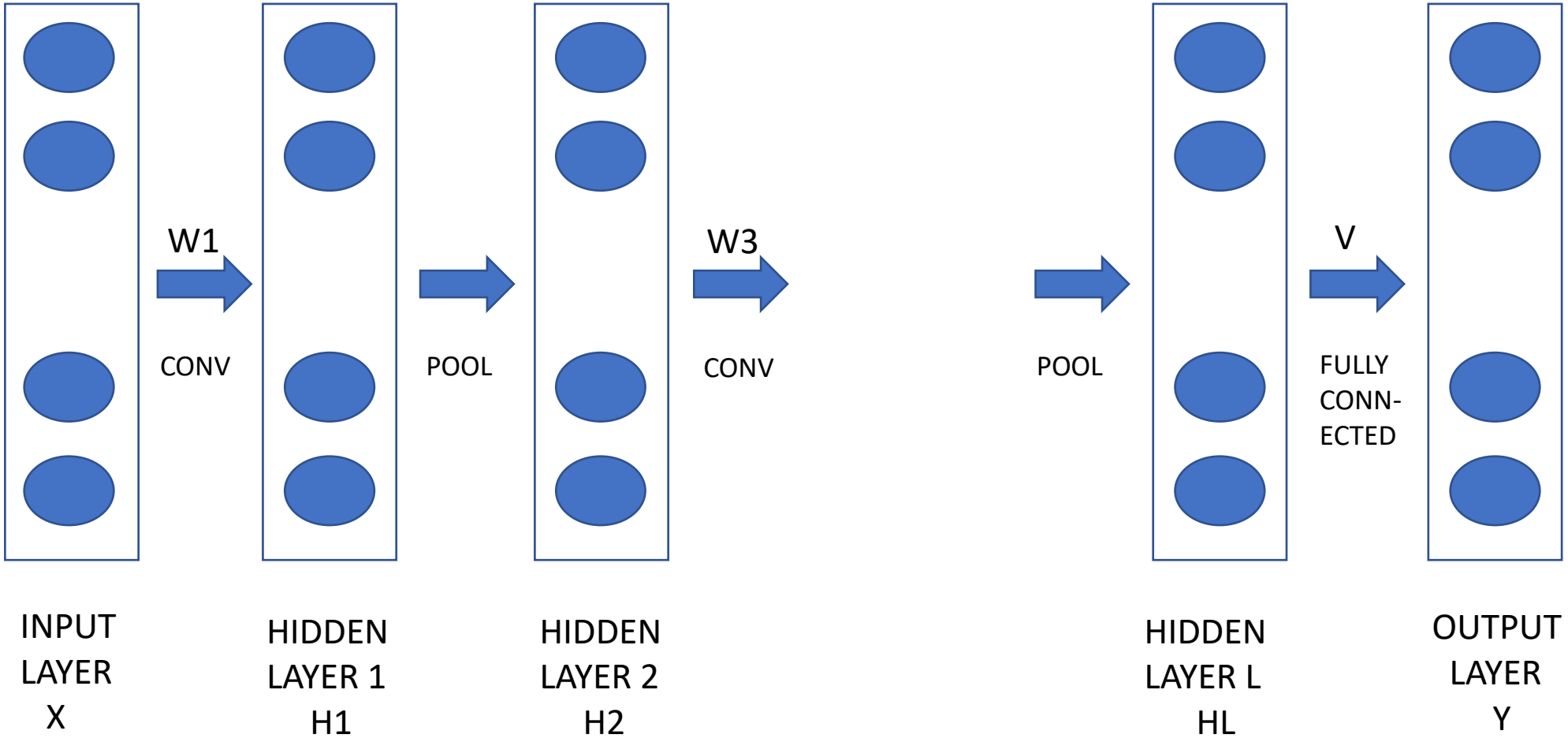
OUTPUT



REPEATING STRUCTURE

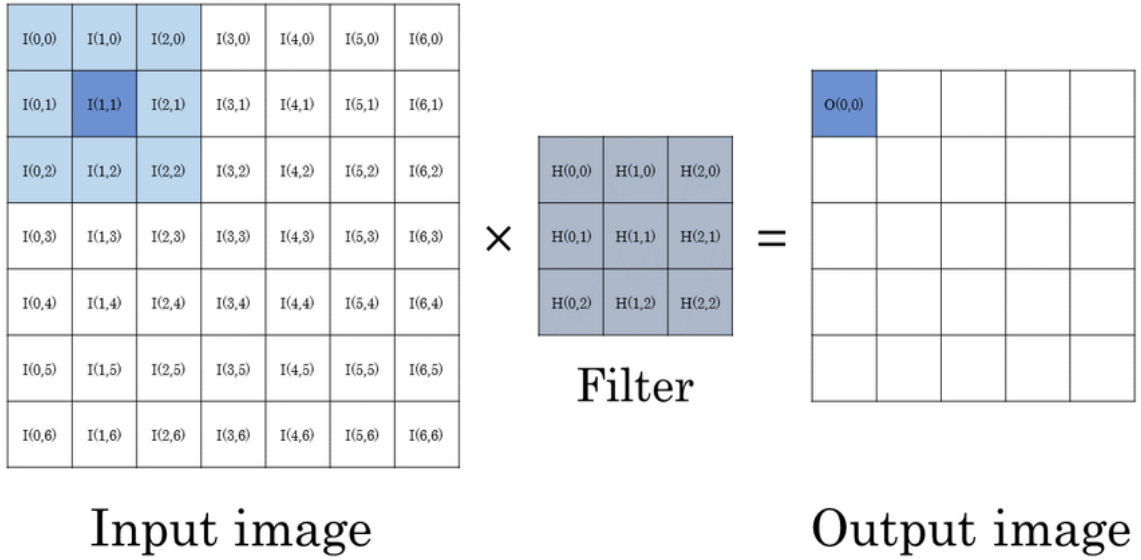
Convolutional Neural Network

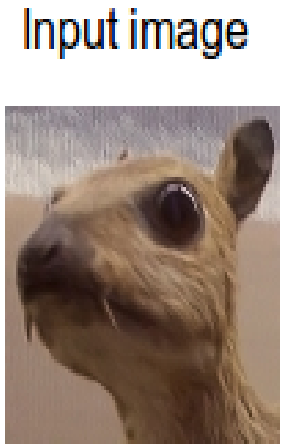
- A convolutional neural network has many “convolution layers”
- Usually, each convolutional layer is followed by a pooling layer



Deep Learning for Images

- Convolution – a standard operation on images
- Useful to identify local structures/orientations in images
- One pass of a convolution filter over the whole image provides the location of a certain kind of orientations



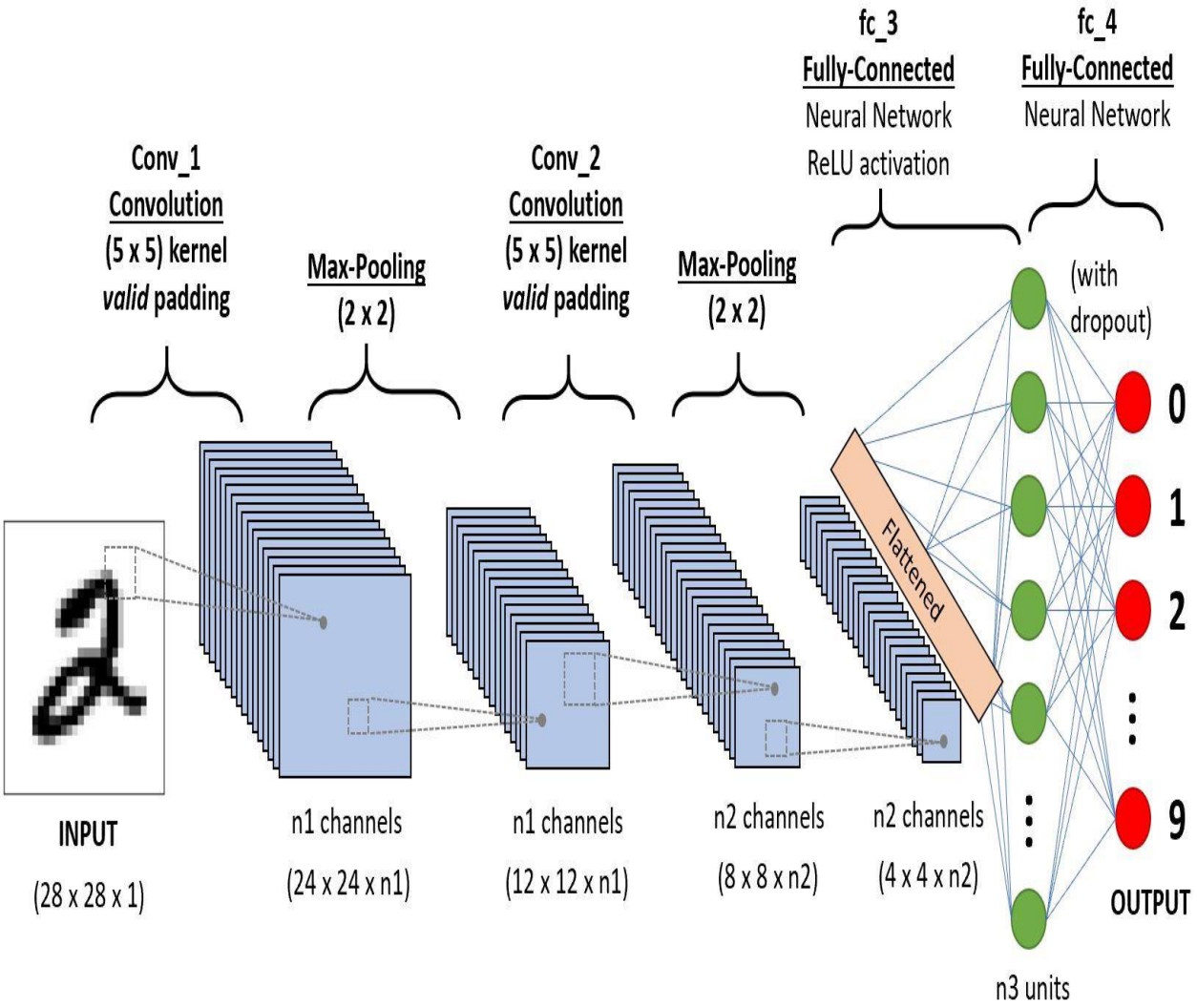


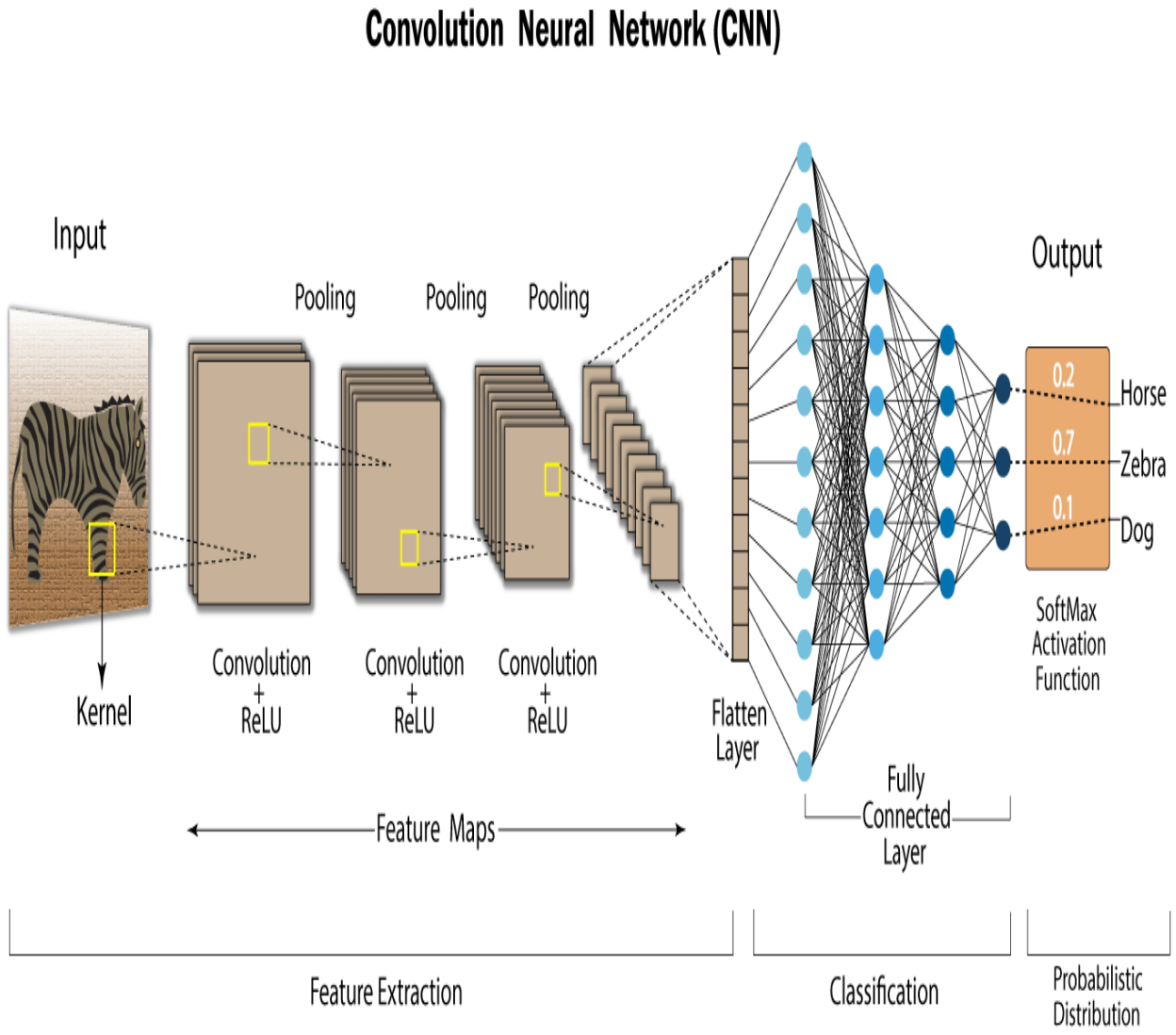
Convolution Kernel

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

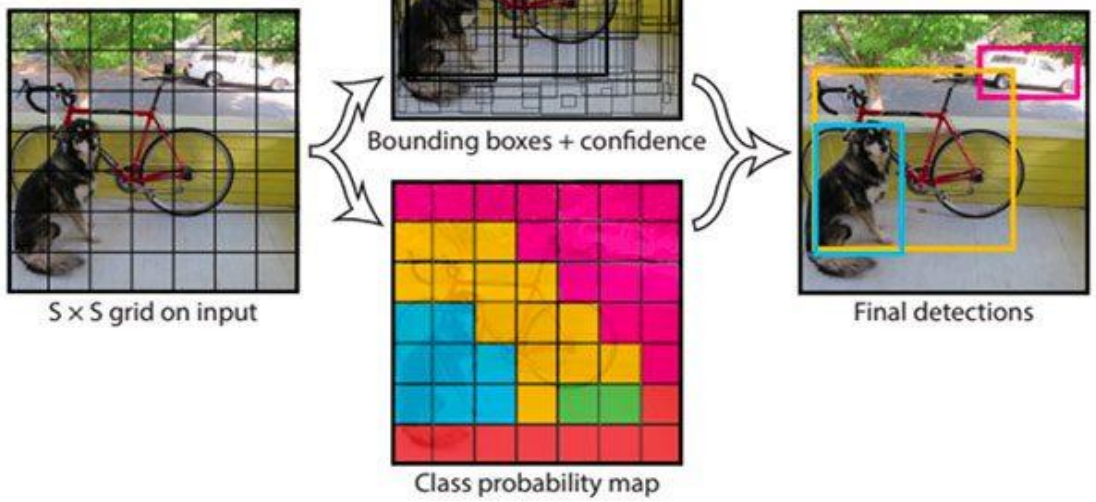
- Can be done with a specially structured neural network (repeating edge weights)

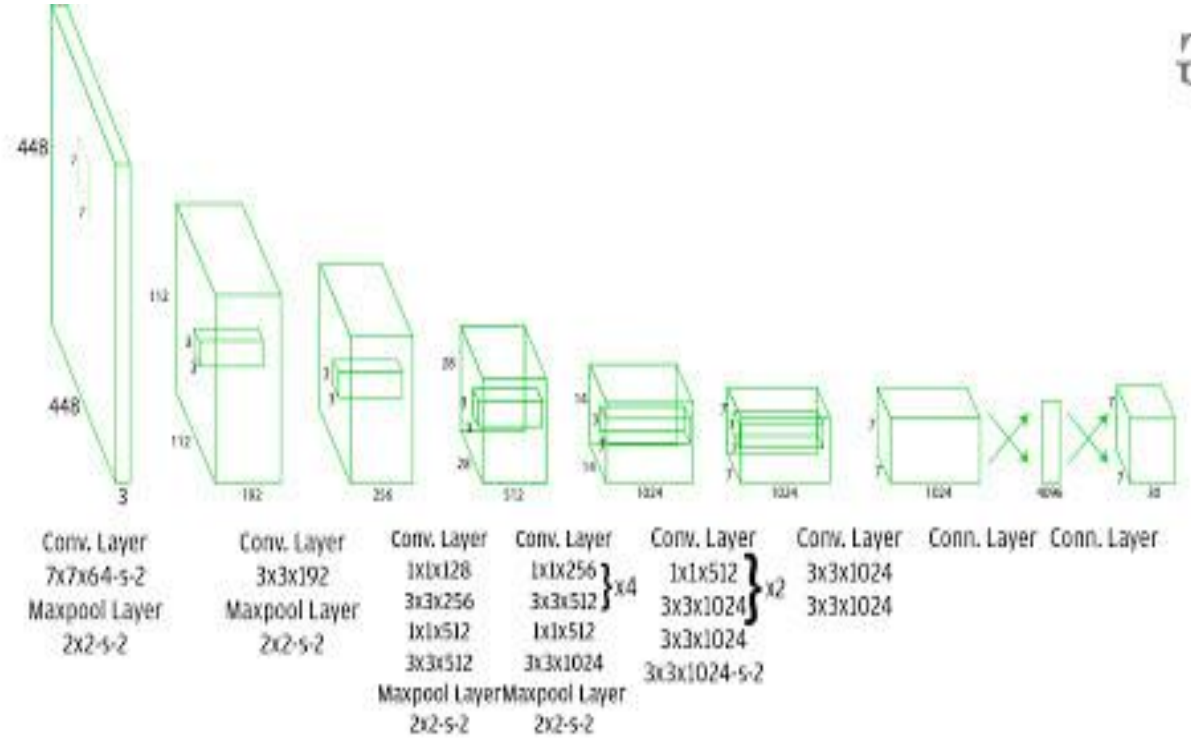
Deep Learning for Images





Object Detection with YOLO – You Only Look Once





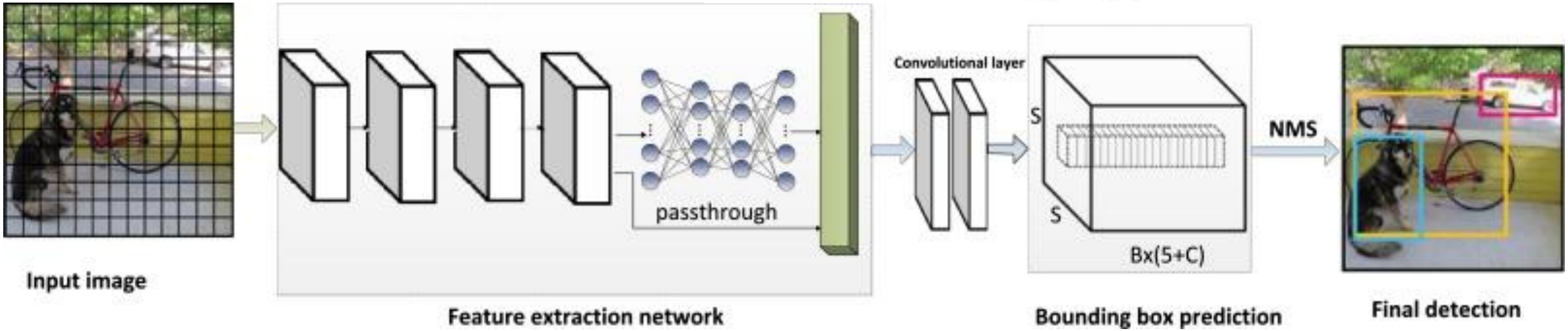
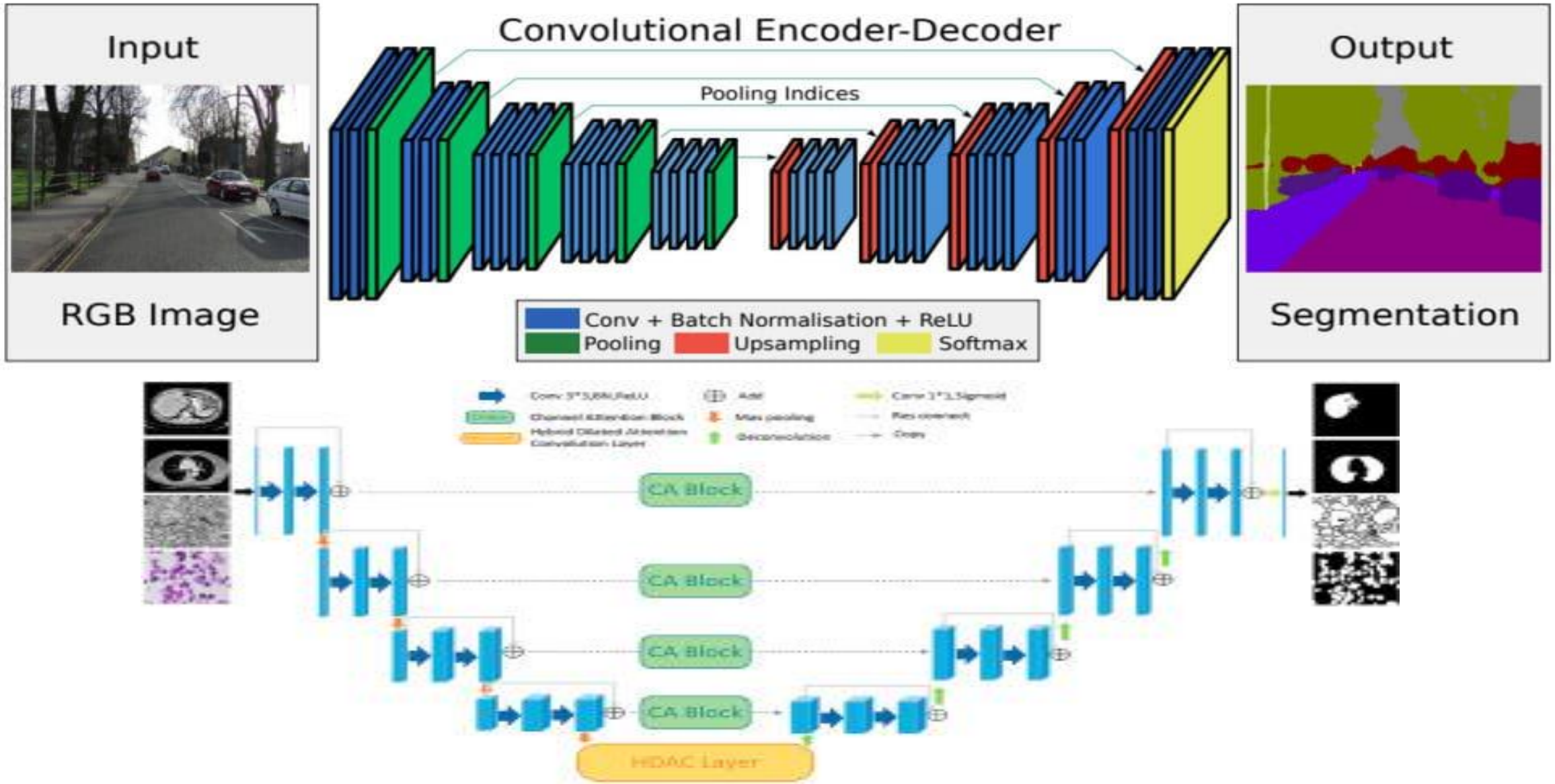


Image Semantic Segmentation



Object Detection in Satellite Images

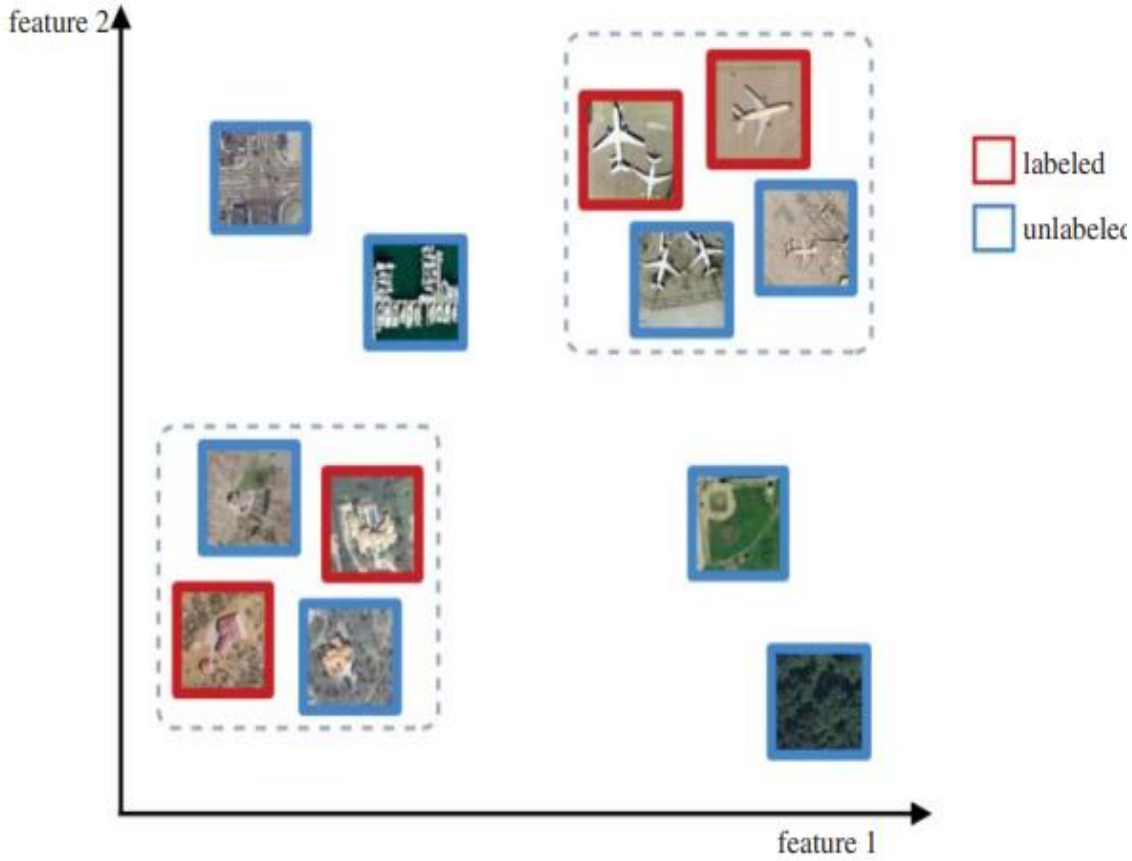


Figure 4.1 Schematic illustration of different learning paradigms and their use of labeled (red) and unlabeled (blue) data samples. In contrast to semi-supervised learning (data samples used shown in dotted boxes), self-taught learning also uses unlabeled data, which need not belong to the same classes as the labeled data. Images are from the UC Merced dataset (Yand and Newsam 2010).

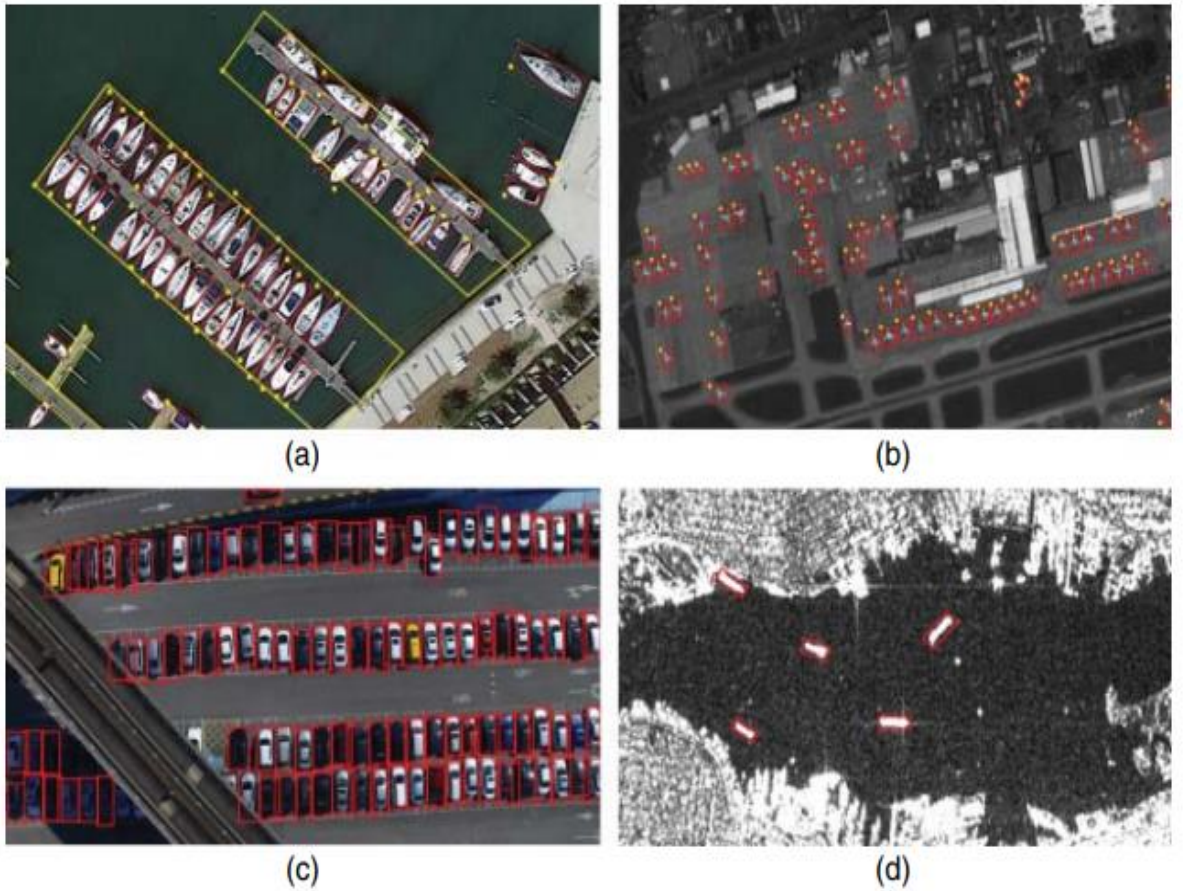


Figure 6.1 Examples of remote sensing images containing objects of interest. (a) An image from Google Earth, containing ships and harbors. (b) An image form JL-1 satellite, including planes. (c) An drone-based image containing many vehicles. (d) A SAR image, containing ships.

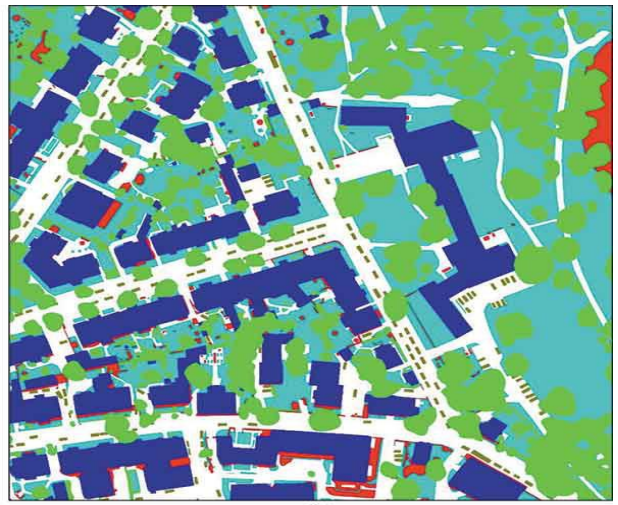
Segmentation of Satellite Images

(a) Original image

(b) Ground truth

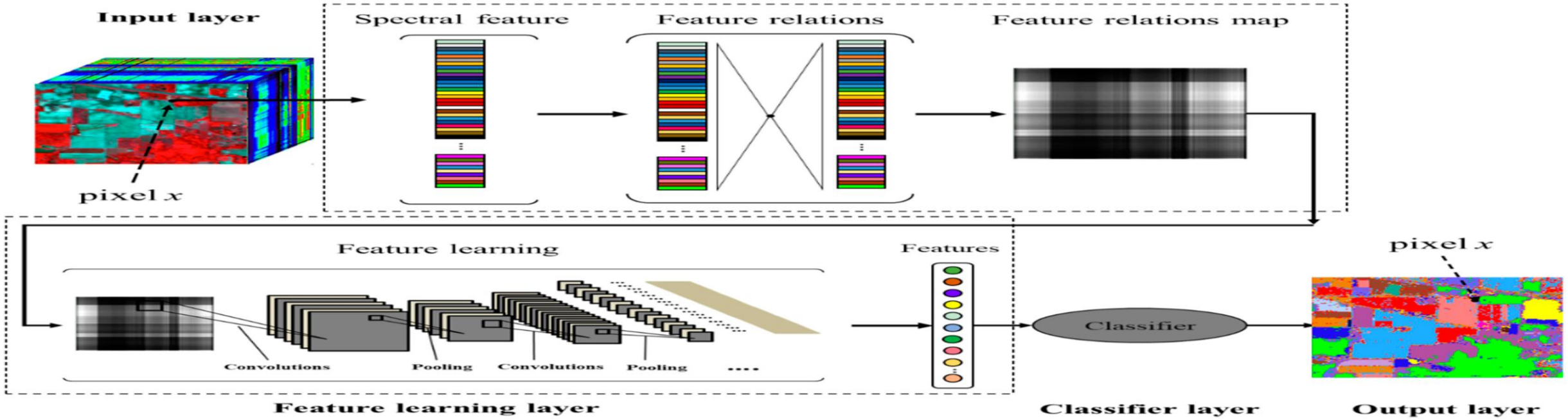
Land-Use Land Cover mapping

(a)

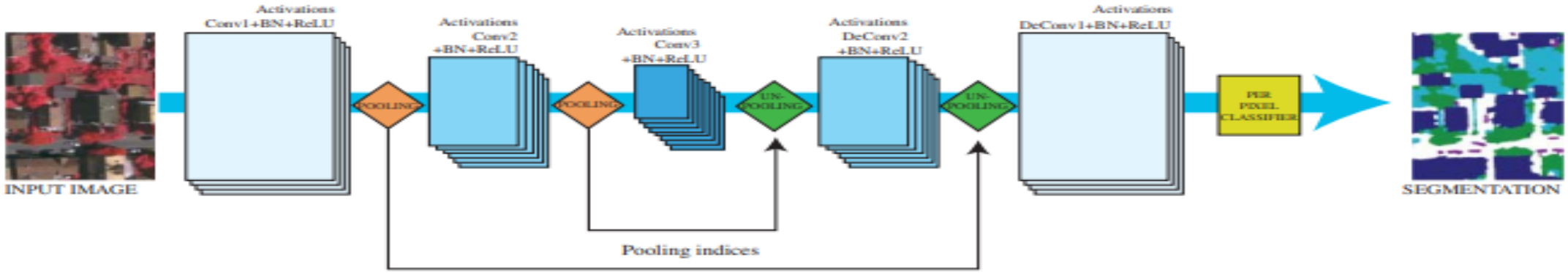


(b)

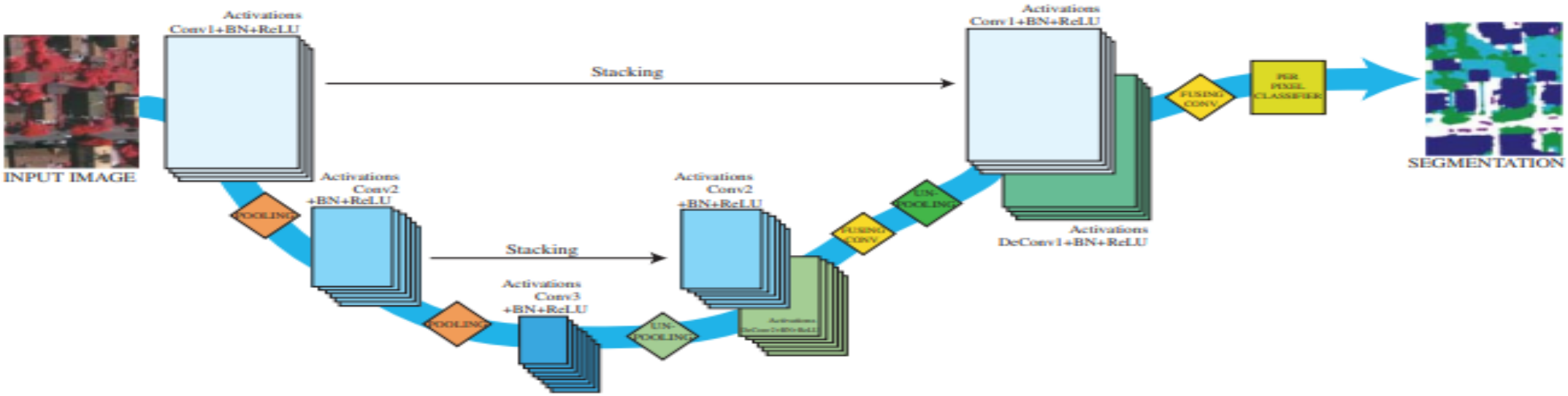
Road Network Extraction



Segmentation of Satellite Images



(a) SegNet (Badrinarayanan et al. 2017), propagating pooling indices.



(b) U-Net (Ronneberger et al. 2015a), propagating activation maps.

Figure 5.3 Semantic segmentation architectures learning the upsampling.

Deep Domain Adaptation

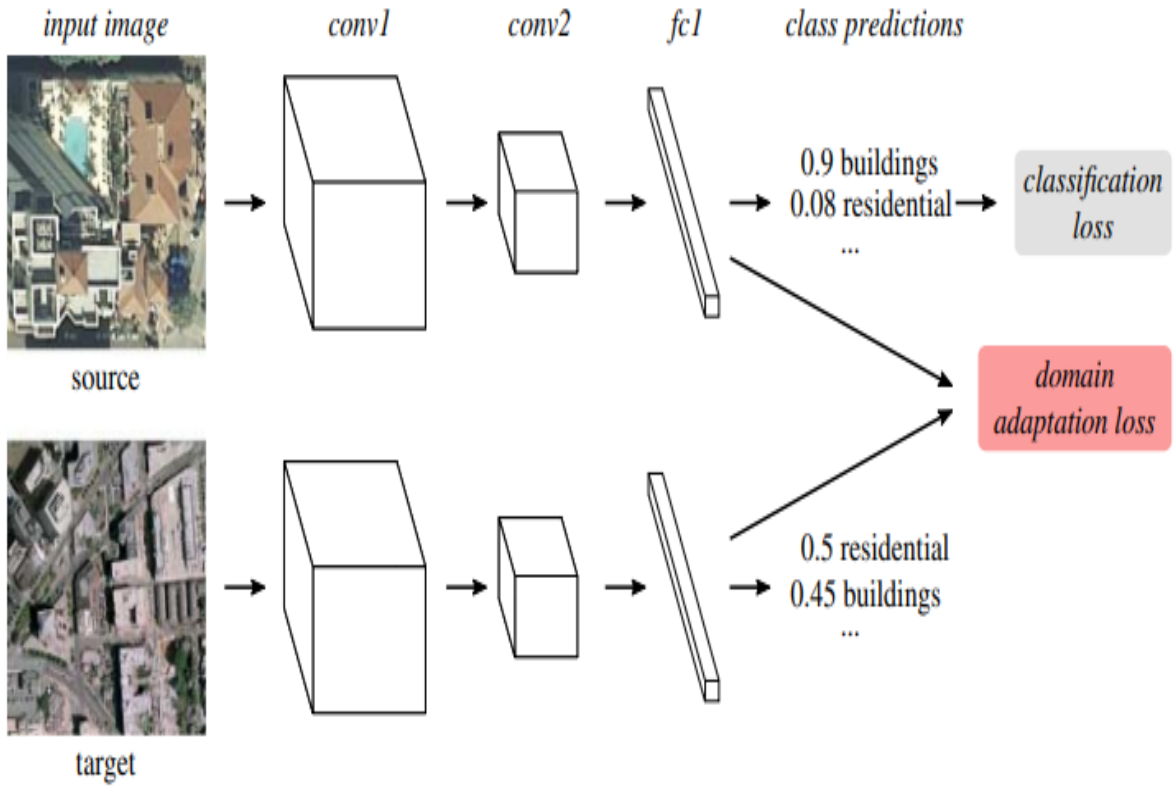


Figure 7.1 Domain adaptation loss (red) imposed on a CNN's feature vectors produced by the penultimate layer ("fc1").

Figure 7.2 Examples from the UC Merced (top) and WHU-RS19 (bottom) datasets.

SkyScapes – Fine-Grained Semantic Understanding of Aerial Scenes

Seyed Majid Azimi[1] Corentin Henry[1] Lars Sommer[2] Arne Schumann[2] Eleonora Vig[1]

[1]German Aerospace Center (DLR), Wessling, Germany [2]Fraunhofer IOSB, Karlsruhe, Germany

https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-12760

Corresponding author: seyedmajid.azimi@dlr.de

Aerial image with overlaid annotation: dense (19 classes) and lane markings (12 classes); the dataset covers 5.7 km^2.

Abstract

Understanding the complex urban infrastructure with centimeter-level accuracy is essential for many applications from autonomous driving to mapping, infrastructure monitoring, and urban management. Aerial images provide valuable information over a large area instantaneously; nevertheless, no current dataset captures the complexity of aerial scenes at the level of granularity required by real-world applications. To address this, we introduce SkyScapes, an aerial image dataset with highly-accurate, fine-grained annotations for pixel-level semantic labeling. SkyScapes provides annotations for 31 semantic categories ranging from large structures, such as buildings, roads and vegetation, to fine details, such as 12 (sub-)categories of lane markings. We have defined two main tasks on this dataset: dense semantic segmentation and multi-class lane-marking prediction. We carry out extensive experiments to evaluate state-of-the-art segmentation methods on SkyScapes. Existing methods struggle to deal with the wide range of classes, object sizes, scales, and fine details present. We therefore propose a novel multi-task model, which incorporates semantic edge detection and is better tuned for feature extraction from a wide range of scales. This model achieves notable improvements over the baselines in region outlines and level of detail on both tasks.

accuracy are of great aid in handling their growing complexity. Applications of such accurate maps include urban management, city planning, and infrastructure monitoring/maintenance. Another prominent example is the creation of *high definition (HD)* maps for autonomous driving. Applications here include the use of a general road network for navigation and more advanced automation tasks in *Advanced driver assistance systems (ADAS)*, such as lane departure warnings, which rely on precise information about lane boundaries, sidewalks, etc. [37, 40, 33, 51, 31].

Currently, the data collection process to generate HD maps is mainly carried out by so-called mobile mapping systems, which comprise of a vehicle equipped with a broad range of sensors (*e.g.*, Radar, LiDAR, cameras) followed by automated analysis of the collected data [17, 18, 5, 24]. The limited field-of-view and occlusions due to the oblique sensor angle make this automated analysis complicated. In addition, mapping large urban areas in this way requires a lot of time and resources. An aerial perspective can alleviate many of these problems and simultaneously allow for processing of much larger areas of cm-level georeferenced data in a short time. Existing aerial semantic segmentation datasets, however, are limited in the range of their annotations. They either focus on a few individual classes, such as roads or building footprints in the IN-

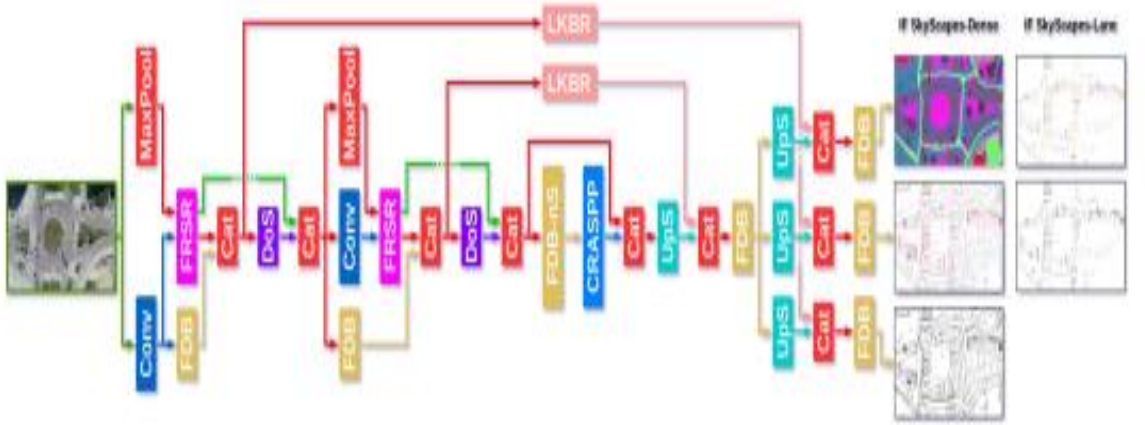


Figure 3: The architecture of SkyScapesNet. Three branches are used to predict dense semantics and multi-class/binary edges. For multi-class lane-marking prediction, two branches are used to predict multi-class and binary lane-markings.

Figure 9: Segmentation result samples of our model trained on SkyScapes and tested on an aerial image over Perth, Australia, with GSD adjustment and without fine-tuning.

CLOUD-NET: AN END-TO-END CLOUD DETECTION ALGORITHM FOR LANDSAT 8 IMAGERY

Sorour Mohajerani, Parvaneh Saeedi

School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

ABSTRACT

Cloud detection in satellite images is an important first-step in many remote sensing applications. This problem is more challenging when only a limited number of spectral bands are available. To address this problem, a deep learning-based algorithm is proposed in this paper. This algorithm consists of a Fully Convolutional Network (FCN) that is trained by multiple patches of Landsat 8 images. This network, which is called Cloud-Net, is capable of capturing global and local cloud features in an image using its convolutional blocks. Since the proposed method is an end-to-end solution no complicated pre-processing step is required. Our experimental results prove that the proposed method outperforms the state-of-the-art method over a benchmark dataset by 8.7% in Jaccard Index.

from Cirrus band of Landsat 8 to increase the accuracy of the detected clouds and is currently utilized to produce cloud masks of the Landsat Level-1 data products [14]. Qui et al. in [9] integrated Digital Elevation Map (DEM) information into Fmask and improved its performance in mountainous areas.

Haze Optimized Transformation (HOT) [10] is among the most famous handcrafted algorithms for identification of clouds [11]. In this algorithm, Zhang et al. utilized the correlation between two spectral bands of Landsat images to distinguish thin clouds from clear regions.

In recent years, deep learning-based methods have been proved to deliver good performance in many image processing applications. Researchers, in the remote sensing field, have also proposed such algorithms to address the problem of cloud detection. For instance, Xie et al. [12] utilized two

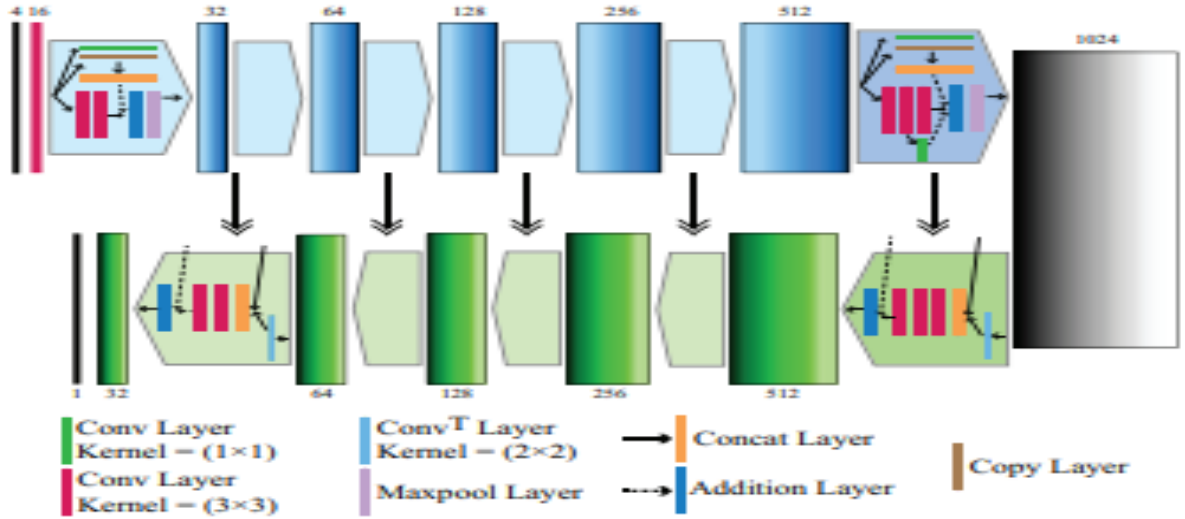


Fig. 1. Cloud-Net architecture. Conv^T, Concat, and Maxpool refer to convolution transposed, concatenation, and maxpooling, respectively. The bars with gradient shading represent the feature maps. The numbers on the top and the bottom of the bars are the corresponding depth of each feature map.

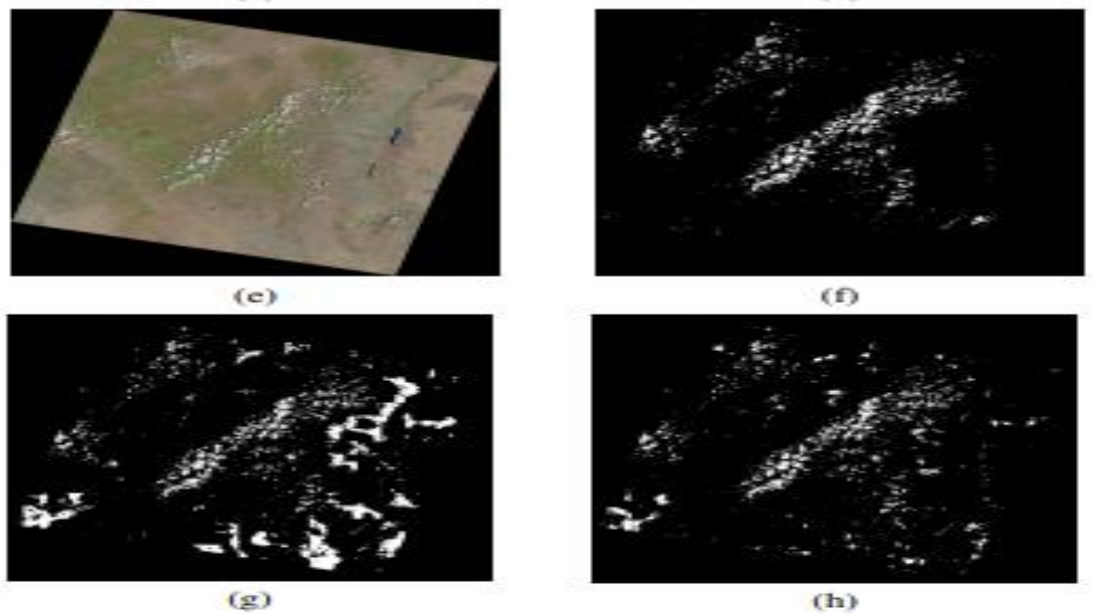


Fig. 2. Cloud-Net visual results: (a),(e) natural color images from 38-Cloud test set, (b),(f) corresponding GTs (c),(g) results of FCN [13] (d),(h) results of Cloud-Net.

Semantic Labeling of Aerial and Satellite Imagery

Sakrapee Paisitkriangkrai, Jamie Sherrah, Pranam Janney, and Anton van den Hengel

Abstract—**Inspired by the recent success of deep convolutional neural networks (CNNs) and feature aggregation in the field of computer vision and machine learning, we propose an effective approach to semantic pixel labeling of aerial and satellite imagery using both CNN features and hand-crafted features. Both CNN and hand-crafted features are applied to dense image patches to produce per-pixel class probabilities. Conditional random fields (CRFs) are applied as a postprocessing step. The CRF infers a labeling that smooths regions while respecting the edges present in the imagery. The combination of these factors leads to a semantic labeling framework which outperforms all existing algorithms on the International Society of Photogrammetry and Remote Sensing (ISPRS) two-dimensional Semantic Labeling Challenge dataset. We advance state-of-the-art results by improving the overall accuracy to 88% on the ISPRS Semantic Labeling Contest. In this paper, we also explore the possibility of applying the proposed framework to other types of data. Our experimental results demonstrate the generalization capability of our approach and its ability to produce accurate results.**

Index Terms—**Aerial imagery, conditional random fields, convolutional neural networks, deep learning, satellite imagery and remote sensing, semantic labeling.**

I. Introduction

AUTOMATED annotation of urban areas from overhead imagery plays an essential role in many photogrammetry and remote sensing applications, e.g., environmental modeling and monitoring, building and updating a geographical database, gathering of military intelligence, infrastructure planning, land cover and change detection. Pixel labeling of aerial photo-

Numerous photogrammetry and remote sensing applications which make use of high-resolution geospatial images, have bee developed as a result of the hardware improvement and faste imaging methods [5]–[14]. Some of these applications includ land use, land cover [15], [16], scene classification [5], coarse grained classification [5], [14], building and tree detection [12 object-class detection [11], oil tank detection [13], object track ing [17], crop classification [8], identification of water-bod types [6], visualization of bridges [10], and anomaly detection [9], [18]. In this paper, we address a problem of semantic pixel labeling of aerial and satellite imagery with a ground sampling distance (GSD) of less than 10 cm.

Semantic labeling is typically applied to multimedia images and involves dense classification followed by smoothing, for example, with a probabilistic graphical model. The traditional visual bag-of-words approach [19] extracts hand-crafted features which are clustered to form visual words, and boosting is used for classification. The success of this method relies on the initial choice of features. More recently, deep convolutional neural networks (CNNs) have been used to learn discriminative image features that are more effective than hand-crafted ones. CNNs have been used for semantic labeling of street scenes in [20].

In this paper, we apply CNNs to overhead imagery. We choose CNN due to the following reasons. First, CNN features can be extracted efficiently. We design our framework such that the en-

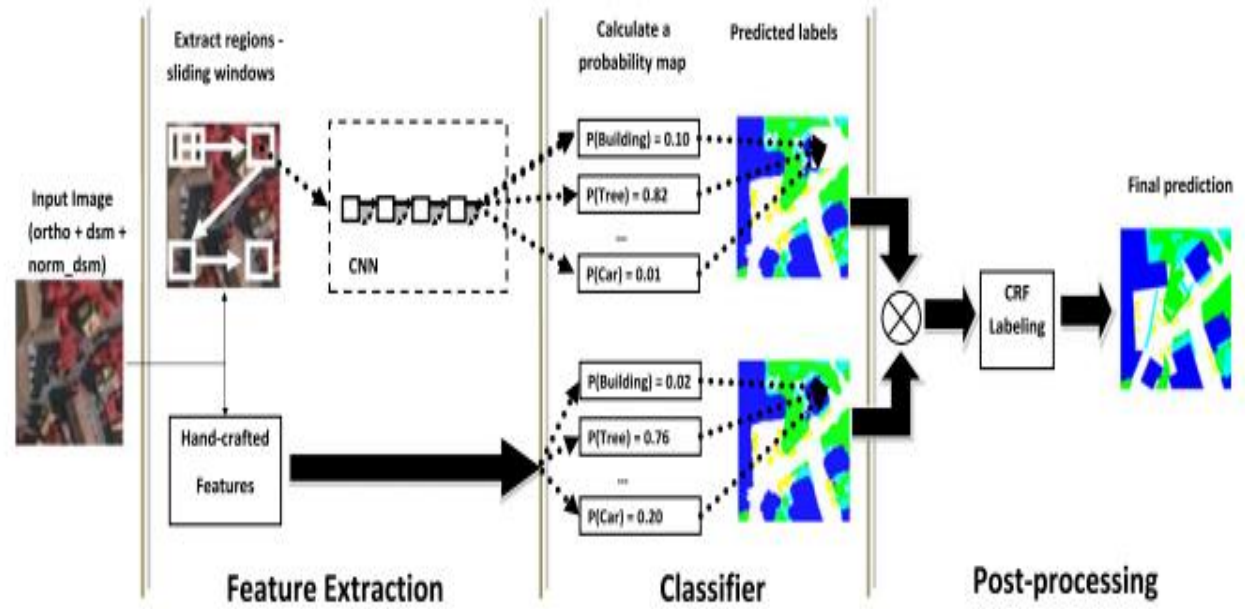


Fig. 1. Overview of the proposed pixels labeling framework.

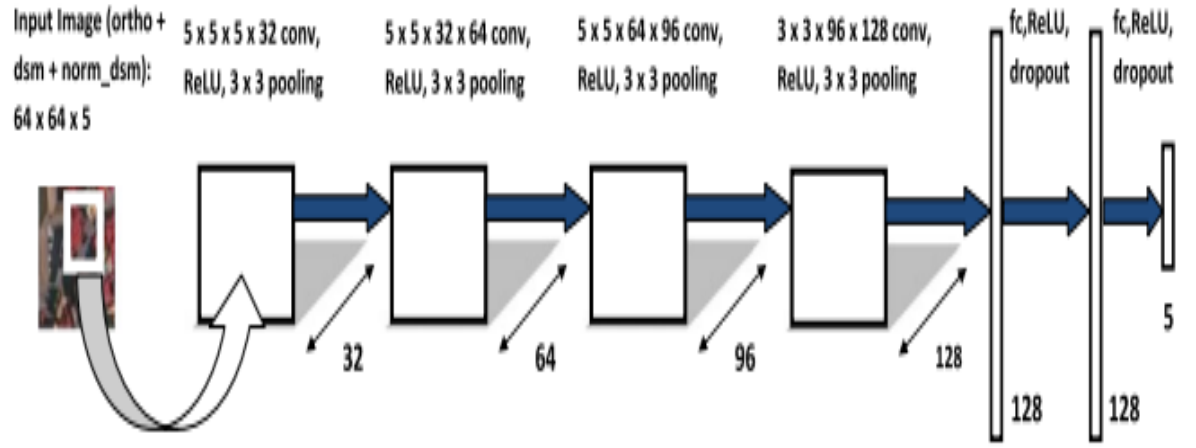
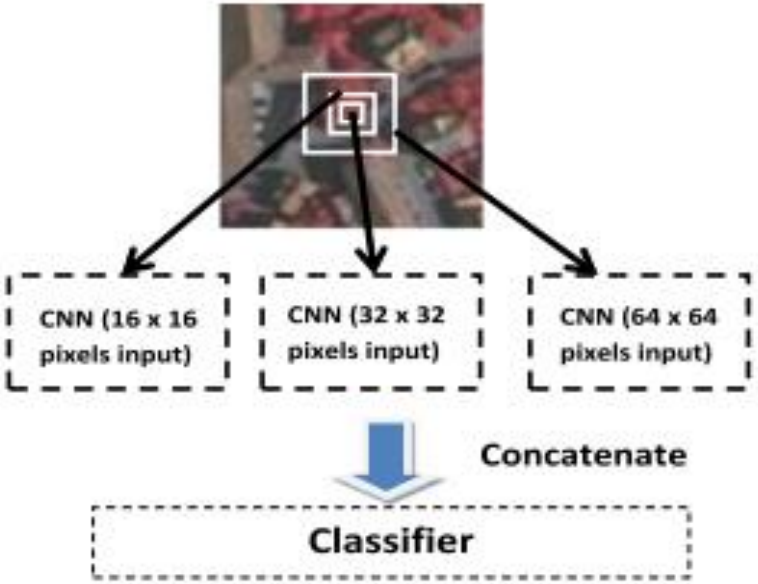


Fig. 3. Illustration of the CNN architecture. In this figure, the networks input is the 64 × 64 × 3-pixels orthophoto, 64 × 64-pixels DSM input image and 64 × 64-pixels normalized DSM image. The networks consist of six layers (four convolutional layers and two fully connected layers) with a final five-way soft-max layer.



Overview of the multiresolution CNN.

C. CRF Labeling

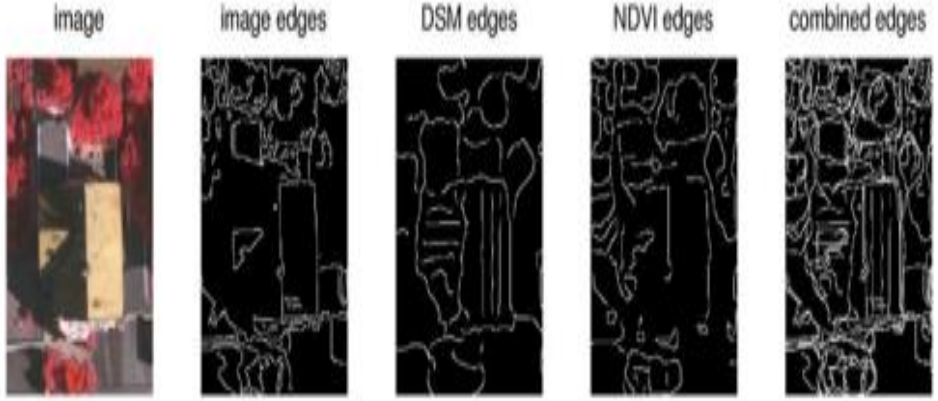
A CRF is a probabilistic graphical model that has been used extensively for semantic labeling of images, for example, see [19], [20]. CRFs are often defined at the super-pixel level rather than the pixel level to improve computational efficiency and robustness [35]. As pointed out in [21], this places an upper limit on the achievable accuracy due to oversegmentation errors (i.e., super-pixels that cover multiple objects). Therefore, we use a pixel-level CRF: a four-connected grid in which each node corresponds to the class label of an image pixel.

Following the standard definition of image labeling CRFs, the energy function consists of unary and pairwise cost terms

$$E = \sum_{i \in \mathcal{V}} \Phi(c_i, \boldsymbol{x}) + \sum_{i,j \in \mathcal{E}} \Psi(c_i, c_j, \boldsymbol{x}) \tag{5}$$

where $\mathcal{V}$ and $\mathcal{E}$ are the nodes and edges of the CRF graph, c_i is the class label of node i and $\boldsymbol{x}$ represents the given data. The unary cost is based on the class probability from the combined CNN and RF classifiers,

$$\Phi(c_i, \boldsymbol{x}) = -\log p_{c_i}^{\text{combo}}. \tag{6}$$



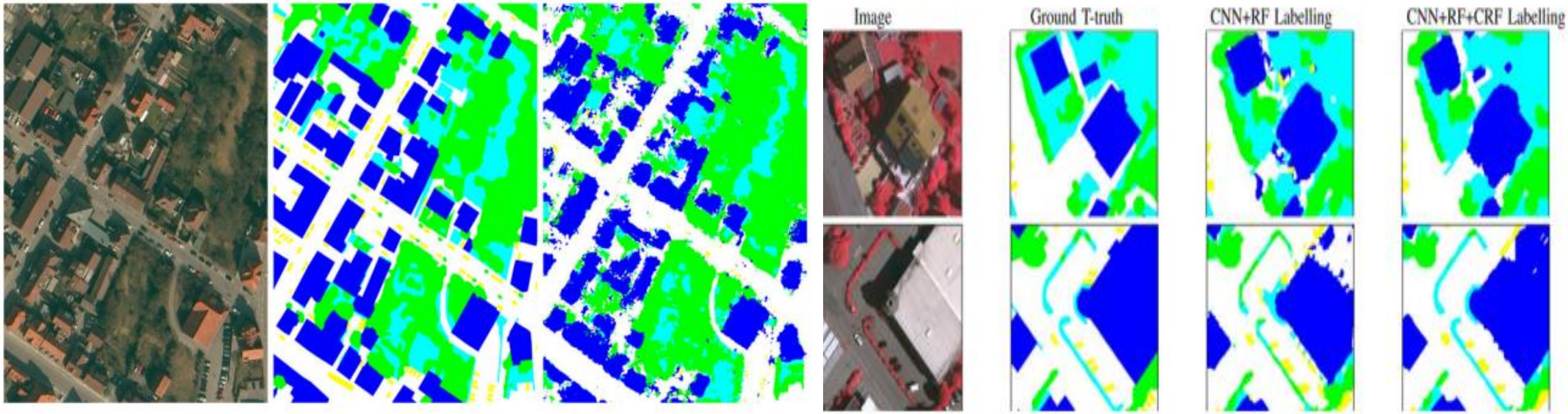


Fig. 13. Classification results on RGB satellite imagery. Left: Satellite imagery, Middle: Ground-truth, Right: Our results.

A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully convolutional neural networks

Chunping Qiu[a], Michael Schmitt[a], Christian Geiß[b], Tzu-Hsin Karen Chen[c], Xiao Xiang Zhu[a,d,*]

[a] Signal Processing in Earth Observation (SIPEO), Technical University of Munich (TUM), Arcisstr. 21, 80333 Munich, Germany
[b] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany
[c] Department of Environmental Science, Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark
[d] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany

ARTICLE INFO

Keywords:
Built-up area
Convolutional neural networks
Human settlement extent
Sentinel-2
Urbanization

ABSTRACT

Human settlement extent (HSE) information is a valuable indicator of world-wide urbanization as well as the resulting human pressure on the natural environment. Therefore, mapping HSE is critical for various environmental issues at local, regional, and even global scales. This paper presents a deep-learning-based framework to automatically map HSE from multi-spectral Sentinel-2 data using regionally available geo-products as training labels. A straightforward, simple, yet effective fully convolutional network-based architecture, Sen2HSE, is implemented as an example for semantic segmentation within the framework. The framework is validated against both manually labelled checking points distributed evenly over the test areas, and the OpenStreetMap building layer. The HSE mapping results were extensively compared to several baseline products in order to thoroughly evaluate the effectiveness of the proposed HSE mapping framework. The HSE mapping power is consistently demonstrated over 10 representative areas across the world. We also present one regional-scale and one country-wide HSE mapping example from our framework to show the potential for upscaling. The results of this study contribute to the generalization of the applicability of CNN-based approaches for large-scale urban mapping to cases where no up-to-date and accurate ground truth is available, as well as the subsequent monitor of global urbanization.

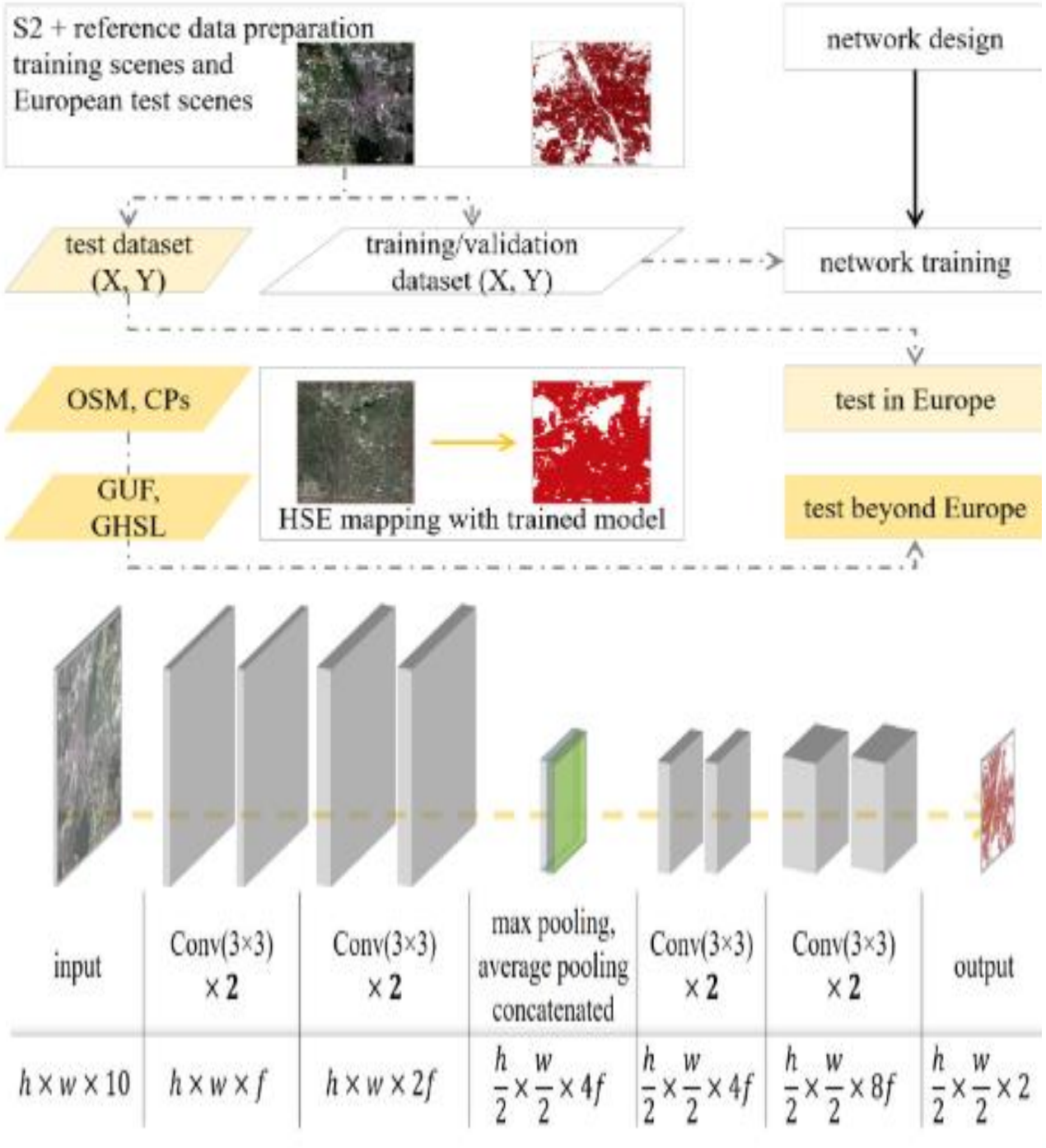
2. HSE mapping with Sen2HSE-Net

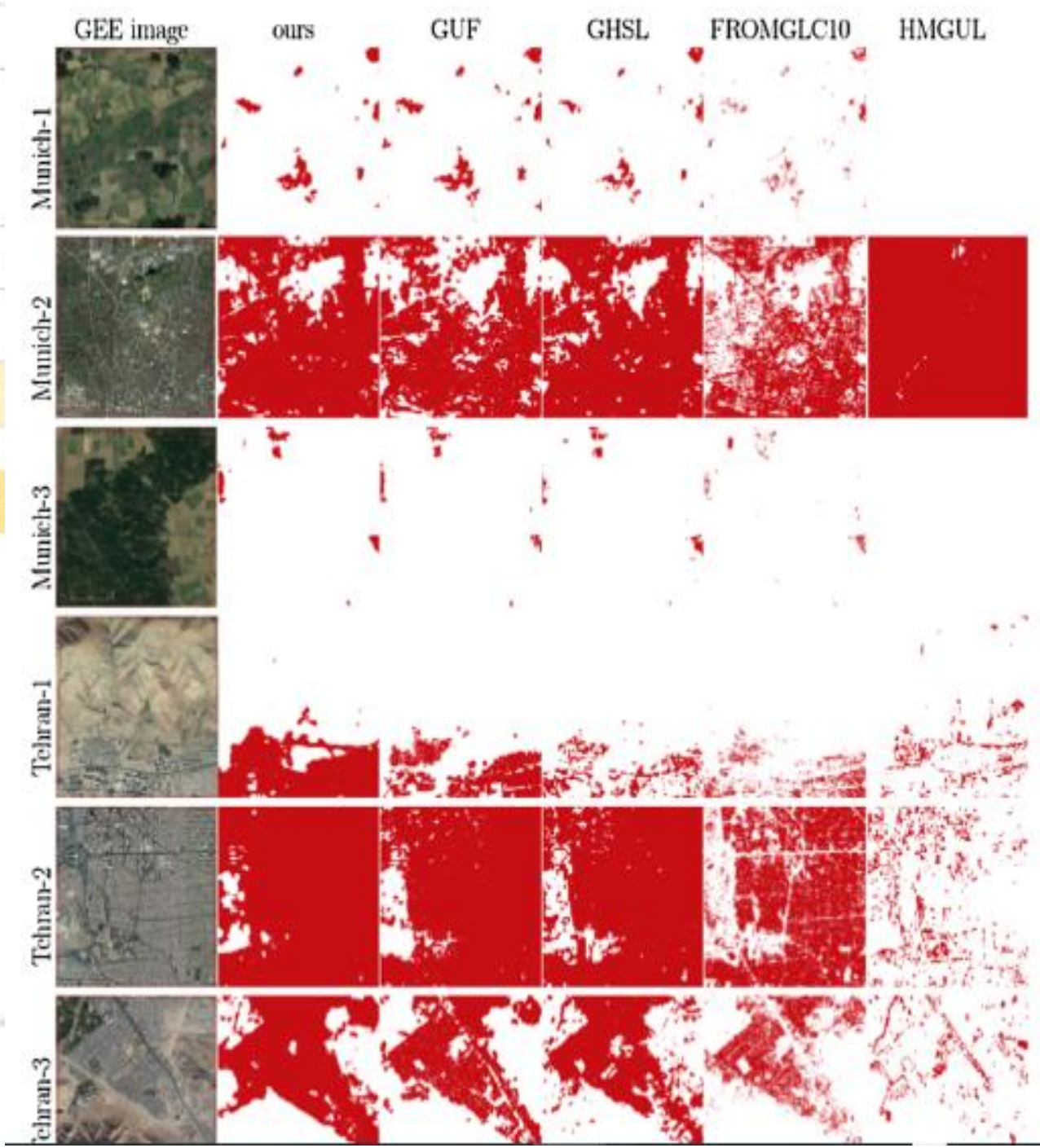
Considering the spatial resolution of available reference data (20 m), the sub-pixel geolocation accuracy of Sentinel-2 data (Drusch et al., 2012), as well as the resolution of existing related products (mostly lower than 20 m), the specific goal of HSE mapping in this study is to detect whether buildings, roads, or other man-made structures are presented—that is, larger than 0% in a 20 × 20 cell. Using this definition, the resulting HSE output from Sentinel-2 imagery will be a binary layer in the Universal Transverse Mercator (UTM) coordinate system, with a ground sampling distance (GSD) of 20 m. This definition is also consistent with the 30 m Global Human Built-up and Settlement Extent (HBASE) dataset derived from Landsat, which consists of human settlement, built-up areas, and roads (Wang et al., 2017).

The procedure used in the proposed HSE mapping framework is illustrated in Fig. 1, which consists of image and reference data preparation, deep neural segmentation network training, and HSE mapping and assessment. Each step will be detailed in the following subsections.

(a) Sentinel-2 image (b) HSE reference

Fig. 2. The processed Sentinel-2 image of central Munich, Germany, and the reference data.


S2 + reference data preparation
training scenes and
European test scenes
network design
test dataset
(X, Y)
training/validation
dataset (X, Y)
network training
OSM, CPs
GUF,
GHSL
HSE mapping with trained model
test in Europe
test beyond Europe
input
Conv(3×3)
× 2
Conv(3×3)
× 2
max pooling,
average pooling
concatenated
Conv(3×3)
× 2
Conv(3×3)
× 2
output
h × w × 10
h × w × f
h × w × 2f
h/2 × w/2 × 4f
h/2 × w/2 × 4f
h/2 × w/2 × 8f
h/2 × w/2 × 2


GEE image
ours
GUF
GHSL
FROMGLC10
HMGUL
Munich-1
Munich-2
Munich-3
Tehran-1
Tehran-2
Tehran-3

Buildings Detection in VHR SAR Images Using Fully Convolution Neural Networks

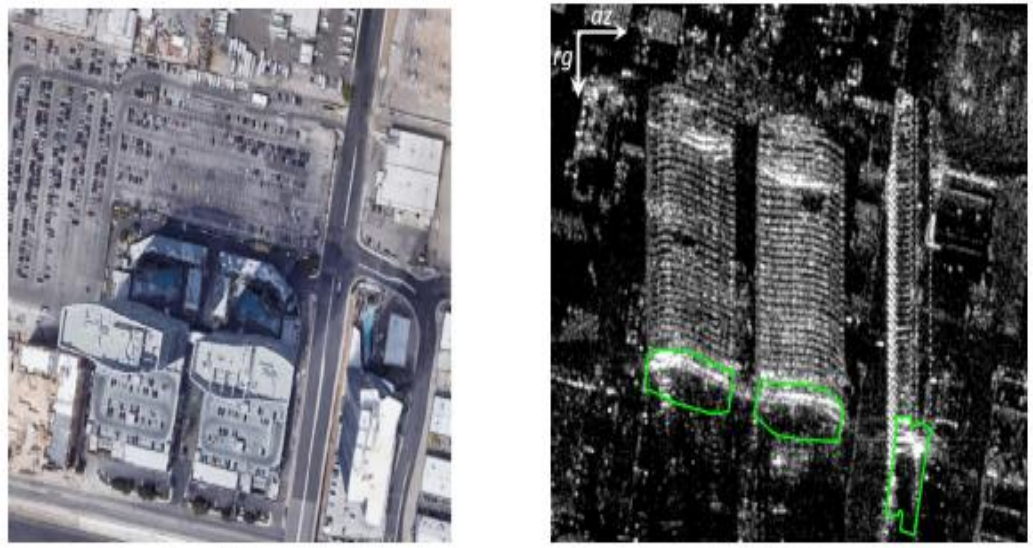
Muhammad Shahzad, *Member, IEEE*, Michael Maurer, Friedrich Fraundorfer, Yuanyuan Wang, *Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*—This paper addresses the highly challenging problem of automatically detecting man-made structures especially buildings in very high-resolution (VHR) synthetic aperture radar (SAR) images. In this context, this paper has two major contributions. First, it presents a novel and generic workflow that initially classifies the spaceborne SAR tomography (TomoSAR) point clouds—generated by processing VHR SAR image stacks using advanced interferometric techniques known as TomoSAR—into buildings and nonbuildings with the aid of auxiliary information (i.e., either using openly available 2-D building footprints or adopting an optical image classification scheme) and later back project the extracted building points onto the SAR imaging coordinates to produce automatic large-scale benchmark labeled (buildings/nonbuildings) SAR data sets. Second, these labeled data sets (i.e., building masks) have been utilized to construct and train the state-of-the-art deep fully convolution neural networks with an additional conditional random field represented as a recurrent neural network to detect building regions in a single VHR SAR image. Such a cascaded formation has been successfully employed in computer vision and remote sensing fields for optical image classification but, to our knowledge, has not been applied to SAR images. The results of the building detection are illustrated and validated over a TerraSAR-X VHR spotlight SAR image covering approximately 39 km^2—almost the whole city of Berlin— with the mean pixel accuracies of around 93.84%.**

I. Introduction

AUTOMATIC detection of man-made objects, in particular buildings from a single very high-resolution (VHR) synthetic aperture radar (SAR) image, is of great practical significance, particularly in applications having stringent temporal restrictions, e.g., emergency responses. However, owing to inherent complexity of SAR images caused by the so-called speckle effect together with radiometric distortions mainly originating due to side-looking geometry, scene interpretation using SAR images is highly challenging. Particularly in urban areas, such distortions render the data to be mainly characterized by multibounce, layover, and shadowing effects consequently giving rise to the need of automatic and robust algorithms for object detection from SAR images.

A variety of algorithms have been published in the literature that aims at the detection and reconstruction of buildings from SAR images. Typically, most of the developed approaches rely on auxiliary information, e.g., the multisensor data provided by the optical [1], [2] and light detection and ranging [3] sensors, geographic information system (GIS) data, e.g., 2-D building footprints [4], multidimensional data, e.g., polarimetric SAR (PolSAR) [5], or multiview/multiaspect data such as



(a) (b)

Fig. 1. Depicting the challenges of SAR image interpretation together with demonstrating the limitations of directly using the 2-D GIS building footprints onto the SAR image. (a) Optical image Google and (b) corresponding SAR image. *rg* and *az* refer to the range and azimuth coordinates, respectively. The three green polygons in (b) are the projections of available 2-D OSM building footprints depicted from top view in (a) onto the SAR image. It can be seen that when the illuminated scene contains elevated objects such as buildings, the so-called "layover" phenomenon (i.e., the superposition of multiple reflection sources in one pixel) occurs as a result of strong reflection of the façade in the SAR image which not only limits the direct usage of 2-D footprint projections for annotation/labeling but also makes the SAR image interpretation of urban areas highly challenging.

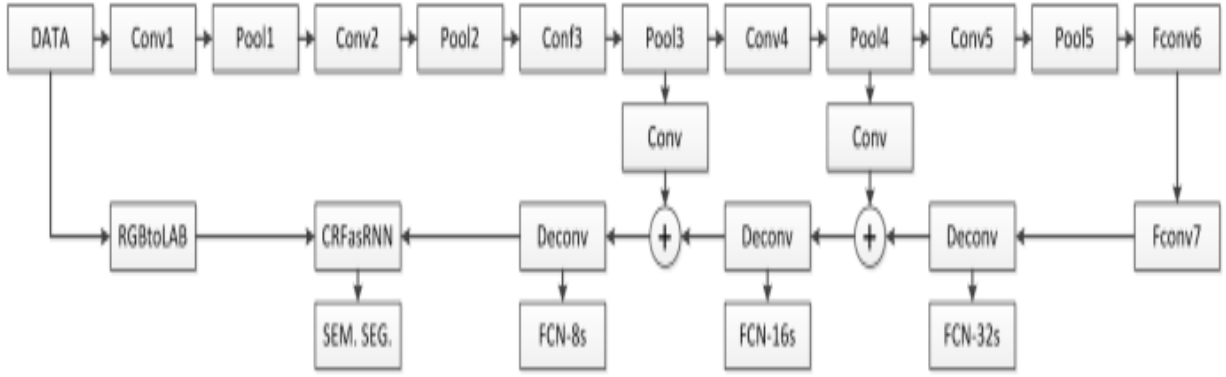


Fig. 5. Overview of the semantic segmentation network. The first part of our network calculates a feature for each input pixel by exploiting an FCN with in-network upsampling and skip-and-fuse architecture to fuse coarse, semantic, and local, appearance information. The second part of the network adds binary potentials (i.e., adding constraints to give neighboring pixels with a similar intensity the same label) by using the dense CRF-RNN as proposed in [32].

Multisource Image Fusion

- Different sources may provide different views of same phenomena
- None of them may be individually accurate, but they can reinforce each other
- Early approaches:

  - extract information from high-spatial-resolution image and inject them into an upsampled low-spatial-resolution image

  - formulate image fusion tasks as optimization problems on various structured models, such as low-rank, sparse, variational, and nonlocal modeling
- Deep Learning: represent complex data transformations from input images to target images (end-to-end system)
- Image-to-image mapping: Neural network takes image as input and produces another image as output!

Multisource Image Fusion

- Multispectral imaging captures image data within specific wavelength ranges across the electromagnetic spectrum
- Hyperspectral imaging, collects and processes information from across the electromagnetic spectrum over hundreds of contiguous spectral bands
- Goal: to obtain the spectrum for each pixel in the image
- Multiband Image Fusion: Hyperspectral (HS) and multispectral (MS) data fusion is a typical example of multiband image fusion, framed as optimization problem by using hand-crafted priors.

Remote Sensing Image Fusion With Deep Convolutional Neural Network

Zhenfeng Shao, *Member, IEEE* and Jiajun Cai

***Abstract*—Remote sensing images with different spatial and spectral resolution, such as panchromatic (PAN) images and multispectral (MS) images, can be captured by many earth-observing satellites. Normally, PAN images possess high spatial resolution but low spectral resolution, while MS images have high spectral resolution with low spatial resolution. In order to integrate spatial and spectral information contained in the PAN and MS images, image fusion techniques are commonly adopted to generate remote sensing images at both high spatial and spectral resolution. In this study, based on the deep convolutional neural network, a remote sensing image fusion method that can adequately extract spectral and spatial features from source images is proposed. The major innovation of this study is that the proposed fusion method contains a two branches network with the deeper structure which can capture salient features of the MS and PAN images separately. Besides, the residual learning is adopted in our network to thoroughly study the relationship between the high- and low-resolution MS images. The proposed method mainly consists of two procedures. First, spatial and spectral features are respectively extracted from the MS and PAN images by convolutional layers with different depth. Second, the feature fusion procedure utilizes the extracted features from the former step to yield fused images. By evaluating the performance on the QuickBird and Gaofen-1 images, our proposed method provides better results compared with other classical methods.**

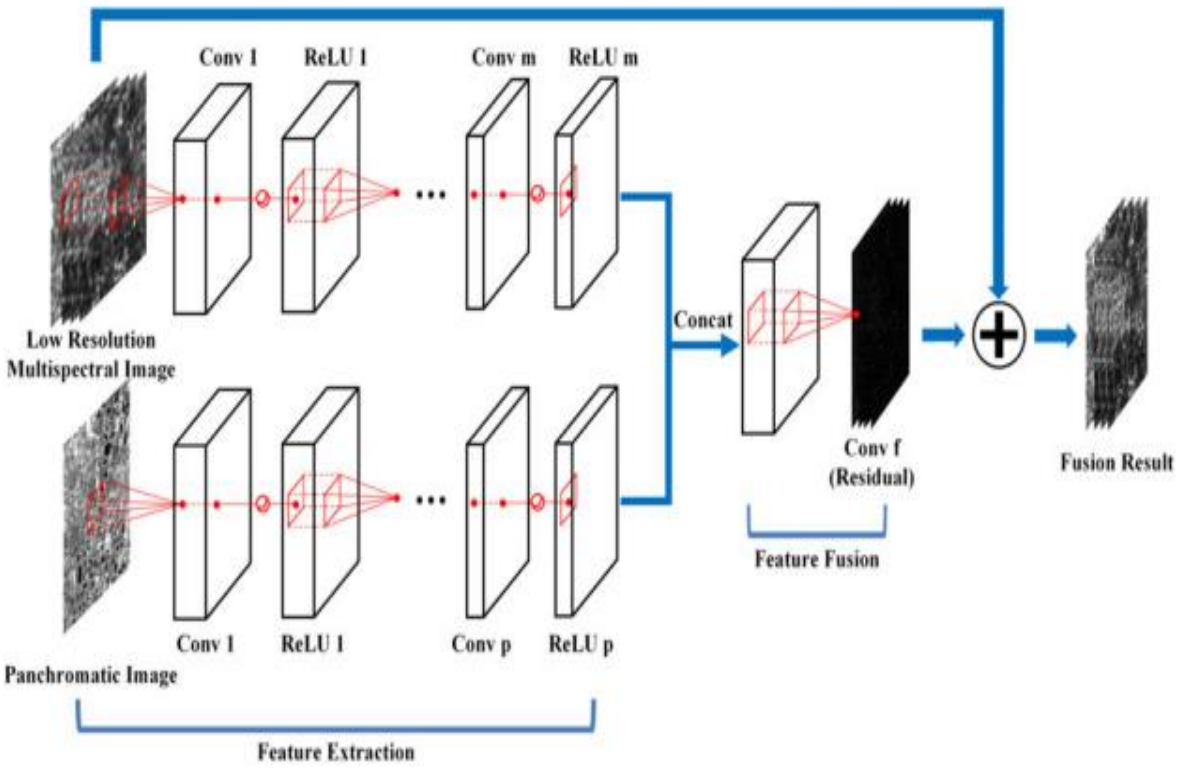
***Index Terms*—Deep convolutional neural network, multispectral image, panchromatic image, remote sensing image fusion.**

earth-observing satellites, such as Landsat, IKON
1, QuickBird, can simultaneously shoot a panchro
and a multispectral image within the same coverage
the reflectance value varies by land covers and sp
multispectral (MS) images can record more informa
surfaces than panchromatic (PAN) images. Howev
signal-noise ratio and tradeoffs of sensors, the sp
tion of MS images is usually lower than that of P
Image fusion is thus designed to make the most o
spectral information by fusing coregistered PAN
MS images. Ideally, the fused images should hav
resolution of PAN images and preserve the spectral
of MS images.

Many remote sensing image fusion methods ha
posed in recent years and can be classified into
component substitution, multiresolution analysis,
representation.

The basic idea of component substitution is to tr
images into another space and replace the principa
by PAN images before inversely transforming the v
to the original domain. This kind of methods is exte
because of their high computation efficiency. Con
implementations of this type of methods include ir

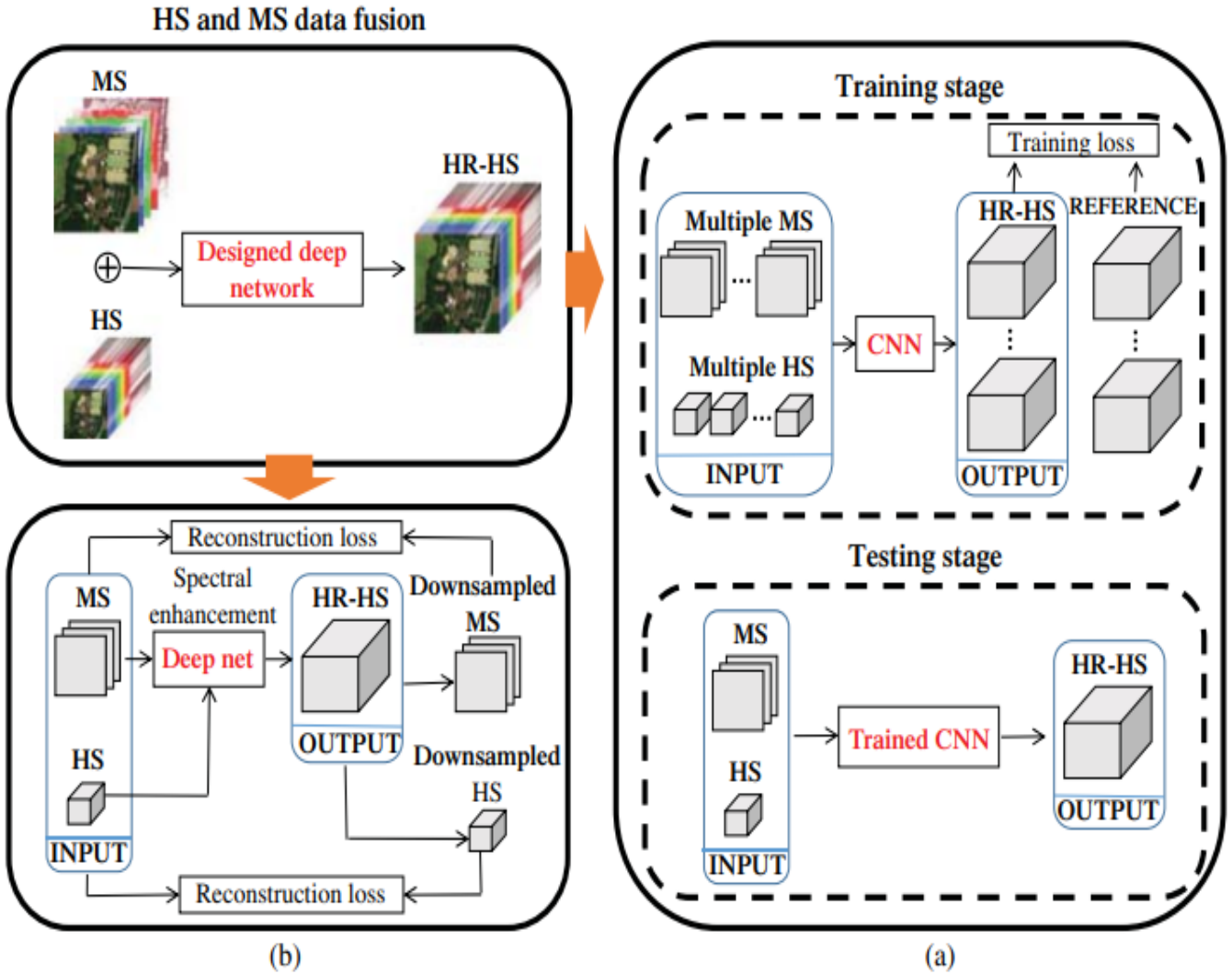
Pansharpening is a process of merging high-resolution panchromatic and lower resolution multispectral imagery to create a single high-resolution color image. Google Maps and nearly every map creating company use this technique to increase image quality. Pansharpening produces a high-resolution color image from three, four or more low-resolution multispectral satellite bands plus a corresponding high-resolution panchromatic band



Architecture of RSIFNN for remote sensing image fusion.

Multispectral and Hyperspectral Image Fusion by MS/HS Fusion Net

Qi Xie[1], Minghao Zhou[1], Qian Zhao[1], Deyu Meng[1,*], Wangmeng Zuo[2], Zongben Xu[1]
[1]Xi'an Jiaotong University; [2]Harbin Institute of Technology
xq.liwu@stu.xjtu.edu.cn woshizhouminghao@stu.xjtu.edu.cn timmy.zhaoqian@gmail.com
dymeng@mail.xjtu.edu.cn wmzuo@hit.edu.cn zbxu@mail.xjtu.edu.cn



$$\mathbf{Y} = \mathbf{X}\mathbf{R} + \mathbf{N}_y, \tag{10.1}$$

$$\mathbf{Z} = \mathbf{C}\mathbf{X} + \mathbf{N}_z, \tag{10.2}$$

where $\mathbf{Y}$ is the observed MS image, $\mathbf{X}$ is the HR-HS image, $\mathbf{R}$ is the spectral response of the multispectral sensor, $\mathbf{Z}$ is the observed HS image, $\mathbf{C}$ is the linear operator that is composed of a cyclic convolution operator and a downsampling operator, and $\mathbf{N}_y$ and $\mathbf{N}_z$ represent noise present in the MS and HS images, respectively. MHF-net formulates a

Abstract

Hyperspectral imaging can help better understand the characteristics of different materials, compared with traditional image systems. However, only high-resolution multispectral (HrMS) and low-resolution hyperspectral (LrHS) images can generally be captured at video rate in practice. In this paper, we propose a model-based deep learning approach for merging an HrMS and LrHS images to generate a high-resolution hyperspectral (HrHS) image. In specific, we construct a novel MS/HS fusion model which takes the observation models of low-resolution images and the low-rankness knowledge along the spectral mode of HrHS image into consideration. Then we design an iterative algorithm to solve the model by exploiting the proximal gradient method. And then, by unfolding the designed algorithm, we construct a deep network, called MS/HS Fusion Net, with learning the proximal operators and model parameters by convolutional neural networks. Experimental results on simulated and real data substantiate the superiority of our method both visually and quantitatively as compared with state-of-the-art methods along this line of research.

Hyperspectral Image Super-Resolution with Optimized RGB Guidance

Ying Fu[1] Tao Zhang[1] Yinqiang Zheng[2] Debing Zhang[3] Hua Huang[1]
[1]Beijing Institute of Technology [2]National Institute of Informatics [3]DeepGlint
{fuying,tzhang,huahuang}@bit.edu.cn yqzheng@nii.ac.jp debingzhang@deepglint.com

Abstract

To overcome the limitations of existing hyperspectral cameras on spatial/temporal resolution, fusing a low resolution hyperspectral image (HSI) with a high resolution RGB (or multispectral) image into a high resolution HSI has been prevalent. Previous methods for this fusion task usually employ hand-crafted priors to model the underlying structure of the latent high resolution HSI, and the effect of the camera spectral response (CSR) of the RGB camera on super-resolution accuracy has rarely been investigated. In this paper, we first present a simple and efficient convolutional neural network (CNN) based method for HSI super-resolution in an unsupervised way, without any prior training. Later, we append a CSR optimization layer onto the HSI super-resolution network, either to automatically select the best CSR in a given CSR dataset, or to design the optimal CSR under some physical restrictions. Experimental results show our method outperforms the state-of-the-arts, and the CSR optimization can further boost the accuracy of HSI super-resolution.

hybrid camera system [1, 2, 3, 12, 14, 24, 28, 30, 31, 45] employ various hand-crafted priors to model the underlying structure of the latent high resolution HSI. Nevertheless, to hammer out proper priors for a specific scene remains to be an art.

Recent alternative approaches [13, 35] leverage on deep learning to alleviate the dependence on hand-crafted priors, and show that the CNN scheme can effectively exploit the intrinsic characteristics of HSIs. Nevertheless, these methods either use the CNN scheme to refine the initialized results in a supervised way [13], or resort to step-by-step alternating optimization [35]. In this work, we present a simple and efficient CNN-based end-to-end method for HSI super-resolution with RGB guidance, which can effectively approximate the spectral nonlinear mapping between the RGB and the spectral space, and utilize the spatial consistency. Neither delicate hand-crafted priors nor training data are needed in our method. This allows our method to handle various scenes more easily.

In addition, all these methods mainly focus on RGB-guided HSI super-resolution under a given CSR function of the RGB camera. Recent researches on HSI super-

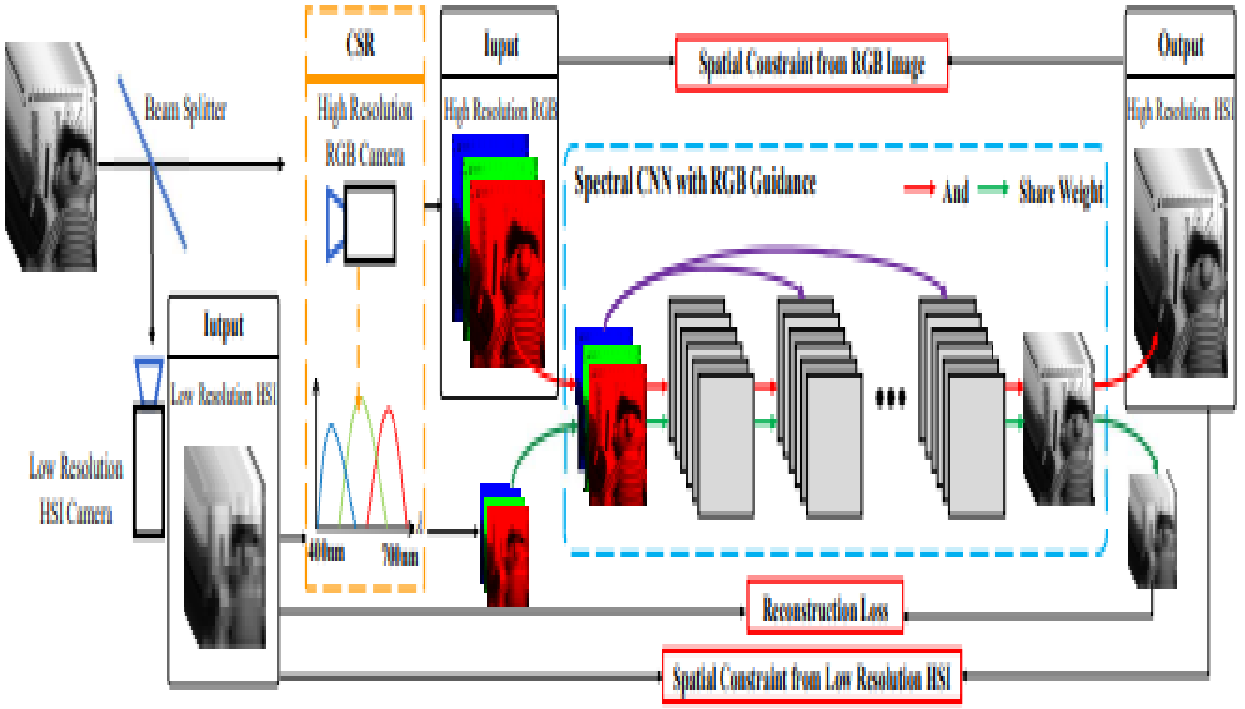


Figure 1. Overview of our CNN-based HSI super-resolution with RGB guidance.

Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL084771

Rainfall Estimation From Ground Radar and TRMM Precipitation Radar Using Hybrid Deep Neural Networks

Haonan Chen[1,2], **V. Chandrasekar**[1], **Haiming Tan**[1], **and Robert Cifelli**[2]

[1]Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA, [2]NOAA/Earth System Research Laboratory, Boulder, CO, USA

Key Points:

- Conventional parametric relationships between radar reflectivity *Z* and rain rate *R* are not sufficient to capture precipitation variabilities
- A hybrid deep neural network system is designed for improved space radar rainfall estimation

Supporting Information:

- Supporting Information S1

Correspondence to:
H. Chen,
haonan.chen@noaa.gov

Citation:
Chen, H., Chandrasekar, V., Tan, H., & Cifelli, R. (2019). Rainfall estimation from ground radar and TRMM Precipitation Radar using hybrid deep neural networks. *Geophysical Research Letters*, *46*, 10,669–10,678. https://doi.org/10.1029/2019GL084771

Received 1 AUG 2019
Accepted 28 AUG 2019
Accepted article online 30 AUG 2019
Published online 11 SEP 2019

Abstract Remote sensing of precipitation is critical for regional, continental, and global water and climate research. This study develops a deep learning mechanism to link between point-wise rain gauge measurements, ground-based, and spaceborne radar reflectivity observations. Two neural network models are designed to construct a hybrid rainfall system, where the ground radar is used to bridge the scale gaps between rain gauge and satellite. The first model is trained for ground radar using rain gauge data as target labels, whereas the second model is for spaceborne Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) using ground radar estimates as training labels. Data from 1 year of observations in Florida during 2009 are utilized to illustrate the application of this hybrid rainfall system. Validation using independent data in 2009, as well as 2-year comparison against the standard PR products, demonstrates the promising performance and generality of this innovative rainfall algorithm.

Plain Language Summary The Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) was the first spaceborne active sensor for observing precipitation over the tropics and subtropics. During its 17 years (1997–2014) in orbit and beyond, PR has been an important tool to characterize tropical precipitation microphysics and quantify rainfall rate over the globe. Ground validation is a critical component in the development of TRMM products. However, the ground-based sensors have different characteristics from PR in terms of resolution, viewing angle, and uncertainties in the sensing environments, which are not taken into account in the operational parametric rainfall relations applied to PR measurements. This study develops a nonparametric machine learning technique for PR rainfall estimation. In the regions where substantial gauge and ground radar data are available, this approach can produce better rainfall estimates compared to the standard PR algorithm. In areas such as ocean and remote regions where no gauge or radar available, the proposed rainfall algorithm is easy to implement, and it can still produce reasonable estimates. With more and more gauges and radars being deployed and many of them become operational, this algorithm can be trained at different locations represented by different

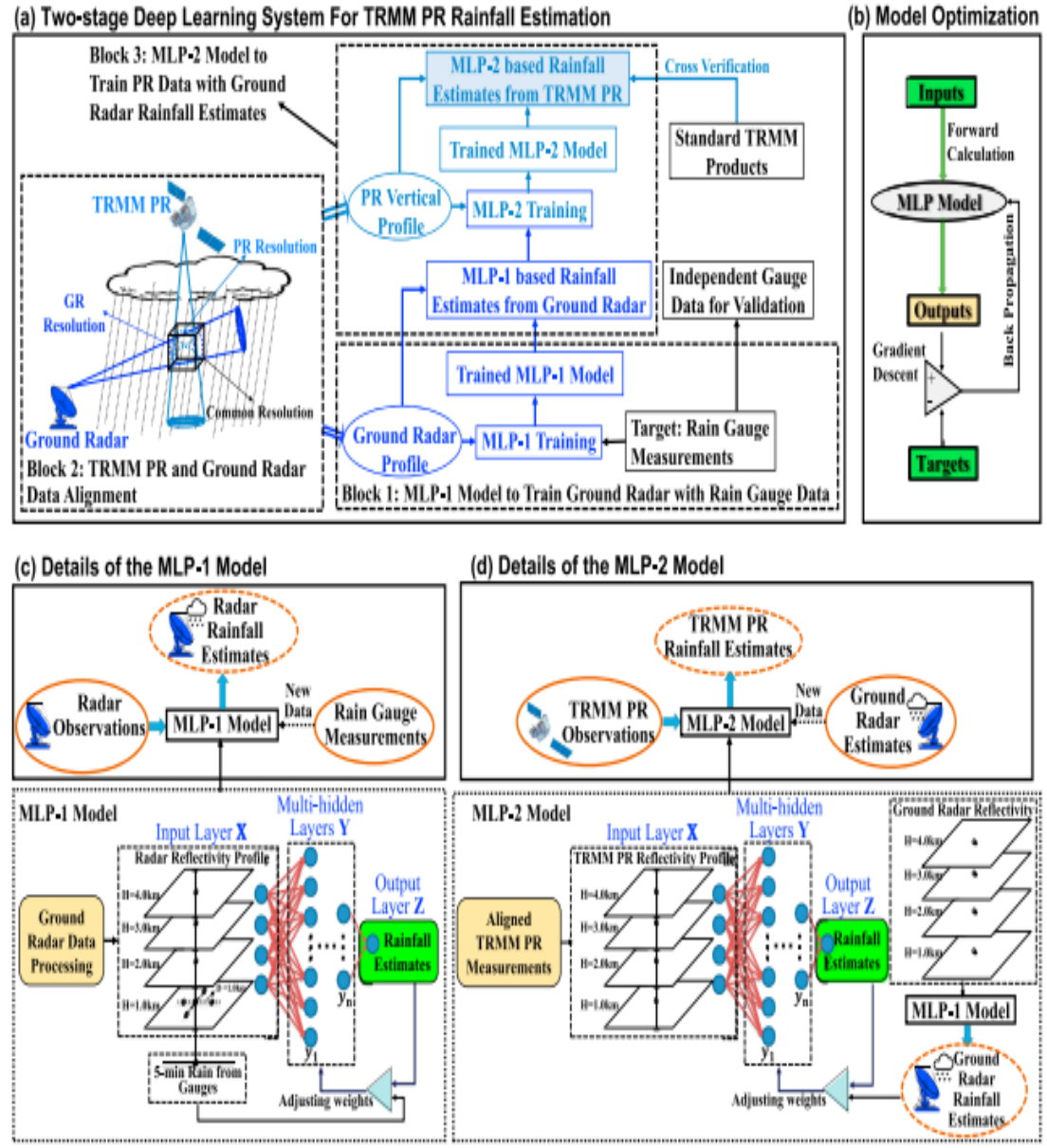


Figure 1. Two-stage deep learning system for Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) rainfall estimation: (a) overall system diagram; Block 1 shows the conceptual diagram of the MLP-1 model designed for ground radar using rain gauge as target labels; Block 2 illustrates the geometry and alignment between ground-based and spaceborne radar measurements; Block 3 sketches the MLP-2 model for PR using ground radar rainfall estimates (from MLP-1) as target labels; (b) MLP model optimization for a predefined hyperparameter; (c) details of the MLP-1 model; (d) details of the MLP-2 model.

A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations

Haonan Chen, *Member, IEEE*, V. Chandrasekar, *Fellow, IEEE*, Robert Cifelli, and Pingping Xie

***Abstract*—Space-based precipitation products are often used for regional and/or global hydrologic modeling and climate studies. A number of precipitation products at multiple space and time scales have been developed based on satellite observations. However, their accuracy is limited due to the restrictions on spatiotemporal sampling of the satellite sensors and the applied parametric retrieval algorithms. Similarly, a ground-based weather radar is widely used for quantitative precipitation estimation (QPE), especially after the implementation of dual-polarization capability and urban scale deployment of high-resolution X-band radar networks. Ground-based radars are often used for the validation of various spaceborne measurements and products. This article introduces a novel machine learning-based data fusion framework to improve the satellite-based precipitation retrievals by incorporating dual-polarization measurements from a ground radar network. The prototype architecture of this fusion system is detailed. In particular, a deep learning multi-layer perceptron (MLP) model is designed to produce the rainfall estimates using the geostationary satellite infrared (IR) data and low earth orbit satellite passive microwave (PMW)-based retrievals as inputs. The high-quality rainfall products from the ground radar network are used as the target labels to train this MLP model. An urban scale demonstration study over the Dallas–Fort Worth (DFW) metroplex is presented. In addition, the Climate Prediction Center morphing technique (i.e., CMORPH) is adopted for preprocessing of the satellite observations. Rainfall products from this deep learning system are evaluated using the standard CMORPH products. The results show that the proposed data fusion framework can be used for generating accurate precipitation estimates and could be considered as an alternative tool for developing future satellite retrieval algorithms.**

***Index Terms*—CMORPH, Dallas–Fort Worth (DFW), deep learning, dual polarization, quantitative precipitation estimation (QPE), radar network, satellite observations.**

I. Introduction

PRECIPITATION plays a key role in understanding the global, continental, and regional water cycles. Accurate precipitation measurements or estimates are vital in various climate, hydrologic, and weather forecast models. Therefore, a large infrastructure has been built around the world over a period of time to measure precipitation and its space–time distributions. Typical instruments include rain gauges that can directly measure rainfall, and remote sensors, such as weather radars and satellites, can indirectly estimate precipitation.

Rain gauges have traditionally been used for precipitation estimation. However, a large number of rain gauges must be deployed in order to capture the complex spatial variabilities of precipitation since gauges only provide pointwise measurements. In the real world, this is neither possible nor necessary due to the arduous nature of deployment and maintenance. A recent study by Kidd *et al.* [1] concluded that the total area measured globally by all currently available rain gauges was surprisingly small, equivalent to less than half a football field.

Compared with rain gauges, satellites have better coverage over the globe, especially over the ocean and polar regions. A number of quasi-global satellite precipitation products at different temporal and spatial resolutions have been

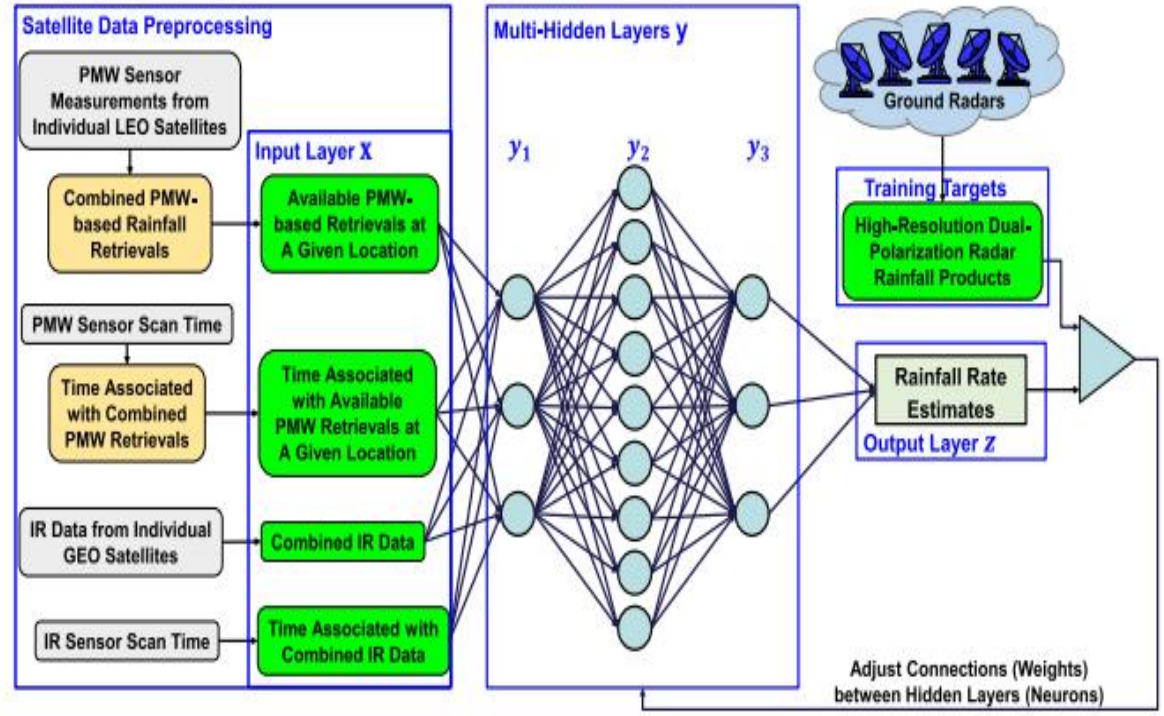
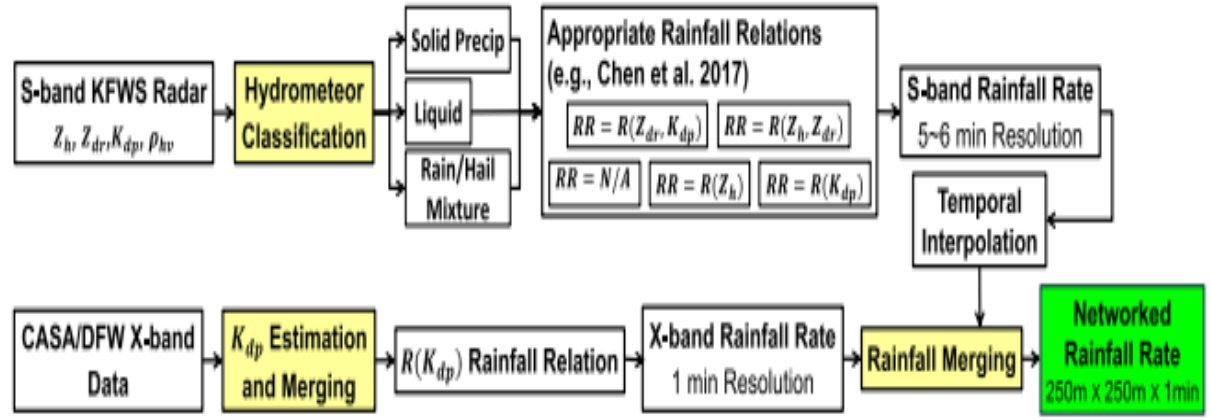


Fig. 2. Architecture of the deep learning model for satellite-based precipitation estimation using ground radar observations as references.



g. 5. Dual-polarization rainfall system for DFW urban radar network. Adapted from [26, Fig. 5].

Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL083015

Using Satellite-Based Vegetation Cover as Indicator of Groundwater Storage in Natural Vegetation Areas

Soumendra N. Bhanja[1,2], Pragnaditya Malakar[1], Abhijit Mukherjee[1], Matthew Rodell[3], Pabitra Mitra[4], and Sudeshna Sarkar[4]

[1]Department of Geology and Geophysics, Indian Institute of Technology Kharagpur, Kharagpur, India, [2]Faculty of Science and Technology, Athabasca University, Athabasca, Alberta, Canada, [3]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [4]Department of Computer Science & Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India

Key Points:
- Widely available, satellite-based vegetation index may be used as a suitable indicator of groundwater storage
- Artificial intelligence estimates show good performance of vegetation index on predicting future groundwater levels
- Vegetation index can be used as an indicator of groundwater storage particularly at natural vegetation-covered areas

Supporting Information:
- Supporting Information S1

Correspondence to:
S. N. Bhanja and A. Mukherjee,
soumendrabhanja@gmail.com;
amukh2@gmail.com

Citation:
Bhanja, S. N., Malakar, P., Mukherjee, A., Rodell, M., Mitra, P., & Sarkar, S. (2019). Using satellite-based vegetation cover as indicator of groundwater storage in natural vegetation areas. *Geophysical Research Letters*, *46*, 8082–8092. https://doi.org/10.1029/2019GL083015

Abstract Normalized Difference Vegetation Index (NDVI) is widely used as an efficient indicator of vegetation cover. Here we assess the possibility of using NDVI as an indicator of groundwater storage. We used groundwater level (GWL) obtained from in situ groundwater observation wells (n > 15,000) in India in 2005–2013. Good correlation (r > 0.6) is observed between NDVI and GWL in natural vegetation-covered areas, that is, forest lands, shrubs, and grasslands. We apply artificial neural network and support vector machine approaches to investigate the relationship between GWL and NDVI using both of the parameters as input. Artificial neural network- and support vector machine-simulated GWL matches very well with observed GWL, particularly in naturally vegetated areas. Thus, we interpret that NDVI may be used as a suitable indicator of groundwater storage conditions in certain areas where the water table is shallow and the vegetation is natural and where in situ groundwater observations are not available.

Plain Language Summary Long-term groundwater resources monitoring is costly at large scales. Satellite-based estimations based on Gravity Recovery and Climate Experiment satellite observations can provide groundwater resource information at coarse spatial and temporal resolution. In this study, we used widely available, high-resolution vegetation index data and investigated the possibility of using it as a proxy of groundwater storage. Artificial intelligence-based approaches that incorporates vegetation index data show good performance in estimating groundwater levels. The results are particularly encouraging in natural vegetation covered areas.

Received 26 MAR 2019

NDVI = (NIR-RED)/(NIR+RED) Red and NIR stand for the spectral reflectance measurements acquired in the red (visible) and near-infrared regions